Ústav aplikovanej informatiky
Fakulta informatiky a
informačných technológií
Slovenská technická univerzita
v Bratislave

Vladimír Kvasnička

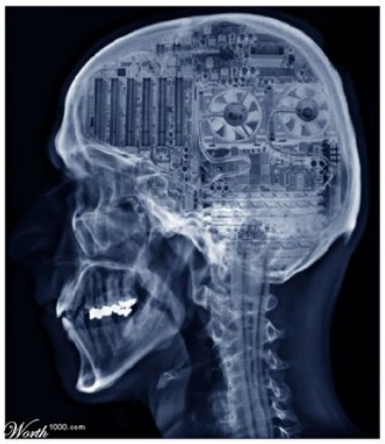
Kognitívna veda pre informatikov

Dočasný učebný text
2011

Obsah

1. prednáška - Úvod do kognitívnej vedy
2. prednáška - Logické neuróny McCullocha a Pittsa
3. prednáška - Darwinovská evolúcia ako algoritmus
4. prednáška - Úloha učenia v evolúcii a memetika
5. prednáška - Učenie s odmenou a trestom a emergencia stratégie hry
6. prednáška - Holografická redukovaná reprezentácia pamäti
7. prednáška - Mentálne modely logiky v kognitívnej vede
8. prednáška - Teória argumentácie
9. prednáška - Revízia poznatkov
10. prednáška - Konštrukcia diagnózy chybného správania sa zložitých systémov
11. prednáška - Kognitívna logika a nemonotónne usudzovanie
12. prednáška - Riadenia zložitých systémov a výpočtové modely okolia
13. prednáška - Symbolická teória Hofstadterovho systému *Copycat* pre tvorbu analógie
14. prednáška - Logický model hry Sudoku
15. prednáška - Teória ohraničenej racionality

Úvod



Tento predbežný učebný text k prednáške „Kognitívna veda“ bol tvorený za pochodu, keď v akad. r. 1997/98 bola zahájená táto výberová prednáška v inžinierskom štúdiu v rozsahu 2 prednášky a 2 hodiny cvičenia. Na webovskej stránke prednášky http://www2.fiit.stuba.sk/ ~kvasnicka/ CognitiveScience každá prednáška má svoju prezentáciu (priesvitky) a riešené príklady na precvičenie preberanej látky z danej prednášky. ***Očakávam spätnú väzbu od študentov tým, že nás upozornia na prípadné chyby, nedôslednosti a preklepy v predkladanom texte (za čšo budú hodnotení pomocnými bodmi, ktoré sa započítajú do celkového záverečného hodnotenia).***

Niekoľko všeobecných poznámok k obsahu predkladaného textu. Niektorým čitateľom sa môže zdať veľmi informaticky orientovaný na umelú inteligenciu, neurónové siete a evolučné algoritmy. Dôvod pre túto orientáciu je veľmi jednoduchý, prednáška je určená pre študentov inžinierskeho (magisterského) štúdia informatiky. Z týchto dôvodov boli zredukované všeobecné „filozufujúce“ partie kognitívnej vedy na minimum. V prezentovanom poňatí je kognitívna veda vlastne totožná s umelou inteligenciou, kde navrhované modely kognitívnych procesov sú mandatórne implementované na neurálnej úrovni (t. j. pomocou neurónových sietí).

U čitateľov (študentov) sa predpokladajú už určité základné vedomosti z matematickej logiky, umelej inteligencie, neurónových sietí a evolučných algoritmov. Pretože posledné dve uvedené oblasti sa prednášajú až v inžinierskom štúdiu ako nepovinné prednášky, snažili sme kapitoly, kde sa využívajú tieto moderné koncepcie (tzv. výpočtovej inteligencie) písať tak, aby boli zrozumiteľné aj pre absolventov bakalárskeho štúdia, ktorí (ešte) neabsolvovali výberové prednášky z týchto dvoch oblastí.

Predbežný učebný text obsahuje 15 prednášok, pričom štandardný semester obsahuje 10-11 týždňov. Z tohto pohľadu sú niektoré kapitoly nadbytočné. Väčší počet sme zvolili preto, aby prednášajúci v spolupráci so študentmi mal možnosť vybrať tie prednášky - kapitoly, ktoré sa javia v danom ročníku najviac zaujímavé. Ďalšia možnosť je použiť neprednesené kapitoly ako predmet prípadových štúdií - esejí, ktorých zaujímavé a tvorivé vypracovanie bude slúžiť ako podklad pre alternatívne absolvovanie skúšky .

August – september 2011

Vlado Kvasnička

1. prednáška

Úvod do kognitívnej vedy

1.1 Úvodné poznámky

Kognitívna veda je charakterizovaná ako teória mysle, ktorá metódami modernej vedy skúma fenomén ľudskej mysli. Používanie moderných, tak teoretických (logika, informatika, teória neurónových sietí, a pod.) ako aj experimentálnych (menovite použitie rôznych typov nukleárnej magnetickej rezonancie k zobrazeniu aktivít mozgu) metód kvantitatívnej psychológie k výskumu kognitívnych aktivít ľudí ju odlišuje od „klasického" špekulatívno-esejistického prístupu, od „ad-hoc"používania fenomenologických pojmov a koncepcií, ktoré nemajú oporu v súčasnej neurovede.

Kognitívna veda má silne interdisciplinárny charakter medzi filozofiou, psychológiou, neurovedou a umelou inteligenciou. Menovite umelá inteligencia má pre kognitívnu vedu zásadný význam, ktorej metódy (pojmy a koncepcie) sa prenášajú do kognitívnej vedy a tvoria je teoretický základ. Môžeme si položiť otázku, aký je rozdiel medzi umelou inteligenciou a kognitívnou vedou. Na túto otázku možno jednoducho odpovedať tak, že modely (prístupy, koncepcie a pod) kognitívnej vedy musia byť kompatibilné s predstavami neurovedy, t. j. musia byť založené na jej základnej paradigme, že vysvetlenie aktivít ľudského mozgu musí byť „zakotvené" na „neurónovom substráte" (obvykle reprezentovanom neurónovými sieťami). Táto vlastnosť „neurónového zakotvenia" sa nepožaduje od prístupov umelej inteligencie, kde je irelevantná podmienka realizácie kognitívnych aktivít ľudského mozgu. Jej snahou je zostrojiť formálny model pre vysvetlenie a modelovanie kognitívnych aktivít (napr. usudzovania, riešenia problémov a pod.), pričom požiadavka ich „neurónovej" realizácie je (nemusí, ale obvykle je) ignorovaná.

Niekoľko poznámok k histórii kognitívnej vedy, ktorá má spoločné počiatky s umelou inteligenciou. Za jednu z prvých prác z kognitívnej vedy a z umelej inteligencie sa pokladá publikácia amerických vedcov Warena McCullocha a Waltera Pittsa z r. 1943 s názvom „A Logical Calculus of the Ideas Immanent in Nervous Activity" v časopise *Bulletin of Mathematical Biophysics.* V tomto pionierskom článku formulovali základy teórie neurónových sietí s použitím binárnych logických neurónov. Dokázali univerzálnu vlastnosť týchto sietí, podľa ktorej každá Boolova funkcia môže byť vyjadrená pomocou neurónovej siete s binárnymi logickými neurónmi.

V súčasnosti je kognitívna veda uznávanou vednou disciplínou, má svoje stavovské organizácie, konferencie, časopisy, na mnohých významných univerzitách sú zavedené programy štúdia tejto vednej disciplíny, existujú univerzitné ústavy kognitívnej vedy. Kognitívna veda spoluexistuje veľmi efektívne s umelou inteligenciou, kognitívnou psychológiou a kognitívnou neurovedou.

1.2 Filozofia a kognitívna veda

Interakcia medzi kognitívnou vedou a filozofiou je veľmi bohatá a plodná, pretože problém mysle je predmetom skúmania filozofie už od čias gréckeho filozofa Aristotela., ktorý v *Etike*

špecifikoval ľudské bytosti ako ***racionálne zvieratá***, pretože má telo s fyzickými potrebami a nutnosťou prijímať potravu a sú racionálnou bytosťou, pretože majú dušu, ktorej časťou je pamäť a schopnosť myslieť a usudzovať (t. j. to, čo môžeme nazvať myseľ). Cieľom tejto bytosti je dosiahnuť šťastie tak, že napĺňa zmysel svojej existencie. Podobne, ako zmyslom existencie noža je, aby krájal a zmyslom existencie očí je, aby videli, tak zmyslom ľudskej existencie – racionálneho zvieraťa je, aby používalo svoju myseľ k usudzovaniu, t. j. správať sa racionálne pri vyhodnocovaní svojich životných skúseností. V tomto momente sa dostávame ku kľúčovému momentu gréckej filozofie, ktorá považovala za jeden z podstatných prejavov racionality ľudských bytostí schopnosť usudzovať pomocou zákonov logiky, ktorej základy boli taktiež vytvorené v starom Grécku.

Francúzský filozof z prvej polovice 17. storočia René Descartes, vychádzajúci z dobových mechanistických predstáv o prírode, formuloval dualistickú teóriu ľudských bytostí. Aj keď príroda, teda aj ľudské telo, je materiálneho charakteru a je teda ovládané základnými princípmi mechaniky, ľudské bytosti sú v tom zvláštne, že sú kombináciou mentálnych substancií (myseľ) a materiálnych substancií (telo). Táto myšlienka tvorí základ Descertesovho ***dualizmu***, ktorý môže byť preformulovaný do súčasného filozofického jazyka tak, že ľudia majú myseľ a telo. Základný problém spočívajúci v tejto paradigme je, akým spôsobom môže interagovať nehmotná myseľ s materiálnym telom.

Materialistický (fyzikalistický) pohľad na tento problém Deascartesovho dualizmu je, že aj myseľ je materiálnej podstaty. Musíme však poznamenať, že aj tento materialistický pohľad na myseľ vyvoláva následné problémy. Fyzikálna myseľ je charakterizovaná špeciálnymi vlastnosti, akými sú intencionalita (zámernosť alebo cielenosť) a vedomie, potom sa môžeme pýtať akým spôsobom vznikli tieto dve vlastnosti ľudskej mysle. K tomuto problému sa ešte vrátime v ďalšej časti tejto prednášky.

Ďalší dôležitý problém filozofie mysle je otázka, čo to je myseľ, ako je odlíšená od ostatných častí človeka, ako je štruktúrovaná a pod. Táto otázka má dva aspekty, jeden pochádza z filozofických snáh riešiť problém vzťahu medzi racionalizmom a empirizmom, ktorý bol veľmi aktuálny v sedemnástom a osemnástom storočí. Racionalizmus a empirizmus sú rôzne pohľady na podstatu a zdroje ľudských vedomostí. Podľa empirikov všetky naše poznatky pochádzajú z našich zmyslov, z empirických alebo experimentálnych interakcií s okolitým svetom. Racionalisti, naopak tvrdia, že existujú určité poznatky, ktoré nie sú odvoditeľné z empírie. Klasický racionalisti, akými sú Descartes alebo Leibniz, ale hlavne Kant tvrdia, že niektoré idey, napr. idea Boha alebo celých prirodzených čísel, nevyplývajú z našej experimentálnej skúsenosti, ale sú vrodenou súčasťou našej mysle. Empirici sa snažili v priebehu ostatných tristo rokov dokázať, ako aj takéto abstraktné idey – poznatky môže vyplývať z našej experimentálnej skúsenosti.

Rôzny pohľad racionalistov a empirikov na zdroje našich poznatkov spôsobil aj ich rôzny pohľad na myseľ. Podľa empirikov myseľ obsahuje len také idey – poznatky, ktoré sú primárne odvodené z našich zmyslov (včítane aj veľmi abstraktných a zložitých idey, akou je idea Boha). Z tejto paradigmy taktiež vyplýva, že myseľ nie je nič iné, ako idey odvoditeľné z pocitov pochádzajúcich z našich zmyslov. Potom kognitívne procesy sú založené na procesoch vnímania, t. j. kognícia a vnímanie sú v podstate identické pojmy, ktoré sa odlišujú len v tom, že kognícia je špecifický sled vnímaní. Podľa škótskeho filozofa Davida Humea, významného predstaviteľa empirizmu a osvietenstva, kognitívne procesy v našej mysli sú riadené asociáciami a podobnosťou. Myseľ je všeobecné zariadenie – orgán, kde všetky typy kognície sú realizované pomocou našej empirickej skúsenosti prostriedkami asociácie a analógie.

Racionalisti, tým, že neakceptujú paradigmu o zdroji našich poznatkov z našich pocitov, musia hľadať alternatívnu odpoveď na otázku, čo je zdrojom našich poznatkov. Tento problém vyriešili tak, že za zdroj našich poznatkov pokladajú vrodenosť za hlavný

zdroj poznatkov a štruktúry našej mysle. Ak idey sú vrodené (a teda nemusia byť odvodená z našej skúsenosti), potom štruktúra našej mysle je vrodená a tak vlastne aj ohraničuje použiteľnosť mysle pri spracovaní našich pocitov. Moderný pohľad na tento problém vrodenosti (nativizmu) je, že len určité schopnosti našej mysle (napr. jazykové schopnosti, inštinkty a pod.) sú nám vrodené a preto sa nemusia učiť. Dôležitý aspekt tohto moderného prístupu k mysli je aj problém jej modularity, t. j. ľudská myseľ je chápaná ako systém navzájom poprepájaných modulov, z ktorých každý vykonáva určité špecifické kognitívne akcie a ktorých štruktúra je vrodená.

Vzťah medzi racionalizmom a empirizmom obsahuje aj ďalší aspekt mysle, a to jej štruktúru a miesto vedomia v nej. Problém vedomia patrí medzi najkomplikovanejšie záležitosti filozofie mysle a taktiež aj medzi tie vlastnosti mysle, ktorá „vzrušuje“ nielen filozofov a psychológov, ale aj informatikov aktívnych v umelej inteligencii a kognitívnej vede. Na čom sa všetci dohodnú v tejto zložitej oblasti je, že vedomie je integrálnou súčasťou mysle, ale taktiež aj na tom, že patrí medzi najzáhadnejšie vlastnosti mozgu. V súčasnej filozofii mysle existuje všeobecne akceptovaný názor, že mentálne stavy vedomia sú integrálnou súčasťou mysle dva krajné názory na vedomie. Každá mentálna (alebo kognitívna) aktivita mozgu je prejavom vedomia alebo je dostupná vedomia (snáď môžeme povedať, že je jeho súčasťou). Ak však akceptujeme „komputačnú“ paradigmu podľa ktorej kognitívne procesy prebiehajúce v mozgu nie sú nič iné ako informačné procesy prebiehajúce v počítači – mysli, tak potom určitá časť týchto informačných procesov nie je dostupná vedomiu (napr. predspracovanie vstupnej vizuálnej informácie zo sietnice oka). Z tohto možno usudzovať, že predstava o tom, že každá mentálna aktivita je dostupná vedomiu je prehnaná, len určité mentálne aktivity sú dostupné nášmu vedomiu.

1.2.1 Myseľ v kognitívnej vede

Vzťah medzi mentálnym a fyzikálnym je centrálnym problémom filozofie mysle. Tento vzťah je aj základným problémom súčasnej kognitívnej vedy. Spôsob riešenia tohto vzťahu v kognitívnej vede sa nazýva funkcionalizmus, podľa ktorého mentálne stavy nie sú ekvivaletné s fyzikálnymi štruktúrami ktoré ich produkujú, ale prejavom-vlastnosťou ich usporiadania a kauzálnych súvislostí medzi nimi. Funkcionalizmus, ktorý vyjadruje určitú predstavu o vzájomnom vzťahu medzi mentálnym a fyzikálnym, je kompatibilný s materializmom (fyzikalizmom), myseľ (mentálne stavy) je prejavom materiálnych – fyzikálnych štruktúr, ktoré sú súčasťou tela a tvoria materiálny substrát (vehikulum) mysle.

Klasický príklad toho, čo nie je identické s tým z čoho je to urobené, sú peniaze. Peniaze nie sú totožné s papierom na ktorom sú vytlačené, ale majú určitú funkcionálnu úlohu v danom ekonomickom systéme. Materiálna reprezentácia peňazí nie je to podstatne, čo ich špecifikuje, či sú alebo nie sú peniaze; podstatné pre ne, je ich nemateriálna - funkcionálna úloha, ktorú hrajú v ekonomických systémoch ako platidlo, ktoré v niektorých prípadoch je reprezentované bankovkami, ale v prípade e-bankingu je reprezentované alfanumerickým reťazcom, ktorý sa prenáša internetom.

Dôležitý dôsledok tohto funkcionálneho pohľadu na myseľ spočíva v tom, že funkcionalistická paradignma umožňuje uvažovať aj nebiologické formy inteligencie. Pretože myseľ je „forma“ a nie „hmota“, ktorá špecifikuje psychické vlastnosti nejakého indivídua. To znamená, že z funkcionalisticej paradigmy vyplýva aspoň teoretická možnosť existencie umelej kreatúry (robot alebo počítač) s mentálnymi stavmi, ak majú tú správnu funkcionálnu štruktúru – organizáciu. Táto idea má ústredné postavenie v tradičnej umelej inteligencii, kde hlavnou víziou tejto informatickej oblasti je implementácia programu s takou algoritmickou štruktúrou, že umožňuje správať sa systému inteligentným spôsobom, ktorý je neodlíšiteľný od správania sa ľudskej bytosti. Tento redukcionistický pohľad na ľudskú inteligenciu je

v súčasnosti kritizovaný mnohými filozofmi a informatikmi, medzi najznámejšieho kritika patrí americký filozof John Searle, ktorý na jednoduchom príklade „čínskej izby“ ukazuje ako taká špičková kognitívna aktivita akou je preklad textu z čínštiny do angličtiny môže byť mechanicky realizovaná človekom, ktorý nevie po čínsky, ale má k dispozícii prekladové slovníky z čínštiny do angličtiny a text mechanicky prekladá znak po znaku, bez uvedomenia, akú činnosť vlastne vykonáva. Sumarizujúc, umelé inteligentné systémy nebudú mať vedomie a preto nebudú úplne rovnocenné ľudskej mysli. .

1.3 Kognitívna veda a psychológia

Psychológia je veda, ktorá sa zaoberá reprezentáciou a spracovaním informácie živočíchmi s vyššou nervovou aktivitou (a teda aj ľuďmi). Tieto sú schopné prijímať informáciu z okolia, vytvárať jej vnútornú reprezentáciu a transformovať ju s cieľom výberu a výkonu akcií. Inteligentné informačné procesy sú vyjadrené schopnosťou získavať a spracovávať informáciu z okolia, s cieľom vykonávania adaptívne akcií, ktoré zvyšujú schopnosť organizmu prežívať a rozmnožovať sa. V tejto súvislosti je dôležitý pojem „adaptívna akcia“, čo znamená, že proces vytvárania akcie je výsledkom učenia organizmu v danom prostredí s cieľom optimalizácie danej akcie. Ústredným problémom študovaným psychológiou sú informačné procesy, ktoré prebiehajú medzi vstupom informácie prostredníctvom zmyslov a výstupom informácie prostredníctvom motorických centier. Zložité formy inteligencie, ktoré sú pozorované u humanoidov, vyžadujú teórie mechanizmov výkonu kognitívnych akcií a vnútornej skúsenosti (existencia pamäti a schopnosti vytvárania vnútorných modelov prostredia). Tieto organizmy sú schopné výkonu zložitých kognitívnych akcií, akými sú rozhodovanie (angl. decision making), riešenie problémov (angl. problem solving) a plánovania (angl. planning); u ľudí pristupuje k týmto aktivitám schopnosť komunikácie prostredníctvom jazyka formálne reprezentovaného generatívnou gramatikou. Ľudia sú schopní najzložitejších a najvšeobecnejších foriem spracovania a transformácie, preto podstatná časť psychológie je orientovaná na pochopenie procesov spracovania informácie a inteligencie u ľudí s cieľom vytvárania kognitívnych architektúr a modelov týchto procesov (vyšších kognitívnych procesov).

Psychológia (menovite tá jej časť, ktorá sa zaoberá kognitívnymi procesmi, t. j. kognitívna psychológia) patrí do jadra kognitívnej vedy. Výskum kognitívnej vedy má vždy bezprostredný alebo sprostredkovaný význam v psychológii, Nie všetky výskumy o inteligentných informačných procesoch sú významné pre psychológiu. Musíme poznamenať, že odborný záber kognitívnej vedy je o mnoho všeobecnejší ako psychológie, preto nie všetky výsledky tohto nového vedného odboru sú relevantné pre psychológiu

Vznik informatiky v 50. rokoch minulého storočia znamenal nový impulz pre rozvoj psychológie, menovite jej odklon od striktného behaviorizmu, ktorý označoval za nevedecké snahy používať v psychológii paradigmu vnútorných mentálnych procesov. Počítače poskytujú jedinečné príklady o tom, že počítač transformuje vstup na výstup zložitou postupnosťou vnútorných procedúr, ktoré postupne transformujú vstupnú informáciu na výstupnú informáciu. Bezprostredne nasledujúci vznik umelej inteligencie znamenal ďalší stimul pre psychológiu, pretože snahou informatikou aktívnych v umelej inteligencii bolo nielen navrhnúť ale aj implementovať algoritmy, ktoré simulujú inteligentné činnosti, ktoré boli pokladané vtedy za výlučnú doménu ľudí (hrať dámu alebo šach, porozumenie prirodzenej reči, určenie diagnózy v medicíne a pod.). Americký vedci (ekonómovia, psychológovia a informatici) Herbert Simon a Allan Newel sa zaslúžili o prienik metód umelej inteligencie do novovznikajúcej kognitívnej psychológie. Vznikla ***komputačná***

hypotéza, podľa ktorej ľudský mozog pracuje ako biologický procesor, ktorý transformuje vstupné symboly pomocou sledu pravidiel typu *ak..., potom...* na výstupné symboly; táto jednoduchá idea je podstatou symbolického prístupu. Skoro súčasne s týmto prístupom vznikol alternatívny prístup nazývaný konekcionizmus (alebo neurónové siete), podľa ktorého ľudský mozog je neurónová sieť, ktorá obsahuje množstvo navzájom poprepájaných elementárnych procesorových jednotiek nazývaných neuróny. Základy tejto paradigmy už boli naformulované v r. 1943 už zmienenou prácou (pozri kapitolu 1.1) Warena McCullocha a Waltera Pittsa. Neurónové siete v súčasnosti patria medzi významnú časť počítačovo orientovanej umelej inteligencie, kde zaujali postavenie univerzálneho matematicko-informatického prístupu k štúdiu a modelovaniu procesov učenia, adaptácie umelých kognitívnych systémov založených na metafore ľudského mozgu.

1.4 Kognitívna neuroveda

„Neuroveda“ študuje nervový systém, jeho vlastnosti, architektúru, fyziológiu a biofyziku neurónov a spojov (synapsií) medzi nimi. Jej špecializovaná časť „kognitívna neuroveda“ študuje mechanizmy realizácie kognitívnych aktivít mozgu. Je to veda o informačných procesoch prebiehajúcich v mozgu: (1) akým spôsobom je táto informácia získavaná z našich zmyslov, (2) ako sú realizované procesy vnímania a rozpoznávania, (3) ako vzniká myslenie a vedomie, (4) ako vznikajú procesy rozhodovania, ktoré zahŕňajú predvídanie budúcich stavov nášho okolia a dôsledkov našej aktivity, (5) čo je to učenie a ako je realizovaná pamäť a konečne, (6) ako sú implementované procesy tvoriace náš jazyk? Tieto činnosti patria do fundamentálneho jadra kognitívnej vedy. Môžeme si položiť otázku, aký je rozdiel medzi kognitívnou vedou a kognitívnou neurovedou? Hľadať odpoveď na túto otázku nie je jednoduché v dôsledku veľkého presahu medzi týmito dvoma „kognitívnymi“ vednými disciplínami. Snáď, pomocou určitého zjednodušeného prístupu k tomuto problému, môžeme povedať, že medzi „kognitívnou neurovedou“ a „kognitívnou vedou“ existuje demrakčná čiara v predmete ich záujmu, predmetom záujmu kognitívnej neurovedy je „hardware“ (t. j. biologická realizácia procesov) a predmetom záujmu „kognitívnej vedy“ je software (t. j. informatická a/alebo algoritmická realizácia procesov). Tu musíme však poznamenať, že toto zakotvenie kognitívnej vedy v biologických vedách (na nervovom substráte) je hlavné kritérium odlíšenia kognitívnej vedy od umelej inteligenciu, ktorá taktiež študuje algoritmickú realizáciu kognitívnych (alebo všeobecnejšie – inteligentných) procesov, ale nie je zakotvená na neurónovom substráte.

Veľmi zaujímavá je história neurovedy (a teda aj kognitívnej neurovedy). Za prvých experimentátorov s ľudským mozgom sú pokladaní Egypťania, ktorí už pred 5000 rokmi vykonávali chirurgické výkony na mozgu. O úspešnosti týchto zákrokov svedčia archeologické nálezy, kde boli nájdené lebky so zahojenou trepanáciou lebečnej kosti. Filozof Aristoteles veril, že centrom našej inteligencie je srdce (podľa neho mozog je orgán na ochladzovanie krvi). Tento názor pretrvával aj v starovekom Ríme, až do čias významného lekára a filozofa Galena (žil v druhom storočí po Kr.), ktorý na základe pozorovaní poranení hlavy gladiátorov usúdil, že sídlom inteligencie (alebo v dobovej terminológii duše) je mozog. K tomuto záveru prišiel pomocou experimentálnej skúsenosti s poraneniami hlavy gladiátorov, ktorí pri špecifických poraneniach strácali schopnosť niektorých kognitívnych akcií (videnia, reči a pod.).

Až počiatkom novoveku v 17. a hlavne v 18. storočí začali rôzni lekári a aj filozofi umiestňovať mentálne aktivity do rôznych častí mmozgu. Ako klasický príklad uvedieme filozofa Descartesa, ktorý umiestnil myseľ/dušu v šišinke mozgovej (hypofýze). Táto snaha lekárov vyvrcholila v polovici 19. storočia, keď francúzky lekár Paul Broca umiestnil centrum

reči do ľavého frontálneho laloku mozgu, čo bolo vykonané na základe jeho pozorovania lokálnej lézie mozgu, ktoré malo za dôsledok stratu reči. Tento prístup hľadať centrum rôznych činnosti mozgu pretrvával až do 30.-40. rokov 20. storočia, keď sa objavili prvé funkčné metódy vyšetrovania mozgu, akými boli elektrické stimulovanie mozgu. Tieto vyšetrovania boli mnohokrát vykonávané ako diagnostické vyšetrenie, pomocou ktorého sa hľadalo centrum ťažkej epilepsie, za účelom jej chirurgického odstránenia. Možno konštatovať, že v tomto pionierskom období boli umiestnené už skoro všetky kognitívne funkcie do rôznych oblastí frontálnych a temporálnych mozgových lalokov.

Pre kognitívnu vedu hrá kľúčovú úlohu vznik a rozvoj neurónovej paradigmy v neurovede. Aj keď už v 2. polovici 19. storočia bolo v podstate jasné, že existuje veľmi úzka korelácia medzi mikroskopickou štruktúrou mozgu a kognitívnymi aktivitami[1], konktrétne vedomosti o mikroskopickej štruktúre mozgu boli veľmi malé. V tomto smere zohral kľúčovú úlohu taliansky lekár C. Golgi z 2. polovice 19. storočia, ktorý vynašiel metódu selektívneho zafarbenia buniek v mozgu pomocou dusičnanu strieborného, aplikovaním tejto metódy sa stali dobre rozpoznateľné mozgové mikroskopické štruktúry. Tento objav inicioval búrlivý rozvoj neurónovej paradigmy hlavne zásluhou španielskeho lekára S. E. y Cajal, ktorý na prelome 19. a 20. storočia vykonal dôležité výskumy pre etablovanie sa neurónovej paradigmy v neurovede; aplikoval Golgiho metódu k charakterizovaniu rôznych mikroštruktúr mozgu na rovnakom mieste (čím získal experimentálne podklady pre špecifikáciu mozgu ako neurónovej siete obsahujúcu neuróny, ktorá obsahujú dendritický systém (zložitý kanálový system pre vstupnú informáciu), axón (kanál pre výstupnú informáciu) a somu (telo neurónu). Taktiež ukázal, že medzi rôznymi neurónmi existujú spoje – synapsie, kde axón z jedného neurónu sa „prilepí“ na dendrit druhého neurónu, pričom tieto spoje sa realizujú pomocou struktúr nazývaných „trne“. Skutočnosť, že medzi neurónmi existujú synapsie, viedlo počiatkom 20. rokov minulého storočia k štúdiu elektrochemických mechanizmov prenosu signálov medzi neurónmi. V nasledujúcich dekádach minulého storočia došlo k prudkému rozvoju štúdia detailov elektrochemického mechanizmu prenosu signálu cez synapsie, čo viedlo k tvorbe detailného popisu aktivity neurónov. V tom istom období anglický lekár Sir E. A. Adrian , ktorý na tú dobu využil moderné prístrojové vybavenie (katódovú trubicu a elektronické zosilňovače), bol schopný zaznamenávať priebeh vnútorného potenciálu neurónu. Pomocou tejto techniky Adrian zistil, že neurón pracuje v móde „všetko alebo nič“: ak suma vstupných signálov (cez dendritický systém neurónu) prevýši prahovú hodnotu vnútorného potenciálnu, potom neurón cez axón začne vysielať výstupné signály.

1.5 Kognitívna veda a umelá inteligencia

Prv než pristúpime k definovaniu umelej inteligencie a špecifikácie jej významu pre kognitívnu vedu, obráťme našu pozornosť na dôležitý „pomocný“ pojem „inteligentný problém“, ktorý bude mať centrálny význam pri definovaní umelej inteligencie. ***Inteligentný problém*** ľudskej kognície je taký problém, ktorého riešenie vyžaduje tieto vyššie aspekty

[1] Americký filozof H. Spencer v knihe “*The Principles of Psychology*” z r. 1855 prezentoval ideu, že pochopenie základných princípov fungovania nervového systému je nevyhnutným predpokladom pre korektnú interpretáciu procesov a fenoménov psychológie. Podobne, ďalší americký filozof W. James v knihe „*The Principles of Psychology*“ z r. 1890 vychádzal z postulátu, že psychologické fenomény sa musia vysvetľovať pomocou aktivít mozgu, a existuje úplný paralelizmus medzi analýzou fungovania nervov a analýzou mentálnych ideí, každá mentálna modifikácia musí byť sprevádzaná zmenami v nervovom systéme mozgu. Pomocou tejto idey rozpracoval teóriu asociatívnej pamäti, ktorá obsahuje jednotlivé idey a ktoré sú medzi sebou prepojené do paralelnej štruktúry tak, že vybavenie si (recall) jednej idey je sprevádzané súčasným pripomenutím si asociovaných ideí.

kognície: (1) percepciu prostredia, (2) vykonávanie akcií nad týmto prostredím, (3) učenie, (4) usudzovanie, (5) plánovanie, (6) rozhodovanie, (7) riešenie úloh, (8) komunikáciu, a pod. Samozrejme, nie vždy sa vyžaduje spoločný výskyt týchto aspektov, obvykle sa vyskytujú vo vzájomnej kombinácie len niekoľkých vyššie uvedených aspektov kognície. Potom, ***umelá inteligencia je tá časť informatiky, ktorá sa zaoberá algoritmizáciou inteligentných problémov.*** Pojem „inteligentný problém" je historicky závislý, to čo sa nám javilo pred 30-40 rokmi ako inteligentný problém, v súčasnosti v dôsledku teoretického rozvoja matematiky a informatiky môže byť pomerne jednoducho algoritmizovateľný a nevyžaduje použitie výpočtových techník umelej inteligencie (napr. konštrukciu stromu riešení pomocou prehľadávania). Ako príklad môžeme uviesť program *DENDRAL*, vytvorený v 60. rokoch minulého storočia pod vedením laureáta Nobelovej ceny za medicínu Joshua Lederberga v spolupráci so známym stanfordským informatikom E. Feigenbaumom. Cieľom tohto programu bola rekonštrukcia vzorcu molekuly zo známeho hmotnostného spektra (ktoré špecifikuje fragmenty vzniknuté z danej molekuly dopadom elektrónov). Tento problém bol algoritmizovaný použitím klasických metód vtedajšej umelej inteligencie. V súčasnosti je problém rekonštrukcie grafu na základe jeho podgrafov štandardný problém počítačovej teórie grafov a nepovažuje sa za „inteligentný problém" vhodný pre umelú inteligenciu.

Kognitívna psychológia používa koncepcie umelej inteligencie a kognitívnej vedy k formulovaniu a testovaniu svojich hypotéz. Mentálne štruktúry a procesy sú popísané pomocou informatických koncepcií, čo umožňuje vykonávať počítačové simulácie modelov vytváraných v rámci kognitívnej psychológie. Otázka o tom, či ľudia vykonávajú danú kognitívnu aktivitu tak, ako je špecifikovaná jej kognitívnym modelom je fundamentálneho charakteru pre kognitívnu psychológiu, práve počítačové simulácie môže byť veľmi nápomocné pri hľadaní odpovede na túto otázku.

V umelej inteligencii už od jej vzniku opakovane diskutuje problém, ako najlepšie modelovať inteligenciu. Tento problém má v zásade dve riešenia, symbolickú umelú inteligenciu a konekcionistickú umelú inteligenciu. Symbolická metafora umelej inteligencie je založená na algoritmoch, ktoré pomocou formálnych pravidiel transformujú vstupné symboly na výstupné symboly. Jedná sa sekvenčné algoritmy, ktoré sa vykonávajú krok po kroku, pričom v danom kroku vykonávania algoritmu využívame výsledky získané v predchádzajúcich krokoch vykonávania algoritmu. Konekcionistická metafora umelej inteligencie, ktorá je založená na používaní neurónových sieti obsahujúcej mnoho jednoduchých procesorov, ktoré sú medzi sebou poprepájané. Neurónovú sieť môžeme chápať ako paralelný počítač, kde aktivita vybraného neurónu v čase t závisí od aktivít neurónov v čase t-1, ktoré sú v siete klasifikované ako „predchádzajúcich" neuróny, t. j. sú spojené spojmi, ktoré z nich vychádzajú a vchádzajú do daného neurónu. Hybridne systémy sú kombináciou symbolického a konkcionistického prístupu k spracovaniu informácie. Tento prístup je v súčasnosti často aplikovaný v umelej inteligencii, kde nižšie kognitívne aktivity sú modelované neurónovou sieťou, zatiaľ čo vyššie kognitívne aktivity sú modelované symbolickým prístupom. Moderná umelá inteligencia sa zameriava na multiagentové systémy, ktoré obsahujú intencionálnych agentov (agenti, ktorý vykonávajú predpísanú aktivitu – intenciu).

1.5.1 Symbolické, konekcionistické a hybridné systémy

Konekcionistické modely mysle sú založené na metafore neurónovej siete obsahujúcej jednoduché procesné elementy – neuróny a jednosmerné spoje medzi nimi – synapsie. Poznamenajme, že konekcionistické počítanie (alebo počítače) sa diametrálne odlišujú klasického sekvenčného počítača modelovaného Turingovým strojom.

Často sa rozdiel medzi klasickým symbolickým prístupm a konekcionistickým prístupom (tiež nazývaným subsymbolickým) považuje za skoro neprekonateľný, ktorý rozdeluje komunitu umelej inteligencie a/alebo kognitívnej vedy na dva „nezmieriteľné tábory“, čo je však veľký omyl. Tento rozdieľ je však veľmi vágny, ba až delikátny. Nie všetký konekcionistické prístupy sú založené na subsymbolickom spracovaní informácie, a naopak, mnohé klasické prístupy umelej inteligencie sú založené na konekcionistickom predspracovaní vstupnej informácie (menovite procesy videnia a pochopenie prirodzeného jazyka). Na prelome 50. a 60. rokov minulého storočia klasická umelá inteligencia sa vyvíjala podstatne rýchlejšie ako konekcionistické modely kognitívnych procesov. V roku 1969 Minsky spolu s Papertom vo svojej slávnej knihe „*The perceptron*“

Aj keď už na prelome 50-tych a 60-tych rokov bolo jasné, že perceptrón má určité obmedzenia, že nie je univerzálne použiteľný na klasifikáciu ľubovoľných obrazcov, až pomocou knihy "*Perceptron*" boli tieto obmedzenia presne opísané a špecifikované. Ukázalo sa a na rôznych príkladoch všeobecného charakteru sa demonštrovalo, čo dokáže perceptrón korektne klasifikovať a čo nie. Minsky a Papert zaviedli fundamentálny pojem *lineárnej separovateľnosti*, pomocou ktorého sa pomerne jednoducho (a tiež názorne) formulujú podmienky pre korektnú klasifikáciu danej triedy objektov. V dobe publikovania knihy bol jedným z najaktuálnejších problémov umelej inteligencie problém rozpoznávania a klasifikácie scén. Práve v tejto dobe vznikali základné "symbolické" algoritmy a postupy, ako riešiť problém analýzy a klasifikácie scén. Preto Minsky a Papert upriamili svoju pozornosť na riešenie týchto problémov pomocou perceptrónov. Zistili napríklad, že perceptrón je schopný rozlíšiť trojuholníky od štvorcov, ale nedokáže rozlíšiť súvislé geometrické telesá od nesúvislých. Neprekonateľné problémy má perceptrón aj s korektnou klasifikáciou nových obrazcov, ktoré vnikli z pôvodného obrazca aplikovaním translácie alebo rotácie. Ich záver sa dal preto očakávať: *perceptrón nie je univerzálne výpočtové zariadenie, existujú jednoduché úlohy, ktoré nie sú korektne klasifikovateľné perceptrónom*. V knihe boli navrhnuté aj rôzne zovšeobecnenia perceptrónu, a to použitie skrytých neurónov a neurónov vyššieho rádu. Autori ukázali, že tieto zovšeobecnenia perceptrónu sú schopné prekonať bariéru lineárnej separovateľnosti. Žiaľ, oba tieto prístupy boli spochybnené samými autormi tým, že nepovažovali za potrebné študovať metódu učenia sa pre nastavenie váhových koeficientov skrytých neurónov alebo neurónov vyšších rádov. Hlavné príspevky Minského a Paperta k teórii neurónových sietí možno zhrnúť takto:

(1) Minsky v knihe "*Computation: Finite and Infinite Machines*" pri štúdiu McCullochových a Pittsových neurónov zaviedol pojem rekurentnej neurónovej siete, ktorá bola zložená z prahových neurónov. Pomocou konštruktívneho dôkazu ukázal, že táto sieť simuluje konečný binárny automat. Podobnú problematiku, lenže na kvalitatívne vyššej úrovni, začali študovať až na prelome 80-tych a 90-tych rokov pomocou dynamických prístupov k teórii rekurentných neurónových sietí. Tieto siete sa chápu ako dynamické systémy so zložitými trajektóriami v stavovom priestore, ktoré môžu simulovať konečné automaty.

(2) Minsky a Papert v knihe "*Perceptron: An Introduction to Computational Geometry*" podrobne študovali výpočtové možnosti perceptrónu. Ukázali, že je schopný klasifikovať korektne len takú množinu objektov, ktorá je v zodpovedajúcom priestore vstupných aktivít lineárne separovateľná. Navrhli tiež dva možné prístupy, ako prekonať túto bariéru, a to buď použitím skrytých neurónov, alebo použitím neurónov vyšších rádov. Pretože nevenovali dostatočnú pozornosť procesu učenia sa takto zovšeobecnených perceptrónov, uvedené prístupy označili za neefektívne, pretože nie je k dispozícii metóda učenia sa, ktorá by určila hodnoty váhových koeficientov priradených týmto neurónom.

Prečo sa venujeme tak podrobne vyčísleniu zásluh týchto dvoch informatikov pre rozvoj neurónových sietí? V literatúre o neurónových sieťach sa už notoricky opakuje tvrdenie, že Minsky s Papertom svojou druhou knihou "*Perceptron*" spôsobili, že nastal na niekoľko desaťročí útlm výskumu v oblasti neurónových sietí. Podľa mienky autorov tejto

kapitoly je tento názor nepravdivý, ba až smiešny. Táto vynikajúca kniha je stále primárnym zdrojom inšpirácie pre teóriu neurónových sietí a isto nespôsobila dlhodobý útlm rozvoja neurónových sietí. Hlavným zdrojom tohto útlmu bola skutočnosť, že vnútorné zdroje symbolickej umelej inteligencie neboli ešte zďaleka vyčerpané. Symbolická umelá inteligencia sa v tej dobe javila ako hlavný generátor nových trendov umelej inteligencie. Ďalším dôvodom bol odklon od problematiky rozpoznávania vzorov, v ktorom boli neurónové siete vtedy najlepšie, k problematike vyšších kognitívnych procesov, ako je napríklad riešenie problémov a dedukcia. Vtedy sa zdalo, že prostriedky klasickej umelej inteligencie dokážu tieto úlohy z vyšších kognitívnych procesov dobre riešiť. K tomu sa pridalo rozčarovanie z počiatočných neopodstatnených nádejí, keď priaznivci neurónových sietí, podobne ako ostatní vedci pracujúci v umelej inteligencii, prisudzovali vtedajším neurónovým sieťam aj kvality, ktoré tieto jednoducho nemali - čo sa skoro ukázalo. No aj tak boli neurónové siete aj naďalej rozvíjané hŕstkou nadšencov, ako bol napríklad Teuvo Kohonen a Stephen Grossberg. Situácia sa diametrálne zmenila až na prelome 80-tych a 90-tych rokov, keď si informatici začali čoraz naliehavejšie uvedomovať obmedzenosť metód symbolickej umelej inteligencie. Nepodarilo sa úplne uskutočniť žiadnu veľkú víziu symbolickej umelej inteligencie zo 60-tych a 70-tych rokov (napr. prekladač z prirodzeného jazyka, komunikácia s počítačom v prirodzenom jazyku, program pre šach na veľmajstrovskej úrovni,...).

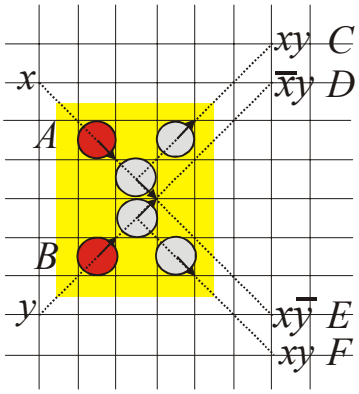
1.6 Počítače a kognícia

Ideové prekurzory umelej inteligenciu môžu byť vysledované až do 3. storočia pred Kr., keď Aristoteles vytváral svoju logiku sylogizmov a stoický filozof Chrysippos vytváral logiku, ktorú by sme z dnešného pohľadu nazvali výrokovou. Táto staroveká grécka logika v podstate s malými zmenami pretrvala až do novoveku. Zo súčasného pohľadu možno konštatovať, že veľmi pozitívnu úlohu pre rozvoj logiky zohral nemecký filozof a matematik Gottfried Wilhelm Leibniz, pokúsil sa vybudovať formálny systém, ktorý by nahradil verbálne metódy usudzovania manipuláciami s formulami. Postuloval formálny systém s dvoma časťami: (1) jazyk logiky *lingua characteristica*, pomocou ktorého je možné reprezentovať každý výrok a (2) kalkulus *calculus ratiocinator*, pomocou ktorého je možné uskutočňovať usudzovanie systematickým a matematicky presným spôsobom. Žiaľ, trvalo ešte ďalších 200 rokov než sa naplnila táto Leibnizova idea, keď v polovici 19 st. anglický matematici A. de Morgan a G. Boole zostrojili „kalkulus“ – výrokovú logiku. Pre Descartesovho súčasník anglického filozofa T. Hobbesa myslenie už nebolo nič iné, ako len špeciálny druhy výpočtu. Táto hypotéza, ktorá v 17. storočí znela veľmi neobvykle ba až exoticky, bola až v súčasnosti plne akceptovaná a realizovaná pomocou umelej inteligencie a kognitívnej vedy, kde má postavenie centrálnej paradigmy. Takto vzniknutú a formulovanú logiku možno považovať za formálnu metódu, ktorá algoritmizuje naše usudzovanie, čiže logika je integrálnou súčasťou umelej inteligencie, menovite tej časti, ktorá sa zaoberá usudzovaním.

Fundamentálna práca z r. 1937 (ktorá má zásadný význam pre vznik umelej inteligencie a kognitívnej vedy), ktorej autorom bol anglický matematik a logik (a snáď aj jeden z prvých britských informatikov) Alan Turing, riešila známy Hilbertov problém rozhodnuteľnosti o tom, či ľubovoľný matematický výrok je pravdivý alebo nepravdivý. Turing pri riešení tohto problému dospel k abstraktnému počítaču, ktorý je schopný vypočítať (alebo aj nevypočítať) funkciu charakterizujúcu daný matematický výrok[2]. Potom môžeme

[2] Napríklad vieme, že pravdivosť (alebo nepravdivosť) ľubovoľnej formuly výrokovej logiky môže byť pomocou tabuľkovej metódy (algoritmu) vypočítaná konečným počtom krokov. Môžeme teda povedať, že pravdivosť formúl výrokovej logiky je vypočítateľná, t. j. existuje algoritmus (aj keď s exponenciálnym nárastom zložitosti),

celú triedu prípustných funkcií rozdeliť na dve potriedy: funkcie, ktoré sú Turingovým počítačom vypočitateľné a funkcie, ktoré nie sú Turingovým počítačom vypočítateľné. K podobné mu výsledku dospel súčasne aj americký logik Alonzo Church pomocou tzv. rekurzívnych funkcií, tieto dva výsledky sú spojené do Turingovej a Churchovej tézy, podľa ktorej pre každú funkciu, ktorá je algoritmizovateľná (pozri poznámku 2) je aj vypočítateľná pomocou Turingovho počítača; t. j. existuje veľmi tesňá väzba medzi algoritmizovateľnosťou nejakej funkcie (problému) a jeho vypočítateľnosťou. Pretože veríme, že naše kognitívne aktivity sú algoritmizovateľné, tak môžu byť aj iomplementované na počítači, čiže naša myseľ a Turingov počítač sú „ekvivalentné“, t. j. že tak myseľ ako aj Turingov počítač, dávajú rovnaký výstup na rovnaký vstup, odlišujú sa len v spôsobe „vypočítania“ tohto výstupu.



Dokonala elastická zrážka dvoch biliardových gúľ simuluje logickú bránu konjunkcie. Binárne premenné x a y špecifikujú, či v čase $t = 1$ sa v daných východiskových polohách nachádza prvá resp. druhá guľa. K elementárnemu aktu zrážky dochádza v čase $t = 2$. Výsledné polohy v čase $t = 3$ sú špecifikované súčinmi (konjunkciami) binárnych premenných $xy, \bar{x}y, x\bar{y}, xz$.

V polovici 80. rokov minulého storočia anglický fyzik a informatik I. Deutsch formuloval fyzikálnu verziu Churchovej a Turingovej tézy: „každý konečný fyzikálny proces simuluje nejaký výpočet“. Podľa tejto verzie Churchovej a Turingovej tézy každý fyzikálny proces môžeme chápať ako špeciálny výpočet, problém je len v tom ako kódovať vstup do tohto fyzikálneho systéme a dekódovať jeho výstup. Známy príklad takéhoto fyzikálneho počítania sú biliardové gule, ktoré bolo navrhnuté americký fyzici Fredkinom a Toffolim v r. 1982.

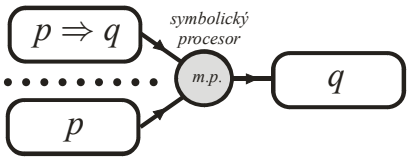
Skoro súčasne s týmto dôležitým teoretickým Turingovým výsledkom bolo na konkrétnom príklade v neurovede ukázané, že počítače špeciálneho typu môžu byť zostrojené aj pomocou jednoduchých logických (alebo prahových) neurónov McCullochom a Pittsom (pozri kapitolu 1.1). Ukázali, že každá Boolova funkcia, ktorá obsahuje spojky konjunkcie, disjunkcie a negácie je vyjadrená pomocou doprednej neurónovej siete. Ak tieto siete obsahujú cykly (hovoríme, že sieť je rekurentného charakteru), potom sieť môže simulovať konečnostavový stroj schopný univerzálneho počítania.

Znázornenie symbolického procesoru, ktorý uskutočnuje pravidlo usudzovania výrokovej logiky „modus ponens“, ktoré je vyjadrené pomocou schémy usudzovania

$$\begin{array}{ll} p \Rightarrow q & \text{1. predpoklad} \\ p & \text{2. predpoklad} \\ \hline q & \text{dôsledok} \end{array}$$

kde p a q sú symboly – výrokové premenné.

pomocou ktorého sme schopný určiť pravdivosť [illegible] jednoduché riešenie pre formuly predikátovej logiky, kde vo všeobecnosti neexistuje algoritmus, ktorý určí pravdivostnú hodnotu každej formuly predikátovej logiky; tento problém je riešiteľný len pre široké triedy formúl predikátovej logiky.



Na prelome 60. a 70. rokoch minulého storočia bola formulovaná ***komputačná teória mysle*** (taktiež nazývaná komputačná metafora alebo komputačná paradigna) kognitívnej vedy, podľa ktorej je myseľ špeciálny symbolický procesor, ktorý vstupné symboly transformuje na výstupné pomocou hierarchického systému pravidiel typu „ak..., potom...“. Je dôsledným prejavom funkcionalizmu, podľa ktorej myseľ nie je priamo totožná s mozgom, ale je prejavom jeho organizácie, štruktúr, ich vzájomného prepojenia. Táto hypotéza má tieto dôležité vedlajšie aspekty:

(1) Výpočet prebieha nad syntakticky špecifikovanými symbolmi (t. j. symboly sú špečifikované pravidlami, ktoré taktiež špecifikujú ich kombinácie), ktoré majú svoj sémantický význam. Tak napríklad, súčet dvoch symbolov, ktoré majú význam – sémantickú interpretáciu – dekadické číslo.
(2) Výpočet je reprezentovaný algoritmom,. ktorý je konečnou krok-po-kroku procedúrou, z ktrých každá je uskutočnená počítačom – mozgom.
(3) Potrebné výpočty neobsahujú len štandardnú aritmetiku, ale taktiež aj logické usudzovanie včítane indukcie, abdukcie a rozhodovanie.
(4) Výpočty pokrývajú relatívne samostatnú úroveň ich psychologickej explanácie, ktorá je nezávislá od neurofyziológie mozgu.

1.6.1 Syntax, sémantika a algoritmy

Základný význam pre pochopenie a interpretáciu komputačnej teórie mysle je rozdiel medzi syntaxou a sémantikou symbolického systému. Spôsob povolenej manipulácie so symbolmi je problém syntaktický a otázka čo znamenajú novovytvorené symboly je problém sémantický. Uvažujme tri rôzne notácie pre čísla: (1) dekadické čísla ‘1’, ‘2’, ‘3’,…, (2) rímske čísla ‘I’, ‘II’, ‘III’,… a (3) binárne čísla ‘1’, ‘10’, ‘11’,…. Tak napríklad, číslo „tri“ má tri rôzne syntaktické reprezentácie 3, III a 11. Pravidlá pre súčet čísel v rôznej reprezentácii sú diametrálne odlišné, ale ich výsledky musia mať rovnakú sémantickú interpretáciu. Keď sa dieťa učí jednoduchú aritmetickú operáciu napr. s celými číslami, potom sa učí používať vhodný algoritmus na realizáciu tejto operácie, pričom tento algoritmus musí odrážať ich sémantiku (menovite táto úloha „sčítať prirodzené čísla“ musí odrážať sémantickú interpretáciu čísel). Ak by sme vykonávali „sčitanie“ pre rímske alebo binárne čísla, musíme dostať rovnaké výsledky, aké dostávame pre dekadické čísla.

Algoritmus je “mechanická” krok-po-kroku procedúra, ktorá transformuje vstupné symboly na výstupné symboly, pričom tieto symboly majú presne špecifikovaný syntax, ktorý odráža ich sémantickú interpretáciu (význam). Turing navrhol všeobecnú konštrukciu tohto algoritmu (ktorý sa nzýva Turingov počítač), ktorá postupne transformuje vstupné symboly na výstupné symboly. Podľa Churchovej a Turingovej tézy, Turingov počítač môže počítať všetko, čo je je „intuitívne“ vypočítateľné. Táto idea bola inšpiráciou pre vznik súčasnej výpočtovej techniky. Komputáčná teória mysle prenáša túto paradigmu do psychológie postulátom, že všetky inteligentné procesy môžu byť študované ako algoritmické procesy realizované na jednoduchých symbolických procesoroch.

1.6.2 Jednoduchý symbolický procesor

Turingov stroj obsahuje „pásku“ na ktorú sa zaznamenávajú symboly medzivýsledky, „skener“, ktorý je schopný s páskou pohybovať dopredu a dozadu a taktiež zapisuje a číta symboly na páske. Toto je však len jedna z možných architektúr Turingobsho stroja, čo je dôležité pre stroj – počítač, že jeho elementárne operácie sú realizované pomocou nejakého fyzikálneho procesu. K tomu, aby komputačný model mozgu bol aspoň vzdialene podobný ľudskému mozgu, použijeme prístup logických neurónových sietí McCullocha a Pittsa, ktoré obsahujú elementárne brány konjunkcie, disjunkcie a negácie. Budeme študovať elementárnu úlohu súčtu dvoch jednobitových čísel

$$\alpha_1 + \alpha_2 = \beta_1\beta_2$$

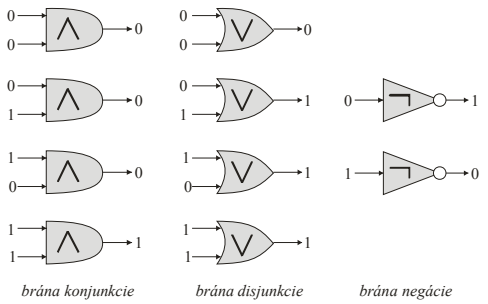
ktorej všetky možné prípady sú uvedené v nasledujúcej tabuľke

α_1	α_2	β_1	β_2
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

kde prvá dva stĺpce reprezentujú vstupné čísla – bity a posledná dna stĺpce reprezentujú výstupné čísla – bity. Z tejto tabuľky bezprostredne vyplýva, že výstupné byty β_1 a β_2 sú určené ako jednoduché funkcie vstupných bitov

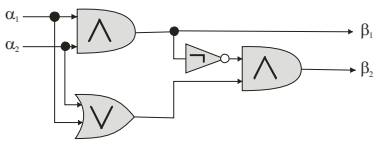
$$\beta_1 = \alpha_1 \wedge \alpha_2\,,\ \beta_2 = \left(\neg\alpha_1 \wedge \alpha_2\right) \vee \left(\alpha_1 \wedge \neg\alpha_2\right) = \left(\alpha_1 \vee \alpha_2\right) \wedge \neg\left(\alpha_1 \wedge \alpha_2\right) \qquad (*)$$

kde pravá strana druhej formuly môže byť jednoducho verifikovaná aplikáciou de Morganovej formuly na konjunkcie $\alpha_1 \wedge \alpha_2$ a potom následným „roznásobením“ oboch zátvoriek s disjunkciou. Týmto sme ukázali, že kognitívna úloha súčtu dvoch jednobitových čísel môže byť vyjadrená funkciou ktorá je jednoducho vypočítateľná (ako ukážeme v ďalšom texte, táto vypočítateľnosť je uskutočnená pomocou obvodu – neurónovej siete, ktorá obsahuje brány konjunkcie, disjunkcie a negácie.



Logické brány konjunkcie, disjunkcie a negácie, ktoré na ľavej strane majú vstupné „binárne“ hodnoty a na pravej strane výstupnú „binárnu“ hodnotu. Vstupné hodnoty obsahujú všetky možné rôzne prípady vstupov. Fyzikálna realizácia týchto brán je v súčasnosti rutinnou technologickou záležitosťou.

Ak tieto elementárne logické brány prepojíme do funkčného celku nazývaného neurónová sieť tak, aby boli “implementované” Boolove funkcie (∗) pre jednotlivé komponenty výsledného súčtu $\beta_1\beta_2$.



Logický obvod (neurónová sieť) realizujúci súčet dvoch 1-bitových čísel α_1 a α_2 (tvoria vstup do siete) na výsledné 2-bitové číslo $\beta_1\beta_2$ (jeho komponenty tvoria výstup siete). Tento obvod (neurónová sieť) môže byť formálne chápaný ako počítač realizujúcu úlohu súčtu dvoch 1-bitových čísel.

V rámci *konekcionistického prístupu* ku vzťahu medzi mozgom a mysľou je tento problém riešený pomocou koncepcií neurónových sietí. Podľa jednej zo základných paradigiem kognitívnej vedy *je mozog/myseľ charakterizovaný ako počítač,* (1) ktorý transformuje symboly pomocou syntaktických pravidiel na iné symboly, pričom (2) myšlienky sú symbolické reprezentácie implementované pomocou jazyka myslenia, a (3) mentálne procesy sú postupnosti symbolov (medzi ktorými sú príčinné vzťahy) generované syntaktickými pravidlami. Použitie termínu „počítač“ obvykle evokuje predstavu sekvenčného počítača von neumannovskej architektúry (napr. personálne počítače majú túto architektúru), kde je možné striktne oddeliť hardware od software; kde na tom istom počítači – hardware môže byť vykonávaných nepreberné množstvo diametrálne rozdielnych programov – softvérov. Pre tieto počítače existuje striktná dichotómia medzi počítačom a programom – t. j. medzi hardvérom a softvérom. Žiaľ, metafora mysle ako počítača implikuje u mnohých ľudí predstavu, že je možné oddeliť mozog od mysle, ako dva „nezávislé“ fenomény, kde mozog hrá úlohu hardvéru, zatiaľ čo myseľ je softvér (vykonávaný na hardvéru – mozgu). V tejto súvislosti sa tiež hovorí o sociokultúrnom zdroji ľudskej mysle, kde mozog akoby bol len irelevantným „hardwarovým“ vehiklom mysle.

Obráťme našu pozornosť na súčasný prístup k chápaniu vzťahu medzi mozgom a mysľou, ktorý je založený na konekcionistickom poňatí tak mozgu, ako aj mysle. Základná predstava o mozgu (založená na experimentálnych neurovedných poznatkoch) je, že je tvorený z neurónov navzájom poprepájaných pomocou jednosmerných synaptických spojov. Ľudský mozog vykazuje neobyčajnú plasticitu, v priebehu učenia neustále vznikajú (ale taktiež aj zanikajú) synaptické spoje. Architektúra mozgu je určená spojmi medzi neurónmi, ich inhibičným, alebo excitačným charakterom, a taktiež aj ich intenzitou. Možno konštatovať, *že schopnosť mozgu vykonávať nielen kognitívne aktivity, ale byť aj pamäťou, je plne zakódovaná do jeho architektúry*. Na základe týchto neurovedných poznatkov bazálneho charakteru môžeme konštatovať, že počítačová paradigma ľudského mozgu/mysle sa musí formulovať tak, že ***mozog je paralelný distribuovaný počítač*** (obsahujúci mnoho miliárd neurónov, elementárnych procesorov, ktoré sú medzi sebou poprepájané do zložitej neurónovej siete). Program v tomto paralelnom počítači je priamo zabudovaný do architektúry neurónovej siete, t. j. ľudský mozog je ***jednoúčelový*** paralelný počítač reprezentovaný neurónovou sieťou, ktorý nie je možné preprogramovať bez zmeny jeho architektúry. Z týchto všeobecných úvah vyplýva, *že myseľ s mozgom tvoria jeden integrálny celok; myseľ je v tomto prístupe chápaná ako program vykonávaný mozgom, avšak tento program je špecifikovaný architektúrou distribuovanej neurónovej siete reprezentujúcej mozog.* Mozog a myseľ tvoria dva rôzne pohľady na ten istý objekt: (1) keď hovoríme o mozgu, myslíme tým „hardwarovú“ štruktúru, biologicky realizovanú neurónmi a ich synaptickými spojmi (formálne reprezentovanú neurónovou sieťou), v opačnom prípade, (2) keď hovoríme o mysli, myslíme tým kognitívne a iné aktivity mozgu, realizované výpočtami neurónovej siete reprezentujúcej mozog.

Literatúra

2. prednáška

Logické neuróny McCullocha a Pittsa

2.1 Úvodné poznámky

Ako už bolo poznamenané v predchádzajúcej prednáške, publikácia Warrena McCullocha a Waltera Pittsa [3] „*A logical calculus of the ideas immanent to nervous activity*" z r. 1943 je medzníkom v rozvoji konekcionizmu. V tejto práci bolo s geniálnou jasnozrivosťou ukázané, že neurónové siete sú mocným modelovým prostriedkom, napr., že siete zložené z logických neurónov simulujú Boolove funkcie.

Elementárnou jednotkou McCullochovej a Pittsovej neurónovej siete je *logický neurón* (pozri ref. [4]) s binárnymi stavmi 1 a 0. Vstupný systém logického neurónu obsahuje *excitačné vstupy* (opísané binárnymi premennými x_1, x_2, ..., x_n) a *inhibičné vstupy* (opísané binárnymi premennými x_{n+1}, x_{n+2}, ..., x_m). Aktivita logického neurónu je jednotková, ak *vnútorný potenciál* neurónu definovaný ako rozdiel medzi sumou excitačných vstupných aktivít a inhibičných vstupných aktivít je väčší alebo rovný prahu ϑ, v opačnom prípade je nulová

$$y = \begin{cases} 1 & \left(x_1 + \ldots + x_n - x_{1+n} - \ldots - x_m \geq -\vartheta\right) \\ 0 & \left(x_1 + \ldots + x_n - x_{1+n} - \ldots - x_m < -\vartheta\right) \end{cases} \tag{2.1}$$

Pomocou jednoduchej krokovej funkcie (1.1b) môžeme aktivitu z vyjadriť takto:

$$y = s\left(\underbrace{x_1 + \ldots + x_n - x_{1+n} - \ldots - x_m}_{\xi} + \vartheta\right) \tag{2.2}$$

Tento vzťah pre aktivitu logického neurónu môžeme alternatívne interpretovať tak, že excitačné aktivity vstupujú do neurónu cez spoje, ktoré sú ohodnotené jednotkovým váhovým koeficientom ($w = 1$), zatiaľ čo inhibičné aktivity vstupujú do neurónu cez spoje so záporným jednotkovým váhovým koeficientom ($w = -1$). Potom aktivitu logického neurónu môžeme vyjadriť takto

$$y = s\left(\underbrace{w_1 x_1 + \ldots + w_m x_m}_{\xi} + \vartheta\right) = s\left(\sum_{i=1}^{m} w_i x_i + \vartheta\right) \tag{2.3}$$

Jednoduchá implementácia elementárnych Boolových funkcií disjunkcie, konjunkcie, implikácie a negácie je znázornená na obr. 2.1. Ako ilustratívny príklad študujme funkciu disjunkcie pre $n = 2$, použitím formúl z definície logického neurónu dostaneme

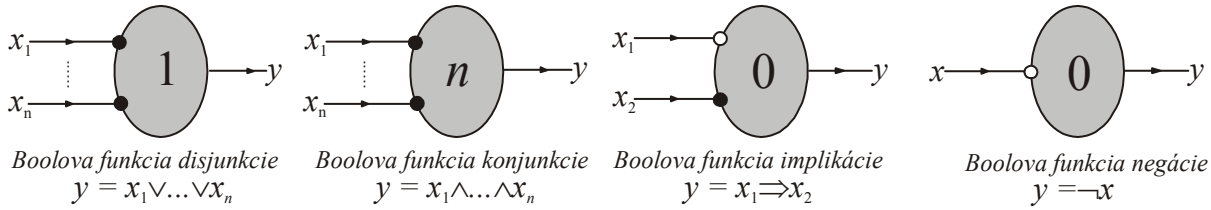
$$y_{OR}\left(x_1, x_2\right) = s\left(x_1 + x_2 - 1\right) \tag{2.4}$$

Funkčné hodnoty tejto Boolovej funkcie sú ukázané v tabuľke 2.1.

Tabuľka 2.1. Binárna Boolova funkcia disjunkcie

#	x_1	x_2	$y_{OR}(x_1,x_2)$	$x_1 \vee x_2$
1	0	0	$s(-1)$	0
2	0	1	$s(0)$	1
3	1	0	$s(0)$	1
4	1	1	$s(1)$	1

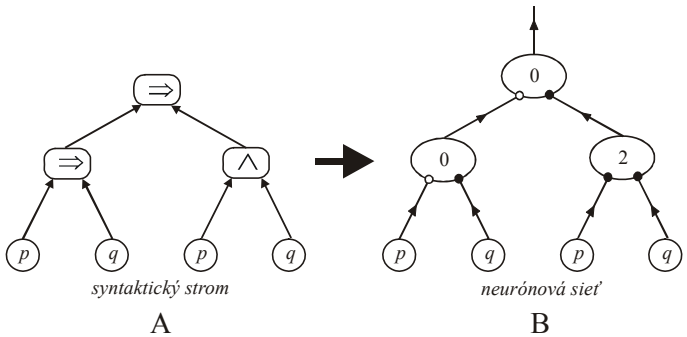
Z tabuľky vyplýva, že Boolova funkcia y_{OR} simuluje Boolovu funkciu disjunkcie.



Obrázok 2.1. Tri realizácie logických neurónov na implementáciu Boolových funkcií disjunkcií, konjunkcií a negácie. Excitačné spoje sú znázornené plným krúžkom, inhibičné prázdnym krúžkom. Pre Boolovej funkcie implikácie je použité vetvenie excitačnej aktivity na dva vstupy, to znamená, že táto excitačná aktivita je započítaná dvakrát v (4.4) pre výpočte výstupnej aktivity neurónu.

2.2 Neurónové siete

Každá Boolova funkcia [1,2] je reprezentovaná pomocou syntaktického stromu, ktrorý reprezentuje jej rekurentnú výstavbu inicializovanú Boolovými premennými a končiacu danou Boolovou funkciou (funkciou výrokovej logiky), pozri obr. 2.2, diagram A. Syntaktický strom je dôležitý keď hľadáme podformuly danej formuly, každý uzol stromu špecifikuje podformulu: najnižšie položené uzly reprezentujú triviálne podformuly p a q, nasledujúce dva vrcholy reprezentujú podformuly $p \Rightarrow q$ a $p \wedge q$, najvyššie položený vrchol – koreň stromu – reprezentuje samotnú formulu $(p \Rightarrow q) \Rightarrow (p \wedge q)$.



Obrázok 2.2. (A) Syntaktický strom Boolovej funkcie (výrokovej formuly) $(p \Rightarrow q) \Rightarrow (p \wedge q)$. Koncové vrcholy stromu reprezentujú Boolove premenné (výrokové premenné) p a q, vrcholy z nasledujúcich vrstiev sú priradené spojkám implikácie a konjunkcie. Vyhodnocovanie tohto stromu prebieha postupne zdola nahor. (B) Neurónová sieť obsahujúca logické neuróny spojok, ktoré sa vyskytujú v príslušnom syntaktickom strome diagramu A. Vidíme, že medzi syntaktickým stromom a príslušnou neurónovou sieťou existuje veľmi tesná previazanosť, ich topológia je identická, odlišujú sa len vo vrcholoch. Obrazne môžeme povedať, že neurónovú sieť pre Boolovu funkciu φ zostrojíme pomocou jej syntaktického stromu tak, že vrcholy zo syntaktického stromu, ktoré reprezentujú logické spojky, nahradíme príslušnými logickými neurónmi, koreň syntaktického stromu je zamenený .

Pomocou syntaktického stromu ľahko zostrojíme neurónovú sieť tak, že jednotlivé vrcholy reprezentujúce Boolove elementárne funkcie (logické spojky) nahradíme príslušnými logickými neurónmi podľa obr. 2.1. Na obr. 2.2 je znázornená táto konštrukcia neurónovej siete pre formulu $(p \Rightarrow q) \Rightarrow (p \wedge q)$. Tento postup konštrukcie neurónovej siete zosumarizujeme pomocou nasledujúcej vety.

Veta 2.1. Každá Boolova funkcia môže byť vyjadrená pomocou „neurónovej siete“ zloženej z logických neurónov vyjadrujúcich logické spojky.

Táto veta patrí medzi základné výsledky McCullocha a Pittsa [3,4], ukázali, že pomocou niekoľkých jednoduchých logických neurónov je možné zostrojiť neurónovú sieť, ktorá simuluje danú Boolovu funkciu. Môžeme teda hovoriť o tom, že neurónové siete s logickými neurónmi majú univerzálny charakter v doméne Boolových funkcií, vyššie naznačený konštruktívny postup reprezentuje všeobecne platnú metódu pre zostrojenie neurónovej siete, ktorá simuluje Boolove funkcie.

Aktivita i-teho neurónu v neurónovej sieti je vyjadrená zovšeobecnenou formulou (2.3)

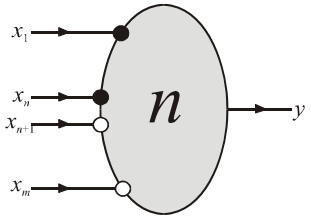
$$x_i = t\left(\sum_j w_{ij} x_j + \vartheta_i\right) \tag{2.5}$$

kde sumácia prebieha nad všetkými neurónmi, ktoré v neurónovej sieti predchádzajú i-ty neurón. Váhové koeficienty w_{ij} sú definované takto

$$w_{ij} = \begin{cases} 1 & (spoj\ j \to i\ má\ exitačný\ charakter) \\ -1 & (spoj\ j \to i\ má\ inhibičný\ charakter) \\ 0 & (spoj\ j \to i\ neexistuje) \end{cases} \tag{2.6}$$

To znamená, že v neurónovej sieti váhové koeficienty sú fixne a dané a sú určené topológiou syntaktického stromu špecifikujúceho Boolovu funkciu.

Architektúra neurónovej siete, ktorá je zostrojená pre danú Boolovu funkciu môže byť podstatne zjednodušená na tzv. 3-vrstvovú neurónovú sieť, t. j. obsahuje vrstvu vstupných neurónov (ktoré len kopírujú vstupné aktivity, nie sú výpočtovými jednotkami), vrstva skrytých neurónov a posledná vrstva obsahujúca výstupný neurón. Táto architektúra je minimalistická a už zrejme nemôže byť zjednodušená. Ukážeme ako zostrojiť takúto neurónovú sieť pre danú Boolovu funkciu.



Obrázok 2.3. Logický neurón, ktorý simuluje klauzulu, ktorá obsahuje konjunkciu konečného prvku výrokových premenných alebo ich negácií, $y = x_1 \wedge \ldots \wedge x_n \wedge \neg x_{n+1} \wedge \ldots \wedge \neg x_m$.

Jednoduchým zovšeobecnením logických spojok možno ukázať, že logický neurón je schopný simulovať aj Boolovu funkciu, ktorá obsahuje konjunkcie premenných alebo ich negácií, $x_1 \wedge \ldots \wedge x_n \wedge \neg x_{n+1} \wedge \ldots \wedge \neg x_m$, pozri obr. 2.3. Táto Boolova funkcia sa rovná 1 len pre $x_1 = \ldots = x_n = 1$ a $x_{n+1} = \ldots = x_m = 0$, vo všetkých ostatných pravdivostných kombináciách argumentov jej hodnota je 0 (nepravdivý výrok)

$$val_\tau\left(x_1 \wedge ... \wedge x_n \wedge \neg x_{n+1} \wedge ... \wedge \neg x_m\right) = \begin{cases} 1 & \left(pre\ \tau = \tau_0\right) \\ 0 & \left(pre\ \tau \neq \tau_0\right) \end{cases} \tag{2.7}$$

kde $\tau_0 = \left(x_1/1,...,x_n/1,x_{n+1}/0,...,x_m/0\right)$ je špecifikácia pravdivostných hodnôt premenných. Ľahko sa presvedčíme o tom, že táto klauzula je simulovaná logickým neurónom znázorneným na obr. 2.3. jeho výstupná aktivita je určená formulou

$$y = s\left(x_1 + ... + x_n - x_{n+1} - ... - x_m - n\right) \tag{2.8a}$$

Funkčná hodnota tejto funkcie sa rovná 1 vtedy, ak

$$x_1 + ... + x_n - x_{n+1} - ... - x_m \geq n \tag{2.8b}$$

Táto podmienka je dosiahnutá len vtedy, keď prvých *n* vstupných (excitačných) aktivít sa rovná 1 a ďalších (m-n) vstupných (inhibičných) aktivít sa rovná 0.

V teórii Boolovych funkcií je dokázana dôležitá vlastnosť Boolových funkcií, podľa ktorej každá Boolova funkcia môže byť prepísaná do ekvivalentného disjunktívneho tvaru [1,2]

$$\varphi = \bigvee_{\substack{\tau \\ (val_\tau(\varphi)=1)}} x_1^{(\tau)} \wedge x_2^{(\tau)} \wedge ... \wedge x_n^{(\tau)} \tag{2.9a}$$

kde

$$x_i^{(\tau)} = \begin{cases} x_i & \left(\text{ak } val_\tau\left(x_i\right) = 1\right) \\ \neg x_i & \left(\text{ak } val_\tau\left(x_i\right) = 0\right) \end{cases} \tag{2.9b}$$

Ako ilustračný príklad použitia tejto vlastnosti, študujme Boolovu funkciu, ktorej funkčné hodnoty sú zadané tabuľkou 2.2. V tejto tabuľkw v riadkoch 3, 4 a 6 sú jednotkové funkčné hodnoty (Boolova funkcia je pre tieto tri hodnoty premenných pravdivá), v poslednom stĺpci sú uvedené v týchto riadkoch aj príslušné klauzuly zostrojene pomocou (2.9b). Použitím (2.9a) dostanmeme „analytický“ tvar Boolovej funkcie určenej len tabuľkou jej funkčných hodnôt

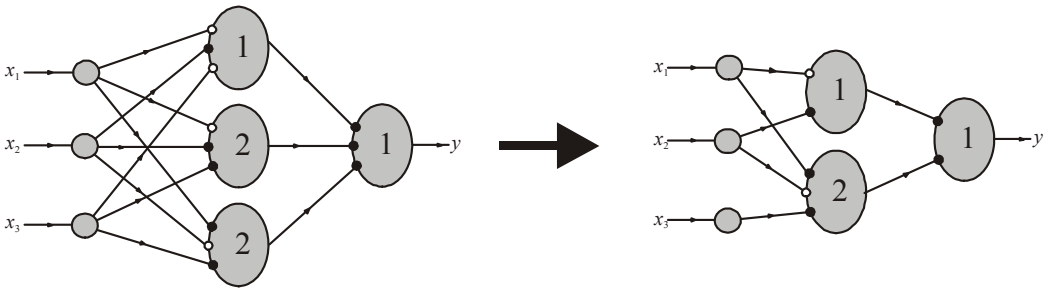
$$y = f\left(x_1, x_2, x_3\right) = \left(\overline{x}_1 \wedge x_2 \wedge \overline{x}_3\right) \vee \left(\overline{x}_1 \wedge x_2 \wedge x_3\right) \vee \left(x_1 \wedge \overline{x}_2 \wedge x_3\right) \tag{2.10}$$

Táto Boolova funkcia môže byť zjednodušená tak, že prvá a druhá klauzula sa zjednodušia

$$\left(\overline{x}_1 \wedge x_2 \wedge \overline{x}_3\right) \vee \left(\overline{x}_1 \wedge x_2 \wedge x_3\right) = \overline{x}_1 \wedge x_2 \wedge \underbrace{\left(\overline{x}_3 \vee x_3\right)}_{1} = \overline{x}_1 \wedge x_2 \tag{2.11}$$

Potom

$$y = f\left(x_1, x_2, x_3\right) = \left(\overline{x}_1 \wedge x_2\right) \vee \left(x_1 \wedge \overline{x}_2 \wedge x_3\right) \tag{2.12}$$



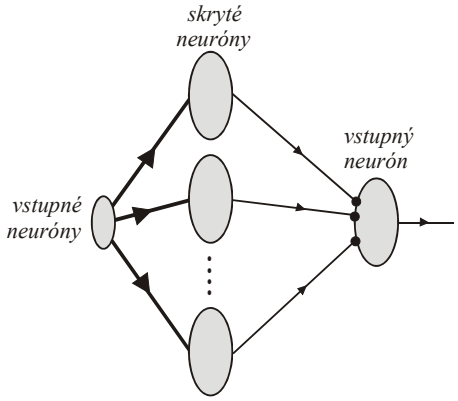
Obrázok 2.4. Trojvrstvová neurónová sieť, ktorá simuluje Boolovu funkciu zadanou primárne Tab. 2.2, z ktorej je zostrojený pomocou formule (2.9) ja „analytický“ tvar (2.10). Skryté neuróny reprezentujú jednotlivé klauzuly z Tab. 2.2, ich disjunkcia je realizovaná pomocou výstupného neurónu.. Táto neurónová sieť môže byť zjednodušená tak, že prvé dve klauzule sa spoja do jednej jednoduchšej klauzule, pozri (2.11-12).

Klauzule $x_1^{(\tau)} \wedge x_2^{(\tau)} \wedge ... \wedge x_n^{(\tau)}$ môžeme vyjadriť jedným logickým neurónom, pozri obr. 2.3. Výstupy z týchto neurónov spojíme do disjunkcie pomocou neurónu reprezentujúceho disjunkciu (pozri obr. 2.1). Potom neurónová sieť, reprezentujúca Boolovu funkciu (2.10) má tvar znázornený na obr. 2.4.

Výsledky z tohto ilustračného príkladu sú zosumarizované pomocou tejto vety.

Veta. 2.2. *Ľubovoľná Boolova funkcia f je simulovaná pomocou 3-vrstvovej neurónovej siete.*

Všeobecný tvar 3-vrstvovej neurónovej siete ja znázornený na obr. 2.5.



Obrázok 2.5. Schématické znázornenie 3-vrstvovej neurónovej siete. Idúc zľava doprava, prvá je vstupná vrstva obsahujúca vstupné neuróny, ktoré nie sú výpočtové elementy, ale len formálne reprezentujú vstupné aktivity. Druhá vrstva obsahuje skryté neuróny, ktoré reprezentujú jednotlivé klauzule danej Boolovej funkcie. Tretia (posledná) vrstva obsahuje výstupný neurón, ktorý vykonáva disjunkciu aktivít zo skrytých neurónov.

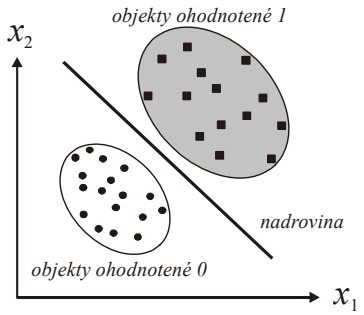
Tabuľka 2.2. Zadanie funkčných hodnôt Boolovej funkcie.

#	x_1	x_2	x_3	$y = f(x_1, x_2, x_3)$	*klauzula*
1	0	0	0	0	-
2	0	0	1	0	-
3	0	1	0	1	$\overline{x}_1 \wedge x_2 \wedge \overline{x}_3$
4	0	1	1	1	$\overline{x}_1 \wedge x_2 \wedge x_3$
5	1	0	0	0	-
6	1	0	1	1	$x_1 \wedge \overline{x}_2 \wedge x_3$
7	1	1	0	0	-
8	1	1	1	0	-

Poznamenajme, že podľa vety 2.2 3-vrstvové neurónové siete obsahujúce logické neuróny sú *univerzálne výpočtové zariadenia* pre doménu Boolových funkcií. Tento výsledok predznačil moderný výsledok neurónových sietí z počiatku 90. rokov minulého storočia, podľa ktorého trojvrstvové dopredné neurónové siete so spojitou aktivačnou funkciou (1.1) majú vlastnosť univerzálneho aproximátora (pozri kapitolu X.X). Navyše, pretože dôkaz vety bol konštruktívny, poznáme jednoduchý postup, ako túto sieť systematickým spôsobom zostrojiť pre ľubovoľnú Boolovu funkciu. Žiaľ, problém minimálnosti takto zostrojenej siete nie je vetou riešený. Vo všeobecnosti môže existovať sieť, ktorá pre danú Boolovu sieť obsahuje menej logických (skrytých) neurónov ako tá, ktorá bola zostrojená systematickým postupom z dôkazu vety. V teórii Boolových funkcií boli vypracované metódy (napr. Quinova a McCluskeyho metóda [xx]), ktoré minimalizujú Boolovu funkciu zostrojené pomocou (2.9a)

vzhľadom k veľkosti konjunktívnych členov (klauzúl). Potom môžeme zostrojiť „optimalizovanú“ neurónovú sieť pomocou takto minimalizovanej Boolovej funkcii..

Môžeme si položiť otázku, aké Boolove funkcie je schopný logický neurón vyjadriť [5]? Táto otázka sa dá pomerne jednoducho vyriešiť pomocou geometrickej interpretácie výpočtu prebiehajúceho v logickom neuróne. Výpočtová funkcia logického neurónu rozdeľuje priestor vstupov na dva polpriestory pomocou roviny $w_1x_1 + w_2x_2 +...+ w_nx_n = \vartheta$, pre koeficienty w_i=0,±1. Hovoríme, že Boolova funkcia $f(x_1, x_2,..., x_n)$ je *lineárne separovateľná*, ak existuje taká rovina $w_1x_1 + w_2x_2 + ...+ w_nx_n = \vartheta$, ktorá separuje priestor vstupných aktivít tak, že v jednej časti priestoru sú vrcholy ohodnotené 0, zatiaľ čo v druhej časti priestoru sú vrcholy ohodnotené 1 (pozri obr. 2.6).



Obrázok 2.6. Schematické znázornenie pojmu lineárnej separovateľnosti, kde okrúhle a štvorcové objekty sú separované nadrovinou $w_1x_1+...+w_nx_n-\vartheta = 0$ tak, že v jednom polopriestore sú objekty jedného druhu, zatiaľ čo v druhom polopriestore sú objekty druhého druhu.

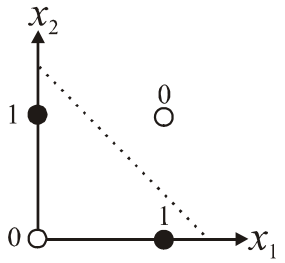
Veta 2.3. *Logický neurón je schopný simulovať len tie Boolove funkcie, ktoré sú lineárne separovateľné.*

Klasický príklad Boolovej funkcie, ktorá nie je lineárne separovateľná, je výroková spojka exkluzívna disjunkcia, ktorá je formálne definovaná ako negácia ekvivalencie $(x \oplus y) \Leftrightarrow \neg(x \equiv y)$, v informatickej literatúre sa obvykle označuje ako Boolova funkcia *XOR*, $\varphi_{XOR}(x,y) = x \oplus y$, jej funkčné hodnoty sú uvedené v tab. 2.3.

Tabuľka 2.3. Boolova funkcia XOR

#	x	y	$\varphi_{XOR}(x,y)$
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

Ak si jej funkčné hodnoty vynesieme do stavového priestoru x - y dostaneme obr. 2.7, z ktorého jasne vyplýva, že táto funkcia nie je lineárne separovateľná.

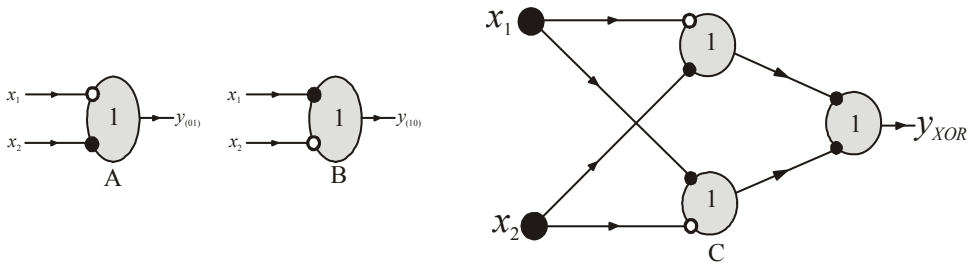


Obrázok 2.7. Znázornenie objektov Boolovej funkcie *XOR* v stavovom priestore jej argumentov. Z obrázku jasne plynie, že neexistuje priamka, ktorá by separovala celú rovinu na dve polroviny tak, že objekty (otvorené krúžky) s ohodnotením 0 budú v jednej polrovine, zatiaľ čo objekty s ohodnotením 1 (čierne krúžky) budú v druhej polrovine.

Použitím trechniky z predchádzajúcej časti tejto kapitoly, zostrojíme neurónovú sieť, ktorá simuluje túto elementárnu lineárne neseparovateľnú Boolovu funkciu. Pomocou Tab. 2.3. zostrojíme jej ekvivaletnú funkciu pomocou klauzúl

$$\varphi_{XOR}(x_1, x_2) = (\neg x_1 \wedge x_2) \vee (x_1 \wedge \neg x_2) \tag{2.13}$$

Potom neurónová sieť zostrojená pomocou tejto fukcie má tvar znázornený na obr. 2.8.

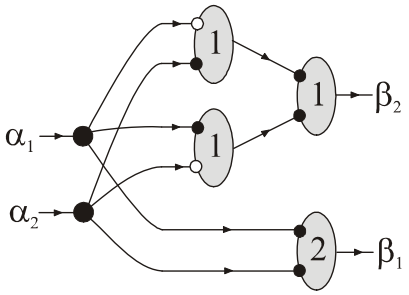


Obrázok 2.8. Diagramy A a B znázorňujú jednotlivé klauzuly z formuly (2.11). Diagram C reprezentuje 3-vrstvovú neurónovú sieť, ktorá ako skryté neuróny obsahuje neuróny z diagramu A a B, výstupný neurón reprezentuje disjunkciu výstupných aktivít zo skrytých neurónov.

Príklad 2.1. Zostrojme neurónovú sieť, ktorá simuluje sčítanie dvoch binárnych číslic:

$$\begin{array}{r} \alpha_1 \\ \alpha_2 \\ \hline \beta_1\beta_2 \end{array}$$

kde jednotlivé binárne premenné výsledku – sumy sú určené takto: $\beta_2 = \alpha_1 \oplus \alpha_2$ a $\beta_1 = \alpha_1 \wedge \alpha_2$. Druhú výstupnú premennú β_2 prepíšeme do disjunktívnej formy $\beta_2 = (\neg\alpha_1 \wedge \alpha_2) \vee (\alpha_1 \wedge \neg\alpha_2)$, príslušná neurónová sieť je znázornená na obrázku 2.9.



Obrázok 2.9. Neurónová sieť vykonávajúca súčet dvoch binárnych číslic.

V predchádzajúcej časti tejto kapitoly bolo ukázané, že samotné logické neuróny sú schopné korektne klasifikovať len lineárne separovateľné Boolove funkcie. Toto podstatné obmedzenie logických neurónov môže byť odstránené pomocou logických neurónov vyšších rádov [5], ktorých aktivita je určená zovšeobecnením formuly (2.3) o členy vyšších rádov

$$y = s\left(\underbrace{\sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j + \ldots + \vartheta}_{\xi} \right) \tag{2.14}$$

Ak vnútorný potenciál neurónu ξ je určený len ako lineárna kombinácia vstupných aktivít (t.j. len prvou sumou), potom logický neurón je štandardný a nazýva sa "*logický neurón prvého rádu*". Ak tento potenciál neurónu ξ obsahuje aj kvadratické a prípadne aj ďalšie členy, potom sa nazýva "*logický neurón vyššieho rádu*". Podľa Minského a Paperta [5] platí nasledujúca veta, ktorá hovorí o tom, že perceptróny vyššieho rádu sú schopné simulovať aj množiny objektov, ktoré nie sú lineárne separovateľné.

Veta 2.4. Ľubovolná Boolova funkcia *f* je simulovaná logickým neurónom vyššieho rádu.

Táto veta hovorí o tom, že každá Boolova funkcia môže byť simulovaná logickým neurónom vyššieho rádu, existujú také váhové koeficienty a prah, že pre každú špecifikáciu premenných $x_1, x_2, ..., x_n$ vypočítaná binárna aktivita sa rovná požadovanej hodnote.

Príklad 2.2. Ako ilustračný príklad vety študujme Boolovu funkciu *XOR*, ktorá nie je lineárne separovateľná a ktorej funkčné hodnoty sú uvedené v tabuľke 4.1. Nech aktivita logického neurónu je určená pomocou kvadratického potenciálu (čiže sa jedná o logicky neurón druhého rádu)

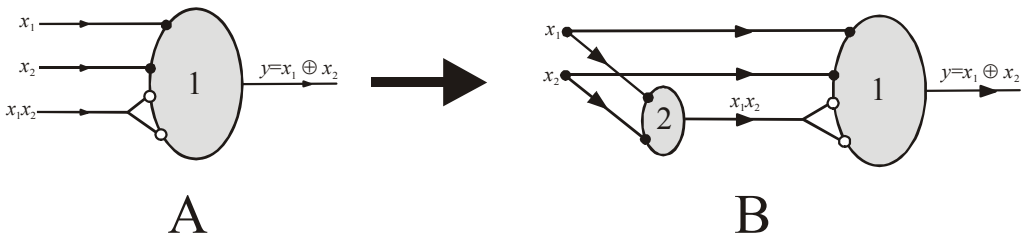
$$y = s\left(\underbrace{w_1 x_1 + w_2 x_2 + w_{12} x_1 x_2}_{\xi} - \vartheta \right) \tag{2.15}$$

Pre jednotlivé riadky tabuľky 2.3 funkcie *XOR* platí

$$\begin{aligned} & & -\vartheta < 0 \\ & w_2 & -\vartheta \geq 0 \\ w_1 & & -\vartheta \geq 0 \\ w_1 + w_2 + w_{12} & & -\vartheta < 0 \end{aligned} \tag{2.16}$$

Postupným riešením týchto nerovníc dostaneme, že váhové koeficienty a prah majú napr. takéto riešenie

$$\vartheta = 1,\ w_1 = w_2 = 1,\ w_{12} = -2 \tag{2.17}$$



Obrázok 2.10. (A) Znázornenie logického neurónu druhého rádu simulujúceho funkciu *XOR*, kde excitačné vstupné aktivity sú binárne premenné x_1 a x_2, inhibičná aktivita je priradená premennej druhého rádu x_1x_2. Výstupná aktivita z je určená pomocou krokovej funkcie, $z = s(x_1 + x_2 - 2x_1x_2 - 1)$, priamou verifikáciou sa presvedčíme, že simuluje funkciu *XOR*. Na záver poznamenajme, že "vidlička" inhibičného vstupu znamená, že vstupná aktivita je počítaná dvakrát. (C) Transformácia logického neurónu druhého rádu, ktorý implementuje spojku XOR (diagram A), na neurónovú sieť, ktorá obsahuje len logické neuróny prvého rádu (digram B). Táto transformácia je založená na výpočte súčinu x_1x_2 pomocou logického neurónu pre spojku konjunkcie, výstup

z tohto neurónu je potom dvojnásobný inhibičný vstup do pravého neurónu. Táto architektúra je asi najjednoduchšou neurónovou sieťou, ktorá implementuje spojku XOR pomocou logických neurónov prvého rádu (porovnaj s obr. 4.7).

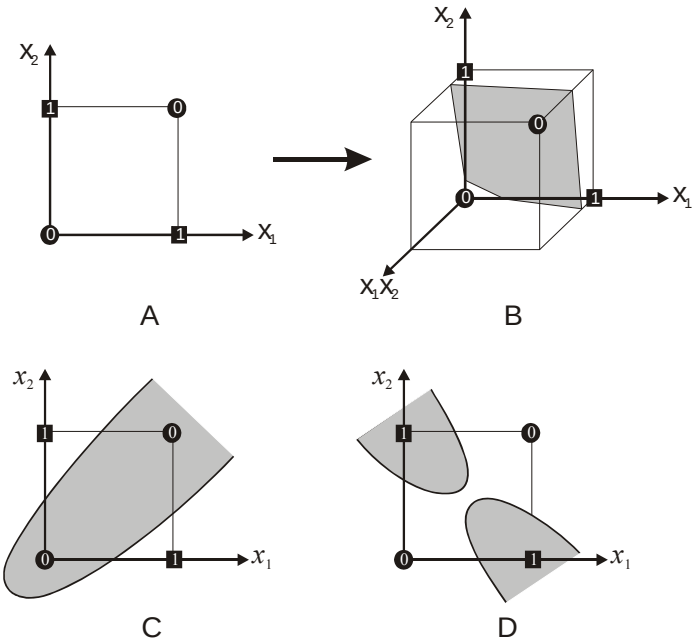
V tomto ilustračnom príklade bolo ukázané, že lineárne neseparovateľná binárna spojka disjunkcie s vylúčeným (XOR) je implementovateľná pomocou logického neurónu, ktorý má tri vstupy pre x_1, x_2 a x_1x_2. V tejto súvislosti sa vynára otázka výpočtu súčinu x_1x_2, ktorý však môže byť jednoducho vykonaný pomocou spojky konjunkcie, $x_1x_2 = x_1 \wedge x_2$. Ak táto operácia bude uskutočnená pomocou logického neurónu pre konjunkciu (pozri obr. 2.10, diagram B), potom môžeme vytvoriť jednoduchú sieť obsahujúcu len dva neuróny, pričom na vstupe máme len vstupné aktivity x_1 a x_2.

Výsledky tohto príkladu 2.2 ukázujú, že logický neurón druhého rádu je schopný korektne klasifikovať Boolovu funkcia *XOR*, ktorá je lineárne neseparovateľná v 2-rozmernom fázovom priestore x_1-x_2, ale je už lineárne separovateľná v 3-rozmernom fázovom priestore x_1-x_2- x_1x_2 , pozri obr. 2.11.

Pojem lineárnej separovateľnosti Boolových funkcií je ľahko zovšeobecniteľný na kvadratickú (kubickú,…) separovateľnosť pomocou kvadratickej (kubickej,…) nadplochy.

Definícia 2.1. Boolova funkcia f sa nazýva *kvadraticky separovateľná*, ak existujú také váhové koeficienty w_i, w_{ij} a prahový faktor ϑ, že pre každú špecifikáciu premenných $x_1, x_2, ..., x_n$ platí

$$y_{req}(x_1, x_2, ..., x_n) = 1 \Rightarrow \sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j \geq \vartheta$$
$$y_{req}(x_1, x_2, ..., x_n) = 0 \Rightarrow \sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j < \vartheta \qquad (2.18)$$



Obrázok 2.11. Diagramatická reprezentácia Boolovej funkcie *XOR*. (A) Ak je funkcia *XOR* reprezentovaná v 2-rozmernom stavovom priestore x_1-x_2, objekty s jednotkovou klasifikáciou nie sú lineárne separovateľné od objektov s nulovou klasifikáciou. (B) Ak znázorníme *XOR* funkciu v 3-rozmernom fázovom priestore x_1-x_2-x_1x_2 ,

potom existuje v tomto priestore nadrovina, ktorá navzájom separuje objekty s rôznou klasifikáciou. Priemet tejto nadroviny do roviny poskytuje kvadratickú čiaru, ktorá separuje objekty, pozri diagramy C a D.

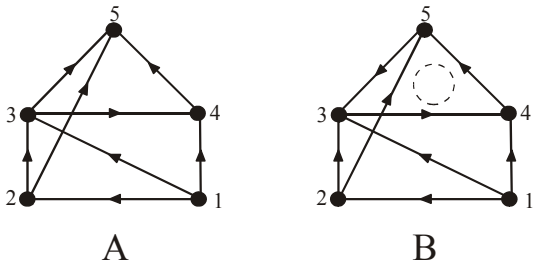
Princíp použiteľnosti logického neurónu pre klasifikáciu Boolových funkcií, ktoré nie sú lineárne separovateľné, spočíva v skutočnosti, že zavedením nových "kvadratických premenných" sa objekty stanú lineárne separovateľnými. Naznačený postup môže byť zovšeobecnený do nasledujúcej vety.

2.2.1 Formálna špecifikácia neurónovej siete

Z našich predchádzajúcich úvah vyplýva, že koncepcia neurónovej siete patrí medzi fundamentálne pojmy teórie neurónových sietí (nielen tých, ktoré obsahujú logické neuróny). Neurónová sieť je definovaná pomocou usporiadanej trojice

$$\mathcal{N} = (G, \boldsymbol{w}, \vartheta) \tag{2.19}$$

kde G je súvislý orientovaný graf, $\boldsymbol{w}$ je matica váhových koeficientov a ϑ je vektor prahových koeficientov.



Obrázok 2.12. Orientované súvislé grafy, ktoré reprezentujú topológiu neurónovej siete. V týchto grafoch vrchol 1 reprezentuje vstupný neurón, vrcholy 2, 3, 4 reprezentujú skryté neuróny a vrchol 5 reprezentuje výstupný neurón. Diagram A znázorňuje acyklický graf, diagram B reprezentuje cyklický graf (vznikol z ľavého grafu A zmenou orientácie hrany 3-4) .

Zatiaľ sme nepoužili čas v explicitnej forme pri popise siete zloženej z logických neurónov. Budeme predpokladať, že čas v neurónovej sieti je diskrétny a je špecifikovaný nezáporným celým číslom, aktivity neurónovej siete v čase t označíme $\boldsymbol{x}^{(t)}$, v čase $t = 0$ vektor $\boldsymbol{x}^{(0)}$ špecifikuje *počiatočné aktivity* neurónovej siete. v. Vzťah (2.5) pre aktivitu i-teho neurónu v čase t je má tvar

$$x_i^{(t)} = s\left(\sum_j w_{ij} x_j^{(t-1)} - \vartheta_i \right) \tag{2.20}$$

kde sumácia prebieha cez neuróny, ktoré v sieti "predchádzajú" i-ty neurón, pričom aktivity týchto neurónov sú z predchádzajúceho času t-1. Pre ilustráciu študujme neurónovú sieť znázornenú na obr. 2.12, kde neurónová sieť je špecifikovaná acyklickým orientovaným grafom. Aktivity jednotlivých neurónov sú určené pomocou (2.20) takto

$$\begin{aligned}
x_1^{(t)} &= \textit{externý vstup} \\
x_2^{(t)} &= s\left(-x_1^{(t-1)} - 0\right) \\
x_3^{(t)} &= s\left(x_1^{(t-1)} + x_2^{(t-1)} - 2\right) \qquad \left(t = 1, 2, \ldots, t_{max}\right) \\
x_4^{(t)} &= s\left(-x_1^{(t-1)} + x_3^{(t-1)} - 1\right) \\
x_5^{(t)} &= s\left(x_2^{(t-1)} + x_3^{(t-1)} + x_4^{(t-1)} - 1\right)
\end{aligned} \tag{2.21}$$

Poznamenajme, že ako dôsledok acykličnosti neurónovej siete je skutočnosť, že pri výpočte $x_i^{(t)}$ potrebujeme poznať aktivity neurónov, ktoré daný neurón predchádzajú a sú z predošlého časového momentu. Táto skutočnosť je hlavným dôvodom prečo sa neurónové siete reprezentované acyklickým orientovaným grafom nazýva *dopredné neurónové siete.*

Binárne aktivity jednotlivých neurónov môžu byť vyjadrené pomocou vektora

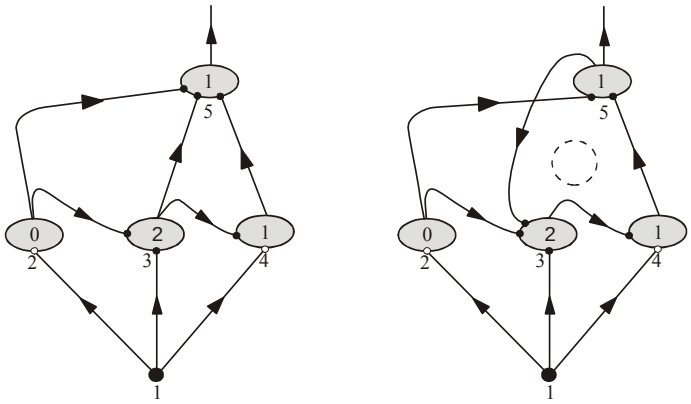
$$\boldsymbol{x} = \boldsymbol{x}_I \oplus \boldsymbol{x}_H \oplus \boldsymbol{x}_O \tag{2.22}$$

kde $\boldsymbol{x}_I$ je vektor aktivít vstupných neurónov, $\boldsymbol{x}_H$ je vektor aktivít skrytých neurónov a $\boldsymbol{x}_O$ je vektor výstupných aktivít. Neurónovú sieť $\mathcal{N}$ môžeme chápať ako funkciu, ktorá zobrazuje vektor aktivít $\boldsymbol{x}^{(t-1)}$ v čase t-1 na vektor aktivít $\boldsymbol{x}^{(t)}$ v čase t

$$\boldsymbol{x}^{(t)} = F\left(\boldsymbol{x}^{(t-1)}; \mathcal{N}\right) \tag{2.23}$$

kde funkcia F obsahuje ako „parameter“ špecifikáciu neurónovej siete $\mathcal{N}$.

Neurónové siete môžeme rozdeliť na dve veľké podskupiny podľa topológie grafu G z (2.19). Ak graf G je acyklický, potom neurónová sieť sa nazýva *dopredná*; v opačnom prípade, ak graf G je cyklický, potom sieť sa nazýva *rekurentná* (pozri obr. 2.13).



Obrázok 2.13. Neurónové siete, ktoré sú špecifikované orientovanými grafmi z obr. 2.12. (A) Dopredná špecifikovaná acyklickým grafom znázorneným diagramom A na obr. neurónová sieť 2.12, (B) rekurentná neurónová sieť špecifikovaná cyklickým grafom B na obr. 2.12.

Pre úplnosť budeme predpokladať, že aktivity neurónov 2-4 pre t=1 sú nulové. Nech vstupné aktivity sú špecifikované binárnym reťazcom dĺžky t_{max}=10, $\boldsymbol{x}_1$=(1101101010), aktivity ostatných neurónov v doprednej neurónovej sieti z obr. 2.13, diagram A sú uvedené v nasledujúcej tabuľke pre časy $1 \le t \le 10$.

t	x_1	x_2	x_3	x_4	x_5
1	1	0	0	0	□ 0
2	1	0	0	0	0
3	0	1	0	0	0
4	1	1	0	0	1
5	1	0	1	0	1
6	0	0	0	0	1
7	1	1	0	0	0
8	0	0	1	0	1
9	1	1	0	1	1
10	0	0	1	0	1

Vo všeobecnosti môžeme konštatovať, že neurónová sieť tvorí "parametrické" (excitačný alebo inhibičný charakter spojov a prahy jednotlivých neurónov) zobrazenie postupnosti vstupných aktivít na nejakú postupnosť výstupných aktivít

$$(1101101010) = \tilde{F}(\square 001110111, \textit{parametre siete}) \tag{2.24}$$

Rekurentné neurónové siete sú určené cyklickým orientovaným grafom, pozri obr. 2.13, diagram B. V tomto prípade môžeme povedať, že neurónová sieť má pamäť, v dôsledku existencie orientovaných cyklov v rekuretných sieťach, niektoré spôsoby závislosti aktivít od aktivít iných neurónov sa môžu neustále opakovať. Napríklad, pri výpočte aktivity neurónu 2 v čase *t* potrebujeme poznať aktivity neurónov 1, 2 a 5 v čase *t*-1, avška v dôsledku existencie cyklu, aktivita neurónu 5 v čase *t*-1 vyžaduje poznať aktivitu neurónov 2 a 4 v čase *t*-2. Z tejto jednoduchej diskusie vyplýva, že aktivita neurónu 5 v čase *t* je určená aktivitami neurónu v časoch *t*-1 a *t*-2. V ďalších krokoch sa môžeme toto „okno do minulosti“ zväčšovať, túto skutočnosť špecifickú pre rekuretné siete nazýva *„pamäť“ rekurentých sietí*. Aktivity neurónov z rekurentnej siete sú určené pomocou (2.19) takto

$$\begin{aligned}
x_1^{(t)} &= \textit{externý vstup} \\
x_2^{(t)} &= s\left(-x_1^{(t-1)} - 0\right) \\
x_3^{(t)} &= s\left(x_1^{(t-1)} + x_2^{(t-1)} + x_5^{(t-1)} - 2\right) \qquad (t = 1,2,\ldots) \\
x_4^{(t)} &= s\left(-x_1^{(t-1)} + x_3^{(t-1)} - 1\right) \\
x_5^{(t)} &= s\left(x_2^{(t-1)} + x_4^{(t-1)} - 1\right)
\end{aligned} \tag{2.25}$$

Pre podobnú sekvenciu vstupných aktivít, aká bola použitá v predchádzajúcom príklade, $\boldsymbol{x}_1$=(1101101010), dostaneme použitím vzťahov (2.22) aktivity neurónov siete, ktoré sú uvedené v nasledujúcej tabuľke

t	x_1	x_2	x_3	x_4	x_5
1	1	0	0	0	□
2	1	0	0	0	0
3	0	0	0	0	0
4	1	1	0	0	0
5	1	0	1	0	1
6	0	0	1	0	0
7	1	1	0	1	0
8	0	0	1	0	1
9	1	1	0	1	0
10	0	0	1	0	1

Podobne ako v predchádzajúcom príklade doprednej neurónovej siete (pozri obr. 2.13, diagram A), aj rekuretná neurónová sieť môže byť interpretovaná ako zobrazenie vstupného reťazca $\boldsymbol{x}_1$=(1101101010) na výstupný reťazec $\boldsymbol{x}_5$=(□000100101), kde symbol □ reprezentuje prázdny znak..

2.3 Konečnostavový stroj [6]

Konečnostavový stroj je znázornený na obr. 2.14., tento stroj pracuje v diskrétnych časových 1, 2,...,*t* , *t*+1 ,... okamžikoch. Obsahuje dve pásky, vtupných symbolov resp. výstupných symbolov, pričom výstupné symboly sú určené na základe vstupných symbolov a vnútorných stavov *s* stroja

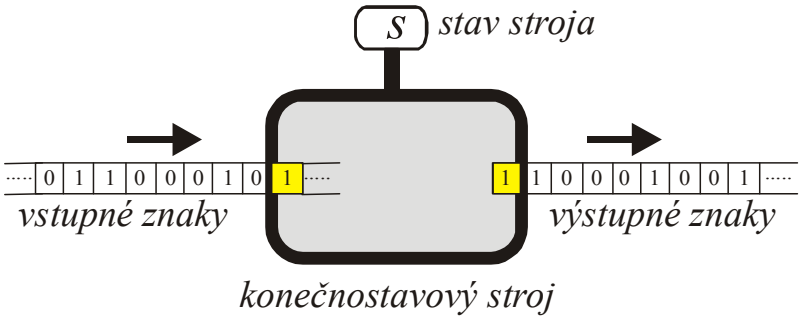
$$stav_{t+1} = f\left(stav_t, vstupný\ symbol_t\right) \tag{2.26a}$$

$$výstupný\ symbol_{t+1} = g\left(stav_t, vstupný\ symbol_t\right) \tag{2.26b}$$

kde funkcie *f* a *g* špecifikujú daný stroj a sú chápané ako jeho základná charakteristika:

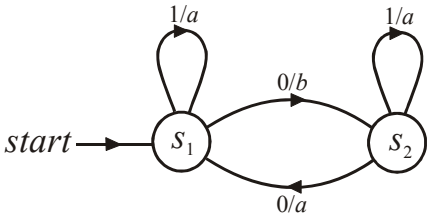
(1) prechodová funkcia f určuje nasledujúci stav na základe aktuálneho stavu a vstupného symbolu,

(2) výstupná funkcia g určuje výstupný symbol na základe aktuálneho stavu a vstupného symbolu.



Obrázok 2.14. Konečnostavový stroj pracuje v diskrétnych časových okamžikoch 1, 2,...,*t*, *t*+1, ,... . Obsahuje dve hlavice, jednu na čítanie vstupnej pásky, pomocou ktorej v danom časovom okamžiku *t* identifikuje vstupný symbol, pomocou výstupnej hlavice vytlačí v čase *t*+1 výstupný symbol. V každom časovom okamžiku *t* je stroj v určitom vnútornom stave *s*. pričom v čase *t*+1 je vnútorný stav určený vnútorným stavom v čase *t* a vstupným symbolom taktiež v čase *t* (pozri vzťahy (2.26a-b)).

Definícia 2.2. Konečnostavový stroj (s výstupom, nazývaný taktiež *Mealyho automat*) je definovaný pomocou usporiadanej šestice $M = \left(S, I, O, f, g, s_{ini}\right)$, kde $S = \left\{s_1, ..., s_m\right\}$ je konečná množina stavov, $I = \left\{i_1, i_2, ..., i_n\right\}$ je konečná množina vstupných symbolov, $O = \left\{o_1, o_2, ..., o_p\right\}$ je konečná množina výstupných symbolov, $f : S \times I \to S$ je prechodová funkcia, $g : S \times I \to O$ je výstupná funkcia a $s_{ini} \in S$ je počiatočný stav.



Obrázok 2.15. Príklad konečnostavového stroja, ktorý obsahuje dva stavy, $S = \left\{s_1, s_2\right\}$, dva vstupné znaky, $I = \left\{0,1\right\}$, dva výstupné znaky, $O = \left\{a, b\right\}$ a počiatočný stav je s_1. Prechodová a výstupná funkcia je špecifikovaná pomocou Tabuľky 2.4

Tabuľka 2.4. Prechodová a výstupná funkcia konečnostavového stroja z obr. 2.15.

stav	*f* *prechodová funkcia*		*g* *výstupná funkcia*	
	0	1	0	1
s_1	s_2	s_1	*b*	*a*
s_2	s_1	s_2	*a*	*a*

Prechodová a výstupná funkcia môžu byť použité na konštrukcia formálneho obvodu, ktorý simuluje daný konečnostavový stroj, pozri obr. 2.16.

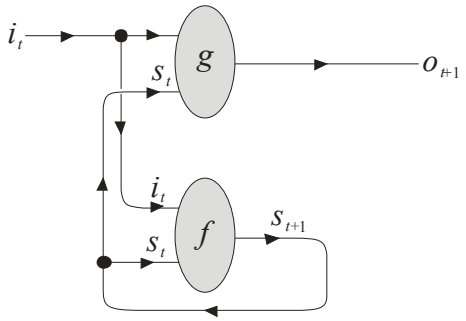
Postupnosti stavov a výstupných znakov pre konečnostavový stroj z obr. 2.15 sú špecifikované Tabuľkou 2.5 pre vstupnú postupnosť znakov (100111010...). Zariadenie môže byť interpretované ako zobrazenie vstupného reťazca znakov na výstupný reťazec znakov

$$G\left(\underbrace{100111010...}_{\text{vstupný reťazec } x}; f, g\right) = \underbrace{\square abaaaabaa...}_{\text{výstupný reťazec } y}$$

kde symbol □ vo výstupnom reťazci znamená „prázdny znak“, znaky výstupného reťazca sú posunuté o jeden časový krok – takt, vzhľadom k vstupnému reťazcu. Zobrazenie *G* obsahuje dve funkcie *f* a *g*, ktoré určujú „topológiu“ konečnostavového stroja. Pre konštrukciu vzťahu medzi neurónovou sieťou a konečnostavového stroja budeme tento prístup špecifikovať takto: nech $\boldsymbol{i} = i^{(1)}i^{(2)}i^{(3)}...i^{(t)}...$, $\boldsymbol{o} = o^{(2)}o^{(3)}o^{(4)}...o^{(t+1)}...$ a $\boldsymbol{s} = s^{(1)}s^{(2)}s^{(3)}...s^{(t)}...$ sú reťazce vstupných znakov, výstupných znakov, resp. stavov stroja (pozri Tabuľku 2.5). Jednotlivé znaky z týchto reťazcov sú navzájom v dvoch funkčných reláciách (pozri obr. 2.16)

$$s^{(t+1)} = f\left(s^{(t)}, i^{(t)}\right) \tag{2.27a}$$

$$o^{(t+1)} = g\left(s^{(t)}, i^{(t)}\right) \tag{2.27b}$$



Obrázok 2.16. Schématická reprezentácia konenčnostavového stroja reprezentovaného prechodovou a výstupnou funkciou *f* resp. *g* (pozri formuly (2.27a-b)).

Prvá formula (2.27a) určuje nasledujúci stav $s^{(t+1)}$ pomocou prechodovej funkcie *f*, vstupného symbolu $i^{(t)}$ a stavu $s^{(t)}$. Podobne, druhá formula (2.27b) určuje nový výstupný symbol $o^{(t+1)}$ pomocou výstupnej funkcie *g,* predchádzajúceho stavu $s^{(t)}$ a vstupného symbolu $i^{(t)}$. Hovoríme, že neurónová sieť je *ekvivalentná* s konenčnostavovým strojom, ak odozva výstupom oboch výpočtových zariadení je identická na ľubovolný rovnaký vstup. Pre ekvivalenciu neurónovej siete s konečnostavovým strojom je nepodstatné akým spôsobom sa vstupné symboly zobrazujú na výstupné symboly, t. j. akým spôsobom prebiehajú výpočty pri konštrukcii výstupného reťazca, podstatné je rovnosť výstupných reťazcov pre rovnaký vstupný reťazec pre obe zariadenia.

Tabuľka 2.5. Postupnosť vstupných znakov, stavov a výstupných znakov pre konečnostavový stroj z obr. 2.15.

vstupný symbol	1	0	0	1	1	1	0	1	0	..
stav	s_1	s_1	s_2	s_1	s_1	s_1	s_1	s_2	s_2	..
výstupný symbol	□	*a*	*b*	*a*	*a*	*a*	*a*	*b*	*a*	*a*

Veta 2.5 [4]**.** Každá neurónová sieť môže byť reprezentovaná ekvivalentným konečnostavovým strojom s výstupom.

Dôkaz tejto vety je konštruktívny a jednoduchý, ukážeme ako pre neurónovú sieť možno zostrojiť jednotlivé zložky z definície 2.2, $M = (S, I, O, f, g, s_{ini})$.

(1) Množina stavov S je vytvorená všetkými možnými binárnymi vektormi $\boldsymbol{x}_H$, $S = \{\boldsymbol{x}_H\}$. Nech neurónová sieť má n_H skrytých neurónov a n_O výstupných neurónov, potom kardinalita (počet stavov) množiny S je 2^{n_H}.

(2) Množina vstupných symbolov je tvorená všetkými možnými binárnymi vektormi $\boldsymbol{x}_I$, $I = \{\boldsymbol{x}_I\}$, kardinalita tejto množiny je 2^{n_I}, kde n_I je počet vstupných neurónov.

(3) Množina výstupných symbolov je tvorená všetkými možnými binárnymi vektormi $\boldsymbol{x}_O$, $O = \{\boldsymbol{x}_O\}$, kardinalita tejto množiny je 2^{n_O}.

(4) Funkcia $f : S \times I \to S$ priradí každému stavu a vstupnému symbolu nový stav. Táto funkcia je špecifikovaná zobrazením (2.23) produkovaného neurónovou sieťou

$$\boldsymbol{x}_H^{(t+1)} = F\left(\boldsymbol{x}_I^{(t)} \oplus \boldsymbol{x}_H^{(t)}; \mathcal{N}\right)$$

Toto zobrazenie priradí dvojici stavu a vstupu v čase *t*-1 nový stav v čase *t*.

(5) Funkcia $g : S \times I \to O$ priradí každému stavu a vstupu výstupný symbol. Podobne ako aj v predchádzajúcom prípade, aj táto funkcia je plne špecifikovaná zobrazením (2.23)

$$\boldsymbol{x}_O^{(t+1)} = \tilde{F}\left(\boldsymbol{x}_I^{(t)} \oplus \boldsymbol{x}_H^{(t)}; \mathcal{N}\right)$$

(6) Počiatočný stav s_{ini} sa obvykle vyberie taký, že aktivity skrytých neurónov sú nulové.

Týmto spôsobom, na základe známej neurónovej siete, sme jednoznačne špecifikovali konečnostavový stroj, ktorý ekvivalentný s danou neurónovou sieťou, každá neurónová sieť môže byť teda reprezentovaná ekvivalentným konečnostavovým strojom, Q.E.D.

Dôkaz inverznej vety k vete 2.5 (t. j. každý konečnostavový stroj môže byť reprezentovaný ekvivalentnou neurónovou sieťou) nie je triviálny, prvý kto ju dokázal bol Minsky v r. 1967 v knihe "*Computation: Finite and Infinite Machines*" [4] pomocou sofistikovaného konštruktívneho dôkazu. Našim cieľom je teda zostrojiť na základe daného konečnostavového stroja takú neurónovú sieť, ktorá je ekvivaletná s konečnostavovým strojom.

Veta 2.6 [4]**.** Každý konečnostavový stroj s výstupom (Mealyho automat) môže byť reprezentovaný ekvivalentnou rekurentnou neurónovou sieťou.

Príklad 2.2. Ilustračný dôkaz tejto vety vykonáme pomocou jednoduchého príkladu konečnostavového stroja so stavovým diagramom znázorneným na obr. 2.15. Tento stroj je určený prechodovou a výstupnou funkciou f a g, (pozri tab. 2.4), ktoré môžeme vyjadriť ako Boolove funkcie pomocou dvoch tabuliek:

(1) Prechodová funkcia $stav_{t+1} = f\left(stav_t, vstupný\ symbol_t\right)$:

stav, vstupný znak	*prechodová funkcia f*
$(s_1,0) \rightarrow (0,0)$	$(b) \rightarrow (1)$
$(s_1,1) \rightarrow (0,1)$	$(a) \rightarrow (0)$
$(s_2,0) \rightarrow (1,0)$	$(a) \rightarrow (0)$
$(s_2,1) \rightarrow (1.1)$	$(a) \rightarrow (0)$

(2) Výstupná funkcia $výstupný\ symbol_{t+1} = g\left(stav_t, vstupný\ symbol_t\right)$:

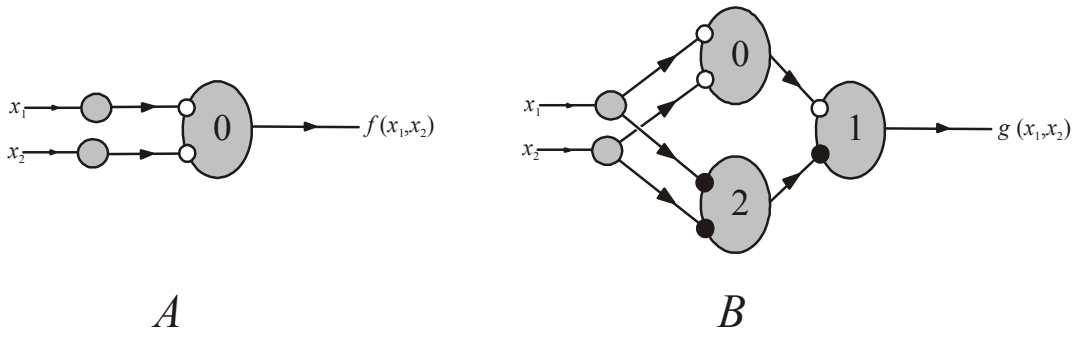
stav, vstupný znak	*výstupná funkcia g*
$(s_1,0) \rightarrow (0,0)$	$(s_2) \rightarrow (1)$
$(s_1,1) \rightarrow (0,1)$	$(s_1) \rightarrow (0)$
$(s_2,0) \rightarrow (1,0)$	$(s_1) \rightarrow (0)$
$(s_2,1) \rightarrow (1.1)$	$(s_2) \rightarrow (1)$

To znamená, že obe funkcie *f* a *g* sú špecifikované ako Boolove funkcie

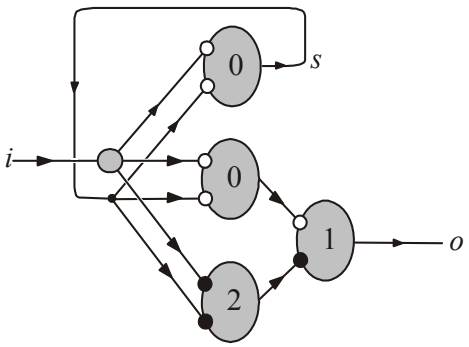
$$f\left(x_1, x_2\right) = \neg x_1 \wedge \neg x_2$$

$$g\left(x_1, x_2\right) = \left(\neg x_1 \wedge \neg x_2\right) \vee \left(x_1 \wedge x_2\right)$$

Reprezentácia týchto funkcií pomocou logických neurónov má tvar (pozri kapitolu 2.1) je znázornená na obr. 2.17. Substitúciou týchto neurálnych reprezentácií do schémy z obr. 2.16 dostaneme rekurentnú neurónovú sieť (pozri obr. 2.18), ktorá je ekvivalentná s konečnostavovým strojom so stavoým diagramom znázorneným na obr. 2.15.



Obrázok 2.17. Znázornenie Boolových funkcii *f* a *g* z príkladu 2.1, ktoré špecifikujú binárnu reprezentáciu prechodovej resp. výstupnej funkcie.



Obrázok 2.18. Rekurentná neurónová sieť reprezentujúca konečnostavový stroj z obr. 2.15. Táto sieť vznikla substitúciou neurónových reprezentácii Boolových funkcií *f* a *g* z obr. 2.17 do diagramu na obr. 2.16, tento diagram reprezentuje všeobecný konečnostavový stroj so zadanými funkciami *f* a *g*.

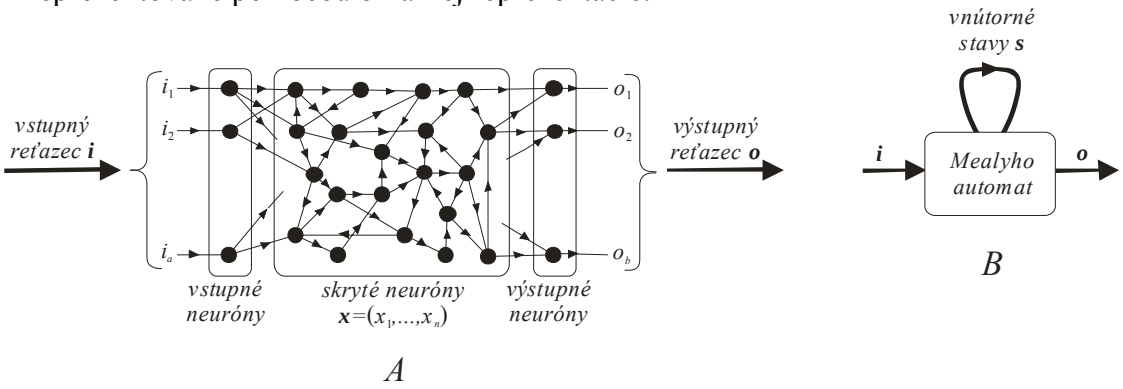
Poznamenajme, že tento jednoduchý príklad nám dostatočne podrobne reprezentuje všeobecný postup konštrukcie rekurentnej neurónovej siete, ktorá je ekvivalentná s daným konečnostavovým strojom, ktorý je špecifikovaný funkciami f a g. V prvom kroku zostrojíme neurálnu reprezentáciu týchto funkcií metódou špecifikovanou pre Boolove funkcie v kapitole 2.1 a v druhom kroku tieto neurálne reprezentácie substituujeme do všeobecného diagramu z obr. 2.16 konenčnostavového stroja špecifikovaného funkciami f a g. Týmto príkladom sme zavŕšili dôkaz vety 2.6, ukázali sme konštruktívnu metódu ako zostrojiť rekurentnú neurónovú sieť, ktorá je ekvivalentná s požadovaným konečnostavovým strojom.

Záverom je potrebne zdôrazniť, že neurónové siete zložené z logických neurónov sú silným výpočtovým prostriedkom: (1) neurónové siete reprezentované acyklickým grafom sú univerzálnym aproximátorom Boolových funkcií a (2) medzi konečnostavovým strojom a neurónovou sieťou existuje princíp ekvivalencie, ľubovoľný konečný automat môže byť simulovaný rekurentnou neurónovou sieťou, a podobne, ľubovoľná neurónová sieť môže byť simulovaná konečnostavovým strojom. Tieto dve vlastnosti sú prototypom výpočtovej mohutnosti pre iné typy neurónových sietí používajúce neuróny so zložitejšou aktivačnou funkciou. Podstatným ohraničením neurónových sietí je fakt, že spoje medzi neurónmi a ich rozdelenie na excitačné a inhibičné a tiež aj prahové faktory sú určené voľbou architektúry siete. Ináč povedané, siete s logickými neurónmi nie sú schopné adaptačného procesu (učenia), Boolova funkcia (alebo Boolove funkcie v prípade viacerých výstupných neurónov) určená sieťou je nemenná v celom priebehu používania siete.

Vety 2.5 a 2.6 umožňujú študovať klasický problém konekcionizmu (neurónový sietí), aký je vzťah medzi ***subsymbolickou reprezentáciou*** (neurálnou, ktorá je reprezentovaná obrazcami zloženými z aktivít neurónov) a ***symbolickou reprezentáciou*** (na ktorej je založená klasická umelá inteligencia):

(1) Pomocou vety 2.5 každá subsymbolická neurónová sieť môže byť pretransformovaná na symbolický konečnostavový stroj (Mealyho automat, pozri obr. 2.19), pričom symboly môžu byť tvorené pomocou prirodzených čísel, ktoré sú priradené binárnym vektorom vstupu, výstupu a aktivitám skrytých neurónov.

(2) Pomocou vety 2.6 každý symbolický konečnostavový stroj (Mealyho automat) môže byť pretransformovaný na subsymbolickú neurónovú sieť, ktorá je ekvivaletná s daným konečnostavoým strojom. Symboly z reprezentácie konečnostavového stroja sú reprezentované pomocou binárnej reprezentácie.



Obrázok 2.19. Diagram A znázorňuje neurónovú sieť, ktorá pomocou vrstvy skrytých neurónov transformuje vstupné binárne signály na výstupné binárne signály. Stav skrytých neur ónov je reprezentovaný binárnym vektorom $\boldsymbol{x}$, ktorý obsahuje aktivity jednotlivých skrytých neurónov. Poznamenajme, že vrstva skrytých neurónov je reprezentovaná cyklickým grafom, t. j. má rekurentný charakter, ktorý umožnuje študovať aj procesy vyžadujúce pamäť. Tento diagram tvorí základom *subsymbolickej reprezentácie*, ktorá je synonymom pre konekcionistický (neurálny) prístup k štúdiu procesov (kognitívnych) prebiehajúcich v mozgu. Diagram B reprezentuje *symbolický prístup* k štúdiu procesov v mozgu, ktorý je „prepisom" konekcionisticej schémy znázornenej na diagrame A pomocou konečnostavového stroja nazývanej Mealyho automat. Vstupné ***i*** a výstupné

o veličiny a vnútorné stavy ***s*** sú v tomto prístupe reprezentované symbolmi, ktoré v najjednoduchšom prístupe môžu byť reprezentované prirodzenými číslami, ktoré sú priradené binárnym vektom ***i***, ***o***, resp. ***x***, z diagramu A.

Naznačený postup pre prechod od subsymbolickej k symbolickej reprezentácii a naopak pomocou neurónových sietí s logickými neurónmi a konečnostavových strojov, môže slúžiť štúdiu vlastností týchto modelov a prechodu medzi nimi.

2.4 Konštrukcia binárnych obvodov

Logické neuróny môžu byť použité na konštrukciu jednoduchých binárnych obvodov [4], ich kombináciou môžeme zostrojiť už zložité zariadenia, akými sú napr. rôzne sčítačky, násobičky a podobne.

2.4.1 Časový posunovač

Časový posunovač je taký binárny obvod, ktorý posunie tok signálu o niekoľko vopred zvolených jednotiek. Študujme jednoduchý obvod znázornený na obr. 2.12, diagram A, ktorý realizuje posun o jednotku (jednotkový posunovač). Aktivitu neurónu v čase t môžeme vyjadriť takto

$$y^{(t)} = t\left(x^{(t-1)} - 1\right) \tag{2.27}$$

Vstupný signál v čase t sa transformuje na výstupný signál v čase t+1

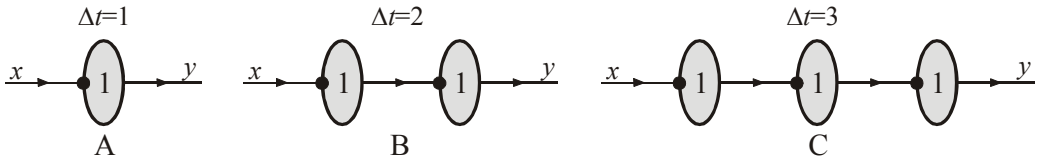
$$\begin{aligned} x^{(t)} = 0 &\Rightarrow y^{(t+1)} = t(-1) = 0 \\ x^{(t)} = 1 &\Rightarrow y^{(t+1)} = t(0) = 1 \end{aligned} \tag{2.28}$$

Nech sekvencie (dĺžky 10) vstupných signálov má tvar (0011000111), potom výstupný symbol má tvar sekvencie (dĺžky 11) (#0011000111), kde prvý znak # je prázdny znak, pretože v počiatočnom čase t = 1 nevieme špecifikovať výstupný signál, pozri Tab. 2.4

Tabuľka 2.4. Vstupná a výstupná sekvencia pre jednotkový časový posunovač

čas	1	2	3	4	5	6	7	8	9	10	11	12
vstup x	0	0	1	1	0	0	0	1	1	1	..	..
výstup y	#	0	0	1	1	0	0	0	1	1	1	..

Sériovým zapojením jednotkových posunovačov dostaneme posunovače o dve, tri, .. časové jednotky, pozri diagramy B a C na obr. 2.12.

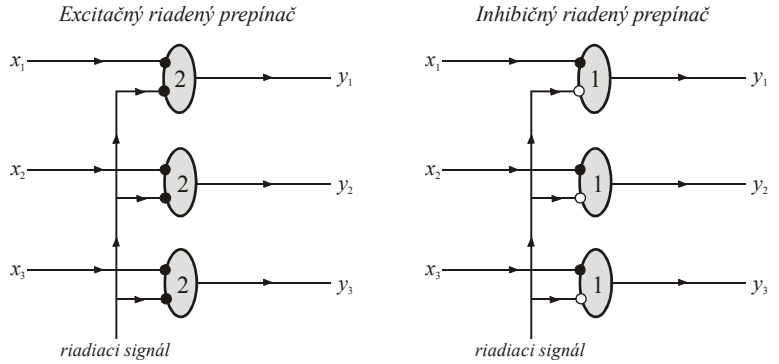


Obrázok 2.12. Neurónové siete realizujúce posun vstupného signálu o 1, 2 a 3 jednotky.

2.4.2 Riadený prepínač

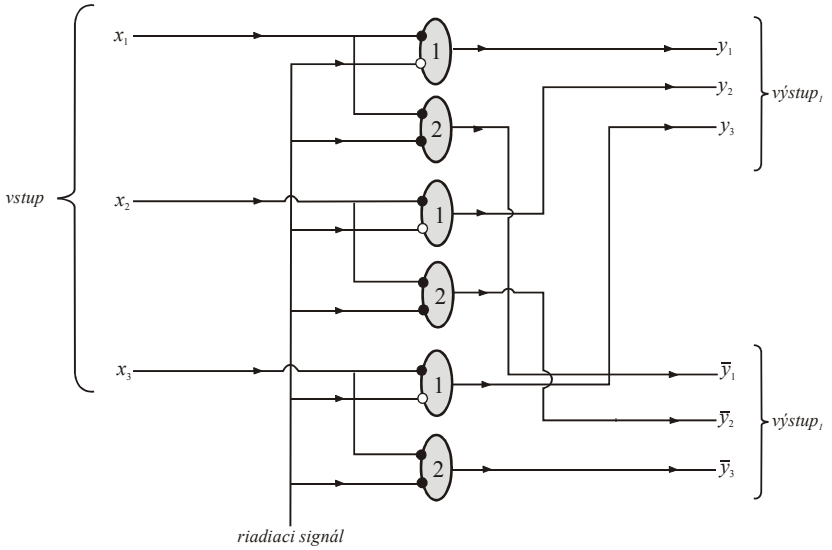
Riadený prepínač pomocou riadiaceho signálu zapína alebo vypína tok signálu. Tak ak v excitačnom riadenom prepínači je riadiaci signál jednotkový, potom prepínač prepúšťa signál na výstup prepínača. V alternatívnom prípade, pre inhibične riadený prepínač, nulový

riadiaci signál spôsobuje prenos signálu na výstup. Tieto dva prepínače sú znázornené na obr. 2.13.



Obrázok 2.13. Znázornenie neurónových sietí, ktoré sú excitačný resp. inhibičný riadiaci prepínač.

Kombináciou excitačného a inhibičného prepínača zostrojíme riadené rozdvojovacie zariadenie, pozri obr. 2.14.



Obrázok 2.14. Neurónová sieť riadeného rozdvojovacieho zariadenia. Ak riadiaci signál je jednotkový (nulový), potom vstup je riadený na 1. výstup (2.výstup). Obrazne môžeme povedať, že riadiaci signál pôsobí ako „prepínač“ do 1. alebo 2. polohy.

2.4.3 Sumátor dvoch binárnych čísel

Sumátor binárnych čísel rieši úlohu sčítania dvoch binárnych čísel – reťazcov rovnakej dĺžky

$$\begin{array}{r} 110011 \\ 101010 \\ \hline 1011101 \end{array}$$

Realizácia tohto problému sa môže uskutočniť úplne analogický procesu sčítania dvoch dekadických čísel, t. j. musíme používať techniku prenosu „zostatku“ (angl. *carry*) do nasledujúcej polohy. Tento princíp si vysvetlíme na všeobecnom príklade

$$\begin{array}{r} \alpha_n \ldots \alpha_2 \alpha_1 \\ \beta_n \ldots \beta_2 \beta_1 \\ \hline \gamma_{n+1} \gamma_n \cdots \gamma_2 \gamma_1 \end{array} \qquad (2.29)$$

sčítania dvoch binárnych čísel $\alpha_n \ldots \alpha_2 \alpha_1$ a $\beta_n \ldots \beta_2 \beta_1$ dĺžky n. Zostatok δ_i je rekurentne definovaný takto:

$$\delta_1 = 0 \tag{2.30a}$$

$$\delta_2 = \begin{cases} 1 & (ak\ \alpha_1 + \beta_1 = 2) \\ 0 & (inák) \end{cases} \tag{2.30b}$$

$$\delta_{i+1} = \begin{cases} 1 & (ak\ \alpha_i + \beta_i + \delta_i \geq 2) \\ 0 & (ak\ \alpha_i + \beta_i + \delta_i \leq 1) \end{cases} \qquad (i = 2, 3, ..., n) \tag{2.30c}$$

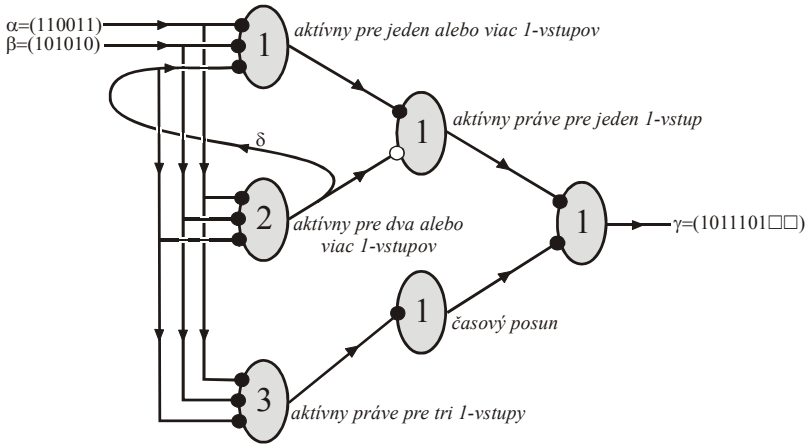
kde (2.20a-b) je inicializácia postupného výpočtu zostatkov. Pre ilustračný príklad zo začiatku tejto podkapitoly zostatky sú špecifikované binárnym reťazcom dĺžky 7 tvaru $\delta = 1000100$. Schému (2.19) prepíšeme do nového tvaru

$$\begin{array}{r} \alpha_n ... \alpha_2 \alpha_1 \\ \beta_n ... \beta_2 \beta_1 \\ \delta_{n+1} \delta_n ... \delta_2 \delta_1 \\ \hline \gamma_{n+1} \gamma_n \cdots \gamma_2 \gamma_1 \end{array} \tag{2.31}$$

Potom výsledok $\gamma = \gamma_{n+1} ... \gamma_2 \gamma_1$ je určený vzťahmi

$$\gamma_i = \begin{cases} 1 & (ak\ \alpha_i + \beta_i + \delta_i = 1 \vee 3) \\ 0 & (ak\ \alpha_i + \beta_i + \delta_i = 0 \vee 2) \end{cases} \quad (i = 1, 2, ..., n) \tag{2.32a}$$

$$\gamma_{n+1} = \delta_{n+1} \tag{2.32b}$$



Obrázok 2.15. Neurónová sieť sumátora dvoch binárnych čísel, ktorá obsahuje šesť prahových neurónov.

Príklad zo začiatku tejto podkapitoly môže byť prepísaný do tvaru

$$\begin{array}{r} 110011 \\ 101010 \\ 1000100 \\ \hline 1011101 \end{array}$$

kde výsledný reťazec bol určený formulami (2.32). Jednotlivé kroky z tohto príkladu sú rozpísané takto:

0. krok. $\delta_1 = 0$
1. krok. $\alpha_1 = 1, \beta_1 = 0, \delta_2 = 0, \boxed{\gamma_1 = 1}$
2. krok. $\alpha_2 = 1, \beta_2 = 0, \delta_3 = 1, \boxed{\gamma_2 = 0}$

3. krok. $\alpha_3 = 0, \beta_3 = 0, \delta_4 = 0, \boxed{\gamma_3 = 1}$

4. krok. $\alpha_4 = 0, \beta_4 = 1, \delta_5 = 0, \boxed{\gamma_4 = 1}$

5. krok. $\alpha_5 = 1, \beta_5 = 0, \delta_6 = 0, \boxed{\gamma_5 = 1}$

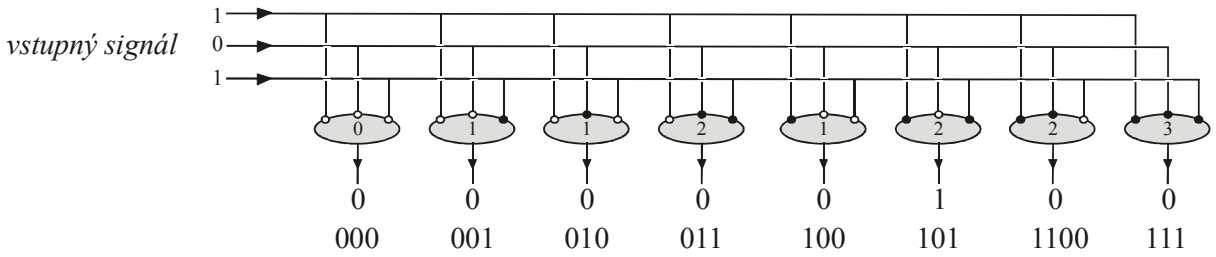
6. krok. $\alpha_6 = 1, \beta_6 = 1, \delta_7 = 1, \boxed{\gamma_6 = 0}$

7. krok. $\alpha_7 = 0, \beta_7 = 0, \quad \boxed{\gamma_7 = 1}$

Implementácia sumátora je znázornená na obr. 2.15.

2.4.4 Binárny paralelný dekóder

Budeme študovať paralelný dekodér pre tri binárne číslice. Predpokladáme, že informácia do dekódera prichádza prostredníctvom impulzov cez tri paralelne vlákna, pozri obr. 2.16. Dekóder obsahuje osem nezávislých elementov – neurónov, z ktorých každý dekóduje jednu možnú binárnu kombináciu prichádzajúcich signálov.



Obrázok 2.16. Binárny paralelný dekóder, ktorý obsahuje osem prahových neurónov, ktoré dekódujú separátne vstupný signál od kombinácie 000 až po kombináciu 111.

2.5 Záver

Neurónové siete zložené z logických neurónov sú silným výpočtovým prostriedkom: (1) neurónové siete reprezentované acyklickým grafom sú univerzálnym aproximátorom Boolových funkcií a (2) ľubovolný konečný automat môže byť simulovaný rekurentnou neurónovou sieťou. Vlastnosť (1) môže byť podstatne zosilnená tvrdením, že každá Boolova funkcia môže byť reprezentovaná doprednou 3-vrstvovou neurónovou sieťou. Podstatným ohraničením logických neurónov je, že klasifikujú len Boolove funkcie, ktoré sú lineárne separovateľné. Táto skutočnosť môže byť prekonaná pomocou logických neurónov vyššieho rádu alebo 3-vrstvovými neurónovými sieťami.

Literatúra

[1] Kolář, J., Štěpánková, O., Chytil, M.: *Logika, algebry a grafy*. SNTL, Praha, 1989.

[2] Kvasnička, V., Pospíchal, J.: *Matematická logika*, Vydavateľstvo STU, Bratislava, 2006.

[3] McCulloch, W. S., Pitts, W. H.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, **5** (1943), 115-133.

[4] Minsky, M. L.: *Computation. Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

[5] Minsky, M. and Papert, S.: *Perceptrons. An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, 1969.

[6] Molnár Ľ., Češka, M., Melichar, B.: *Gramatiky a jazyky*. Bratislava, Alfa, 1987.

3. prednáška

Darwinovská evolúcia ako algoritmus

3.1 Úvodné poznámky

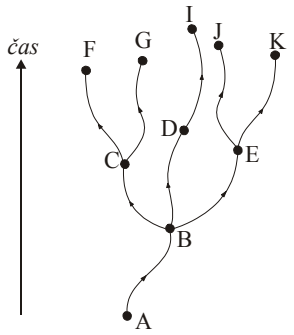
Evolučná teória spočíva v jednoduchej myšlienke, že život je v čase premenlivý a mladšie formy života nasledujú za staršími formami. Táto myšlienka už existovala pred Darwinom, ale až on a jeho nasledovníci ju rozvinuli do podoby umožňujúcej vysvetliť nielen rôznorodosť foriem života, ale aj adaptáciu živých organizmov vzhľadom k meniacemu sa okoliu. Podľa filozofujúceho biológa Mayra [40], koncepcia evolúcie spočíva v kombinácii nasledujúcich piatich princípoch:

1. Živé organizmy sú *časovo premenné*, podliehajú evolučným zmenám.
2. Evolučné zmeny majú charakter *vetvenia*. Súčasné druhy sú nasledovníci predchodcov existujúcich v minulosti. Podobne, súčasný druh môže tvoriť predchodcu jednému alebo viacerým budúcim druhom (pozri obr. 1).
3. Nové druhy vznikli tak, že *populácia sa rozštiepila* na izolované podpopulácie, ktoré potom navzájom evolučne divergovali za vzniku nových druhov.
4. Evolučné zmeny sú *postupné – graduálne*. Niektoré organizmy, ktoré sa odlišujú od svojich rodičov, sú schopné prežitia. Z týchto organizmov si niektoré vytvoria svoju vlastnú podpopuláciu zloženú z podobných organizmov, v ktorej sa zachovajú ich špecifické odlišnosti.
5. Mechanizmom adaptívnej zmeny je *prírodný výber*.

Darwinovým veľkým úspechom [7] bolo, že pomerne rýchlo presvedčil svojich súčasníkov o existencii evolúcie, ktorej priebeh sa dá názorne reprezentovať stromom života (pozri obrázok 1). Bol evolučným gradualistom, ktorý veril v nahromadenie (akumuláciu) malých zmien v priebehu dlhých časových úsekov, ktoré sú interpretované ako evolučná zmena. Toto jeho „gradualistické“ evolučné presvedčenie bolo všeobecne prijaté až v prvej polovici 20. storočia hlavne zásluhou trojice matematikov – štatistikov Fishera, Wrighta a Haldanea, ktorý spojili populačnú genetiku s evolučnou teóriou za vzniku modernej evolučnej teórie nazývanej *neodarwinizmus*. Dôležitosť izolácie populácie vznikajúceho nového druhu zostala kontroverznou otázkou evolučnej teórie ešte o mnoho dlhšie. Tento pohľad na mechanizmus vzniku nových druhov sa stal všeobecne akceptovaným až zásluhou prác Mayra z polovice 20. storočia.

Súčasná evolučná biológia stojí pred dvoma dôležitými problémami, ktoré sa snaží vysvetliť. Prvým problémom je *neobyčajná rôznorodosť* foriem života na našej planéte. Až darwinizmus poskytol vysvetlenie tejto skutočnosti, kde vedľa seba existuje mnoho druhov, ktoré sú morfologicky zásadne odlišné a ktoré na prvý pohľad nemajú medzi sebou skoro nič spoločné. Druhý, nemenej dôležitý problém je *adaptácia*. Štruktúrovaná zložitosť organizmov, ich adaptácia na životné prostredie, je rovnako prekvapujúca ako rôznorodosť organizmov. Zmyslové orgány sú klasickým príkladom zložitej a jemne naladenej adaptácie.

Medzi najväčšie úspechy darwinizmu patrí prírodovedecké vysvetlenie adaptácie pomocou prírodného výberu založeného na prežívaní lepšie prispôsobených organizmov (s väčším fitnes) a dedení získaných vlastností. Evolučné zmeny sú určené variáciami v populácii, rozdielmi vo fitnes spôsobenými týmito variáciami a genetickým prenosom. Adaptívne zmeny sa uskutočňujú aj napriek často zmieňovanej skutočnosti, že mechanizmy spôsobujúce variabilnosť populácie (hlavne mutácie) nesúvisia priamo s výhodnosťou adaptívnych zmien. Tak napr. adaptívny posun farby peria vtáka prebieha postupne, v priebehu mnohých generácií a závisí od kumulatívneho výberu. Ak výber má vysvetliť adaptáciu, potom musí byť kumulatívny. Inovácia je výsledkom dlhej postupnosti epizód výberu, pretože pravdepodobnosť požadovanej náhlej adaptačnej zmeny prostredníctvom mutácie je zanedbateľne malá. Súčasne s adaptačnými zmenami prebieha aj proces diverzifikácie populácie, adaptácia na nové prostredie a postupná divergencia jednotlivých podpopulácií od seba.



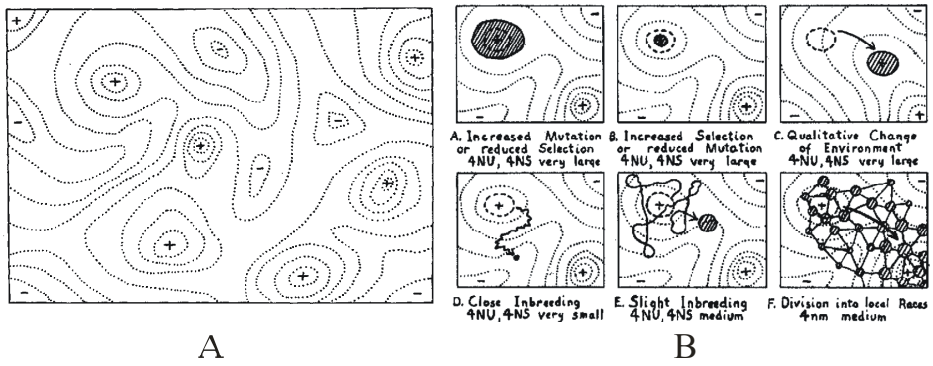
Obr. 3.1. Schematické znázornenie vzniku nových druhov (označených písmenami A, B, ...) v priebehu evolúcie. Druh A je predchodcom druhu B, druh B je spoločným predchodcom nových druhov C, D a E. Priebeh evolúcie sa dá názorne reprezentovať pomocou tohto „stromu života“.

Určité pochybnosti pri akceptovaní Darwinovej evolučnej teórie neustále pretrvávajú [4] v súvislosti s vysvetlením mechanizmov vzniku nových zložitostí. Tieto zložitosti majú komplexnú štruktúru, pričom ich architektúra je obvykle ireducibílneho charakteru, t.j. nemôžeme odstrániť ani jeden element štruktúry bez straty funkčnosti celej zložitosti. V tomto smere existuje v súčasnosti najviac námietok proti evolučnej teórii, ako principiálne neschopnej vysvetliť pomocou prírodného výberu a adaptácie vznik nových kvalít, kreativitu živej prírody. Evolučný vznik nových kvalít, nových zložitostí, je obvykle sprevádzaný aj veľkou prestavbou genotypu organizmu, zmenou jeho modulárnej štruktúry a interakcie jednotlivých modulov medzi sebou. Gradualizmus klasického darwinizmu (aj neodarwinizmu) vychádza implicitne z predpokladu, že štruktúra genotypu je nemenná, môže sa meniť len zloženie jednotlivých alel genotypu, bez zmeny jeho modulárnej štruktúry a pod. Tento problém stojí pred moderným darwinizmom [5,30,31,49] ako dôležitá neriešená úloha, ktorej riešenie už bude vyžadovať úzku spoluprácu s informatikmi modelujúcimi *in-silico* evolúciu. Evolučný nárast zložitosti je formálne zložitý problém, ktorý sa nedá riešiť len na verbálnej kvalitatívnej úrovni, ale vyžaduje spoluprácu s teoretickými biológmi a informatikmi, ktorí sa venujú počítačovým simuláciám evolúcie.

V tejto sekcii uvedieme v chronologickom poradí vybrané názory, koncepcie na darwinovskú evolúciu, ktoré podľa názoru autorov zohrali významnú úlohu pri vzniku a rozvíjaní univerzálneho darwinizmu. Ako informatici pracujúci v umelej inteligencii a v kognitívnej vede, budeme preferovať také názory, ktoré viedli k pohľadu na evolúciu ako na univerzálne platný optimalizačný algoritmus, ktorého „materiálna“ realizácia - implementácia je nezávislá od pôvodnej biologickej realizácie evolúcie.

3.2 Wrightov pohľad na evolúciu ako na optimalizáciu [51]

V prírodnom výbere má dominantné postavenie koncepcia *fitnes* organizmov, ktorý je špecifikovaný ako relatívna schopnosť prežitia a reprodukovania sa v danom *prostredí* a v danej *populácii*. Vo všeobecnosti, fitnes je chápaný nielen ako atribút genotypu ale aj populácie. Tento postulát odráža naše predstavy o fitnes, ako o veličine, ktorá kvantifikuje nielen prežívanie a schopnosť reprodukcie jednotlivých organizmov, ale aj ich vzájomnú interakciu v populácii. Vlastnosti určujúce fitnes sa aspoň čiastočne prenášajú z rodičov na potomkov, t.j. sú zafixované v genotype jedincov populácie. Potomkovia rodičov s väčším fitnes budú početnejší a taktiež aj "silnejší" ako potomkovia "slabších" rodičov.



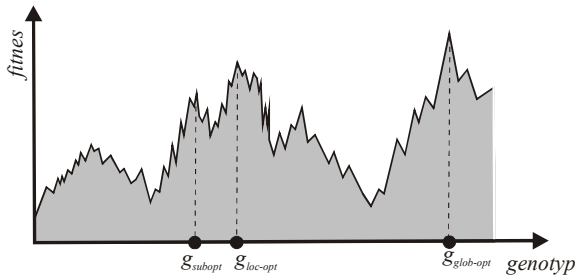
Obr. 3.2. Jeden z najslávnejších obrázkov neodarwinizmu, ktorý pochádza od amerického populačného genetika a evolučného biológa Sewalla Wrighta [51]. (A) Znázorňuje povrch fitnes funkcie obsahujúci množstvo lokálnych extrémov (miním a maxím). Evolúciu charakterizoval ako adaptívny proces (v matematickej terminológii "optimalizačný proces"), kde genotyp populácie je adaptovaný tak, že dosiahne lokálne (snáď aj globálne) optimum – maximum na povrchu fitnes funkcie. (B) Wright navrhoval rôzne mechanizmy, ako prekonať hlboké údolia medzi dvoma maximami. Svoju pozornosť v tomto smere sústredil hlavne na možnosť genetického driftu malej populácie, ktorá je dočasne izolovaná od hlavnej populácie a kde drift nadobúda význam s poklesom významu prirodzeného výberu. V takomto evolučnom „taviacom kotli“ dochádza k neočakávaným zmenám genotypu, čo sa môže prejaviť aj preskokom z jedného maxima na druhé maximum. Týmto spôsobom vysvetľoval Wright mechanizmus toho (pozri motto tohto príspevku), ako sa genotyp populácie môže presunúť z nižšieho vrcholu na vyšší vrchol povrchu fitnes.

Koncepcia *povrch fitnes* (pôvodne „adaptive fitness landscape“) (pozri obr. 2, diagramy A a B) pochádza od jedného zo zakladateľov neodarwinizmu Sewalla Wrighta, ktorý v r. 1932 navrhol [51] túto koncepciu, ktorá predstavuje jednoduchý heuristický prístup ako vizualizovať proces evolúcie populácie organizmov. Evolúcia populácie môže byť pomocou tohto obrázku interpretovaná ako pohyb populácie smerom k vrcholu s najväčším fitnes. V dôsledku existencie mutácií, tento pohyb populácie na povrchu fitnes má stochastický charakter. V tejto súvislosti sa zdôrazňuje, že práve vďaka mutáciám, organizmy populácie majú šancu „preskúmať“ väčšiu oblasť na povrchu fitnes a tak potenciálne dosiahnuť v procese evolúcie globálne maximum na povrchu fitnes. To znamená, že evolučný proces môžeme charakterizovať ako *optimalizačný proces*. Formálne túto skutočnosť môžeme vyjadriť pomocou diskrétneho optimalizačného problému

$$\boldsymbol{g}_{opt} = arg \max_{\boldsymbol{g}\in\{a,b,..\}^n} f\left(\boldsymbol{g}\right) \tag{1}$$

kde $f(\boldsymbol{g})$ je fitnes funkcia, ktorá priradí každému genotypu $\boldsymbol{g}$ zostrojenému nad abecedou $\{a,b,..\}$ kladné reálne číslo - fitnes, výsledné riešenie $\boldsymbol{g}_{opt}$ reprezentuje genotyp nachádzajúci sa na globálnom maxime povrchu funkcie fitnes. Evolučným cieľom organizmov je optimálna adaptácia v danom prostredí. Vlastnosti povrchu fitnes (extrémna multimodálnosť) určujú ako

ťažko sa požadované optimálne riešenie nájde „prirodzeným“ evolučným procesom. Vychádzajúc z tejto metafory povrchu fitnes, Wright formuloval aj ciele evolúcie takto „*Ústredným problémom evolúcie je navrhnúť taký mechanizmus, ktorým organizmy môžu nájsť spojitú cestu z nižšie položeného vrcholu na vyššie položený vrchol*“ (pozri obrázok 3).



Obr. 3.3. Povrch fitnes bol prvý krát použitý v Darwinovej teórii americkým biológom Sewalom Wrightom v r. 1932 [51], ktorý charakterizoval evolúciu ako optimalizačný problém (v jeho terminológii adaptačný problém) nad povrchom fitnes, ktorý je charakterizovaný funkciou f . Výsledný genotyp populácie je priradený bodu – lokálne optimálnemu genotypu $\boldsymbol{g}_{loc\text{-}opt}$ − s vysokou (ale nie najväčšou) hodnotou fitnes. Globálne optimálne riešenie označené $\boldsymbol{g}_{glob\text{-}opt}$ reprezentuje také riešenie, ktoré ma maximálnu funkčnú hodnotu na povrchu fitnes. Obrazne povedané, Darwinova evolúcia musí mať prostriedky ako prekonať hlboké údolie pri pohybe z suboptimálneho riešenia $\boldsymbol{g}_{subopt}$ k blízkemu lokálno-optimálnemu riešeniu $\boldsymbol{g}_{loc\text{-}opt}$. Pretože Darwinova evolúcia je stochastický optimalizačný algoritmus prebiehajúci na povrchu fitnes funkcie, pravdepodobnosť získania globálne optimálneho stavu nie je stopercentná, nie vždy je toto riešenie dosiahnuteľné z danej počiatočnej populácie a v reálnom čase.

3.3 Hollandov genetický algoritmus

Genetické algoritmy boli vynájdené informatikom Johnom Hollandom [25] počiatkom 70-tych rokov minulého storočia. Po určitej nábehovej 10-ročnej perióde rozpakov a mlčania v komunite informatikov, stali sa jednou z najbúrlivejšie rozvíjajúcich sa oblastí informatiky a umelej inteligencie [15,37]. Možno konštatovať, že spolu s neurónovými sieťami tvoria jadro novovznikajúcej oblasti nazývanej *počítačová inteligencia*, ktorá je schopná riešiť praktické inžinierske problémy z informačných technológií, ktoré vyžadujú vysoký stupeň „inteligencie[1]“. V tejto podkapitole ukážeme základné princípy genetických algoritmov.

Genetický algoritmus nemanipuluje priamo so živými organizmami, ale s ich genetickou reprezentáciou. Pôvodný Hollandov názov pre túto reprezentáciu bol „chromozóm“, žiaľ táto pôvodná terminológia vyvolávala medzi biológmi rôzne asociácie s čisto biologickým významom tohto termínu, ktoré nesúviseli s problematikou genetického algoritmu, preto budeme tieto informačné entity nazývať „*replikátory*“. Postulujme, že replikátory majú schopnosť sa rozmnožovať – replikovať, a to buď unárne (asexuálne) alebo binárne (sexuálne). Pri unárnej replikácii sa zúčastňuje procesu jeden replikátor - rodič, zatiaľ čo pri binárnej replikácii sa zúčastňujú procesu dva replikátory – rodičia. Rodičia (rodič) sú kvázináhodne vybraní z populácie v závislosti na ich fitnes (replikátory s väčším fitnes s väčšou pravdepodobnosťou vstupujú do replikácie) a produkujú nové replikátory - potomkov. Budeme rozlišovať tieto tri zložky replikačného procesu:

(1) *selekcia* rodičov,
(2) *replikácia* rodičov, pričom vznikajú potomkovia, a
(3) *návrat* potomkov do populácie.

[1] V modernej informatike sa pod termínom „inteligentný system” rozumie taký „agent“, ktorého kognitívny orgán (reprezentovaný *subsymbolickou* neurónovou sieťou alebo *symbolickým* expertným systémom) je schopný adaptácie (učenia) pri zmene prostredia alebo cieľov, ktoré sa spontánne vynárajú pri pohybe v tomto prostredí.

V prvom kroku binárnej replikácie sa vyberú pomocou stochastického operátora O_{select} z populácie P dva replikátory $\boldsymbol{x}_1^{old} = O_{select}(P)$, $\boldsymbol{x}_2^{old} = O_{select}(P)$ tak, že pravdepodobnosť ich výberu je úmerná ich fitnes. V druhom kroku použitím stochastického operátora replikácie O_{repli} z rodičov dostaneme potomkov $\boldsymbol{x}_1^{new}$ a $\boldsymbol{x}_2^{new}$.

$$\left(\boldsymbol{x}_1^{new}, \boldsymbol{x}_2^{new}\right) = O_{repli}\left(\boldsymbol{x}_1^{old}, \boldsymbol{x}_2^{old}\right)$$

V unárnej (asexuálnej) replikácii sa na tvorbe potomkov podieľa len jeden replikátor – rodič. Formálne, tento proces môžeme vyjadriť takto

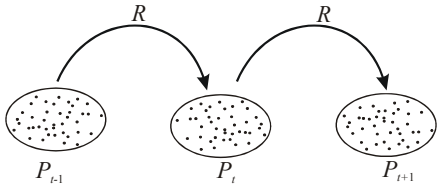
$$\boldsymbol{x}^{old} = O_{select}(P) \text{ a } \boldsymbol{x}^{new} = O_{repli}\left(\boldsymbol{x}^{old}\right)$$

V prvom kroku je kvázináhodne vybraný rodič, v druhom aplikáciou stochastického operátora reprodukcie O_{repro} sa z replikátora - rodiča vytvorí replikátor – potomok. Prečo hovoríme, že operátor reprodukcie replikátorov je stochastický? Tento operátor z rodiča (rodičov) vyrobí potomka (potomkov) určitým procesom, ktorý nie je striktne deterministický, takže potomkovia nie sú presnými kópiami rodičov, ale majú určitú malú variabilitu vzhľadom k rodičom, čím je zabezpečená variabilita populácie. V treťom kroku sa rieši problém návratu potomka do populácie. Pseudopascalovský kód takto určenej Darwinovej evolúcie je daný nasledujúcim algoritmom:

Algoritmus 1.

```
P:=náhodne vygenerovaná populácia replikátorov;
t:=0;
while t<tmax do
begin t:=t+1;
      Q:=∅;
      while |Q|<|P| do
      begin x1:=Oselect(P); x2:=Oselect(P);
            (x1',x2'):=Orepli(x1,x2);
             Q:=Q∪{x1',x2'};
      end;
      P:=Q;
end;
```

Algoritmus je inicializovaný náhodnou generáciou populácie `P`. Evolúcia populácie je obsiahnutá vo vnútri **`pokiaľ-urob`**-cyklu, ktorý sa opakuje $t_{\max}$-krát. V rámci jednej epochy sa vytvorí nová populácia `Q`, ktorá sa tvorí opakovanou replikáciou dvoch rodičovských genotypov, ktoré majú vysoký fitnes. Nová populácia `Q` nahradí pôvodnú populáciu `P` (pozri obrázok 4).

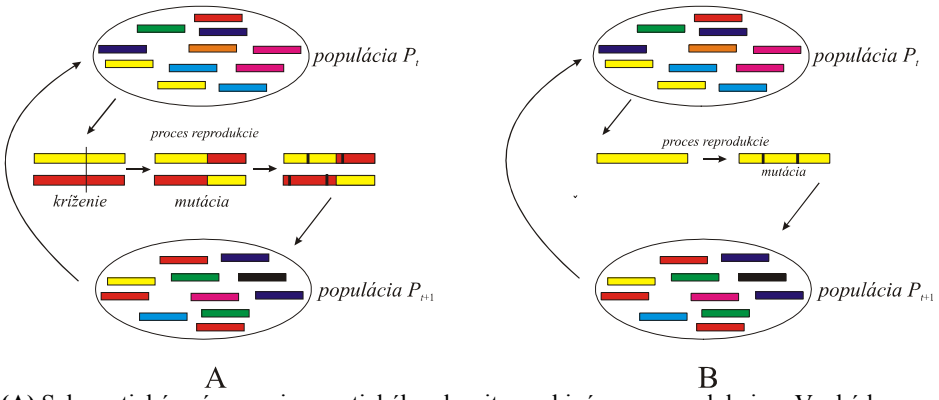


Obr. 3.4. Schematické znázornenie rekurzívneho charakteru Darwinovej evolúcie, nová populácia sa vytvára z predchádzajúcej populácie aplikáciou stochastického operátora reprodukcie R.

Darwinova evolúcia môže byť interpretovaná ako *rekurentný proces*, v ktorom nasledujúca populácia je vytvorená reprodukciou predchádzajúcej populácie (pozri obrázok 4)

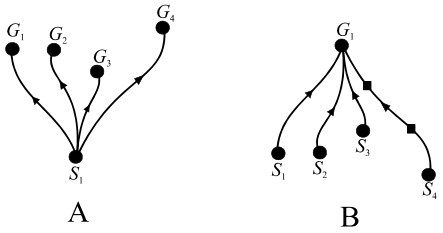
$$P_{t+1} = R(P_t)$$

Funkcia R kvázináhodne (vzhľadom k fitnes replikátorov) priradí k populácii P_t nasledujúcu populáciu P_{t+1}. Cieľom tejto rekurzie – evolúcie darwinovského systému – je spontánna emergencia replikátorov s vysokým fitnes, ktoré sa vyvinú z počiatočných, náhodne generovaných replikátorov, tvoriacich populáciu P_0. Hybnou silou evolúcie darwinovského systému je (1) prírodný výber a (2) spontánne mutácie. Z tohto pohľadu môžeme povedať, že evolúcia systému prebieha na hrane chaosu a poriadku. Mechanizmy selekcie najlepšie prispôsobených replikátorov pomocou prírodného výberu a neustála zmena replikátorov v populácii dôsledkom stochastických mutácií pri replikačnom procese, sú hlavným zdrojom evolúcie systému, postupného zvyšovania priemerného fitnes v celej populácii (pozri obrázok 5).



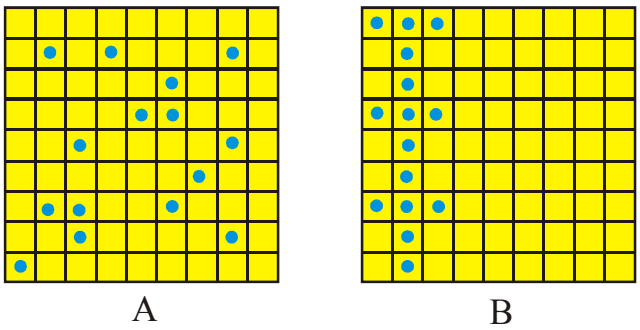
Obr. 3.5. (A) Schematické znázornenie genetického algoritmu s binárnou reprodukciou. Vychádza sa z aktuálnej populácie P_t , z ktorej opakovane vyberáme dvojice rodičovských replikátorov (pravdepodobnosť výberu je úmerná ich fitnes). V reprodukčnom procese ich replikácia je založená na tzv. krížení, ktoré spočíva v tom, že pri kopírovaní replikátorov najprv využívame pôvodnú časť replikátora 1. rodiča a potom (s určitou pravdepodobnosťou) aj časť replikátora 2. rodiča. Vzniknuté replikátory ešte prechádzajú procesom mutácie, pri ktorom s malou pravdepodobnosťou sa zavedú do replikátora chyby (na obrázku sú znázornené malými vertikálnymi čiarami). Takto vytvorené replikátory vytvárajú novú populáciu potomkov, ktorá, keď obsahuje rovnaký počet replikátorov ako rodičovská populácia, sa osamostatní a podobný proces reprodukcie sa opakuje. **(B)** Genetický algoritmus s unárnou reprodukciou, replikátory potomkov sa vytvárajú len jednoduchým kopírovaním rodičovských replikátorov spolu s málo pravdepodobnými mutáciami.

Zatiaľ sme bližšie nešpecifikovali reprodukčný proces. Holland zaviedol pre binárnu (sexuálnu) reprodukciu proces kríženia, ktorý spočíva v tom, že z dvoch rodičovských replikátorov sa kopírovaním vytvoria dva nové replikátory potomkov, avšak pri tomto kopírovacom procese prvá časť sa kopíruje z 1. rodiča, zatiaľ čo nasledujúca časť sa kopíruje z 2. rodiča. Táto možnosť použitia časti replikátora 2. rodiča pri vytváraní nového replikátora sa využíva len s určitou pravdepodobnosťou (preto môže nastať situácia, že replikátor potomka je kópiou replikátora len jedného rodiča). Súčasťou reprodukčného procesu sú aj tzv. mutácie, ktoré pri tvorbe replikátorov potomkov vytvárajú s malou pravdepodobnosťou náhodné zmeny, pozri obr. 5, diagram A. Genetický algoritmus s unárnou (asexuálnou) reprodukciou je zjednodušenou verziou predošlej verzie s binárnou reprodukciou, kríženie je v tomto prípade vylúčené (jeden rodič sa nemôže sám so sebou krížiť), pozri obrázok 5, diagram B.



Obr. 3.6. Dva rôzne priebehy evolúcie v genetickom algoritme. Diagram A znázorňuje prípad, v ktorom sledujeme rôzne priebehy evolúcie pre jeden východiskový bod S_1, v dôsledku stochastického charakteru evolúcie z tohto východiskového stavu môžeme dostať rôzne finálne stavy G_1, G_2, G_3 a G_4. Predpovedať, ktorý z týchto finálnych stavov dostaneme je skoro nemožné pre enormnú kombinatoriálnu zložitosť problému (obrazne môžeme povedať, že nepoznáme také výpočtové zariadenie, ktoré by bolo schopné v reálnom čase vykonať vyčerpávajúcim spôsobom rekonštrukciu všetkých možných evolučných trajektórií. Diagram B odpovedá opačnej situácií, poznáme konečný stav evolúcie a hľadáme možné evolučné trajektórie s počiatočnými stavmi S_1, S_2, S_3 a S_4, ktoré by viedli k danému finálnemu stavu G_1. Podobne, ako v predchádzajúcom inverznom prípade, aj táto rekonštrukcia evolúcie so zadaným finálnym stavom je enormne zložitý kombinatoriálny problém. Určité záchytné body (reprezentované plnými štvorcami na trajektórii vedúcej z počiatočného stavu S_4) nám môžu podstatne zjednodušiť rekonštrukciu evolučnej trajektórie; tie trajektórie, ktoré neobsahujú tieto body sú automaticky vylúčené z našich úvah.

Genetický algoritmus má silne stochastický charakter. V dôsledku existencie náhodných mutácií, kvázináhodného výberu do procesu reprodukcie (a taktiež aj možnosti náhodného vyhynutia replikátorov populácie s vysokým fitnes pred ich vstupom do reprodukčného procesu), evolučné algoritmy nemajú deterministický charakter, t.j. nemôžeme presne predvídať evolúciu populácie, pozri obrázok 6.



Obr.3.7. Dva ilustračné príklady zložitostí rôzneho typu. Diagram A znázorňuje náhodne rozmiestnené body na mriežke, diagram B znázorňuje body umiestené do tvaru trojkríža.

Na druhej strane, jednoduchý organizmu, napr. baktéria je adaptívna zložitosť, pretože vykazuje tak znaky zložitosti, ako aj adaptivity na zmenu prostredia. Existencia adaptívnej zložitosti je *hlavným kritériom existencie života*. Ak na nejakej planéte nájdeme nejaké hmotné štruktúry, ktoré môžeme klasifikovať ako adaptívnu zložitosť, potom môžeme deklarovať s vysokým stupňom istoty, že na planéte existuje (alebo existoval) život. Cieľom každej evolučnej teórie je vysvetliť vznik adaptívnej zložitosti, t.j. experimentálny fakt, ktorý niektorí pokladajú za dôkaz existencie Tvorcu.

Budeme študovať jednoduchý ilustračný príklad zložitosti pomocou 15 objektov, ktoré sú rozmiestnené na mriežke 9×9, pozri obr. 7. Diagram A nám znázorňuje typické náhodné rozmiestnenie objektov na mriežke, môžeme konštatovať, že pravdepodobnosť vzniku podobného "chaotického" rozmiestnenia je skoro jednotková, p_A=1-ε, kde ε je malé kladné číslo. Potom pravdepodobnosť vzniku rozmiestnenia B (alebo podobného posunom alebo rotáciou) bude p_B=ε. Informačný obsah nejakého obrazca, ktorý vzniká s pravdepodobnosťou

p je definovaný vzťahom $I = -ln_2\, p$. Potom informačný obsah našich obrazcov A a B má tieto hodnoty

$$I_A = -ln_2\,(1-\varepsilon),\;\; I_B = -ln_2\,\varepsilon \tag{2a}$$

Pre jednoduchosť odhadneme len asymptotické hodnoty týchto informačných obsahov pre malé kladné ε, dostaneme

$$\lim_{\varepsilon\to 0_+} I_A = 0\;,\; \lim_{\varepsilon\to 0_+} I_B \to \infty \tag{2b}$$

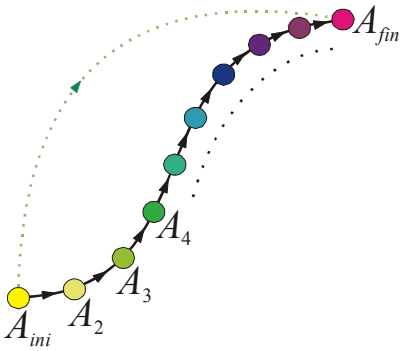
To znamená, že v súhlase s našim očakávaním, obrazec A má nulový informačný obsah, zatiaľ čo obrazec B má skoro nekonečný informačný obsah. K lepšiemu pochopeniu tohto výsledku predstavme si nasledujúcu situáciu: Prechádzame sa po morskom pobreží na neznámom ostrove a nájdeme obrazec A, na ktorom sú objekty - kamene náhodne rozmiestnené. Na základe našich predchádzajúcich úvah vieme, že tento obrazec má nulový informačný obsah, nedáva nám žiadnu správu o svojom tvorcovi. Naopak, ak by sme našli obrazec B, na ktorom kamene sú umiestnené pravidelne do tvaru troj-kríža, získali sme správu o veľkom informačnom obsahu, vieme skoro so 100% istotou, že na ostrove musí existovať tvorca obrazca.

Tento príbeh je parafrázou argumentácie anglikánskeho teológa Williama Paleyho z prelomu 18. a 19. storočia [44], ktorý vo svojej knihe „*Natural Theology*“ dôvodil, že každá adaptívna zložitosť musí mať svojho tvorcu (každé hodiny majú svojho hodinára – tvorcu). Tento argument sa neustále opakuje v rôznych verziách ako argument proti evolučnej teórii už skoro dvesto rokov. V súčasnosti je zmodernizovaný do podoby tzv. ireducibílnej zložitosti [4], ktorej príkladom môžu byť rôzne zložité orgány vnímania fungujúce len vtedy, ak obsahujú všetky svoje časti v správnom nastavení. Ak chýba čo len jedna súčasť tejto ireducibílnej zložitosti, stáva sa totálne nefunkčnou[2]. Keď rozmýšľame zjednodušene evolucionisticky, „nedokončený“ orgán vnímania nepredstavuje pre organizmus žiadnu evolučnú výhodu, čiže nemôže existovať ani selekčný tlak na jeho vznik.

Nedávno sa diskutoval v evolučnej psychológii a v kognitívnej vede podobný problém evolučného pôvodu „univerzálnej gramatiky“, pôvodne postulovanej americkým lingvistom Noamom Chomskym [6] v 60. rokoch minulého storočia, ako nutnej štruktúry našej mysle (mozgu) k tomu, aby bol schopný vysvetliť akvizíciu jazyka malým dieťaťom počúvaním jednoduchých gramatických viet, bez nutnosti poznania gramatických pravidiel danej reči. Evolúcia nevytvorila schopnosť ľudskej reči úplne od počiatku v priebehu ostatných niekoľkých miliónov rokov, ale použila materiál, ktorý už bol k dispozícii, ktorý bol podstatne evolučne starší, ako práve vznikajúci jazyk. Evolúcia využíva už existujúce štruktúry pre nové, často prekvapujúce ciele. Anatomická štruktúra mozgu humanoidných opíc je veľmi podobná anatomickej štruktúre ľudského mozgu, medzi opicami a ľuďmi môžu byť pomerne presne identifikované mnohé vzájomne si odpovedajúce oblasti mozgu. Oblasti ľudského mozgu, ktoré sú označené ako za centrum reči, u opíc sú identifikované ako oblasti pre kontrolu svalov tváre a pre analýzu zvukového vstupu. Táto skutočnosť asi podstatne uľahčila evolučné „predrôtovanie“ týchto oblastí pre účely vzniku ľudskej reči. Z týchto úvah vyplýva, že ľudský jazykový inštinkt nie je dielom „zázračnej“ mutácie, alebo povedané poetickým jazykom Dawkinsa, *náhla inšpirácia slepého hodinára riadiaceho evolúciu* [9], ale dôsledkom niekoľko sto milión rokov, v ktorých evolúcia experimentovala so živočíšnou kogníciou. Adaptívny význam ľudského jazyka je nespochybniteľný – umožňuje hovoriť. Spolupráca pri lovení, plánovanie, koordinovanie aktivít, podieľanie sa na spoločných

[2] Klasickým „úsmevným“ príkladom neredukovateľnej zložitosti je pasca na myši, ktorá obsahuje 5 súčastí: drevenú doštičku, pružinu, drôtový rám, drôtovú závlačku a plechovú poistku. Ak chýba z týchto súčastí čo len jedna časť, pasca sa stáva nefunkčnou.

úlohách, sociálne väzby, manipulácia alebo podvádzanie, toto všetko môže mať úžitok zo zvyšovania schopnosti komunikovať ľudským jazykom.



Obr. 3.8. Znázornenie evolučnej tvorby komplexného orgánu (napr. oka stavovcov), proces je inicializovaný evolučným vznikom jednoduchého „protoorgánu“ A_{ini} , cez prechodné riešenia A_i, až po konečné riešenie A_{fin}. Zložitosť orgánu postupne rastie, dva evolučne susedné orgány sú si podobné, evolučný prechod $A_i \rightarrow A_{i+1}$ je dobre interpretovateľný pomocou mutácie genotypu organizmu. Každý orgán má špecifikovanú úlohu, ktorá určuje aj selekčný tlak na jeho vznik, tieto úlohy sa môžu v priebehu evolúcie podstatne meniť. Preto môžeme evolučnú úlohu vzniku komplexného orgánu formulovať ako zložitú optimalizačnú úlohu pohyblivého cieľa. Počiatočný a konečný stav je spojený čiarou, ktorá reprezentuje možnosť priameho „evolučného skoku“ z A_{ini} na A_{fin}. Táto evolučná zmena by vyžadovala zásadnú prestavbu genotypu, ktorá je jednou mutáciou nerealizovateľná, preto pravdepodobnosť takého evolučného skoku je skoro nulová.

Záver vyplývajúci z obr. 8 môžeme zovšeobecniť tak, že evolučný vznik zložitých živočíšnych orgánov neprebiehal priamočiaro za vzniku hotového orgánu požadovanej zložitosti, ale postupne využíval[3] staršie štruktúry organizmu, ktoré boli prispôsobené na nové potreby. Týmto spôsobom je zabezpečený neustály selekčný tlak vedúci v konečnom dôsledku na vznik komplexného orgánu prostredníctvom medzistupňov, ktorých význam môže byť veľmi vzdialený významu konečného orgánu (pozri obrázok 8).

Niekoľko poznámok k problémom reprodukcie a dedičnosti, ktorých významnou časťou je tzv. embryogenéza, čo je spôsob vzniku a vývoja nových jedincov zo zárodkov. Budeme rozlišovať dva druhy embryogenézy:

(1) *Epigenetická embryogenéza*, kde nový organizmus vzniká delením jednej "zárodočnej" bunky, tento druh embryogenézy sa vyskytuje na našej planéte pri reprodukcii biologických systémov. Môže byť zjednodušene chápaná ako *recept* na prípravu koláča, kde je uvedené nielen to čo koláč obsahuje ale aj jeho príprava.

(2) *Preformačná embryogenéza*, kde nový organizmus je reprezentovaný zárodkom, ktorý obsahuje všetky detaily dospelého organizmu. Tento typ embryogenézy môžeme formálne prirovnať k plánu budovy alebo stroja, kde sú podrobne uvedené všetky detaily. Existencia tohto druhu embryogenézy je pre nás hypotetická, na našej planéte sa nevyskytuje. Snáď v budúcnosti sa preformačná

[3] Podľa francúzskeho molekulárneho biológa F. Jacoba [28], darwinovská evolúcia organizmov so zložitou hierarchickou štruktúrou postupuje tak, že prebieha dominantne len na najvyšších hierarchických úrovniach, pričom nižšie hierarchické úrovne sú v podstate evolučne uzavreté – invariantné. Evolúcia môže zasahovať do nižších hierarchických úrovní len veľmi vzácne a aj to len formou tzv. „bricolage“ (fušerstva), t.j. používa štruktúry, ktoré sú na danej úrovni bezprostredne k dispozícii. Napríklad, biochémia bunky bola evolučne uzavretá už pred miliardami rokov, po vzniku mnohobunkových organizmov evolúcia len veľmi vzácne zasahuje na tejto nízkej hierarchickej úrovni (preto môže existovať univerzálne platný vedný odbor biochémia). V prípade fatálnej potreby, evolúcia sa pokúša aj na tejto úrovni pomocou „bricolage“ riešiť vzniklý problém. Takouto „bricolage“ môže byť aj evolučný vznik univerzálnej gramatiky, keď evolúcia použila „predpripravené“ štruktúry v mozgu, ktoré pôvodne vykonávali iné funkcie. V súčasnej evolučnej teórii je „bricolage“ pokladaná za jeden z hlavných mechanizmov vzniku zložitých orgánov zo starších štruktúr, ktoré mohli pôvodne hrať úlohu celkom odlišnú od výsledku „bricolage“. Niektorí evoluční biológovia idú tak ďaleko, že evidentnú existenciu „bricolage“ vo fenotype druhu pokladajú za jeden z hlavných dôkazov evolučného vzniku druhov.

embryogenéza bude vyskytovať aj na našej planéte v súvislosti s reprodukciou umelých robotických systémov.

Pre evolúciu je primárne dôležitá tzv. *reverzibilita embryogenézy*, t.j. vzťah medzi genotypom a fenotypom, genotyp $\rightleftarrows$ fenotyp. Formálne povedané, či sa genotyp jednoznačne zobrazuje na fenotyp a naopak (t.j. či z fenotypu vieme dekódovať genotyp organizmu). Tento vzťah medzi genotypom a fenotypom je veľmi dôležitý pri diskusii takej evolúcie, ktorá umožňuje prenos získaných vlastností na potomkov (lamarckovská evolúcia). Preformačná embryogenéza umožňuje prenos získaných vlastností do genotypu, pretože vzťah medzi genotypom a fenotypom je jedno-jednoznačný. Táto reverzibilita neexistuje pre epigenetickú embryogenézu. Ak použijeme metaforu, podľa ktorej je genotyp niečo ako recept na prípravu koláča, potom z výslednej štruktúry koláča sa nám nepodarí jednoznačne zostrojiť recept podľa ktorého bol vytvorený. Z tohto dôvodu lamarckovská evolúcia (dedenie získaných vlastností) nemôže existovať na našej planéte, kde živé systémy používajú výhradne len epigenetickú embryogenézu. Niekde vo vesmíre môžu existovať živé systémy, ktoré využívajú preformačnú embryogenézu, potom je možné, že medzi týmito živými systémami prebieha lamarckovská evolúcia založená na prenose získaných vlastností z rodiča na potomkov.

Dawkins poukázal vo svojej eseji [8] na fakt, ktorý sa často ignoruje, že aj keď pripustíme možnosť lamarckovskej evolúcie, ktorá predpokladá genetický prenos získaných vlastností z rodičov na potomkov, neexistencia prírodného výberu[4] spôsobuje malú špecifičnosť evolúcie adaptívnych zložitostí. Ak aj doplníme Lamarckovu evolúciu o postulát prírodného výberu, táto „zmeska“ darwinovskej a lamarckovskej evolúcie je priveľmi „oportunistická, vždy vyberá len najbližšie optimálne riešenie, čo spôsobuje „zamrznutie“ evolučnej trajektórie v lokálnom minime. Lamarckovská evolúcia s prírodným výberom už môže poskytovať adaptívne zložitosti, ale tieto sú ďaleko od globálne – optimálnych riešení. Budú vznikať „evolučné kreatúry“, ktoré evolučne „zamrzli“ v nejakom suboptimálnom riešení bez šance ďalšieho evolučného vylepšovania.

3.4 Dennettov pohľad na evolúciu ako na algoritmus

Dennett v knihe „*Darwin's Dangerous Idea*“ [11] anticipoval Dawkinsove názory na univerzálnosť Darwinovej teórie tak, že zdôraznil jej algoritmický charakter (pozri odsek venovaný genetickému algoritmu), ktorý je nezávislý od materiálnej realizácie. Evolúcia, chápaná ako algoritmus, môže byť aplikovaná na rôznorodé systémy, v ktorých sme schopní identifikovať replikátory a ktorých podstata nemusí byť biologická. Dôvod prečo Dennett zdôrazňoval algoritmický charakter darwinovskej evolúcie, jej univerzálnosť v nezávislosti na materiálnej realizácii replikátorov bol ten, že chcel položiť solídne základy Dawkinsovej memetiky [10], v ktorej bola použitá paradigma Darwinovej evolúcie k vysvetleniu

[4] Lamarckova evolučná teória [42] je založená na nasledujúcich dvoch postulátoch: (1) Orgány používaním (nepoužívaním) zvyšujú svoju perfektnosť (zanikajú), (2) získané vlastnosti sa prenášajú geneticky z rodičov na potomkov. Takto formulovaná evolučná teória je nepostačujúca na vznik adaptívnych zložitostí, pretože neobsahuje prírodný výber. Predpoklady o rastúcej perfektnosti orgánov častým používaním a o možnosti genetického prenosu získaných vlastností nie sú dostatočne špecifické k tomu, aby vznikali adaptívne zložitosti. Naviac, ako ukazujú simulačné výpočty realizované pomocou modifikovaného genetického algoritmu, aj keď doplníme Lamarckovu evolúciu o prírodný výber, prenos získaných vlastností z rodičov na potomkov zapríčiňuje vznik „unáhlenej“ konvergencie evolúcie, konečné riešenia sú obvykle pomerne vzdialené od globálnych riešení, ktoré sú (ako postuloval už Wright) cieľom darwinovskej evolúcie. Obrazne povedané, lamarckovská evolúcia je veľmi oportunistická, využíva vždy najbližšie optimálne riešenie, čo môže podstatne negatívne ovplyvniť priebeh evolučnej trajektórie, jej „zamrznutie“ v nejakom lokálnom riešení.

niektorých fenoménov šírenia kultúry. Pretože sa jedná o abiotickú oblasť aplikácie myšlienok Darwinovej evolúcie, považoval Dennett za potrebné najprv sa vyrovnať s otázkou jej aplikovateľnosti mimo biologické systémy. U nás sa filozofickými aspektmi evolúcie systematicky zaoberá Dušan Gálik [21,22].

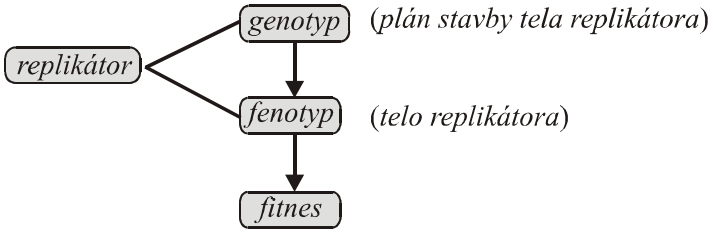
3.5 Darwinovský systém

Najvšeobecnejšia formulácia základnej idey univerzálneho darwinizmu je uskutočnená pomocou koncepcie ***darwinovského systému*** založeného na nasledujúcich dvoch postulátoch (určité náznaky tejto všeobecnej formulácie už boli vykonané pri popise Hollandovho genetického algoritmu):

(1) *Darwinovský systém sa skladá z populácie replikátorov* – jedincov/objektov, ktoré za určitých vhodných podmienok sú schopné *replikácie* - rozmnožovania. Replikačný proces spočíva v „kopírovaní“ jedincov do populácie, pričom toto „kopírovanie“ sa uskutočňuje s *určitými malými chybami*. Z tohto predpokladu nepresnej replikácie (s malými chybami) vyplýva, že *populácia nie je homogénna*, obsahuje jedincov – replikátory, ktoré vykazujú malú variabilitu.

(2) Každý replikátor populácie je ohodnotený funkciou *fitnes*, ktorá vyjadruje schopnosť replikátora prežiť a úspešne vstupovať do replikačného procesu. *Prírodný výber* v darwinovskom systéme spočíva v tom, že jedinci populácie nie sú vyberaní do replikačného procesu náhodne, ale s *pravdepodobnosťou úmernou ich fitnes* (hovoríme, že výber sa deje kvázináhodne).

V definícii darwinovského systému sa vyskytuje niekoľko pojmov, ktoré musia byť bližšie objasnené. Pojem „replikátor“ je potrebné chápať ako relatívne samostatnú entitu, ktorá existuje v nejakom prostredí a vytvára spolu s ostatnými replikátormi populáciu. Najjednoduchší prístup k implementácii replikátora je jeho stotožnenie s nejakou informáciou, ktorá kóduje jeho „telo“. Z tohto dôvodu môžeme využívať biologickú terminológiu a rozlišovať dve rôzne špecifikácie replikátora, jeho *genotyp* a *fenotyp*. Genotyp je informácia o architektúre replikátora, zatiaľ čo fenotyp je „telo“ replikátora[5]. Vzťah medzi genotypom a fenotypom nebudeme bližšie špecifikovať, obvykle je silne závislý na typu darwinovského systému. Čo bude pre nás dôležité, je skutočnosť, že fitnes replikátora je určený schopnosťou fenotypu prežívať v danom prostredí a vstupovať do výhodných interakcií s inými replikátormi z populácie, pozri obr. 9.



Obr. 3.9. Replikátor je charakterizovaný tak genotypom, ako aj fenotypom. Fitnes replikátora je určený schopnosťou fenotypu prežívať v danom prostredí a vstupovať do výhodných interakcií s inými replikátormi z populácie.

[5] Podobne, ako vo fyzike mikrosveta, kde častice majú duálny častico-vlnový charakter, môžeme aj fenotyp a genotyp chápať ako duálne charakteristiky replikátora. V určitých situáciách pod replikátorom rozumieme jeho genotyp, zatiaľ čo v iných situáciách pod replikátorom myslíme jeho fenotyp.

Proces replikácie jedinca – replikátora je formálne chápaný ako kopírovanie jeho genotypu a vytvorenie nového fenotypu určeného kopírovaným genotypom. To znamená, že fenotyp – organizmus replikátora – môžeme chápať ako nosič (vehikel) genotypu, ktorý umožňuje jeho replikáciu. Pre zjednodušenie našich úvah chápeme proces replikácie len ako kopírovanie genotypu, pričom tento proces kopírovania je „fyzicky“ uskutočnený fenotypom replikátora. Je potrebné poznamenať, že sa jedná o veľmi silnú idealizáciu, ale umožní nám zaviesť pomerne jednoduchú algoritmizáciu univerzálneho darwinizmu (pozri algoritmus 1).

Musíme však podotknúť, že môžu existovať darwinovské systémy, kde odlíšenie fenotypu od genotypu neplatí, kde sa genotyp kopíruje – replikuje iným zariadením, ako vlastným fenotypom. Dobrým príkladom tejto situácie sú biologické a počítačové vírusy, ktoré k vlastnej replikácii využívajú systémy v ktorých parazitujú.

Postulujme, že replikátor je reprezentovaný svojím *genotypom* $\boldsymbol{x}$, ktorý, ako už bolo poznamenané, obsahuje zakódovanú informáciu o konštrukcii replikátora (pripomeňme si Dawkinsove prirovnanie genotypu k receptu na koláč). *Populácia* replikátorov je množina genotypov

$$P = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$$

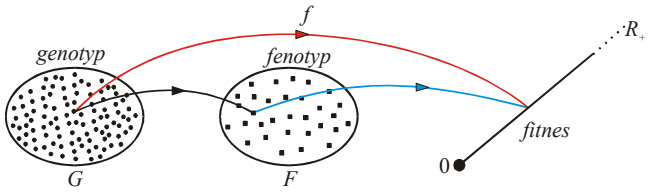
Vzájomný vzťah medzi triádou „genotyp – fenotyp – fitnes“ je reprezentovaný postupnosťou dvoch zobrazení (pozri obr. 10)

$$G \xrightarrow{\text{fenotyp}} F \xrightarrow{\text{fitnes}} [0,\infty)$$

kde G je množina všetkých genotypov a F je množina všetkých fenotypov. To znamená, že prvotnou veličinou je genotyp, ktorý je zobrazený na fenotyp, potom fenotyp je zobrazený na fitnes (reprezentovaný kladným reálnym číslom). Prvé zobrazenie priradí každému genotypu telo – organizmus, nazývaný fenotyp. V biológii toto zobrazenie predpokladá existenciu procesu nazývaného embryogenéza, t. j. tvorba organizmu zo zárodku – genotypu. Skrátená forma dvojstupňového zobrazenia má tvar (pozri obr. 10)

$$G \xrightarrow{f} [0,\infty)$$

kde nové zobrazenie f vznikne zložením dvoch zobrazení, zobrazenia genotypu na fenotyp a zobrazenia fenotypu na fitnes.



Obr. 3.10. Znázornenie jednoduchých zobrazení genotypu na fenotyp a fenotypu na fitnes, a zloženého zobrazenia genotypu priamo na fitnes.

V tejto súvislosti si môžeme položiť otázku, prečo je užitočné explicitne uvažovať fenotyp ako sprostredkovateľa medzi genotypom a fitnes. Táto otázka je plne legitímna z pohľadu matematiky, avšak je potrebné poznamenať, že koncepcia fenotypu reprezentuje veľmi efektívnu a plodnú heuristiku[6] pre interpretáciu univerzálneho darwinizmu. Cieľom evolúcie nie je obvykle požadovaný genotyp, ale fenotyp vykazujúci určité vlastnosti. Z tohto pohľadu môžeme povedať, že darwinovská evolúcia je reprezentovaná postupnosťou

[6] Heuristika v informatike znamená pravidlo, väčšinou intuitívne zavedené, ktoré poskytuje určitý návod, ako sa správať v zložitom prostredí, ktorého presný model nepoznáme. Tak napríklad, v neznámom lese je veľmi užitočnou heuristikou pravidlo, že sa používajú len označené chodníky. V mnohých prípadoch tieto označené chodníky nereprezentujú optimálnu cestu, ale zabezpečia nám, že sa dostaneme do cieľa.

fenotypov, ktorých vlastnosti sú bližšie a bližšie k vlastnostiam cieľového hypotetického fenotypu.

Fitnes replikátora $\boldsymbol{x} \in P$ je určený dvoma rôznymi prístupmi. Prvý spôsob chápe fitnes ako atribút replikátora (vloženého do nejakého prostredia E), interakcie daného replikátora s populáciou sa ignorujú, potom

$$fitnes(\boldsymbol{x}) = F(\boldsymbol{x}; E)$$

kde prostredie E je chápané ako parameter funkcie, ktorá replikátor $\boldsymbol{x}$ ohodnotí fitnes, táto funkcia je určená len pomocou jeho genotypu (a prípadného vplyvu prostredia). Druhý spôsob určenia fitnes je založený na predstave, že táto veličina je dominantne určená len interakciou daného replikátora $\boldsymbol{x}$ s ostatnými replikátormi v populácii (plus vplyv prostredia E)

$$fitnes(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{x}' \in P \\ \boldsymbol{x}' \neq \boldsymbol{x}}} F(\boldsymbol{x}, \boldsymbol{x}'; E)$$

Tento druhý prístup postuluje, že dominantná časť fitnes je určená práve interakciou replikátorov v populácii, kde nie je dôležitá architektúra toho-ktorého replikátora (môžeme povedať, že táto prispieva nemennou časťou k fitnes). Takto je určené fitnes replikátorov v aplikáciách evolučnej teórie hier [2] (napr. pri štúdiu evolučného vzniku kooperácie v populácii). V týchto aplikáciách zahájených Maynard Smithom [39] a Axelrodom [2] už skoro pre 30 rokmi, fitnes replikátorov je určený schopnosťou spolupracovať alebo nespolupracovať s inými replikátormi z populácie, a nie ich vnútornou architektúrou. V tomto prípade funkcia fitnes nie je fixne zadaná, ale závisí od zloženia celej populácie. Z tohto pohľadu potom môžeme tento druh evolučnej optimalizácie interpretovať ako tzv. „problém pohyblivého cieľa“, kde povrch fitnes funkcie sa dynamicky mení s evolúciou populácie. Problém emergencie sociálnych štruktúr v multiagentových systémoch bude študovaný v sekcii 4.

3.6 Koevolúcia

Teória darwinovskej evolúcie obvykle implicitne predpokladá existenciu jednej populácie, ktorá je predmetom evolúcie uskutočňovanej mechanizmami popísanými vyššie. Existujú však prípady, keď dva rôzne druhy replikátorov medzi sebou silne interagujú (napr. dve populácie rôznych druhov, ktoré sú v symbiotickom vzťahu, jeden druh produkuje látku, ktorá je vitálne dôležitá pre prežitie druhého druhu). V abiotických aplikáciách sú koevolučné modely dôležité pre štúdium rôznych kultúrnych fenoménov.

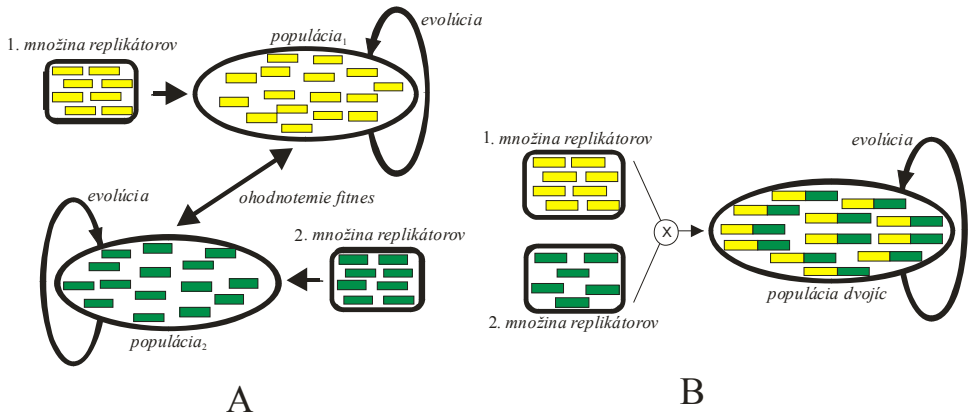
Budeme študovať dva rôzne modely koevolúcie, pozri obrázok 11.

Prvý model (diagram A) vyjadruje tzv. *kompetitívny koevolučný model*, kde existujú dve kvázinezávislé populácie, ktoré sú zložené z replikátorov rôznych druhov. Interakcia medzi týmito rôznymi populácia spočíva vo výpočte fitnes replikátorov z jednotlivých populácií, fitnes replikátora z prvej populácie je určený nielen jeho zložením (alebo jeho interakciou s ostatnými replikátormi z danej populácie), ale aj jeho interakciou s replikátormi z druhej populácie, napr.

$$fitnes(\boldsymbol{x}) = F_1(\boldsymbol{x}) + \sum_{\boldsymbol{y} \in P_2} F_{12}(\boldsymbol{x}, \boldsymbol{y}) \qquad (\forall \boldsymbol{x} \in P_1)$$

$$fitnes(\boldsymbol{y}) = F_2(\boldsymbol{y}) + \sum_{\boldsymbol{x} \in P_1} F_{21}(\boldsymbol{y}, \boldsymbol{x}) \qquad (\forall \boldsymbol{y} \in P_2)$$

Ak obe populácie chápeme ako jeden (ko)evolučný systém, potom evolučná optimalizácia môže byť taktiež interpretovaná ako „problém pohyblivého cieľa“, pretože funkcia fitnes sa dynamicky mení s evolúciou populácií.



Obr. 3.11. Dva rôzne koevolučné modely. Diagram A vyjadruje kompetitívny model koevolúcie obsahujúci dve kvázinezávislé populácie, tieto populácie vzájomne interagujú prostredníctvom výpočtu fitnes. Ak počítame fitnes replikátorov z jednej populácie, berieme do úvahy aj interakciu s replikátormi z druhej populácie. Diagram B vyjadruje kooperatívny model koevolúcie, v tomto prípade populácia je zložená z dvojíc odlišných replikátorov. Fitnes týchto dvojíc je určený jednak ich zložkami a taktiež aj ich vzájomnou interakciou.

Druhý model (obrázok 11, diagram B) sa nazýva *kooperatívny koevolučný model*, v ktorom existuje len jedna populácia, ale táto populácia je zložená z komplexov (usporiadaných dvojíc) replikátorov rôzneho druhu. Tieto komplexy sú chápané ako nezávislé objekty evolúcie, ich fitnes je určený napr. takto

$$fitness(\boldsymbol{x},\boldsymbol{y}) = F_1(\boldsymbol{x}) + F_2(\boldsymbol{y}) + G(\boldsymbol{x},\boldsymbol{y})$$

kde funkcie F_1 a F_2 určujú fitnes jednotlivých zložiek komplexu, funkcia G špecifikuje príspevok k fitnes na základe interakcie medzi zložkami komplexu. Tento model sa používa vtedy, keď jednotlivé komponenty neexistujú nezávislé, sú na sebe silne závislé. Ako pekný ilustračný príklad pre kooperatívnu koevolúciu môže slúžiť „kultúrna“ aplikácia [36] zaoberajúca sa koevolúciou génov a mémov.

3.7 Eigenova a Schustrova teoria molekulárneho darwinizmu

Eigenova a Schustrova teória molekulárneho darwinizmu [13,14] postihuje elementárne aspekty evolúcie na úrovni biomakromolekúl. Tento model evolúcie poskytuje semikvantitatívnu teóriu, ktorá je schopná interpretovať mnohé pojmy a fenomény evolúcie, ako napr. úlohu neutrálnych mutácií v priebehu adaptačného procesu. Dôležitú úlohu v našom prístupe hrajú metódy ***umelej chémie*** [3,12,16-20,23, 32-36], ktorá je v súčasnosti považovaná za oblasť úzko súvisiacu s umelým životom. Tento prístup je založený na metafore chemostatu (chemického reaktora), ktorý obsahuje "molekuly" reprezentované abstraktnými objektmi (reťazcami znakov, grafmi, formulami z nejakého jazyka, atď.). Molekuly sú v chemostate transformované pomocou "chemických reakcií" na iné prípustné objekty, pričom pravdepodobnosť týchto transformácií je striktne určená štruktúrou reagujúcich objektov. Kinetika procesov prebiehajúcich v chemostate je popísaná Eigenovými replikátorovými diferenciálnymi rovnicami, ktoré majú svoj „zdravý“ základ vo fyzikálnej chémii (zákon účinných hmotností [1]). Hlavným predmetom záujmu umelej chémie je (*i*) štúdium formálnych systémov založených na metafore chemostatu, ktoré sú schopné vykonávať špeciálne vysoko paralelné výpočty a (*ii*) návrh formálne autokatalytických systémov (molekuly sú reprezentované štruktúrovanými objektmi) pre účely *in-silico* simulácií emergencie "živých" systémov.

V súčasnosti je molekulárna evolúcia teoreticky študovaná pomocou "fyzikálneho" modelu RNA a jej foldingu [16-20,45-46]. Pre úplnosť je potrebné poznamenať, že existujú aj iné modely molekulárnej evolúcie [43], menovite pripomenieme tie, ktoré sú založené na stochastickej Kauffmanovej *KN* funkcii [27,28], ktoré sú veľmi všeobecné a poskytujú podobné výsledky ako metódy založené na RNA foldingu. Ukážeme, že použitie tohto jednoduchého modelu povrchu fitnes pomocou RNA foldingu poskytuje výsledky, ktoré sú formálne veľmi blízke tým, ktoré sú predpovedané replikátorovými diferenciálnymi rovnicami, ktoré boli postulované už pred viac ako 30 rokmi Eigenom [13] ako odozva na Spiegelmanove *in-vitro* experimenty [47] s evolúciou RNA molekuly[7].

3.7.1 Eigenove replikátory

Na prelome 60. a 70. rokov publikoval Manfred Eigen významnú prácu "*Self organization of matter and the evolution of biological macro molecules*" [13], v ktorej študoval vlastnosti chemického systému zloženého z tzv. *replikátorov* (priamo v texte tejto práce uvádza, že bol stimulovaný k jej napísaniu Spiegelmanovými [47] experimentmi *in-vitro* darwinovskej evolúcie). Replikátory sú hypotetické biomakromolekuly, ktoré majú schopnosť replikovať sa. Eigen ukázal, že takýto jednoduchý chemicko-kinetický systém formálne simuluje darwinovskú evolúciu na molekulárnej úrovni. Ďalšie zovšeobecnenie modelu replikátorov bolo vykonané Eigenom a Schusterom [14] v sérii prác so spoločným názvom "*The Hypercycles: A Principle of Natural Evolution*", v ktorých sa položili základy teórie *hypercyklov*, ktoré sú v súčasnosti považované za jeden z nemnohých formálnych systémov schopných simulovať nárast zložitosti (komplexity) na rozhraní živých a neživých systémov.

Študujme molekuly - replikátory $X_1, X_2, ..., X_n$, ktoré sú schopné nasledujúcich dvoch chemických reakcií

$$X_i \xrightarrow{k_i} X_i + X_i \quad (i = 1,2,...,n)$$

$$X_i \xrightarrow{\phi} \varnothing \quad (i = 1,2,...,n)$$

Prvá reakcia znamená, že molekula X_i sa replikuje s rýchlostnou konštantou k_i a druhá reakcia znamená, že molekula X_i zaniká s rýchlostnou konštantou ϕ (ktorá sa nazýva *zrieďovací tok* a bude špecifikovaná neskoršie). Ak použijeme k týmto reakciám štandardný fyzikálno-chemický zákon účinných hmotností [1], dynamika kinetického systému obsahujúceho obe reakcie je určená diferenciálnymi rovnicami

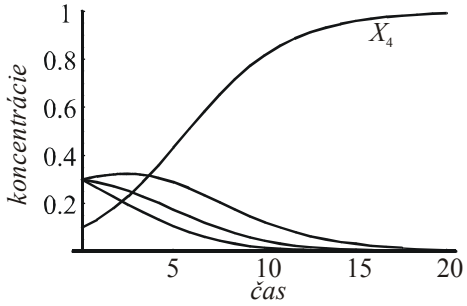
$$\dot{x}_i = x_i (k_i - \phi) \quad (i = 1,2,...,n)$$

kde $0 \le x_i \le 1$ je koncentrácia molekuly X_i. Zrieďovací tok ϕ je chápaný ako "voľný parameter", ktorý sa nastaví tak, aby suma časových derivácií koncentrácií bola nulová, $\sum \dot{x}_i = 0$

$$\dot{x}_i = x_i \left(k_i - \sum_{j=1}^{n} k_j x_j \right) \quad (i = 1,2,...,n)$$

kde podmienka $\sum x_i = 1$ bola použitá bez straty všeobecnosti našich úvah.

[7] Pozri prednášku Ľ. Tomáška a J. Noseka [52] na seminári Kognitívne vedy III (CogSci2000), CHTF STU, Bratislava. Text prednášky je dostupný na http://math.chtf.stuba.sk/ CogSci_2000.htm.

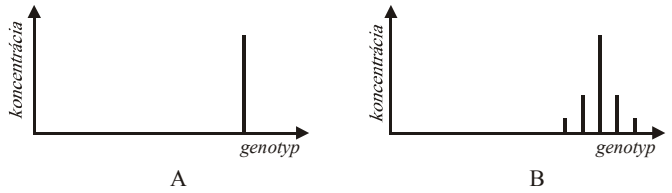


Obr. 3.12. Časový priebeh koncentrácií 4-zložkového systému s rýchlostnými konštantami k_1=1, k_2=2, k_3=3 a k_4=4. Molekuly X_4 víťazia, aj keď ich počiatočná koncentrácia bola najmenšia, v čase t>15 systém už obsahuje skoro výlučne len molekuly X_4, ostatné vymizli.

Riešenie tohto systému diferenciálnych rovníc vyhovuje asymptotickej podmienke, kde "prežíva" len tá molekula, ktorá má maximálnu rýchlostnú konštantu k_{max}, zatiaľ čo ostatné molekuly vymiznú

$$\lim_{t\to\infty} x_i(t) = \begin{cases} 1 & \left(\text{pre } k_i = k_{max} = max\{k_1,...,k_n\}\right) \\ 0 & \left(\text{ostatné prípady}\right) \end{cases}$$

Každý typ molekuly je formálne chápaný ako zvláštny biologický druh s fitnes špecifikovanou rýchlostnou konštantou k, potom prežívajú len tie molekuly - biologické druhy, ktoré sú najlepšie prispôsobené, t.j. majú najväčšiu rýchlostnú konštantu k_{max}; všetky ostatné molekuly - druhy s rýchlostnými konštantami - fitnes menšími než ako k_{max} vymiznú v priebehu časovej evolúcie chemicko-kinetického systému, pozri obr. 12. Podmienka konštantnosti sumy koncentrácie (t.j. $\sum x_i = 1$) zavádza do systému dodatočný "selekčný tlak", replikujúce sa molekuly eliminujú ostatné molekuly.



Obr. 3.13. Diagram A znázorňuje asymptotickú situáciu $t\to\infty$ pre systém replikátorov bez mutácií, ktoré sa presne replikujú. V tomto prípade asymptoticky stabilný stav systému obsahuje len replikátory s maximálnou rýchlostnou konštantou k_{max}. Diagram B znázorňuje asymptoticky stabilný stav systému, kde replikácia nie je presný proces, kde sa už uvažujú mutácie pri replikačnom procese. Stabilný stav je zložený z niekoľkých replikátorov, pričom najväčšia koncentrácia prislúcha replikátoru s maximálnou rýchlostnou konštantou k_{max}, ostatné replikátory sa vyskytujú s podstatne menšou koncentráciou. Replikátory vyskytujúce sa súčasne s hlavným víťazom X_{max} sú tie, ktoré sú mu blízke a sú z neho tvorené mutáciami.

Pôvodná forma Eigenovho modelu replikátorov neobsahovala mutácie, replikácia molekúl bol presný proces, pri replikácii molekuly X_i vznikala len táto molekula a žiadna iná sa týmto replikačným procesom nevytvárala. Eigen jednoducho modifikoval svoj model tak, že obsahoval už aj mutácie

$$X_j \xrightarrow{k_{ji}} X_j + X_i \quad (i,j = 1,2,...,n)$$

kde k_{ji} je rýchlostná konštanta tejto reakcie, pričom sa predpokladá, že matica rýchlostných konštánt $\boldsymbol{K} = \left(k_{ij}\right)$ má dominantné diagonálne elementy, t.j. nediagonálne elementy sú podstatne (rádovo) menšie ako diagonálne elementy ($k_{ij} \ll k_{ii}$, pre i,j=1,2,...,n a $i \neq j$). Tento dôležitý predpoklad vyplýva priamo zo skutočnosti, že "nepresné" replikácie pre $i \neq j$ sú

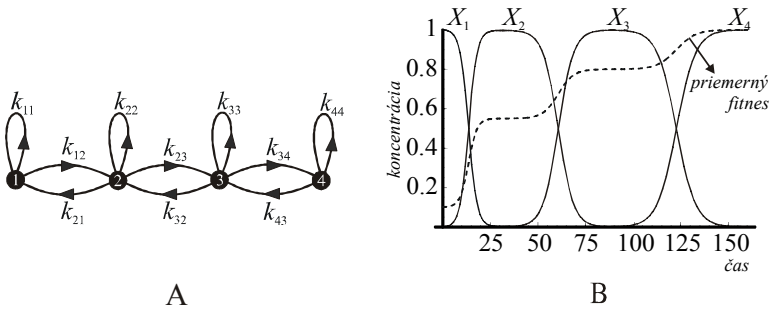
veľmi vzácne, t.j. produkt X_j je uvažovaný ako mutácia eduktu X_i, $X_j = O_{mut}(X_i)$. Systém diferenciálnych rovníc popisujúci dynamické vlastnosti takto modifikovaného systému replikátorov má tvar

$$\dot{x}_i = x_i(k_{ii} - \phi) + \sum_{j \neq i} k_{ji} x_j \quad (i = 1,2,...,n)$$

Zrieďovací tok ϕ je určený tak, aby aj v tomto všeobecnejšom prípade suma časových derivácií koncentrácií bola nulová ($\sum \dot{x}_i = 0$, alebo $\sum x_i = 1$)

$$\phi = \sum_{i,j=1}^{n} k_{ij} x_j$$

Zavedenie slabých mutácií nemení dramaticky všeobecné vlastnosti pôvodného jednoduchého systému replikátorov bez mutácií. Menovite, v asymptotickom čase $t \to \infty$ systém obsahuje skoro výlučne len molekuly majúce najväčšie rýchlostné konštanty k_{max}, ktoré sú vzácne sprevádzané inými replikátormi tvorenými z víťazného replikátora (s rýchlostnou konštantou k_{max}) pomocou nepresného replikačného procesu, pozri obrázok 13.



Obr. 3.14. Diagram (A) reprezentuje 4-replikátorový systém, kde replikátor X_i vytvára pomocou replikácie s chybami susedné replikátory X_{i-1} a X_{i+1}. Orientované hrany diagramu sú ohodnotené rýchlostnými konštantami, pričom diagonálne rýchlostné konštanty sú medzi sebou dobre separované a o mnoho väčšie ako nediagonálne rýchlostné konštanty. Diagram (B) znázorňuje koncentračné profily replikátorov, ktoré tvoria postupnosť koncentračných vĺn. Diagram taktiež obsahuje priebeh priemerného fitnes, ktorý je špecifikovaný $\bar{k} = k_{11}x_1 + ... + k_{44}x_4$, tvorí neklesajúcu skokovú funkciu, ktorá je typická pre Darwinovu evolúciu.

3.7.2 Replikátory a molekulová Darwinova evolúcia

Eigenov systém replikátorov s mutáciou (t.j. replikácie s chybami) je vhodným prostriedkom na simuláciu *molekulárnej Darwinovej evolúcie*. Študujme hypoteticky reakčný systém obsahujúci postupnosť n replikátorov X_1, X_2, X_3, ..., X_n. Tieto replikátory vyhovujú podmienkam, že X_i môže produkovať pomocou chybnej replikácie susedné replikátory $X_{i\pm1}$, pozri obrázok 14, diagram A. Ak chemostat v čase t=0 obsahuje len replikátor X_1, $x_1(0)$=1, potom v priebehu evolúcie existujú koncentračné vlny, ktoré sú postupne priradené replikátorom X_2, X_3, ..., X_n, pozri obrázok 14, diagram B. Táto skutočnosť je interpretovaná ako prejav molekulovej Darwinovej evolúcie, kde fitnes jednotlivých „druhov“ sú špecifikované diagonálnymi rýchlostnými konštantami k_{ii}. Evolučný proces bol naštartovaný čistou populáciou obsahujúcou len X_1. Pretože jeho replikácia nie je perfektná, príležitostne môže vznikať nasledujúci replikátor X_2 (špecifikovaný rýchlostnou konštantou k_{12}), ktorý má väčší fitnes ako jeho predchodca X_1 ($k_{11}<k_{22}$). To znamená, že v tejto etape evolúcie "druh" X_2 zvíťazí (t.j. jeho koncentrácia sa postupne zvýši skoro na jednotkovú hodnotu, zatiaľ čo koncentrácia predchodcu sa asymptoticky blíži k nulovej hodnote. Tento proces sa opakuje

pre replikátor X_2, ktorý je teraz chápaný ako počiatočný replikátor (hrá rovnakú úlohu ako replikátor X_1 v predchádzajúcej etape). Pretože replikácia X_2 nie je taktiež perfektná, s malou pravdepodobnosťou vzniká aj nový „druh“ X_3, ktorý má väčší fitnes ako jeho predchodca X_2 ($k_{33}>k_{22}$), t.j. tento nový "druh" X_3 zvíťazí. Tento postupný proces je ukončený, keď sa začne objavovať posledný replikátor X_n ako dôsledok neperfektnej mutácie predchodcu X_{n-1}, v záverečnej etape evolúcie sa zvýši koncentrácia tohto finálneho produktu na jednotkovú hodnotu ako dôsledok jeho autoreplikácie.

Sumarizujúc naše výsledky, môžeme konštatovať, že Eigenov systém replikátorov simuluje Darwinovu evolúcie vtedy, keď (1) diagonálne rýchlostné konštanty sú navzájom dobre separované a (2) suma „vychádzajúcich“ rýchlostných konštánt musí byť podstatne menšia ako suma vchádzajúcich rýchlostných konštánt. Táto druhá podmienka znamená, že „pravdepodobnosť“ vzniku daného replikátora zo susedných replikátorov ako dôsledok ich nepresných replikácii je o mnoho väčšia než „pravdepodobnosť“ zániku daného replikátora v dôsledku jeho nepresnej replikácie.

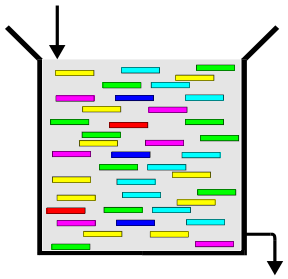
3.7.3 Metafora chemického reaktoru - chemostat

Uvažujme chemický reaktor - *chemostat*, ktorý obsahuje formálne objekty nazývané „*molekuly*“, ktoré sú reprezentované reťazcami znakov. Chemostat nie je priestorovo štruktúrovaný, v chémii sa hovorí, že je s dobrým miešaním (well stirred reactor), pozri obr. 15. Akt chemickej reakcie - replikácie spočíva v nasledujúcich troch krokoch:

(1) náhodný výber replikátora M (ktorý je totožný s X_i) z chemostatu,

(2) pre takto vybraný replikátor sa uskutoční s pravdepodobnosťou úmernou rýchlostnej konštante k_{ii} proces replikácie, pričom vznik chybného replikátora M' (ktorý je totožný s X_j) je uskutočnený s pravdepodobnosťou úmernou rýchlostnej konštante k_{ij}.

(3) Vzniknutým novým replikátorom M' (bez mutácie alebo s mutáciou) sa nahradí iný náhodne vybraný replikátor z chemostatu (t.j. vo všetkých elementárnych krokoch máme v chemostate konštantný počet replikátorov).



Obr. 3.15. Chemostat obsahuje "zmes" molekúl – replikátorov. Replikačný akt spočíva v tom, že náhodne vybraná molekula sa zreplikuje (buď perfektne alebo s malou mutáciou), vzniknutý produkt – molekula vytesní inú náhodne vybranú molekulu (chemostat obsahuje v celej svojej histórii konštantný počet molekúl).

Počítačová simulácia chemostatu a v ňom prebiehajúce replikačné reakcie sa dá pomerne jednoducho algoritmizovať pomocou stochastického prístupu [23]. Tento prístup je založený na skutočnosti, že ak náhodne vyberieme z chemostatu nejaký replikátor, potom pravdepodobnosť toho, že tento replikátor je totožný s nejakým daným replikátorom X sa jednoducho rovná koncentrácii tohto replikátora v chemostate, $0 \le c(X) \le 1$.

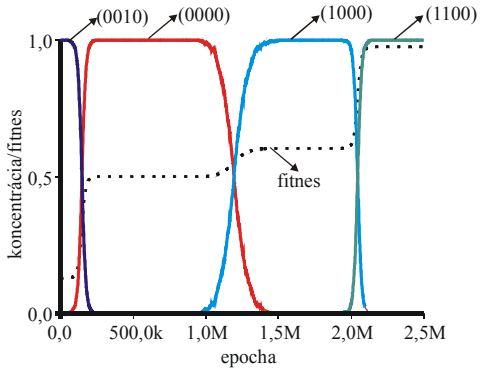
Pravdepodobnosť uskutočnenia replikácie sa rieši tak, že táto pravdepodobnosť je úmerná replikačnej rýchlostnej konštante. Algoritmicky môžeme chemostat implementovať pomocou pseudo-Pascalu týmto jednoduchým algoritmom:

Algoritmus 2.

```
chemostat:=náhodne vygenerovaná množina replikátorov;
t:=0;
while t<t_max then
begin t:=t+1;
      X:=O_select(chemostat);
      if random<prob_rep(X) then
      begin X':=O_mut(X);
            X'':=O_select(chemostat);
            chemostat:=(chemostat+{X'})-{X''};
      end;
end;
```

Algoritmus je inicializovaný náhodným vygenerovaním chemostatu. Pravdepodobnosť $\mathtt{prob}_{\mathtt{rep}}\mathtt{(X)}$ je určená tak, že je úmerná rýchlostnej konštante replikácie *X*, pravdepodobnosť mutácie je zahrnutá priamo v operátore $\mathtt{O}_{\mathtt{mut}}\mathtt{(X')}$.

Všeobecné idey chemostatu budú ilustrované jeho použitím ako optimalizátora binárnej funkcie $f : \{0,1\}^n \to [0,1]$, ktorá zobrazuje binárne reťazce $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ dĺžky *n* na reálne čísla z intervalu $\langle 0,1 \rangle$. Našim cieľom bude nájsť optimálne riešenie $\boldsymbol{x}_{opt}$, ktoré maximalizuje binárnu funkciu *f* nad množinou všetkých možných binárnych reťazcoch dĺžky *n*.



Obr. 3.16. Priebeh frekvencií výskytu dominantných binárnych reťazcov (16) dĺžky 4. Chemostat bol inicializovaný 1000 reťazcami (0010), po 2.5×10^6 časových krokoch dominantným konečným riešením je reťazec $\boldsymbol{x}_{fin}$=(1100), kde *real*($\boldsymbol{x}_{opt}$)=12/15 a *f*(12/15)=0.9755. Toto konečné riešenie je blízke optimálnemu riešeniu $\boldsymbol{x}_{opt}$=(1011), kde ($\boldsymbol{x}_{opt}$)=11/15 a *f*(11/15)=0.9973, ale s „veľkou“ Hammingovou vzdialenosťou *d*=3 (v genetickom algoritme tento efekt sa nazýva „Hammingov útes“). Táto relatívne veľká Hammingova vzdialenosť je hlavný dôvod prečo replikátorový algoritmus nie je schopný poskytnúť globálne riešenie (1011).

Nech chemostat obsahuje molekuly, ktoré sú realizované binárnymi reťazcami $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. Uvažujme nasledujúcu monomolekulárnu reakciu – replikáciu

$$\boldsymbol{x} \xrightarrow{f(\boldsymbol{x})} \boldsymbol{x} + \boldsymbol{x}'$$

pričom molekula - reaktant $\boldsymbol{x}'$ nahradí náhodne vybranú molekulu z chemostatu. Funkčná hodnota $f(\boldsymbol{x})$ priradená chemickej reakcii je interpretovaná ako pravdepodobnosť (rýchlostná konštanta) replikácie. Podobne ako v genetickom algoritme, selekčný tlak v chemostate je vytvorený pomocou fitnes (rýchlostné konštanty), replikátory s väčším fitnes majú väčšiu

šancu zúčastňovať sa reprodukčného procesu (miera kvality replikátorov), a naopak, replikátory s malým fitnes len vzácne vstupujú do reprodukčného procesu. Tento jednoduchý prejav princípu Darwinovho prirodzeného výberu zabezpečuje postupnú monotónne neklesajúcu gradáciu fitness v priebehu evolúcie populácie. V tomto prístupe založenom na metafore chemostatu, princíp prírodného výberu na základe fitness je rozšírený o dodatočný selekčný tlak založený na skutočnosti, že populácia v chemostate má konštantný počet jedincov. Molekula vstupujúca do reakcie je náhodne vybraná z chemostatu, kvázináhodne je rozhodnuté, či vstupuje do reakcie alebo nie (s pravdepodobnosťou úmernou jej funkčnej hodnote), vzniknutá molekula vytesní z chemostatu inú, náhodne vybranú molekulu. Produkt $\boldsymbol{x}'$ je mutácia vstupujúcej molekuly $\boldsymbol{x}$, $\boldsymbol{x}' = O_{mut}(\boldsymbol{x})$, kde O_{mut} je stochastický operátor mutácie, ktorý zmení jednotlivý bit s pravdepodobnosťou P_{mut}. Pseudo-Pascalovský kód tohto optimalizačného algoritmu dostaneme jednoduchou modifikáciou Algoritmu 2.

Ako ilustračný príklad študujme chemostat pre unimodálnu funkciu určenú nad binárnymi reťazcami dĺžky 4, chemostat je množina $P = \{...,(1100),...\} \subset \{0,1\}^4$. Každý binárny vektor $\boldsymbol{x}$ je ohodnotený racionálnym číslom z intervalu $\langle 0,1\rangle$

$$real(\boldsymbol{x}) = \frac{1}{2^4 - 1} int(\boldsymbol{x})$$

kde $int(\boldsymbol{x})$ je nezáporné celé číslo vyjadrujúce dekadickú interpretáciu binárneho reťazca $\boldsymbol{x}$. Rýchlostná konštanta k priradená reťazcu

$$k(\boldsymbol{x}) = f\left(real(\boldsymbol{x})\right) = \frac{1}{2}\left(1 + sin\left(2\pi \cdot real(\boldsymbol{x})\right)\right)$$

s optimálnym riešením $\boldsymbol{x}_{\text{opt}}$=(1011), kde $real(\boldsymbol{x})$=11/15 and f(11/15)=0.9973. Chemostat bol inicializovaný 1000 binárnymi reťazcami (0010), mutačný operátor O_{mut} je špecifikovaný 1-bitovou pravdepodobnosťou P_{mut}=0.01. Z numerických výsledkov znázornených na obrázku 16 vyplýva, že v chemostate spontánne emerguje binárny reťazec, ktorý je blízky k optimálnemu riešeniu.

Z tohto ilustračného príkladu vyplýva, že metafora Eigenových replikátorov poskytuje efektívny *replikátorový optimalizačný algoritmus*, ktorý je veľmi podobný štandardnému genetickému algoritmu [15,25,37], avšak je založený na úplne odlišnej metafore ako GA, metafora darwinovskej evolúcie v GA je nahradená metaforou chemostatu replikátorov. Môžeme teda na záver tejto kapitoly konštatovať, že ***Eigenova fenomenologická teória replikátorov tvorí vhodný teoretický rámec pre počítačové simulačné štúdia molekulárnej darwinovskej evolúcie*** (t.j. biomakromolekúl, ako sú RNA alebo DNA, ktoré sú schopné replikačných procesov).

3.7.4 Folding (skladanie) binárnych reťazcov

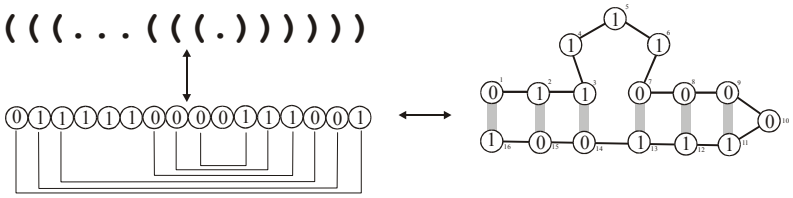
Na základe analógie s RNA molekulami zavedieme tzv. folding (sekundárnu štruktúru) aj pre binárne reťazce (pozri obrázok 17), kde úlohu komplementárnych báz hrajú dvojice 0 a 1. Folding binárneho reťazca je špecifikovaný ako zoznam spárovaných binárnych komplementárnych báz *i-j* (pre $i<j$) a nespárovaných jedincov k

$$folding(\boldsymbol{x}) = \left\{i_1 - j_1, i_2 - j_2, ..., i_p - j_p; k_1, k_2, ..., k_q\right\}$$

kde jednotlivé páry sú ohraničené nasledujúcimi dvoma podmienkami:

(1) pre pár *i-j* platí $j - i \geq 2$,

(2) pre dva páry *i-j* a *k-l,* kde $i \leq k$, platí $i = k \Leftrightarrow j = l$ alebo $k < j \Rightarrow i < k < l < j$.



Obr. 3.17. Binárny reťazec môže byť poskladaný do 2-rozmernej štruktúry (je nazývaná sekundárna štruktúra) takým spôsobom, že komplementárne binárne dvojice sú navzájom spojené. Zátvorková reprezentácia foldingu je ukázaná v hornej časti obrázku, pričom bodky sú priradené neinteragujúcim elementom reťazca a zátvorky '(' a ')' sú priradené interagujúcim elementom.

Prvá podmienka znamená, že minimálna dĺžka tzv. „*vlasovej sponky*“ v každom foldingu musí byť aspoň 2 (pozri pravú stranu sekundárnej štruktúry na obrázku 17). Ináč povedané, jednotlivé väzby – hrany nemôžu byť ohnuté, prípustná je len rotácia dvoch susedných väzieb okolo rovnakého vrcholu. Z druhej podmienky vyplýva, že každý element reťazca sa môže podieľať maximálne na jednej väzbe, alebo, že tzv. „pseudouzly“ sú zakázané (t.j. každý folding sa dá vyjadriť rovinným grafom, kde sa jednotlivé hrany – väzby nepretínajú). Definícia foldingu má ešte podmienku, že obsahuje maximálny možný počet dvojíc s komplementárnymi bázami; táto podmienka odráža fyzikálny význam foldingu, ako najstabilnejšej sekundárnej štruktúry. Podmienky (1-2) sú veľmi dôležité, pretože umožňujú použitie rýchleho algoritmu dynamického programovania [50] pre konštrukciu foldingu, založeného na predpoklade, že hľadanie foldingu je úloha rozložiteľná na menšie podúlohy, ktoré sú použiteľné aj pre konštrukciu riešení väčších podúloh.

Ako už bolo poznamenané, konštrukcia foldingu binárnych reťazcov bude uskutočnená pomocou metódy dynamického programovania [50]. Vo všeobecnosti, túto konštrukciu by sme mohli uskutočniť aj jednoduchým algoritmom spätného prehľadávania, avšak jeho časová náročnosť rastie exponenciálne s dĺžkou reťazca, čo sa pokladá v informatike za málo efektívny algoritmus. K tomu, aby sa zvýšila efektívnosť spätného prehľadávania, zaznamenávajú sa optimálne výsledky pre menšie podreťazce a tieto sa využívajú pre konštrukciu foldingu väčších podreťazcov, tento rekurentný proces je ukončený vtedy, keď dosiahneme celý reťazec, ktorého folding sme hľadali. Táto jednoduchá modifikácia spätného prehľadávania sa nazýva v informatike dynamické programovanie, je použiteľná však len vtedy, keď riešená úloha je rozložiteľná na menšie podúlohy, z ktorých môžeme skladať väčšie podúlohy. Časová náročnosť konštrukcie foldingu pomocou dynamického programovania je n^3, t.j. exponenciálna zložitosť je zredukovaná na kubickú zložitosť.

Ukážeme elementárnu formuláciu algoritmu na konštrukciu foldingu binárnych reťazcov. Nech S_{ij} je podreťazec zátvorkovej reprezentácie, ktorý je priradený foldingu medzi uzlami i a j (včítane), tieto podreťazce sú inicializované $S_{ii} = '.'$ alebo $S_{i,i+1} = '..'$. Nasledujúce elementy S_{ij}, pre j-$i \geq 2$, sú rekurentne zostrojené pomocou formuly

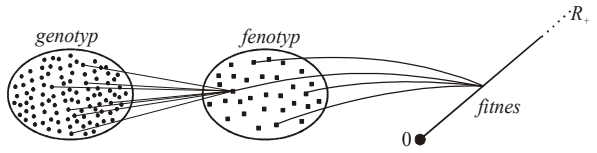
$$S_{i,j} = \max_{i \leq k \leq j-1} S_{ik} \oplus S_{k+1,j}$$

kde symbol $\oplus$ reprezentuje spojenie podreťazcov S_{ik} a $S_{k+1,j}$, ktoré už boli zostrojené v predchádzajúcich krokoch algoritmu (pozri Algoritmus 3). Ak S_{ik} ($S_{k+1,j}$) obsahuje v pravej (ľavej) krajnej polohe symbol ‘.’ , pričom oba tieto symboly sú priradené dvojici komplementárnych binárnych elementov reťazca, potom tieto symboly sú nahradené symbolmi **(** a **)**. Symbol “*max*” znamená, že vyberieme taký index k, ktorý produkuje maximálne párovanie v S_{ij}. Výsledný folding v zátvorkovej reprezentácii je umiestnený v elemente S_{1n}.

Algoritmus 3

```
for i:=1 to n S[i,i]:='.';
for i:=1 to n-1 do S[i,i+1]:='..';
for d:=2 to n-1 do
for i:=1 to n-d do
begin j:=d+i;
      S[i,j]:=max{i≤k≤j-1,S[i,k]⊕S[k+1,j]};
end;
folding:=S[1,n];
```

Základný problém pri implementácii tohto algoritmu spočíva v tom, že takto špecifikovaný algoritmus zostrojí len jeden folding z mnohých možných. Z týchto dôvodov algoritmus musí byť charakterizovaný ako nejaká aktuálna „embryogenéza" fenotypu z daného genotypu a tiež aj ako integrálna časť ohodnocovania genotypu fitnes. Všetky možné foldingy pre reťazce dĺžky n=7 sú uvedené v Tabuľke 1.



Obr. 3.18. Schematické znázornenie zobrazení genotypu na fenotyp a fenotypu na fitnes pre prípad binárnych reťazcov. Obe zobrazenia sú typu mnoho-na-jeden, t.j. existuje silná redundancia v kódovaní fenotypu pomocou genotypu a v kódovaní fitnes pomocou fenotypu. Táto vlastnosť redundancie je dôležitý predpoklad pre existenciu neutrálnych etáp darwinovskej evolúcie, v ktorých je fitnes konštantný.

Tabuľka 1. Všetky možné genotypy reprezentované binárnymi reťazcami dĺžky n=7 a im priradené fenotypy (foldingy) s hodnotami fitnes

No.	Genotyp	fenotyp	fitnes
1	(0000000)(1111111)	`.......`	1
2	(0000001)(0000100)(1111011)(1111110)	`....(.)`	1
3	(0000010)(1111101)	`...(.).`	0
4	(0000011)(0001001)(0001101)(0010010)(0010110) (0011000)(0011100)(0100011)(0100111)(0101001) (0110010)(0110110)(0111000)(0111100)(1000011) (1000111)(1001001)(1001101)(1010110)(1011000) (1011100)(1100011)(1100111)(1101001)(1101101) (1110010)(1110110)(1111100)	`..((.))`	0
5	(0000101)(0010001)(0101100)(0101110)(0110000) (0111010)(1000101)(1001111)(1010001)(1010011) (1101110)(1111010)	`.((.).)`	3
6	(0000110)(0011010)(1100101)(1111001)	`.((.)).`	5
7	(0000111)(0001111)(0010011)(0011011)(0100101) (0101101)(0110001)(0111001)(1000110)(1001110) (1010010)(1011010)(1100100)(1101100)(1110000) (1111000)	`(((.)))`	7
8	(0001000)(1110111)	`...(..)`	0
9	(0001010)(1110101)	`((.).).`	0
10	(0001011)(0001100)(0001110)(0011001)(0011110) (0100001)(0100100)(0100110)(0110011)(0110100)	`.(.)(.)`	0

10	(1001011)(1001100)(1011001)(1011011)(1011110) (1100001)(1100110)(1110001)(1110011)(1110100)		
11	(0010000)(1101111)	..(...)	1
12	(0010100)(0111011)(0111110)(1000001)(1000100) (1101011)	(.).(.)	1
13	(0010101)(0101010)(1010101)(1101010)	.((..))	1
14	(0010111)(0011101)(0100010)(0101000)(1010111) (1011101)(1100010)(1101000)	.(.(.))	0
15	(0011111)(1100000)	((.)..)	0
16	(0100000)(1011111)	.(....)	1
17	(0101011)(1010100)	(..)(.)	0
18	(0101111)(1010000)	(.(.).)	5
19	(0110101)(0110111)(1001000)(1001010)	(.)(..)	0
20	(0111101)(1000010)	(..(.))	3
21	(0111111)(1000000)	(.....)	3

Koncepcia foldingu umožňuje zaviesť aj triádu základných pojmov evolučnej biológie: *genotyp*, *fenotyp*, a *fitnes*. Genotyp $\boldsymbol{x}$ je reprezentovaný binárnym reťazcom dĺžky n, $\boldsymbol{x} = (x_1 x_2 ... x_n) \in \{0,1\}^n$, fenotyp $ph(\boldsymbol{x})$ je určený foldingom reťazca $\boldsymbol{x}$; formálne, fenotyp môže byť vyjadrený pomocou zátvorkovej reprezentácie, $ph(\boldsymbol{x}) = fold(\boldsymbol{x}) \in \{(,),.\}^n$. Fitnes je reprezentovaný kladným reálnym číslom, je atribútom genotypu $\boldsymbol{x}$, jeho ohodnotenie fitnes je realizované pomocou pridruženého fenotypu. Pre naše nasledujúce úvahy fitnes bude špecifikovaný pomocou podobnosti medzi fenotypom $ph(\boldsymbol{x})$ a ad-hoc cieľovým fenotypom ph_{req}. Postulujme, že maximálny fitnes rovný n je dosiahnutý vtedy, ak podobnosť medzi fenotypom $ph(\boldsymbol{x})$ a požadovaným fenotypom ph_{req} je maximálna (t.j. hľadáme taký cieľový genotyp $\boldsymbol{x}$ pre ktorý platí $ph(\boldsymbol{x}) = ph_{req}$). V opačnom prípade, ak platí $s(ph(\boldsymbol{x}), ph_{req}) < n \Leftrightarrow ph(\boldsymbol{x}) \neq ph_{req}$, potom fitnes je menšie ako n; v limitnom prípade; keď foldingy sú úplne odlišné (t.j. $s(ph(\boldsymbol{x}), ph_{req}) = 0$), potom fitnes je nulový, $fitnes(\boldsymbol{x}) = 0$. Presná definícia pojmu podobnosti je uskutočnená v našej práci venovanej molekulárnej evolúcii [35]. Tabuľka 1 obsahuje ilustratívne výsledky pre binárne reťazce dĺžky $n=7$ a im pridružené fenotypy – foldingy. Z týchto výsledkov vyplýva, že mnoho binárnych reťazcov je ohodnotených jedným fenotypom, a podobne, mnoho fenotypov je ohodnotených rovnakým fitnes. Toto jednoduché pozorovanie zovšeobecníme tak, že zobrazenia genotypu na fenotypu a fenotypu na fitnes sú typu mnoho-na-jeden, t.j. tieto zobrazenia majú mohutnú redundanciu[8] (pozri obrázok 18). Fitnes fenotypov je určený pomocou podobnosti, $fitnes(\boldsymbol{x}) = s(ph(\boldsymbol{x}), ph_{req})$.

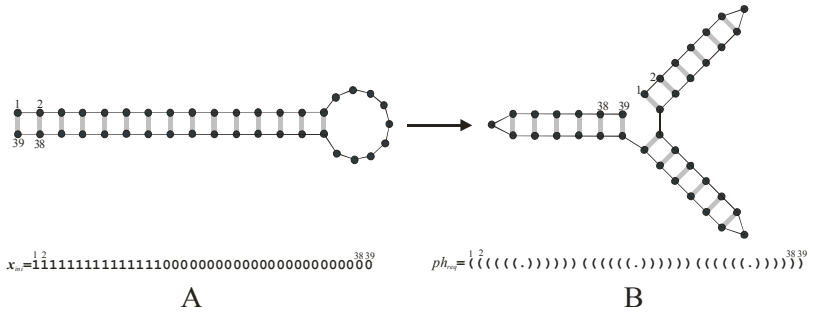
Simulácia molekulovej Darwinovej evolúcie bude uskutočnená pomocou chemostatového algoritmu (pozri Algoritmus 2), pričom počiatočné zloženie chemostatu je tvorené len binárnymi reťazcami dĺžky $n=39$ tvaru

$$\boldsymbol{x}_{ini} = \mathbf{111111111111111000000000000000000000000}$$

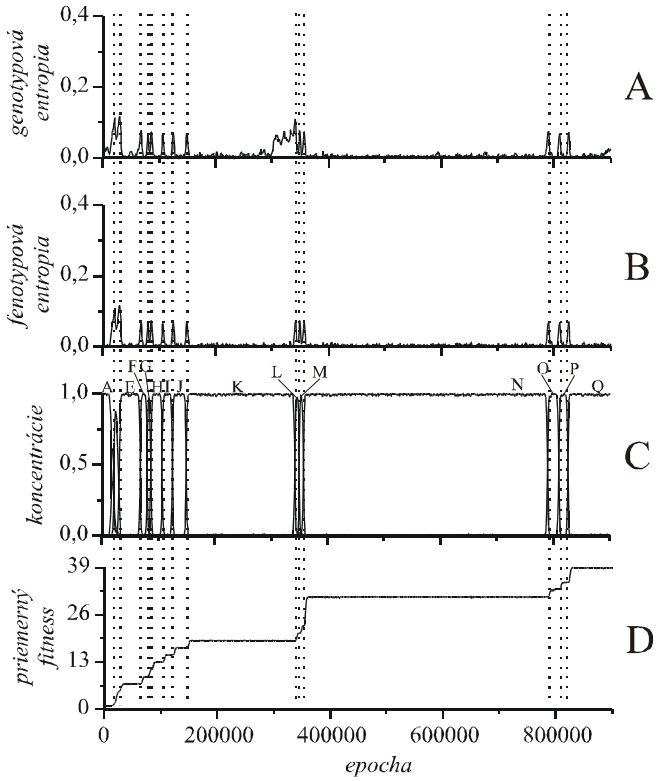
Cieľový fenotyp je špecifikovaný v zátvorkovej reprezentácii takto (pozri obrázok 19)

$$ph_{req} = \mathbf{((((((.))))))((((((.))))))((((((.))))))}$$

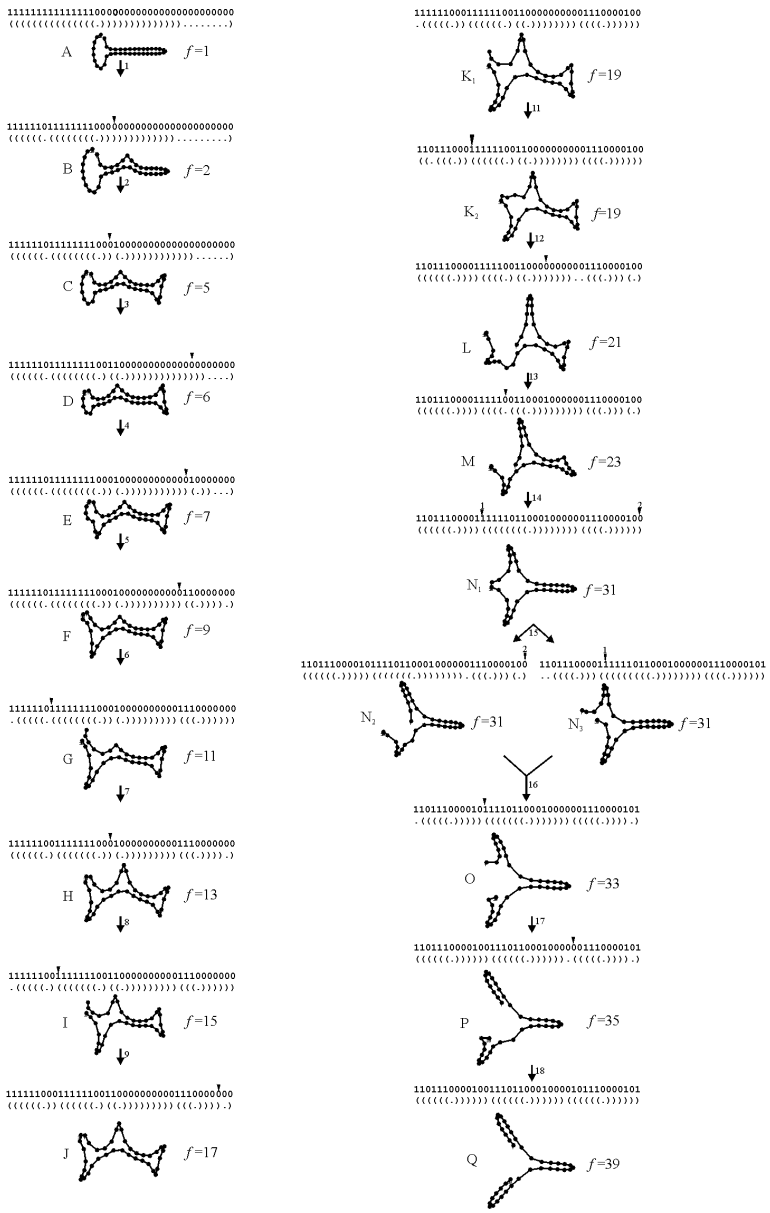
[8] *Redundantnosť* zobrazenia genotypu na fenotyp znamená, že daný obraz – fenotyp má mnoho rôznych argumentov - genotypov, alebo ináč povedané, spôsob kódovania fenotypu je vysoko redundantný - nadbytočný. V opačnom prípade hovoríme, že zobrazenie $\boldsymbol{y}=f(\boldsymbol{x})$ *nie je redundantné*, ak každému obrazu $\boldsymbol{y}$ existuje len jeden argument $\boldsymbol{x}$, t.j. ak zobrazenie je 1-1-značné.



Obr. 3.19. (A) Počiatočný binárny reťazec $\boldsymbol{x}_{ini}$ vyjadrený binárnym reťazcom dĺžky 39, v hornej časti diagramu je taktiež uvedený folding – fenotyp tohto reťazca. (B) Cieľový fenotyp pre genotypy vyjadrené binárnymi reťazcami dĺžky n=39. V dolnej časti obrázku je uvedená zátvorková reprezentácia fenotypu. Fenotyp „vrtuľa“ obsahuje 3×6=18 väzieb, ktoré sú tvorené dvojicami komplementárnych čísel 01 a 10, tento fenotyp môže byť priradený aspoň 2^{18}=262144 binárnym reťazcom. Šípka na obrázku znamená „evolučnú transformáciu“ počiatočného fenotypu na cieľový fenotyp, ktorá je uskutočnená pomocou chemostatového algoritmu (pozri Algoritmus 2).



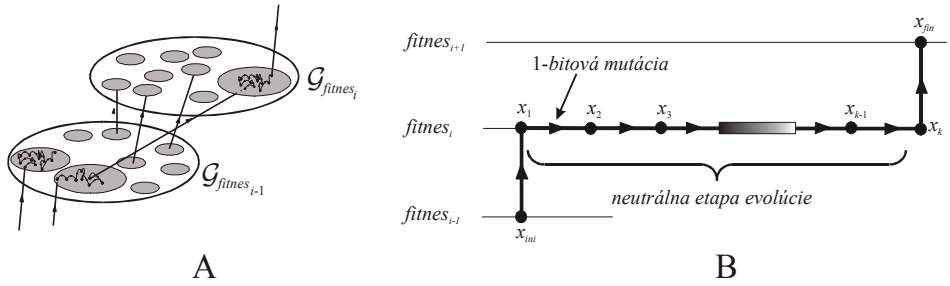
Obr. 3.20. Priebehy (A) genotypovej entropie, (B) fenotypovej entropie, (C) koncentračných profilov dominantných genotypov a (D) priemerného fitnes. Priemerný fitnes je neklesajúca funkcia, ktorá má „skokový“ charakter, obsahuje relatívne dlhé neutrálne etapy (stázie). Skoky na funkcii priemerného fitnes sú dobre indikované obidvoma entropiami. Napríklad, ukončenie neutrálnej etapy (okolo epochy 320000) je indikované masívnou zmenou genotypovej entropie, zatiaľ čo zmeny fenotypovej entropie sú skoro zanedbateľné. Táto skutočnosť indikuje v tejto evolučnej etape existenciu veľa evolučných mutantov s rovnakým fitnes, ktoré tvoria predpoklad evolučného skoku z fenotypu *K* na fenotyp *L*. Diagram C znázorňuje priebeh koncentračných profilov dominantných genotypov s fitnes špecifikovaným priemernou hodnotou z diagramu D, každý koncentračný profil je identifikovaný znakom (písmena A – Q) fenotypu z obrázku 21.



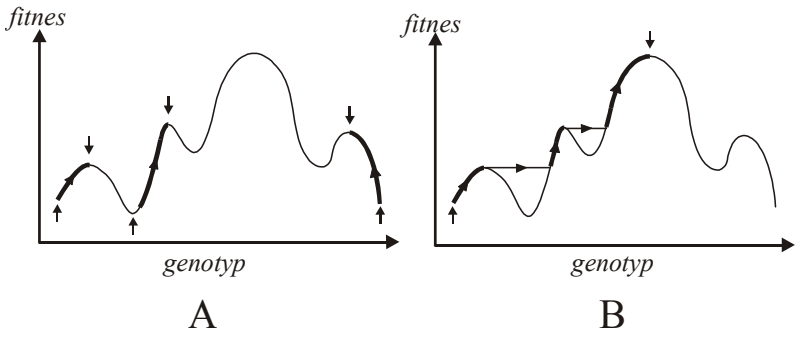
Obr. 3.21. Priebeh evolúcie v chemostate, kde počiatočný genotyp je špecifikovaný binárnym reťazcom $\boldsymbol{x}_{ini}$ a cieľový fenotyp je špecifikovaný útvarom znázorneným na obrázku 19. Identifikovalo sa 18 prechodných stavov, ktoré sú zostrojené z predchádzajúcich pomocou 1-bitovej mutácie (t.j. medzi danými binárnymi reťazcami je jednotková Hammingova vzdialenosť). Existujú dva prípady neutrálnych mutácií, prechody 11 a 15. Tieto prechody sú rôzneho typu, prvý prechod je priradený postupnosti neutrálnych mutácií, zatiaľ čo, druhý prechod reprezentuje dve paralelné neutrálne mutácie.Táto „vrtuľová“ sekundárna štruktúra obsahuje tri „listy“, tento cieľový fenotyp bude slúžiť na ohodnocovanie genotypov (binárnych reťazcov) pomocou fitnes. To znamená, že očakávame v chemostate evolučnú emergenciu takých genotypov, ktorých folding – fenotyp je špecifikovaný na obrázku 19. Operátor mutácie O_{mut} je špecifikovaný pravdepodobnosťou 1-bitovej mutácie P_{mut}=0.0001. Táto pravdepodobnosť je dostatočne malá k tomu, aby v chemostate vznikali len blízke genotypy (s malou Hammingovou vzdialenosťou). Veľkosť chemostatu (počet reťazcov) bola 500. Zloženie chemostatu môže byť charakterizované pomocou dvoch rôznych typov entropie, a to genotypovej entropie a fenotypovej entropie, ktoré sú určené takto: nech $0 \le w(x) \le 1$ je koncentrácia genotypov, ktorých daná veličina X (genotyp alebo fenotyp) je ohodnotená

číslom x, potom $S_X = -\sum_x w(x) ln\left[w(x) \right]$. V prípade, že chemostat je homogénny, t.j. obsahuje len genotypy jedného druhu (teda aj fenotypy sú rovnaké), potom tieto entropie sú nulové, $S_{genotyp}=S_{fenotyp}=0$. V opačnom prípade, ak chemostat nie je homogénny, obsahuje aspoň dva rôzne genotypy a fenotypy, potom entropie sú kladné, $S_{genotyp}>0$, $S_{fenotyp}>0$.

Numerické výsledky našich počítačových simulácií [35] molekulovej Darwinovej evolúcie sú znázornené na obrázkoch 20 a 21. Vidíme, že vygenerovaný priebeh priemerného fitnes (diagram D, obrázok 20) je neklesajúca funkcia s niekoľkými relatívne dlhými neutrálnymi etapami – stázami. Diagramy A a B znázorňujú priebehy genotypovej a fenotypovej entropie, ktoré jednoznačne indukujú na konci prvej neutrálnej etapy (okolo epochy 300 000), že existuje podstatná odchýlka od genotypovej homogenity, t.j. v tejto evolučnej etape chemostat je tvorený z niekoľkých fenotypov s rovnakým fenotypom. Táto odchýlka od chemostatov homogénnosti môže byť považovaná ako nutná predbežná fáza pred evolučným prechodom do nového „stavu“ špecifikovaného novým fenotypom s vyšším fitnes. Rovnaké situácie môžu byť identifikované taktiež pre fenotypové prechody A→B a E→F. Diagram C vyjadruje koncentračné profily dominantných fenotypov, ich tvary majú podobný tvar ako koncentračné profily s koncentračnými profilmi vyplývajúcimi z Eigenovej teórie (pozri obrázok 13). Tieto profily tvoria postupnosť dobre separovaných jednotkových maxím. Bodkované vertikálne pomocné čiary indikujú evolučné prechody pomocou entropií. Obrázok 21 znázorňuje postupnosť fenotypov (foldingov), od počiatočného z obrázku 19A, až po konečný – cieľový z obrázku 19B. Podobnosť jednotlivých fenotypov z tejto postupnosti s cieľovým fenotypom určuje fitnes odpovedajúcich binárnych reťazcov, $fitnes(\boldsymbol{x}) = s\left(ph(\boldsymbol{x}), ph_{req} \right)$.



Obr. 3.22. Schematické znázornenie skoku medzi dvoma neutrálnymi etapami, kde sa fitnes mení skokom $f \rightarrow f'$, symboly $\mathcal{G}_f$ a $\mathcal{G}_{f'}$ reprezentujú „neutrálne“ množiny genotypov (binárnych reťazcov) ohodnotených rovnakým fitnes f resp. f'. (A) V prvej neutrálnej etape, evolúcia vykonáva lokálne prehľadávanie v rámci množiny $\mathcal{G}_f$, kde počiatočný genotyp reprezentuje riešenie z posledného „skoku“ a finálne riešenie je schopné „preskočiť“ na riešenie z množiny $\mathcal{G}_{f'}$. Potom sa neutrálna etapa znovu opakuje, až je nájdené v rámci množiny $\mathcal{G}_{f'}$ nové riešenie, ktoré je schopné preskočiť na nasledujúcu neutrálnu množinu. (B) Podrobnejší „rozpis“ neutrálnej etapy s nemenným fitnes $fitnes_i$, postupnosť genotypov $\boldsymbol{x}_1$ až $\boldsymbol{x}_k$ tvorí neutrálnu stáziu, v ktorej sa pripravuje evolúcia na ďalší „skok“ do množiny genotypov s väčším fitnes.



Obr. 3.23. (A) Ilustratívny priebeh povrchu funkcie fitnes pre molekulovú Darwinovu evolúciu, keď neutrálne mutácie neexistujú. Evolučná adaptácia vždy končí v najbližšom lokálnom maxime (v tomto prípade môže byť Darwinova evolúcia charakterizovaná ako lokálny optimalizátor, kde fitnes populácie je „zamrznuté“ na najbližšom maxime). (B) Ak neutrálne mutácie existujú (zobrazenie genotypu na fenotyp je redundantné), potom povrch fitnes funkcie obsahuje množstvo „chodníčkov“ s konštantným fitnes, ktoré umožňujú evolúcii prekonanie hlbokých údolí na povrchu fitnes funkcie. Darwinova evolúcia má dobrú šancu k dosiahnutiu globálneho maxima v priebehu adaptačného procesu. (Parafráza na obrázok z publikácie P. Schustra [45].)

Zo získaných numerických výsledkov vyplýva dôležitosť neutrálnych mutácií v molekulovej Darwinovej evolúcii, umožňujú evolúcii uniknúť z lokálnych riešení na povrchu funkcie fitnes, pozri obrázok 23. V mnohých prípadoch, keď daný genotyp $\boldsymbol{x}$ nie je priamo spojený prostredníctvom 1-bitovej mutácie s iným genotypom $\boldsymbol{x}'$ s vyšším fitnes (t.j. prechod $\boldsymbol{x} \to \boldsymbol{x}'$ reprezentuje evolučný skok), potom evolúcia stochasticky hľadá taký genotyp $\tilde{\boldsymbol{x}}$, ktorý leží v susedstve genotypu $\boldsymbol{x}$ s rovnakým fitnes, avšak tento nový genotyp umožňuje už evolučný skok $\tilde{\boldsymbol{x}} \to \boldsymbol{x}'$ pomocou 1-bitovej mutácie, pričom $fitnes(\boldsymbol{x}') > fitnes(\tilde{\boldsymbol{x}})$, pozri obrázok 22B.

3.8 Evolúcia v multiagentových systémoch

Cieľom tejto časti je aplikovať myšlienky univerzálneho darwinizmu k evolúcii multiagentových systémov, ktorých fitnes je určený hlavne schopnosťou špecificky interagovať s ostatnými agentmi populácie. To znamená, že genotyp agentov špecifikuje spôsob interakcie s inými agentmi, pričom detaily jeho fenotypu – „telesnej schránky“ – sú ignorované. Tieto aplikácie multiagentových systémov spolu s univerzálnym darwinizmem znamenajú v súčasnosti významný prienik informatických metód k simuláciám sociálnych javov, akými sú evolučná emergencia kooperácie, altruizmu, etnickej kooperácie alebo konfliktu, normy správania sa, jazyka, mémov, atď.

Študujme populáciu agentov (organizmov), $P = \{A_1, A_2, ..., A_n\}$, kde agent A je špecifikovaný genotypom $\boldsymbol{x} = \boldsymbol{x}(A)$. Fitnes agentov z populácie je určený vzťahom (pozri sekciu 2)

$$fitnes(A) = \sum_{\substack{A' \in P \\ A' \neq A}} F(A, A')$$

ktorý určuje fitnes agenta A ako sumu „platieb“, ktoré sú priradené interakciám daného agenta s ostatnými agentmi populácie. Fitnes agentov je plne určený „stratégiou“ vzájomnej interakcie, ktorá je dominantná pre ich prežívanie a reprodukciu, zatiaľ čo, architektúra samotného agenta je v tomto prístupu nemenná. V tomto prístupe je fenotyp agenta reprezentovaný stratégiou interakcie s ostatnými agentmi. V zjednodušenom prístupe tieto

stratégie splývajú priamo s genotypom agenta, takže koncepcia dichotómie medzi genotypom a fenotypom agenta v tomto prístupe neplatí.

3.8.1 Sociálne dilemy

Ako vyplýva z teórie hier v ekonómii, aplikovanej tiež často na analýzu pretekov v **zbrojení medzi štátmi**, zjednodušenej zo skúmania ďalej preberaného "problému väzňov" (prisoner's dilemma – väzňova alebo väzenská dilema), kooperácia sa môže vynoriť zo sebeckého záujmu, keď je výhodná pre obidvoch. Vo všeobecnosti, keď hľadáme racionálne dôvody, či pri našej interakcii s niekým iným (s našim obchodným alebo iným partnerom) pristúpiť na kooperáciu, alebo nie, naše rozhodnutie by malo byť založené na racionálnych argumentoch a nie na predsudkoch a ideologických dôvodoch [2].

Medzi sociálne dilemy napríklad patrí oblasť životného prostredia (keď ho znečistím, negatívny vplyv na mňa je malý a úspora veľká, aj keď súčet negatívnych vplyvoch pre všetkých je významnejší ako moja úspora), obchodné vojny, vyjednávanie medzi krajinami a sociálne interakcie. Tu všade je výber medzi altruistickým správaním, kooperatívnym, alebo egoistickým. Na modelovanie toho sa používajú hry, ktoré dovoľujú analýzu rôznych scenárov.

Ľudské spoločnosti sú veľmi komplexné kooperatívne systémy, porovnateľné iba so sociálnym hmyzom. Napriek vonkajšiemu zdaniu spoločnosť okolo nás nie je v svojom rozhodovaní riadená iba krátkodobými sebeckými cieľmi. Ako myšlienkový dôkaz si môžete predstaviť situáciu, že by človeka vysadili do voľnej prírody. Ako dlho by asi prežil a ako by sa mu darilo? Z tejto predstavy môžeme usúdiť, že kooperácia je v rozvinutej spoločnosti nevyhnutná.

Je najskôr potrebné hľadať niečo podobné ako "neviditeľnú ruku trhu", ibaže vedúcu namiesto súťaženia k spolupráci. Tieto vplyvy najskôr nebudú určené prevažne geneticky na rozdiel od hmyzu, kde je víťazstvo kooperácie udržiavané genetickou homogénnosťou spoločenstiev. Z pohľadu tzv. sebeckého génu [10] pomoc blízkemu príbuznému znamená pomoc sebe samému. Ľudská spoločnosť je ale z tohto pohľadu príliš heterogénna, vysvetlenie spolupráce "sebeckým génom" asi nie je to pravé.

V žargónu teórie hier poznáme súťaživé interakcie ako "hry s nulovým súčtom", čo jeden získa, druhý stratí. Tento model je ale pre prax príliš obmedzujúci, zdroje nie sú konštantné. Keď sebeckosť jedincov spôsobuje, že ich zdatnosť ako jednotlivcov aj ako celku sa znižuje, namiesto toho aby sa zvyšovala, ide o "hry so záporným súčtom". Zmena systému tiež môže viesť k využití nevyužívaných zdrojov, kooperácia alebo symbióza odpovedá "hre s kladným súčtom".

V tejto kapitole ukážeme na jeden z možných prístupov k vysvetleniu vzniku kooperácie, ktorý je založený na simulačných výpočtoch pomocou hry nazývanej „***väzňova dilema***".

Nech dvaja zlodeji boli zatknutí políciou pre vylúpenie *Národnej Poctivej Sporiteľne*, a boli separovane umiestnení do cely predbežného zadržania. Každému z páchateľov záleží o mnoho viac na tom, aby bol oslobodený on, ako na tom, aby bol oslobodený jeho spolupáchateľ. Pretože polícia nemala dostatok dôkazov k súčasnému usvedčeniu oboch páchateľov, prokurátor dal každému tento „diabolský" návrh: „*Môžete si vybrať, buď sa priznať k lúpeži, alebo mlčať. Ak sa prizná te vy a váš komplic sa neprizná, tak potom celú vinu za lúpež dám na vášho komplica, ktorý bude odsúdený na mnohoročný trest a vy budete oslobodený. Podobne, ak vy budete mlčať a váš komplic sa prizná k činu, potom on bude oslobodený a vy dostanete mnohoročný trest. Ak budete obaja mlčať a nepriznáte sa, potom, aj keď nerád, navrhnem súdu, aby vás oboch odsúdil iba na podmienečný trest. V opačnom prípade, ak sa obaja prizná te, navrhnem vaše odsúdenie, ale sľubujem vám trest, ktorý tiež bude mnohoročný, ale kratší, ako v prípade, že by sa len jeden z vás priznal a druhý mlčal. Ak*

sa mienite priznať, dozorcovi nič nepovedzte, na druhý deň ráno vás znovu navštívim a môžete mi oznámiť vaše rozhodnutie“.

“Dilema”, pred ktorou stoja obaja väzni, spočíva v tom, že nech sa druhý z nich rozhodne akokoľvek, vždy je lepšie nemlčať a priznať sa, než nepriznať svoju vinu. Táto dilema ilustruje konflikt medzi individuálnou a skupinovou racionálnosťou. Skupina ktorej členovia preferujú svoj vlastný záujem obvykle skončí horšie, ako tá skupina, ktorej členovia uprednostňujú skupinové záujmy. Tento problém bol navrhnutý a široko diskutovaný už počiatkom 50-tych rokov 20 storočia Merrilom Floodom a Melvinom Dresherom vo vtedy veľmi známej Rand Corporation inštitúcii, kde sa vedci snažili použiť teóriu hier k odhaleniu zákonitostí globálnej nukleárnej stratégie. Väzňova dilema sa stala často používaným prístupom k vysvetleniu vzniku kooperácie v rôznych sociálnych, ekonomických a iných systémoch, kde sa pomocou matematickej teórie hier a/alebo počítačových simulácií hľadajú „racionálne dôvody“ k tomu, prečo a za akých okolností je výhodnejšie spolupracovať, než ako nespolupracovať.

3.8.2 Základné princípy hry "väzňovej dilemy"

"Väzňova dilema" **(**VD) [2] je hrou medzi dvoma hráčmi, kde každý z nich má výber z dvoch možností: spolupracovať (cooperate – *C*) alebo podvádzať (defect – *D*). Matica platieb tejto hry je daná takto

$$f = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \Leftrightarrow \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$$

Keď oba hráči spolupracujú, potom dostanú „odmenu“ *R* (reward), ktorá by mala byť väčšia ako bodový zisk „trest“ *P* (punishment) získaný obidvoma hráčmi, keď podvádzajú. Ale keď jeden hráč podvádza (hrá *D*) zatiaľ čo ten druhý spolupracuje (hrá C), potom podvádzajúci získa výplatu bodov *T* (temptation - pokušenia) ktorá je väčšia ako *R*, zatiaľ čo spolupracujúci dostáva počet bodov *S* (sucker – ten, čo naletel) ktorá je menšia ako *P*. Platby spĺňajú nerovnosti $T > R > P > S$ a $2R > T+S$. Druhá nerovnosť znamená, že celková výplata pre dvoch hráčov je väčšia keď obaja spolupracujú v porovnaní so súčtom bodov keď jeden spolupracuje a druhý nie.

Uvažujme hráča – agenta *A*, jeho stratégia $\boldsymbol{s}(A)$ je súbor pravidiel, ktoré použije pri výbere ťahu *C* alebo *D* (spolupráce alebo podvádzania) proti druhému hráčovi, na základe predošlej histórie hry. Pri iba jednej iterácii by samozrejme bolo výhodné zradiť pri polovičnej šanci stretnúť spolupracujúceho, no keď sa iterácie s tým istým spoluväzňom opakujú a vopred nie je známy počet iterácií, metóda ako sa zachovať nie je jasná. Hra sa opakuje t_{max}-krát (dôležitý parameter hry), na záver hry sa spočítajú platby jednotlivých hráčov. Funkcia `Stratégia(x,t)` v nasledujúcom algoritme poskytuje stratégiu hráča *x*(= *a*, *b*) v čase *t*. Pre jednoduchosť v našich ďalších úvahách budeme používať stratégiu, pri ktorej hráč zvolí ťah na základe predchádzajúceho ťahu súpera. Stratégia je určená pomocou 3-bitového vektora (t.j. hráči si pamätajú predchádzajúci ťah oponenta) $\boldsymbol{s}=\{s_1, s_2, s_3\} \in \{0,1\}^3$, kde jednotlivé zložky (0 – kooperácie, 1 – nekoooperácia) majú túto interpretáciu:

(1) s_1 určuje prvý ťah hráča: ak $s_1=0$, potom hráč kooperuje, v opačnom prípade, ak $s_1=1$, nekooperuje.

(2) s_2 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku kooperoval: ak $s_2=0$, potom hráč kooperuje, v opačnom prípade, ak $s_2=1$, nekooperuje.

(3) s_3 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku nekooperoval: ak $s_3=0$, potom hráč kooperuje, v opačnom prípade, ak $s_3=1$, nekooperuje.

Takáto stratégia vyžaduje len krátkodobú pamäť hráčov, požaduje len znalosť posledného kroku protivníka - druhého hráča. Celkový počet rôznych 1-stratégií je $2^3=8$.

Keby si väzni mali pamätať viac krokov do minulosti, s každým takým krokom by exponenciálne narastala zložitosť zápisu stratégie.

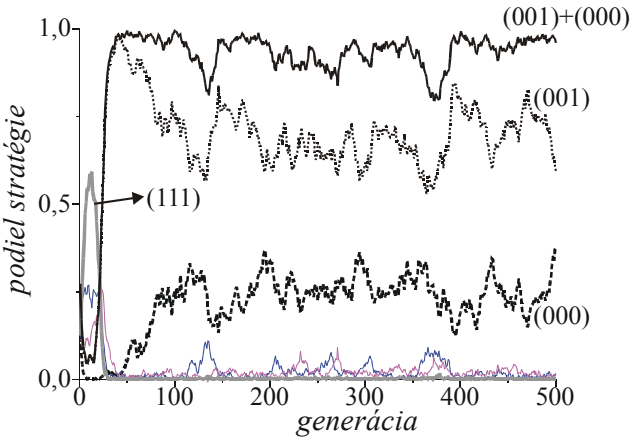
Majme dve stratégie $s(A)$=(010) a $s(B)$=(101), potom nasledujúcich 5 ťahov hráčov *a* a *b* a taktiež aj jednotlivé platby na základe matice platieb (1) sú určené pomocou nasledujúcej tabuľky

ťah	agent *A*		agent *B*	
	ťah	platba	ťah	platba
1	0	0	1	5
2	0	3	0	3
3	1	5	0	0
4	1	1	1	1
5	0	0	1	5
suma:		9		14

V prvom ťahu oba hráči použijú ťahy, ktoré sú určené prvou zložkou ich stratégií. V nasledujúcich ťahoch už sú determinovaní ťahom súpera v predchádzajúcom ťahu.

3.8.3 Simulácia emergencie stabilnej stratégie pomocou genetického algoritmu

Pri plánovaní stratégie na základe predchádzajúcich skúseností s protihráčom sa často používa genetický algoritmus. Táto aplikácia genetického algoritmu prvýkrát vznikla zo spolupráce amerického politológa Roberta Axelroda [2] s informatikom a tvorcom genetických algoritmov Johnom Hollandom a rozvíjala sa ďalej pre zložitejšie modely medzinárodnej bezpečnosti odpovedajúce napr. na otázku, koľko z rozpočtu vydávať na obranu. V časovom priebehu je zaujímavé sledovať vývoj kooperatívnych a nekooperatívnych stratégií. Kooperatívne stratégie sa spočiatku ukázali menej výhodné, no v neskorších štádiách prevážili.



Obr. 3.24. Výsledky genetického algoritmu použitého pre emergenciu výslednej stratégie, pričom populácia stratégií bola inicializovaná náhodne. V priebehu evolúcie algoritmu spontánne emergovali dve „kooperatívne" stratégie (001) a (000), ktoré zahajujú hru kooperáciou a na kooperáciu súpera v predchádzajúcom ťahu odpovedajú kooperáciou v nasledujúcom ťahu. V dôsledku toho, že väčšinová stratégia (001) neprodukuje „podvod", menšinová stratégia (000) nemá príležitosť reagovať spoluprácou na podvod.

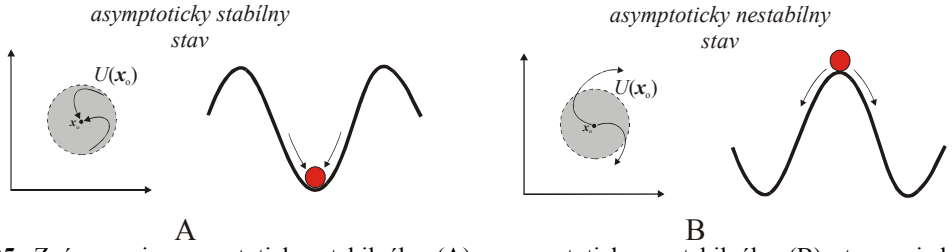
Základné princípy genetického algoritmu použitého k štúdiu emergencie „najvýhodnejšej" stratégie sú rovnaké ako u klasického genetického algoritmu. Populácia je

zložená z chromozómov - stratégií. Algoritmus je inicializovaný náhodne vytvorenou populáciou stratégií. Fitnes chromozómov sa určí pomocou "megaturnaja", kde všetky možné dvojice hrajú VD hru, celková platba každého chromozómu získaná v tomto megaturnaji je fitnes. Štandardná realizácia reprodukčného procesu obsahuje selekciu dvoch stratégií na základe ich fitnes (stratégie s väčším fitnes majú väčšiu pravdepodobnosť byť vybraté do reprodukčného procesu), ktorá sa obvykle realizuje pomocou Goldbergovej rulety. Vybraté stratégie vstupujú do vlastného procesu reprodukcie, ktorý obsahuje kríženie a mutáciu. Výsledné stratégie – potomkovia tvoria novú populáciu, ktorá keď obsahuje rovnaký počet potomkov, ako mala pôvodná populácia počet rodičov, nahradí rodičovskú populáciu populáciou potomkov. Numerické výsledky genetického algoritmu sú znázornené na obr. 24. Ako výsledná stratégia emergujú dve podobné stratégie (001) a (000), ktoré obe zahajajú hru kooperáciou a na kooperáciu odpovedajú kooperáciou. Líšia sa v reakcii na podvod, prvá stratégia (001) reaguje na podvod podvodom, zatiaľ čo druhá stratégia (000) „hlúpo“ reaguje na podvod kooperáciou. Stratégia (001) sa nazýva podľa Axelroda tit-for-tat (vo voľnom preklade, ako ty mne, tak ja tebe). Menšinová stratégia „dobrák“ (000) má šancu prežívať iba popri stratégii tit-for-tat.[9]

3.8.4 Evolučne stabilné stratégie

Z evolučnej teórie hier zavádza veľmi užitočný pojem „evolučne stabilná stratégia“, ktorý je vo veľmi blízkej analógii s pojmom „asymptoticky stabilný stav“, pochádzajúcom z dynamickej teórie systémov [41], pozri obr. 25.

Asymptoticky stabilný stav je charakterizovaný podmienkou, že ak ho trochu vychýlime z jeho rovnovážnej polohy, potom v dynamickom systéme nastanú procesy, ktoré ho vrátia do pôvodnej polohy. Asymptoticky nestabilný stav sa už nevráti po vychýlení do pôvodnej polohy. Aby sme mohli použiť túto analógiu z dynamických systémov, pojem stav nahradíme novým pojmom „*stratégia*“, ktorý špecifikuje interakcie agenta s inými agentami populácie. Na základe analógie z teórie dynamických systémov, definujme *evolučne stabilnú stratégiu* (ESS) [39] tak, že ak populácia P obsahuje agentov prevažne len s referenčnou stratégiu a taktiež podstatne menej iných agentov s menšinovou stratégiou, potom evolučné procesy spontánne prebiehajúce v populácii po určitom čase túto menšinovú stratégiu eliminujú z populácie.



Obr. 3.25. Znázornenie asymptoticky stabilného (A) a asymptoticky nestabilného (B) stavu s jednoduchou fyzikálnou interpretáciou guľôčky, ktorá je buď v jamke (príklad asymptoticky stabilného stavu), alebo je na vrchole kopčeka (príklad asymptoticky nestabilného stavu). Ľavé časti obrázkov odpovedajú presnej matematickej definícii stavov, pre asymptoticky (ne)stabilný stav x_0 existuje také okolie $U(x_0)$, že ak trajektória má počiatočný stav v tomto okolí, potom asymptoticky (pre $t\to\infty$) trajektória vždy končí v bode x_0 (vždy opúšťa okolie).

[9] ďalšia úspešná stratégia založená na trocha inom princípe sa volá Pavlov. Keď jedinec dostane odmenu T alebo R, opakuje svoj krok, keď dostane P alebo S, prepne z nespolupráce na spoluprácu alebo naopak. Ešte zložitejšie sú pravdepodobnostné stratégie.

Nech populácia P obsahuje prevažne väčšinovú referenčnú stratégiu $\boldsymbol{s}_0$ a v malom rozsahu aj menšinovú stratégiu $\boldsymbol{s}_1$.

$$P = P_0 \cup P_1,\ P_0 = \{\boldsymbol{s}_0, \boldsymbol{s}_0, ..., \boldsymbol{s}_0\},\quad |P_0| = p,\quad P_1 = \{\boldsymbol{s}_1, \boldsymbol{s}_1, ..., \boldsymbol{s}_1\},\quad |P_1| = q$$

pričom predpokladáme, že počet stratégií $\boldsymbol{s}_0$ je podstatne viac ako menšinových stratégií $\boldsymbol{s}_1$, $p>>q>>0$ (pozri obr. 26). Fitnes týchto dvoch stratégií je určený takto

$$f(\boldsymbol{s}_0) = (p-1)F(\boldsymbol{s}_0, \boldsymbol{s}_0) + qF(\boldsymbol{s}_0, \boldsymbol{s}_1)$$

$$f(\boldsymbol{s}_1) = (q-1)F(\boldsymbol{s}_1, \boldsymbol{s}_1) + pF(\boldsymbol{s}_1, \boldsymbol{s}_0)$$

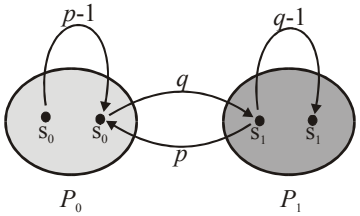
Pretože evolúcia populácie prebieha na základe prirodzeného výberu, podmienka evolučnej stability stratégie $\boldsymbol{s}_0$ má tvar

$$f(\boldsymbol{s}_0) > f(\boldsymbol{s}_1) \Rightarrow (p-1)F(\boldsymbol{s}_0, \boldsymbol{s}_0) + qF(\boldsymbol{s}_0, \boldsymbol{s}_1) > (q-1)F(\boldsymbol{s}_1, \boldsymbol{s}_1) + pF(\boldsymbol{s}_1, \boldsymbol{s}_0)$$

Z predpokladu $p>>q>>0$ vyplýva, že stratégia $\boldsymbol{s}_0$ je evolučne stabilná, ak platí buď prvá podmienka (dominantná) alebo druhá podmienka (vedľajšia, keď dominantná neplatí)

$$F(\boldsymbol{s}_0, \boldsymbol{s}_0) > F(\boldsymbol{s}_1, \boldsymbol{s}_0)$$

$$F(\boldsymbol{s}_0, \boldsymbol{s}_0) = F(\boldsymbol{s}_1, \boldsymbol{s}_0) \text{ a } F(\boldsymbol{s}_0, \boldsymbol{s}_1) > F(\boldsymbol{s}_1, \boldsymbol{s}_1)$$



Obr. 3.26. Výpočet fitnes pre referenčnú stratégiu $\boldsymbol{s}_0$ a menšinovú stratégiu $\boldsymbol{s}_1$. Podpopulácia P_0 obsahuje p referenčných stratégií $\boldsymbol{s}_0$ a podpopulácia P_1 obsahuje q menšinovej stratégie $\boldsymbol{s}_1$.

Podmienka ESS pre referenčnú stratégiu $\boldsymbol{s}_0$ vzhľadom k menšinovej stratégii obsahuje dve kritéria, buď je splnené prvé kritérium, alebo ak nie je splnené (podmienka rovnosti), potom musí platiť splnené druhé kritérium, ktoré je však podstatne subtílnejšie vzhľadom k prvému kritériu (rezultujúci fitnes stratégií je násobkom platieb parametra q, ktorý je však podstatne menší ako parameter p, od ktorého je odvodené prvé kritérium).

Týmto sme dostali jednoduché podmienky, ktoré špecifikujú či daná stratégia $\boldsymbol{s}_0$ je v prítomnosti malého počtu votreleckých stratégií $\boldsymbol{s}_1$ evolučne stabilná. Podmienky požadujú len znalosť funkcie platieb pre jednotlivé stratégie. Tieto platby urobené ako súčet platieb z piatich ťahov za sebou pre všetky možné dvojice stratégií sú ukázané v nasledujúcej tabuľke

	(000)	(001)	(010)	(011)	(100)	(101)	(110)	(111)
(000)	15	15	3	3	12	12	0	0
(001)	15	15	12	6	4	10.	9	4
(010)	23	12	11	4	18	9	0	0
(011)	23	11	19	7	20	8	16	4
(100)	17	14	8	5	13	1	4	1
(101)	17	15	14	8	15	5	10	5
(110)	25	14	25	6	19	10	9	1
(111)	25	9	25	9	21	5	21	5

Dvojica stratégií 001 a 111 vyhovuje podmienke $F(\boldsymbol{s}_0, \boldsymbol{s}_0) > F(\boldsymbol{s}_1, \boldsymbol{s}_0)$, teda by teoreticky mohli byť stabilné, keby nebolo dvojíc stratégií 000 – 001 a 101 – 111. Tieto dvojice stratégií majú pri vzájomnej hre rovnaké ohodnotenie. Stratégia 000 sa teda popri stratégii 001 môže

náhodnými mutáciami rozšíriť. Potom, čo sa rozšíri, môže náhodne vzniknúť ďalšia stratégia, ktorá ich využije na svoje rozšírenie. Podobne to platí aj pre dvojicu stratégií 101 – 111.

Keby stratégie 000 a 101 boli zakázané, potom by stratégie 001 a 111 boli evolučne stabilné. Každá dvojica by bola v tzv. Nashovej rovnováhe[10], teda stratégia, ktorá by nahradila jednu stratégiu z dvojice 001 – 001 alebo z dvojice 111 – 111 by si zhoršila platbu. V prípade nahradenia tit-for-tat nespolupracujúcim by síce nespolupracujúci skončil horšie ako predtým, ale relatívne lepšie ako jeho spoluhráč. Vzhľadom k tomu, že sa ale fitnes porovnáva v rámci celej populácie, nespolupracujúci by na tom bol stále zle, jeho spoluhráč tit-for-tat by nabral fitnes v interakciách s ostatnými tit-for-tat (pozri obrázok 27).

Podobne sa dajú študovať aj stratégie agentov, ktorí si pamätajú dva alebo viac krokov do minulosti. Spolu so zložitosťou stratégií ale narastá tendencia k nespolupráci a chaotické správanie populácie, aj keď je možné tieto tendencie prekonať napríklad ustanovením „administrátora“, ktorý by stochasticky trestal nespoluprácu.

Pomocou tohto výsledku pokúsime sa odvodiť podmienku pre ESS stratégie $\boldsymbol{s}_0$, ktorá obsahuje len diagonálne platby. Nech nediagonálna platba pre $\boldsymbol{s}_0$ a $\boldsymbol{s}_1$ je približne určená vzťahom

$$F\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right) \doteq \frac{1}{2}\left[F\left(\boldsymbol{s}_0,\boldsymbol{s}_0\right)+F\left(\boldsymbol{s}_1,\boldsymbol{s}_1\right)\right]\left(1-\frac{d\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)}{d_{max}}\right)$$

kde $d\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)$ je „vzdialenosť“ medzi stratégiami $\boldsymbol{s}_0$ a $\boldsymbol{s}_1$ a d_{max} je jej postulovaná maximálna hodnota. Nediagonálna platba $F\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)$ je teda určená ako priemer príslušných diagonálnych hodnôt krát člen špecifikujúci podobnosť medzi stratégiami. Dosadením tohto vzťahu do prvej podmienky ESS po jednoduchých algebraických úpravách dostaneme

$$F\left(\boldsymbol{s}_0,\boldsymbol{s}_0\right) > F\left(\boldsymbol{s}_1,\boldsymbol{s}_1\right)\frac{d_{max}-d\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)}{d_{max}+d\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)}$$

Nech stratégia s_1 vznikne zo stratégie s_0 1-bodovou (minimálnou) mutáciou, potom môžeme položiť $d\left(\boldsymbol{s}_1,\boldsymbol{s}_0\right)=1$, podmienka pre evolučnú stabilitu s_0 vzhľadom k s_1 má tento konečný tvar

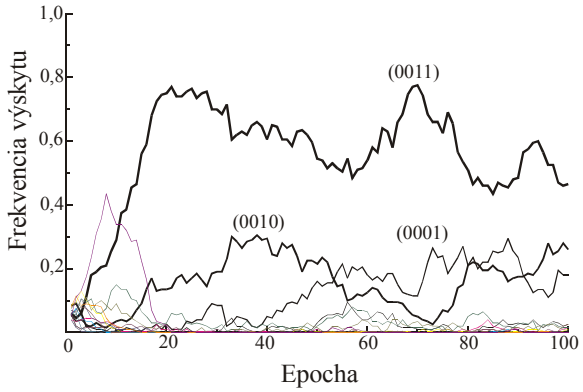
$$F\left(\boldsymbol{s}_0,\boldsymbol{s}_0\right) > F\left(\boldsymbol{s}_1,\boldsymbol{s}_1\right)\frac{d_{max}-1}{d_{max}+1}$$

Táto podmienka je podstatne jednoduchšie aplikovateľná ako jej pôvodný tvar požadujúci aj znalosť nediagonálnych platieb. Jeden z dôležitých dôsledkov tejto formule je výsledok, že ak $\boldsymbol{s}_0$ je nejaká evolučne výsledná dominantná stratégia a predpokladáme, že vedľajšie stratégie z nej odvoditeľné pomocou 1–bodovej mutácie sú veľmi blízke pôvodnej dominantnej stratégii $\boldsymbol{s}_0$, $F\left(\boldsymbol{s}_0,\boldsymbol{s}_0\right)=F\left(\boldsymbol{s}_1,\boldsymbol{s}_1\right)$, potom dominantná stratégie je evolučne stabilná, lebo $d_{max}-1<d_{max}+1$. K tomu, aby sa vedľajšia stratégie stala dominantnou je potrebné, aby diagonálna platba $F\left(\boldsymbol{s}_1,\boldsymbol{s}_1\right)$ bola podstatne väčšia ako platba $F\left(\boldsymbol{s}_0,\boldsymbol{s}_0\right)$, potom aj napriek pomerne veľkej zotrvačnosti pretrvávania pôvodnej stratégie $\boldsymbol{s}_0$ presadí sa evolučne (po peripetiách jej fluktuačného vracania sa do dominantného postavenia) odvodená stratégia $\boldsymbol{s}_1$.

[10] Keď existuje súbor stratégií s takou vlastnosťou, že žiaden hráč nemôže profitovať na zmene svojej stratégie, zatiaľ čo ostatní hráči svoju stratégiu nezmenia, potom tento súbor stratégií je v Nashovej rovnováhe (spomeňte si na film Čistá duša – Beautiful mind).

3.8.5 Tragédia spoločného

Modelový príklad egoistických tendencií erodujúcich spoluprácu je tzv. tragédia spoločného definovaná Garrettom Hardinom r. 1968 [24]: "Predstavte si pastvinu prístupnú pre všetkých. Dá sa očakávať, že každý pastier sa pokúsi držať na spoločnej pastvine toľko dobytka, ako sa len dá. Ako racionálny jedinec, každý pastier sa snaží maximalizovať svoj zisk. Explicitne alebo implicitne sa pýta: "Akú zmenu výnosu pre mňa znamená prídavok jedného alebo viac zvierat k môjmu stádu?"
1. Pozitívna časť je funkciou prínosu jedného zvieraťa. Pretože pastier dostáva všetok zisk z predaja zvieraťa, možný príspevok sa blíži + 1.
2. Negatívna časť je funkciou prídavku prílišného spásania pridaným zvieraťom. Pretože ale výsledky prílišného spásania sú zdieľané všetkými pastiermi, negatívny príspevok pre ktoréhokoľvek rozhodujúceho sa pastiera je iba zlomkom z - 1.
Keď si racionálny pastier spočíta pozitívnu a negatívnu časť prínosu, vychádza mu, že jediné rozumné riešenie je preňho pridať ďalšie zviera k stádu. A ešte ďalšie ...
Ale toto je záver každého racionálneho pastiera zdieľajúceho pastvinu. V tom je tá tragédia. Každý človek je uväznený v systéme, ktorý ho núti zväčšovať svoje stádo bez obmedzenia - v svete ktorý je limitovaný. ... Sloboda v užívaní spoločnej pastviny prináša skazu pre všetkých."



Obr. 3.**27.** Tragédia spoločnej pastviny a jej modelovanie troma pastiermi. Výsledok genetického algoritmu – zobrazenie frekvencie výskytu stratégií pre spoluprácu troch hráčov. Ako je vidno, vyhráva stratégia 0011, ktorá znamená, že na začiatku je výhodné spolupracovať, a pokiaľ všetci spolupracujú, aj pre mňa je výhodné spolupracovať. Pokiaľ ale iba jeden nespolupracuje, je aj pre mňa výhodné nespolupracovať. Ku koncu vývoja sa vyskytuje aj stratégia 0001 a 0010, ktorá pripúšťa možnosť spolupráce, aj keď jeden alebo dvaja spoluhráči z turnaja nespolupracujú. To je ale výchylka spôsobená tým, že ku koncu vývoja sa vyskytujú väčšinou jedinci zo stratégiou 0011, a teda situácia stretnutia s nespolupracujúcimi jedincami prakticky nenastáva.

Vyše uvedený príklad sa dá modelovať tiež pomocou väzňovej dilemy, kedy sa namiesto dvoch väzňov interagujú traja alebo viac. Zovšeobecnená väzňova dilema pre viac hráčov (konkrétna stratégia je pre troch) má potom stratégiu $\boldsymbol{s} = (s_1, s_2, s_3, s_4) \in \{0,1\}^4$ špecifikovanú takto

$s1$	..... prvý ťah
$s2$	..... ťah ak všetci 3 oponenti kooperovali
$s3$	..... ťah ak aspoň 2 oponenti kooperovali
$s4$	..... ťah ak aspoň 1 oponent kooperoval

Potom napríklad stratégia (0 1 1 0) v prvom kroku kooperuje, nekooperuje, keď v predošlom kroku všetci ostatní kooperovali, nekooperuje, keď v predošlom kroku jeden spoluhráč

nekooperoval, kooperuje, keď v predošlom kroku všetci ostatní kooperovali. Výsledok je obr. 27.

Všeobecné vlastnosti víťaznej stratégie: hra je ***zahájená kooperáciou***, ak v predchádzajúcom ťahu ***všetci oponenti kooperovali, potom kooperuj***, ak v predchádzajúcom ťahu ***aspoň jeden oponent nekooperoval, potom nekooperuj***.

Pretože vo väčších celkoch sa vždy nájde nekooperujúci jedinec, nastáva všeobecná nespolupráca - tragédia spoločnej pastviny napr. u rybolovu v medzinárodných vodách, pri rúbaní pralesov, starosti o spoločné zavodňovanie alebo znečisťovaní atmosféry. Politológ Axelrod tvrdí, že k tomu, aby v spoločnosti vznikla kooperácia, musia v nej existovať inštitúcie, ktoré kontrolujú - penalizujú nekooperáciu.

Tabuľka pre prvých päť ťahov zovšeobecnenej väzňovej dilemy pre troch hráčov, ktorých stratégie sú určené ako (0110), (1100), (0011)

No.	Ťah			Skóre		
	1.hráč	2. hráč	3. hráč	1.hráč	2. hráč	3. hráč
1	0	1	0	3	10	3
2	1	1	1	2	2	2
3	0	0	1	3	3	10
4	1	0	0	10	3	3
5	1	0	1	6	0	6
Celkové skóre:				24	18	24

Okrem zavedenia administrátora trestajúceho nespoluprácu sa emergencia spolupráce dá dosiahnuť aj priestorovým rozmiestnením stratégií, ktoré potom môžu interagovať iba s blízkymi stratégiami, alebo zavedením "dobrého mena" jedincov, kedy si jedinci vymieňajú informácie o spolupráci či nespolupráci ostatných, a k tým, čo nespolupracovali s ostatnými, sa správajú, ako keby nespolupracovali s nimi - tzv. moralistická agresia. Ďalšou možnosťou je aj tzv. skupinový výber, kde súťaž prebieha nielen medzi jedincami v rámci populácie, ale aj medzi populáciami. Populácia s príliš veľa sebcami má oproti ostatným populáciám nižšie priemerné skóre u jedincov, preto zanikne. U biológov je táto predstava uznávaná iba obmedzene, pretože rýchlosť výmeny jedincov je väčšia ako rýchlosť výmeny populácií, a teda populácia sa skôr nakazí egoizmom, ako by mala zvíťaziť nad inou už egoistickou populáciou. U spoločností ale toto vysvetlenie môže mať svoj význam.

Evolučné štúdium väzňovej dilemy sa dá použiť aj na modelovanie vnútro- a medzietnických vzťahov [38]. U najjednoduchšieho modelu s dvoma etnikami medzi jedincami populácie *P* existujú tri typy interakcií: vnútroetnické interakcie v rámci etnickej skupiny *A*, vnútroetnické interakcie v rámci etnickej skupiny *B*, a konečne, medzietnické interakcie medzi jedincami z *A* a *B*. Postulujeme, že frekvencia medzietnickej interakcie je oveľa menšia ako frekvencia vnútroetnických interakcií v *A* alebo *B*. Jedinec je vybavený zafixovanou stratégiou väzenskej dilemy, dokáže rozpoznať etnickú skupinu druhého a jeho penalizáciu, čo je označenie nespolupracujúcich jedincov (alebo etnika) spolupracujúcimi jedincami. Evolúcia populácie prebieha nezávisle pre obidve etnické skupiny *A* a *B*. Fitnes jedincov je vypočítaný pomocou "turnaja", kde je predpísaný počet-krát náhodne vybraná dvojica agentov, ktorí spolu hrajú jeden ťah hry väzňovej dilemy. K štúdiu medzietnických interakcií boli použité dve rozdielne evolučné stratégie.

V prvej stratégii, ktorá sa nazýva *kolektívna vina*, je potrestaná celá etnická skupina za akt nespolupráce svojho člena voči členovi druhej etnickej skupiny. Zo simulačných výpočtov a teoretických úvah vyplýva, že tento typ stratégie nevedie k emergencii spolupráce medzi rozdielnymi etnickými skupinami, pokiaľ nezavedieme administrátora pokutujúceho akty nespolupráce. Keď etnické skupiny majú veľmi rozdielne veľkosti, princíp kolektívnej viny

môže byť veľmi devastujúci pre menšiu etnickú skupinu, v ktorej môže zaniknúť aj vnútroetnická spolupráca.

Druhou použitou stratégiou pre podporu spolupráce je tzv. *osobná zodpovednosť*, kedy člen skupiny, ktorý nespolupracuje v rámci medzietnických interakcií, je potrestaný osobne v rámci svojej etnickej skupiny. To znamená, že na rozdiel od princípu kolektívnej viny, v tejto stratégii existuje len jeden typ "potrestania", členovia sú označení penalizáciou osobne. Simulačné výpočty spolu s teoretickou analýzou ukazujú, že v tomto prípade môže emergovať medzietnická spolupráca.

3.9 Závery

Metafora univerzálneho darwinizmu rozširuje princípy darwinizmu ako univerzálneho prístupu k štúdiu evolúcie nielen v biológii, ale aj v iných vedných oblastiach (chémia, sociálne vedy, informatika, umelá inteligencia, kognitívna veda,...). V informatike univerzálny darwinizmus poskytuje jedinečnú možnosť simulácií darwinovskej evolúcie prostriedkami „in silico“ (t. j. počítačom), môžeme konštatovať tieto všeobecné závery:

- Evolučné algoritmy tvoria robustný optimalizačný prostriedok pre rôzne kombinatoriálne úlohy (napr. z teórie grafov, operačného výskumu, inžinierskej praxe návrhu zložitých zaradení, a pod.).
- Evolučné multiagentové systémy umožňujú počítačové simulácie sociálnych systémov, emergenciu kooperácie a komunikácie vypracovať evolučné základy memetiky ako špeciálneho prípadu multiagentových systémov, študovať etnické konflikty v multiagentových systémoch, a pod.
- Univerzálny darwinizmus umožňuje informatike aktívne interagovať so sociálnymi a behaviorálnymi vedami.

Z ilustračných výpočtov molekulovej Darwinovej evolúcie vyplývajú tieto dôležité všeobecné závery pre univerzálny darwinizmus:

(1) Eigenova ***teória replikátorov*** tvorí dobrý fenomenologický základ molekulovej darwinovej evolúcie. Priebehy koncentračných profilov získaných riešením diferenciálnych rovníc popisujúcich dynamiku chemostatu replikátorov, ktoré boli získané vhodným výberom rýchlostných konštánt, poskytujú priebehy koncentrácií, ktoré sú veľmi podobné experimentálne pozorovaným [47] (pozri obrázok 13).

(2) Darwinovská evolúcia prebieha ako Monodovská [42] vzájomná hra medzi ***náhodou a nutnosťou***, medzi deterministickým a stochastickým charakterom, ktoré sú jej integrálnou súčasťou. Evolúcia obsahuje časti, ktoré sú plne deterministické – plne predpovedateľné (napr. zobrazenie genotypu na fenotyp), a taktiež aj stochastické časti, ktoré nemôžu byť dobre predpovedateľné, môže hovoriť len o ich štatistických charakteristikách (napr. mutácie). V tejto súvislosti sa občas používa terminológia zavedená americkým informatikom a zakladateľom umelého života Christopherom Langtonom „na hrane chaosu“ (angl. *on the edge of chaos*). Používa sa termín „evolúcia prebieha na hrane chaosu“, ktorý sa interpretuje takto: Hlavnou hnacou silou inovácií v evolúcii sú mutácie, preto by sme sa mohli domnievať, že čím častejší je výskyt mutácií pri reprodukčnom procese, tým väčšia je pravdepodobnosť výskytu inovácie. Avšak, ak miera mutácií prekročí určitú kritickú hodnotu, efektívnosť evolúcie sa zničí a stáva sa z nej niečo čo môžeme nazvať „slepé hľadanie“, bez vplyvu stratégie prirodzeného výberu, ktorý je schopný efektívne selektovať pozitívne mutácie. V opačnom prípade, ak miera výskytu mutácií je veľmi malá, evolúcia sa v podstate zastaví a zamrzne v aktuálnom stave. Preto môžeme povedať, že frekvencia výskytu mutácií musí

byť dostatočne veľká, ale zas nie moc, aby sa evolúcia nerozpadla, hovoríme, že musí byť na hrane chaosu.

(3) Koncepcia Sewalla Wrighta [51] ***povrchu funkcie fitnes*** môže byť považovaná ako jeden z najväčších úspechov Darwinovej evolučnej teórie a taktiež aj ako jeden zo zdrojov jej univerzálneho charakteru. Táto koncepcia má veľký heuristický význam v tom, že Darwinova evolúcia môže byť interpretovaná ako špeciálny druh evolučného algoritmu [15,25,37], ktorý je schopný riešiť extrémne komplikované kombinatoriálne optimalizačné problémy. V prvej sekcii bol formulovaný Wrightov názor na evolúciu ako na optimalizačný proces na povrchu funkcie fitnes (pozri obrázok 2). V tejto súvislosti Wright poznamenal, že jeden z hlavných cieľov teórie evolúcie je popísať mechanizmy, ktoré umožňujú evolúcii nájsť „spojité cesty" z jedného vrcholu do druhého vrcholu na povrchu fitnes funkcie, k tomu, aby evolúcia bola schopná nájsť globálne riešenie (maximum). Ak by takáto možnosť neexistovala, potom by evolúcia vždy „zamrzla" v lokálnom riešení, ktoré by nebola schopná opustiť. Naše modelové výpočty pomocou molekulovej Darwinovej evolúcie umožňujú popísať evolučný mechanizmus úniku z lokálnych riešení (toto riešenie je široko študované viedenským teoretickým chemikom a biológom Petrom Schusterom so spolupracovníkmi [45,46,15-20]) pomocou existencie neutrálnych mutácií, ktoré menia genotyp ale neprejavujú sa zmenou fenotypu (foldingu).

(4) Ako už bolo poznamenané v predchádzajúcom bode, použitý model molekulovej evolúcie umožňuje vysvetliť ***úlohu neutrálnych mutácií*** v Darwinovej evolučnej teórii. (pozri obrázky 22 a 23). Existencia neutrálnych mutácií na povrchu funkcie fitnes je významná pre vymanenie sa z evolučných pascí lokálnych miním. Darwinova evolúcia obsahuje dlhodobé neutrálne stázy, kde priemerný fitnes populácie je konštantný (môže vykazovať malé fluktuácie), ale zloženie populácie stochasticky „vandruje" smerom ku genotypu, z ktorého je možný „skok" na genotyp s väčším fitnes, pomocou 1-bitovej mutácie. Dĺžka týchto neutrálnych evolučných epoch je úmerná pravdepodobnosti náhodného „skoku" na genotyp s väčším fitnes. Ak je táto pravdepodobnosť veľmi malá, potom neutrálna epocha môže trvať veľmi dlho. V opačnom prípade, ak je pravdepodobnosť „skoku" veľká, tak potom neutrálna epocha netrvá dlho. Existencia neutrálnych mutácií je podmienená typom zobrazenia genotypu na fenotyp, ak je toto zobrazenie vysoko redundantné, tak potom pre daný genotyp existuje mnoho neutrálnych mutácií, ktoré zachovávajú nemennosť fenotypu.

(5) ***Časová orientácia*** Darwinovej evolúcie je jednoznačná, táto skutočnosť sa prejavuje napr. existenciou neklesajúceho priebehu priemerného fitnes. Genotypy s väčším fitnes sa replikujú častejšie, tieto genotypy majú určitú selekčnú výhodu pred ostatnými genotypmi s malým fitnes. Inými slovami, Darwinova evolúcia je taká progresívna zmena priemerného populačného genotypu, ktorá sa prejaví neklesajúcim charakterom priemerného fitnes v priebehu celej evolúcie.

(6) Dve rôzne časové merítka môžu byť v evolúcii identifikované. ***Adaptívna epocha*** spočíva v náhlej zmene priemerného fitnes (pozri obrázok 20, diagram D), kde koexistujú dva rozdielne fenotypy; pôvodný fenotyp má menší fitnes, zatiaľ čo nový fenotyp má väčší fitnes. Pretože pravdepodobnosť replikácie genotypov je úmerná ich fitnes, genotypy priradené novému fenotypu majú väčšiu pravdepodobnosť sa replikovať, než pôvodné genotypy. Pretože nové genotypy v populácii rýchlo zvíťazia, adaptívna fáza, kde staré genotypy sú nahradzované novými genotypmi, sa zdá externému pozorovateľovi ako extrémne rýchly proces. Na druhej strane, Darwinova evolúcia obsahuje ***aj neutrálnu epochu***, ktorá spočíva v dlhodobej stáze, kde existujú stochastické mutácie pripravujúce nasledujúcu adaptívnu epochu.

(7) Evolučné vlastnosti chemostatového modelu sú inštruktívne špecifikované pomocou entropií dvoch typov [35]. ***Fenotypová entropia*** postihuje zmeny v zložení binárnych reťazcov vzhľadom k ich fenotypu (fitnes), to znamená, že tento typ entropie nepostihuje neutrálne mutácie. Druhý typ entropie, ***genotypová entropia***, je veľmi citlivá na každé zmeny v chemostate, včítane aj neutrálnych mutácií. Preto súčasné sledovanie oboch entropií v priebehu evolúcie nám umožňuje odlíšenie medzi neutrálnymi a inými mutáciami na evolučnej trajektórie chemostatu.

(8) V evolúcii neexistuje zámer, ktorý by ju riadil smerom k požadovanému cieľu. Ako bolo uvedené v predchádzajúcej časti tejto práce, fitnes replikátorov odráža aktuálny stav replikátorov, ich aktuálnu schopnosť vstupovať do reprodukčného procesu a prežívať v populácii. Môžeme povedať, že Darwinova evolúcia je oportunistická, kde rekurzívne obnovovanie populácie je založené len na parametroch aktuálnej populácie. Táto skutočnosť platí len v ohraničenej miere v multiagentových systémoch, kde dochádza ku koevolúcii génov a mémov, ktoré môžeme charakterizovať ako určitý manuál na prežívanie v danom prostredí. V tomto momente môže dochádzať k určitému „kultúrnemu“ zámeru, ktorý môže ovplyvňovať fitnes génov.

(9) ***Aké sú limity chemostatového modelu***? Použitý tvar genotypu ako binárnych reťazcov a ich zobrazenie na fenotyp pomocou foldingu reprezentuje extrémne jednoduchý model živých organizmov, ktorý postihuje také jednoduché živé organizmy, akými sú vírusy, bakteriofágy a jednoduché prokaryotické baktérie. Zložitejšie organizmy model genotypu (a jeho zobrazenie na fenotyp) musia byť študované zložitejšími modelmi genotypu, ktoré majú premenlivú dĺžku, hierarchickú štruktúru a iné možné mutácie, ako sú použité 1-bitové mutácie. Jeden z najdôležitejších problémov súčasnej teórie Darwinovej evolúcie je vysvetlenie vzniku tzv. „***neredukovateľných zložitostí***“ [4] . Tento model nie je schopný vysvetliť tento problém ani na elementárnej úrovni. Moderné snahy v „umelom živote“ sú sústredené do oblastí, ktoré sú potenciálne dôležité pre tvorbu všeobecnejšej teórie darwinovej evolúcie, ktorá už v nejakej miere bude postihovať tvorbu neredukovateľných zložitostí. Intenzívne sú študované modulárne aspekty genotypu [5,30,31,49] a taktiež aj modelovanie symbiózy [29,49], oba tieto problémy už vyžadujú genotypy zložitejšieho tvaru, ako ten, ktorý bol použitý v našich simuláciách (lineárny reťazec symbolov konštantnej dĺžky).

Literatúra

[1] Atkins, P. W.: *Physical Chemistry*. Oxford: Oxford University Press 1998.

[2] Axelrod, R.: *The Evolution of Cooperation*. New York: Basic Books 1984.

[3] Banzhaf, W., Dittrich, P., Eller, B.: Topological Interactions in a Binary String System. *Physica D* **125** (1999) 85-104.

[4] Behe, M. J. : *Darwin's black box*. New York: Simon & Schuster 1996 (existuje český preklad *Darwinova černá skříňka*, Praha: Návrat domů 2001).

[5] Callebout, W., Raskin-Gutman, D. (eds.): *Modularity: Understanding the development and evolution of complex natural systems.* Cambridge, MA.: MIT Press 2002.

[6] Chomsky, N.: A*spects of the Theory of Syntax*. Cambridge, MA: MIT Press 1965.

[7] Darwin, C.: *Pôvod druhov*. Bratislava: Kalligram, 2006 (preložil Cedzo, R.).

[8] Dawkins, R. : Universal Darwinism. In Bendall D. S. (ed.) *Evolution from Molecules to Men*. Cambridge Univesity Press, London, 1982, pp. 403 -425.

[9] Dawkins, R.: *The Blind Watchmaker*. London and New York: Penguin Books 1988 (existuje český preklad: *Slepý hodinář*. Praha: Paseka 2002).

[10] Dawkins, R.: *The Selfish Gene*. Oxford, UK: Oxford University Press 1976 (existuje český preklad: *Sobecký gen*. Praha: Mladá Fronta 2003).

[11] Dennett, D. C. : *Darwin's Dangerous Idea - Evolution and the Meaning of Life.* London: Penguin Press 1995.

[12] Dittrich, P.: *Artificial Chemistries* (tutorial material). A tutorial held at *ECAL'99*, 13-17 September 1999, Lausanne, Switzerland.

[13] Eigen, M. : Self organization of matter and the evolution of biological macro molecules. *Naturwissenshaften* **58** (1971) 465-523.

[14] Eigen, M., Schuster, P.: The Hypercycles: A Principle of Natural Evolution. *Naturwissenschaften* **64** (1977) 541-565; **65** (1978) 7-41; **65** (1978) 341-369.

[15] Fogel, D. B.: *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. New York: The IEEE Press 1995.

[16] Fontana, W. and Schuster, P.: Continuity in Evolution: On the Nature of Transitions. *Science* **280** (1998) 1451-1455.

[17] Fontana, W. and Schuster P.: A computer model of evolutionary optimization. *Biophysical Chemistry* **26** (1987) 123-147.

[18] Fontana, W., Schnabl, W., Schuster, P.: Physical aspects of evolutionary optimization and adaptation. *Physical Review A* **40** (1989) 3301-3321.

[19] Fontana, W.: Algorithmic Chemistry. In: Langton C.G. (ed.) *Artificial Life II* Reading, MA: Addison Wesley 1992, pp. 159- 210.

[20] Fontana, W.: Pulling Strings. *Science* **314** (2006) 1552-1553.

[21] Gálik, D.: Univezálny darvinizmus (Môže byť darvinizmus teóriou všetkého_), Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na internetovskej adrese http://math.chtf.stuba.sk/ CogSci_2000.html.

[22] Gálik, D.: Univerzálna teória evolúcie a problém mémov. In Nosek, J. (ed.), *Memy ve vědě a filosofii*. Praha: Filosofia 2004, pp. 133-146.

[23] Gillepsie, D. T.: Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **81** (1977) 2340- 2361.

[24] Hardin, G.: The Tragedy of the Commons. *Science* **162** (1968) 1243-1248.

[25] Holland, J. H.: *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press 1975.

[26] Jacob, F.: *The Possible & The Actual*. New York: Pantheon Books 1982 (existuje český preklad: *Hra s možnostmi*. Praha: Karolinum 1999)

[27] Kauffman, S. A. : *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press 1993.

[28] Kauffman, S.A., Hohnsen, S.: Coevolution to the edge of chaos: Coupled fitnes landscape, poised states, and coevolutionary avalanches. *J. Theor. Biology* **149** (1991) 467-506.

[29] Kvasnička, V.: An evolutionary model of symbiosis. In: P. Sinčák, J.Vasčák (eds.): *Quo Vadis Computational Intelligence*? Physica-Verlag, Heidelberg, 2000, p. 293-304.

[30] Kvasnička, V.: An evolutionary simulation of modularity emergence of genotype-phenotype mappings. *Neural Network World* **5** (2001) 473- 491.

[31] Kvasnička, V.: A modularity emergence of genotype-phenotype mappings. *Artificial Life* **8** (2002) 295-310.

[32] Kvasnička, V., Pospíchal, J.: A study of autoreplicators and hypercycles by typogenetics. In: Kelemen, J., Sosik, P.: *Advances in Artificial Life*, ECAL 2001, LNAI 2159, Springer 2001, pp. 37-54.

[33] Kvasnička, V., Pospíchal, J.: Autoreplicators and Hypercycles in Typogenetics, *Journal of Molecular Structure (Theochem)* **547** (2001) 119- 138.

[34] Kvasnička, V.: Artificial Chemistry, Replicators, and Molecular Darwinian Evolution *In Silico.* In: P. Sinčák, J.Vasčák (eds.): *Machine Intelligence, Quo Vadis?* Heidelberg: Physica-Verlag 2002, pp. 403-425.

[35] Kvasnička, V., Pospíchal, J.: Artificial Chemistry and Molecular Darwinian Evolution in silico. *Collection of Czechoslovak Chemical Communication*. **68** (2003) 139-177.

[36] Kvasnička, V., Pospíchal, J.: Artificial Chemistry and Replicator Theory of Coevolution of Genes and Memes. *Collection of Czechoslovak Chemical Communication*. **72** (2007) 223-251.

[37] Kvasnička, V., Pospíchal, J., Tiňo, J.: *Evolučné algoritmy*. Bratislava: Vydavateľstvo STU 2000.

[38] Kvasnička, V., Pospíchal, J.: A Multi-agent Study of Interethnic Cooperation. In *Multi-Agent Systems and Applications*, LNCS, Vol. 2086. Berlin: Springer Verlag 2001, pp. 415-435.

[39] Maynard Smith, J.: *Evolution and the Theory of Games*. Oxford, UK: Oxford University Press 1982.

[40] Mayr, E.: *Towards a New Philosophy of Biology.* Cambridge, MA: Harvard University Press 1988.

[41] Medveď, M.: *Dynamické systémy*. Bratislava: Veda 1988.

[42] Monod, J.: *Chance and Necessity. An Essay on the Natural Philosophy of Modern Biology.* New York: Vintage 2000 (existuje český preklad: *Náhoda a nutnosť*, Praha: vydav. Pavel Mervart 2008)

[43] Newman, M. E. J., Engelhardt, R.: Effects of neutral selection on the evolution of molecular species. *Proc. Roy. Soc. London B* **265** (1998) 1333- 1338.

[44] Paley, W.: *Natural Theology*, with an introduction and notes by Eddy, M. D. and Knight, D. M., London and Oxford: Oxford University Press 2006.

[45] Schuster, P.: Molecular Insight into Evolution of Phenotypes. In Crutchfield, J.P. and Schuster, P. (eds.): *Evolutionary Dynamics - Exploring the Interplay of Accident, Selection, Neutrality, and Function*. Oxford: Oxford University Press 2002.

[46] Schuster, P., Fontana, W.: Chance and Necessity in Evolution: Lessons from RNA. *Physica D* **133** (1999) 427-452.

[47] Spiegelman, S.: An Approach to the Experimental Analysis of Precellular Evolution. *Quart. Rev. Biophysics* **4** (1971) 213-253.

[48] Tomáška, Ľ., Nosek, J.: Molekulárna darwinovská evolúcia: In vivo, in vitro a in silico. Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na internetovskej adrese http://math.chtf.stuba.sk/ CogSci_2000.htm).

[49] Watson, A. R.: *Compositional Evolution. The Impact of Sex, Symbiosis, and Modularity on the Gradualist Framework of Evolution*. Cambridge, MA: MIT Press 2006.

[50] Waterman, M. S.: *Introduction to Computational Biology. Maps, sequences and genomes*. New York: Chapman & Hall / CRC, 1995.

[51] Wright, S.: The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution. *Proceedings of The Six International Congress of Genetics,* Vol. I, 1932, pp. 356-366. (Táto klasická publikácia je dostupná na http://www. blackwellpublishing.com/ridley/classictexts/wright.pdf).

4. prednáška

Úloha učenia v evolúcii a memetika

4.1 Baldwinov efekt a úloha učenia v evolúcii

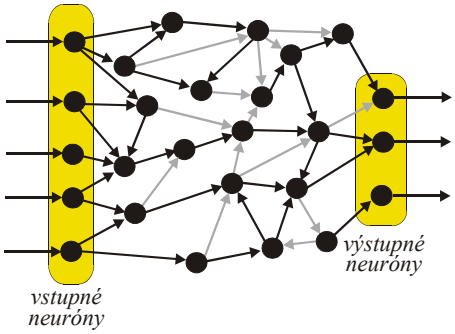
V r.1896 americký psychológ J. M. Baldwin formuloval hypotézu podľa ktorej, jedinci z populácie ***schopní učiť sa majú väčšiu šancu prežitia***, ich fitnes je väčšie ako fitnes jedincov, ktorí túto vlastnosť nemajú. Ak prostredie evolúcie zostáva dlhú dobu nemenné, potom prirodzený výber môže viesť k ***emergencii*** takých jedincov populácie, ktorí majú túto ***schopnosť učiť sa geneticky zafixovanú***, Baldwin nazval tento mechanizmus „***organická selekcia***". V súčasnosti sa na počesť Baldwina tento jav nazýva ***Baldwinov efekt***.

Môžeme teda konštatovať, že pri evolúcii kognitívneho aparátu vystupuje do popredia ***aspekt učenia***, t.j. schopnosť jedincov adaptovať si svoj kognitívny orgán tak, aby odrážal v určitom rozsahu zmeny v prostredí alebo nové kognitívne funkcie dôležité pre prežitie a reprodukciu. Význam učenia v darwinovskej evolúcii spočíva v tom, že organizmus s lepším kognitívnym aparátom má väčšiu šancu na prežitie a reprodukciu. Poznamenajme, že základná architektúra a snáď aj počiatočné nastavenie vybraných váhových koeficientov kognitívneho orgánu sa dedične prenášajú na potomkov, avšak neprenášajú sa vedomosti získané rodičmi v priebehu ich života.

Obrázok 4.1. Americký psychológ J. M. Baldwin (1861-1934), ktorý v r. 1896 publikoval prácu „A New Factor in Evolution" v časopise *The American Naturalist*, v ktorej formuloval úlohu učenia v Darwinovej evolúcii, čo môže vo vybraných prípadoch viesť až ku genetickej fixácii vlastnosti, ktorá bola predmetom učenia.

4.1.1 GA s učením

Moderný pohľad na Baldwinov efekt pochádza od ***Hintona a Nowlana*** [xx], ktorí pomocou obyčajného genetického algoritmu modelovali tento efekt, pričom chápali učenie ako prehľadávanie v určitom okolí aktuálneho stavového bodu parametrov kognitívneho orgánu.



Obrázok 4.2. Znázornenie jednoduchej neurónovej siete, hrany znázornené čiernou čiarou majú už fixne nastavený váhový koeficient na binárnu hodnotu 0 alebo 1, ostatne hrany znázornené svetlou čiarou sú len hypotetické spoje, ktoré potenciálne môžu existovať a sú reprezentované symbolom #.

Budeme predpokladať, že chromozóm reprezentuje váhové koeficienty neurónovej sieti znázornenej na obrázku 4.2. Niektoré váhové koeficienty *w* sú už nastavené geneticky (čo je reprezentované binárnymi hodnotami 0 a 1 v genotype agenta), iné parametre siete nie sú nastavené (čo je reprezentované symbolmi # v genotype agenta), tak napríklad chromozóm dĺžky 16 je lineárny reťazec obsahujúci symboly 0, 1, a #,

$$\boldsymbol{x} = (1100\#\#\#\#01\#\#0011).$$

V priebehu evolúcie agenta sa genotyp agenta (chromozóm) mení pomocou mutácii tak, že sú mutované tak binárne hodnoty genotypu ako aj symboly #. Pod ***učením agenta*** rozumieme proces, kde agent metódou „hill climbing“ (t. j. náhodným lokálnym prehľadávaním) skúma, ktoré symboly # môžu byť nahradené nejakou binárnou hodnotou 0 alebo 1. Ukazuje sa, že väčšina symbolov # je nahradená binárnymi hodnotami, čo môžeme interpretovať tak, že kognitívny orgán agenta sa stáva vhodnejším pre realizáciu požadovaných kognitívnych aktivít. Tento proces prebieha aj bez učenia, ale je podstatne pomalší.

Hlavný výsledok týchto numerických experimentov je ten, že ak do procesu evolúcie zapojíme učenie, potom ***emergencia*** chromozómov v populácii s malou frekvenciou výskytu allel # je podstatne rýchlejšia ako keď učenie nie je zapojené. Výsledkom evolúcie je agent, ktorý má vo svojom kognitívnom orgáne určité poznatky buď

(1) už ***priamo zabudované*** v jeho architektúre (inštinkty) alebo

(2) jeho kognitívny orgán má ***určité predispozície*** vykonávať špeciálne kognitívne aktivity (napr. schopnosť reči).

4.1.2 Plasticita genotypu

Pristúpime k formulácii koncepcie efektívneho fitnes, ktorý hrá dôležitú úlohu pri formulácii Baldwinovho efektu pomocou genetického algoritmu. ***Plasticita fenotypu*** je vyjadrená skutočnosťou, že fitnes fenotypu $f(\boldsymbol{x})$ môže byť prehľadávaný v okolí $U(\boldsymbol{x})$ metódou „hill climbing“. V okolí $U(\boldsymbol{x})$ hľadáme „lepšie riešenie“ $\boldsymbol{x}' \in U(\boldsymbol{x})$, pre ktoré platí $f(\boldsymbol{x}') > f(\boldsymbol{x})$. Použitie koncepcie ***efektívneho fitnes*** spočíva v nasledujúcej možnosti vyjadriť fitnes chromozómu $\boldsymbol{x}$: Majme dva chromozómy $\boldsymbol{x}_1$ a $\boldsymbol{x}_2$ s rovnakým fitnes, $f(\boldsymbol{x}') = f(\boldsymbol{x})$, pričom v blízkom okolí týchto chromozómov existujú riešenia – chromozómy

$$\boldsymbol{x}_{1,opt} = arg \min_{\boldsymbol{x} \in U(\boldsymbol{x}_1)} f(\boldsymbol{x}), \quad \boldsymbol{x}_{2,opt} = arg \min_{\boldsymbol{x} \in U(\boldsymbol{x}_2)} f(\boldsymbol{x}) \tag{4.1}$$

Ak tieto chromozómy vyhovujú podmienke

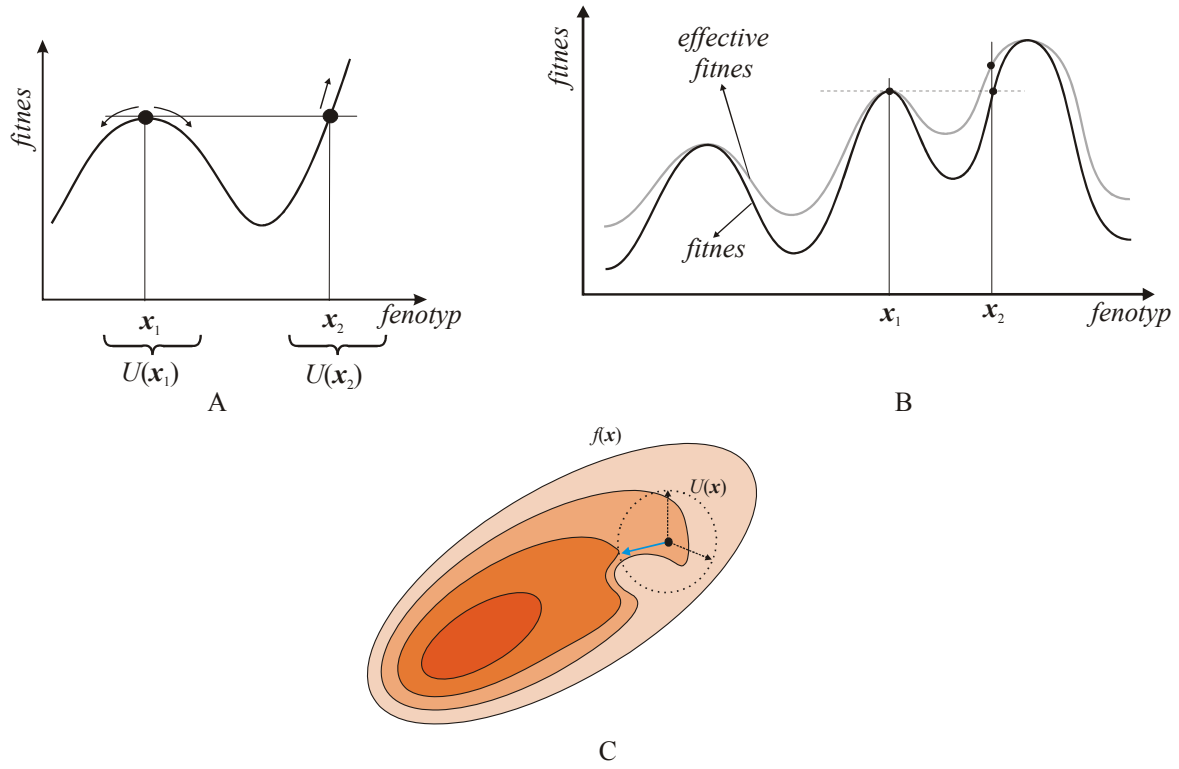
$$f(\boldsymbol{x}_{1,opt}) > f(\boldsymbol{x}_{2,opt}) \tag{4.2}$$

lebo pomocou ***efektívnych fitnes*** $F(\boldsymbol{x}_1) = f(\boldsymbol{x}_{1,opt})$ a $F(\boldsymbol{x}_2) = f(\boldsymbol{x}_{2,opt})$ túto vlastnosť vyjadrím ako

$$F(\boldsymbol{x}) > f(\boldsymbol{x}) \tag{4.3}$$

Môžeme teda konštatovať, že ***efektívne fitnes*** odráža vlastnosti fenotypu v jeho blízkom na povrchu funkcie fitnes $f(\boldsymbol{x})$

1. $f(x_1) = f(x_2) \Rightarrow F(x_1) = F(x_2)$, potom fenotypy $\boldsymbol{x}_1$ a $\boldsymbol{x}_2$ sú ***efektívne rovnocenné***,
2. $f(x_1) = f(x_2) \Rightarrow F(x_1) > F(x_2)$, potom fenotyp $\boldsymbol{x}_1$ je ***efektívne lepší*** ako fenotyp $\boldsymbol{x}_2$.



Obrázok 4.3. Grafická interpretácia koncepcie efektívneho fitnes. (A) Tento diagram znázorňuje situáciu, keď dva chromozómy $\boldsymbol{x}_1$ a $\boldsymbol{x}_2$ majú rovnaké fitnes, avšak v okolí chromozómu $\boldsymbol{x}_2$ existujú „stavy" s vyšším fitnes ako je $f(\boldsymbol{x}_2)$. To znamená, že pre evolučný proces je tento chromozóm podstatne dôležitejší ako chromozóm $\boldsymbol{x}_1$, v ktorého blízkosti neexistujú chromozómy s vyšším fitnes. Túto skutočnosť môžeme formulovať pomocou efektívneho fitnes tak, že efektívne fitnes v $\boldsymbol{x}_1$ je menšie ako efektívne fitnes v $\boldsymbol{x}_2$. Ak je selekcia riadená efektívnym fitnes, potom druhý chromozóm je v prirodzenom výbere preferovaný pred chromozómom $\boldsymbol{x}_1$. (B) Znázornenie priebehu efektívneho fitnes $F(\boldsymbol{x})$ vzhľadom k fitnes $f(\boldsymbol{x})$, ako vyplýva z obrázku, vždy platí $F(\boldsymbol{x}) \geq f(\boldsymbol{x})$. (C) Alternatívne znázornenie významu efektívneho fitnes na jeho povrchu. V štandardnom genetickom algoritme generovanie nového chromozómu $\boldsymbol{x}'$ z pôvodného chromozómu je náhodné, avšak ak použijeme efektívnu fitnes, potom určitý smer je nádejný a ten je zvolený ako následovník pôvodného $\boldsymbol{x}$.

Môžeme teda konštatovať, že koncepcia efektívneho fitnes nám slúži k zefektívneniu genetického algoritmu: (1) Hovoríme, že ***učenie*** nám pomáha nájsť lokálne nádejný smer (evolvabilitu) pre nasledujúcu evolúciu, (2) V štandardnom prístupe Darwinovej evolúcie evolvabilita sa ponecháva výlučne na prírodný výber, a (3) Zapojenie Baldwinovho efektu do evolúcie môže slúžiť ako významný akceleračný faktor Darwinovej evolúcie. Samozrejme, tieto tri dôležité aspekty zapojenia učenia do genetického algoritmu sú efektívny akceleračným faktorom len vtedy, ak „učenie" je možné efektívne realizovať ako lokálne prehľadávanie metódou „hill climbing". Ako jednoduchý kontrapríklad nevhodnosti učenia

pre zefektívnenie genetického algoritmu je možný prípad, keď aktivovanie funkcie $f(\boldsymbol{x})$ (t. j. výpočet fitnes chromozómu $\boldsymbol{x}$) je časovo veľmi náročný proces.

4.1. 2 Formulácia GA s učením

Zopakujeme si základné princípy genetického algoritmu s kapitoly 1.2. Nech chromozóm $\boldsymbol{x} = (x_1, x_2, ..., x_n) \in \{0,1,\#\}^n$ je lineárny reťazec dĺžky n, ktorý obsahuje symboly 0, 1 a #. Nech si optimálny (cieľ evolúcie GA) chromozóm $\boldsymbol{x}_{\text{opt}}$ je náhodne generovaný binárny vektor dľžky n

$$\boldsymbol{x}_{opt} = \left(x_1^{opt} x_2^{opt} ... x_n^{opt}\right) \in \{0,1\}^n \tag{4.4}$$

Populácia P je multimnožina obsahujúca p chromozómov

$$P = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_p\} \subseteq \{0,1,\#\}^n \tag{4.5}$$

Každý chromozóm $\boldsymbol{x} \in P$ je ohodnotený ***fitnes***. Tento ohodnocovací proces úzko súvisí s procesom učenia a je určený takto: Nech okolie chromozómu $\boldsymbol{x} \in P$ je

$$U(\boldsymbol{x}) = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_q\} \subseteq \{0,1\}^n \tag{4.6}$$

Toto okolie obsahuje q binárnych vektorov, ktoré sú náhodne vytvorené z chromozómu $\boldsymbol{x}$ tak, že jeho „hash“ symboly sú náhodne nahradené 0 alebo 1

$$\boldsymbol{y} = (y_1 y_2 ... y_n) \in \{0,1\}^n \quad \text{a} \quad y_i = \begin{cases} x_i & (\text{if } x_i = 0 \text{ or } x_i = 1) \\ \text{random}(2) & (\text{if } x_i = \#) \end{cases} \tag{4.7}$$

Kardinalita $q=|U(\boldsymbol{x})|$ sa nazýva *rozsah učenia*. Funkcia $d(\boldsymbol{x})$, pre $\boldsymbol{x} \in P$, je špecifikovaná ako minimálna vzdialenosť medzi optimálnym chromozómon $\boldsymbol{x}_{\text{opt}}$ a binárnym vektorom z okolia $U(\boldsymbol{x})$

$$d(\boldsymbol{x}) = \min_{\boldsymbol{y} \in U(\boldsymbol{x})} \left|\boldsymbol{y} - \boldsymbol{x}_{opt}\right| \tag{4.8}$$

Efektívne fitnes chromozómu $\boldsymbol{x}$ je určené formulou

$$F(\boldsymbol{x}) = e^{-\xi d(\boldsymbol{x})} \tag{4.9}$$

kde ξ je tzv. koeficient strmosti, ktorý špecifikuje „ostrosť“ fitnes funkcie. Proces učenia je špecifikovaný dvoma parametrami: (1) kardinalitou q okolia $U(\boldsymbol{x})$ a (2) parameter strmosti ξ. Pre asymptotické hodnoty $F(\boldsymbol{x})$ platí

$$\lim_{\xi \to \infty} F(\boldsymbol{x}) = \begin{cases} 1 & (\text{if } \boldsymbol{x} = \boldsymbol{x}_{opt}) \\ 0 & (\text{if } \boldsymbol{x} \neq \boldsymbol{x}_{opt}) \end{cases} \quad \text{a} \quad \lim_{\xi \to 0} F(\boldsymbol{x}) = f_{max} \tag{4.10}$$

Obyčajné fitnes $f(\boldsymbol{x})$ má tento jednoduchý priebeh

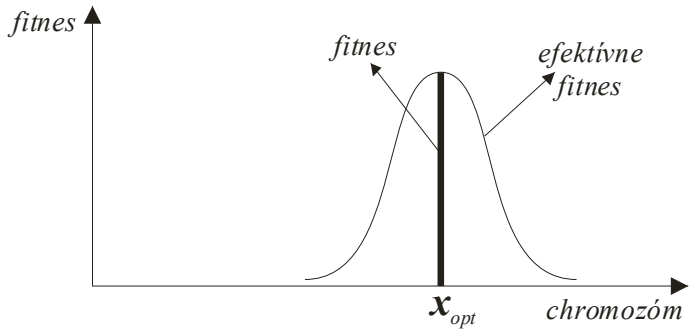
$$f(\boldsymbol{x}) = \begin{cases} 1 & (pre\ \boldsymbol{x} = \boldsymbol{x}_{opt}) \\ 0 & (pre\ \boldsymbol{x} \neq \boldsymbol{x}_{opt}) \end{cases} \tag{4.11}$$

Potom platí pre každý chromozóm $\boldsymbol{x}$ podmienka (pozri obr. 4.4)

$$f(\boldsymbol{x}) \geq F(\boldsymbol{x}) \tag{4.12}$$

Podľa Hintona a Nowlana [xx] študovaný problém môže byť charakterizovaný ako hľadanie ihly v kope sena (ang. *needle in a haystack*) dôsledku skutočnosti, že existuje len jedno korektné riešenie v priestore 2^n riešení. Fitness je obvykle významnou heuristikou, ktorá nám uľahčuje toto hľadanie optimálneho riešenia. Obyčajné fitnes $f(\boldsymbol{x})$ nie je nám však nápomocné v tomto prehľadávaní, z jeho definície (4.11) vypláva, že je jednotkové v prípade, že $\boldsymbol{x} = \boldsymbol{x}_{opt}$, v opačnom prípade je nulové (pozri obr. 4.4). To znamená, že k tomu, aby sme

našli správne riešenie (t. j. ihlu v kope sena), musíme prekontrolovať všetky prípustné riešenia, čo je vlastne ten najhorší možný algoritmus riešenia daného problému. Avšak, ak používame efektívne fitnes, ktoré je nenulové aj v blízkom okolí hľadaného riešenia, máme podstatne väčšiu šancu nájsť vsprávne riešenie. Môžeme povedaž, že efektívne fitnes nás proces prehľadávania naviguje smerom, kde sú väčšie hodnoty efektívneho fitnes.



Obrázok 4.4. Priebeh fitnes f(x) a efektívneho fitnes $F(\boldsymbol{x})$. Poznamenajme, že fitnes $f(\boldsymbol{x})$ je nenulové len pre jeden konkrétny binárny chromozóm $\boldsymbol{x}_{opt}$, pre všetky ostatné chromozómy je nulové. Efektívne fitnes je nenulové aj pre chromozómy z blízkeho okolia optimálneho chromozómu $\boldsymbol{x}_{opt}$.

Podľa Hintona a Nowlana študovaný problém môže byť charakterizovaný ako hľadanie ihly v kope sena; pretože existuje len jedno korektné riešenie v priestore 2^n riešení. Ak je do GA zahrnuté aj učenie, potom GA môže byť interpretovaný ako metóda prehľadávania, kde „niekto“ radí chromozómu v ktorom „smere je ihla“. Ak do GA nie je zahrnuté učenie, GA môže byť interpretované zhruba ako metóda prehľadávania, kde nik neradí "chromozómu", ktorým „smerom je ihla“. Naznačenú teóriu GA učením sme realizovali pre tieto parametre:

(1) n, dĺžka chromozómu (n=20),
(2) p, veľkost populácie (p=500),
(3) P_{cross}, pravdepodobnosť kríženia (P_{cross}=0.1),
(4) P_{mut}, pravdepodobnosť mutácie (P_{mut}=0.01),
(5) t_{max}, počet epoch (t_{max}=200),
(6) q, veľkosť učenia (q=10),
(7) ξ, parameter strmosti (ξ=0.1 - 1.0),
(8) maximálna hodnota fitnes f_{max}=2.0,
(9) minimálna hodnota fitnes f_{min}=1.0,
(10) prvotná populácia je náhodne generovaná tak, že jej chromozómy obsahujú 50% symbolov #, 25% symbolov 0, and 25% symbolov 1.

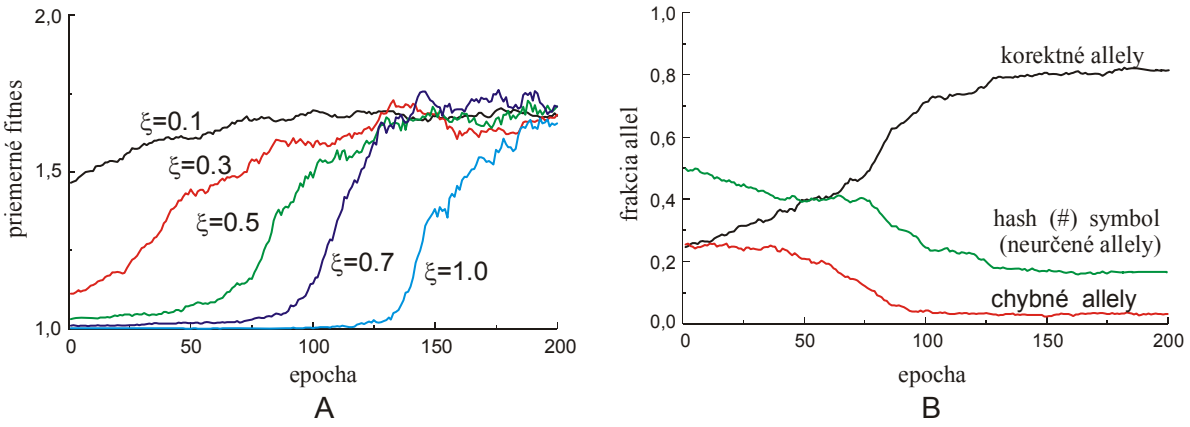
Numerické výpočty realizované pomocou týchto parametrov sú znázornené na obrázku 4.4. Z týchto výsledkov jednoznačne vyplýva, že zapojenie učenia do GA podstatne akceleruje túto metódu. Naviac, jednoduché kognitívne úlohy môžu byť pomocou Bladwinovho efektu kanalizované (geneticky zakódované - vznik inštinktov sociálneho hmyzu) do genotypu organizmu. Zložité kognitívne úlohy nemôžu byť však pomocou Baldwinovho efektu kanalizované, pretože dĺžka života nie je postačujúca k prejaveniu sa tohto efektu.

Pomocou metafory učenia môžeme genetické algoritmy rozdeliť do troch skupín:

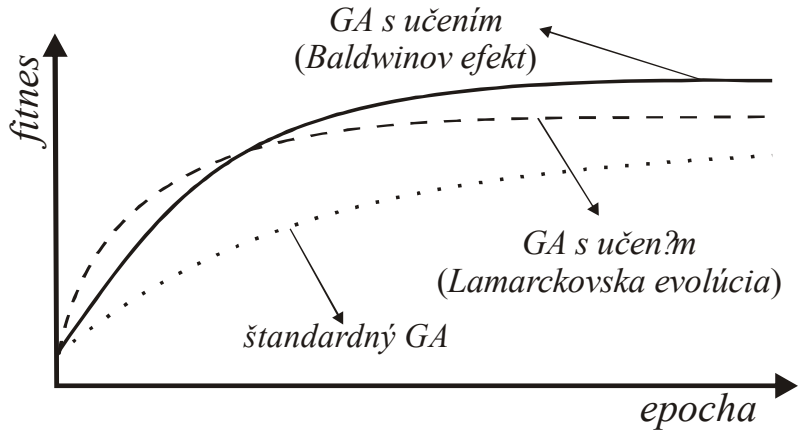
(1) ***Štandardné genetické algoritmy***, kde selekcia chromozómov do reprodukčného procesu je vyberaná na základe ich obyčajného fitnes, ktoré je určené ako zobrazenie chromozómov (binárnych reťazcov fixnej dĺžky) na kladné reálne číslo.
(2) ***Modifikované genetické algoritmy s učením***, kde selekcia chromozómov do reprodukcie je riadená pomocou efektívnej fitnes (t. j. GA s Baldwinovym efektom),

avšak kde vstupné chromozómy do reprodukčného procesu sa nezamieňajú za optimálne chromozómy z ich blízkeho okolia, t. j. genetické algoritmy si zachovávajú svoj darwinovský charakter.

(3) Genetické algoritmy s procesom učenia, kde pri reprodukčnom procese vybrané chromozómy sú už lokálne optimalizované. Takto modifikovaný genetický algoritmus už nemá charakter darwinovskej evolúcie, ale má charakter tzv. ***lamarckovskej evolúcie***.



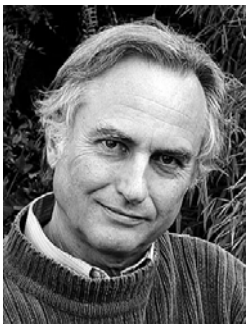
Obrázok 4.4. (A) Päť rôznych priebehov fitness pre rôzne hodnoty parametrov strmosti, od podporovaného učenia (ξ=0.1) až po potlačené učenie (ξ=1.0). (B) Priebehy frakcií korektných allel, nekorektných allel a „hash" allel (#) pre GA s parametrom strmosti ξ=0.5.



Obrázok 4.5. Znázornenie rôznych typov evolúcie v rámci genetických algoritmov. (A) Priebeh fitnes štandardného GA (bodkovaná čiara ·····) ukazuje, že priemerné fitnes rastie pomaly ale monotónne do maximálnej hodnity. (B) Priebeh fitnes GA s učením (spojitá čiara ————) ukazuje, že algoritmus je podstatne rýchlejší ako štandardný GA. (C) Priebeh fitnes GA s lamrckovskou evolúciou (prerušovaná čiara – – – – –) rastie najrýchlejšie, avšak algoritmus obvykle skončí v lokálnom minime.

Znázornenie evolúcie vo všetkých týchto troch prípadoch je ukázané na obrázku 4.5. Ak rozšírime štandardný genetický algoritmus o učenie (t. j. zahrnieme Baldwinov efekt) vznikne nová forma genetického algoritmu, ktorá je akcelerovaná pred pôvodnou štandardnou verziou a naviac nemá tendenciu konvergovať do blízkeho lokálneho minima. Na obrázku je znázornená aj tretia verzia genetického algoritmu, kde pri prehľadávaní lokálneho minima v rámci okolia vybraného rodičovského chromozómu dochádza k jeho zámene za lokálne optimálny chromozóm. Toto rozšírenie genetického algoritmu je mimo rámca darwinovskej

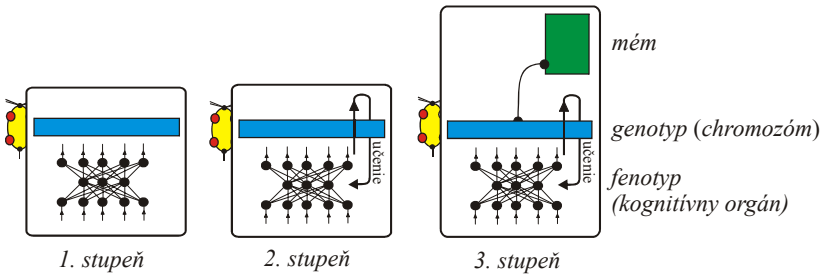
evolúcie a môže byť považované za lamarckovskú verziu evolúcie genetického algoritmu, kde chromozómy dedia získané vlastnosti v priebehu ich existencie (táto metafora lamarckovskej evolúcie je často vyjadrená diktumom „svaly kováča dedia jeho deti“. Z obr. 4.5 vidieť, že táto lamarcovská verzia genetického algoritmu konverguje najrýchlešie z oboch predchádzajúcich cerzií, ale žial veľmi často do lokálneho extrému.



Obrázok 4.6. Anglický evolučný biológ Richad Dawkins (1941-), ktorý v r. 1978 publikoval slávnu knihu „The Selfish Gene“ v ktorej formuloval mémemtiku - vedu o mémoch, ktoré sú reprezentované informačnými jednotkami kopírujúce sa z jedného mozgu na druhý mozog pomocou imitácie. Taktiež postulov, že mémy sú replikátory, ktoré môže byť predmetom evolúcie. Memetika poskytuje konceptuálne jednoduché vysvetlenie o podstate a evolúcii ľudskej kultúry, paradigma mémov ako replikátorov je veľmi zaujímavým prístupom pre komunitu umelej inteligencie a kognitívnej vedy.

4.2. Dawkinsova memetika

Môžeme sa zamýšľať nad tým, ako ďalej akcelerovať darwinovskú evolúciu. V predchádzajúcej časti tejto kapitoly už sme študovali možnosť akcelerácie pomocou Baldwinovho efektu, kde štandardná fitnes je v procese selekcie nahradená efektívnou fitnes. Ďalšia možnosť akcelerácie darwinovskej evolúcie existuje pomocou Dawkinsových mémov, ktoré obsahujú čiatočnú informáciu o hľadanom optimálnom riešení. Z tohot pohľadu môžeme teda rozlišovať tri stupne Darwinovej evolúcie (pozri obrázok 4.6):



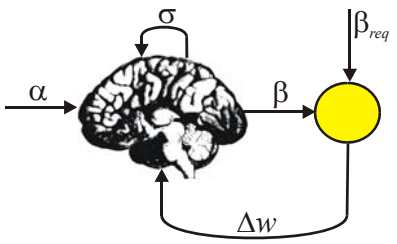
Obrázok 4.6. Schématické znázornenie troch stupňov darwinovskej evolúcie.pomocou hypotetického agenta s kognitívnym orgánom reprezentovaným neurónovou sieťou. Táto sieť je špecifikovaná fenotypom – chromozómom. V 1. stupni neexistuje spätná interakcia (učenie) medzi chromozómom a neurónovou sieťou, chromozóm „na tvrdo“ špecifikuje neurónovú sieť, zmeny v architektúre tejto siete sú možné len pomocou „neriadenej“ mutácie. V 2. stupni existuje už proces učenia, čiže môže dochádzať k zmenám neurónovej siete pomocou „riadených mutácií. V 3. stupni darwinovskej evolúcie okrem učenia je do procesu výpočtu efektívneho fitnes zapojený aj mém, ktorý reprezentuje „informáciu“ o optimálnom riešení.

1. **stupeň.** Fitnes genotypu agenta je určený ***výlučne len jeho zložením***, žiadne iné efekty (učenie sa, mémy, ...) nie sú zahrnuté pri ohodnotení genotypu fitnesom, tento prvý stupeň je reprezentovaný štandardným genetickým algoritmom..

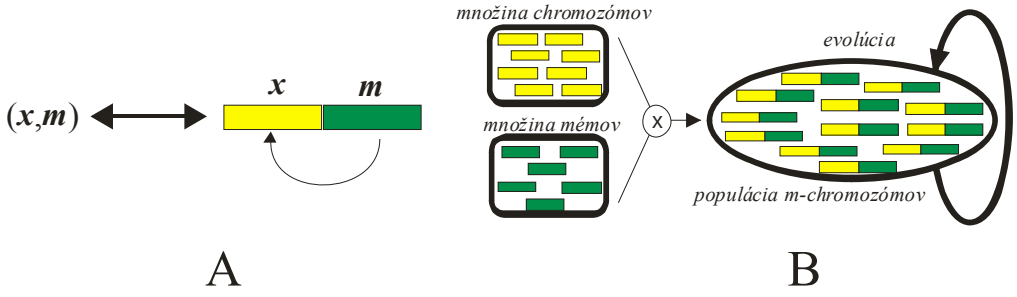
2. **stupeň.** Fitnes genotypu je určené nielen jeho štruktúrou, ale aj ***schopnosťou organizmu učiť sa*** (Baldwinov efekt). Elementárne kognitívne aktivity môžu byť zabudované (kanalizované) do genotypu organizmu (napr. pri vzniku sociálneho hmyzu), tento stupeň je reprezentovaný genetickým algoritmom, ktorý v procese selekcie využíva efentívne fitnes.

3. **stupeň.** Fitnes genotypu je určené nielen jeho štruktúrou a schopnosťou organizmu učiť sa, ale aj určitým ***„informačným balíkom" (mémom)***, ktorý podstatne ovplyvňuje fitnes.

Ako môžu ovplyvnovať mémy kognitívny orgán (neurónovú sieť) agenta.Pre jednoduchosť predpokladajme, že mém – informácia je poznatok zakódovaný vo forme implikácie $\alpha \Rightarrow \beta_{req}$, potom neurónovú sieť môžeme adaptovať tak, že na vstup α reaguje výstupom β, ktorý je skoro totožný s požadovaným výstupom β_{req}, pozri obrázok 4.7.



Obrázok 4.7. Učenie kognitívneho organu (mozgu) pomocou mému štruktúry $\alpha \Rightarrow \beta_{req}$. Premisa mému α je vstupom do kognitívneho orgánu, a nech jeho odozva ne tento vstup je β. Požadovaný výstup β_{req} je porovnaný s vypočítaným výstupom β (úloha učiteľa / tútora), toto porovnanie vyústi v konštrukciu Δw, pomocou ktorého sa upraví kognitívny orgán tak, aby sa zmenšil rozdiel medzi vypočítaným β a požadovaným β_{req}. Vnútorná (skrytá) premenná σ reprezentuje vnútorný stav mozgu, ktorý ovplyvňuje odozvu mozgu na daný vstup, $\beta=F(\alpha,\sigma;w)$ (odozva mozgu na dva rovnaké vstupy $\alpha_1=\alpha_2$, avšak pre rôzne vnútorné stavy $\sigma_1 \neq \sigma_2$, môže byť diametrálne odlišná, $\beta_1 \neq \beta_2$).

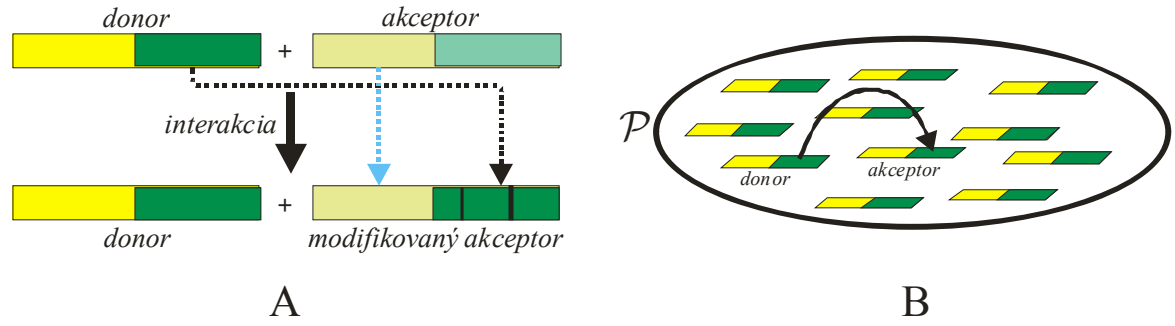


Obrázok 4.8. (A) Vznik m-chromozómu ako komplexu chromozómu *x* a mému ***m***. (B) Predmetom evolúcie v memetickom prístupe je populácia m-chromozómov.

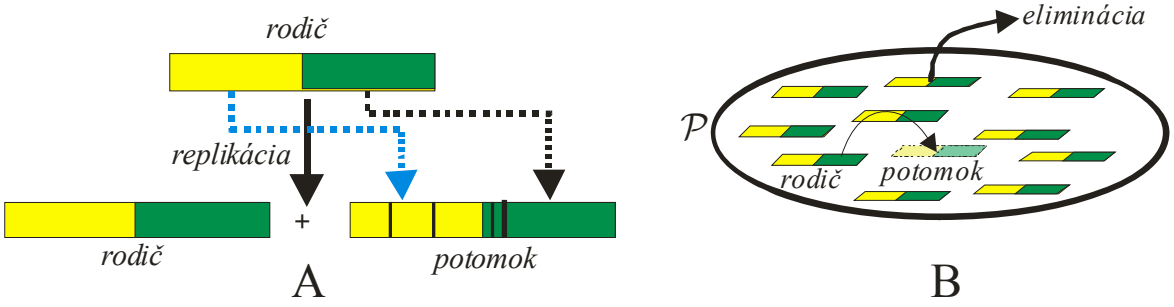
4.2.1 Formulácia GA s učením a memetickým prenosom informácie

Použijeme veľmi zjednodušený ***model mémov***, ktorý môže byť jednoducho zahrnutý do GA s učením. Mémy budeme chápať ako lineárne reťazce (rovnakej dĺžky ako gény), ktoré spolu s génmi vytvárajú komplex nazývaný ***m-chromozóm,*** $\alpha = (x,m)$ (pozri diagram A na obrázku 4.8.). Mémy sú nositeľom informácie o najlepšom lokálnom riešení, ktoré bolo doposiaľ nájdené v priebehu evolúcie GA. Využitie mémov je dôležité v procese ohodnotenia m-chromozómov ***fitness***, ako ďalší zdroj informácie o aktuálnom najlepšom riešení. Na začiatku evolučného procesu sú mémy prázdne (majú nulový informačný obsah). V priebehu evolučného procesu mémy sa stávajú nositeľom informácie, ktorá sa šíri v polulácii pomocou ***mémentickej interakcie***. Budeme rozlišovať dva typy memetickej interakcie:

(1) ***Horizontálna memetická interakcia***, kde kvázináhone vybraný (s vysokým fitness) m-chromozóm - donor s príslušným mémom vytvára správu – mém, ktorá je ***prenesená*** (mierne zmutovaná) do náhodne vybraného m-chromozóm – akceptoru, ktorý príjme túto správu a zahrnie ju do svojho m-chromozómu (pozri obrázok 4.9).
(2) ***Vertikálna memetická interakcia,*** v rámci reprodukčného procesu vytváraný potomkovia obsahujú mém (mierne zmutovaný) od jedného zo svojich rodičov; budeme používať model, kde "matka" určuje mém svojho potomka (hovoríme, že matka vychováva potomka) (pozri obrázok 4.10)



Obrázok 4.9. Horizontálna interakcia medzi dvoma m-chromozómami donorom a akceptporom, pričom donor je kvázináhodne vybraný s pravdepodobnosťou úmernou fitnes, jeho mém je v ľahko zmutovanej forme prenesený do m-chromozómu



Obrázok 4.10. Vertikálna interakcia medzi rodičovským m-chromozómom a potomkovským m-chromozómom, kde mém z rodiča v ľahko zmutovanej forme prechádza na potomka. V tomto prípade musíme z populácie odstrániť náhodne vybraný m-chromozóm, aby sme zachovali celkový počet m-chromozómov v populácii.

Pristúpime teraz k formulácii mémov a ich komplexov s chomozómami. ***Mém*** je reťazec n symbolov 0, 1 a $*$

$$\boldsymbol{m} = (m_1 m_2 ... m_n) \in \{0,1,*\}^n \tag{4.13}$$

Šírka mému $\boldsymbol{m}$, označená $|\boldsymbol{m}|$, je určená počtom binárnych (nehviezdičkových) symbolov, napr. mém $\boldsymbol{m}$=(01**01***) má šírku $|\boldsymbol{m}|$=4. Mém budeme používať ako pomocný nositeľ informácie získanej v rámci procesu učenia. Ak šírka mému je nulová, potom hovoríme, že mém neobsahuje vhodnú informáciu a jeho aplikácia v rámci evolučného procesu je bezpredmetná.

Mémetický chromozóm (*m-chromosome*) je určený ako usporiadaná dvojica chromozómu a mému o rovnakej dĺžke

$$\alpha = (\boldsymbol{x}, \boldsymbol{m}) = ((x_1 x_2 ... x_n), (m_1 m_2 ... m_n)) \in \{0,1,\#\}^n \times \{0,1,*\}^n \tag{4.14}$$

Kde $\boldsymbol{\alpha} = (\boldsymbol{x}, \boldsymbol{m}) = ((x_1 x_2 ... x_n), (m_1 m_2 ... m_n))$ je m-chromozóm s časťami $\boldsymbol{x}$ a $\boldsymbol{m}$. Populácia P je multimnožina obsahujúca p m-chromozómov

$$P = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ... \boldsymbol{\alpha}_p\} \subseteq \{0,1,\#\}^n \times \{0,1,*\}^n \tag{4.15}$$

Proces učenia v ktorom sú m-chromozómy ohodnotené fitnes obsahuje dve etapy:

- V ***prvej etape*** mémetická časť $\boldsymbol{m}$ m-chromozómu $\boldsymbol{\alpha}$ je použitá na opravu (modifikáciu) chromozómovej časti $\boldsymbol{x}$ Tento proces spočíva v tom, že „hash" symboly v $\boldsymbol{x}$ sú nahradené príslušnými binárnymi symbolmi (ak existujú) 0/1 z mému. Pre ilustráciu tejto etapy uvažujme m-chromozóm

$$\boldsymbol{\alpha} = \left(\underbrace{00\#\#1\#0\#}_{x} \right) \left(\underbrace{11***0**}_{m} \right) \tag{4.16}$$

  Opravený chromozóm $\hat{\boldsymbol{x}}$ vznikne z chromozómu $\boldsymbol{x}$ náhradením symbolov # binárnymi symbolmi z mému $\boldsymbol{m}$

$$\begin{array}{r} \boldsymbol{x} = (0\,0\,\#\,\#\,1\,\#\,0\,\#) \\ \boldsymbol{m} = (1\,1\,*\;*\,*\,0\,*\,*) \\ \hline \hat{\boldsymbol{x}} = (\,0\,0\,\#\,\#\,1\;0\,0\,\#) \end{array}$$

  Kde sme v šiestej polohe symbol # nahradili binárnym symbolom 0 z mému $\boldsymbol{m}$.

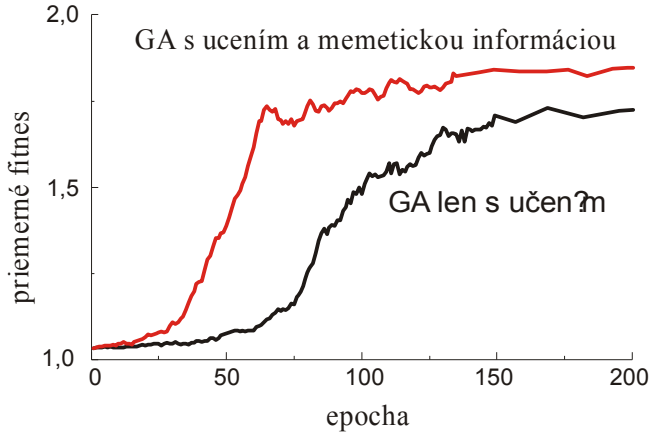
- V ***druhej etape***. pre modifikovaný chromozóm $\hat{\boldsymbol{x}}$ vytvoríme okolie $U(\hat{\boldsymbol{x}})$, pomocou tohto okolia spočítame minimálnu vzdialenosť od $\boldsymbol{x}_{opt}$

$$d_{min} = \min_{\boldsymbol{y} \in U(\hat{\boldsymbol{x}})} (\boldsymbol{y} - \boldsymbol{x}_{opt}) \tag{4.17}$$

  Potom efektívne fitnes priradené m-chromozómu $\boldsymbol{\alpha}$ je určené formulou

$$f(\boldsymbol{\alpha}) = e^{-\xi \hat{d}_{min}} \tag{4.18}$$

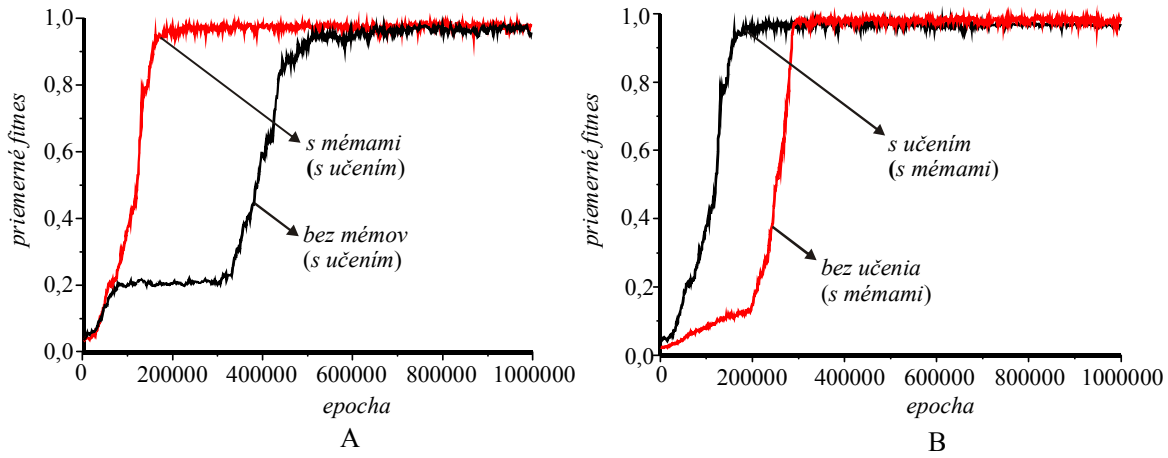
Týmto sme zavŕšili špecifikáciu genetického algoritmu, kde populácia $\boldsymbol{P}$ je tvorená m-chromozómami a proces selekcie m-chromozómov do reprodukčného procesu je založený na efektívnom fitnes (4.18).



Obrázok 4.11. Priebeh priemerného fitnes pre dve rôzne verzie GA s a bez memetickej informáciou, keď učenie (Baldwinov efekt) je zahrnutý v oboch verziách, pričom parameter strmosti učenia je ξ=0.5. Vidíme, že GA s učením a s memetickou informáciou podstatne rýchlejšie konverguje k maximálnej hodnote fitnes než ako GA le s učením bez zahrnutia memetickej informácie.

Na obrázku 4.11 sú znázornené výsledky našich simulačných výpočtov v dvoch verziách so zahrnutím učenia, ale jedna vezia neobsahuje memetickú informáciu, zatiaľ čo druhá verzia ju

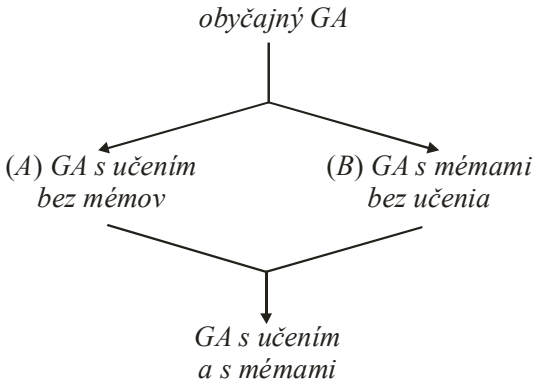
obsahuje . Vidíme, že zahrnutie mémov do GA s učením podstatne zvyšuje jeho efektívnosť. Môžeme povedať, že memy do určitej miery substitujú proces učenia. Ak je process učenia "drahý", potom prístup mémov je dôležitý factor prekonania barier učenia. Podobné výsledky sú prezentované aj na obrázku 4.12. Diagram A obsahuje porovnanie GA algoritmu s mémami a bez mémov (v oboch prípadoch s učením). Podobne, diagram B obsahuje dva priebehy priemerného fitnes s učením a bez učenie (v oboch prípadoch mémy sú zahrnuté).



Obrázok 4.12. (A) Dva priebehy priemerného fitnes pre GA so zahrnutím učenia (Baldwinovým efektom), ktoré sa odlišujú tým, že mémy sú alebo nie sú zahrnuté. (B) . Dva priebehy priemerného fitnes pre GA so zahrnutím mémov, ktoré sa odlišujú len tým, či učenie je al bo nie je zahrnuté.

Prezentované numerické výsledky ukazujú, že GA môžeme usporiadať do postupnosti s monotónnym zvyšovaním efektivity (pozri obrázok 4.13)

(a) Štandardný GA, fitnes je určené len štruktúrou chromozómu,
(b) GA s učením (Baldwinov efekt), fitnes chromozómu je určené prehľadávaním okolia daného chromozómu (tento process sa nazýva učenie),
(c) GA s učením a mémami (Baldwinov efekt plus mémy).



Obrázok 14.13. Rozdelenie GA do rôznych kategórií so zahrnutí/ bez zahrnutia učenia a mémov, idúc zhora nadol efktívnosť GA monotónne rastie.

Výsledky môžeme zhrnúť do týchto záverov:

1. Zahrnutie učenia (Baldinovho efektu) a Dawkinsových mémov do GA je významný intenzifikačným faktorom genetických algoritmov. Zhruba možno konštatovať, že efektívnosť akcelerácia GA týchto diametrálne odlišných prístupov je rovnaké. Oba

prístupu odlišnými mechanizmami napomáhajú prekonať medzigeneraečné informačné bariéry prostriedkami, ktoré zachovávajú charakter darwinovej evolúcie.

2. Pomocou memetického prístupu môžeme interpretovať prostriedkami multiagentových systémov význam kultúry v sociálnych systémov ako významného faktora akcelerácie evolúcie pomocou medzigeneračného prenosu poznatkov (význam kultúrnej alebo spoločnej pamäti) v sociálnych vedách.

Literatúra

[1] Chomsky, N.: As*pects of the Theory of Syntax*. Cambridge, MA: MIT Press 1965.

[2] Dawkins, R.: *The Selfish Gene*. Oxford, UK: Oxford University Press 1976 (existuje český preklad: *Sobecký gen*. Praha: Mladá Fronta 2003).

[3] Dennett, D. C. : *Darwin's Dangerous Idea - Evolution and the Meaning of Life.* London: Penguin Press 1995.

[4] Gálik, D.: Univezálny darvinizmus (Môže byť darvinizmus teóriou všetkého_), Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na internetovskej adrese http://math.chtf.stuba.sk/ CogSci_2000.html.

[5] Gálik, D.: Univerzálna teória evolúcie a problém mémov. In Nosek, J. (ed.), *Memy ve vědě a filosofii*. Praha: Filosofia 2004, pp. 133-146.

[6] Holland, J. H.: *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press 1975.

[7] Kauffman, S.A., Hohnsen, S.: Coevolution to the edge of chaos: Coupled fitnes landscape, poised states, and coevolutionary avalanches. *J. Theor. Biology* **149** (1991) 467-506.

[8] Kvasnička, V., Pospíchal, J., Tiňo, J.: *Evolučné algoritmy*. Bratislava: Vydavateľstvo STU 2000.

[9] Maynard Smith, J.: *Evolution and the Theory of Games*. Oxford, UK: Oxford University Press 1982.

[10] Mayr, E.: *Towards a New Philosophy of Biology.* Cambridge, MA: Harvard University Press 1988.

[11] Monod, J.: *Chance and Necessity. An Essay on the Natural Philosophy of Modern Biology.* New York: Vintage 2000 (existuje český preklad: *Náhoda a nutnosť*, Praha: vydav. Pavel Mervart 2008)

5. prednáška

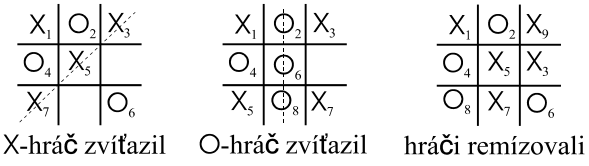
Učenie s odmenou a trestom a emergencia stratégie hry

5.1 Úvodné poznámky

Metódy riešenia zložitých úloh patria v umelej inteligencii medzi základné problémy už od jej vzniku v 50. rokoch. Prostriedkami klasickej (symbolickej) umelej inteligencie tieto metódy sa najčastejšie redukujú na rôzne prehľadávacie metódy stromov riešení, ktoré sú akcelerované rôznymi pravidlami a heuristikami. Aplikácie týchto symbolických metód na netriviálne problémy je ohraničená dimenziou ich stavových priestorov. Ak je táto dimenzia príliš veľká (N>10^5), potom prehľadávacie metódy v dôsledku existencie „kombinatoriálnej explózie poskytujú len veľmi približné výsledky v dôsledku nutnosti redukcie hĺbky prehľadávania v stromoch riešení. Preto sa v umelej inteligencii hľadajú nové prístupy a metafory k prekonaniu "kombinatoriálnej bariéry". Cieľom tohto článku je ukázať efektívnosť moderných metód umelej inteligencie, ktoré sú založené na metaforách Darwinovho prirodzeného výberu (evolučné metódy), ľudského mozgu (neurónové siete) a „reinforcement learning“ (veľká trieda metód učenia multiagentových systémov, ktorá sa nazýva „reinforcement learning“). Všetky tieto prístupy, samotné alebo ich kombinácia, sú v súčasnosti používané pre hľadanie riešenia zložitých úloh. Ako ilustráciu použitia týchto metód ukážeme na jednoduchej hre Tic-Tac-Toe (piškvorky) ako zostrojiť program, ktorý sa túto hru učí hrať. prvá dva prístupy (model 1 a model 2) budú vyžadovať , aby bol k dispozícii určitý model tejto hry. Potom na kvalite modelu závisia aj získané výsledky. Počítačový program hrá s týmto model neustále hry, na základe

5.2 Hra Tic-Tac-Toe (Piškvorky)

Detská hra Tic-Tac-Toe (TTT) má dvoch hráčov, prvý hráč je označený symbolom X a druhý hráč symbolom O. Snahou každého hráča je umiestniť na štvorcovej 3×3 hracej doske svoje symboly tak, aby tvorili buď riadok, stĺpec, alebo uhlopriečku. Hra je zahájená hráčom X, potom nasleduje hráč O, toto striedanie hráčov sa opakuje tak dlho, až je vytvorená požadovaná „riadková“ pozícia alebo na hracej doske je umiestnených deväť symbolov. Hru vyhráva ten hráč, ktorý prvý vytvoril požadovaný obrazec, alebo hra končí remízou, ak hracia doska obsahuje deväť symbolov a žiaden hráč nevytvoril požadovaný obrazec (pozri obr. 1).



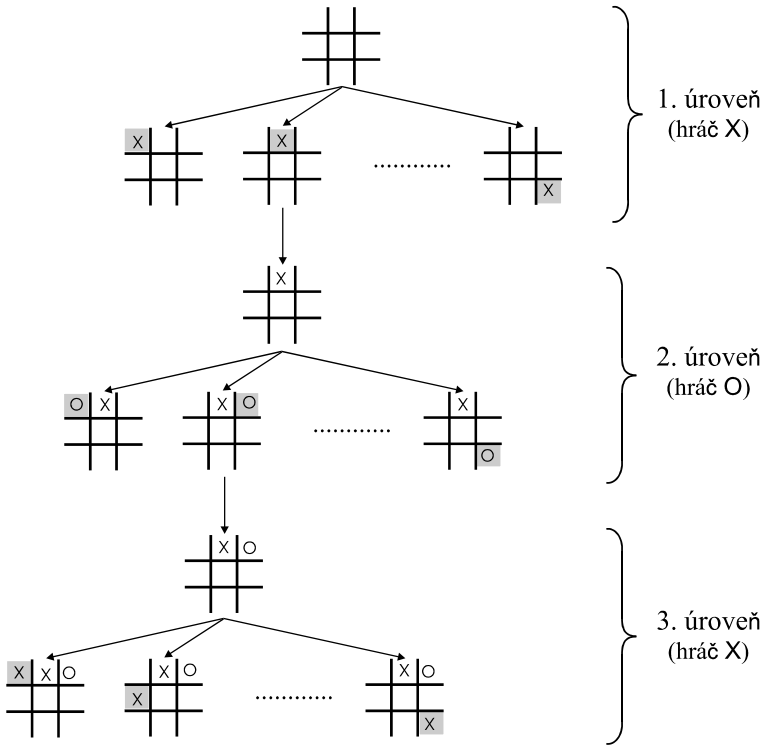
Obrázok 5.1. Znázornenie 3 partií hry TTT. Indexy pri jednotlivých symboloch X a/alebo O znamenajú poradie ťahu. V prvých dvoch partiách bolo dosiahnuté víťazstvo, hráč umiestnil svoje symboly do "riadku" naznačeného prerušovanou čiarou. Tretia partia skončila remízou, žiadnemu hráčovi sa nepodarilo umiestniť svoje symboly do "riadku", po 9 ťahoch, keď sú obsadené všetky pozície hracej dosky, hra končí remízou.

Hra TTT patrí medzi tzv. symetrické hry, z pohľadu druhého hráča je hra identická s hrou prvého hráča (hráči sú rovnocenní, odlišujú sa len v tom, že jeden z nich zahajuje hru). Obaja hráči riešia rovnaký strategický problém, vytvoriť čo najrýchlejšie „riadkovú“ pozíciu svojich znakov a súčasne zabrániť súperovi vytvorenie jeho „riadkovej“ pozície.

Dimenzia stavového priestoru hry TTT sa zostrojiť pomocou jednoduchých kombinatorických úvah, dostaneme

$$N = \sum_{p=1}^{5} \binom{9}{p} \left[\binom{9-p}{p-1} + \binom{9-p}{p} \right] - 2 \times 6 \times 13 = 5889 \qquad (1)$$

kde napr. pre p=1 dostaneme 9×(1+8) pozícií, ktoré obsahujú buď len jeden symbol X, alebo dva symboly X a O. Posledný člen na pravej strane odpovedá skutočnosti, že niektoré členy, obsahujúce jeden stľpec (riadok) znakov X a jeden stĺpec (riadok) znakov O, nemôžu byť zahrnuté v povolených pozíciách, týchto "nepovolených" pozícií je 156.



Obrázok 5.2. Znázornenie stromu riešení, ktorého vrchol (koreň) je prázdna hracia doska. Po prvom ťahu, ktorý hral hráč X, je obsadená jedna pozícia (zdôraznená vytieňovaným) symbolom X. V druhej úrovni, hranej hráčom O, je obsadená pozícia (vytieňovaná) symbolom O. Vetva stromu riešení končí vtedy, ak jeden hráč zvíťazil, alebo sú obsadené všetky pozície na hracej doske - hra skončila remízou. Konštrukcia tohto stromu riešení môže byť algoritmicky jednoducho realizovaná pomocou metódy spätného prehľadávania, výsledky tejto metódy sú uvedené v predchádzajúcej tabuľke.

Horný úsek stromu riešení je znázornený na obr. 2. Koreňom tohto stromu riešení je prázdna hracia doska, v druhej úrovni je 9 pozícií, ktoré obsahujú jeden X znak. Pomocou metódy spätného prehľadávanie sme zostrojili celý strom riešení a zistili sme, že má nasledujúci počet koncových pozícií, ktoré sú charakterizované takto

No.	Počet	Typ

1	131184	víťazstvo hráča X
2	77904	víťazstvo hráča O
3	46080	remíza hráčov X a O
	255168	celkový počet

Z tejto tabuľky vyplýva, že prvý hráč X má väčšiu šancu hru vyhrať. Podrobnou analýzou sa dá ukázať, že aj hráč O môže hru forsírovať tak, že remizuje. Celkový počet koncových vetví v strome riešení možno jednoducho odhadnúť ako 9!=362880.

Pre naše ďalšie štúdie bude potrebné mať k dispozícii určitý referenčný algoritmus hry TTT, ktorý aj keď nie je najefektívnejší, vedie obvykle aspoň k remíze. Uvedieme jednoduchý algoritmus, ktorý sa zakladá na tom, že hráč vykoná ťah, pričom tento je určený nasledujúcimi štyrmi pravidlami s klesajúcou prioritou

Model hry TTT.

1. pravidlo. *Hráč vykoná ťah, ktorý vedie k jeho víťazstvu.*

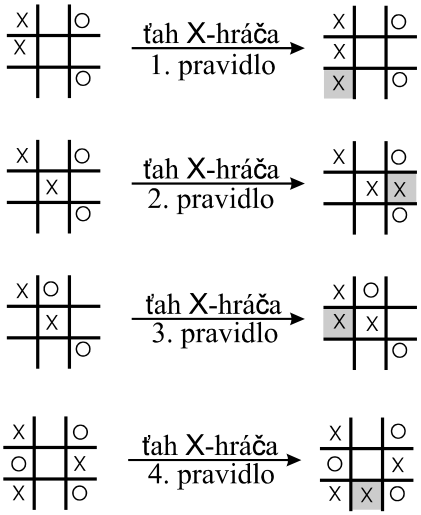
2. pravidlo. *Hráč vykoná ťah, ktorý zabráni víťazstvu oponenta v nasledujúcom ťahu.*

3. pravidlo. *Hráč vykoná ťah, ktorým si pripraví možnosť dvojitého použitia 1. pravidla* (*tzv. vidlička*).

4. pravidlo. *Hráč obsadí stredové pole.*

5. pravidlo. *Hráč obsadí rohové pole.*

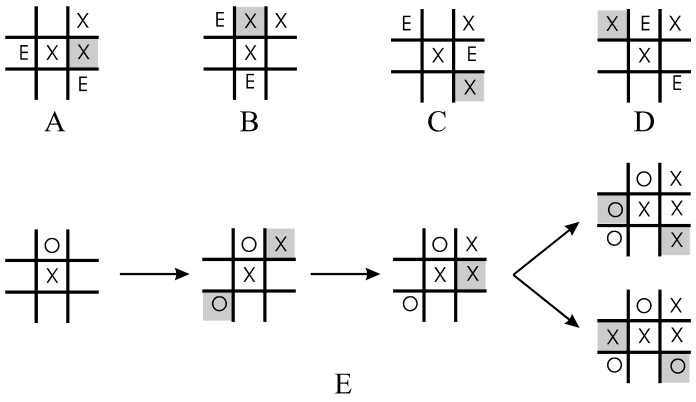
6. pravidlo. *Hráč obsadí volné pole.*



Obrázok 5.3. Ilustračné príklady aplikácie jednotlivých pravidiel z nášho modelu hry TTT. V prvom riadku je znázornená pozícia v ktorej hráč X môže v nasledujúcom ťahu vyhrať. Použitím pravidla 1, ktoré má najvyššiu prioritu, obsadí vytieňovanú pozíciu, hra končí víťazstvom hráča X. V druhom riadku je ilustrované pravidlo 2, podľa ktorého, ak hráč nemôže použiť prvé pravidlo, tak znemožní protihráčovi vyhrať tak, že mu obsadí pozíciu, ktorá viedla k jeho bezprostrednému víťazstvu. V treťom riedku je ilustrované pravidlo 3, ktoré umožňuje hráčovi vytvoriť tzv. vidličkovú pozíciu, ktorá vedie k jeho víťazstvu po dvoch ťahoch. V poslednom riadku je ilustrované pravidlo 4, ktoré sa aplikuje len vtedy, ak prvé tri pravidlá sú neaplikovateľné, hráč náhodne obsadí volnú pozíciu.

Model nie je plne deterministický, v prípade, že podľa niektorého aplikovaného pravidla existuje niekoľko možností, tak potom z nich vyberieme náhodne jednu z nich. Prvé

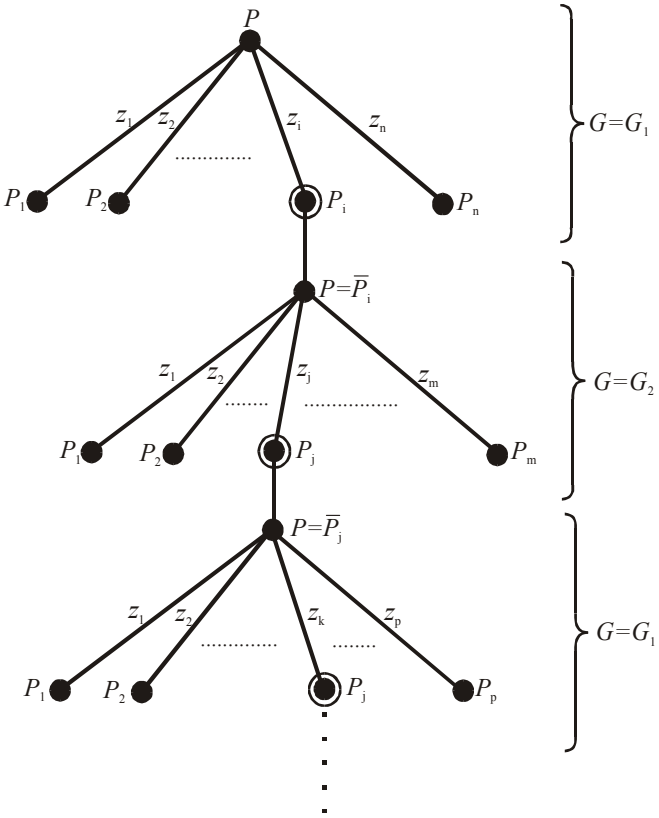
dve pravidlá odpovedajú jednoduchej rekognoskácii aktuálnej pozícii o jeden ťah vopred. Rekognoskácia o dva ťahy vopred je zahrnutá až v treťom previdle, pripravuje možnosť použitia 1. Pravidla. Jednotlivé pravidlá sú ilustrované na obr. 3. Tieto pravidlá poskytujú veľmi efektívny model hry TTT, ktorý dáva prvému hráčovi pomerne dobrú šancu zvíťaziť, druhý hráč pri presnej hre môže len forsírovať remízu. Na obr. 4, diagramy A-D, sú uvedené vybrané východzie pozície, kde je možné vytvoriť vetvenie pozícií, diagram E ilustruje priebeh hry, ktorá už po druhom ťahu druhého hráča je prehraná, prvý hráč môže vynútiť výhru.



Obrázok 5.4. Diagramy A-D znázorňujú základné typy vidličkových pozícií, ktoré sú aplikovateľné použitím pravidla 3. Diagram E ukazuje pozíciu, ktorá je prehraná pre hráča O už po druhom ťahu. Pravidlá nášho modelu hry TTT nepostihujú túto možnosť, jedná sa o predpoveď o tri ťahy dopredu.

5.3 Formalizácia hry TTT

V tejro kapitole sa budeme zaoberať formalizáciou hry TTT, ktorá je aplikovateľná pre všetky symetrické hry dvoch hráčov (šach, dáma, go, backgamon - vrhcáby, atď.). Nech aktuálne pozícia hry je popísaná premennou P, na túto pozíciu môžu byť aplikované prípustné akcie - ťahy tvoriace množinu $A(P)$. Použitím ťahu $a' \in A(P)$ pretransformujeme pozíciu P na novú pozíciu P', $P \xrightarrow{a'} P'$. Inverznú pozíciu $\overline{P}$ dostaneme z pozície P tak, že symboly X (O) zameníme za symboly O (X). Pojem inverznej pozície bude dôležitý pre formuláciu jednotného algoritmu pre symetrické hry, ktorý bude jednotný tak pre prvého ako aj druhého hráča. Použijeme multiagentový prístup, budeme predpokladať, že hru hrajú dvaja agenti G_1 a G_2, ktorý sú vybavený tzv. kognitívnym orgánom, pomocou ktorého sú schopných ohodnocovať jednotlivé nasledujúce pozície. Formulácia algoritmu je nasledovná (pozri obr. 5):



Obrázok 5.5. Diagramatická reprezentácia symetrickej hry pre dvoch hráčov. Vybraný hráč G, z aktuálnej pozície vytvorí všetky možné nasledujúce pozície P_1, P_2, … . Použitím svojho kognitívneho orgánu (reprezentovaného napríklad neurónovou sieťou) ohodnotí každú novú pozíciu číslom λ, ako svoj nasledujúci ťah vyberie si tú pozíciu, ktorá má maximálne ohodnotenie. Tento elementárny akt rozhodovania sa opakuje tak, že hráči sa striedajú. Oba hráči majú rovnocenný kognitívny orgán, to znamená, že ich deklarácia na prvého a druhého hráča je náhodná. Z tejto skutočnosti vyplýva, že druhý hráč používa svoj kognitívny orgán na invezné pozície $\overline{P}$, pristupuje k svojej aktuálnej pozícii ako keby bol prvý hráč. V opačnom prípade, keby hráč nerozlišoval či je prvý alebo druhý hráč, musel by mať dva kognitívne orgány, ktoré by používal v závislosti od toho, či je prvý alebo druhý hráč.

Algoritmus 0.

Krok 1. Hra je zahájená prvým hráčom, $G \leftarrow G_1$, a počiatočnou pozíciou, $P \leftarrow P_{\text{ini}}$.

Krok 2. *Hráč G vytvorí z pozície P množinu nasledujúcich pozícií* $A(P) = \{P_1, P_2, \ldots, P_n\}$.

(a) Ak množina je prázdna, potom oba hráči G_1 a G_2 remizujú a hra pokračuje krokom 4.

(b) V prípade, že hráč je prvý a hrá sa prvý ťah, tak náhodne vyberieme jednu pozíciu z A(P). V opačnom prípade každá pozícia P_i z množiny nasledujúcich pozícií je ohodnotená koeficientom $0<z_i<1$. Hráč vyberie za nasledujúcu pozíciu takú $P' \in A(P)$, ktorá je ohodnotená maximálnym koeficientom z, $P \leftarrow P'$. Ak pozícia P je víťazná, potom hráč G víťazí a hra pokračuje krokom 4.

Krok 3. *Hra prechádza na druhého hráča,* $G \leftarrow G_2$*, pozíciu P si vytvorí inverziou aktuálnej pozície,* $P \leftarrow \overline{P}$ *, hra pokračuje krokom 2.*

Krok 4. *Koniec hry.*

Podobne, ako pre model hry TTT (pozri Kapitolu 2), tento algoritmus nie je plne deterministický, v prípade prvého hráča a jeho prvého ťahu, ťah sa vyberie z množiny možných ťahov náhodne, potom je už hra plne deterministická. Kľúčovú úlohu v algoritme má výpočet koeficientov $z=z(P')$ pre pozície $P' \in A(P)$. Toto môže byť vykonané buď metódami klasickej umelej inteligencie, ktoré sú založené na kombinácií metód spätného prehľadávania a rôznych heuristík. Našu pozornosť usmerníme na moderný prístup multiagentových systémov, ktorý sa zakladá na predpoklade, že správanie sa agenta v prostredí a/alebo vykonávanie určitých činností je plne determinované jeho kognitívnym orgánom, ktorý vykazuje určitú plasticitu (t.j. je schopný učenia). Kognitívny orgán agenta G je obykle numericky realizovaný parametrickým zobrazením $G(w):P \rightarrow R$, kde w je parameter (parametre) zobrazenia a ktoré priradzuje každej pozícii (prostrediu) reálne číslo z, formálne $z=G(P,w)$. Zmenou w menia sa vlastnosti kognitívneho orgánu (t.j. dostávame iné parametrické zobrazenie), týmto spôsobom máme zabezpečenú plasticitu kognitívneho orgánu, čo je nutná podmienka procesu učenia agenta v danom prostredí, v procesu ktorého stále lepšie a lepšie zvláda požadovanú úlohu. Vo všeobecnosti parametrické zobrazenie $G(w):P \rightarrow R$ môže byť realizované mnohými rôznymi spôsobmi, od ktorých sa požaduje len splnenie podmienky „univerzálneho aproximátora“. Táto všeobecná podmienka je automaticky splnená v rámci požadovanej presnosti pre mocninné rozvoje alebo pre neurónové siete s dopredným šírením signálu a aspoň s jednou vrstvou skrytých neurónov.

Budeme predpokladať, že agent je schopný svojim kognitívnym orgánom ohodnocovať vznikajúce nasledujúce pozície tak v prípade, že je prvý, alebo aj druhý hráč. Tento predpoklad je plazibilný pre symetrické hry, kde strategické pravidlá tak pre prvého, ako aj druhého hráča sú rovnaké, určitý malý rozdiel však existuje v tom, že hru zahajuje prvý hráč a tak môže v určitom rozsahu vnucovať štýl hry druhému hráčovi. Vo vyššie uvedenom algoritme je táto požiadavka univerzálnosti kognitívneho orgánu (t.j. agent nemá dva kognitívne orgány, jeden pre prípad, že je prvým hráčom a iný pre prípad, že je druhým hráčom) realizovaná tým, že druhý hráč svojim kognitívnym orgánom neohodnocuje priamo pozície $P' \in A(P)$ ale ich inverzie $\overline{P'}$, t.j. agent aj ako druhý hráč ohodnocuje nasledujúce pozície z pohľadu prvého hráča, čiže kognitívny orgán je nastavený na to, že agent je akoby vždy prvým hráčom, v prípade, že je druhým hráčom, „ohodnocuje“ inverzné pozície.

5.4 Štruktúra kognitívneho orgánu - neurónovej siete

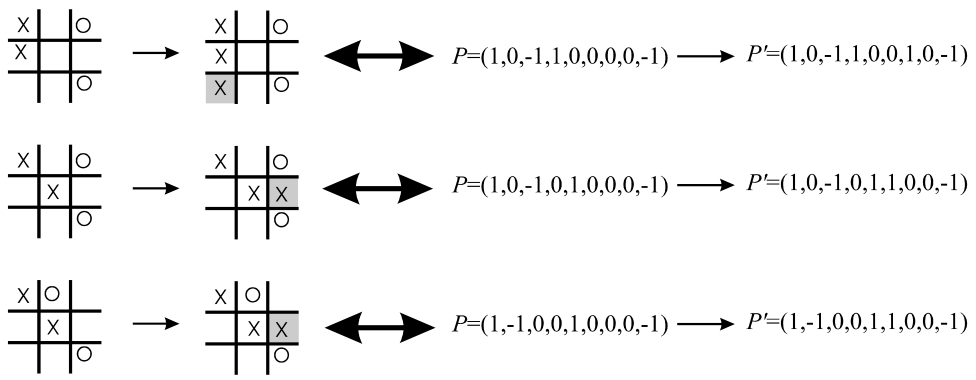
Prv ako pristúpime k špecifikácii kognitívneho orgánu agentov, musíme zaviesť tzv. numerickú reprezentáciu pozícií. Pozícia je reprezentovaná 9-rozmerným vektorom

$$x(P) = (x_1, x_2, \ldots, x_9) \in \{0,1,-1\}^9 \tag{2a}$$

kde jednotlivé zložky určujú jednotlivé políčka v pozícii P

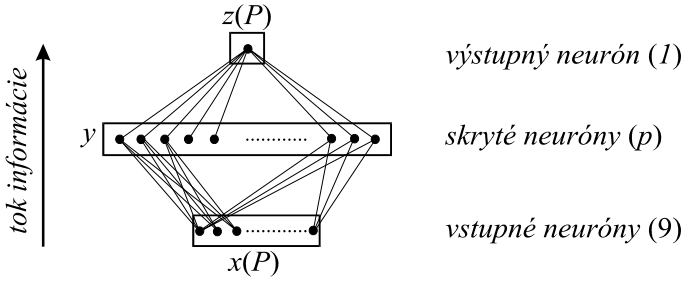
$$x_i = \begin{cases} 0 & (i - \text{té pole je neobsadené}) \\ 1 & (i - \text{té pole je obsadené X}) \\ -1 & (i - \text{té pole je obsadené O}) \end{cases} \tag{2b}$$

Tak napríklad, numerická reprezentácia pozícií na obr. 3 je znázornená na obr. 6.



Obrázok 5.6. Pozície sú reprezentované 9-rozmerným vektorom z $\{-1,0,1\}^9$. Ak je políčko v pozícii obsadené symbolom X (O) , potom príslušná zložka vektora sa rovná 1 (-1). V prípade, že políčko je prázdne, potom zložka vektora je nulová.

Použitá neurónová sieť má architektúru typu „feed-forward“, s jednou vrstvou skrytých neurónov. Aktivity vstupných neurónov sú určené numerickou reprezentáciou $x(P)$ danej pozície P, výstupná aktivita sa rovná ohodnoteniu pozície $\lambda(P)$ (pozri obr. 7), počet parametrov neurónovej siete je $11p+1$, kde p je počet skrytých neurónov. Plasticita takto špecifikovaného kognitívneho aparátu pomocou neurónovej siete je je špecifikovaná dvoma spôsobmi: (a) š**trukturálna plasticita,** ktorá znamená, že sa mení počet skrytých neurónov (t.j. mení sa štruktúra – topológia neurónovej siete), a (b) **parametrická plasticita**, ktorá odpovedá zmene váh spojov a prahových koeficientov skrytých a výstupných neurónov.



Obrázok 7. Dopredná (feed forward) neurónová sieť s jednou vrstvou skrytých neurónov. Vstupné aktivity sú rovné 9-rozmernému vektoru $\boldsymbol{x}(P)$, ktorý kóduje pozície hry. Výstupná aktivita sa rovná reálnemu číslu $z(P)$ z otvoreného intervalu (0,1), toto číslo je ohodnotenie vstupnej pozície.

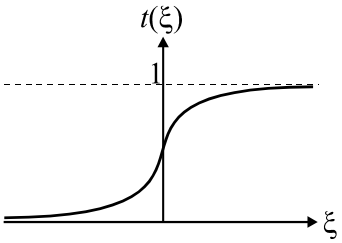
Aktivity neurónov skrytých neurónov a výstupného neurónu sú určené vzťahmi

$$y_i = t\left(\sum_{j=1}^{9} w_{ij} x_j + \vartheta_i\right) \quad (i = 1,2,\ldots,p) \tag{3a}$$

$$z = t\left(\sum_{i=1}^{p} \tilde{w}_j y_j + \tilde{\vartheta}\right) \tag{3b}$$

kde $t(\xi)$ je prechodová funkcia tvaru jednoduchej sigmoidy (pozri obr.8)

$$t(\xi) = \frac{1}{1+e^{-\xi}}, \qquad t: R \to (0,1) \quad \Rightarrow \quad 0 < \lambda < 1 \tag{4}$$



Obrázok 7. Priebeh funkcie sigmoidy definovanej vzťahom (4). Táto funkcia je kladná, monotónne rastúca a vyhovujúca asymptotickým podmienkam $t(\xi)\to 1$, pre $\xi\to\infty$ a $t(\xi)\to 0$, pre $\xi\to -\infty$.

Pri adaptácii neurónových sietí sa obvykle využíva parciálne derivácie výstupnej aktivity $\lambda(P)$ vzhľadom k parametrom neurónovej siete. Tieto parciálne derivácie vzhľadom k parametrom výstupného neurónu sú určené takto

$$\frac{\partial z(P)}{\partial \tilde{\vartheta}} = z(P)\left[1 - z(P)\right], \quad \frac{\partial z(P)}{\partial \tilde{w}_j} = \frac{\partial z(P)}{\partial \tilde{\vartheta}} y_j \tag{5a}$$

Podobným spôsobom sú určené aj parciálne derivácie vzhľadom k parametrom skrytých neurónov

$$\frac{\partial z(P)}{\partial \vartheta_i} = y_i(1 - y_i)\frac{\partial z(P)}{\partial \tilde{\vartheta}} \tilde{w}_i, \quad \frac{\partial z(P)}{\partial w_{ij}} = \frac{\partial z(P)}{\partial \vartheta_i} x_j \tag{5b}$$

Parciálne derivácie môžeme jednoducho počítať rekurentným postupom nazývaný „metóda spätného šírenia“, najprv sa spočítajú parciálne derivácie pre parametre výstupného neurónu, potom môžeme spočítať parciálne derivácie vzhľadom k parametrom skrytých neurónov.

5.5 Klasický prístup k adaptácii neurónových sietí reprezentujúcich kognitívny orgán agenta

Klasický prístup k adaptácii neurónových sietí je založený na použití tzv. tréningovej množiny, ktorá obsahuje dvojice vektorov vstupných a požadovaných výstupných aktivít (poznamenajme, že vektor vstupných aktivít obvykle je numerickou reprezentáciou objektu klasifikovaného neurónovou sieťou, pričom požadované výstupné aktivity klasifikujú daný objekt)

$$\mathrm{A}_{train} = \left\{ \boldsymbol{x} = (x_1, x_2, ..., x_n) / z_{req} \right\} \tag{6}$$

To znamená, že neurónová sieť je parametrické zobrazenie *n*-rozmerných vektorov na reálne číslo z otvoreného intervalu (0,1)

$$G(w): R^n \to (0,1) \tag{7}$$

Adaptácie neurónovej siete je realizovaná pomocou *minimalizácie* účelovej funkcie vyjadrujúcej sumu kvadrátov rozdielov medzi vypočítanými a požadovanými výstupnými aktivitami

$$E(w) = \frac{1}{2} \sum_{\boldsymbol{x}/z_{req} \in A_{train}} \left(z_{req} - G(\boldsymbol{x}; w) \right)^2 \tag{8}$$

Váhové koeficienty neurónovej siete sú upravované pomocou gradientovej metódy najprudšieho spádu (steepest descent)

$$w := w - \alpha \frac{\partial E}{\partial w} = w + \Delta w \tag{9a}$$

$$\Delta w = \alpha \sum_{\boldsymbol{x}/z_{req} \in A_{train}} \left(z_{req} - G(\boldsymbol{x}, w) \right) \frac{\partial G(\boldsymbol{x}, w)}{\partial w} \tag{9b}$$

kde α je kladný parameter nazývaný „krok učenia“. Výpočet parciálnych derivácií výstupných aktivít vzhľadom k parametrom neurónovej siete je popísaný vzťahmi (5a) a (5b).

Po všeobecnom úvode o štandardnom spôsobe adaptácie neurónových sietí obrátime našu pozornosť o použití tohto typu adaptácie neurónových sietí, ktoré slúžia ako kognitívny orgán aganta hrajúceho symetrickú hru pre dvoch hráčov. Pre väčšiu konkrétnosť našich úvah budeme konkrétne vždy hovoriť o TTT hre, avšak naše úvahy budú dostatočne všeobecné, aby boli aplikovateľné aj pre iné typy symetrických hier. Množinu $\mathtt{P}$ všetkých možných pozícií rozdelíme na dve disjunktné podmnožiny, ktoré obsahujú párny a nepárny počet znakov

$$\mathrm{P} = \mathrm{P}_{odd} \cup \mathrm{P}_{even} \tag{7a}$$

$$\mathrm{P}_{odd} = \left\{ (x_1, x_2, ..., x_9);\ \sum_{i=1}^{9} |x_i| = 2k+1,\ k = 0,1,...,4 \right\} \tag{7b}$$

$$\mathrm{P}_{even} = \left\{ (x_1, x_2, ..., x_9);\ \sum_{i=1}^{9} |x_i| = 2k,\ k = 1,...,4 \right\} \tag{7c}$$

Pozície z $\mathtt{P}_{\mathrm{odd}}$, obsahujúce nepárny počet znakov, sa vyskytujú v priebehu hry TTT u prvého hráča, zatiaľ čo, pozície z $\mathtt{P}_{\mathrm{even}}$, obsahujúce párny počet znakov, sa vyskytujú v priebehu hry TTT u druhého hráča. Potom z pohľadu toho-ktorého hráča, pozície z príslušnej tréningovej podmnožiny môžu byť ohodnotené reálnym číslom z_{req}, ktoré špecifikuje či daná pozícia vedie k víťaztvu (z_{req}=1), remíze (z_{req}=1/2) a prehre (z_{req}=0). Pre lepšie pochopenie týchto myšlienok obrátime našu pozornosť na obr. 4, kde je uvedená pozícia P=(0,-1,0,0,1,0,0,0,0), ktorá pri optimálnej hre prvého hráča vedie k jeho víťazstvu (táto skutočnosť je nezávislá od hry druhého hráča). To znamená, že pozícia P=(0,-1,0,0,1,0,0,0,0)$\in$ $\mathtt{P}_{\mathrm{even}}$ je ohodnotená z_{req}=1. Vo všeobecnosti, hodnotenie pozícií je netriviálna záležitosť a môže byť realizovaná s ohraničenou presnosťou "expertom", ktorý na základe svojich vedomostí, skúseností a intuície ohodnotí každú pozíciu z tréningovej množiny číslom z_{req}. Pre hru TTT, ktorá obsahuje „len“ okolo 8000 prípustných pozícií, je možné ohodnotenie pozícií vykonať pomocou „presnej“ metódy spätného prehľadávania. Žiaľ, tento priamočiary prístup je nerealizovateľný pre hry s podstatne väčším stavovým priestorom (napr. šach, dáma,…), časová náročnosť metódy spätného prehľadávania prudko rastie (exponenciálne) s počtom možných pozícií.

Predpokladajme, že pozície z množiny $\mathtt{P}$ máme ohodnotené číslami $z_{\mathrm{req}}\in[0,1]$, potom môžeme zostrojiť tréningovú množinu takto

$$\mathsf{A}_{train} = \left\{ \boldsymbol{x} / z_{req};\ x \in \mathsf{P} \right\} \tag{8}$$

to znamená, že každá pozícia z množiny $\mathtt{P}$ je ohodnotená reálnym číslom $z_{\mathrm{req}}\in[0,1]$ v závislosti od toho, či pozícia pri optimálnej hre daného hráča vedie buď k víťazstvu, remíze, alebo prehre. Tréningová množina (8) môže byť použitá ako základ pre realizáciu adaptačného procesu neurónovej siete – kognitívneho orgánu, ktorého parametre v priebehu učenia sa nastavia tak, aby kognitívny orgán správne klasifikoval pozície, t.j. predpokladá sa, že výstupná aktivita neurónovej siete aproximuje čísla z

$$z_{req} \approx G(\boldsymbol{x}, w) \qquad \forall\, \boldsymbol{x} / z_{req} \in \mathsf{A}_{train} \tag{9}$$

Obrazne povedané, úspešne adaptovaný kognitívny orgán klasifikuje pozície podobne, ako „expert“, pomocou ktorého sme vytvorili tréningovú množinu.

Niekoľko záverečných poznámok k tomuto priamočiaremu použitiu neurónovej siete k implemenmtácii algoritmu pre symetrickú hru dvoch hráčov. Aj keď je koncepčne veľmi jednoduchý a jedná sa o bezprostrednú aplikáciu neurónových sietí k problematike počítačovej implementácie symetrických hier, obsahuje vážnu reštrikciu ako je konštrukcia tréningovej množiny pre účely adaptačného procesu. Ohodnotenie každej pozície číslom λ pomocou „experta“ je realizovateľné s dostatočnou presnosťou len pre hry s relatívne malým stavovým priestorom. Podobne, použitie systematickej metódy „spätného prehľadávania“ je

aplikovateľné podobne len k hrám s malým stavovým priestorom, časová náročnosť „spätného prehľadávania“ prudko rastie so zväčšovaním stavového priestoru. Naznačený prístup bol úspešne použitý pre hru backgammon Geraldom Tesaurom z IBM v r. 1989 v programe *Neurogammon*, kde rozsiahla tréningová množina bola vytvorená analýzou mnoho tisíc „majstrovských partií“. Vytvorený program používajúci neurónovú sieť na ohodnotenie pozícií tejto hry patrila svojho času k najlepším programom pre hru backgammon a v r. 1989 vyhral olympiádu programov pre túto hru.

5.6 Adaptácia kognitívneho orgánu agenta pomocou „temporal difference“ TD(λ)-metódy „reinforcement learning“

V tejto kapitole uvedieme základné princípy modernej metódy „reinforcement learning“, ktorá v súčasnosti vo verzii „temporal difference“ TD(λ) patrí medzí účinné algoritmické prostriedky adaptácie kognitívnych orgánov multiagentových systémov. Základné princípy „reinforcement learning“ sú tieto: Agent sleduje závislosť medzi vstupným obrazcom a výstupným signálom jeho kognitívneho orgánu (ktorý sa často nazýva „akcia“ alebo „riadiaci signál“). Na základe externého skalárneho signálu „odmeny“ (reward) vyhodnocuje kvalitu výstupného signálu. Cieľom učenia je taká modifikácia kognitívneho orgánu agenta, aby výstupné signály maximalizovali príjem externých „reward“ signálov. V mnohých prípadoch signál „odmeny“ je časovo oneskorený, prichádza po dlhom slede akcií až na záver a môže byť chápaný ako ohodnotenie celej sekvencie akcií, toho či viedla k požadovanému výsledku alebo nie. V tomto prípade agent musí riešiť tzv. problém „temporal difference“ priradenia, kde učenie je založené na diferenciách medzi dočasne vykonaných predikciách pre jednotlivé elementy celkovej sekvencie akcií.

V tejto kapitole naznačíme konštrukciu metódy TD(λ) ako určité zovšeobecnenie štandardnej metódy učenia neurónových sietí popísanej v predchádzajúcej kapitole. Predpokladajme, že poznáme sekvenciu pozícií a jej ohodnotenie reálnym číslom, ktoré odpovedajú pozíciám daného agenta – hráča, ktoré musel ohodnocovať číslom z

$$P_1, P_2, \ldots, P_m, z_{reward} \tag{10}$$

kde z_{reward} je vonkajšie ohodnotenie sekvencie postupnosti a odpovedá skutočnosti či posledná pozícia P_m je pre daného agenta – hráča víťazná, remízová, alebo prehratá.

$$z_{reward} = \begin{cases} 1 & (sekvencia\ pozícií\ je\ vítazná) \\ 0.5 & (sekvencia\ pozícií\ je\ remízová) \\ 0 & (sekvencia\ pozícií\ je\ prehraná) \end{cases} \tag{11}$$

Zo sekvencie (10) vytvoríme m dvojíc pozícií a ich ohodnotenie číslom z_{reward}, vytvoríme nasledujúcu účelovú funkciu

$$E(w) = \frac{1}{2}\sum_{t=1}^{m}\left(z_{reward} - G(\boldsymbol{x}_t; w)\right)^2 \tag{12}$$

Budeme hľadať také váhové koeficienty neurónovej siete - kognitívneho orgánu, ktoré minimalizujú účelovú funkciu. v prípade, že sa nám podarí nájsť také váhové koeficienty siete, pre ktoré je účelová funkcia nulová, potom každá pozícia zo sekvencie (10) je ohodnotená číslom z_{reward}. Rekurnetná formula pre obnovu váhových koeficientov má tvar (pozri (9a) a (9b))

$$w := w - \alpha\frac{\partial E}{\partial w} = w + \Delta w \tag{13a}$$

$$\Delta w = \alpha \sum_{t=1}^{m} \left(z_{reward} - z_t \right) \frac{\partial z_t}{\partial w} \tag{13b}$$

kde $z_t = G(P_i, w)$ je ohodnotenie t-tej pozícii P_t pomocou neurónovej sieti - kognitívnejo orgánu číslom z_t. Našim cieľom bude, aby všetky pozície zo sekvencie (10) boli ohodnotené rovnakým číslom z_{reward}, ktoré nám špecifikuje, či hra daného hráča pozostávajúca zo sekvencií (10) bola víťazná, remízová, alebo viedla k prehre. Výraz v zátvorke v (13b) dá sa jednoducho prepísať do nasledujúceho tvaru

$$\begin{aligned} z_{reward} - z_t &= z_{m+1} - z_t \\ &= z_{m+1} - z_m + z_m - z_t \\ &= z_{m+1} - z_m + z_m - z_{m-1} + z_{m-1} - P_t \\ &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= \sum_{k=t}^{m} \left(z_{k+1} - z_k \right) \end{aligned} \tag{14}$$

K prepisu formule (13b) použijeme vzťah kde $z_{m+1} = z_{\text{rewar.d}}$. Dosadením vzťahu (14) do (13b) a použitím jednoduchej algebraickej identity

$$\sum_{t=1}^{m} \sum_{k=t}^{m} A_{kt} = \sum_{t=1}^{m} \sum_{k=1}^{t} A_{tk} \tag{15}$$

dostaneme

$$\begin{aligned} \Delta w &= \alpha \sum_{t=1}^{m} \left(\sum_{k=t}^{m} \left(z_{k+1} - z_k \right) \right) \frac{\partial z_t}{\partial w} \\ &= \alpha \sum_{t=1}^{m} \left(z_{t+1} - z_t \right) \sum_{k=1}^{t} \frac{\partial z_t}{\partial w} \end{aligned} \tag{16}$$

Sumarizujúc, inkrement obnovovacej formule váhového koeficienta má tvar

$$\Delta w = \sum_{t=1}^{m} \Delta w_t \tag{17a}$$

$$\Delta w_t = \alpha \left(P_{t+1} - P_t \right) \sum_{k=1}^{t} \frac{\partial z_k}{\partial w} \tag{17b}$$

Tento dôležitý výsledok môže byť "zovšeobecnený" na formulu, ktorá tvorí základ TD(λ) rodiny učiacich metód

$$\Delta w_t = \alpha \left(z_{t+1} - z_t \right) \sum_{k=1}^{t} \lambda^{t-k} \frac{\partial z_k}{\partial w} \tag{18}$$

kde parameter $0 \le \lambda \le 1$. Pre $\lambda = 1$ formula (18) poskytuje pôvodný výsledok (16), zatiaľ čo pre $\lambda = 0$ dáva

$$\Delta w_t = \alpha \left(z_{t+1} - z_t \right) \frac{\partial z_t}{\partial w} \tag{19}$$

t.j. inkrement váhového koeficienta je určený len posledným "pozorovaním".

Formule (17) a (18) umožňujú rekurentý výpočet inkrementu Δw. Zavedieme nový symbol $e_t(\lambda)$, ktorý sa dá jednoducho počítať rekurentne

$$e_t(\lambda) = \sum_{k=1}^{t} \lambda^{t-k} \frac{\partial z_k}{\partial w} \Rightarrow e_{t+1}(\lambda) = \lambda e_t(\lambda) + \frac{\partial z_{t+1}}{\partial w} \tag{20}$$

kde $e_1(\lambda) = \partial z_1 / \partial w$. Potom jednotlivé "parciálne" inkrementy Δw_t sú určené takto

$$\Delta w_t = \alpha \left(z_{t+1} - z_t \right) e_t(\lambda) \tag{21}$$

5.7 Adaptácia kognitívneho orgánu agenta pomocou „temporal difference“ TD(λ)-metódy „reinforcement learning“

V tejto kapitole budeme študovať prvý ilustratívny príklad, kde agent hrá proti modelu hry TTT, popísaného v kapitole 2. Pripomeňme, že sa jedná o pomerne kvalitný model, ktorý je schopný „predvídania dva ťahy vopred, menovite je schopný plánovania „vidličkových“ pozícií, proti ktorým protihráč, ak ich pripustí, je bezmocný. Ilustratívny príklad je realizovaný pomocou nasledujúceho algoritmu:

Algoritmus 1.

Krok 1. *Váhové koeficienty neurónovej siete sú náhodne vygenerované z intervalu [-1,1].*

Krok 2. *Polož* t:=1.

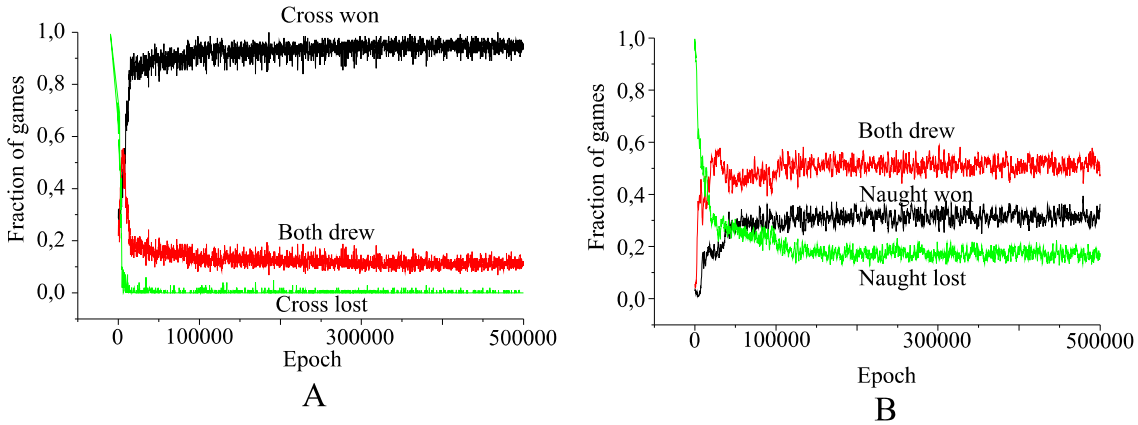
Krok 3. *S 50% pravdepodobnosťou deklaruj agenta ako prvého* X*-hráča a model ako druhého* O*-hráča (v opačnom prípade je agent deklarovaný ako druhý* O*-hráč a model ako prvý* X*-hráč). Na záver hry pomocou metódy TD(λ) opraví váhové koeficienty kognitívneho orgánu agenta.*

Krok 4. *Polož* t:=t+1.

Krok 5. *Ak* $t<t_{max}$*, potom pokračuj krokom 3, v opačnom prípade prejdi na krok 6.*

Krok 6. *Koniec algoritmu.*

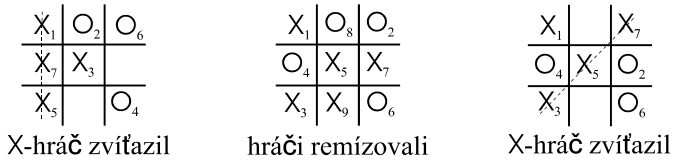
Tento algoritmus bol použitý na naše simulačné výpočty, pričom „reinforcement learning“ konštanta λ=0.3, krok učenia α=0.1, a počet skrytých neurónov N_H=50. Získané numerické výsledky sú sumarizované na obr. 8, diagramy A a B vyjadrujú frekvenciu výskytu víťazstva, prehry a remízy, ktoré v priebehu algoritmu boli počítané každých 200 epoch. Z diagramu A vyplýva, že asi po 100000 epochách kognitívny aparát agenta už bol tak adaptovaný, že ako prvý hráč neprehráva, približne 90% hier vyhráva a asi 10% remizuje. Žiaľ, výsledky pre agenta deklarovaného ako druhého hráča (diagram B) nie sú tak výborné, agent vyhráva len okolo 30% hier, 50% remizuje, a 20% prehráva. Tento pomer sa už nemení, aj keď pokračujeme v adaptácii systému až do 500000 epoch.



Obrázok 8. Diagram A znázorňuje priebeh výsledkov hry pre prvého hráča X vzhľadom k počtu adaptačných krokov, diagram B znázorňuje podobnú závislosť pre druhého hráča O. Z diagrama A vyplýva, že kognitívny orgán agenta vystupujúceho ako prvý hráč sa relatívne rýchlo pomocou TD(λ) učenia rýchlo adaptuje do formy, kde vyhráva skoro všetky hry, niekoľko málo remizuje a žiadne neprehráva. Ako vyplýva z diagramu B, učenie

kognitívneho orgánu pre hru druhého hráča nie je také úspešné ako v predchádzajúcom prípade, v priebehu celého učiaceho procesu zostáva agent v postavení druhého hráča neustálne prehráva okolo 30% hier.

Stojíme pred problémom, ako vysvetliť rozdielne výsledky získané pre agenta, keď je deklarovaný ako prvý alebo ako druhý hráč. V prípade, že agent je prvý hráč (obr.8, diagram A), učením kognitívneho orgánu relatívne rýchlo dostane stav, v ktorom neprehráva a podstatnú časť hier vyhráva. Za predpokladu, že agent je deklarovaný ako druhý hráč (obr. 8, diagram B), neustále určitú časť hier prehráva. Táto skutočnosť môže byť jednoducho vysvetlená nasledujúcim spôsobom: V prípade, že agent je deklarovaný ako prvý hráč, rôznorodosť hry je podstatne ohraničená skutočnosťou, že nepárne ťahy (prvý, tretí, …) sú deterministicky určené neurónovou sieťou agenta, t.j. produkovaný strom riešení je podstatne ohraničený skutočnosťou, že neurónová sieť stále produkuje rovnaký prvý ťah (pozri napr. obr. 9). V opačnom prípade, ak je agent deklarovaný ako druhý hráč, potom nepárne ťahy sú určené modelom, ktorý prvé ťahy určuje náhodne. Týmto spôsobom produkovaný strom riešení má podstatne bohatšiu štruktúru, než ako v predchádzajúcom prípade, keď nepárne ťahy boli deterministicky určené neurónovou sieťou agenta. Z týchto jednoduchých úvah vyplýva, že nároky na učenie neurónovej siete v prvom prípade sú podstatne menšie ako v druhom prípade, keď agent je deklarovaný ako druhý hráč. Možno očakávať, že pri enormnom množstve krokov učenia (epoch) sa bude podstatne zmenšovať počet prehier aj v druhom prípade, za predpokladu, že neurónová sieť agenta je dostatočne zložitá, aby mohla postihnúť aj heterogénnejšie vzory učenia. Na obr. 9 sú uvedené tri jednoduché ilustračné príklady hry agenta deklarovaného ako prvý hráč s adaptovanou neurónovou sieťou. V súhlase s našou vyššie uvedenou diskúziou, prvý ťah agenta je deterministicky určený ako vždy X_1, táto skutočnosť podstatne redukuje veľkosť produkovaného stromu riešení. Z uvedených príkladov vyplýva, že neurónová sieť suverénne zvládla „trik“ vzniku „vidličkovej“ pozície, ktorá automaticky vedie k jeho výhre.



Obrázok 9. Tri ilustratívne partie hrané adaptovaným agentom ako prvým hráčom proti modelu hry TTT. Prvá a tretia partia manifestujú skutočnosť, že adaptovaný kognitívny orgán agent je schopný efektívne využívať "vidličkové" pozície.

6. prednáška

Holografická redukovaná reprezentácia v umelej inteligencii a kognitíivnej vede

6.1 Úvodné poznámky

Moderný pohľad na vzťah medzi mozgom a mysľou vychádza z neurovednej paradigmy [3], podľa ktorej architektúra mozgu je určená spojmi medzi neurónmi, ich inhibičným, alebo excitačným charakterom a taktiež aj ich intenzitou. Ľudský mozog vykazuje neobyčajnú plasticitu, v priebehu učenia neustále vznikajú (ale taktiež aj zanikajú) synaptické spoje. Možno konštatovať, *že schopnosť mozgu vykonávať nielen kognitívne aktivity, ale byť aj pamäťou a aj riadiacim centrom pre našu motoriku, je plne zakódovaná jeho architektúrou*. Môžeme konštatovať, že predstava o ľudskom mozgu ako o počítači sa musí formulovať tak, že mozog je *paralelný distribuovaný počítač* (obsahujúci mnoho miliárd neurónov, elementárnych procesorov, ktoré sú medzi sebou poprepájané do zložitej neurónovej siete). Program v tomto paralelnom počítači je priamo zabudovaný do architektúry neurónovej siete, t. j. ľudský mozog je jednoúčelový paralelný počítač reprezentovaný neurónovou sieťou, ktorý nie je možné preprogramovať bez zmeny jeho architektúry.

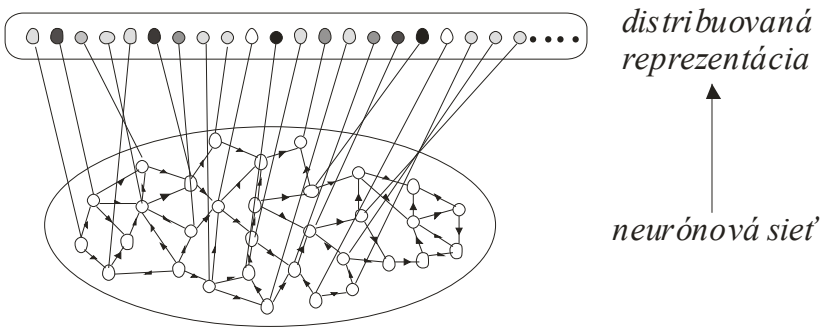
Z vyššie uvedených všeobecných úvah vyplýva, že myseľ s mozgom tvoria jeden integrálny celok, ktorý je charakterizovaný *komplementárnym dualizmon*. Myseľ je v tomto prístupe chápaná ako program vykonávaný mozgom, pričom tento program je špecifikovaný architektúrou distribuovanej neurónovej siete reprezentujúcej mozog. Mozog a myseľ tvoria dva rôzne pohľady na ten istý objektu:

(1) Keď hovoríme o mozgu, myslíme tým „hardwarovú“ štruktúru, biologicky realizovanú neurónmi a ich synaptickými spojmi (formálne reprezentovanú neurónovou sieťou), v opačnom prípade,

(2) keď hovoríme o mysli, myslíme tým kognitívne a iné podobné aktivity mozgu, ktoré sú vykonávané na symbolickej úrovni, kde prebieha transformácia symbolickej informácie na základe (jednoduchých) pravidiel.

Komplementárny dualizmus medzi mozgom a mysľou spôsobuje určité ťažkosti pri interpretácii kognitívnych aktivít mysle. Čisto neurálny prístup k interpretácii kognitívnych aktivít mysle sa upriamuje na hľadanie neurálnych korelátov medzi aktivitami neurónov a kognitívnymi aktivitami (konekcionizmus). Použitie neurálnej paradigmy k interpretácii symbolických kognitívnych aktivít má „vedľajší“ efekt v tom, že tieto sa nám „rozpúšťajú“ v ich mikroskopickom popise, symboly sa nám akoby strácajú v detailnom popise aktivít neurónov, intenzít synaptických spojov a pod. V opačnom prípade, absolutizovanie symbolickej paradigmy pri interpretácii kognitívnych aktivít mysli (kognitivizmus), ignorovanie skutočnosti, že myseľ je pevne ukotvená v mozgu, vedie ku koncepčnej sterilite, k snahe vysvetľovať kognitívne procesy ľudskej mysle len pomocou fenomenologických

termínov odvodených od koncepcie symbolu. Navrhujú sa umelé symbolické konštrukty (metódy, algoritmy a pod.), pre ktoré obvykle neexistuje žiadna experimentálna podpora v neurovede. Cieľom tejto práce je poukázať na alternatívny prístup, ktorý prekonáva „priepasť" medzi konekcionistickým a kognitivistickým prístupom k popisu a interpretácii kognitívnych aktivít ľudského mozgu [10-13]. Ukážeme, že použitie distribuovanej reprezentácie umožňuje integrovať konekcionizmus a kognitivizmus, kde mentálne reprezentácie (symboly) sú špecifikované pomocou distribuovaných obrazcov neurálnych aktivít, pričom nad týmito distribuovanými obrazcami môžeme zaviesť formálne algebraické operácie, ktoré umožňujú matematicky modelovať nielen kognitívne operácia, ale aj simulovať procesy prebiehajúci pri ukladaní a vyberaní informácie uloženej v pamäti.



Obrázok 6.1. Znázornenie prechodu od neurónovej siete k distribuovanej reprezentácii. Stav neurónovej siete v čase *t* je zadaný aktivitami jednotlivých neurónov, ktoré sú určené aktivitami v predchádzajúcom čase *t*-1 a váhovými koeficientmi jednotlivých orientovaných spojov. V určitej abstrakcie, môžeme tieto aktivity usporiadať do veľkého jednorozmerného poľa (vektora) útvaru, ktorého komponenty sú reálne čísla (ich veľkosť je určená stupňom šede príslušnej zložky – neurónu. V distribuovanej reprezentácii je ignorovaná architektúra neurónovej siete, t.j. dve distribuované reprezentácie musia byť chápané ako úplne nezávislé bez vzájomných súvislostí, ich kauzálna súvislosť pochádzajúca z neurónovej siete je teraz úplne ignorovaná. V distribuovanej reprezentácii zavádzame nové unárne a binárne operácie, pomocou ktorých vznikajú z pôvodných nové distribuované reprezentácie.

Našu pozornosť usmerníme na netradičný štýl uskutočňovania výpočtov pomocou distribuovaných obrazcov, ktorý sa diametrálne líši od klasických numerických a symbolických výpočtov a je vhodným modelovým prostriedkom pre pochopenie globálnych vlastností neurónových sietí. Ukážeme, že takéto „neuropočítanie" je založené na rozsiahlych náhodne vytvorených obrazcoch (reprezentovanými mnohorozmernými vektormi s náhodnými zložkami), pozri obr. 1. Tento prístup, ktorého základné prístupy boli formulované už koncom 60-tich rokov [2,4,5,9,14], bol zavŕšený sériou prác Tony Plateom [7-9] o „*holografickej redukovanej reprezentácii*" (HRR). Ukážeme, aké typy výpočtov sú týmto prístupom uskutočniteľné a či nám umožnia pochopiť procesy prebiehajúce v mozgu pri kognitívnych aktivitách. Náš príspevkom k rozvoju HRR spočíva v jeho aplikácia k modelovaniu kognitívnych procesov usudzovania pomocou pravidiel modus ponens a modus tollens. Kanerva [16-18] v polovici 90-tich rokov navrhol určitú alternatívu k HRR, ktorá je založená na náhodne generovaných binárnych vektoroch.

6.2 Matematická formulácia holografickej reprezentácie

Cieľom tejto kapitoly je prezentácia základných vlastností holografickej reprezentácie, ktorá bola vypracovaná A. Plateom [7-9]. Základný pojem je *konceptuálny vektor*, ktorý je reprezentovaný *n*-rozmerným vektorom

$$\boldsymbol{a} \in R^n \Rightarrow \boldsymbol{a} = \left(a_0, a_2, ..., a_{n-1}\right) \tag{1}$$

pričom jeho komponenty sú náhodné čísla so štandardným normálnym rozdelením

$$a_i = N\left(0,1/n\right) \quad \forall \in \left\{0,1,...,n-1\right\} \tag{2}$$

kde $N(0,1/n)$ je náhodné číslo so stredom v 0 a so strednou odchýlkou $1/n$.

Nad konceptuálnymi vektormi definujeme binárnu operáciu „*konvolúciu*“, ktorá dvojici vektorov priradí tretí vektor, $\otimes \;\; : \;\; R^n \times R^n \to R^n$, alebo

$$\boldsymbol{c} = \boldsymbol{a} \otimes \boldsymbol{b} \tag{3}$$

Zložky výsledného vektora $c = \left(c_0, c_1, ..., c_{n-1}\right)$ sú určené vzťahom

$$c_i = \sum_{j=0}^{n-1} a_j b_{[i-j]} \quad \left(i = 0,1,...,n-1\right) \tag{4}$$

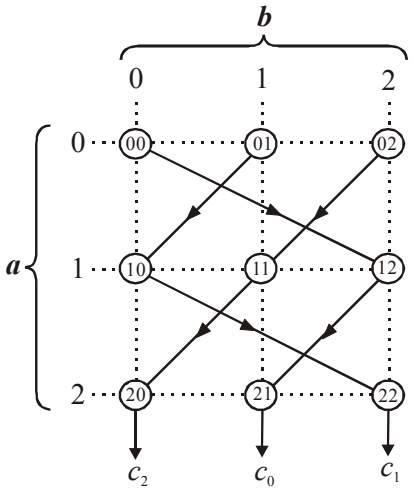
kde index v hranatej zátvorke, $[k]$, pomocou operácie *modulo n* takto[1]

$$k' = k \; mod \; n \tag{5a}$$

$$\left[k\right] = \begin{cases} k' & \left(if \;\; k' \geq 0\right) \\ n + k' & \left(if \;\; k' < 0\right) \end{cases} \tag{5b}$$

Konvolúcia dvoch vektorov $\boldsymbol{a}$ a $\boldsymbol{b}$ pre n=3 má tento tvar (pozri obrázok 2)

$$\begin{aligned} c_0 &= a_0 b_0 + a_1 b_2 + a_2 b_1 \\ c_1 &= a_0 b_1 + a_1 b_0 + a_2 b_2 \\ c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \end{aligned} \tag{6}$$



Obrázok 2. Znázornenie konvolúcie dvoch vektorov $\boldsymbol{a}$ a $\boldsymbol{b}$ pre n=3 (pozri (6)).

Konvolúcia vyhovuje týmto vlastnostiam:

(1) komutatívnosť, $\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{b} \otimes \boldsymbol{a}$

(2) asociatívnosť, $\left(\boldsymbol{a} \otimes \boldsymbol{b}\right) \otimes \boldsymbol{c} = \boldsymbol{a} \otimes \left(\boldsymbol{b} \otimes \boldsymbol{c}\right)$

(3) distributívnosť, $\boldsymbol{a} \otimes \left(\alpha \boldsymbol{b} + \beta \boldsymbol{c}\right) = \alpha\left(\boldsymbol{a} \otimes \boldsymbol{b}\right) + \beta\left(\boldsymbol{a} \otimes \boldsymbol{c}\right)$

(4) existencia jednotkového vektora, $\mathbf{1} \otimes \boldsymbol{a} = \boldsymbol{a} \quad \left(\mathbf{1} = \left(1,0,...,0\right)\right)$

[1] Štandardná definícia aritmetickej operácie *k modulo n* je určená ako zbytok po celočíselnom delení číslom *n*, Je potrebné poznamenať, že použitá definícia operácia *k modulo n* sa líši od tejto štandardnej definície pre záporné čísla *k*, ktorá poskytuje záporný výsledok. V našom prípade, ak je tento výsledok záporný, potom je transformovaný tak, že sa k nemu pripočíta *n*.

Konvolúcia môže byť taktiež vyjadrená pomocou *cirkulantnej matice*[1]

$$\left.\begin{aligned} c_0 &= a_0b_0 + a_1b_2 + a_2b_1 \\ c_1 &= a_0b_1 + a_1b_0 + a_2b_2 \\ c_2 &= a_0b_2 + a_1b_1 + a_2b_0 \end{aligned}\right\} \Rightarrow \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \underbrace{\begin{pmatrix} a_0 & a_2 & a_1 \\ a_1 & a_0 & a_2 \\ a_2 & a_1 & a_0 \end{pmatrix}}_{circ(a)} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \Rightarrow a \otimes b = circ(a)b \tag{7}$$

Tento konkrétny príklad je zovšeobecnený pre ľubovolnú dimenziu *n* takto

$$c = a \otimes b \Leftrightarrow c = circ(a)b \Leftrightarrow circ(a) = \begin{pmatrix} a_0 & a_{n-1} & .. & a_2 & a_1 \\ a_1 & a_0 & .. & a_3 & a_2 \\ .. & .. & .. & .. & .. \\ a_{n-2} & a_{n-3} & .. & a_0 & a_{n-1} \\ a_{n-1} & a_{n-2} & .. & a_1 & a_0 \end{pmatrix} \tag{8}$$

kde všeobecná cirkulantná matica má maticové elementy

$$\left(circ(a)\right)_{ij} = a_{[i-j]} \tag{9}$$

Cirkulantná matica má tieto vlastnosti

$$circ(a \otimes b) = circ(a)\,circ(b) \tag{10}$$

pretože konvolúcia je komutatívna operácia, potom cirkulantné matice navzájom komutujú

$$circ(a)\,circ(b) = circ(b)\,circ(a) \tag{11}$$

Nech X je inverzná matica k cirkulantnej matici $circ(a)$

$$X\,circ(a) = circ(a)\,X = E \tag{12}$$

Jej alternatívny tvar je

$$X = circ^{-1}(a) \tag{13}$$

Nech a^{-1} je *inverzný vektor* vzhľadom k vektoru a, $a^{-1} \otimes a = 1 = (1,0,...,0,0)$, potom za predpokladu, že cirkulantná matica je regulárna, $|circ(a)| \neq 0$, dostaneme

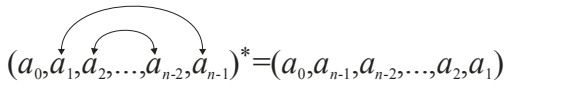
$$circ^{-1}(a) = circ\left(a^{-1}\right) \tag{14}$$

Definujme unárnu operáciu *involúcie* (pozri obrázok 3)

$$(\ )^* \quad : \quad R^n \to R^n \tag{15}$$

vzťahom

$$b = a^* = \left(a_{[0]}, a_{[-1]}, ..., a_{[-n+2]}, a_{[-n+1]}\right) \tag{16}$$



Obrázok 3. Znázornenie unárnej operácie involúcie.

Operácia involúcie spĺňa tieto vzťahy

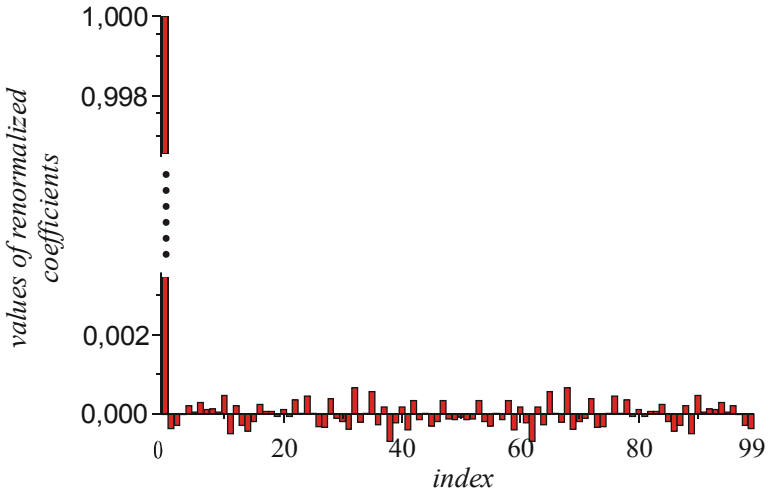
$$(\boldsymbol{a} + \boldsymbol{b})^* = \boldsymbol{a}^* + \boldsymbol{b}^* \tag{17a}$$

$$(\boldsymbol{a} \otimes \boldsymbol{b})^* = \boldsymbol{a}^* \otimes \boldsymbol{b}^* \tag{17b}$$

$$(\boldsymbol{a} \otimes \boldsymbol{b}^*) \cdot \boldsymbol{c} = \boldsymbol{a} \cdot (\boldsymbol{b} \otimes \boldsymbol{c}) \tag{17c}$$

$$\boldsymbol{a}^{**} = \boldsymbol{a} \tag{17d}$$

$$circ(a^*) = circ^T(a) \tag{17e}$$



Obrázok 4. Histogram jednotlivých zložiek súčinu $\boldsymbol{c}^* \otimes \boldsymbol{c}$, kde $\boldsymbol{c}$ je náhodne generovaný conceptuálny vektor pre n=100. Absolútna hodnota „prvej“ zložky $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_0$ je väčšia o niekoľko rádov ako absolútne hodnoty ostatných zložiek $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_i$, pre $i \geq 1$. To znamená, že súčin $\boldsymbol{c}^* \otimes \boldsymbol{c}$ po vhodnej normalizácii je približne rovný jednotkovému vektoru $(\boldsymbol{c} \cdot \boldsymbol{c})^{-1}\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right) \approx \mathbf{1} = (1,0,...,0,0)$.

Dolážeme, že involúcia $\boldsymbol{c}^*$ je približne rovná inverznému vektoru $\boldsymbol{c}^{-1}$, $\boldsymbol{c}^* \otimes \boldsymbol{c} \approx \mathbf{1}$. Študujme i-tú zložku konvolúcie $\boldsymbol{c}^* \otimes \boldsymbol{c}$

$$\begin{aligned}\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_i &= \sum_{k=0}^{n-1} c_k^* c_{[i-k]} = \sum_{k=0}^{n-1} c_{[-k]} c_{[i-k]} \\ &= \begin{cases} \boldsymbol{c} \cdot \boldsymbol{c} & \left(\text{pre } i = 0\right) \\ \sum_{k=0}^{n-1} c_{[-k]} c_{[i-k]} & \left(\text{pre } i > 0\right) \end{cases}\end{aligned} \tag{18}$$

Nultá zložka konvolúcie $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_0$ sa rovná skalárnemu súčinu $\boldsymbol{c} \cdot \boldsymbol{c}$, ktorý je vyhňjadrený pomocou kladných “diagonálnych“ členov c_i^2, zatiaľ čo ostatné zložky $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_i$, pre $i \geq 1$, sú určené sumami „nediagonálnych“ členov $c_i c_j$, ktoré majú náhodné znamienko. Dôsledkom tejto skutočnosti je, že $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_0$ je podstatne väčšia než ako $\left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right)_i$, pre $i \geq 1$, z čoho priamo vyplýva požadovaná skutočnosť $\boldsymbol{c}^* \otimes \boldsymbol{c} \approx \mathbf{1}$, (pozri obr. 4).

Jedna zo základných vlastností holografickej reprezentácie je možnosť rekonštrukcie pôvodných konceptuálnych vektorov, ktoré boli použité pri konštrukcii konvolúcie dvoch vektorov. Táto možnosť je veľmi dôležitá, pretože umožňuje dekódovať pôvodnú informáciu zo zložených konceptuálnych vektorov. *Rekonštrukcia* $\boldsymbol{x}$ z $\boldsymbol{c}\otimes\boldsymbol{x}$ je založená na skutočnosti, že že involúcia vektora je približne rovná jeho inverznej hodnote, t.j. $\boldsymbol{c}^* \otimes \boldsymbol{c} \approx \mathbf{1}$

$$\tilde{\boldsymbol{x}} = \boldsymbol{c}^* \otimes \left(\boldsymbol{c} \otimes \boldsymbol{x}\right) = \left(\boldsymbol{c}^* \otimes \boldsymbol{c}\right) \otimes \boldsymbol{x} \approx \frac{1}{\boldsymbol{c} \cdot \boldsymbol{c}} \mathbf{1} \otimes \boldsymbol{x} = \frac{1}{\boldsymbol{c} \cdot \boldsymbol{c}} \boldsymbol{x} \tag{19}$$

podľa ktorej, konvolúcia $\boldsymbol{c}^*$ s vektorom $\boldsymbol{c}\otimes\boldsymbol{x}$ produkuje vektor $\tilde{\boldsymbol{x}}$, ktorý je podobný pôvodnému vektoru $\boldsymbol{x}$, $\tilde{\boldsymbol{x}} \approx \boldsymbol{x}$. Tento výsledok vyjadríme takto

$$\frac{1}{(\boldsymbol{c}\cdot\boldsymbol{c})}\tilde{\boldsymbol{x}} = \begin{pmatrix} x_0 \\ x_1+\eta_1 \\ ..... \\ x_{n-1}+\eta_{n-1} \end{pmatrix} = \boldsymbol{x}+\boldsymbol{\eta} \tag{20}$$

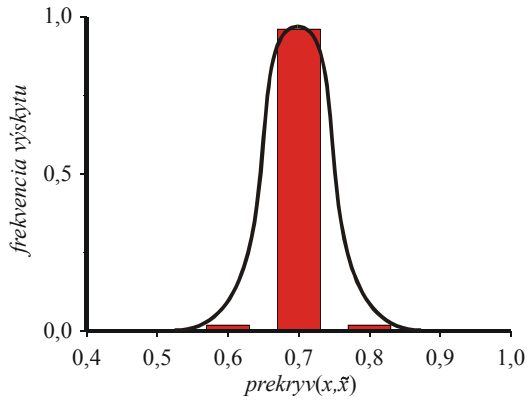
kde vektor $\boldsymbol{\eta}$ je interpretovaný ako náhodný šum s normálnou distribúciou so strtednou hodnotou v nule a so štandardnou odchýlkou omnoho menšou než akú má $\boldsymbol{x}$.

Prekryv rezultujúceho vektora $\tilde{\boldsymbol{x}}$ s pôvodným vektorom $\boldsymbol{x}$ je určený pomocou skalárneho súčinu takto

$$-1 \le prekryv(\boldsymbol{x},\tilde{\boldsymbol{x}}) = \frac{\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}}{|\boldsymbol{x}||\tilde{\boldsymbol{x}}|} \le 1 \tag{21}$$

kde nerovnosti vyplývajú priamo zo Schwartzovej nerovnosti z lineárnej algebry. Čím je táto hodnota bližšie k maximálnej hodnote 1, tým sú si vektory $\tilde{\boldsymbol{x}}$ a $\boldsymbol{x}$ *podobnejšie*[2].

Na obrázku 4 je znázornený histogram prekryvov pre súčin $\boldsymbol{c}\otimes\boldsymbol{x}$, obsahujúci dvojicu náhodne vygenerovaných rôznych konceptuálnych vektorov $\boldsymbol{c}$ a $\boldsymbol{x}$ dimenzie n=1000. Z obrázku vyplýva, že najčastejší prekryv medzi $\tilde{\boldsymbol{x}} = \boldsymbol{c}^*\otimes\boldsymbol{c}\otimes\boldsymbol{x}$ a $\boldsymbol{x}$ je okolo 0.7, z čoho vyplýva, že vektory $\tilde{\boldsymbol{x}}$ a $\boldsymbol{x}$ sú si podobné, $\tilde{\boldsymbol{x}} \approx \boldsymbol{x}$.



Obrázok 5. Histogram prekryvov medzi vektormi $\tilde{\boldsymbol{x}}$ a $\boldsymbol{x}$ (dimenzie n=1000) je s najväčšou frekvenciou okolo 0.7, z čoho vyplýva, že vektory $\tilde{\boldsymbol{x}}$ a $\boldsymbol{x}$ sú si podobné.

Obráťme pozornosť na druhú možnosť verifikácie formuly (20b) pomocou prístupu nazývaného „*aditívna pamäť*“. Majme množinu obsahujúcu p+q náhodne vygenerovaných konceptuálnych vektorov, $X = \{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_p,\boldsymbol{x}_{p+1},...,\boldsymbol{x}_{p+q}\}$, pričom p<q. Pomocou prvých p vektorov z X definujme *pamäťový vektor* $\boldsymbol{t}$ ako ich súčet

$$\boldsymbol{t} = \sum_{i=1}^{p}\boldsymbol{x}_i \tag{22}$$

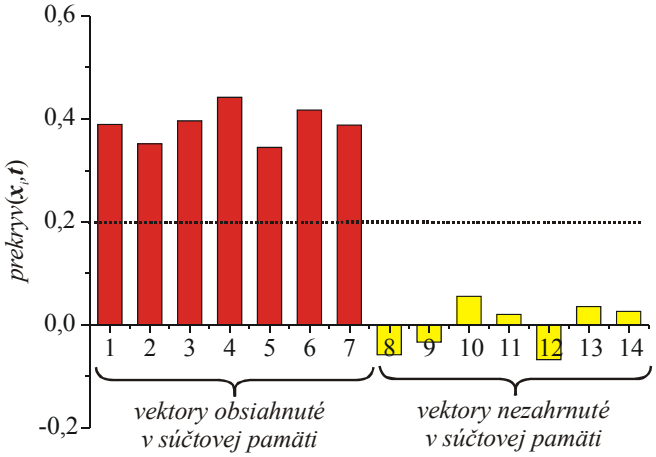
Vektor $\boldsymbol{t}$ reprezentuje aditívnu pamäť, ktorá jednoduchým aditívnym spôsobom obsahuje vektory z množiny X. Rozhodnutie o tom, či nejaký vektor $\boldsymbol{x}\in X$ je obsiahnutý v $\boldsymbol{t}$ musí byť založené na hodnote prekryvu (21)

$$prekryv(\boldsymbol{x},\boldsymbol{t}) = \frac{\boldsymbol{x}\cdot\boldsymbol{t}}{|\boldsymbol{x}||\boldsymbol{t}|} \tag{23}$$

Ak je táto hodnota väčšia ako vopred zvolená prahová hodnota, $prekryv(\boldsymbol{x},\boldsymbol{t}) \ge \vartheta$, potom

[2] V prípade, že prekryv sa blíži k hodnote -1, potom vektory $\tilde{\boldsymbol{x}}$ a $\boldsymbol{x}$ sú si taktiež podobné, aj keď sú opačne orientované (antikolineárne).

vektor $\boldsymbol{x}$ je zahrnutý v aditívnej pamäti $\boldsymbol{t}$, v opačnom prípade, ak $prekryv(\boldsymbol{x},\boldsymbol{t}) < \vartheta$, potom vektor $\boldsymbol{x}$ nie je zahrnutý v $\boldsymbol{t}$ (pozri obr. 5).



Obrázok 6. Znázornenie súčtovej pamäti pre prvých 7 vektorov množiny X, ktorá obsahuje 14 náhodne vygenerovaných konceptuálnych vektorov dimenzie n=1000. Prahovú hodnotu ϑ v tomto prípade môžeme zvoliť 0.2.

V oboch predchádzajúcich príkladoch (pozri obr. 3 a 4) bol použitý rovnaký spôsob určenia konceptuálnych vektorov, ktoré sa „vyskytujú“ v nejakom inom komplexnom konceptuálnom vektore (ktorý môže byť výsledkom zložitých predchádzajúcich výpočtov – transformácií). Použitý spôsob sa nazýva „*čistenie*“ (angl. *clean-up*), ktorý je formálne špecifikovaný takto: Majme množinu vektorov $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ a nejaký vektor $\boldsymbol{t}$. Stojíme pre problémom, rozhodnúť, či pamäťový vektor $\boldsymbol{t}$ obsahuje aditívnu komponentu, ktorá je podobná (alebo nie je podobná) nejakému vektoru z množiny X. Tento problém riešime pomocou výpočtu tzv. prekryvu (23), formálne

$$\boldsymbol{x} \approx \boldsymbol{t} = \begin{cases} áno & \left(prekryv(\boldsymbol{x},\boldsymbol{t}) \geq \vartheta\right) \\ nie & \left(prekryv(\boldsymbol{x},\boldsymbol{t}) < \vartheta\right) \end{cases} \tag{24}$$

kde ϑ je zvolená *prahová hodnota* akceptovania veľkosti prekryvu ako pozitívnej odpovedi. Výsledkom tohto procesu čistenia je podmnožina vektorov

$$X(\boldsymbol{t}) = \{\boldsymbol{x} \in X; \boldsymbol{x} \approx \boldsymbol{t}\} \subseteq X \tag{25}$$

Môžeme si položiť otázku aj v trochu inej podobe, a to, či pamäťový vektor $\boldsymbol{t}$ je podobný nejakému vektoru z množiny X? Odpoveď na túto všeobecnejšiu otázku rozhodneme podľa maximálnej hodnoty prekryvu

$$prekryv(\boldsymbol{t}, X) = \max_{x \in X} prekryv(\boldsymbol{t}, \boldsymbol{x}) \tag{26}$$

Potom môžeme prepísať (24) do tvaru

$$\boldsymbol{x} \approx X = \begin{cases} áno & \left(prekryv(\boldsymbol{x}, X) \geq \vartheta\right) \\ nie & \left(prekryv(\boldsymbol{x}, X) < \vartheta\right) \end{cases} \tag{27}$$

6.3 Asociačná pamäť

Konštrukcia asociačnej pamäti patrí medzi hlavné výsledky holografickej redukovanej reprezentácie, ktorá môže byť ďalej zovšeobecnená pomocou tzv. agregácie (angl. chunking).

Majme množinu konceptuálnych vektorov $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_n\}$ a tréningovú množinu $A_{train}=\{\boldsymbol{c}_i/\boldsymbol{x}_i\,;i=1,2,...,m\}$, ktorá obsahuje $m<n$ asociačných dvojíc konceptuálnych vektorov $\boldsymbol{c}_i/\boldsymbol{x}_i$, kde $\boldsymbol{c}_i$ je *vstup* do asociačnej pamäti (angl. *cue*) a $\boldsymbol{x}_i$ je výstup z pamäti. Zostrojíme pamäťový vektor $\boldsymbol{t}$, ktorý reprezentuje *asociačnú pamäť* nad tréningovou množinou A_{train}

$$\boldsymbol{t}=\boldsymbol{c}_1\otimes\boldsymbol{x}_1+...+\boldsymbol{c}_m\otimes\boldsymbol{x}_m=\sum_{i=1}^{m}\boldsymbol{c}_i\otimes\boldsymbol{x}_i \tag{28}$$

Predpokladajme, že vopred poznáme len vstupy asociačnej pamäti $\boldsymbol{c}_i$, nepoznáme možné výstupy z množiny $X_{train}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_m\}$. Odozva asociačnej pamäti na vstup – narážku $\boldsymbol{c}_i$ je určená pomocou procesu „čistenia“ reprezentovaného formulou (27). V prvom kroku spočítame vektor $\tilde{\boldsymbol{x}}_i=\boldsymbol{c}_i^*\otimes\boldsymbol{t}$, potom pomocou procesu založenom na maximálnej hodnoty prekryvu zistíme či $\tilde{\boldsymbol{x}}_i\approx\boldsymbol{x}_i\in X$

$$prekryv\left(\tilde{\boldsymbol{x}}_i,X\right)=\max_{x\in X_{train}}prekryv\left(\tilde{\boldsymbol{x}}_i,\boldsymbol{x}\right) \tag{29}$$

Asociačnú pamäť bude ilustrovaná nasledujúcimi dvoma príkladmi.

Prvý príklad

Tento príklad využíva len tréningovú množinu $A_{train}=\{\boldsymbol{c}_i/\boldsymbol{x}_i\,;i=1,2,...,m\}$, ktorá je náhodne vygenererovaná pre m=8, pričom dimenzia konceptuálnych vektorov je n=1000. Pre každú asociačnú dvojicu $\boldsymbol{c}_i/\boldsymbol{x}_i$ spočítame $\boldsymbol{t}_i=\boldsymbol{c}_i\otimes\boldsymbol{x}_i$. Hodnoty $prekryv\left(\boldsymbol{c}_i^*\otimes\boldsymbol{t}_i,\boldsymbol{x}_j\right)$ sú uvedené v tabuľke

	1	2	3	4	5	6	7	8
1	0.71703	-0.01820	0.01452	0.02776	-0.01488	-0.01922	-0.02442	0.01358
2	-0.03998	0.73804	0.01510	0.01430	0.00276	0.02346	-0.00545	-0.01626
3	-0.02757	-0.01736	0.64667	0.00474	-0.11580	-0.00812	0.01476	0.00379
4	0.00785	0.00374	-0.01899	0.68728	-0.15340	0.00005	-0.00561	0.00136
5	-0.00466	0.00426	-0.01831	-0.00827	0.70767	0.04175	-0.03384	-0.00668
6	-0.01467	0.02522	-0.01403	-0.01316	-0.03000	0.71444	0.00078	-0.00526
7	0.02966	0.00892	-0.00301	-0.00358	0.01285	0.00971	0.70790	0.01816
8	-0.00344	-0.01080	0.00843	-0.01871	0.00324	-0.02629	0.00851	0.58957

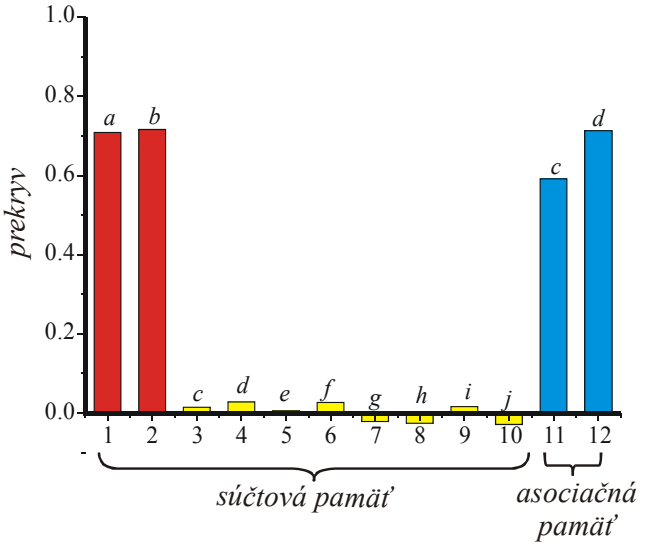
Z tejto tabuľky jasne vyplýva, že prekryvy sú dostatočne veľké len pre diagonálne hodnoty, zatiaľ čo nediagonálne prekryvy sú rádovo menšie. Môžeme teda jednoznačne rozhodnúť pomocou prekryvu, či $\boldsymbol{c}_i^*\otimes\boldsymbol{t}_i\approx\boldsymbol{x}_i$ je asociantom s „narážkou“ $\mathbf{c}_i$.

Druhý príklad

V tomto ilustračnom príklade budeme využívať tréningovú množinu $A_{train}=\{\boldsymbol{c}_i/\boldsymbol{x}_i\}$, generovanú pre m=10 asociačných párov – vektorov dimenzie n=1000. Táto pamäť je reprezentovaná pamäťovým vektorom $\boldsymbol{t}=\boldsymbol{c}_1\otimes\boldsymbol{x}_1+...+\boldsymbol{c}_m\otimes\boldsymbol{x}_m$. Nasledujúca tabuľka obsahuje 20 pokusov „čistenia“, pri ktorých sme s 50% pravdepodobnosťou použili ako asociačný vstup vektor $\boldsymbol{c}_i$ z tréningovej množiny alebo náhodne vygenerovaný konceptuálny vektor. Tabuľka obsahuje maximálne hodnoty prekryvov (29), pomocou ktorých môžeme jednoznačne identifikovať, či použitý vstup má asociáciu v rámci tréningovej množiny.

#	max. prekryv	index vstupu	index výstupu s max. prekryvom
1	0.311	6	6
2	0.047	náhod.gener.	neexistuje
3	0.383	5	5
4	0.373	10	10
5	0.316	3	3
6	0.397	4	4
7	0.074	náhod.gener.	neexistuje
8	0.065	náhod.gener.	neexistuje
9	0.069	náhod.gener.	neexistuje
10	0.039	náhod.gener.	neexistuje
11	0.344	7	7
12	0.402	8	8
13	0.032	náhod.gener.	neexistuje
14	0.073	náhod.gener.	neexistuje
15	0.017	náhod.gener.	neexistuje
16	0.004	náhod.gener.	neexistuje
17	0.033	náhod.gener.	neexistuje
18	0.056	náhod.gener.	neexistuje
19	0.373	10	10
20	0.037	náhod.gener.	neexistuje

Z tabuľky vyplýva, že asociačná pamäť pomocou procesu čistenia je jednoznačne identifikovateľná, pre konceptuálne vektory hodnoty maximálneho prekryvu dobre špecifikujú existenciu (alebo neexistenciu) príslušných asociačných výstupov.



Obrázok 6.6. Znázornenie rekognoskácie kombinácie súčtovej a asociačnej pamäti. V prvom kroku sme vykonali proces „čistenia“, pomocou ktorého sme zistili, že pamäť obsahuje len dve položky $\boldsymbol{a}$ a $\boldsymbol{b}$, ktoré v ďalšom kroku slúžia ako vstup pre rekognoskáciu asociačnej pamäti, kde sme zistili ako výstupy vektory $\boldsymbol{c}$ a $\boldsymbol{d}$. Pri procese „čistenia“ sme ešte naviac verifikovali s negatívnym výsledkom, či súčtová časť pamäti neobsahuje aj ďalšie vektory $\boldsymbol{c},\boldsymbol{d},...,\boldsymbol{i},\boldsymbol{j}$. S veľkou pravdepodobnosťou môžeme teda rekognoskovať pamäťový vektor $\boldsymbol{t}$ ako kombináciu súčtovej a asociačnej pamäti $\boldsymbol{t} = \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{a} \otimes \boldsymbol{c} + \boldsymbol{b} \otimes \boldsymbol{d}$. Dimenzia použitých vektorov je n=1000.

Kombinácia súčtovej a asociačnej pamäti

Ukážeme, že aj kombinácia súčtovej a asociačnej pamäti je spoľahlivo rekognoskovateľná, čo sa ukáže v našich ďalších aplikáciách ako výhodná črta holografickej reprezentácie. Predpokladajme, že máme 10 konceptuálnych vektorov $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d},...,\boldsymbol{i},\boldsymbol{j}$, pomocou prvých štyroch zostrojíme kombináciu súčtovej a asociatívnej pamäti takto

$$\boldsymbol{t} = \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{a} \otimes \boldsymbol{c} + \boldsymbol{b} \otimes \boldsymbol{d} \tag{30}$$

Pomocou procedúry „čistenia“ získame, že vektor $\boldsymbol{t}$ obsahuje ako zložky vektory $\boldsymbol{a}$ a $\boldsymbol{b}$, ktoré v následnom kroku odstránim z vektora $\boldsymbol{t}$

$$\boldsymbol{t}' = \boldsymbol{t} - \boldsymbol{a} - \boldsymbol{b} = \boldsymbol{a} \otimes \boldsymbol{c} + \boldsymbol{b} \otimes \boldsymbol{d} \tag{31}$$

Z odstránenej aditívnej časti použitím techniky rekognoskácia asociatívnej pamäti zistíme, že obsahuje dve dvojice $\boldsymbol{a}/\boldsymbol{c}$ a $\boldsymbol{b}/\boldsymbol{d}$, pozri obr. 6.

6.4 Postupnosť symbolov

Konštrukcia asociačnej pamäti neumožňuje uchovať štruktúrovanú informácia, cieľom tejto kapitoly je ukázať, že holografická distribuovaná reprezentácia je schopná spracovať lineárny reťazec znakov, ktoré sú reprezentovanú postupnosťou konceptuálnych vektorov.

Pre konkrétnosť našich úvah, študujme postupnosť 6 konceptuálnych vektorov dimenzie n=1000

$$\textit{postupnosť} = \{\boldsymbol{a} \to \boldsymbol{b} \to \boldsymbol{c} \to \boldsymbol{d} \to \boldsymbol{e} \to \boldsymbol{f}\} \tag{32}$$

Pre tieto vektory zostrojíme pamäťový vektor

$$\boldsymbol{t}_0 = \boldsymbol{a} + \boldsymbol{a} \otimes \boldsymbol{b} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d} \otimes \boldsymbol{e} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d} \otimes \boldsymbol{e} \otimes \boldsymbol{f} \tag{33}$$

Vieme, že tento vektor obsahuje postupnosť vektorov zakódovanú vektorom (33), avšak nevieme ktorých vektorov a v akom poradí. Ukážeme, že pomocou procedúry „čistenia“ z vektora $\boldsymbol{t}_0$ môžeme postupne rekonštruovať pôvodnú postupnosť (32) nasledujúcou procedúrou (pozri obr. 7):

1. krok: $\boldsymbol{a} = clean\_up(\boldsymbol{t}_0)$, $\boldsymbol{t}_1 := \boldsymbol{t}_0 - \boldsymbol{a}$,

   $\tilde{\boldsymbol{t}}_1 := \boldsymbol{a}^* \otimes \boldsymbol{t}_1$,

2. krok: $\boldsymbol{b} = clean\_up(\tilde{\boldsymbol{t}}_1)$, $\boldsymbol{t}_2 := \boldsymbol{t}_1 - \boldsymbol{a} \otimes \boldsymbol{b}$,

   $\tilde{\boldsymbol{t}}_2 := (\boldsymbol{a} \otimes \boldsymbol{b})^* \otimes \boldsymbol{t}_2$,

3. krok: $\boldsymbol{c} = clean\_up(\tilde{\boldsymbol{t}}_2)$, $\boldsymbol{t}_3 := \boldsymbol{t}_2 - \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c}$,

   $\tilde{\boldsymbol{t}}_3 := (\boldsymbol{y}_1 \otimes \boldsymbol{y}_2 \otimes \boldsymbol{y}_3)^* \otimes \boldsymbol{t}_3$,

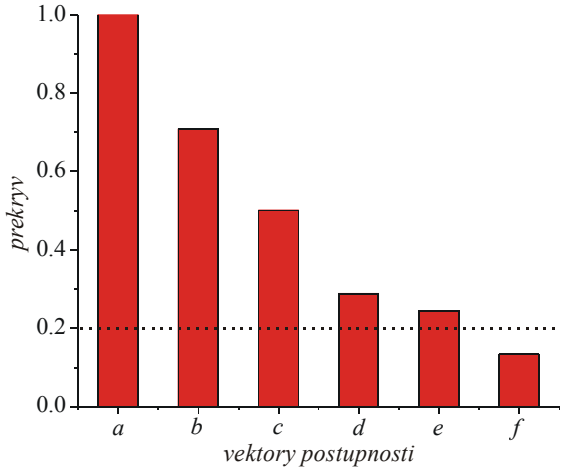
4. krok: $\boldsymbol{d} = clean\_up(\tilde{\boldsymbol{t}}_3)$, $\boldsymbol{t}_4 := \boldsymbol{t}_3 - \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d}$,

   $\tilde{\boldsymbol{t}}_4 := (\boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d})^* \otimes \boldsymbol{t}_4$,

5. krok: $\boldsymbol{e} = clean\_up(\tilde{\boldsymbol{t}}_4)$, $\boldsymbol{t}_5 := \boldsymbol{t}_4 - \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d} \otimes \boldsymbol{e}$,

   $\tilde{\boldsymbol{t}}_5 := (\boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \otimes \boldsymbol{d} \otimes \boldsymbol{e})^* \otimes \boldsymbol{t}_5$,

6. krok: $\boldsymbol{f} = clean\_up(\tilde{\boldsymbol{t}}_5)$.



Obrázok 6.8. Prekryv pre jednotlivé vektory s postupnosti z (30), ktoré boli získané rekonštrukciou z vektora $\boldsymbol{t}_0$. Z obrázku vyplýva dôležitý poznatok, že nastáva pomerne rýchla degradácia rekonštrukcie, už šiesty vektor f je rekonštruovaný s pravdepodobnosťou menšou ako 0.20.

Funkcia *clean_up*(·) vykonáva „čistenie" pre daný vektor $\boldsymbol{t}$ vzhľadom k množine vektorov $X = \{\boldsymbol{a}, \boldsymbol{b}, ..., \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h}, ...\}$. Jednotlivé kroky rekonštrukcie postupnosti konceptuálnych vektorov – znakov sú znázornené na obr. 8, z ktorého vyplýva, že proces rekonštrukcie postupnosti symbolov pomerne rýchlo degraduje, už pri šiestom vektore je prekryv menší ako 0.2.

Postupnosť symbolov môže byť kódovaná taktiež pomocou asociačnej pamäti, kde vektor vstupu $\boldsymbol{c}_i$ špecifikuje i-tú pozíciu daného symbola. Vyššie uvedený ilustračný príklad je reprezentovaný pamäťovým vektorom

$$\boldsymbol{t} = \boldsymbol{c}_1 \otimes \boldsymbol{a} + \boldsymbol{c}_2 \otimes \boldsymbol{b} + \boldsymbol{c}_3 \otimes \boldsymbol{c} + \boldsymbol{c}_4 \otimes \boldsymbol{d} + \boldsymbol{c}_5 \otimes \boldsymbol{e} + \boldsymbol{c}_6 \otimes \boldsymbol{f} \tag{34}$$

Potom rekognoskácia tejto postupnosti spočíva v hľadaní asocianta k vstupnému vektoru $\boldsymbol{c}_i$, pomocou procesu čistenia zostrojíme „tréningovú množinu"

$$A_{train} = \{\boldsymbol{c}_1/\boldsymbol{a}, \boldsymbol{c}_2/\boldsymbol{b}, \boldsymbol{c}_3/\boldsymbol{c}, \boldsymbol{c}_4/\boldsymbol{d}, \boldsymbol{c}_5/\boldsymbol{e}, \boldsymbol{c}_6/\boldsymbol{f}\} \tag{35}$$

ktorá jednoznačne špecifikuje postupnosť vektorov. Výhoda tohto postupu je v tom, že presnosť rekognoskácia nedegraduje tak rýchlo ako pri pôvodnom postupe založenom na pamäťovom vektore (33).

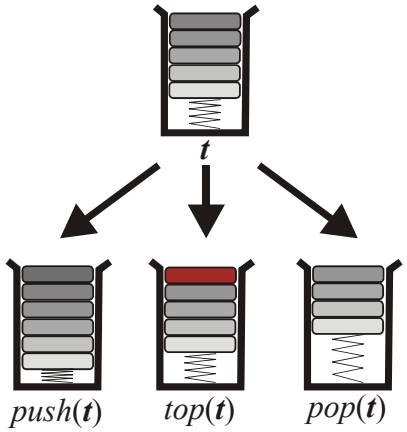
Asociačný prístup k implementácii pamäti pre postupnosť symbolov (reprezentovných vektormi $\boldsymbol{a}$, $\boldsymbol{b}$,) je možné jednoducho zmodifikovať do tvaru tzv. „*zásobníkovej pamäti*". Asociačné vstupy $\boldsymbol{c}_i$ sú určené takto

$$\boldsymbol{c}_i = \boldsymbol{p}^i \quad (i = 1, 2, ...) \tag{34}$$

kde $\boldsymbol{p}^i$ je i-tá (konvolučná) mocnina náhodne generovaného konceptuálneho vektora $\boldsymbol{p}$, $\boldsymbol{p}^i = \boldsymbol{p}^{i-1} \otimes \boldsymbol{p}$. Potom pamäťový vektor (32) má tvar

$$\boldsymbol{t} = \boldsymbol{p} \otimes \boldsymbol{x}_1 + \boldsymbol{p}^2 \otimes \boldsymbol{x}_2 + \boldsymbol{p}^3 \otimes \boldsymbol{x}_3 + ... + \boldsymbol{p}^n \otimes \boldsymbol{x}_n \tag{35}$$

kde $\boldsymbol{x}_i$ sú jednotlivé položky pamäti {$\boldsymbol{a}$, $\boldsymbol{b}$,...}. Takto interpretovaná asociatívna pamäť pre postupnosť symbolov sa nazýva „*zásobníková pamäť*", pomocou mocninových vstupov $\boldsymbol{p}^i$ môžeme jednoducho meniť jej obsah, pozri obr. 9. Nad touto pamäťou definujeme tri rôzne operácie, pomocou ktorých môžeme meniť jej obsah:



Obrázok 6.9. Tri možné operácie nad zásobníkovou pamäťou reprezentovanou vektorom.

(1) $push(\boldsymbol{t},\boldsymbol{x}) = \boldsymbol{p} \otimes \boldsymbol{x} + \boldsymbol{p} \otimes \boldsymbol{t}$, nová položka $\boldsymbol{x}$ je zavedená na vrchol zásobníka.

(2) $top(\boldsymbol{t}) = clean\_up(\boldsymbol{p}^* \otimes \boldsymbol{t})$, rekognoskuje vrchnú položku v zásobníku.

(3) $pop(\boldsymbol{t}) = \boldsymbol{p}^{-1} \otimes \boldsymbol{t} - top(\boldsymbol{t})$, zo zásobníka je odstránená vrchná položka.

Najproblematickejšia je tretia operácia, pomocou ktorého je zo zásobníka odstránená vrchná položka. Korektná implementácia požaduje použitie presného inverzného vektora $\boldsymbol{p}^{-1}$, aproximácia tohto inverzného vektora pomocou involúcie, $\boldsymbol{p}^{-1} \approx \boldsymbol{p}^*$, vedie k rýchlej degradácii zásobníkovej pamäti.

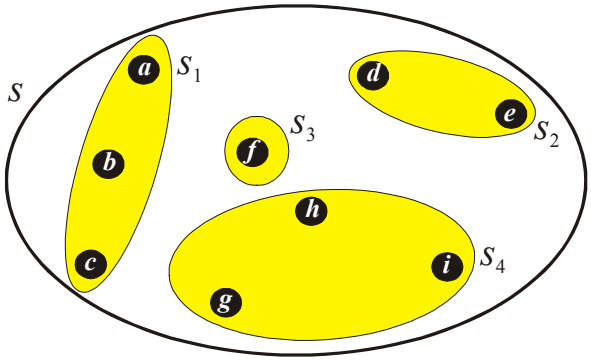
6.5 Agregovaná pamät

Agregovaná pamäť pomáha prekonávať problémy s degradáciou pamäti pre postupnosť symbolov (pozri kap. 4). Majme množinu konceptuálnych vektorov $S = \{\boldsymbol{a},\boldsymbol{b},...,\boldsymbol{k},\boldsymbol{l},...\}$, túto množinu rozložíme na disjunktné podmnožiny - *agregáty*

$$S = S_1 \cup S_2 \cup S_3 \cup S_4 \cup ... \qquad (S_i \cup S_j = \varnothing, \text{ pre } i \neq j) \qquad (36)$$

Študujme množinu $S = \{a,b,c,d,e,f,g,h\}$, jej rozklad na agregáty vypadá takto (pozri obr. 8)

$$S_1 = \{a,b,c\},\ S_2 = \{d,e\},\ S_3 = \{f\},\ \text{and}\ S_4 = \{g,h\} \qquad (37)$$



Obrázok 10. Ilustrácia rozkladu množiny *S* na disjunktné agregáty, pozri (36).

Agregovaná pamäť je reprezentovaná pamäťovým vektorom, ktorý reprezentuje postupnosť agregátov $\{\boldsymbol{s}_1 \rightarrow \boldsymbol{s}_2 \rightarrow \boldsymbol{s}_3 \rightarrow \boldsymbol{s}_3\}$

$$\boldsymbol{t} = \boldsymbol{s}_1 + \boldsymbol{s}_1 \otimes \boldsymbol{s}_2 + \boldsymbol{s}_1 \otimes \boldsymbol{s}_2 \otimes \boldsymbol{s}_3 + \boldsymbol{s}_1 \otimes \boldsymbol{s}_2 \otimes \boldsymbol{s}_3 \otimes \boldsymbol{s}_4 \tag{38b}$$

pričom jednotlivé agregáty sú definované takto pomocou príslušných postupnosti vektorov (pozri obr. 9)

$$\boldsymbol{s}_1 = \boldsymbol{a} + \boldsymbol{a} \otimes \boldsymbol{b} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \tag{38c}$$

$$\boldsymbol{s}_2 = \boldsymbol{d} + \boldsymbol{d} \otimes \boldsymbol{e} \tag{38d}$$

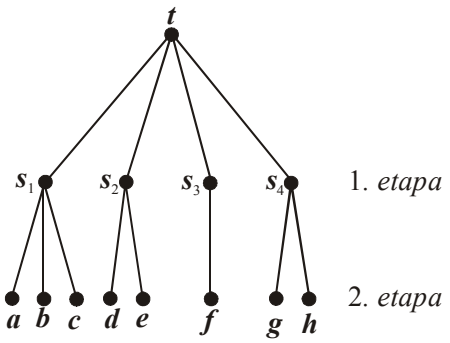
$$\boldsymbol{s}_3 = \boldsymbol{f} \tag{38e}$$

$$\boldsymbol{s}_4 = \boldsymbol{g} + \boldsymbol{g} \otimes \boldsymbol{h} \tag{38f}$$

Proces spracovania agregovanej pamäti obsahuje dve etapy:

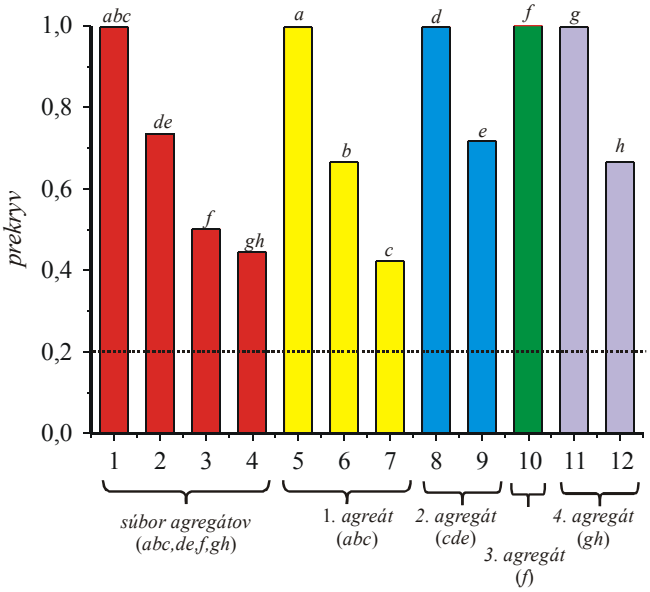
1. etapa – pomocou procesu „čistenia“ identifikujeme agregáty vyskytujúce sa v $\boldsymbol{t}$ (predpokladáme, že „upratovanie“ má množinu $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ z konca 2. kapitoly, kde bol tento proces špecifikovaný, rozšírenú aj o agregáty $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4$, t.j. v našom ilustračnom príklade $X = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h}, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4\}$).

2. etapa – identifikované agregáty sú ďalej analyzované pomocou procesu upratovania.



Obrázok 11. Agregácia 8 conceptuálnych vektorov na 4 agregáty. V 1. etape proces „upratovania“ identifikuje agregáty, ktoré sú potom v 2. etape ďalej identifikované na elementárne konceptuálne vektory.

Výsledky dvoj-etapového procesu „čistenia“ sú znázornené na obr. 11. Z tohto obrázku vyplýva, že v prípade dlhej sekvencie konceptuálnych vektorov rýchla degradácia procesu „čistenia“ môže čiastočne prekonaná pomocou agregácia konceptov na agregáty, ktoré sú na najvyššej úrovni samostatne kódované.



Obrázok 12. Znázornenie dvoj-etapového procesu „čistenia“, pričom v prvej etape sú identifikované agregáty, zatiaľ čo v druhej etape sú identifikované elementárne konceptuálne vektory z už identifikovaných agregátov.

Agregačná metóda môže byť zovšeobecnená tak, že sa vytvárajú agregáty vyšších rádov, t.j. agregáty zložené z agregátov, pozri obr. 12, kde jednotlivé agregáty sú definované takto (použité konceptuálne vektory $\boldsymbol{a}$, $\boldsymbol{b}$, ..., $\boldsymbol{c}$ majú dimenziu n=1000)

$$\boldsymbol{s}_1 = \boldsymbol{a} + \boldsymbol{a} \otimes \boldsymbol{b} + \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \tag{39a}$$

$$\boldsymbol{s}_2 = \boldsymbol{d} + \boldsymbol{d} \otimes \boldsymbol{e} \tag{39b}$$

$$\boldsymbol{s}_3 = \boldsymbol{f} \tag{39c}$$

$$\boldsymbol{s}_4 = \boldsymbol{g} + \boldsymbol{g} \otimes \boldsymbol{h} \tag{39d}$$

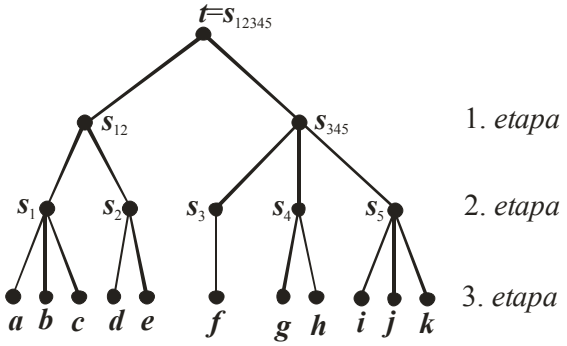
$$\boldsymbol{s}_5 = \boldsymbol{i} + \boldsymbol{i} \otimes \boldsymbol{j} + \boldsymbol{i} \otimes \boldsymbol{j} \otimes \boldsymbol{k} \tag{39e}$$

$$\boldsymbol{s}_{12} = \boldsymbol{s}_1 + \boldsymbol{s}_1 \otimes \boldsymbol{s}_2 \tag{39f}$$

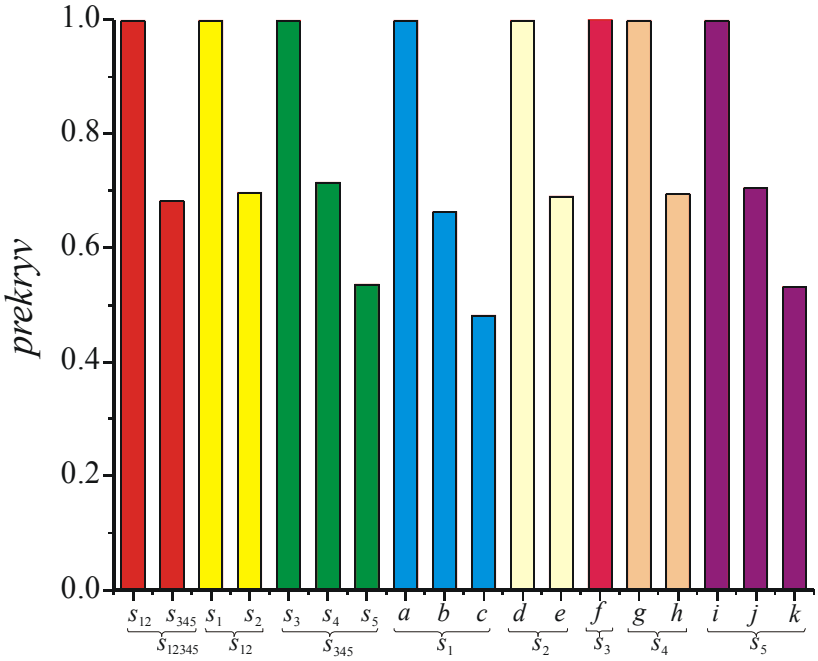
$$\boldsymbol{s}_{345} = \boldsymbol{s}_3 + \boldsymbol{s}_3 \otimes \boldsymbol{s}_4 + \boldsymbol{s}_3 \otimes \boldsymbol{s}_4 \otimes \boldsymbol{s}_5 \tag{39g}$$

$$\boldsymbol{t} = \boldsymbol{s}_{12345} = \boldsymbol{s}_{12} + \boldsymbol{s}_{12} \otimes \boldsymbol{s}_{345} \tag{39h}$$

Proces „čistenia“ agregovaného vektora $\boldsymbol{t}$ špecifikovaného (39h) je znázornená na obr. 13. V 1. etape rekognoskujeme vektor $\boldsymbol{t}$, zistíme, že obsahuje dva agregáty $\boldsymbol{s}_{12}$ a $\boldsymbol{s}_{345}$. V 2. etape analyzujeme agregáty z predchádxajúcej 1, etapy, zistíme, že obsahujú agregáty $\boldsymbol{s}_1$, $\boldsymbol{s}_2$,..., $\boldsymbol{s}_5$. V poslednej tretej etape analyzujeme agregáty z predchádzajúceho kroku, ktoré už obsahujú elementárne konceptuálne vektory $\boldsymbol{a}$, $\boldsymbol{b}$, ..., $\boldsymbol{k}$. Prekryvy rezultujúcich vektorov v procesu „čistenia“ sú znázornené na obr. 14.



Obrázok 13. Znázornenie agregácie vyššieho rádu, kde sú použité agregáty zložené z agregátov.



Obrázok 14. Znázornenie 3-etapového čistenia agregovaného vektora $\boldsymbol{t}$ špecifikovaného vzťahom (39h). V prvej etape sa rekognoskuje vektor $\boldsymbol{t}$, zistí sa, že obsahuje dva agregáty $\boldsymbol{s}_{12}$ a $\boldsymbol{s}_{345}$. V druhej etape sa rekognoskujú dva agregáty z predchádzajúcej etapy, zistí sa, že obsahujú agregáty $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, ..., $\boldsymbol{s}_5$. V poslednej tretej etape sa rekognoskuje postupne 5 agregátov identifikovaných v predchádzajúcej etape, obsahujú konceptuálne vektory $\boldsymbol{a}$, $\boldsymbol{b}$, ..., $\boldsymbol{k}$.

Z uvedených ilustračných príkladov vyplýva, že prístup agregovanej pamäti reprezentuje efektívny spôsob prekonania rýchlej degradácie pôvodnej verzie postupnej rekognoskácie vektora (33). Tým, že združíme niekoľko konceptuálnych vektorov do agregátu, získame jednoduchú možnosť rozšírenia našich možností korektne rekognoskovať väčšie množiny konceptuálnych vektorov. Proces agregácie môže mať niekoľko hierarchických úrovní, čím sa naše možnosti zapamätania a rekognoskácie konceptuálnych vektorov stávajú skoro neohraničené.

6.6 Kódovanie relácií

Holografická redukovaná reprezentácia je vhodným prostriedkom aj na kódovanie relácií (predikátov). Študujme binárnu reláciu $P(x,y)$, použijúc Pascalovský zápis, táto relácia je formálne špecifikovaná touto hlavičkou

$$function\ P\left(x:type_1;\ y:type_2\right):type_3 \tag{40}$$

Jednotlivé argumenty relácie sú špecifikované typmi $type_1$ a $type_2$, ktoré špecifikujú doménu nad ktorou sú tieto premenné určené; podobne aj samotná relácia P je chápaná ako funkcia, ktorej doména hodnôt je špecifikovaná typom $type_3$. V mnohých prípadoch doména premenných a aj samotnej relácie je jedna, preto ich špecifikácie môžu byť vynechané, čo podstatne zjednodušuje holografickú reprezentáciu relácií. Redukovaný tvar relácie (40) vypadá takto

$$function\ P\left(x;\ y\right) \tag{41}$$

kde už vopred vieme akého typu sú premenné x, y, a taktiež aj samotná relácia P. Holografická reprezentácia relácie (40) môže mať tento tvar

$$\begin{aligned} \boldsymbol{t} &= \boldsymbol{P} + \boldsymbol{variable}_1 + \boldsymbol{variable}_2 + \\ &\boldsymbol{P} \otimes \left(\boldsymbol{type}_3 + \boldsymbol{variable}_1 \otimes \left(\boldsymbol{x} + \boldsymbol{type}_1 \right) + \boldsymbol{variable}_2 \otimes \left(\boldsymbol{y} + \boldsymbol{type}_2 \right) \right) \end{aligned} \tag{42}$$

Jej dekódovanie prebieha postupne. V prvej etape pomocou procedúry „čistenia“ rekognoskujeme názov (identifikátor) relácie ***P*** a taktiež aj názvy (identifikátory) jej jednotlivých premenných ***x*** a ***y***. V druhom kroku rekognoskujeme typ $\boldsymbol{type}_3$ relácie ***P***, v záverečnom treťom kroku pomocou predchádzajúcich výsledkov rekognoskujeme premenné ***x***, ***y*** a taktiež aj ich typy $\boldsymbol{type}_1$ a $\boldsymbol{type}_2$. V mnohých prípadoch reprezentácia predikátora $P(x,y)$ je postačujúca v tomto zjednodušenom tvare (pozri (41))

$$\boldsymbol{t} = \boldsymbol{P} + \boldsymbol{variable}_1 \otimes \boldsymbol{x} + \boldsymbol{variable}_2 \otimes \boldsymbol{y} \tag{43}$$

Zvolená metóda holografickej reprezentácie relácie je ľahko zovšeobecniteľná aj zložitejšie relácie, kde premenné sú taktiež predikáty, napr. $P\left(x, Q\left(y,z\right)\right)$, kde „vnútorný“ predikát Q je charakterizovaný

$$function\, Q\left(y : type_3; z : type_4\right) : type_5 \tag{44}$$

K tomu, aby sme mohli vytvoriť zloženú reláciu $P\left(x, Q\left(y,z\right)\right)$, musíme predpokladať typovú kompatibilitu druhej premennej relácie P a typu relácie Q, t.j. $type_2 = type_4$.V zjednodušenom prístupe, kde všetky typy sú rovnaké, nie je potrebné rozlišovať typy jednotlivých premenných a samotných relácií. Zjednodušená holografická reprezentácia relácie (44) má tento tvar

$$\boldsymbol{t}' = \boldsymbol{Q} + \boldsymbol{variable}_1 \otimes \boldsymbol{y} + \boldsymbol{variable}_2 \otimes \boldsymbol{z} \tag{45}$$

Dosadením reprezentácie (45) za premennú ***y*** v reprezentácii (43) dostaneme túto výslednú reprezentáciu zloženej relácie $P\left(x, Q\left(y,z\right)\right)$

$$\begin{aligned} \boldsymbol{t} &= \boldsymbol{P} + \boldsymbol{variable}_1 \otimes \boldsymbol{x} + \boldsymbol{variable}_2 \otimes \left(\boldsymbol{Q} + \boldsymbol{variable}_3 \otimes \boldsymbol{y} + \boldsymbol{variable}_4 \otimes \boldsymbol{z} \right) \\ &= \boldsymbol{P} + \boldsymbol{variable}_1 \otimes \boldsymbol{x} + \boldsymbol{variable}_2 \otimes \boldsymbol{Q} + \\ &\quad \boldsymbol{variable}_2 \otimes \boldsymbol{variable}_3 \otimes \boldsymbol{y} + \boldsymbol{variable}_2 \otimes \boldsymbol{variable}_4 \otimes \boldsymbol{z} \end{aligned} \tag{46}$$

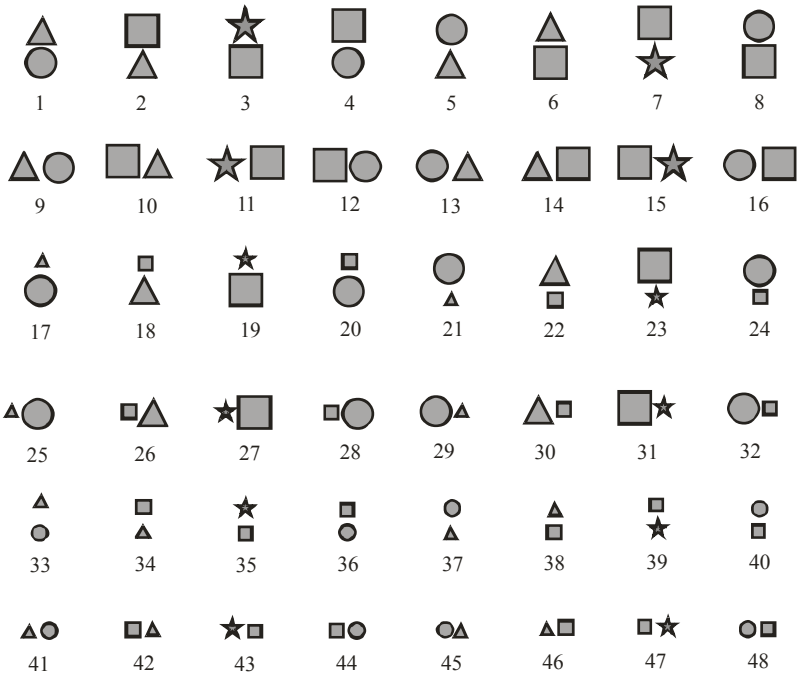
Prvý ilustračný príklad – podobnosť medzi geometrickými obrázkami

Na obrázku 14 je znázornených 48=6×8 geometrických útvarov, ktoré obsahujú v horizontálnom alebo vertikálnom usporiadaní dva objekty, ktoré naviac môžu byť ešte malé alebo veľké. Označme holografické reprezentácie príslušných atomárnych koncepcií takto:

Obrazce: ***tr*** (triangle), ***sq*** (square), ***ci***(circle), ***st*** (star)
Unárne relácie: ***sm*** (small), ***lg*** (large)
Binárne relácie: ***hor*** (horizontal), ***ver*** (vertical)
Premenné: $\boldsymbol{ver\_var_1}$ (1st variable for binary relation ***ver***), $\boldsymbol{ver\_var_1}$ (2nd variable for binary relation ***ver***), $\boldsymbol{hor\_ver_1}$ (1st variable for binary relation ***hor***), $\boldsymbol{hor\_ver_2}$ (2nd variable for binary relation ***hor***)

Jednotlivé obrázky z obr. 14 sú charakterizované reláciami uvedenými v nasledujúcej tabuľke.

riadok	špecifikácia
1	***ver***(***lg***(***x***),***lg***(***y***))
2	***hor***(***lg***(***x***),***lg***(***y***))
3	***hor***(***sm***(***x***),***lg***(***y***)) a ***hor***(***lg***(***x***),***sm***(***y***))
4	***ver***(***sm***(***x***),***lg***(***y***)) a ***ver***(***lg***(***x***),***sm***(***y***))
5	***ver***(***sm***(***x***),***sm***(***y***))
6	***hor***(***sm***(***x***),***sm***(***y***))



Obrázok 6.15. Množina 48 podobných obrázkov, ktoré obsahujú dva objekty, umiestnené buď vedľa seba alebo nad sebou a objekty sú buď malé alebo veľké. Každý stĺpec obsahuje dvojicu podobných objektov, ktoré sa líšia len svojim umiestnením alebo veľkosťou.

Holografické reprezentácie jednotlivých prípadov z tejto tabuľky majú tento tvar (porovnaj s rovnicou (43)).

$$
\begin{aligned}
\boldsymbol{t}_{1,\boldsymbol{x},\boldsymbol{y}} &= \boldsymbol{ver} + \left\langle \boldsymbol{ver\_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{ver\_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \\
\boldsymbol{t}_{2,\boldsymbol{x},\boldsymbol{y}} &= \boldsymbol{hor} + \left\langle \boldsymbol{hor\_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{hor\_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \\
\boldsymbol{t}_{3,\boldsymbol{x},\boldsymbol{y}} &= \begin{cases} \boldsymbol{ver} + \left\langle \boldsymbol{ver\_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{ver\_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\ \boldsymbol{ver} + \left\langle \boldsymbol{ver\_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{ver\_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \end{cases} \\
\boldsymbol{t}_{4,\boldsymbol{x},\boldsymbol{y}} &= \begin{cases} \boldsymbol{hor} + \left\langle \boldsymbol{hor\_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{hor\_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\ \boldsymbol{hor} + \left\langle \boldsymbol{hor\_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{hor\_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \end{cases} \\
\boldsymbol{t}_{5,\boldsymbol{x},\boldsymbol{y}} &= \boldsymbol{ver} + \left\langle \boldsymbol{ver\_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{ver\_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\
\boldsymbol{t}_{6,\boldsymbol{x},\boldsymbol{y}} &= \boldsymbol{hor} + \left\langle \boldsymbol{hor\_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{hor\_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle
\end{aligned}
\tag{48}
$$

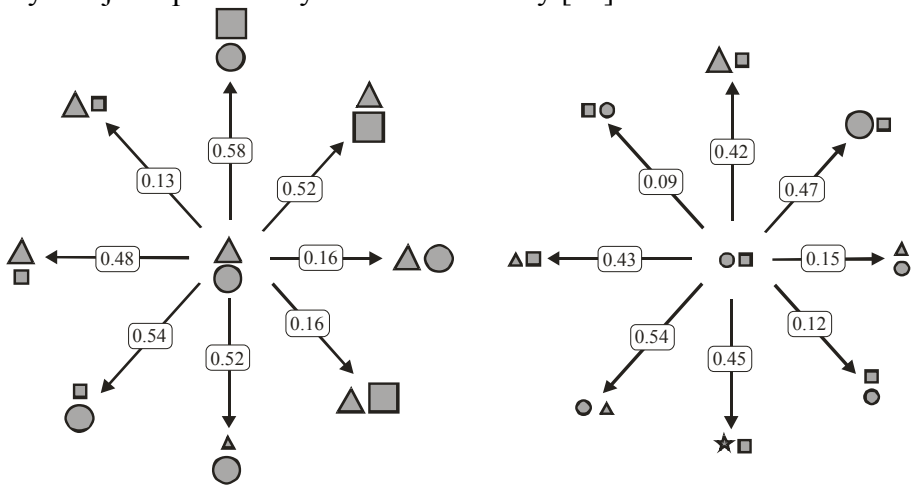
kde $\boldsymbol{x}$ a $\boldsymbol{y}$ sú holografické reprezentácia jednotlivých obrazcov (***tr***, ***sq***, ***ci***, ***st***) a zátvorka $\langle \boldsymbol{u} \rangle$ indikuje, že vektor $\boldsymbol{u}$ je normalizovaný. Podobnosť medzi jednotlivými obrázkami je určená prekryvom ich holografickými reprezentáciami

$$
\textit{podobnosť}\left(X, X'\right) = overlap\left(\boldsymbol{t}, \boldsymbol{t}'\right) \tag{48}
$$

Získané výsledky sú znázornené na obr. 15. Dominantným determinantom podobnosti je ich vertikálne alebo horizontálne usporiadanie. Prekryv (t.j. aj podobnosť) medzi dvoma obrázkami, ktoré majú rôzne usporiadanie je obvykle menší ako 0.1.

Vo všeobecnosti, holografická redukovaná reprezentácia umožňuje pomerne jednoduché stanovenie podobnosti medzi objektmi, ktoré sú špecifikované predikátovou štruktúrou (40) alebo jej zovšeobecnením pomocou ďalších vložených (nested) predikátov (pozri (46)) . Táto možnosť otvára úplne nové možnosti ako hľadať podobné objekty alebo

riešiť problém analógie, ktorý je štandardne považovaný za veľmi obtiažny problém umelej inteligencie vyžadujúci špeciálne symbolické techniky [xx].



Obrázok 16. Znázornenie podobných obrázkov pre dva vybrané obrázky 1 a 48 (pozri obr. 16). Jednotlivé šípky obsahujú prekryv medzi obrázkami spočítaný vzťahom (48).

Druhý ilustračný príklad – podobnosť medzi binárnymi číslami

Budeme študovať podobnosť medzi binárnymi číslami dĺžky 3, ktoré sú reprezentované reťazcom $(\alpha_1\alpha_2\alpha_3) \in \{0,1\}^3$. Toto číslo môžeme chápať ako usporiadanú trojicu binárnych symbolov, ktoré v distribuovanej reprezentácii sú reprezentované takto (pozri (33))

$$\boldsymbol{t}_{(\alpha_1\alpha_2\alpha_3)} = \boldsymbol{t}_{\alpha_1} + \boldsymbol{t}_{\alpha_1} \otimes \boldsymbol{t}_{\alpha_2} + \boldsymbol{t}_{\alpha_1} \otimes \boldsymbol{t}_{\alpha_2} \otimes \boldsymbol{t}_{\alpha_3} \tag{49a}$$

$$\boldsymbol{t}_{\alpha} = \begin{cases} \boldsymbol{nula} & (pre\ \alpha = 0) \\ \boldsymbol{jedna} & (pre\ \alpha = 1) \end{cases} \tag{49b}$$

kde ***nula*** a ***jedna*** sú distribuované reprezentácie čísiel '0' resp. '1'. Tak napríklad, binárne číslo (101) je reprezentované takto

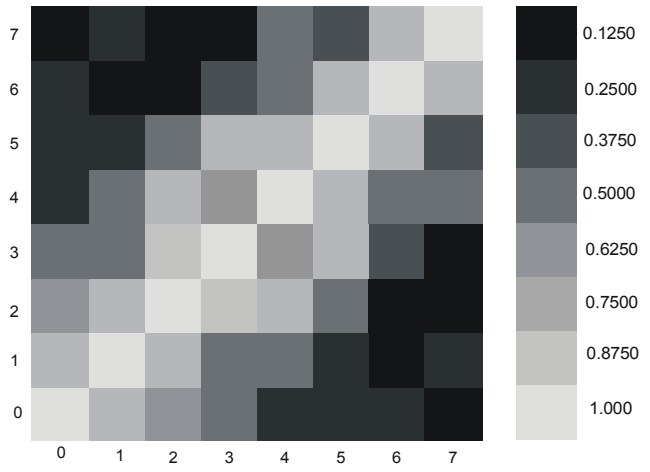
$$\boldsymbol{t}_{(101)} = \boldsymbol{jedna} + \boldsymbol{jedna} \otimes \boldsymbol{nula} + \boldsymbol{jedna} \otimes \boldsymbol{nula} \otimes \boldsymbol{jedna} \tag{50}$$

Podobnosť medzi jednotlivými reprezentáciami odráža aj podobnosť medzi príslušnými binárnymi číslami; reprezentácia dvoch blízkych binárnych čísel je nepriamo úmerná ich vzdialenosti (napr. absolútnej hodnote rozdielu). V nasledujúcej tabuľke sú uvedené podobnosti medzi reprezentáciami binárnych čísel, ktoré boli počítane pomocou ich prekryvu (23).

	000	001	010	011	100	101	110	111
000	1.00	0.74	0.50	0.46	0.21	0.16	0.14	0.11
001	0.74	1.00	0.71	0.42	0.45	0.15	0.12	0.15
010	0.50	0.71	1.00	0.75	0.70	0.43	0.09	0.08
011	0.46	0.42	0.75	1.00	0.46	0.70	0.37	0.04
100	0.21	0.45	0.70	0.46	1.00	0.74	0.43	0.39
101	0.16	0.15	0.43	0.70	0.74	1.00	0.70	0.35
110	0.14	0.12	0.09	0.37	0.43	0.70	1.00	0.69
111	0.11	0.15	0.08	0.04	0.39	0.35	0.69	1.00

Maximálna podobnosť je medzi dvojicami reprezentácií, ktoré sú priradené dvom susedným celým číslam, minimálna podobnosť medzi dvoma reprezentáciami existuje vtedy, ak

priradené čísla majú maximálnu vzdialenosť 7. Tieto údaje z tabuľky sú graficky znázornené na obr. 17.



Obrázok 17. Grafické znázornenie podobnosti medzi distribuovanými reprezentáciami binárnych čísel, veľkosť prekryvu medzi jednotlivými reprezentáciami je úmerná svetlosti odpovedajúcej štvorcovej oblasti (oblasti na vedľajšej diagonále sú najsvetlejšie). V pravom dolnom a v ľavom hornom rohu sú oblasti najtmavšie, čo odpovedá skutočnosti, že tieto oblasti sú priradené reprezentáciam maximálne vzdialených čísel.

Z tohoto jednoduchého ilustračného príkladu vyplýva, že v rámci holografickej distribuovanej reprezentácie môžeme navrhovať (aspoň potenciálne) asociačné reprezentácie typu (28), kde asociačné „narážky“ odpovedajú číslam. To znamená, že v tomto distribuovanom prístupe existuje možnosť asociačnej simulácie ľubovolnej funkcie, čo podstatne zvyšuje možnosti metódy smerom k jej univerzálnosti.

6.7 Uvažovanie pomocou modus ponens a modus tollens

Simulácia procesov uvažovania (inferencie) patrí medzi základné problémy, ktoré sa znovu a znovu riešia v umelej inteligencii a v kognitívnej vede [xx]. Fodorova kritika konekcionizmu [xx] bola práve založená na unáhlenom závere, že umelé neurónové siete nie sú schopné simulovať vyššie kognitívne aktivity, ktoré sú vraj výhradnou doménou klasického symbolického prístupu. Tento Fodorov názor sa ukázal ako nesrávny, ďalší rozvoj teórie neurónových sietí ukázal, že konekcionizmus je univerzálnym výpočtovým prostriedkom, ktorý nemá hranice aplikovateľnosti, neexistujú preň domény neaplikovateľnosti, ktoré by boli preň zakázané, samozrejme môže sa ukázať, že v niektorých doménach jeho použitie je neobyčajne ťažkopádne a zbytočne komplikované, že existujú iné prístupy, ktoré v danej doméne poskytujú podstatne jednoduchšie a priamočiarejšie riešenie, než ako umožňujú neurónové siete.

V tejto kapitole ukážeme na možnosť reprezentácie dvoch základných módov deduktívneho uvažovania výrokovej logiky,

$$\frac{p \Rightarrow q \quad p}{q} \qquad \text{a} \qquad \frac{p \Rightarrow q \quad \overline{q}}{\overline{p}} \tag{51}$$

ktoré sa nazývajú *modus ponens* resp. *modus tollens*. Tieto módy uvažovania sú ekvivalentné týmto tautológiam výrokovej logiky

$$\big((p \Rightarrow q) \wedge p\big) \Rightarrow q \tag{52a}$$

$$\left(\left(p \Rightarrow q\right) \wedge \overline{q}\right) \Rightarrow \overline{p} \tag{52b}$$

Implikáciu $'\Rightarrow'$ môžeme chápať ako binárnu reláciu, ktorú je možno holografickým distribuovaným spôsobom reprezentovať takto (pozri vzťah (45))

$$\boldsymbol{t}_{p\Rightarrow q} = \boldsymbol{op} \otimes \boldsymbol{impl} + \boldsymbol{var}_1 \otimes \boldsymbol{p} + \boldsymbol{var}_2 \otimes \boldsymbol{q} \tag{53}$$

ktorá obsahuje spčet troch častí, prvá časť špecifikuje typ relácie (implikáciu), druhá a tretia časť špecifikuje prvú (antedecedent) resp. druhú (konsekvent) premennú relácie implikácie. Tento konceptuálny vektor reprezentujúci reláciu implikácie môže byť transformovaný takto

$$\tilde{\boldsymbol{t}}_{p\Rightarrow q} = \boldsymbol{t}_{p\Rightarrow q} \otimes \boldsymbol{T} \tag{54a}$$

kde

$$\boldsymbol{T} = \boldsymbol{var}_1^* \otimes \boldsymbol{p}^* \otimes \boldsymbol{p}^* \otimes \boldsymbol{q} + \boldsymbol{var}_2^* \otimes \boldsymbol{q}^* \otimes \overline{\boldsymbol{q}}^* \otimes \overline{\boldsymbol{p}} \tag{54b}$$

Transformovaná reprezentácie implikácie je reprezentuje súčet dvoch asociačných dvojíc

$$\tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \boldsymbol{p}^* \otimes \boldsymbol{q} + \overline{\boldsymbol{q}}^* \otimes \overline{\boldsymbol{p}} \tag{55}$$

ktorá poskytuje holografickú reprezentáciu pravidiel *modus ponens* a *modus tollens*

$$\boldsymbol{p} \otimes \tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \boldsymbol{q} \tag{56a}$$

$$\overline{\boldsymbol{q}} \otimes \tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \overline{\boldsymbol{p}} \tag{56b}$$

Prvý vzťah (56a) môžeme chápať ako holografickú reprezentáciu modus ponens (pozri (51) a (52a)), zatiaľ čo druhý vzťah je holografická reprezentácia modus tollens (pozri (51) a (52b)).

Podobný výsledok môže byť získaný aj iným alternatívnym spôsobom, ktorý je založený na disjunktnom tvare implikácie

$$\left(p \Rightarrow q\right) \equiv \left(\overline{p} \vee q\right) \tag{57}$$

Distribuovaná reprezentácia implikácie v tomto alternatívnom tvare ma túto podobu

$$\boldsymbol{t}_{\overline{p}\vee q} = \boldsymbol{op} \otimes \boldsymbol{disj} + \boldsymbol{var}_1 \otimes \overline{\boldsymbol{p}} + \boldsymbol{var}_2 \otimes \boldsymbol{q} \tag{58}$$

Transformáciou tejto reprezentácie dostaneme (porovnaj (53))

$$\tilde{\boldsymbol{t}}_{\overline{p}\vee q} = \boldsymbol{t}_{\overline{p}\vee q} \otimes \boldsymbol{T} \approx \overline{\boldsymbol{p}} \otimes \overline{\boldsymbol{q}}^* + \boldsymbol{q} \otimes \boldsymbol{p}^* \tag{59a}$$

kde

$$\boldsymbol{T} = \boldsymbol{var}_1^* \overline{\boldsymbol{q}}^* + \boldsymbol{var}_2^* \boldsymbol{p}^* \tag{59a}$$

Táto transformácia je podstatne jednoduchšia než ako v predchádzajúcom prípade (54b). Pravidlá modus ponens a mosus tollens sú teraz realizované vzťahmi podobnými s (56a-b). Naviac dostávame ešte tieto dve „pravidlá“

$$\boldsymbol{q}^* \otimes \tilde{\boldsymbol{t}}_{\overline{p}\vee q} \approx \boldsymbol{p}^* \tag{60a}$$

$$\overline{\boldsymbol{p}}^* \otimes \tilde{\boldsymbol{t}}_{\overline{p}\vee q} \approx \overline{\boldsymbol{q}}^* \tag{60b}$$

ktoré pripomínajú dobre známe pravidlá chybného uvažovania

$$\begin{array}{l} p \Rightarrow q \\ q \\ \hline p \end{array} \quad \text{a} \quad \begin{array}{l} p \Rightarrow q \\ \overline{p} \\ \hline \overline{q} \end{array} \tag{61}$$

ktoré sa nazývajú „*potvrdenie dôsledku*“ resp. „*poprenie predpokladu*“ . Táto skutočnosť je spôsobená tým, že transformované reprezentácie implikácie $\tilde{\boldsymbol{t}}_{p\Rightarrow q}$ a $\tilde{\boldsymbol{t}}_{\overline{p}\vee q}$ nie sú identické, reprezentácia $\tilde{\boldsymbol{t}}_{\overline{p}\vee q}$ vedie práve k neočakávaným výsledkom (61), ktoré reprezentujú nesprávne módy usudzovania (ktoré sú však často, aj napriek tomu, často používané ľuďmi nepoznajúcimi zásady logiky).

6.8 Predikátová logika

Budeme sa zaoberať jednoduchou formou predikátovej logiky, ktorá je založená na unárnych predikátoch, $P(x)$, ktorých distribuovaná reprezentácia má tvar (pozri kapitolu 6)

$$\boldsymbol{t}_{P(x)} = \boldsymbol{pred} \otimes \boldsymbol{P} + \boldsymbol{pred\_var} \otimes \boldsymbol{x} \tag{62}$$

Spojenie tohoto predikátu s univerzálnym kvantifikátorom, $(\forall x)P(x)$, môže byť reprezentované nasledujúcim spôsobom

$$\boldsymbol{t}_{(\forall x)} = \boldsymbol{uni\_quant} \otimes \boldsymbol{uni} + \boldsymbol{uni\_quant\_var} \otimes \boldsymbol{x} \tag{63a}$$

$$\boldsymbol{t}_{(\forall x)P(x)} = \boldsymbol{t}_{(\forall x)} + \boldsymbol{t}_{(\forall x)} \otimes \left(\boldsymbol{pred} \otimes \boldsymbol{P} + \boldsymbol{pred\_var} \otimes \boldsymbol{x} \right) \tag{63b}$$

Oba konceptuálne vektory $\boldsymbol{t}_{P(x)}$ a $\boldsymbol{t}_{(\forall x)P(x)}$ môžu byť rekognoskované pomocou „čistiacej" procedúry.

Pre naše ďalšie účely štúdia procesov uvažovania v rámci predikátovej logiky a ich distribuovanej reprezentácie je tento prístup zbytočne zložitý, používanie konceptuálneho vektora $\boldsymbol{t}_{(\forall x)}$ pre reprezentáciu symbolu $(\forall x)$ v podstate len zbytočne komplikuje proces rekognoskácie zložených konceptuálnych vektorov obsahujúcich ako zložku $\boldsymbol{t}_{(\forall x)}$. Preto opustíme od explicitného používania symbolu $(\forall x)$, jeho význam bude substituovaný používaním „*univerzálnej premennej*" x, t.j. predikát $P(x)$ obsahujúci univerzálnu premennú x sa interpretuje ako $(\forall x)P(x)$, môžeme teda s určitou obozretnosťou používať „formulu" $(\forall x)P(x) \equiv P(x)$.

V predikátovej logike platí pravidlo univerzálneho odlúčenia, pomocou ktorého sa konkretizuje predikát s univerzálnym kvantifikátorom na predikát s konkrétnou premennou *a,* $(\forall x)P(x) \Rightarrow P(a)$, ktorý je dôsledkom jednoduchej tautológie výrokovej logiky $\big((p \wedge q) \wedge p\big) \Rightarrow q$. Použitím konvencie univerzálnej premennej x túto konkretizáciu prepíšeme do jednoduchšieho tvaru

$$P(x) \Rightarrow P(a) \tag{64}$$

Zostrojíme distribuovanú reprezentáciu tejto univerzálnej eliminácie jednoduchého unárneho replikátora pomocou transformačného vektora $\boldsymbol{T}$, pozri formule (54-56). Distribuovaná reprezentácia (64) má tvar

$$\boldsymbol{t}_{P(a)} \approx \boldsymbol{t}_{P(x)} \otimes \boldsymbol{T} \tag{65}$$

kde $\boldsymbol{T} = \boldsymbol{x}^* \otimes \boldsymbol{a}$, potom

$$\boldsymbol{t}_{P(a)} \approx \boldsymbol{t}_{P(x)} \otimes \boldsymbol{x}^* \otimes \boldsymbol{a} \tag{66}$$

Sumarizujúc naše úvahy, distribuovaná reprezentácia pravidla univerzálnej eliminácie (konkretizácie) sa uskutočňuje pomocou transformačného vektora $\boldsymbol{T}$, pomocou ktorého sa univerzálna premenná $\boldsymbol{x}$ substituuje „konkrétnou" premennou a.

Pomocou tohto zjednodušenej reprezentácie kvantifikovaných predikátov budeme študovať tzv. *zovšeobecnený modus ponens* a *zovšeobecnený modus tollens*

$$\frac{(\forall x)\big(P(x) \Rightarrow Q(x)\big) \quad P(a)}{Q(a)} \quad \text{a} \quad \frac{(\forall x)\big(P(x) \Rightarrow Q(x)\big) \quad \overline{Q}(a)}{\overline{P}(a)} \tag{67a}$$

alebo v zjednodušenej podobe pomocou univerzálnej premennej x

$$\frac{P(x)\Rightarrow Q(x)}{\underline{P(a)}\qquad}\quad\text{a}\quad\frac{P(x)\Rightarrow Q(x)}{\underline{\bar{Q}(a)}\qquad}\tag{67b}$$
$$Q(a)\qquad\qquad\bar{P}(a)$$

Tieto zovšeobecnené schémy deduktívneho uvažovania vyplývajú priamo z ich štandardnej výrokovej formy (51) a konkretizácie (61). Distribuovaná reprezentácia hlavnej (vrchnej) premisy týchto pravidieľ má tvar

$$\begin{aligned}\boldsymbol{t}_{P(x)\Rightarrow Q(x)} = \boldsymbol{op}\otimes\boldsymbol{impl} + \boldsymbol{var}_1\otimes\left(\boldsymbol{pred}\otimes\boldsymbol{P}+\boldsymbol{pred\_var}\otimes\boldsymbol{x}\right)+\\ \boldsymbol{var}_2\otimes\left(\boldsymbol{pred}\otimes\boldsymbol{Q}+\boldsymbol{pred\_var}\otimes\boldsymbol{y}\right)\end{aligned}\tag{68}$$

Konkretizácia implikácie $P(x)\Rightarrow Q(x)$ na $P(a)\Rightarrow Q(b)$, je formálne vyjadriteľná pomocou implikácie (pozri (64))

$$\left(P(x)\Rightarrow Q(x)\right)\Rightarrow\left(P(a)\Rightarrow Q(a)\right)\tag{69}$$

kde práva strana má nasledujúcu distribuovanú reprezentáciu

$$\boldsymbol{t}_{P(a)\Rightarrow Q(a)} = \boldsymbol{op}\otimes\boldsymbol{impl}+\boldsymbol{var}_1\otimes\left(\boldsymbol{P}+\boldsymbol{a}\right)+\boldsymbol{var}_2\otimes\left(\boldsymbol{Q}+\boldsymbol{b}\right)\tag{70}$$

Podobne ako v úvodnom ilustračnom príklade (pozri (65)), tento prechod vyjadrený implikáciou (69), môže byť v distribuovanej reprezentácii vyjadrený pomocou transformácie

$$\tilde{\boldsymbol{t}}_{P(a)\Rightarrow Q(a)} = \boldsymbol{t}_{P(a)\Rightarrow Q(a)}\otimes\boldsymbol{T}\approx\left(\boldsymbol{P}+\boldsymbol{a}\right)^*\otimes\left(\boldsymbol{Q}+\boldsymbol{b}\right)+\left(\bar{\boldsymbol{Q}}+\boldsymbol{b}\right)^*\otimes\left(\bar{\boldsymbol{P}}+\boldsymbol{a}\right)\tag{71}$$

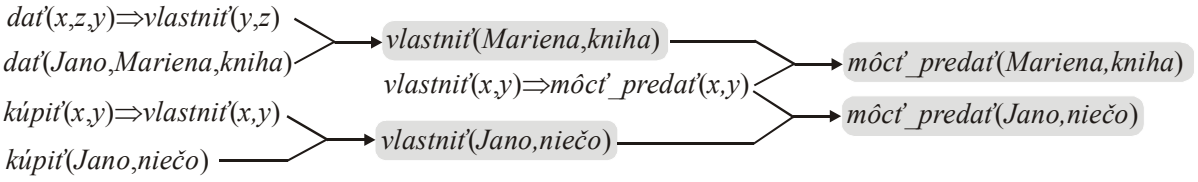
kde nová transformovaná distribuovaná reprezentácia $\tilde{\boldsymbol{t}}_{P(a)\Rightarrow Q(a)}$ vyhovuje vzťahom, ktoré reprezentujú pravidlá (67) modus ponens resp. modus tollens

$$\left(\boldsymbol{P}+\boldsymbol{a}\right)\otimes\tilde{\boldsymbol{t}}_{P(a)\Rightarrow Q(a)}\approx\left(\boldsymbol{Q}+\boldsymbol{b}\right)\tag{72a}$$

$$\left(\bar{\boldsymbol{Q}}+\boldsymbol{b}\right)\otimes\tilde{\boldsymbol{t}}_{P(a)\Rightarrow Q(a)}\approx\left(\bar{\boldsymbol{P}}+\boldsymbol{a}\right)\tag{72b}$$

Ilustračný príklad – modelovanie reflexívneho uvažovania

V tomto ilustračnom príklade ukážeme, že holografická distribuovaná reprezentácia poskytuje formálne prostriedky, pomocou ktorého môžeme simulovať proces usudzovania založeného na zovšeobecnenom modus ponens (67). Tento proces bol široko študovaný Shastrim a Ajjanagaddom [xx] pomocou konekcionistického systému nazývaného SHRUTI, ktorý bol schopný simulovať reflexívne uvažovanie založenom na predikátovej logike. Podobné výsledky budú dosiahnuté aj pomocou formalizmu holografickej distribuovanej reprezentácie.



Obrázok 18. Znázornenie použitia zovšeobecneného pravidla modus ponens (64) pre vyvodenie poznatkov (označených tmavým podkladom a taktiež aj výstupnými šípkami) z implikácií (1-3) a zo vstupných pozorovaní (a-c), ktoré sú označené vstupnými šípkami.

Majme formálny systém obsahujúcou tieto tri všeobecné pravidlá (pozri obr.(18))

(1) *dať*$\left(x,y,z\right)\Rightarrow$ *vlastniť*$\left(y,z\right)$, *typ x* : darca; *typ y* : príjemca; *typ z* : vec,

(2) *kúpiť*$\left(y,z\right)\Rightarrow$ *vlastniť*$\left(y,z\right)$, *type y* : kupujúci; *typ z* : vec,

(3) $vlastniť(y,z) \Rightarrow môcť\_predať(y,z)$, *typ y* : vlastník; *typ z* : vec,

a tri pozorovania (skutočnosti – fakty)

(a) $dať(Jano, Mariena, kniha)$,

(b) $vlastniť(Mariena, kniha)$

(c) $kúpiť(Jano, niečo)$.

Aké sú deduktívne závery z tohto systému? Výsledky sú naznačené na obrázku 15, teraz pristúpime k ich odvodeniu pomocou distribuovanej reprezentácie založenej na konceptuálnych vektoroch a operáciách nad nimi.

Sledujme prvý zovšeobecnený modus ponens z obr. 16

$$\frac{dať(x,y,z) \Rightarrow vlastniť(y,z) \qquad dať(Jano, Mariena, kniha)}{vlastniť(Mariena, kniha)} \tag{73}$$

Pomocou postupu (68-72) môžeme realizovať túto schému uvažovania pomocou reprezentácie konceptuálnych vektorov, jej jednotlivé zložky (idúc zhora nadol) sú reprezentované takto

$$\boldsymbol{t}_1^{(1)} = \boldsymbol{op} \otimes \boldsymbol{impl} + \boldsymbol{var}_1 \otimes \boldsymbol{t}_{dať(x,y,x)} + \boldsymbol{var}_2 \otimes \boldsymbol{t}_{vlastniť(y,x)} \tag{74a}$$

$$\boldsymbol{t}_2^{(1)} = \boldsymbol{t}_{dať(Jano,Mariena,kniha)} = \boldsymbol{dať} + \boldsymbol{dať\_var}_1 \otimes \boldsymbol{Jano} + \boldsymbol{dať\_var}_2 \otimes \boldsymbol{Mariena} + \boldsymbol{dať\_var}_3 \otimes \boldsymbol{kniha} \tag{74b}$$

$$\boldsymbol{t}_3^{(1)} = t_{vlastniť(Mariena,kniha)} = \boldsymbol{vlastniť} + \boldsymbol{vlastniť\_var}_1 \otimes \boldsymbol{Mariena} + \boldsymbol{vlastniť\_var}_2 \otimes \boldsymbol{kniha} \tag{74c}$$

kde konceptuálne vektory $\boldsymbol{t}_{dať(x,y,x)}$ a $\boldsymbol{t}_{vlastniť(y,x)}$ sú zostrojené analogickým spôsobom ako v (43). V prvom kroku musíme vykonať konkrétizáciu implikácie $dať(x,y,z) \Rightarrow vlastniť(y,z)$, tak že všeobecné premenné *x*, *y*, *z* sú substituované konkrétnymi premennými *Jano*, *Mariena*, *kniha*. Táto konkretizácia sa vykoná pomocou transformácie $\boldsymbol{T}$, ktorá je špecifikovaná formulami (65-66)

$$\hat{\boldsymbol{t}}_1^{(1)} \approx \boldsymbol{t}_1^{(1)} \otimes \boldsymbol{T} \tag{75}$$

kde

$$\hat{\boldsymbol{t}}_1^{(1)} = \boldsymbol{op} \otimes \boldsymbol{impl} + \boldsymbol{var}_1 \otimes \boldsymbol{t}_{dať(Jano,Mariena,kniha)} + \boldsymbol{var}_2 \otimes \boldsymbol{t}_{vlastniť(Mariena,kniha)} \tag{76a}$$

$$\boldsymbol{T} = \boldsymbol{t}_1^{*(1)} \otimes \hat{\boldsymbol{t}}_1^{(1)} \tag{76b}$$

Takto konkrétizovaná reprezentácia $dať(Jano, Mariena, kniha) \Rightarrow vlastniť(Mariena, kniha)$ v následnom kroku je použiteľná pre modus ponens uskutočnenej pomocou formúl (71-72)

$$\tilde{\boldsymbol{t}}_1^{(1)} \approx \hat{\boldsymbol{t}}_1^{(1)} \otimes \boldsymbol{T}' \tag{77a}$$

kde rezultujúci konceptuálny vektor $\tilde{\boldsymbol{t}}_1^{(1)}$ už reprezentuje modus ponens, t.j. platí

$$\boldsymbol{t}_{dať(Jano,Mariena,kniha)} \otimes \tilde{\boldsymbol{t}}_1^{(1)} \approx \boldsymbol{t}_{vlastniť(Mariena,kniha)} \tag{78}$$

Podobným spôsobom môžu byť uskutočniteľné aj ostatné tri zovšeobecnené modus ponens z obr. 16.

Ilustračný príklad –generalizácie pomocou indukcie

Uvažujme „tréningovú“ množinu tvorenú postupnosťou jednoduchých unárnych predikátorv, ktoré môžeme interpretovať ako pozorovania

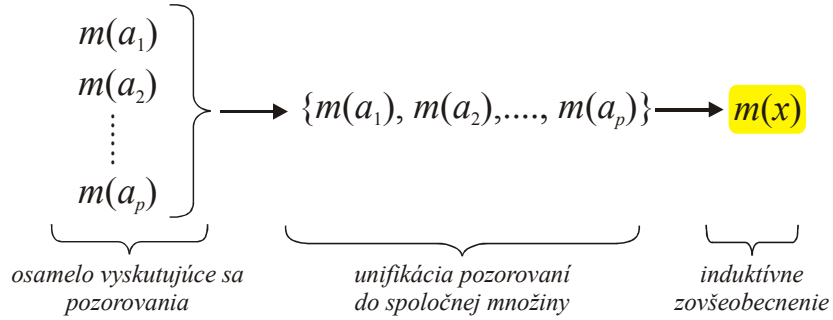
$$A_{train} = \{m(a_i)\,;\, i = 1,2,...,q\} \tag{79}$$

Našim cieľom bude zovšeobecniť tieto partikulárne predikáty do tvaru s univerzálnym kvantifikátorom (alebo v našom jedenoduchšom formalizme, s univerzálnou premennou)

$$(\forall x)[m(x)] \equiv m(x) \tag{80}$$

To znamená, že jednotlivé partikulárne prípady $m(a_i)$ z tréningovej množiny (75) sú zovšeobecnené do formule, ktorá nie je bezprostrerne odvoditeľná z nich (pozri obr. 19)

$$m(a_1) \wedge ... \wedge m(a_p) \overset{?}{\Rightarrow} m(x) \tag{81}$$



Obrázok 19. Znázornenie induktívneho zovšeobecnenia, ktoré obsahuje tri etapy. V prvej etape máme izolované pozorovania $m(a_i)$, ktoré medzi sebou nesúvisia. V druhej etape izolovanosť jednotlivých pozorovaní je nahradená ich unifikáciou do jednej množiny, čo vyjadruje skutočnosť, že pozorovania nie sú izolované a nezávislé, ale majú niečo spoločné. V záverečnej tretej etape, unifikovaná forma pozorovaní je induktívne zovšeobecnená pomocou univerzálnej premennej *x*, ktorá reprezentuje triedu objektov majúcich rovnakú vlastnosť *m*.

Nech predikát $m(a_i)$ z tréningovej množiny A_{train} je holograficky reprezentovaný takto

$$\boldsymbol{t}_{m(a_i)} = \boldsymbol{rel} \otimes \boldsymbol{m} + \boldsymbol{m\_var} \otimes \boldsymbol{a}_i \tag{82}$$

Tieto reprezentačné konceptuálne vektory môžu byť navzájom transformované pomocou transformačných vektorov $T_{i,j}$

$$\boldsymbol{t}_{m(a_j)} \approx \boldsymbol{t}_{m(a_i)} \otimes \boldsymbol{T}_{i,j} \tag{83a}$$

$$\boldsymbol{T}_{i,j} = \boldsymbol{t}^*_{m(a_i)} \otimes \boldsymbol{t}_{m(a_j)} \approx \boldsymbol{a}^*_i \otimes \boldsymbol{a}_j \tag{83b}$$

Definujme nový transformačný vektor $\tilde{\boldsymbol{T}}_i$

$$\tilde{\boldsymbol{T}}_i = \sum_{j=1}^{q} \boldsymbol{T}_{i,j} = \boldsymbol{a}^*_i \otimes \boldsymbol{x} \tag{84a}$$

$$\boldsymbol{x} = \boldsymbol{a}_1 + \boldsymbol{a}_2 + ... + \boldsymbol{a}_p \tag{84b}$$

kde konceptuálny vektor $\boldsymbol{x}$ je priradený novému ”zovšeobecnenému” vektoru, ktorý reprezentuje každý argument a_i z tréningovej množiny A_{train} (ktorý je v našich ďalších úvahách o indukcii chápaný ako samostatná konceptuálna entita rovnocenná s reprezentáciami pôvodných objektov $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_p$)

$$\boldsymbol{t}_{m(a_i)} \otimes \tilde{\boldsymbol{T}}_i \approx \boldsymbol{t}_{m(\boldsymbol{x})} \tag{85}$$

Táto formula môže byť interpretovaná ako indukčné zovšeobecnenie, kde partikulárne objekty a_i z tréningovej množiny v reprezentáciách $\boldsymbol{t}_{m(a_i)}$ sú substituované novým objektom x, ktorý môže byť interpretovaný ako nový univerzálny objekt.

6.9 Formálne jazyky

Vo štvrtej kapitole sme študovali použitie holografickej distribuovanej reprezentácie k štúdiu postupnosti symbolov, ktoré môžeme chápať ako najjednoduchší typ štruktúrovanej informácie. Majme konečnú abecedu $\mathcal{A} = \{a, b, ..., p, q, ...\}$, ktorý obsahuje malé písmena a ktoré sú nazývané terminálne symboly. Množina všetkých možných reťazcov zostrojených nad touto abecedou obsahuje je označená $L = \mathcal{A}^+ = \{a, b, ..., aa, ab, ..., aaa, aab, ..., ppp, ...\}$. Táto nekonečná (ale spočitateľná) množina môže byť rozdelená na dve disjunktné podmnožiny, $L = L_G \cup L_{NG}$, pričom stojíme pred problémom ako jednoducho charakterizovať reťazce z jednotlivých podmnožín. Formálne riešenie tohoto problému spočíva v definovaní *charakteristickej funkcie*, ktorá špecifikuje elementy podmnožín takto

$$\eta_G(x) = \begin{cases} 1 & (\text{ak } x \in L_G) \\ 0 & (\text{ak } x \in L_{NG}) \end{cases} \tag{88}$$

potom $L_G = \{x \mid x \in L \wedge \eta_G(x) = 1\}$. Tento prístup špecifikuje reťazce z podmnožiny L_G pomocou ich enumerácie, t.j. taxatívnym vymenovaním, či reťazec $x \in L$ patrí do podmnožiny L_G alebo do L_{NG}.

Predpokladajme, že reťazce z L_G sú hierarchicky klastrované na dve alebo viac zložiek, ktoré naviac môžu byť do seba vnorené, pričom na záver tohoto procesu klastrovania získame jeden globálny klaster. Tento proces môže byť diagramaticky vyjadrený mnohými spôsobmi (pozri obr. 20, kde tri jednoduché spôsoby klastrovania sú znázornené). Prístup koreňového stromu z obrázku patrí medzi najjednoduchšie spôsoby formalizovania klastrovaci proces reťazca na hierarchicky organizovanú štruktúru, jeho formalizácia môže byť vykonaná pomocou tzv. prepisovacích praviedieľ. Tak napríklad, ilustračný príklad z prvého riadku obr. 20 tieto pravidlá majú tvar

$$\begin{aligned} P_1 &: S \to BC \\ P_2 &: B \to Ac \\ P_3 &: C \to ab \\ P_4 &: A \to ab \end{aligned} \tag{89}$$

kde veľké písmena sa nazývajú neterminálne symboly (volne povedané, tieto symboly reprezentujú „podklastre“, ktoré sú súčasťou finálneho globálneho klastra). Písmeno S sa nazýva „štartovací“ neterminál, je priradený hierarchicky najvyššiemu klastru, ktorý zahrňuje všetky terminálne symboly z daného reťazca). Použitím pravidiel z (89) môžeme postupne vytvoriť reťazec $x = (abcab)$ takto

$$S \to BC \begin{cases} \to AcC \begin{cases} \to abcC \\ \to Acab \end{cases} \\ \to Bab \ \to \ Acab \end{cases} \to abcab \tag{90}$$

We see that there exist three different ways how to construct the string $x = (abcab)$.

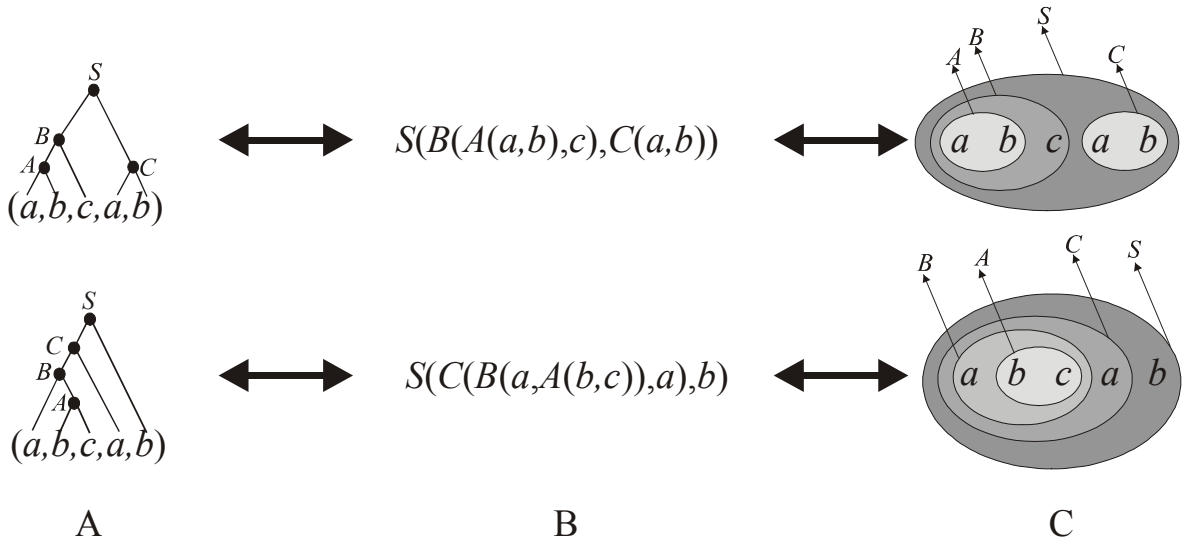


Figure 20. Two illustrative conversions of a sequence (*a,b,c,a,b*) into different „nested“ structures: (A) Rooted tree structure that aggregates simple sequence elements into higher-level hierarchic structures, (B) composite two argument functions, and (C) nested subsets encompass one or more symbols such that the resulting structure correspond to an aggregation of symbols.

The production system (89) may be considerably simplified such that we removed all nonterminal symbols that are redundant (in this specific case we may remove the nonterminal *C*, which is identical to *A*.

$$\begin{aligned} P_1 &: \; S \to BA \\ P_2 &: \; B \to Ac \\ P_3 &: \; A \to ab \end{aligned} \tag{91}$$

Moreover, the production system may be extended about one or more productions such that the construction of the original string $x = (abcab)$ is saved, but we may potentially construct many other new different strings $x \in L$, e.g.

$$\left.\begin{aligned} P_1 &: \; S \to BA \\ P_2 &: \; B \to Ac \\ P_3 &: \; A \to ab \\ P_4' &: \; A \to Aa \end{aligned}\right\} P_{34} : \; A \to ab \mid Ab \tag{92}$$

where new production $P_4' : \; A \to Aa$ introduces a generative feature of the resulting used scheme.

The above considerations may be generalized onto a concept called the *grammar* [xx] (initially introduced by Noam Chomsky [20] almost 50 years ago)specified as follows

$$G = (N, T, S, P) \tag{93}$$

where N is a set composed of nonterminal symbols (in our above preliminary considerations represented by capital letters), $T \subset \mathcal{A}$ is a set composed of terminal symbols (represented by lower case letters), $S \in N$ is a *starting symbol* especially distinguished from elements of nonterminals *N*, and finally, *P* is a set composed of the so-called *productions* (often called also production rules or rewriting rules)

$$P = \{\alpha \to \beta\} \tag{94}$$

where α is a substring composed of terminal and at least one nonterminal symbols and β is a substring composed of terminal or nonterminal symbols. In particular, the set *P* contains a

production of the form $S \to \beta$, this production will initialized the procedure of creation of strings that are induced by the grammar *G*. These strings, initialized by the starting symbol *S* and composed entirely of terminal symbols form a subset denoted by L_G, which is called the *language*. It means that the characteristic function (88) is now unambiguously specified by the production set (94) such its elements are (1) composed only of terminal symbols and (2) its creation procedure is initialized the nonterminal starting symbol *S*.

We have seen that each grammatical string $x \in L_G$ may be interpreted by a rooted tree $t(x)$ called the *derivation tree*, where a top vertex is labeled by the stating symbol *S*, its leafs (terminal vertices) are labeled by terminal symbols, and all other remaining intermediate vertices are labeled by nonterminal symbols (cf. Fig. 20). Derivation trees have an extraordinary position in artificial intelligence and cognitive science [21,22], a semantic meaning of the grammatical string $x \in L_G$ is specified not only by the string itself but also by its derivation tree $t(x)$. If a string has two or more derivation trees, then we may say that it has more than one different semantic meanings. An actual semantic meaning of $x \in L_G$ will depend on the used derivation tree $t(x)$, which specifies an actual chunking of string symbols onto hierarchically higher clusters that contribute to the whole semantic meaning by a specific contribution.

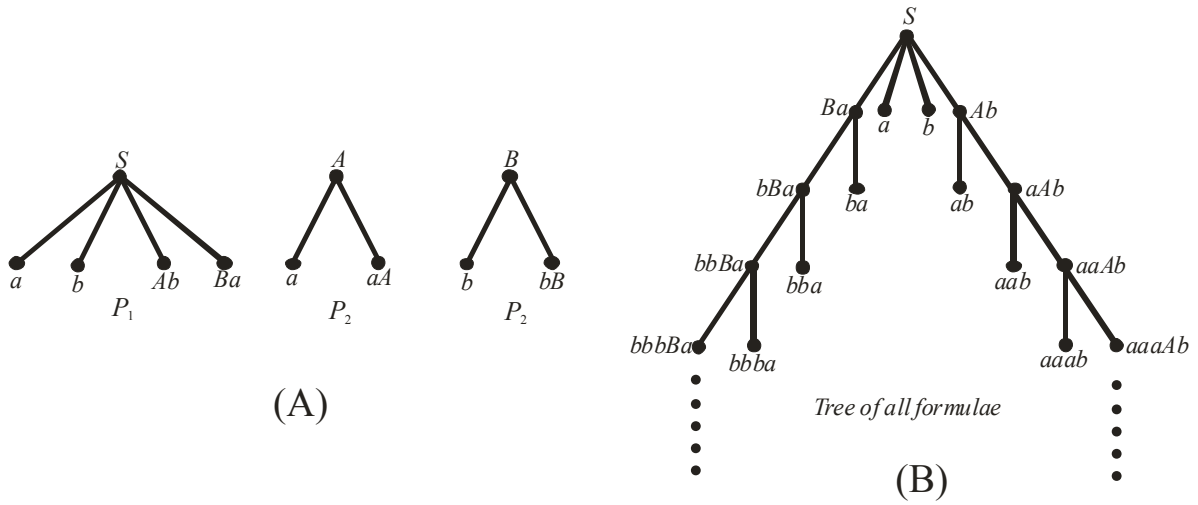


Figure 21. (A) Diagrammatic representation of productions (1).. (B) An infinite tree of all possible solutions – formulae.

Let is consider the following simple grammar $G = \left(N = \{S, A, B\}, T = \{a, b\}, S, P\right)$ with productions specified by

$$\begin{aligned} P_1 &: S \to a \mid b \mid Ab \mid Ba \\ P_2 &: A \to a \mid aA \\ P_3 &: B \to b \mid bB \end{aligned} \tag{95}$$

with productions diagrammatically visualized by These productions are diagrammatically visualized in Fig. 21, diagram A.

From fig. 21 we may construct all possible formulae that are specified by the productions (95)

$$\begin{aligned} L_G &= \left\{b^n a; n \geq 0\right\} \cup \left\{a^n b; n \geq 0\right\} = \\ &= \{a, b, ab, ba, bba, aab, bbba, aaab, \ldots\} \end{aligned} \tag{96}$$

For each grammatical formula there exists on tree of all formulae an oriented path beginning in the root of tree (labeled by a starting nonterminal S) and ending at a leaf evaluated by the respective formula, e.g.

$$\begin{aligned} a \in L_G &\Rightarrow S \to a \\ b \in L_G &\Rightarrow S \to b \\ ab \in L_G &\Rightarrow S \to Ab \to ab \\ ba \in L_G &\Rightarrow S \to Ba \to ba \\ aab \in L_G &\Rightarrow S \to Ab \to aAb \to aab \\ bba \in L_G &\Rightarrow S \to Ba \to bBa \to bba \end{aligned} \tag{97}$$

The corresponding derivation trees are listed in Fig. 22.

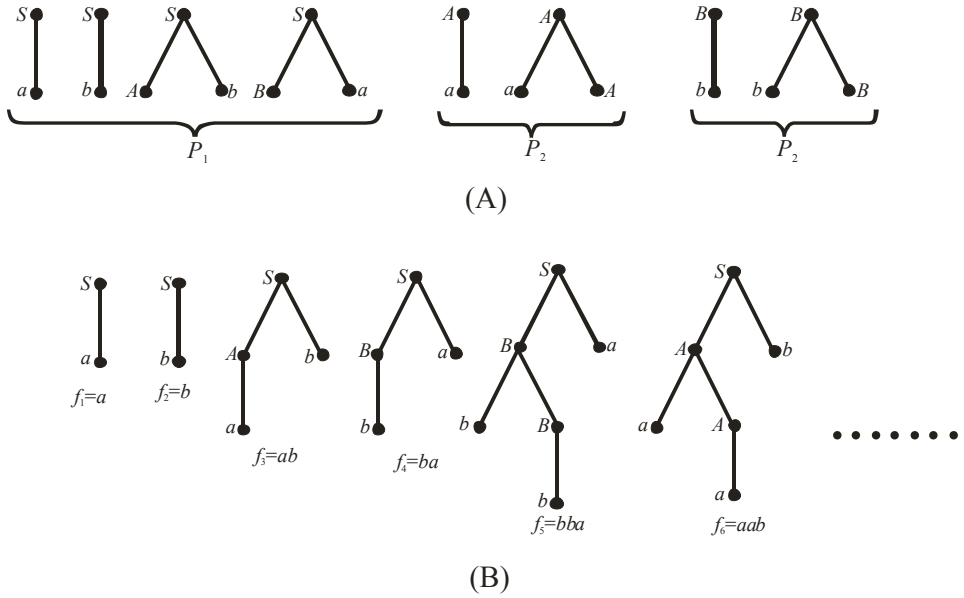


Figure 22. (A) Detailed diagrammatic representation of productions (1). (B) Derivation trees assigned to the first six formulae.

HRR representations of derivation trees in Fig. 22 are (cf. (44))

$$\begin{aligned} \boldsymbol{t}_1 &= \boldsymbol{left} \otimes \boldsymbol{a} \\ \boldsymbol{t}_2 &= \boldsymbol{left} \otimes \boldsymbol{b} \\ \boldsymbol{t}_3 &= \boldsymbol{left} \otimes \left(\boldsymbol{A} + \boldsymbol{left} \otimes \boldsymbol{a} \right) + \boldsymbol{right} \otimes \boldsymbol{b} \\ \boldsymbol{t}_4 &= \boldsymbol{left} \otimes \left(\boldsymbol{B} + \boldsymbol{left} \otimes \boldsymbol{b} \right) + \boldsymbol{right} \otimes \boldsymbol{a} \\ \boldsymbol{t}_5 &= \boldsymbol{left} \otimes \left(\boldsymbol{B} + \boldsymbol{left} \otimes \boldsymbol{b} + \boldsymbol{right} \otimes \left(\boldsymbol{B} + \boldsymbol{left} \otimes \boldsymbol{b} \right) \right) + \boldsymbol{right} \otimes \boldsymbol{a} \\ \boldsymbol{t}_6 &= \boldsymbol{left} \otimes \left(\boldsymbol{A} + \boldsymbol{left} \otimes \boldsymbol{a} + \boldsymbol{right} \otimes \left(\boldsymbol{A} + \boldsymbol{left} \otimes \boldsymbol{a} \right) \right) + \boldsymbol{right} \otimes \boldsymbol{b} \end{aligned} \tag{98}$$

where we have omitted the vector representation $\boldsymbol{S}$ of the nonterminal starting symbol, which appears as the first in each representation of derivation trees.

HRR representation of the first grammatical strings (96) are (cf. (34))

$$
\begin{aligned}
\boldsymbol{f}_1 &= \boldsymbol{first} \otimes \boldsymbol{a} \\
\boldsymbol{f}_2 &= \boldsymbol{first} \otimes \boldsymbol{b} \\
\boldsymbol{f}_3 &= \boldsymbol{first} \otimes \boldsymbol{a} + \boldsymbol{second} \otimes \boldsymbol{b} \\
\boldsymbol{f}_4 &= \boldsymbol{first} \otimes \boldsymbol{b} + \boldsymbol{second} \otimes \boldsymbol{a} \\
\boldsymbol{f}_5 &= \boldsymbol{first} \otimes \boldsymbol{b} + \boldsymbol{second} \otimes \boldsymbol{b} + \boldsymbol{third} \otimes \boldsymbol{a} \\
\boldsymbol{f}_6 &= \boldsymbol{first} \otimes \boldsymbol{a} + \boldsymbol{second} \otimes \boldsymbol{a} + \boldsymbol{third} \otimes \boldsymbol{b}
\end{aligned}
\tag{99}
$$

The used holographic representations of derivation trees and string of symbols have a serious shortcoming; different $\boldsymbol{t}$'s have great similarities between them (the same property is also true for vectors $\boldsymbol{t}$'s), see Fig. 23. This fact causes that an immediate application of (28) to form an analogy of associative memory vector $\boldsymbol{T} = \boldsymbol{t}_1 \otimes \boldsymbol{f}_1 + ... + \boldsymbol{t}_6 \otimes \boldsymbol{f}_6$ could not be well performed, since this particular memory vector does not satisfy the following two important properties $\boldsymbol{t}_i^* \otimes \boldsymbol{T} \approx \boldsymbol{t}_i$ and $\boldsymbol{f}_i^* \otimes \boldsymbol{T} \approx \boldsymbol{f}_i$ that are of basic importance for correct recognition of $\boldsymbol{T}$. Therefore we have to use slightly another approach to the construction of associative memory vector $\boldsymbol{T}$ based on the so-called *label vectors*.

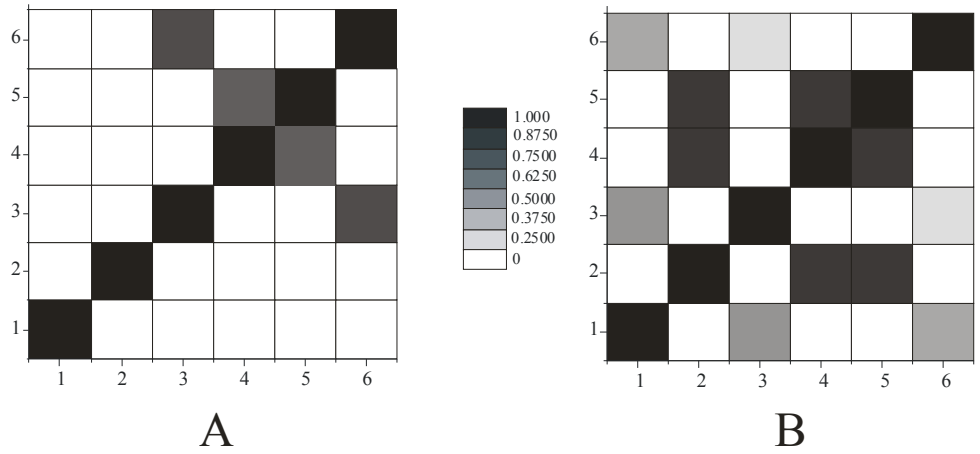


Figure 23. Graphic representation of similarities (A) between holographic distributed representations of derivation trees (98) and (B) between their terminal symbol saequences.

Let us have a set of six randomly generated label vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_6$, which will serve in our forthcoming considerations as labels for distinguishing of vector representations of derivation trees (that may be very similar from the standpoint of their mutual overlap specified by (21)). Applying these "label" vectors we construct the following new modified vectors

$$\tilde{\boldsymbol{t}}_i = \boldsymbol{x}_i \otimes \boldsymbol{t}_i \quad \text{and} \quad \tilde{\boldsymbol{f}}_i = \boldsymbol{x}_i \otimes \boldsymbol{f}_i \tag{100}$$

A transformation vector is determined as follow

$$\boldsymbol{T} = \sum_{i=1}^{6} \tilde{\boldsymbol{t}}_i^* \otimes \tilde{\boldsymbol{f}}_i \tag{101}$$

The modified vectors are specified by

$$\tilde{\boldsymbol{t}}_i \otimes \boldsymbol{T} \approx \tilde{\boldsymbol{f}}_i \quad \text{and} \quad \tilde{\boldsymbol{f}}_i \otimes \boldsymbol{T}^* \approx \tilde{\boldsymbol{t}}_i \tag{102}$$

for i=1,2,…,6. It means that the transformation vector $\boldsymbol{T}^*$ plays a role of parser, which assigns to a „formula" $\boldsymbol{f}$ the respective „derivation tree" $\boldsymbol{t}$. In other words, we have „algebraized" the highly symbolic parsing process of formulae, which assigns derivation trees to formulae, where the trees are very important for a construction of semantic interpretations of the

formulae (i.e. the semantic meaning of formulae is specified not only by its terminal symbols but also by its respective derivation tree).

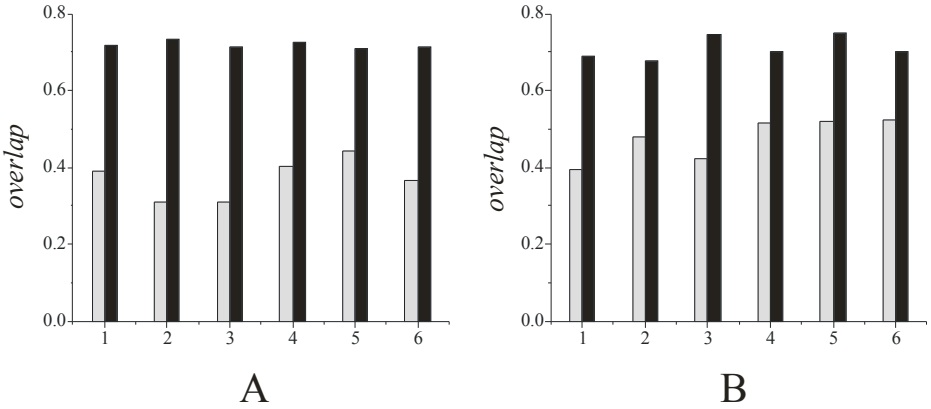


Figure 24. Reconstruction of holographic representations of derivation tress (A) and symbol sequences (B) induced by the grammar (95). The light darkness columns correspond to overlap between reconstructed and original vectors

6.10 Závery

Holografická redukovaná reprezentácia ponúka nové nekonvenčné riešenie jedného zo základných problémov umelej inteligencie a kognitívnej vedy, ktorý spočíva v distribuovanom kódovaní štruktúrovanej informácie (postupnosť znakov, vnorené relačné štruktúry, a pod.). Použitá distribuovaná reprezentácia je založená na dvoch binárnych operáciách: unárnej operácii „involúcie“ a binárnej operácii „konvolúcia“ nad doménou n-rozmenrých náhodne generovaných konceptuálnych vektorov, ktorých elementy spĺňajú normálnu distribúciu $N(0,1/n)$. Pomocou tejto distribuovanej reprezentácie môžeme modelovať rôzne typy asociatívnej pamäti, ktoré sú reprezentované konceptuálnym vektorom a taktiež dekódovať pamäťový vektor, t.j. určiť konceptuálne vektory, ktoré ho vytvorili. Táto rekognoskácia pamäťového vektora sa vykonáva pomocou procedúry čistenia (clean-up), ktorá nám určí pomocou vzájomného prekryvu vektor, ktorý je mu najviac podobný. Popísali sme taktiež proces agregácie vektorov, ktorý umožňuje prekonať ťažkosti rýchlej degradácie rekognoskácie pamäťového vektora tak, niekoľko vektorov sa združí do agregátu a tento sa zapamätáva prostredníctvom pamäťového vektora. Holografická redukovaná reprezentácia umožňuje hľadať podobnosť medzi dvoma štruktúrovanými konceptami pomocou jednoduchej algebraickej operácie skalárneho súčinu medzi ich distribuovanými reprezentáciami. Táto skutočnosť môže byť veľmi užitočnou pre modelovanie procesov prehľadávania pamäti na základe podobnosti (analógie) jej jednotlivých položiek. V záverečnej časti práce sme ukázali, že holografická redukovaná reprezentácia môže byť taktiež použitá k modelovaniu procesov usudzovania založených na pravidlách modus ponens a modus tollens, taktiež sme ukázali plodnosť tohto prístupu pri modelovaní procesu zovšeobecnenia pomocou indukcie.

Literatúra

[1] Davis, P. J.: *Circulant Matrices*. Chelsea Publishing, New York, 1999.

[2] Gabor, D.: Holographic model for temporal recall. *Nature*, **217** (1968), 1288–1289.

[3] McClelland, J. L. M., Rumelhart, D. E., and the PDP research group (eds.): *Parallel distributed processing: Explorations in the microstructure of cognition*, volumes 1 and 2. The MIT Press, Cambridge, MA (1986).

[4] Metcalfe Eich, J.: Levels of processing, encoding specificity, elaboration, and charm. *Psychological Review*, **92** (1985), 1–38.

[5] Murdock, B. B.: A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89** (1982), 316–338.

[6] Plate, T. A.: *Distributed Representations and Nested Compositional Structure*. Department of Computer Science, University of Toronto. Ph.D. Thesis, 1994.

[7] Plate, T. A.: *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*, CSLI Publications, Stanford, CA, 2003.

[8] Plate, T. A.: Holographic Distributed Representations. *IEEE Transaction on Neural Networks*, **6** (1995), 623-641.

[9] Slack, J. N.: The role of distributed memory in natural language processing. In O'Shea, T. (ed.): *Advances in Artificial Intelligence: Proceedings of the Sixth European Conference on Artificial Intelligence, ECAI-84*. Elsevier Science Publishers, New York, 1984.

[10] Smolensky, P., Legendre, G., Miyata, Y.: *Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition***.** Technical Report CU-CS-600-92, Department of Computer Science and 92-8, Institute of Cognitive Science. University of Colorado at Boulder, 1992, 75 pages.

[11] Smolensky, P.: On the proper treatment of connectionism. *The Behavioral and Brain Sciences,* **11** (1988), 1-74.

[12] Smolensky, P.: Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence,* **46** (1990)., 159-216.

[13] Smolensky, P: Neural and conceptual interpretations of parallel distributed processing models**.** In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. MIT Press - Bradford Books, Cambridge, MA, 1986, pp. 390-431.

[14] Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C.: Non-holographic associative memory. *Nature*, **222** (1969), 960–962.

[15] Shastri, L., Ajjanagadde, V.: From simple associations to systematic reasoning: connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, **16** (1993), 417-494.

[16] Kanerva, P.: The Spatter Code for Encoding Concepts at Many Levels. In Marinaro, M., Morasso, P. G. (eds.): *ICANN '94, Proceedings of the International Conference on Artificial Neural Networks*. Springer Verlag, London, 1994, pp. 226–229.

[17] Kanerva, P.: Binary Spatter-Coding of Ordered K-tuples. In von der Malsburg, C., von Seelen, W., Vorbruggen, J. C., Sendhoff, B. (eds.): *ICANN'96, Proceedings of the International Conference on Artificial Neural Networks*. Springer Verlag, Berlin, 1996, pp. 869-873.

[18] Kanerva, P.: Analogy as a Basis of Computation, In Uesaka, Y., Kanerva, P., Asoh, H.: *Foundations of real-world intelligence*. CSLI Publications, Stanford, CA, 2001.

[19] Fodor, J.A., Pylyshyn, Z.W.: Connectionism and cognitive architecture: A critical analysis., *Cognition*, **28** (1988) 3-71.

[20] Chomsky, N.: Three models for the description of language. *IEEE Trans. Info. Theory* **2** (1956), 113 - 124.

[21] Johnson – Laird, P.N.: *Mental Models. Toward a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge 1983.

[22] Russel,. S., Norvig, P.: *Artificial Intelligence. A Modern Approach*. Prentice Hall, New Jersey 1995.

7. prednáška

Mentálne modely logiky v kognitívnej vede

7.1 Úvodné poznámky

Postavenie logiky v kognitívnej vede je veľmi kontroverzné, je v protiklade s intuitívnym pohľadom na logiku, ako na *posvätný grál* kognitívnej vedy a umelej inteligencie. Logika nám slúži ako hlavný teoretický symbolický prostriedok na modelovanie usudzovania (Návrat, 2001; Russel, Norvig, 1995). Avšak na druhej strane, pri analýze ľudských schopností používať logiku na elementárnej úrovni (Johnson – Laird, 1983; Stenning, van Lambalgen, 2007), pozorujeme závažné nedostatky. Tieto experimentálne závery sú v kontradikcii s našou všeobecnou vierou, že človek je bytosť racionálna, ktorá uvážlivo používa zásady logiky pri vyvodzovaní nových poznatkov z daných predpokladov – poznatkov a viedli spolu s inými argumentmi ku vzniku koncepcie *ohraničenej racionality* (Cosmides, Tooby, 1992; Gigerenzer, Selten, 2001; Johnson – Laird, 1983; Stenning, van Lambalgen, 2007).

Cieľom tejto práce je prezentovať pohľad súčasnej kognitívnej vedy na tvorbu mentálneho *modelu* logiky. Kognitívna veda sa už niekoľko desaťročí zaoberá problémom, ako ľudia používajú jednoduché logické zákony a aké sa pritom vyskytujú chyby. Inovatívnosť kognitivistického pohľadu na túto zaujímavú problematiku (Johnson – Laird, 1983; Stenning, van Lambalgen, 2007) spočíva v tom, že sa navrhujú mentálne modely kognitívnych aktivít človeka pri riešení jednoduchých logických úloh. Požaduje sa, aby tieto modely boli schopné postihnúť nielen správne riešenia ale aj interpretovať chyby pri konštrukcii riešenia. Mentálny model má určitú zložitosť, ktorá je premenlivá pri aplikovaní na rôzne problémy (napr. na rôzne kategorické sylogizmy) (Johnson – Laird, 1983), pričom očakávame, že táto zložitosť modelu koreluje s výskytom chýb. Mentálne modely (alebo mentálne modelovanie) boli prvý krát postulované škótskym psychológom Kenneth Craikom v r. 1943, ktorý predpokladal, že ľudská myseľ obsahuje malé modely reality, pomocou ktorých sme schopný prijať - anticipovať udalosti, uvažovať a vysvetlovať. Model existuje v pracovnej pamäti a je tvorený ako výsledok komunikácie alebo predstavivosti. Základnou črtou mentálnych modelov je, že sú „izomorfné“ s objektom, ktorý reprezentujú (podobne, ako sú chemické molekulárne modely podobné molekulám). Podľa Johnson-Lairda, použitie mentálnych modelov v kognitívnej psychológii odstránilo jej deskriptívny charakter, bez snahy interpretovať pozorované výsledky. Existuje dve možnosti pre tvorbu mentálneho modelu logiky:

(1) *Syntaktický prístup*. Mentálny model je stotožnený s Gentzenovou prirodzenou dedukciou (Gentzen, 1935), ktorá obsahuje okolo tuctu pravidiel usudzovania vychádzajúcich z elementárnych zákonov logiky s jednoduchou a jasnou interpretáciou ich významu. Z množiny predpokladov $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$ použitím vyššie zmienených pravidiel prirodzenej dedukcie odvodíme záver ψ. Hovoríme, že tento záver logicky vyplýva z predpokladov (alebo, že existuje logický dôkaz formuly ψ z množiny predpokladov $\{\varphi_1, \varphi_2, ..., \varphi_n\}$), čo zapisujeme pomocou „relácie“ $\vdash$ takto: $\Phi \vdash \psi$. Syntaktický prístup k tvorbe mentálneho modelu logiky je vhodný pre ľudí, ktorí už absolvovali určité základné

vzdelanie z logiky a preto môžu suverénne používať prirodzenú dedukciu ku konštrukcii dôkazov zložitejších formúl.

(2) *Sémantický prístup.* Mentálny model je stotožnený s tvorbou *sémantického* modelu množiny predpokladov, Φ, spolu s tautologickým dôsledkom ψ, čo zapisujeme $\Phi \models \psi$. Sémantický prístup je založený na grafickej metóde nazývanej *sémantické tablá*. K ich konštrukcii potrebujeme poznať len elementárne základné pojmy sémantickej interpretácie logických spojok.

7.2 Syntax, sémantika a pragmatika výrokovej logiky

Jeden z prvých formálnych systémov v ľudskej histórii bol Euklidov axiomatický systém geometrie, ktorý sa stal vzorom pre rozvíjanie ďalších systémov hlavne v rámci matematiky, informatiky a logiky. V teórii formálnych systémov sa obvykle od seba striktne separujú tieto tri aspekty:

(1) *syntax* – spôsob konštrukcie jazyka (formúl) teórie,
(2) *sémantika* – významová interpretácia jazyka (jednotlivých formúl), a
(3) *pragmatika* – zmena sémantiky pri zmene prostredia/situácie v ktorej prebieha významová interpretácia formúl.

Pri špecifikácii formálneho systému vystupujú do popredia nasledujúce dve otázky: (1) Za akých podmienok je dôkaz formuly - vety korektný a (2) aké metódy môžu byť použité pri konštrukcii týchto dôkazov.

Veta (teorém, formula, výrok, zákon, argument, ...) je výrok o ktorom môže byť ukázané, že je pravdivý. V tejto súvislosti hovoríme o *dôkaze* vety, ktorý spočíva v postupnosti jednotlivých „medzikrokov“, ktoré sú odvodené buď z množiny jednoduchých postulátov, nazývaných *axiómy*, alebo z predchádzajúcich viet (pomocných viet, často nazývaných lemy) danej postupnosti. Komplikované dôkazy sú obvykle jasnejšie formulované, keď ich dôkaz je rozdelený na jednotlivé medzikroky, ktoré sú formulované ako samostatné vety. Tieto medzikroky - vety v postupnosti sú vytvárané pomocou *pravidiel odvodzovania* (*pravidiel usudzovania*), ktoré z niekoľkých pravdivých tvrdení - argumentov vytvorí nové pravdivé tvrdenie - argument.

Metódy dôkazu logiky sú dôležité nielen pre tvorbu korektných dôkazov v samotnej logike, ale aj v matematike a v informatike. V teoretickej informatike sa napr. študujú rôzne metódy verifikácie korektnosti programu, alebo či operačný systém je bezpečný. V umelej inteligencii pri odvodzovaní nových faktov z danej databázy poznatkov (množiny výrokových formúl, ktorá sa vo výrokovej logike nazýva teória) je dôležité mať zabezpečené, aby daná databáza bola konzistentná (korektná), teda aby z nej súčasne nevyplýval nejaký výrok a taktiež aj jeho negácia. Môžeme teda konštatovať, že zvládnutie metód matematického dôkazu je dôležité tak v samotnej logike, ako aj v matematike a v informatike.

Na záver tejto úvodnej kapitoly uvedieme ešte niekoľko poznámok k pragmatike formálneho systému logiky. Obvykle sa toto hľadisko vôbec nespomína pri špecifikácii logiky, verí sa, že logika je rigidný formálny systém, kde nemôže nastať zmena sémantickej interpretácie pri zmene prostredia (napr. sociálneho). Medzi logikmi existuje pevná viera, ktorá pochádza už od starovekých gréckych logikov, že logika je len jedna a používanie jej zákonov nezávisí od prostredia v ktorom užívateľ vykonáva usudzovania a pod.

Oblasť pragmatiky logiky bola otvorená v psychológii ruským neuropsychológom z prvej polovice minulého storočia Alexandrom R. Luriom (1979), ktorý v 30. rokoch vykonal so svojimi spolupracovníkmi expedíciu do stredoázijských oblastí vtedajšieho Ruska,

kde skúmali kultúrnu závislosť používania jednoduchých logických zákonov. Miestnemu negramotnému obyvateľovi najprv vysvetlili vstupnú premisu úlohy, že každé zviera, ktoré žije za polárnym kruhom má bielu farbu kožušiny, potom nasledovala konkrétna otázka, akú farbu kožuchu má ľadový medveď, ktorý žije za polárnym kruhom, čo formálne môžeme vyjadriť pomocou schémy

každé zviera žijúce za polárnym kruhom má bielu farbu kožušiny
ľadový medveď žije za polárnym kruhom

?

Väčšina negramotných respondentov odpovedala, že nevie, pretože nikdy neboli za polárnym kruhom. Odpoveď gramotných respondentov (hlavne tých, čo absolvovali vzdelanie v ruskom jazyku) na túto otázku bola diametrálne odlišná, odpovedali, že polárny medveď má biely kožuch

každé zviera žijúce za polárnym kruhom má bielu farbu kožušiny
ľadový medveď žije za polárnym kruhom

ľadový medveď má bielu farbu kožušiny

A. Luria z týchto a podobných pozorovaní vyvodzoval, že existuje *kultúrna závislosť* pri používaní schém usudzovania logiky, že ich používanie je závislé od vzdelania respondentov a pod. Tieto experimenty sú neobyčajne zaujímavé a boli v podstate zreprodukované aj v iných oblastiach (napr. v Afrike a v Pacifiku). Žiaľ, existuje aj ich veľmi tvrdá kritika, že sa nejedná o žiadnu kultúrnu závislosť používania elementárnych schém usudzovania, ale beží o problém komunikácie medzi negramotným respondentom a riešiteľom, medzi ktorými existuje veľká kultúrna bariéra, pričom je vôbec otázne, či respondenti správne pochopili položenú otázku. Gramotní respondenti, ktorí zvládli v škole základné trívium (čítanie, písanie a počítanie) obvykle v ruskom jazyku, zrejme už podstatne lepšie chápu riešený problém a sú schopní použiť pravidlo modus ponens na konštrukciu správnej odpovede – riešenia.

7.3 Špecifikácia jazyka výrokovej logiky (syntax)

Jazyk L výrokovej logiky je tvorený formulami výrokovej logiky, ktoré sú definované pomocou množiny atomických výrokových premenných $\{p,q,...,p',q',...\}$ a množiny logických spojok $\{\wedge,\vee,\Rightarrow,\equiv,\neg\}$.

Formuly výrokovej logiky, ktoré tvoria jazyk L, sú rekurentne definované ako minimálna množina, ktorá vyhovuje týmto vlastnostiam

(1) $\{p,q,...,p',q',...\}\subseteq L$,

(2) ak $(\varphi\in L)$, potom $(\neg\varphi)\in L$,

(3) ak $(\varphi,\psi\in L)$, potom $(\varphi\wedge\psi),(\varphi\vee\psi),(\varphi\Rightarrow\psi),(\varphi\equiv\psi)\in L$.

7.4 Špecifikácia významu výrokovej logiky (sémantiky)

Ďalší pojem dôležitý pre výrokovú logiku je *sémantika*. Pojem pochádza z teórie prirodzených jazykov, kde sémantika špecifikuje význam danej vety (ktorá má tiež aj svoju syntax). Vo výrokovej logike, ktorá sa zaoberá len pravdivostnými hodnotami premenných

a ich formúl, *sémantika* nie je veľmi bohatá. Sémantika výrokovej formuly je vlastne tabuľka pravdivostných hodnôt formuly pre rôzne hodnoty jej atomických výrokov. Formuly (elementy jazyka L výrokovej logiky) majú pravdivostný význam, ktorý je špecifikovaný takto: symbol ´1´ reprezentuje pravdivostnú hodnotu ´pravda´ a symbol ´0´ reprezentuje pravdivostnú hodnotu ´nepravda´.

(1) *Konjunkcia* $\alpha \wedge \beta$ je pravdivá vtedy a len vtedy (vtt) ak obe jej komponenty sú pravdivé, v opačnom prípade je nepravdivá

$$val(\alpha \wedge \beta) = 1 \text{ vtt } val(\alpha) = 1 \text{ a } val(\beta) = 1$$

$$val(\alpha \wedge \beta) = 0 \text{ vtt } val(\alpha) = 0 \text{ alebo } val(\beta) = 0$$

(2) *Disjunkcia* $\alpha \vee \beta$ je pravdivá vtt, ak aspoň jedna jej komponenta je pravdivá, v opačnom prípade je nepravdivá

$$val(\alpha \vee \beta) = 1 \text{ vtt } val(\alpha) = 1 \text{ alebo } val(\beta) = 1$$

$$val(\alpha \vee \beta) = 0 \text{ vtt } val(\alpha) = 0 \text{ a } val(\beta) = 0$$

(3) *Implikácia* $\alpha \Rightarrow \beta$ je pravdivá vtt, ak prvá jej komponenta (α) je nepravdivá alebo druhá komponenta (β) je pravdivá, v opačnom prípade (prvá komponenta je pravdivá a druhá komponenta je nepravdivá) je nepravdivá

$$val(\alpha \Rightarrow \beta) = 1 \text{ vtt } val(\alpha) = 0 \text{ alebo } val(\beta) = 1$$

$$val(\alpha \Rightarrow \beta) = 0 \text{ vtt } val(\alpha) = 1 \text{ a } val(\beta) = 0$$

(4) *Ekvivalencia* $\alpha \equiv \beta$ je pravdivá vtedy vtt, ak obe komponenty majú rovnakú pravdivostnú hodnotu, v opačnom prípade je nepravdivá

$$val(\alpha \equiv \beta) = 1 \text{ vtt } val(\alpha) = val(\beta)$$

$$val(\alpha \equiv \beta) = 0 \text{ vtt } val(\alpha) \neq val(\beta)$$

(5) *Negácia* $\neg\alpha$ je pravdivá vtt, ak jej komponenta je nepravdivá, v opačnom prípade je nepravdivá

$$val(\neg\alpha) = 1 \text{ vtt } val(\alpha) = 0$$

$$val(\neg\alpha) = 0 \text{ vtt } val(\alpha) = 1$$

Pri tejto príležitosti je vhodné spomenúť psychológov B. Inhelderovú a J. Piageta, ktorí dôvodia (Inhelder, Piaget, 1964), že už deti vo veku 10 rokov zvládajú zásady sémantického prístupu k výrokovej logike. Sú schopné korektne reprodukovať tabuľku pravdivostných hodnôt pre disjunkciu, konjunkciu a negáciu. Problémy so sémantickou interpretáciou implikácie pretrvávajú až do dospelosti, správna interpretácia vyžaduje základné logické vzdelanie. Z týchto dôvodov sémantický prístup k tvorbe mentálneho modelu logiky je vhodný pre „laikov v logike“, ktorí sú schopní správne zreprodukovať pravdivostné tabuľky logických spojok a nemusia mať žiadne iné špeciálne vedomosti z logiky. Teraz len toľko, interpretácia implikácie je silne textovo závislá, v niektorých extrémnych prípadoch v prirodzenom jazyku sa interpretuje ako konjunkcia, čo môže spôsobovať v iných prípadoch určité problémy s korektnou sémantickou interpretáciou implikácie. Preto jediné možné riešenie tohto problému vidíme v tom, že už malým deťom by sa malo „drilom“ zafixovať, že implikácia $p \Rightarrow q$ je pravdivá práve vtedy, ak p je nepravdivé alebo q je pravdivé a taktiež, že implikácia $p \Rightarrow q$ je nepravdivá práve vtedy, ak p je pravdivé a q je nepravdivé (tieto dve skutočnosti je možné vyjadriť Aristotelovou dikciou, že pravda implikuje len pravdu a nepravda implikuje čokoľvek, čiže výrok $1 \Rightarrow 0$ musí byť nepravdivý). Pravdivostnú hodnotu formuly φ môžeme jednoducho určiť pomocou

tabuľkovej metódy (pozri tabuľku 1), kde postupne počítame pravdivostné hodnoty jej podformúl, ktoré v ďalších krokoch využívame pre konštrukciu pravdivostných hodnôt „podformúl“, ktoré obsahujú predchádzajúce podformuly ako svoju časť.

Formula φ sa nazýva *tautológia* (alebo *zákon*, čo vyjadríme $\models \varphi$), ak pre každú interpretáciu τ platí $val_\tau(\varphi) = 1$

$$(\models \varphi) =_{def} \forall(\tau)\left(val_\tau(\varphi) = 1\right)$$

V tab. 1 formula φ_2 je tautológia, je pravdivá pre všetky možné interpretácie τ_i, i = 1, 2, 3, 4. V opačnom prípade, ak pre každú interpretáciu τ platí $val_\tau(\varphi) = 0$, formula sa nazýva *kontradikcia*. Ak existuje aspoň jedna interpretácia τ taká, že $val_\tau(\varphi) = 1$, potom formula φ je *splniteľná* (to znamená, že tautológia je špeciálny prípad splniteľnosti, čo zapisujeme $\models \varphi$). V tab. 1 formula φ_1 je splniteľná pre dve interpretácie τ_1 a τ_4. Môžeme teda povedať, že všetky formuly, ktoré nie sú kontradikcie sú splniteľné a tautológie sú také splniteľné formuly, ktoré sú pre všetky možné interpretácie τ pravdivé.

Tabuľka 1. Tabuľková metóda pre výpočet pravdivostných hodnôt formúl

$\varphi_1 = (p \vee q) \Rightarrow (p \wedge q)$ a $\varphi_2 = (p \wedge q) \Rightarrow (p \vee q)$

τ	p	q	$p \vee q$	$p \wedge q$	$\varphi_1 = (p \vee q) \Rightarrow (p \wedge q)$	$\varphi_2 = (p \wedge q) \Rightarrow (p \vee q)$
τ_1	0	0	0	0	1	1
τ_2	0	1	1	0	0	1
τ_3	1	0	1	0	0	1
τ_4	1	1	1	1	1	1

Tautológie majú vo výrokovej logike mimoriadne postavenie *zákonov* logiky, tieto formule sú vždy pravdivé pre ľubovoľné pravdivostné hodnoty premenných. Niektoré tautológie sa často používajú nielen v samotnej výrokovej logike, ale aj v bežnom usudzovaní a sú obvykle označované aj vlastným menom. Väčšinou ide o tautológie tvaru ekvivalencie, ktoré umožňujú nahradzovať jedny formuly inými bez straty vlastnosti ich tautologičnosti.

Ľubovoľná neprázdna množina formúl, $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$, sa nazýva *teória* výrokovej logiky. Ak pre teóriu Φ existuje taká interpretácia τ, pre ktorú sú všetky formuly pravdivé, $val_\tau(\varphi_i) = 1$, pre $i = 1, 2, ..., n$, potom táto interpretácia τ sa nazýva *model teórie*. Teória Φ sa nazýva *konzistentná*, ak má model.

Tabuľka 2. Konštrukcia modelu pre tri formuly výrokovej logiky

$\varphi_1 = (p \vee q) \Rightarrow (p \wedge q)$ $\qquad$ $\varphi_2 = (p \Rightarrow q) \wedge (q \Rightarrow p)$

1	2	3	4	5
p	q	$p \vee q$	$p \wedge q$	$3 \Rightarrow 4$
0	0	0	0	1
0	1	1	0	0
1	0	1	0	0
1	1	1	1	1

1	2	3	4	5
p	q	$p \Rightarrow q$	$q \Rightarrow p$	$3 \wedge 4$
0	0	1	1	1
0	1	1	0	0
1	0	0	1	0
1	1	1	1	1

$\varphi_3 = (\neg p \wedge \neg q) \Rightarrow (p \Rightarrow q)$

1	2	3	4	5	6	7
p	q	$\neg p$	$\neg q$	$3 \wedge 4$	$p \Rightarrow q$	$5 \Rightarrow 6$
0	0	1	1	1	1	1
0	1	1	0	0	1	1
1	0	0	1	0	0	1
1	1	0	0	0	1	1

Príklad 1. Nech

$$\Phi = \{\varphi_1 = (p \vee q) \Rightarrow (p \wedge q), \varphi_2 = (p \Rightarrow q) \wedge (q \Rightarrow p), \varphi_3 = (\neg p \wedge \neg q) \Rightarrow (p \Rightarrow q)\}$$

chceme zistiť, či táto teória má model. Pomocou tabuľkovej metódy určíme pravdivostné hodnoty týchto formúl pre všetky možné interpretácie, pozri tab. 2.

Z týchto tabuliek vyplýva, že existujú dve interpretácie premenných, $\tau_1 = (p/0, q/0)$ a $\tau_4 = (p/1, q/1)$, pre ktoré všetky formuly z Φ sú pravdivé, t.j. interpretácie τ_1 a τ_4 sú modelom teórie Φ. Tiež môžeme povedať, že teória Φ je konzistentná, čo vyplýva priamo zo skutočnosti, že má model.

Formula φ sa nazýva *tautologický dôsledok teórie* Φ (čo označíme $\Phi \models \varphi$) práve vtedy, ak každý model teórie Φ je aj modelom formuly φ (t.j. formula φ je v ňom pravdivá). Nech φ je tautologickým dôsledkom teórie Φ, potom pre každý model – interpretáciu τ platí: $val_\tau(\varphi) = val_\tau(\varphi_i) = 1$, pre $i = 1, 2, ..., n$.

Príklad 2. Nech teória Φ je definovaná rovnako ako v predchádzajúcom príklade, má dva modely určené interpretáciami premenných $\tau_1 = (p/0, q/0)$ a $\tau_4 = (p/1, q/1)$. Uvažujme formulu φ v tvare $p \wedge q$, potom táto formula nie je tautologickým dôsledkom teórie Φ, pretože len pre model τ_2 je formula pravdivá, $val_{\tau_2}(\varphi) = 1$, pre model τ_1 už nie je pravdivá, $val_{\tau_1}(\varphi) = 0$.

V predchádzajúcich kapitolách boli formulované dva nezávislé pojmy:

(1) V syntaktickom prístupe pojem logického dôsledku formuly ψ z množiny predpokladov Φ, čo formálne zapisujeme $\Phi \vdash \psi$.

(2) V sémantickom prístupe pojem tautologického dôsledku formuly ψ z teórie Φ, čo formálne zapisujeme $\Phi \models \psi$.

Aj keď sú tieto pojmy definované nezávislým spôsobom (syntaktickým a sémantickým), existuje medzi nimi tesná väzba. Podľa Postovej vety výrokovej logiky (Jirků, Vejnarová, 2000; Kvasnička, Pospíchal, 2006; Peregrin, 2004; Sochor, 2001; Švejdar, 2002) platí tvrdenie:

Pre ľubovoľnú formulu φ vzťah $\Phi \vdash \psi$ platí práve vtedy, ak $\Phi \models \psi$,

$$(\Phi \models \psi) \Rightarrow (\Phi \vdash \psi)$$

Opačná implikácia

$$(\Phi \vdash \psi) \Rightarrow (\Phi \models \psi)$$

je triviálna (ľahko skontrolujeme, že axiómy sú tautológie a použitie pravidla *modus ponens* z tautologických predpokladov poskytuje tautologický záver). Spojením týchto dvoch implikácií dostaneme ekvivalenciu medzi logickým a tautologickým vyplývaním.

$$(\Phi \models \psi) \equiv (\Phi \vdash \psi)$$

7.5 Teória dôkazu – prirodzená dedukcia (syntaktický prístup)

Klasický spôsob riešenia logického dôkazu vo výrokovej logike je Hilbertova axiomatická metóda (Gahér, 1998; Jirků, Vejnarová, 2000; Kvasnička, Pospíchal, 2006; Peregrin, 2004; Sochor, 2001; Švejdar, 2002), ktorá postuluje systém 10 axióm a jedného pravidla usudzovania. Aj keď je tento syntaktický prístup elegantný a v zásade jednoduchý, jeho použitie na dôkaz nových zákonov výrokovej logiky (ktoré nie sú obsiahnuté v množine axióm) v mnohých prípadoch je netriviálna záležitosť, ktorá obvykle vyžaduje množstvo „jemných“ trikov a postupov, aby sme dokázali aj pomerne jednoduché zákony výrokovej logiky (napr. $\vdash p \Rightarrow p$). Alternatívny prístup ku konštrukcii teórie dôkazu je Gentzenova prirodzená dedukcia z r. 1935 (Szabo, 1969), ktorá je založená na jednej triviálnej axióme (napr. $p \Rightarrow p$ alebo $p \vee \neg p$) a okolo tuctu pravidiel odvodzovania). Pravidlá usudzovania v prirodzenej dedukcii sú tvorené schémou (pozri tab. 3)

$$\left|\begin{array}{l} predpoklad_1 \\ \ldots\ldots\ldots\ldots \\ predpoklad_n \\ \hline záver \end{array}\right| \Rightarrow \left|\begin{array}{l} \varphi_1 \\ \ldots\ldots \\ \varphi_n \\ \hline \psi \end{array}\right|$$

ktorá obsahuje *n predpokladov* $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ a jeden *záver* ψ. Táto schéma usudzovania je formalizovaná pomocou „relácie“ $\vdash$ *logického dôkazu* (alebo *vyplývania*)

$$\{predpoklad_1, \ldots, predpoklad_n\} \vdash záver \quad \text{alebo} \quad \{\varphi_1, \ldots, \varphi_n\} \vdash \varphi \quad \text{alebo} \quad \Phi \vdash \varphi$$

Pomerne jednoduchými prostriedkami môžeme dokázať, že relácia „logického dôkazu – vyplývania) je ekvivalentná formulám

$$\vdash \varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow \varphi \equiv \vdash \varphi_1 \Rightarrow (\varphi_2 \Rightarrow (\Rightarrow \ldots (\varphi_n \Rightarrow \varphi)))$$

V oboch týchto formulách používame symbol $\vdash$ s prázdnou množinou predpokladov, $\Phi = \varnothing$, čo znamená, že formula $\varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow \varphi$ (alebo $\varphi_1 \Rightarrow (\varphi_2 \Rightarrow (\Rightarrow \ldots (\varphi_n \Rightarrow \varphi)))$) je zákonom výrokovej logiky, ku ktorého konštrukcii nepotrebujeme žiadne predpoklady. Podľa ľavej formuly, konjunkcia predpokladov implikuje záver.

Naše úvahy zformalizujeme pomocou nasledujúceho sledu definícií:

(1) Formula φ sa nazýva *bezprostredným logickým dôsledkom* množiny formúl $\Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$ vtedy a len vtedy, ak vznikne aplikáciou jedného z pravidiel logického dôkazu na formuly z Φ.

(2) Formula φ sa nazýva *logický dôsledok* množiny formúl Φ (čo označíme $\Phi \vdash \varphi$ vtedy a len vtedy, ak $\varphi \in \Phi$ alebo je bezprostredným dôsledkom Φ alebo je bezprostredným dôsledkom Φ rozšírenej o niektoré jej dôsledky.

(3) Konečná postupnosť formúl $\varphi_1, \varphi_2, \ldots, \varphi_n$ sa nazýva ***dôkaz*** formuly φ z množiny Φ vtedy

a len vtedy, ak $\varphi = \varphi_n$ a každá formula φ_i z tejto postupnosti je buď bezprostredným logickým dôsledkom niektorých formúl z Φ alebo formúl $\varphi_1, \varphi_2, ..., \varphi_{i-1}$.

Tabuľka 3. Pravidlá usudzovania prirodzenej dedukcie

#	Schéma usudzovania	Ekvivalentný zákon výrokovej logiky	Názov
1	$\dfrac{p}{p \vee q}$	$p \Rightarrow (p \vee q)$	*adícia*
2	$\dfrac{p \wedge q}{p}$, $\dfrac{p \wedge q}{q}$	$(p \wedge q) \Rightarrow p$, $(p \wedge q) \Rightarrow q$	*simplifikácia*
3	$\dfrac{\begin{array}{l}p\\q\end{array}}{p \wedge q}$	$p \Rightarrow (q \Rightarrow (p \wedge q))$	*introdukcia konjunkcie*
4	$\dfrac{\begin{array}{l}p\\p \Rightarrow q\end{array}}{q}$	$p \Rightarrow ((p \Rightarrow q) \Rightarrow q)$	*modus ponens*
5	$\dfrac{\begin{array}{l}\neg q\\p \Rightarrow q\end{array}}{\neg p}$	$\neg q \Rightarrow ((p \Rightarrow q) \Rightarrow \neg p)$	*modus tollens*
6	$\dfrac{q}{p \Rightarrow q}$	$q \Rightarrow (p \Rightarrow q)$	*introdukcia implikácie*
7	$\dfrac{\begin{array}{l}p \Rightarrow q\\q \Rightarrow r\end{array}}{p \Rightarrow r}$	$(p \Rightarrow q) \Rightarrow ((q \Rightarrow r) \Rightarrow (p \Rightarrow r))$	*hypotetický sylogizmus*
8	$\dfrac{\begin{array}{l}p \vee q\\\neg p\end{array}}{q}$, $\dfrac{\begin{array}{l}p \vee q\\\neg q\end{array}}{p}$	$((p \vee q) \wedge \neg p) \Rightarrow q$ $((p \vee q) \wedge \neg q) \Rightarrow p$	*eliminácia disjunkcie*
9	$\dfrac{p \Rightarrow q}{\neg p \vee q}$, $\dfrac{\neg p \vee q}{p \Rightarrow q}$	$(p \Rightarrow q) \equiv (\neg p \vee q)$	*disjunktívny sylogizmus*
10	$\dfrac{p \Rightarrow q}{\neg q \Rightarrow \neg p}$	$(p \Rightarrow q) \Rightarrow (\neg q \Rightarrow \neg p)$	*inverzia implikácie*
11	$\dfrac{\begin{array}{l}p \Rightarrow q\\p \Rightarrow \neg q\end{array}}{\neg p}$	$(p \Rightarrow q) \Rightarrow ((p \Rightarrow \neg q) \Rightarrow \neg p)$	*reductio ad absurdum*
12	$\dfrac{\neg\neg p}{p}$, $\dfrac{p}{\neg\neg p}$	$\neg\neg p \equiv p$	*dvojitá negácia*

Príklad 3. Majme dva jednoduché výroky

(1) Keď bude pršať, potom pôjdem do kina

(2) Keď bude pršať, potom pôjdem do kaviarne

Aký záver z nich vyplýva?

V prvom kroku vykonáme formalizáciu týchto výrokov, zavedieme tri atomické výrokové premenné

p = ´bude pršať´, q = ´pôjdem do kina´, r = ´pôjdem do kaviarne´

Potom množina predpokladov má tvar $\Phi = \{p \Rightarrow q, p \Rightarrow r\}$, zaujíma nás, aký netriviálny dôsledok vyplýva z týchto predpokladov, $\Phi = \{p \Rightarrow q, p \Rightarrow r\} \vdash ?$ Množinu predpokladov rozšírime o pomocný predpoklad (hovoríme, že je aktivovaný).

1.	$p \Rightarrow q$	(1. predpoklad φ_1)
2.	$p \Rightarrow r$	(2. predpoklad φ_2)
3.	p	(aktivácia pomocného predpokladu φ)
4.	q	(použitie pravidla 4 *modus ponens* na predpoklady 1 a 3)
5.	r	(použitie pravidla 4 *modus ponens* na predpoklady 2 a 3)
6.	$q \wedge r$	(introdukcia konjunkcie na dôsledky 4 a 5)
7.	$p \Rightarrow q \wedge r$	(deaktivácia pomocného predpokladu 3 pomocou dôsledku 6)

Záver, ktorý vyplýva z predpokladov je $p \Rightarrow q \wedge r$ = ´ak bude pršať pôjdem do kina a kaviarne´.

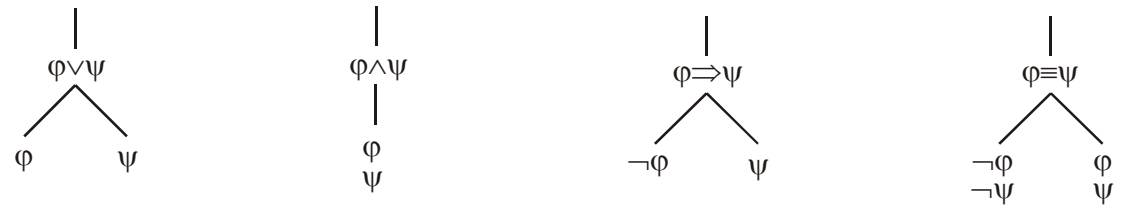
Na záver tejto kapitoly zosumarizujeme hlavné závery syntaktického prístupu pre logický dôkaz formúl výrokovej logiky:

(1) *Pôvodný Hilbertov formálny systém* logického dôkazu je pre kognitívnu vedu bezcenný, je veľmi zložitý na to, aby sa dal použiť ako mentálny model usudzovania.

(2) *Gentzenov systém prirodzenej dedukcie* (ktorý je formálne rovnocenný Hilbertovmu systému) nesie všetky znaky jednoduchosti a intuitívnosti, preto je vhodný ako *mentálny model logiky* pre tých, čo už absolvovali základné vzdelanie v logike.

7.6 Model výrokovej logiky – sémantické tablá (sémantický prístup)

V predchádzajúcej kapitole boli formulované základné princípy syntaktického prístupu k výrokovej logike. Cieľom tejto kapitoly bude ukázať realizáciu sémantického prístupu pomocou diagramatickej techniky nazývanej sémantické tablá (Smullyan, 1968). V rámci tejto metódy riešime reláciu $\Phi = \{\varphi_1, ..., \varphi_n\} \models \psi$ tak, že formulu $\varphi_1 \wedge ... \wedge \varphi_n \Rightarrow \neg\psi$ rozkladáme postupne na podformuly, pričom tento rekurentný proces končíme vtedy, keď podformuly obsahujú len atomické premenné alebo ich negácie (tzv. *literály*), pozri obr. 1.



A (*disjunkcia*) *B* (*konjunkcia*) *C* (*implikácia*) *D* (*ekvivalencia*)

Obrázok 1. Tri základné rozklady pre tvorbu sémantického tabla. V prípade, že vytvorená podformula má tvar $\neg(\varphi \wedge \psi)$ alebo $\neg(\varphi \vee \psi)$, použijeme De Morganove vzťahy na ich ekvivalentný prepis na $(\neg\varphi \vee \neg\psi)$ resp.

na $(\neg\varphi \wedge \neg\psi)$, pre ktoré už existujú diagramatické rozklady. Diagram D reprezentuje rozklad pre ekvivalenciu, kde bola použitá definičná formula $(\varphi \equiv \psi) =_{def} (\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi)$, pričom implikácie boli rozložené pomocou diagramu C a dve takto vzniknuté uzavreté vetvy boli odstránené ako nepodstatné.

Pomocou metódy sémantických tabiel budeme riešiť úlohu, či daná výroková formula ψ má model, t. j. či existuje aspoň jedna interpretácia τ taká, že $val_{\tau}(\psi) = 1$. Táto úloha je vo všeobecnosti riešiteľná pomocou tabuľkovej metódy (pozri tab. 2), avšak metóda sémantických tabiel predstavuje veľmi efektívnu alternatívu k tabuľkovej metóde, kde musíme vyšetrovať 2^n riadkov (ak formula ψ má n atomických výrokových premenných). Metóda sémantických tabiel je založená na postupnej diagramatickej transformácii formuly ψ do ekvivalentného DNF (Kvasnička, Pospíchal, 2006) tvaru ψ_{DNF} (disjunktívna normálna forma, ktorá obsahuje disjunkcie konjunkcií literálov).

Príklad 4. Prepíšeme formulu $\psi = (p \equiv q) \wedge (r \vee \neg(r \vee \neg p))$ do ekvivalentného DNF tvaru

$$\psi' = \psi_{DNF} = \underbrace{(\neg p \wedge \neg q \wedge r)}_{\square} \vee \underbrace{(\neg p \wedge p \wedge r)}_{\times} \vee \underbrace{(q \wedge \neg q \wedge r)}_{\times} \vee \underbrace{(q \wedge p \wedge r)}_{\square} \vee$$
$$\underbrace{(\neg p \wedge \neg q \wedge \neg r \wedge p)}_{\times} \vee \underbrace{(\neg p \wedge p \wedge \neg r \wedge p)}_{\times} \vee \underbrace{(q \wedge \neg q \wedge \neg r \wedge p)}_{\times} \vee$$
$$\underbrace{(q \wedge p \wedge \neg r \wedge p)}_{\square}$$
$$= \underbrace{(\neg p \wedge \neg q \wedge r)}_{\tau_1=(0,0,1)} \vee \underbrace{(q \wedge p \wedge r)}_{\tau_2=(1,1,1)} \vee \underbrace{(q \wedge p \wedge \neg r)}_{\tau_3=(1,1,0)}$$

Vidíme, že v takto upravenej DNF formule existujú klauzuly, ktoré sú a-priori nepravdivé (obsahujú konjunkciu premennej a jej negácie), ktoré sú označené symbolom ´×´. Pre zostávajúce klauzule, ktoré sú označené symbolom ´□´ existuje vždy interpretácia τ, pre ktorú sú pravdivé (pozri obr. 2 a tab. 4).

Tabuľka 4. Tabuľka pravdivostných hodnôt formuly $\psi = (p \equiv q) \wedge (r \vee \neg(r \vee \neg p))$

#	p	q	r	$p \equiv q$	$r \vee \neg p$	$\neg(r \vee \neg p)$	$r \vee \neg(r \vee \neg p)$	ψ	
1	0	0	0	1	1	0	0	0	
2	0	0	1	1	1	0	1	**1**	τ_1=(0,0,1)
3	0	1	0	0	1	0	0	0	
4	0	1	1	0	1	0	1	0	
5	1	0	0	0	0	1	1	0	
6	1	0	1	0	1	0	1	0	
7	1	1	0	1	0	1	1	**1**	τ_3=(1,1,0)
8	1	1	1	1	1	0	1	**1**	τ_2=(1,1,1)

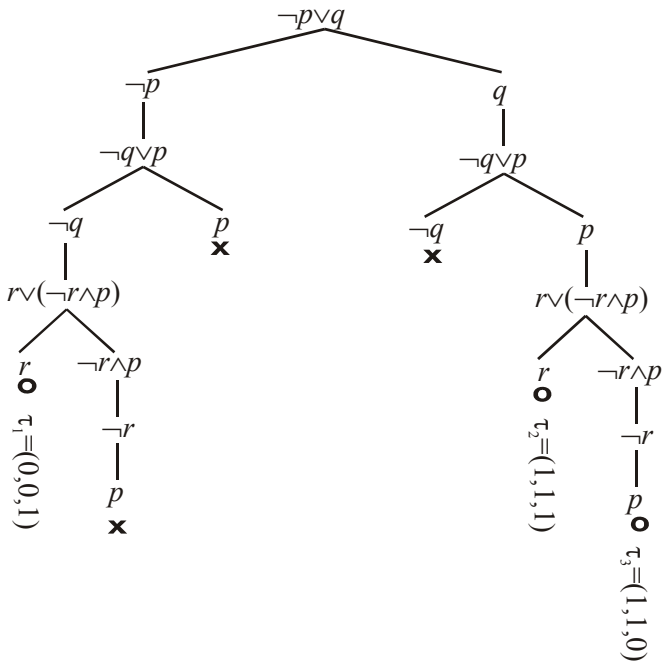
V príklade 4 bola formula $\psi = (p \equiv q) \wedge (r \vee \neg(r \vee \neg p))$ prepísaná do DNF tvaru (Disjunktívna Normálna Forma, ktorá obsahuje disjunkcie konjuktívnych klauzúl). ***Jednotlivé klauzuly z DNF tvaru formuly sú jednoznačne identifikovateľné pomocou vetví sémantického tabla***. Poznamenajme, že v prípade, že nejaká vetva už na počiatku obsahuje dvojicu atomickej premennej a jej negácie, môže byť predčasne uzavretá, pretože ďalšie predlžovanie tejto väzby už nemôže zmeniť pravdivostnú hodnotu.

Význam sémantického tabla je založený na nasledujúcej vete:

Formula ψ je tautologickým dôsledkom formúl $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$, $\Phi \models \psi$, vtedy a len vtedy, ak sémantické tablo zostrojené pre formuly $\{\varphi_1, \varphi_2, ..., \varphi_n, \neg\psi\}$ obsahuje len uzavreté vetvy.

Táto veta nám umožňuje verifikovať či formula ψ je tautologickým dôsledkom teórie s modelom $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$. Podstatne zaujímavší problém je riešenie relácie $\{\varphi_1, ..., \varphi_n\} \models ?$, t. j. hľadáme taký záver ψ, ktorý je tautologickým dôsledkom teórie Φ. Riešenie ψ zostrojíme pomocou nasledujúceho postupu:

Zostrojíme sémantické tablo pre formuly z teórie $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$, pričom predpokladáme, že existuje aspoň jedna otvorená vetva tabla (t. j. teória Φ je konzistentná). Potom riešenie ψ, $\Phi \models \psi$, zostrojíme ako disjunkciu konjunkcií literálov pre vrcholy otvorených vetví.



Obrázok 2. Konštrukcia sémantického tabla pre formulu $\psi = (p \equiv q) \wedge (r \vee \neg(r \vee \neg p))$. Vetvy tabla, ktoré sú označené symbolom × sa neuvažujú, pretože reprezentujú nepravdivé klauzuly. Vetvy označené symbolom ○ reprezentujú klauzuly s pravdivou interpretáciou (pozri tab. 4).

Príklad 5. Pomocou sémantického tabla budeme verifikovať reláciu tautologického dôsledku $\{p \Rightarrow q, q \Rightarrow r\} \models (p \Rightarrow r)$. Pomocou sémantického tabla budeme študovať konjunkciu $(p \Rightarrow q) \wedge (q \Rightarrow r) \wedge \neg(p \Rightarrow r)$, ak sa nám podarí ukázať, že príslušné sémantické tablo má všetky vetvy uzavreté, potom platí relácia $\{p \Rightarrow q, q \Rightarrow r\} \models (p \Rightarrow r)$, pozri obr. 3.

Príklad 6. Pomocou sémantického tabla budeme hľadať záver z teórie $\Phi = \{p \Rightarrow q, p \Rightarrow r\}$, t. j. budeme riešiť reláciu $\Phi = \{p \Rightarrow q, p \Rightarrow r\} \vdash ?$. Výsledky sú znázornené na obr. 4, potom

teória $\Phi = \{p \Rightarrow q, p \Rightarrow r\}$ má štyri rôzne interpretácie – modely, pre ktoré sú predpoklady teórie pravdivé

$$\tau_1 = (p/0, q/?, r/?)$$
$$\tau_2 = (p/0, q/?, r/1)$$
$$\tau_3 = (p/0, q/1, r/?)$$
$$\tau_4 = (p/?, q/1, r/1)$$

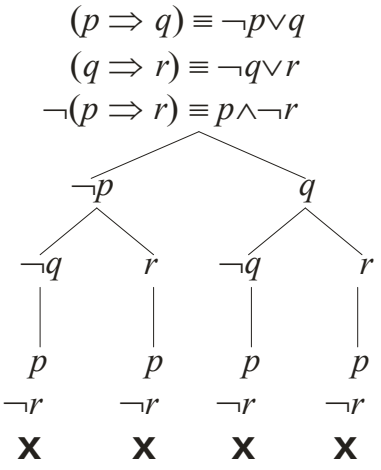
kde symbol ′?′ znamená, že pravdivostná hodnota danej premennej nie je špecifikovaná (čiže môže byť ľubovoľná). Každému modelu priradíme riešenie, ktoré je pravdivé

$$\psi_1 = \neg p$$
$$\psi_2 = \neg p \wedge r$$
$$\psi_3 = \neg p \wedge q$$
$$\psi_4 = q \wedge r$$

To znamená, že máme štyri riešenia relácie $\Phi = \{p \Rightarrow q, p \Rightarrow r\} \vdash ?$ v tvare $\Phi = \{p \Rightarrow q, p \Rightarrow r\} \vdash \psi_i$, pre i = 1, 2, 3, 4. Tieto riešenia môžeme „skladať“ pomocou disjunkcie do nového riešenia, disjunkciou všetkých ψ_i dostaneme riešenie, ktoré je pravdivé pre všetky modely teórie $\Phi = \{p \Rightarrow q, p \Rightarrow r\}$

$$\begin{aligned}\psi' &= \neg p \vee (\neg p \wedge r) \vee (\neg p \wedge q) \vee (q \wedge r) \\ &\equiv \neg p \wedge \underbrace{(1 \vee r \vee q)}_{1} \vee (q \wedge r) \\ &\equiv \neg p \vee (q \wedge r) \equiv p \Rightarrow (q \wedge r)\end{aligned}$$

Potom platí $\{p \Rightarrow q, p \Rightarrow r\} \models (p \Rightarrow q \wedge r)$. Poznamenajme, že toto riešenie je ľahko zostrojiteľné aj pomocou prirodzenej dedukcie. keď množina predpokladov je rozšírená o dodatočný predpoklad p, $\{p \Rightarrow q, p \Rightarrow r, p\} \vdash (q \wedge r)$.



Obrázok 3. Sémantické tablo pre $(p \Rightarrow q) \wedge (q \Rightarrow r) \wedge \neg(p \Rightarrow r)$ (pozri príklad 5), kde každá vetva je uzavretá, potom relácia tautologického dôsledku $\{p \Rightarrow q, q \Rightarrow r\} \models (p \Rightarrow r)$ je platná.

Príklad 7. Pomocou sémantického tabla budeme hľadať riešenie, ktoré vyplýva z $\Phi = \{p \Rightarrow q, q \Rightarrow r\}$, t. j. budeme riešiť reláciu $\Phi = \{p \Rightarrow q, q \Rightarrow r\} \vdash ?$. Výsledky sú znázornené na obr. 5

Zo sémantického tabla vyplýva, že otvorené vetvy produkujú tri interpretácie (modely)

$$\tau_1 = (p/0, q/0, r/?)$$
$$\tau_2 = (p/0, q/?, r/1)$$
$$\tau_3 = (p/?, q/1, r/1)$$

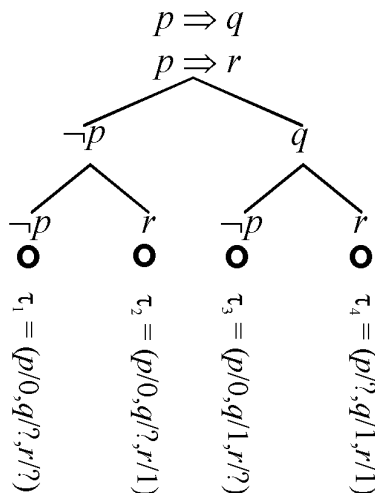
Každému modelu môžeme pripísať riešenie, ktoré je pravdivé

$$\psi_1 = \neg p \wedge \neg q$$
$$\psi_2 = \neg p \wedge r$$
$$\psi_3 = q \wedge r$$

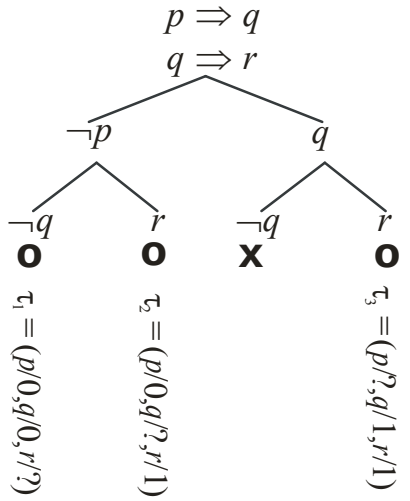
Ak tieto tri nezávislé riešenia spojíme pomocou disjunkcie

$$\psi = (\neg p \wedge \neg q) \vee (\neg p \wedge r) \vee (q \wedge r) \equiv \left(\neg p \wedge \underbrace{(q \Rightarrow r)}_{1} \right) \vee \left(r \wedge \underbrace{(p \Rightarrow q)}_{1} \right) \equiv \neg p \vee r \equiv p \Rightarrow r$$

Týmto sme dokázali, že platí $\Phi = \{p \Rightarrow q, q \Rightarrow r\} \vdash (p \Rightarrow r)$, čo nie je nič iné, ako hypotetický sylogizmus.



Obrázok 4. Sémantické tablo pre teóriu $\Phi = \{p \Rightarrow q, p \Rightarrow r\}$



Obrázok 5. Sémantické tablo pre teóriu $\Phi = \{p \Rightarrow q, q \Rightarrow r\}$

Na záver môžeme konštatovať, že sémantické tablá poskytujú jednoduchý a efektívny prostriedok pre kontrolu vzťahu „tautologického vyplývania“, $\Phi \models \psi$, pričom nemusíme poznať relatívne zložitú syntaktickú teóriu dôkazu „logického vyplývania“ . Taktiež, sémantické tablo je vhodnou technikou na riešenie relácie tautologického vyplývania $\{\varphi_1,...,\varphi_n\} \models ?$, pomocou „partikulárnych“ riešení pre jednotlivé otvorené vetvy dostaneme riešenie, ktoré je tautologickým dôsledkom teórie $\Phi = \{\varphi_1,...,\varphi_n\}$.

7.7 Záver

Tvorba mentálneho modelu (mentálne modelovanie) patrí v kognitívnej vede medzi elementárne aktivity pri modelovaní kognitívnych aktivít. Koncepcia mentálneho modelovania poskytuje kognitívnej vede črty experimentálnej prírodovedy. Pomocou teoretického argumentačného aparátu (väčšinou prevzatého z umelej inteligencie) interpretuje experimentálne merania kognitívnej vedy (kognitívnej psychológie) spôsobom, ktorý silne pripomína vzťah medzi experimentom a teóriou v prírodovede. Mentálny model pre danú kognitívnu aktivitu vystupuje v úlohe jej teórie, pričom kritériom platnosti a efektívnosti danej teórie – mentálneho modelu, je súhlas jeho predikcií s experimentálnymi pozorovaniami, výsledkami pozorovaní kognitívnej psychológie. Požaduje sa, aby mentálne modely boli schopné postihnúť nielen správne riešenia ale aj interpretovať chyby pri konštrukcii riešenia. Tieto modely majú určitú zložitosť, ktorá je premenlivá pri aplikovaní na rôzne problémy (napr. na rôzne kategorické sylogizmy) (Johnson – Laird, 1983), pričom očakávame, že táto zložitosť modelu koreluje s výskytom chýb.

V tomto článku sme venovali pozornosť konštrukcii mentálnych modelov jednoduchej výrokovej logiky. K ich konštrukcii boli použité dva rôzne prístupy:
(1) ***Syntaktický prístup***, kde mentálny model je totožný s Gentzenovým systémom prirodzenej dedukcie, ktorý obsahuje okolo tuctu elementárnych pravidiel usudzovania, ktoré sú totožné so známymi zákonmi logiky (modus ponens a modus tollens, hypotetický sylogizmus, pravidlo inverzie implikácie a pod.). Tento mentálny model sa aplikuje užívateľmi, ktorí sú trénovaní vo formálnej logike aspoň na takej úrovni, že absolvovali základné logické vzdelanie na úrovni úvodu do výrokovej logiky.

(2) ***Sémantický prístup***, kde mentálny model je založený na sémantických tablách. Tento prístup požaduje od užívateľa len schopnosť korektne pravdivostne (sémanticky) interpretovať elementárne logické spojky, k čomu dochádza podľa Inhelderovej a Piageta u detí vo veku okolo 10 rokov. Môžeme teda konštatovať, že tento sémantický prístup ku konštrukcii mentálneho modelu logiky je aplikovaný užívateľmi, ktorí neabsolvovali tréning vo výrokovej logike, ktorých jedinou kognitívnou schopnosťou je korektne pravdivostne (sémanticky) interpretovať elementárne logické spojky. Samozrejme, v tomto prístupe neustále pretrvávajú problémy s korektnou interpretáciou implikácie, ktoré môžu byť zdrojom chýb (experimentálne pozorovaných) v usudzovaní.

Literatúra

[1] Cosmides, L., Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, J., Cosmides, L., Tooby, J. (eds.). *The adapted mind*. Oxford University Press, New York.

[2] Gahér, F. (1998): *Logika pre každého*. IRIS, Bratislava.

[3] Gigerenzer, G., Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox*. MIT Press, Cambridge, (MA).

[4] Inhelder, B., Piaget, J. (1964). *The Early Growth of Logic in the Child: Classification and Seriation.* Routledge and Kegan Paul, London.

[5] Jirků, P., Vejnarová. J. (2000). *Logika - Neformální úvod do formální logiky*. VŠE, Praha.

[6] Johnson – Laird, P.N. (1983). *Mental Models. Toward a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.

[7] Kvasnička V., Pospíchal, J. (2006): *Matematická logika*. Vydavateľstvo STU, Bratislava.

[8] Luria, A. R. (1979). *The Making of Mind* (Cole, M., Cole, S., eds.). Harvard University Press, Cambridge (MA).

[9] Návrat, P., a kolekt. (2001). *Umelá inteligencia*. Vydavateľstvo STU, Bratislava.

[10] Peregrin, J (2004). *Logika a logiky*. Academia, Praha.

[11] Russel, S., Norvig, P.(1995). *Artificial Intelligence. A Modern Approach*. Prentice Hall, New Jersey.

[12] Smullyan, R. (1968). *First Order-Logic*. Springer Verlag, New York (existuje slovenský preklad knihy z r. 1979: *Logika prvého rádu*. Alfa, Bratislava).

[13] Sochor, A. (2001). *Klasická matematická logika*. Karolinum, Praha.

[14] Stenning, K., van Lambalgen, M. (2007). *Semantics and cognition: Semantic insight into the psychology of reasoning* (to be published).

[15] Gentzen, G. (1935). Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* **39**, 405-431. Anglický preklad: Investigations into logical deduction, in *The Collected Papers of Gerhard Gentzen*, edited by Szabo, M. E. (ed.) (1969). North-Holland, Amsterdam, pp. 68-131.

[16] Švejdar, V. (2002). *Logika: neúplnost, složitost a nutnost*. Academia, Praha.

[17] Turney, P., Whitley, D., Anderson, R.W. (eds.) (1996). Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect. A special issue of *Evolutionary Computation* **4**, No. 3.

[18] Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* **20**, 273-281.

8. prednáška

Teória argumentácie

8.1 Argumenty

Argumentácia patrí medzi dôležité aspekty ľudskej inteligencie. Integrálnou súčasťou rôznych profesionálnych aktivít je porovnávanie a analyzovanie argumentácie, hľadanie jej kľúčových zložiek, ktoré sú dôležité pre jej dôkaz alebo vyvrátenie. V súčasnej anglosaskej kultúre existuje už dlhodobá tradícia pestovať argumentáciu v rámci rétoriky a teoretických základov právnych vied, kde sú študované formálne základy argumentácie, t. j. ako je reprezentovaný argument, čo je protiargument k danému argumentu, ako sa vytvára proti-protiargument, aká je závislosť medzi argumentom a protiargumentom, a pod. Teória argumentácie sa do nedávna rozvíjala nezávisle [xx] od logiky, jej pojmového a argumentačného aparátu; bola vytváraná na filozofujúcej fenomenologickej úrovni, pričom poznatky logiky boli zväčša ignorované. V priebehu ostatných 20-30 rokov bol tento tradicionalistický fenomenologický prístup [xx] nahradený aplikáciami metód matematickej logiky, ktoré tvoria efektívny rámec pre vznik logickej teórie argumentácie. Môžeme konštatovať, že moderná teória argumentácie [xx] je aplikácia matematickej logiky k špecifikácii a analýze jednotlivých pojmov, techník a metód argumentácie. Tento prístup umožňuje v rámci matematickej logiky študovať teóriu argumentácie ako aplikovanú logickú disciplínu.

Pod ***argumentom*** budeme rozumieť logickú štruktúru, ktorá obsahuje:

- poznatkovú databázu (***podporu***) reprezentovanú množinou formúl $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\} \subseteq \Delta$, kde univerzálna množina $\Delta = \{\varphi, \varphi', ...\}$ obsahuje formuly, ktoré nie sú kontradikcie, pričom sa nepožaduje, aby táto množina bola konzistentná
- ***metódu inferencie***, pomocou ktorej zostrojíme pomocou formúl z podpory Φ
- ***záver*** α.

Inferenciu môžeme formálne vyjadriť pomocou schémy

$$\frac{\varphi_1, \varphi_2, ..., \varphi_n}{\alpha} \tag{8.1}$$

V hornej časti tejto schémy sú predpoklady a v dolnej časti je záver. Typ inferencie môže byť veľmi rôznorodý, napr. deduktívna inferencia, analogická inferencie, abdukcia, indukcia a pod. V tejto kapitole sa budeme zaoberať len klasickou deduktívnou inferenciou, ktorá je založená na zásadách klasickej výrokovej logiky. To znamená, že naša východisková pozícia pri konštrukcii teórie usudzovania je, že deduktívna schéma usudzovania (8.1), ktorú môžeme stotožniť s nejakou schémou usudzovania prirodzenej dedukcie z kapitoly 4.1. Môžeme teda konštatovať, že záver α je logickým dôsledkom predpokladov z množiny Φ, $\Phi \vdash \alpha$.

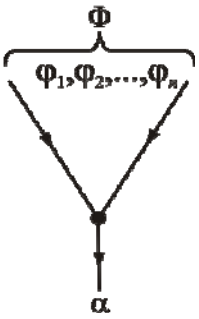
Definícia 8.1. ***Argument*** je usporiadaná dvojica $A = (\Phi, \alpha)$, kde $\Phi \subseteq \Delta$ je podmnožina formúl z Δ, ktorá vyhovuje podmienkam:

(1) $\Phi \nvdash \bot$ (podmienka konzistentnosti),

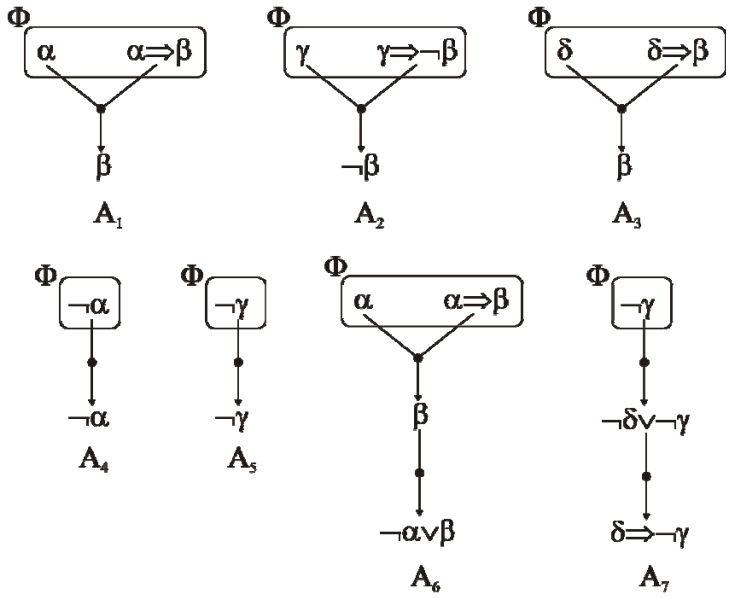
(2) $\Phi \vdash \alpha$ (z podmnožiny Φ logicky vyplýva formula α),

(3) Φ je minimálna množina vyhovujúca podmienke (2) (stačí, ak odstránime jeden prvok podmnožiny Φ a prestane platiť podmienka $\Phi \vdash \alpha$).

Formula α sa nazýva ***dôsledok*** argumentu a podmnožina Φ sa nazýva ***podpora*** (support) argumentu.



Obrázok 8.1. Schematické znázornenie argumentu špecifikovaného podľa definície 8.1. V hornej časti diagramu sú uvedené východiskové predpoklady argumentu (konzistetná podpora argumentu), v dolnej časti diagramu je uvedený dôsledok (výsledná formula) α, ktorý je odvodený pomocou zásad prirodzenej dedukcie výrokovej logiky (táto postupnosť operácií prirodzenej logiky je reprezentovaná šípkami idúcimi zhora nadol).



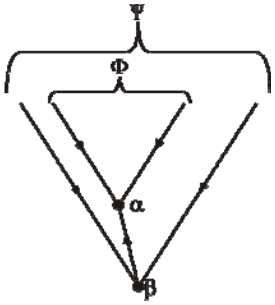
Obrázok 8.2. Znázornenie jednotlivých argumentov A_1-A_7 z príkladu 8.1. Dôsledok každého argumentu je vytvorený pomocou zásad prirodzenej dedukcie aplikovanej na formuly z podpory daného argumentu.

Príklad 8.1. Nech $\Delta = \{\alpha, \alpha \Rightarrow \beta, \gamma \Rightarrow \neg\beta, \gamma, \delta, \delta \Rightarrow \beta, \neg\alpha, \neg\gamma\}$, potom nad touto množinou formúl môžeme definovať tieto argumenty:

- $A_1 = (\{\alpha, \alpha \Rightarrow \beta\}, \beta)$,
- $A_2 = (\{\gamma \Rightarrow \neg\beta, \gamma\}, \neg\beta)$,
- $A_3 = (\{\delta, \delta \Rightarrow \beta\}, \beta)$,
- $A_4 = (\{\neg\alpha\}, \neg\alpha)$,
- $A_5 = (\{\neg\gamma\}, \neg\gamma)$,
- $A_6 = (\{\alpha, \alpha \Rightarrow \beta\}, \neg\alpha \vee \beta)$,
- $A_7 = (\{\neg\gamma\}, \delta \Rightarrow \neg\gamma)$.

Definícia 8.2. Hovoríme, že argument $A=(\Phi,\alpha)$ je ***konzervatívnejší*** ako argument $A'=(\Psi,\beta)$ vtedy a len vtedy, ak $\Phi\subseteq\Psi$ a $\beta\vdash\alpha$.

Obrazne môžeme povedať, že konzervatívnejší argument má všeobecnejší charakter, kladie menšie požiadavky na podporu a menej špecifický na dôsledok α.



Obrázok 8.3. Schematické znázornenie vzťahu konzervatívnosti medzi dvoma argumentami $A=(\Phi,\alpha)$ a $A'=(\Psi,\beta)$. Hovoríme, že argument *A* je konzervatívnejší ako argument *A´* práve vtedy, ak $\Phi\subseteq\Psi$ a formula α je logickým dôsledkom formuly β, $\beta\vdash\alpha$.

Príklad 8.2. Majme dva argumenty $A=\left(\underbrace{\{p,p\Rightarrow q\}}_{\Phi},\underbrace{q}_{\alpha}\right)$ a $A'=\left(\underbrace{\{p,p\Rightarrow q,q\Rightarrow r\}}_{\Psi},\underbrace{q\wedge r}_{\beta}\right)$, pričom platí $\underbrace{q\wedge r}_{\beta}\vdash\underbrace{q}_{\alpha}$. Pretože platí $\Phi\subseteq\Psi$ a $\beta\vdash\alpha$, argument *A* je konzervatívnejší ako argument *A´*. Túto skutočnosť môžeme alternatívne vyjadriť takto

$$\left.\begin{array}{l}(\{p,p\Rightarrow q\},q)\\ \{q,q\Rightarrow r\}\vdash r\wedge q\end{array}\right\}\Rightarrow(\{p,p\Rightarrow q,q\Rightarrow r\},r\wedge q)$$

t. j. rozšírením argumentu $A=(\{p,p\Rightarrow q\},q)$ o ďalší člen $q\Rightarrow r$ dostaneme nový argument $A'=(\{p,p\Rightarrow q,q\Rightarrow r\},r\wedge q)$, pričom v zmysle definície 8.2 argument *A* je konzervatívnejší ako argument *A´*.

Príklad 16.3. Podmienka 2 z definície 16.1 nám zabezpečuje, že zloženie podpory Φ je dostačujúce na to, aby sme dostali požadovaný dôsledok α. Sledujme jednoduchý ilustračný príklad nazývaný v klasickej logike „entyméma“, kde určitá premisa nie je uvedená explicitne, pretože je očividná.

Nech Mariková nemôže byť hosťujúcim mestom zimnej olympiády, pretože je malé mesto, môže byť hosťujúcim mestom olympiády len vtedy, ak je veľkým mestom. Formalizujme tieto poznatky

$$\frac{\begin{array}{l}mar\Rightarrow vm=\text{´Mariková hosťuje olympiádu´}\Rightarrow\text{´Mariková je veľké mesto´}\\ \qquad mm=\text{´Mariková je malé mesto mesto´}\end{array}}{\neg mar=\text{´Mariková nemôže hosťovať olympiádu´}}$$

Podľa klasickej logiky , z uvedených predpokladov nemôžeme dostať uvedený záver, t. j. $\{mar\Rightarrow vm,mm\}\nvdash\neg mar$ Ak predpoklady tejto schémy usudzovania rozšírime o nový (evidentne platný) predpoklad $mm\Rightarrow\neg vm$ *= ´ak Mariková je malé mesto´*, potom *´Mariková*

nie je veľké mesto´, z uvedenej schémy dostaneme požadovaný záver, $\{mm \Rightarrow \neg vm, mar \Rightarrow vm, mm\} \vdash \neg mar$.

Veta 8.1. Výraz (Φ, α) je argumentom vtedy a len vtedy, ak rozšírená podpora $\Phi \cup \{\neg\alpha\}$ je minimálne nekonzistentná.

Dôkaz rozdelíme na dva kroky:
(1⇒) Nech (Φ, α) je argument, potom Φ je konzistentná množina $((\Phi \nvdash \bot) \equiv (\Phi \vdash \top))$ a $\Phi \vdash \alpha$. Potom rozšírenie podpory Φ o negáciu $\neg\alpha$ má za dôsledok, že pôvodne konzistentná množina sa stáva nekonzistentnou, $\Phi \cup \{\neg\alpha\} \vdash \bot$.
(2⇐). Predpokladajme, že rozšírená podpora $\Phi \cup \{\neg\alpha\}$ je minimálne nekonzistentná, ak napr. odstránime z nej formulu $\neg\alpha$, potom podpora Φ je už konzistentná. Taktiež, musí platiť $\Phi \vdash \alpha$, lebo len v tomto prípade je $\Phi \cup \{\neg\alpha\}$ nekonzistentná.

Definícia 8.3. Argumenty (Φ, α) a (Ψ, β) sú ***ekvivalentné*** vtedy a len vtedy, ak množiny Φ a Ψ sú rovné a formuly α a β sú ekvivalentné.

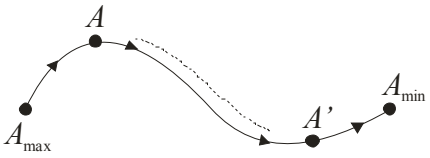
Veta 8.2. Argumenty (Φ, α) a (Ψ, β) sú ***ekvivalentné*** $(A \sim A')$ vtedy a len vtedy, ak argument (Φ, α) je konzervatívnejší ako argument (Ψ, β) a argument (Ψ, β) je konzervatívnejší ako argument (Φ, α).

Dôkaz tejto vety vykonáme v dvoch krokoch:
(1⇒) Nech (Φ, α) je konzervatívnejší ako argument (Ψ, β), potom $\Phi \subseteq \Psi$ a $\beta \vdash \alpha$, čo môžeme prepísať ako tautológiu $\beta \Rightarrow \alpha$.
(2⇐) Nech (Ψ, β) je konzervatívnejší ako argument (Φ, α), potom $\Psi \subseteq \Phi$ a $\alpha \vdash \beta$, čo môžeme opäť prepíšeme ako tautológiu $\alpha \Rightarrow \beta$.
Ak zhrnieme tieto dva výsledky dostaneme $(\Phi \subseteq \Psi) \wedge (\Psi \subseteq \Phi)$ a $(\beta \Rightarrow \alpha) \wedge (\alpha \Rightarrow \beta)$, čo môžeme prepísať ako $\Phi = \Psi$ a $\alpha \equiv \beta$, čo bolo potrebné dokázať.

Nech $\mathcal{A} = \{A, A', ...\}$ je množina argumentov, ktoré sú zostrojené nad rovnakou univerzálnou množinou $\Delta = \{\varphi, \varphi', ...\}$, ukážeme, že množina $\mathcal{A}$ tvorí čiastočne usporiadanú množinu (poset). Definujme binárnu reláciu

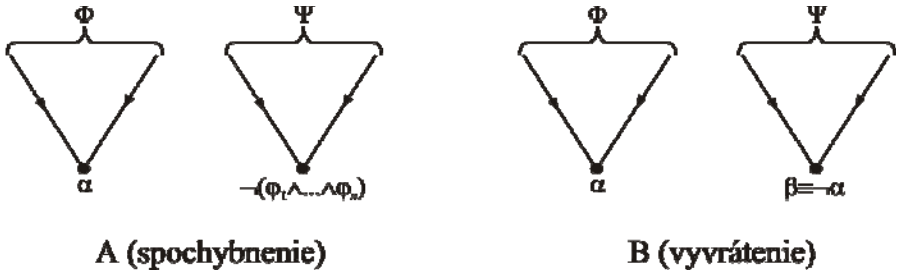
$$A \sqsupseteq A' =_{def} \begin{cases} A \sim A' \\ argument\ A\ je\ konzervatívnejší\ ako\ arg\,ument\ A' \end{cases} \tag{8.2}$$

Pomocou tejto relácie môžeme čiastočne usporiadať množinu $\mathcal{A}$ tak, že ak sú dva argumenty $A, A' \in \mathcal{A}$ sú vo vzájomnej relácii, $A \sqsupseteq A'$, potom ich spojíme orientovanou čiarou vychádzajúcou z argumentu A a končiacou v argumente A'. Ľahko sa ukáže, že takto špecifikovaná relácia, interpretovaná pomocou orientovaného grafu, je tranzitívna a reflexívna. Potom argument A_{max} môžeme nazývať ***maximálne konzervatívny argument***, keď neexistuje taký argument, ktorý by bol konzervatívnejší vzhľadom k argumentu A. Podobne, argument A_{min} môžeme nazývať ***minimálne konzervatívny argument***, keď neexistuje taký argument, ktorý by bol menej konzervatívnejší vzhľadom k argumentu A., pozri obrázok 8.4.



Obrázok 8.4. Diagramatická reprezentácia maximálne konzervatívneho argumentu (A_{max}) a minimálne konzervatívneho argumentu (A_{min}), ktoré sú usporiadané takto: $A_{max} \sqsupseteq A \sqsupseteq ... \sqsupseteq A' \sqsupseteq A_{min}$. Poznamenajme, že dva argumenty A a A' sú spojené orientovanou hranou práve vtedy, ak $A \sqsupseteq A'$. K maximálne konzervatívnemu argumentu neexistuje taký argument, ktorý by bol viac konzervatívnejší ako A_{max}. Podobne k minimálne konzervatívnemu argumentu, ktorý by bol menej konzervatívnejší ako A_{min}.

Centrálnu pozíciu v teórii argumentácie má pojem konfliktu (ataku) jedného argumentu druhým. V „klasickej“ teórii argumentácie sa tieto konflikty medzi argumentmi študujú pomocou relácie, ktorá je definovaná ako podmnožina karteziánskeho súčinu množiny všetkých možných argumentov so sebou samým. Jedná sa o formálny „fenomenologický“ prístup, ktorý ignoruje vnútornú štruktúru argumentov založenú na deduktívnej výrokovej logike. V ďalšom postupe použijeme „nefenomologický“ prístup k štúdiu konfliktov medzi argumentmi, ktorý využíva ich vnútornú logicko-výrokovú štruktúru (pozri obr. 8.5).



Obrázok 8.5. Diagramatická reprezentácia dvoch rôznych konfliktov daného argumentu s iným argumentom. Diagram A reprezentuje konflikt pomocou spochybnenia (ang. *undercut*), kde argument (Ψ,β) spochybňuje argument (Φ,α) tak, že formula β je priamo ekvivalentná $\neg(\varphi_1 \wedge ... \wedge \varphi_n)$, $\Psi \vdash \neg(\varphi_1 \wedge ... \wedge \varphi_n)$, kde $\{\varphi_1,...,\varphi_n\} \subseteq \Phi$, t. j. argument (Ψ,β) je v konflikte s argumentom (Φ,α) tak, že záver β je v kontradikcii aspoň s jedným predpokladom argumentu (Φ,α). Diagram B reprezentuje konflikt pomocou vyvrátenia (angl. *rebuttal*) , kde argument (Ψ,β) je v konflikte s argumentom (Φ,α) tak, že záver β je ekvivalentný negácii záveru α, $\beta \equiv \neg\alpha$. To znamená, že v tomto druhom prípade, jeden argument vyvracia druhý argument tak, že záver β prvého je negáciou záveru druhého argumentu.

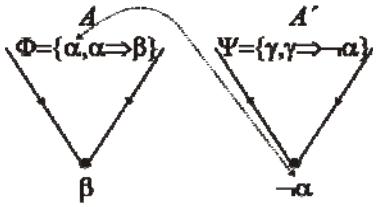
Definícia 8.4.

(1) Hovoríme, že argument (Ψ,β) ***spochybňuje*** (ang. *undercut*) argument (Φ,α) práve vtedy, ak $\beta \equiv \neg(\varphi_1 \wedge ... \wedge \varphi_n)$, pre $\{\varphi_1,...,\varphi_n\} \subseteq \Phi$.

(2) Hovoríme, že argument (Ψ,β) ***vyvrácia*** (ang. *rebuttal*) argumentu (Φ,α) práve vtedy, ak $\beta \equiv \neg\alpha$.

Príklad 8.3. Nech univerzálna množina poznatkov má tvar $\Delta = \{\alpha, \alpha \Rightarrow \beta, \gamma, \gamma \Rightarrow \neg\alpha\}$, z tejto množiny vytvoríme dva argumenty $A = (\{\alpha, \alpha \Rightarrow \beta\}, \beta)$ a $A' = (\{\gamma, \gamma \Rightarrow \neg\alpha\}, \neg\alpha)$. Môžeme povedať, že argument $A' = (\{\gamma, \gamma \Rightarrow \neg\alpha\}, \neg\alpha)$ spochybňuje argument $A = (\{\alpha, \alpha \Rightarrow \beta\}, \beta)$,

pretože záver $\neg\alpha$ argumentu A' atakuje podporu $\Phi = \{\alpha, \alpha \Rightarrow \beta\}$ argumentu $A = (\{\alpha, \alpha \Rightarrow \beta\}, \beta)$, pozri obr. 8.6.



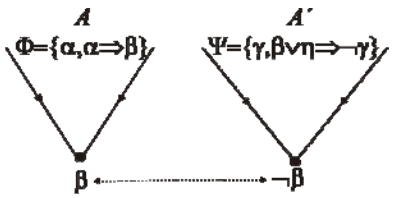
Obrázok 8.6. Schematické znázornenie konfliktu – spochybnenia medzi argumentmi A a A´ z príkladu 8.3.

Príklad 8.4. Nech univerzálna množina poznatkov má tvar $\Delta = \{\alpha, \alpha \Rightarrow \beta, \gamma, \beta \vee \eta \Rightarrow \neg\gamma\}$, z tejto množiny vytvoríme dva argumenty $A = (\{\alpha, \alpha \Rightarrow \beta\}, \beta)$ a $A' = (\{\gamma, \beta \vee \eta \Rightarrow \neg\gamma\}, \neg\beta)$. Pre úplnosť znázorníme pomocou prirodzenej dedukcie odvodenie záveru $\neg\beta$ pre argument A', pozri tabuľku 8.1.

Tabuľka 8.1. Prirodzená dedukciu pre tvorbu dôsledku $\neg\beta$ z predpokladov $\{\gamma, \beta \vee \eta \Rightarrow \neg\gamma\}$ (pozri príklad 8.4)

1	γ	1. predpoklad
2	$\beta \vee \eta \Rightarrow \neg\gamma$	2. predpoklad
3	$\gamma \Rightarrow \neg(\beta \vee \eta)$	aplikácia transpozície implikácie na 2
4	$\gamma \Rightarrow \neg\beta \wedge \neg\eta$	prepis 3 pomocou DeMorganovho zákona
5	$\neg\beta \wedge \neg\eta$	alikácia modus ponens na 1 a 4
6	$\neg\beta$	aplikácia eliminácie konjunkcie na 5

Na obr. 8.7 sú diagramaticky znázornené argumenty A a A', je ukázané, že tieto argumenty sú vo vzájomnom konflikte nazývanom vyvrátenie, ich závery sú vo vzájomnej kontradikcii.



Obrázok 8.7. Schematické znázornenie konfliktu – vyvrátenia medzi argumentmi A a A´ z príkladu 8.4.

Veta 8.3. Argument (Ψ, β) je ***vyvrátenie*** argumentu (Φ, α) vtedy a len vtedy, ak formula $\alpha \equiv \neg\beta$ je tautológiou.

Táto veta je jednoduchým dôsledkom definície konfliktu - vyvrátenia medzi dvojicou argumentov.

Príklad 8.5. Nech $\Delta = \{\alpha, \beta, \neg\alpha, \neg\beta\}$ je univerzálna množina poznatkov, zostrojíme nad ňou tieto argumenty

$$A_0 = (\{\alpha, \beta\}, \alpha \wedge \beta)$$

$$A_1 = \left(\{\neg\alpha\}, \neg(\alpha \wedge \beta)\right)$$
$$A_2 = \left(\{\neg\beta\}, \neg(\alpha \wedge \beta)\right)$$

Argumenty A_1 a A_2 sú spochybnenia argumentu A_0. Pre úplnosť ešte uvedieme v tabuľke 8.2 prirodzené dedukcie pre jednotlivé argumenty, ktoré z predpokladov každého argumentu vedú k jeho záveru.

Tabuľka 8.2. Inferenčné schémy pre jednotlivé argumenty z príkladu 8.5

argument A_0	argument A_1	argument A_2
$\dfrac{\alpha \quad \beta}{\alpha \wedge \beta}$	$\dfrac{\neg\alpha}{\neg\alpha \vee \neg\beta}$ $\neg(\alpha \wedge \beta)$	$\dfrac{\neg\beta}{\neg\alpha \vee \neg\beta}$ $\neg(\alpha \wedge \beta)$

Veta 8.4. Nech argument (Ψ,β) je spochybnenie argumentu (Φ,α), potom aj argument (Φ,α) je spochybnenie argumentu (Ψ,β), t. j. vzťah spochybnenia medzi dvojicou argumentov (Ψ,β) a (Φ,α) je symetrický.

Z predpokladu, že (Ψ,β) je spochybnenie argumentu (Φ,α) vyplýva, pre $\Phi = \{\varphi_1,...,\varphi_n\}$; potom $(\Psi,\beta) = (\Psi, \neg(\varphi_1 \wedge ... \wedge \varphi_n))$. Pre záver β z definície argumentu vyplýva $\Psi \vdash \beta$, alebo $\Psi \vdash \neg(\varphi_1 \wedge ... \wedge \varphi_n)$, čo môžeme prepísať ako implikáciu $\psi_1 \wedge ... \wedge \psi_n \Rightarrow \neg(\varphi_1 \wedge ... \wedge \varphi_n)$, použitím transpozície implikácie dostaneme $\varphi_1 \wedge ... \wedge \varphi_n \Rightarrow \neg(\psi_1 \wedge ... \wedge \psi_n)$. Týmto sme dokázali, že z predpokladu platnosti vety vyplýva $\Phi \vdash \neg(\psi_1 \wedge ... \wedge \psi_n)$. čiže môžeme zostrojiť argument $(\Phi, \neg(\psi_1 \wedge ... \wedge \psi_n))$, ktorý je spochybnenie argumentu (Ψ,β).

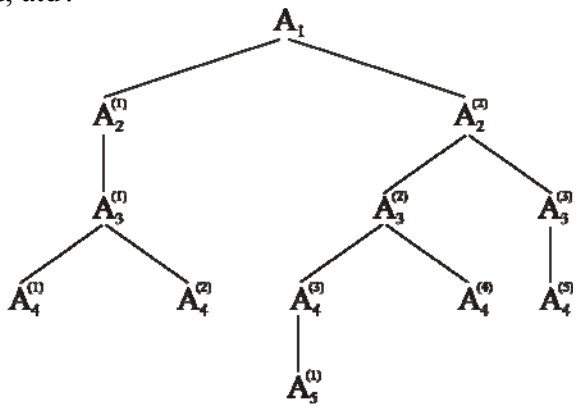
Veta 8.5. Nech argument (Ψ,β) je vyvrátením argumentu (Φ,α), potom aj argument (Φ,α) je vyvrátení argumentu (Ψ,β), t. j. vzťah vyvrátenia medzi dvojicou argumentov (Ψ,β) a (Φ,α) je symetrický.

Z predpokladu, že (Ψ,β) je vyvrátením argumentu (Φ,α) vyplýva ekvivalencia $\beta \equiv \neg\alpha$, potom platí aj $\alpha \equiv \neg\beta$, t. j. argument (Φ,α) je vyvrátením argumentu (Ψ,β), čo bolo potrené dokázať.

8.2. Stromy argumentácie

Pod argumentáciou rozumieme proces, ktorý je zahájený počiatočným argumentom, $A_1 = (\Phi_1, \alpha_1)$, ktorého záver α_1 logicky vyplýva (čo môžeme napr. verifikovať pomocou prirodzenej dedukcie) z podpory Φ_1, formálne $\Phi_1 \vdash \alpha_1$. Námietka (prvý protiargument) je vytvorený pomocou ataku $A_2 = (\Phi_2, \alpha_2)$, kde dôsledok α_2 je v kontradikcii aspoň s jedným predpokladom $\varphi \in \Phi_1$ (spochybnenie), alebo je v kontradikcii so záverom α_1, $\alpha_1 \equiv \neg\alpha_2$. Týchto atakov počiatočného argumentu A_1 môže byť vytvorených niekoľko. Tento proces opakujeme tak dlho, ako je možné vytvárať neekvivalentné ataky argumentov

z predchádzajúcej etapy. Konštrukciu týchto argumentácií a ich protiargumentácií môžeme znázorniť pomocou ***stromu argumentácie*** (pozri obr. 8.8) , ktorý je koreňovým stromom, kde koreňom je počiatočný argument A_1, v nasledujúcej vrstve sú umiestnené protiargumenty k pôvodnému argumentu, atď.



Obrázok 8.8. Schematické znázornenie ***argumentačného stromu***, ktorého vrcholom je počiatočný argument A_1. V nasledujúcej vrstve (idúc zhora nadol) sú umiestnene protiargumenty vzhľadom k počiatočnému argumentu, ktoré sú realizované buď pomocou spochybnení alebo vyvrátení počiatočného argumentu. V nasledujúcej (tretej) vrstve sú umiestnené proti-protiargumenty vzhľadom k protiargumentom z druhej vrstvy, ktoré sú taktiež realizované buď pomocou spochybnení alebo vyvrátení protiargumentov. Tento proces predlžovania argumentačného stromu končí vtedy, keď vzhľadom k argumentom z predchádzajúcej vrstvy sa začnú opakovať protiargumenty už použité v predchádzajúcich vrstvách.

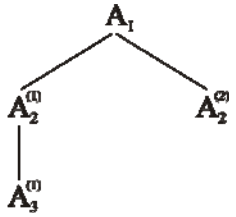
Tabuľka 8.3. Inferenčné schémy
pre jednotlivé argumenty z príkladu 8.6

argument A_1	argument $A_2^{(1)}$	argument $A_2^{(2)}$	argument $A_3^{(1)}$
r $(\neg r \Rightarrow \neg p) \Rightarrow \neg q$ --- $r \vee \neg p$ $\neg r \Rightarrow \neg p$ $\neg q$	s $(p \Rightarrow s) \Rightarrow q$ --- $\neg p \vee s$ $p \Rightarrow s$ q	s $(p \Rightarrow s) \Rightarrow q$ $(\neg r \Rightarrow \neg p) \Rightarrow \neg q$ --- $\neg p \vee s$ $p \Rightarrow s$ q $q \Rightarrow \neg(\neg r \Rightarrow \neg p)$ $\neg(\neg r \Rightarrow \neg p)$ $\neg(p \Rightarrow r)$	$\neg p$ $(\neg r \Rightarrow \neg p) \Rightarrow \neg q$ --- $r \vee \neg p$ $\neg r \Rightarrow \neg p$ $\neg q$

Príklad 8.6. Nech $\Delta = \{r, s, \neg p, (\neg r \Rightarrow \neg p) \Rightarrow \neg q, (p \Rightarrow s) \Rightarrow q\}$. Z tejto univerzálnej databázy poznatkov vytvoríme tieto argumenty

$$A_1 = (\{r, (\neg r \Rightarrow \neg p) \Rightarrow \neg q\}, \neg q),$$
$$A_2^{(1)} = (\{s, (p \Rightarrow s) \Rightarrow q\}, q),$$
$$A_2^{(2)} = (\{s, (p \Rightarrow s) \Rightarrow q, (\neg r \Rightarrow \neg p) \Rightarrow \neg q\}, \neg(p \Rightarrow r)),$$
$$A_3^{(1)} = (\{\neg p, (\neg r \Rightarrow \neg p) \Rightarrow \neg q\}, \neg q).$$

Pre každý argument zostrojíme príslušné schéma usudzovania prirodzenej dedukcie, pomocou ktorého z predpokladov argumentu uvedených v zložených zátvorkách zostrojíme dôsledok (pozri tabuľku 8.3). Strom argumentácie má tvar znázornený na obr. 8.9.



Obrázok 8.9. Strom argumentácie pre argumenty z príkladu 8.6. Koreňom tohto stromu je argument A_1, protiargumenty vzhľadom k tomuto koreňovému argumentu sú argumenty z druhej vrstvy stromu $A_2^{(1)}$ a $A_2^{(2)}$. V tretej vrstve máme jeden argument $A_3^{(1)}$, ktorý je protiargument $A_2^{(1)}$ a proti-protiargument koreňového argumentu A_1. Poznamenajme, že argument $A_2^{(1)}$ je vyvrátením argumentu A_1, argument $A_2^{(2)}$ je spochybnením argumentu A_1. Konečne, argument $A_3^{(1)}$ je spochybnením argumentu $A_2^{(1)}$. Určitú poznámku si zasluhuje naše tvrdenie, že argument $A_2^{(2)}$ je spochybnením argumentu A_1, ako vyplýva z tabuľky 8.3, jeho dôsledok je $\neg(p \Rightarrow r)$, čo môže byť prepísané do ekvivalentného tvaru $p \wedge \neg r$, kde $\neg r$ spochybňuje jeden predpoklad r z podpory koreňového argumentu A_1.

Môžeme si položiť dôležitú otázku, akým spôsobom ukončiť konštrukciu stromu argumentácie? Je jasné, že opakovaním rovnakých sledov argumentov môžeme zväčšovať veľkosť stromu argumentácie do ľubovoľnej veľkosti. Použijeme konvenciu, že v danom slede – ceste argumentov (v strome argumentácie) sa nesmú opakovať argumenty, použitím tejto jednoduchej podmienky dostaneme, že ***veľkosť stromu argumentácie je konečná***.

Ďalším nemej dôležitým problémom je pre danú univerzálnu množinu $\Delta = \{\varphi, \varphi', ...\}$ zostrojiť všetky možné argumenty

$$\mathcal{A} = \{A = (\Phi, \alpha)\,;\, (\Phi \subseteq \Delta) \wedge (\Phi \nvdash \bot) \wedge (\Phi \vdash \alpha)\} \tag{8.3}$$

To znamenám že pre danú podporu $\Phi \subseteq \Delta$ musíme zistiť, či je (1) konzistentná a (2) nájsť takú formulu – záver α, ktorá je (a) logickým dôsledkom podpory, $\Phi \vdash \alpha$, pričom podpora Φ je (b) minimálna vzhľadom k tejto podmienke (pozri definíciu 8.1). Tieto dve podmienky pomerne silne ohraničujú dôsledok α, jeho konštrukcia na základe znalosti podpory Φ je netriviálny problém, ktorého riešenie vyžaduje pomerne sofistikovaný teoretický prístup výrokovej logiky. Zvolíme sémantický prístup k riešeniu tohto problému, t. j. relácia logického vyplývania $\Phi \vdash \alpha$ je nahradená reláciou tautologického vyplývania $\Phi \models \alpha$. Tento prechod od „syntaktického“ prístupu k „sémantickému“ prístupu je umožnený dobre známou skutočnosťou, že výroková logika je korektná a úplná (t. j. $(\Phi \vdash \alpha) \equiv (\Phi \models \alpha)$). V kapitole 3.2.1 je popísané riešene tohto problému konštrukcie dôsledku α ako tautologického vyplývania z podpory – teórie $\Phi = \{\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n\}$ pomocou sémantických tabiel. Túto metódu môžeme rozdeliť do dvoch krokov:

(1) Zostrojíme sémantické tablo pre konjunkciu formúl z podpory, t. j. koreňom sémantického tabla je formula $\varphi = \varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n$. Pomocou otvorených vetiev tohto sémantického tabla vytvoríme množinu modelov podpory $M(\Phi) = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, ..., \boldsymbol{\tau}_a\}$, kde interpretácie – modely $\boldsymbol{\tau}$´s špecifikujú pravdivostné hodnoty výrokových atomických premenných $p_1, p_2, ..., p_n$.

(2) Pomocou formuly (3.3) zostrojíme pre množinu modelov $M(\Phi)$ zostrojíme formulu

$$\alpha = \bigvee_{\tau \in M(\Phi)} p_1^{(\tau_1)} \wedge p_2^{(\tau_2)} \wedge \ldots \wedge p_n^{(\tau_n)} \tag{8.4a}$$

kde pre model $\tau = \left(\tau^{(1)}, \tau^{(2)}, \ldots, \tau^{(n)}\right) \in M\left(\boldsymbol{T}\right)$ nové premenné sú určené takto

$$p_i^{\left(\tau^{(i)}\right)} = \begin{cases} p_i & \left(ak\ \tau^{(i)} = 1\right) \\ \neg p_i & \left(ak\ \tau^{(i)} = 0\right) \\ 1 & \left(ak\ \tau^{(i)} = \#\right) \end{cases} \tag{8.4b}$$

formula α je pravdivá pre každý model $\tau \in M\left(\Phi\right)$, t. j. tautologicky vyplýva z podpory Φ, $\Phi \models \alpha$.

Týmto jednoduchým postupom môžeme pomocou techniky sémantických tabiel zostrojiť dôsledok α pre zvolenú podporu Φ. Vyššie špecifikovaná dvojkroková procedúra zabezpečuje aj platnosť podmienky minimálnosti podpory Φ vzhľadom k dôsledku α. To znamená, že metóda sémantických tabiel nám poskytuje jednoduchú dvojkrokovú procedúru konštrukcie argumentov $\left(\Phi, \alpha\right)$ a teda aj množiny $\mathcal{A}$, pomocou ktorej môžeme pristúpiť ku konštrukcii stromu argumentácií, čo chápeme, ako hlavný cieľ tejto kapitoly.

Tabuľka 8.4. Schémy usudzovania jednotlivých argumentov $A_{0\text{-}4}$ z príkladu 8.7

argument A_0	argument A_1	argument A_2	argument A_3	argument A_4
p $p \Rightarrow q$ $q \Rightarrow r$ --- q r	u $\left(p \Rightarrow q\right) \Rightarrow \neg u$ --- $\neg\left(p \Rightarrow q\right)$	v $\left(q \Rightarrow r\right) \Rightarrow \neg v$ --- $\neg\left(q \Rightarrow r\right)$	u v $u \wedge v \Rightarrow \neg p$ --- $\neg p$	$r \Rightarrow u \wedge v$ $\left(p \Rightarrow q\right) \Rightarrow \neg u$ $p \Rightarrow q$ --- $\neg u$ $\neg u \vee \neg v$ $\neg\left(u \wedge v\right)$ $\neg r$

Príklad 8.7. Nech univerzálna databáza obsahuje tieto formuly – poznatky

$$\Delta = \left\{p, p \Rightarrow q, q \Rightarrow r, u, v, u \wedge v \Rightarrow \neg p, \left(p \Rightarrow q\right) \Rightarrow \neg u, \left(q \Rightarrow r\right) \Rightarrow \neg v, r \Rightarrow u \wedge v\right\}$$

Nad touto univerzálnou databázou zostrojíme tento počiatočný argument

$$A_0 = \left(\left\{p, p \Rightarrow q, q \Rightarrow r\right\}, r\right)$$

Tento argument je spochybnený argumentmi

$$A_1 = \left(\left\{\left(p \Rightarrow q\right) \Rightarrow \neg u, u\right\}, \neg\left(p \Rightarrow q\right)\right)$$

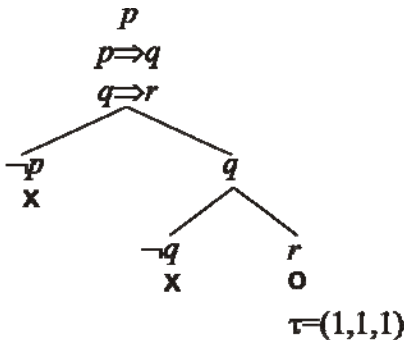
$$A_2 = \left(\left\{\left(q \Rightarrow r\right) \Rightarrow \neg v, v\right\}, \neg\left(q \Rightarrow r\right)\right)$$

$$A_3 = \left(\left\{u, v, u \wedge v \Rightarrow \neg p\right\}, \neg p\right)$$

Vyvrátenie počiatočného argumentu dosiahneme týmto argumentom

$$A_4 = \left(\left\{r \Rightarrow u \wedge v, \left(p \Rightarrow q\right) \Rightarrow \neg u, p \Rightarrow q\right\}, \neg r\right)$$

Schémy usudzovania založené na vybraných argumentoch z Δ sú uvedené v tabuľke 8.4. Na príklade argumentu A_0 vysvetlíme konštrukciu záveru r pomocou sémantického tabla pre konjunkciu formúl z podpory $\Phi = \left\{p, p \Rightarrow q, q \Rightarrow r\right\}$, pozri obr. 8.10.



Obrázok 8.10. Sémantické tablo zostrojene pre konjunkciu $p \wedge (p \Rightarrow q) \wedge (q \Rightarrow r)$ elementov z podpory $\Phi = \{p, p \Rightarrow q, q \Rightarrow r\}$ z príkladu 8.7, ktoré má len jednu vetvu otvorenú. Potom model $M(\Phi)$ obsahuje len jednu interpretáciou $\tau = (1,1,1)$, kde všetky tri výrokové premenné p, q a r z podpory Φ sú pravdivé.

Použitím formuly (8.4) môžeme zostrojiť záver α, ktorý je tautologickým dôsledkom podpory Φ, t. j. $\Phi \models \alpha$

$$\alpha = p \wedge q \wedge r \qquad (*)$$

Predpoklady z Φ obsahujú explicitne p a implicitne q (ako dôsledok aplikácie modus ponens na prvý a druhý predpoklad), potom tieto premenné môžeme formálne pokladať za pravdivé a môžeme zjednodušiť $(*)$ na

$$\alpha = r \qquad (**)$$

Toto riešenie je totožné so záverom argumentu A_0, čo bolo potrebné dokázať.

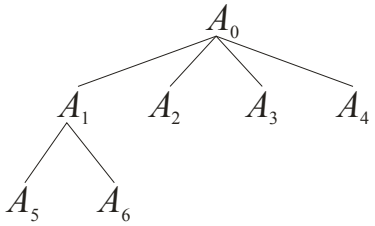
V ďalšom kroku obrátime pozornosť na konštrukciu protiargumentov vzhľadom k argumentu A_1 (t. j. hľadáme proti – protiargumenty vzhľadom k pôvodnému argumentu A_0). Argument A_1 môžeme spochybniť tak, že zostrojíme nový argument, ktorého dôsledok je negáciou vybraného elementu z podpory $\Phi = \{(p \Rightarrow q) \Rightarrow \neg u, u\}$ argumentu A_1. Tak napr, uvažujme nový argument

$$A_5 = (\{p \Rightarrow q, (p \Rightarrow q) \Rightarrow \neg u\}, \neg u)$$

ktorého dôsledok spochybňuje podporu argumentu A_1. Ďalší nový argument

$$A_6 = (\{p \Rightarrow q\}, p \Rightarrow q)$$

vyvracia dôsledok argumentu A_1, $\alpha = \neg(p \Rightarrow q)$, priamo pomocou svojho predpokladu $p \Rightarrow q$. Všetky tieto argumenty nám vytvárajú strom argumentácie znázornený na obr. 8.11.



Obrázok 8.11. Argumentačný strom tvorený vybranými argumentmi z príkladu 8.7.

9. prednáška

Revízia poznatkov

9.1 Úvodné poznámky

Problém revízie poznatkov má vo filozofii a v logike dlhú tradíciu. Stal sa integrálnou súčasťou mnohých traktátov a monografií z filozofie poznania (epistemológie) a logiky od staroveku až po súčasnosť, ktoré obsahovali rozsiahle kapitoly, kde väčšinou na fenomenologicko-špekulatívnej úrovni sa formulovali zásady nášho myslenia, argumentácie a zmeny poznatkov vyplývajúcich zo zmien východiskových predpokladov, ich rozšírenia alebo čiastočnej falzifikácie. Táto zaujímavá problematika sa stala súčasťou aplikovanej matematickej logiky až koncom minulého storočia, kedy jej formálny aparát bol použitý na formuláciu teórie revízie poznatkov. Taktiež, problém revízie poznatkov sa stal v súčasnosti integrálnou časťou informatiky a umelej inteligencie a tým musel prejsť z roviny všeobecno–špekulatívnej k rovine formálno–exaktnej, ktorá môže slúžiť ako základ algoritmizácie problematiky revízie poznatkov na počítačoch.

Tak napríklad, ak chceme špecifikovať základné epistemické princípy vývinu vedy, dynamiku jej vývoja v čase, tento cieľ môže byť realizovaný na abstraktnej úrovni tak, že študujeme danú konzistentnú databázu poznatkov. Táto databáza je v čase postupne modifikovaná elementárnymi operáciami, akými sú dodanie nového poznatku a odstránenie pôvodného poznatku. V obidvoch prípadoch tieto operácie zmeny môžu ovplyvňovať ostatné poznatky, preto sa vykonáva ich revízia, aby sa odstránili prípadné nekonzistentosti. V počiatočnom období vzniku teórie revízie poznatkov (80. roky minulého storočia) základné idey boli formulované švédskym kognitívnym vedcom Petrom Gärdenforsom [xx] a dvojicou amerických logikov Carlosom Alchourrrónom a Davidom Makinsomom [xx], ktorí spoločne v r. 1985 publikovali v *Journal of Symbolic Logic* významnú prácu [xx] (ktorá sa v odbornej literatúre označuje akronymom AGM počiatočných písmen ich priezvisk v abecednom poriadku jej autorov), v ktorej formulovali základné princípy, koncepcie a konštrukcie teórie revízie poznatkov.

Príklad 1. Predpokladajme, že teória obsahuje tieto štyri poznatky:

p_1 = ´*všetky európske labute sú biele*´
p_2 = ´*vták zachytený do siete je labuť*´
p_3 = ´*vták zachytený v sieti pochádza zo Slovenska*´
p_4 = ´*Slovensko je časťou Európy*´

Ak tieto štyri poznatky tvoria vstup do programu simulujúceho logické usudzovanie, potom ako výstup z tohto programu dostaneme nový poznatok

p_5 = ´*vták zachytený v sieti je biely*´

Teraz predpokladajme, že naša databáza bola doplnená o ďalší poznatok

p_6 = ´*vták zachytený v sieti je čierny*´

Tento nový „poznatok“ je v kontradikcii so záverom p_5 , t. j. platí $p_6 = \neg p_5$. Potom musíme vykonať ***revíziu*** databázy, aby sme odstránili túto nekonzistentnosť, t. j. niektorý poznatok z z pôvodnej databázy $\{p_1, p_2, p_3, p_4\}$ musí byť odstránený alebo modifikovaný. Tak napríklad, ak poznatok p_1 nahradíme novým poznatkom

$p_1' =$ *'všetky európske labute sú biele, okrem niektorých zo Slovenska, ktoré sú čierne'*

táto náhrada odstraňuje vyššie spomenutú nekonzistentnosť tak, že jednoznačná špecifikácia v poznatku p_1 o tom, že každá labuť je biela, je nahradená novým poznatkom p_1', že pôvodný predpoklad platí pre labute okrem niektorých slovenských labutí, ktoré sú čierne.

V použitom prístupe k formulácii teórii revízie poznatkov, poznatky sú reprezentované formulami výrokovej logiky, ktoré sú zostrojené nad abecedou atomických výrokov $\mathcal{A} = \{p, q, ..., p', q', ...\}$ a množinou logických spojok $\{\wedge, \vee, \Rightarrow, \equiv, \neg\}$. Jazyk výrokovej logiky zostrojený nad abecedou $\mathcal{A}$ označíme $L_{\mathcal{A}} = \{\varphi, \psi, ..., \varphi', \psi', ..., \mathbf{0}, 1\}$, kde sú uvedené dva symboly (formuly), ktoré reprezentujú kontradikciu resp. tautológiu (napr. $\mathbf{0} \equiv p \wedge \neg p$ a $1 \equiv p \vee \neg p$). Nech $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\} \subseteq L_{\mathcal{A}}$ je množina formúl (teória) výrokovej logiky. Hovoríme, že z teórie $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$ ***logicky vyplýva*** formula ψ, $\Phi \vdash \psi$, práve vtedy, ak existuje postupnosť formúl $\alpha_1, \alpha_2, ..., \alpha_m$ z jazyka $L_{\mathcal{A}}$, pričom pre každé i = 1, 2, ..., m platí jedna z nasledujúcich dvoch podmienok: (1) formula α_i je totožná s nejakou formulou z Φ, alebo (2) formula α_i je logickým dôsledkom formúl $\alpha_1, ..., \alpha_{i-1}$. Teória Φ je ***nekonzistentná*** vtedy a len vtedy, ak existuje taká formula ψ, že súčasne platí $\Phi \vdash \psi$ a $\Phi \vdash \neg\psi$, čo môžeme formálne zapísať pomocou existenčného kvantifikátora, $(\exists\psi)\big((\Phi \vdash \psi) \wedge (\Phi \vdash \neg\psi)\big)$. Potom teória Φ je ***konzistetná*** vtedy a len vtedy, ak pre každú formulu ψ buď platí $\Phi \nvdash \psi$ alebo $\Phi \nvdash \neg\psi$ (nie obe súčasne) , t. j. $(\forall\psi)\big((\Phi \nvdash \psi) \oplus (\Phi \nvdash \neg\psi)\big)$. Nekonzistentnosť teórie Φ vyjadríme symbolom $\Phi \vdash \mathbf{0}$, čo verbálne vyjadríme tak, že teória Φ je nekonzistetná práve vtedy, ak z nej vyplýva formula – kontradikcia, napr. $\mathbf{0} \equiv p \wedge \neg p$. Negáciou tejto definície dostaneme, že teória je konzistetná práve vtedy, ak z nej nevyplýva kontradikcia, $\Phi \nvdash \mathbf{0}$.

Podobné úvahy, ktoré boli vykonané vyššie na syntaktickej úrovni v texte, budú teraz vykonaná aj na sémantickej úrovni. Teória $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\} \subseteq L_{\mathcal{A}}$ sa nazýva konzistentná práve vtedy, ak má neprázdny model, $[\![\Phi]\!] \neq \varnothing$. V opačnom prípade, ak je model prázdny, $[\![\Phi]\!] = \varnothing$, sa nazýva nekonzistentná. Hovoríme, že formula φ ***sémanticky vyplýva*** z teórie Φ práve vtedy, ak je táto formula φ pravdivá pre každú interpretáciu $\tau \in [\![\Phi]\!]$, t. j. $[\![\Phi]\!] \subseteq [\![\varphi]\!]$, vlastnosť tautologického vyplývania formálne zapisujeme $\Phi \vDash \varphi$. V opačnom prípade, neplatnosť inklúzie $[\![\Phi]\!] \not\subseteq [\![\varphi]\!]$ je ekvivalentná vlastnosti $\Phi \nvDash \varphi$, t. j. formula φ sémantický nevyplýva z teórie Φ. Nech Φ a Ψ sú dve teórie, potom model ich zjednotenia $\Phi \cup \Psi$ je prienikom ich modelov

$$[\![\Phi \cup \Psi]\!] = [\![\Phi]\!] \cap [\![\Psi]\!] \tag{1a}$$

Nech Φ a Ψ sú dve teórie, ktoré vyhovujú podmienke $\Phi \subseteq \Psi$, potom vlastnosť monotónnosti implikuje pre ich modely má tvar

$$(\Phi \subseteq \Psi) \Rightarrow ([\![\Phi]\!] \subseteq [\![\Psi]\!]) \tag{1b}$$

Pretože medzi reláciami logického a tautologického vyplývania existuje ekvivalencia je irelevantné, či používame syntaktický alebo sémantický prístup k špecifikácii konzistentnosti teórie alebo vyplývania formuly z teórie, budeme používať oba prístupy v závistlosti od ich vhodnosti k riešeniu daného problému.

Tarského ***operátor konsekvencie*** je definovaný pomocou relácie logického dôsledku

$$Cn(\Phi) =_{def} \{\varphi; \Phi \vdash \varphi\} \tag{2a}$$

kde minimálna množina $Cn(\Phi)$ obsahuje formuly, ktoré sú logickým dôsledkom teórie Φ. Vyhovuje týmto podmienkam

(1) inklúzia: $\Phi \subseteq Cn(\Phi)$, (2b)

(2) monotónnosť: $(\Phi \subseteq \Phi') \Rightarrow (Cn(\Phi) \subseteq Cn(\Phi'))$, (2c)

(3) idempotentnosť: $Cn(\Phi) = Cn(Cn(\Phi))$, (2d)

(4) ak formula φ je logicky odvodená z teórie Φ, $\Phi \vdash \varphi$,
potom $\varphi \in Cn(\Phi)$, (2e)

(5) nech Φ je ***konečná*** konzistentná teória, potom pre každú podmnožinu $\Phi' \subseteq \Phi$ existuje taká formula $\varphi \in \Phi$, že
$(\forall \Phi' \subseteq \Phi)(\exists \varphi \in \Phi)(\varphi \in Cn(\Phi) \Rightarrow \varphi \in Cn(\Phi'))$. (2f)

Posledné pravidlo (2f) je dôsledkom ***princípu kompaktnosti***, ktoré je platné vo výrokovej logike (pozri vetu 2.4): ak konečná množina Φ je konzistentná, potom ľubovoľná podmnožina (včítane prázdnej) $\Phi' \subseteq \Phi$ je taktiež konzistentná.

Veta 15.1. Vybrané vlastnosti operátora konsekvencie

(1) $Cn(\Phi \cup \Psi) = Cn(\Phi \cup Cn(\Psi))$ (3a)

(2) $\Phi \subseteq \Psi \subseteq Cn(\Phi) \Rightarrow Cn(\Phi) \subseteq Cn(\Psi)$ (3b)

(3) $\psi \in Cn(\Phi \cup \{\alpha_1\}) \wedge \varphi \in Cn(\Phi \cup \{\alpha_2\}) \Rightarrow \varphi \in Cn(\Phi \cup \{\alpha_1, \alpha_2\})$ (3c)

(4) $Cn(\{\varphi \vee \psi\}) = Cn(\{\varphi\}) \cap Cn(\{\psi\})$ (3d)

(5) $Cn(\Phi \cup \{\varphi, \psi\}) = Cn(\Phi \cup \{\varphi \wedge \psi\})$ (3e)

Dôkaz týchto vlastností je analogický dôkazu rôznych dôsledkov vety o dedukcii v v kapitole 2.

Pre danú teóriu Φ zostrojíme ***databázu poznatkov*** (ang. ***belief set***) pomocou operátora konsekvencie

$$K = Cn(\Phi) \tag{4}$$

Z definície (2a) operátora konsekvencie vyplýva, že relácia vyplývania $\Phi \vdash \varphi$ je ekvivalentná podmienke $\varphi \in K$,

$$(\Phi \vdash \varphi) \equiv (\varphi \in K) \tag{5}$$

V zmysle vyššie uvedených definícií, databáza poznatkov K je nekonzistentná (čo budeme zapisovať $K_\perp$) práve vtedy, ak obsahuje každú formulu z jazyka L_A, t. j. $K_\perp = L_A$.

Príklad 2. Nech množina formúl $\Phi = \{p, q\}$ obsahuje dva poznatky výroky p a q. Potom množina $Cn(\Phi)$ má tvar

$$Cn(\Phi) = \{p, q, p \vee \varphi, \psi \vee q, \varphi \Rightarrow p, \psi \Rightarrow q, p \wedge q, ...\}$$

kde φ a ψ sú ľubovoľné formuly. Poznamenajme, že táto množina formúl je nekonečná a spočítateľná; ďalšie a ďalšie formuly, ktoré sú dôsledkom použitej teórie $\Phi = \{p, q\}$ môžeme vytvárať napr. použitím idempotentnosti operátorov konjunkcie a disjunkcie.

Budeme študovať tri typy zmeny teórie Φ (pozri obr. 15.1):

(1) ***Expanzia***, nové poznatky, reprezentované formulami z konzistentnej množiny Ψ sú dodané do konzistentnej teórie Φ, za vzniku novej konzistentnej teórie $\Phi' = \Phi \cup \Psi$, čo indukuje rozšírenie jej logických dôsledkov o nové poznatky, ktoré sú logickým dôsledkom rozšírenej teórie Φ'. Databáza poznatkov priradená rozšírenej teórii Φ' je označená

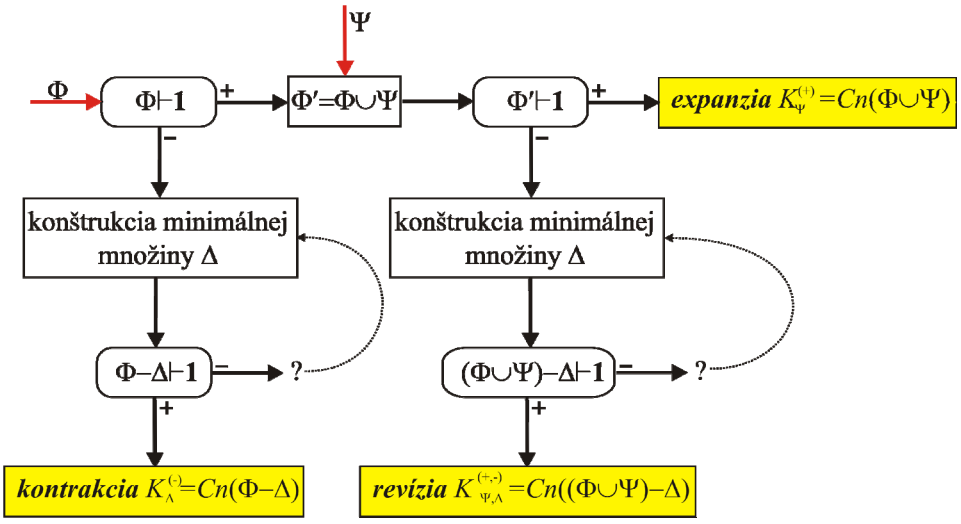
$$K_{\Psi}^{(+)} =_{def} Cn\left(\Phi \cup \Psi\right) \tag{6}$$

(2) ***Kontrakcia***, minimálna podmnožina $\Delta \subseteq \Phi$ je odstránená z nekonzistentnej teórie Φ za vzniku maximálnej konzistentnej podmnožiny – teórie $\Phi' = \Phi - \Delta$. Databáza poznatkov priradená zúženej teórii Φ' je označená

$$K_{\Delta}^{(-)} =_{def} Cn\left(\Phi - \Delta\right) \tag{7}$$

(3) ***Revízia*** je kombináciou predchádzajúcich dvoch operácii expanzie a kontrakcie, pôvodne konzistentná teória Φ je expanziou rozšírená o nové poznatky z Ψ, čím sa stane nekonzistentnou, z takto rozšírenej teórie je pomocou kontrakcie je odstránená podmnožina $\Delta \subseteq \Phi$, výsledkom tohto procesu je konzistentná teória $\Phi' = \left(\Phi \cup \Psi\right) - \Delta$. Pôvodná databáza $K = Cn\left(\Phi\right)$ priradená modifikovanej rozšíreno – zúženej teórii Φ' je označená

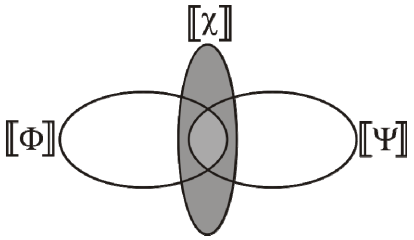
$$K_{\Psi,\Delta}^{(+,-)} =_{def} Cn\left(\left(\Phi \cup \Psi\right) - \Delta\right) \tag{8}$$



Obrázok 9.1. Schématické znázornenie konštrukcie expanzie, kontrakcie a revízie. Hlavný teoretický problém navrhnutej schémy je blok konštrukcie minimálnej množiny Δ, ktorá odstraňuje nekonzistentnosť počiatočnej teórie.

9.2 Expanzia teórie

Operácia expanzie konzistentnej teórie o nové poznatky, pričom konzistentnosť novej rozšírenej teórie sa zachováva, patrí medzi najjednoduchšie operácia zmeny databáz poznatkov. Môžeme povedať, že touto operáciou ja daná databáza neustále rozširovaná o nové a nové poznatky, pričom nemusíme použiť operáciu kontrakcie alebo revízie, pretože nové poznatky sú konzistentné s poznatkami z predchádzajúcich etáp rozširovania databázy poznatkov.



Obrázok 9.2. Diagramatické znázornenie množinovej relácie $[\![\Phi]\!]\cap[\![\Psi]\!]\subseteq[\![\chi]\!]$, ktorá tvorí podmienku pre existenciu tautologického vyplývania $(\Phi\cup\Psi)\models\chi$.

Nech $\Phi=\{\varphi_1,\varphi_2,...,\varphi_n\}$ je konzistentná teória obsahujúca *n* formúl – poznatkov $\varphi_1,\varphi_2,...,\varphi_n$, $[\![\Phi]\!]\neq\varnothing$. Našou úlohou je rozšíriť túto teóriu o množinu nových poznatkov $\Psi=\{\psi_1,\psi_2,...,\psi_m\}$ tak, aby vlastnosť konzistencie zachovala, $\Phi'=\Phi\cup\Psi$, kde $[\![\Phi']\!]=[\![\Phi\cup\Psi]\!]=[\![\Phi]\!]\cap[\![\Psi]\!]\neq\varnothing$. Cieľom tejto expanzie pôvodnej teórie je, aby z rozšírenej teórie vyplývala formula χ, t. j. musí platiť $[\![\Phi]\!]\cap[\![\Psi]\!]\subseteq[\![\chi]\!]$, pozri obr. 9.2

Príklad 2. Nech teória $\Phi=\{\varphi_1,\varphi_2\}$ obsahuje tieto poznatky

$$\varphi_1=\text{´ak }\underbrace{\text{prší}}_{p}\text{, potom }\underbrace{\text{cesta je mokrá}}_{q}\text{´}=(p\Rightarrow q)$$

$$\varphi_2=\text{´ak }\underbrace{\text{cesta je mokrá}}_{q}\text{, potom }\underbrace{\text{jazdím opatrne}}_{r}\text{´}=(q\Rightarrow r)$$

Z tejto teórie Φ vyplýva pomocou hypotetického sylogizmu nový poznatok

$$\chi=p\Rightarrow r=\text{´ak }\underbrace{\text{prší}}_{p}\text{, potom }\underbrace{\text{jazdím opatrne}}_{r}\text{´}$$

Tabuľka 1. Tabuľka pravdivostných hodnôt pre formuly z príkladu 2.

#	p	q	r	$p\Rightarrow q$	$q\Rightarrow r$	$p\Rightarrow r$	$q\wedge r$
1	0	0	0	1	1	1	0
2	0	0	1	1	1	1	0
3	0	1	0	1	0	1	0
4	0	1	1	1	1	1	1
5	1	0	0	0	1	0	0
6	1	0	1	0	1	1	0
7	1	1	0	1	0	0	0
8	1	1	1	1	1	1	1

Pôvodnú teóriu $\Phi=\{\varphi_1,\varphi_2\}$ pomocou expanzie rozšírime o poznatok $\Psi=\{p\}$ (t. j. ´prší´), dostaneme novú rozšírenú teóriu $\Phi'=\Phi\cup\{p\}=\{p,p\Rightarrow q,q\Rightarrow r\}$. Z takto rozšírenej teórie vyplýva nový poznatok

$$\tilde{\chi}=q\wedge r=\text{´cesta je mokra a jazdím opatrne´}$$

Z rozšírenej teórie Φ´ vyplýva taktiež pôvodný poznatok χ, ktorý teraz označíme $\tilde{\chi}'$

$$\tilde{\chi}'=p\Rightarrow r=\text{´ak }\underbrace{\text{prší}}_{p}\text{, potom }\underbrace{\text{jazdím opatrne}}_{r}\text{´}$$

Poznamenajme, že tento záver vyplýva z pôvodnej teórie Φ, čiže musí vyplývať aj z rozšírenej teórie Φ´, t. j. platí implikácia

$$\left(\Phi \subseteq \Phi'\right) \Rightarrow \left(\left(\Phi \models \chi\right) \Rightarrow \left(\Phi' \models \chi\right)\right) \tag{9}$$

Pomocou tabuľky 15.1 môžeme zostrojiť tieto modely:

$$[\![\Phi = \{p \Rightarrow q, q \Rightarrow r\}]\!] = \{\tau_1 = (000), \tau_2 = (001), \tau_4 = (011), \tau_8 = (111)\}$$

$$[\![\Phi']\!] = [\![\Phi]\!] \cap [\![p]\!] = \{p, p \Rightarrow q, q \Rightarrow r\} = \{\tau_8 = (111)\}$$

$$[\![\chi = p \Rightarrow r]\!] = \{\tau_1 = (000), \tau_2 = (001), \tau_3 = (010), \tau_4 = (011), \tau_6 = (101), \tau_8 = (111)\}$$

$$[\![\tilde{\chi} = q \wedge r]\!] = \{\tau_4 = (011), \tau_8 = (111)\}$$

Ľahko skontrolujeme podmienku pre existenciu vyplývania z danej teórie

$$[\![\Phi]\!] \subseteq [\![\chi = p \Rightarrow r]\!]$$

$$[\![\Phi']\!] = [\![\Phi]\!] \cap [\![p]\!] \subseteq [\![\tilde{\chi} = q \wedge r]\!]$$

Kde prvá podmienka – inklúzia implikuje existenciu vyplývania $\Phi \models (p \Rightarrow r)$, zatiaľ čo druhá podmienka – inklúzia implikuje existenciu vyplývania $\Phi' \models q \wedge r$, pozri obrázok 15.2.

Na záver ukážeme prirodzenú dedukciu, ktorá nám potvrdí existenciu záveru $\tilde{\chi} = q \wedge r$ z rozšírenej teórie Φ':

1	p	1. predpoklad
2	$p \Rightarrow q$	2. predpoklad
3	$q \Rightarrow r$	3. predpoklad
4	q	aplikácia modus ponens na 1. a 2. predpoklad
5	r	aplikácia modus ponens na 3. predpoklad a 4. dôsledok
6	$q \wedge r$	spojenie 4. a 5. dôsledku pomocou konjunkcie

To znamená, že sme dokázali pomocou syntaktickej prirodzenej dedukcie, že z rozšírenej teórie Φ' logický vyplýva záver $\tilde{\chi} = q \wedge r$, čo môžeme zapísať ako $\Phi' \vdash q \wedge r$, ktorý môžeme na základe ekvivalentnosti medzi logickým a tautologický vyplývaním prepísať do tvaru $\Phi' \models q \wedge r$, čo bolo potrebné dokázať.

Problém expanzie teórie Φ o formuly z množiny Ψ môžeme formálne chápať ako problém riešenie množinovej rovnice

$$[\![\Phi]\!] \cap [\![\Psi]\!] \subseteq [\![\chi]\!] \tag{10}$$

kde $[\![\Phi]\!]$ a $[\![\chi]\!]$ sú dané modely, model $[\![\Psi]\!]$ musíme zostrojiť tak, aby platila množinová rovnica (10) a súčasne, aby riešenie $[\![\Psi]\!]$ vyhovovalo podmienke $[\![\Phi]\!] \cap [\![\Psi]\!] \neq \emptyset$, t. j. model rozšírenej teórie $\Phi' = \Phi \cup \Psi$ je neprázdny. Poznamenajme, že riešenie rovnice (10) nám ***umožňuje jednoduchú algoritmizáciu*** problému expanzie konzistentnej teórie tak, aby z tejto expanzie vyplývala formula χ.

Príklad 3. Budeme riešiť úlohu 2 pomocou prístupu založenom na riešení množinovej rovnice (10). Nech teória Φ má tvar $\Phi = \{p \Rightarrow q, q \Rightarrow r\}$ a požadovaná formula $\tilde{\chi} = q \wedge r$, pričom hľadáme také rozšírenie danej teórie $\Phi \rightarrow \Phi' = \Phi \cup \Psi$, aby platilo $\Phi' \models \tilde{\chi}$. V príklade 15.2 modely Φ a $\tilde{\chi}$ mali tvar

$$[\![\Phi]\!] = \{(000), (001), (011), (111)\}$$

$$[\![\tilde{\chi}]\!] = \{(011), (111)\} = \{(\#11)\}$$

kde bol použitý „*wilde card*“ symbol #, ktorý sa môže nahradiť tak symbolom 0 ako aj symbolom 1. Ľahko sa presvedčíme, že ak model Ψ má tvar

$$[\![\Psi]\!] = \{(1\#\#)\}$$

potom platí podmienka (10)

$$\underbrace{\{(000),(001),(011),(111)\}}_{[\![\Phi]\!]} \cap \underbrace{\{(1\#\#)\}}_{[\![\Psi]\!]} \subseteq \underbrace{\{(\#11)\}}_{[\![\chi]\!]}$$

To znamená, že pôvodná teória je rozšírená o $\Psi = \{(1\#\#)\} = \{p\}$. Potom sa ľahko presvedčíme, že platí

$$\{p \Rightarrow q, q \Rightarrow r\} \cup \{p\} \vDash q \wedge r$$

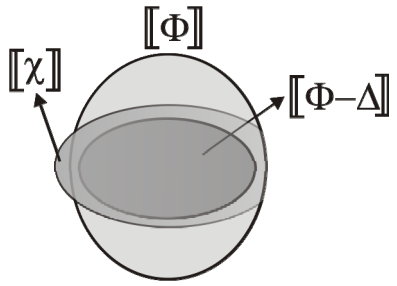
čo bolo potrebné dokázať.

9.3 Kontrakcia teórie

Nech $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$ je nekonzistentná teória, hľadáme takú jej minimálnu podmnožinu $\Delta \subseteq \Phi$, ktorú keď odstránime z pôvodnej teórie Φ, získame novú konzistentnú teóriu $\Phi' = \Phi - \Delta$ z ktorej vyplýva poznatok χ, $\Phi' \vDash \chi$. Táto vlastnosť je splnená práve vtedy, ak platí

$$[\![\Phi - \Delta]\!] \subseteq [\![\chi]\!] \tag{11}$$

kde $\Phi' = \Phi - \Delta$ je nová maximálna konzistentná teória, ktorá vznikla z pôvodnej teórie odobratím poznatkov z Δ, $[\![\Phi - \Delta]\!] \neq \varnothing$. Riešením množinovej rovnice (11) získame takú maximálnu konzistentnú teóriu $\Phi' = \Phi - \Delta$ z ktorej vyplýva daný poznatok χ, pozri obr. 9.3.



Obrázok 9.3.. Kontrakcia pôvodnej teórie Φ vzhľadom k podmnožine $\Delta \subseteq \Phi$, vznikne nová teória $\Phi' = \Phi - \Delta$, podmienka pre tautologické vyplývanie $(\Phi - \Delta) \vDash \chi$ má tvar (11).

Príklad 4. Študujme teóriu tvaru $\Phi = \{p, \neg q, p \Rightarrow q\}$, ktorá je nekonzistentná, $[\![\Phi]\!] = \varnothing$. Jednoduchými úvahami dokážeme, že z tejto teórie vyplýva kontradikcia, $\Phi \vDash q \wedge \neg q$. Vytvoríme tieto štyri tri podmnožiny teórie Φ

$$\Phi'_1 = \Phi - \Delta_1 = \{\neg q, p \Rightarrow q\},\ \Phi'_2 = \Phi - \Delta_2 = \{p \Rightarrow q\},\ \Phi'_3 = \Phi - \Delta_3 = \{p, \neg q\},$$
$$\Phi'_4 = \Phi - \Delta_4 = \{p, p \Rightarrow q\}$$

kde príslušné podmnožiny Δ_i majú tvar

$$\Delta_1 = \{p\}, \Delta_2 = \{p, \neg q\}\ \ \Delta_3 = \{p \Rightarrow q\},\ \Delta_4 = \{\neg q\}$$

a pre nové teórie platí

$$\Phi'_1 \vDash \neg p, \Phi'_2 \vDash p \Rightarrow q,\ \Phi'_3 \vDash \neg q,\ \Phi'_4 \vDash q$$

Tabuľka 2. Pravdivostné hodnoty formúl z príkladu 4

#	p	q	p	$\neg q$	$p \Rightarrow q$	Φ'_i	$[\![\Phi'_i]\!]$	$\Phi'_i \vdash \chi$
1	0	0	0	1	1	$\Phi'_1 = \{\neg q, p \Rightarrow q\}$	$\{(0,\#)\}$	$\chi = \neg p$
2	0	1	0	0	1	$\Phi'_2 = \{p \Rightarrow q\}$	$\{(0,\#),(\#,1)\}$	$\chi = p \Rightarrow q$
3	1	0	1	1	0	$\Phi'_3 = \{p, \neg q\}$	$\{(1,0)\}$	$\chi = p \wedge \neg q$
4	1	1	1	0	1	$\Phi'_4 = \{p, p \Rightarrow q\}$	$\{(\#,1)\}$	$\chi = q$

Sémantický prístup k riešeniu tohto problému je založený na pravdivostnej tabuľke 2, z ktorej okamžite dostávame, že teória Φ je nekonzistentná, t. j. $[\![\Phi]\!] = \varnothing$. Pre všetky možné pravdivostné hodnoty, nikdy nestáva situácia, aby každá formula z teórie bola súčasne pravdivá. Vytvorené „podteórie“ Φ_i majú tieto modely:

$$[\![\Phi'_1]\!] = \{(0,\#)\},\ [\![\Phi'_2]\!] = \{(0,\#),(\#,1)\},\ [\![\Phi'_3]\!] = \{(1,0)\},\ [\![\Phi'_4]\!] = \{(\#,1)\}$$

Pristúpime k „riešeniu“ rovnice inklúzie $[\![\Phi'_i]\!] \subseteq [\![\chi_i]\!]$, dostaneme

$$\Phi'_1 \models \neg p,\ \Phi'_2 \models p \Rightarrow q,\ \Phi'_3 \models \neg q,\ \Phi'_4 \models q$$

Pomocou prirodzenej dedukcie znázorníme prvý výsledok:

1.	$\neg q$	2. predpoklad z Φ
2.	$p \Rightarrow q$	3. predpoklad z Φ
3.	$\neg p$	aplikácia modus tollens na 1. a 2.

Z tohto jednoduchého ilustračného príkladu vyplýva, že proces kontrakcie nie je jednoznačný, dá sa vykonať mnohými spôsobmi, ktoré sú od seba netriviálne odlíšiteľné svojou výhodnosťou alebo nevýhodnosťou. Túto skutočnosť možno pokladať za hlavný tak teoretický, ako aj praktický problém teórie kontrakcie. K odstráneniu tejto nejednoznačnosti sa obvykle používajú mimologické prostriedky, ktoré umožňujú vybrať špecifické predpoklady pôvodnej teórie Φ pre kontrakciu na konzistentnú „podteóriu“ Φ'. V záverečnej časti tejto kapitoly sa k riešeniu tohto problému vrátime pomocou „mimologického“ prístupu nazývaného „***epistemická významnosť***“[1] [xx].

9.4 Revízia teórie

Nech $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$ je konzistentná teória, ktorá je rozšírená o nové poznatky z konzistentnej množiny Ψ, pričom nová teória $\Phi' = \Phi \cup \Psi$ stáva sa nekonzistentnou. Problém revízie potom spočíva v tom, že hľadáme takú minimálnu množinu formúl $\Delta \subseteq \Phi$, ktorú keď odpočítame od Φ', dostaneme maximálnu konzistentnú teóriu $\Phi'' = (\Phi \cup \Psi) - \Delta$, ktorá nie je podmnožinou pôvodnej teórie $\Phi'' \not\subseteq \Phi$. To znamená, že pri kontrakcii sa snažíme uchovať nové poznatky z Ψ a odstraňovať len poznatky z pôvodnej teórie Φ, pozri obrázok

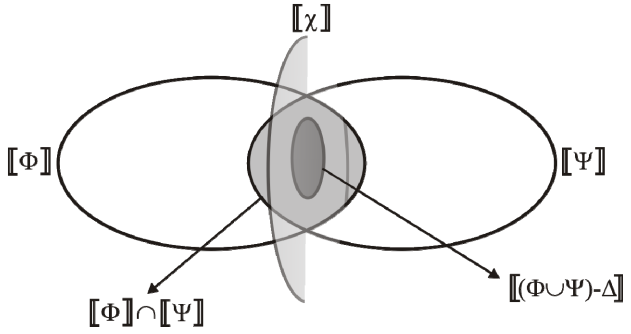
[1] V literatúre sa používa ťažko preložiteľný angl. termín „*epistemic entrenchment*“, čo v prenesenom slova môže byť preložené ako „*epistemická pevnosť – nedobytnosť*“. Z týchto dôvodov používame nový termín „*epistemická významnosť*“, ktorý pomerne presne vystihuje zmysel a význam pôvodného anglického termínu. Poznamenajme, že prístup „epistemickej významnosti“ môžeme chápať ako prejav našej racionálnosti k všeobecnému problému revízie poznatkov mimologickými prostriedkami.

9.4. Z týchto úvodných poznámok opäť vyplýva, že proces revízie nie je jednoznačný a preto musíme použiť „mimologické“ prostriedky na aspoň čiastočné odstránenie tejto nejednoznačnosti.

Podmienka pre existenciu formuly χ, ktorá tautologicky vyplýva z modifikovanej teórie $\big((\Phi\cup\Psi)-\Delta\big)\models\chi$ má tvar množinovej relácie

$$[\![(\Phi\cup\Psi)-\Delta]\!]\subseteq[\![\chi]\!]$$

Pretože $\big((\Phi\cup\Psi)-\Delta\big)\subseteq(\Phi\cup\Psi)$, z monotónnosti modelu (vyjadrenej implikáciou $(A\subseteq B)\Rightarrow[\![A]\!]\subseteq[\![B]\!]$) vyplýva inklúzia $[\![(\Phi\cup\Psi)-\Delta]\!]\subseteq[\![\Phi\cup\Psi]\!]$, tieto podmienky sú znázornené na obrázku 9.3.



Obrázok 9.3. Diagramatické znázornenie podmienky tautologického vyplývania $\big((\Phi\cup\Psi)-\Delta\big)\models\chi$ vyjadrenou pomocou relácie $[\![(\Phi\cup\Psi)-\Delta]\!]\subseteq[\![\chi]\!]$.

Príklad 5. Nech teória Φ má tvar $\Phi=\{p,q,q\Rightarrow p\wedge r,r\Rightarrow s\}$. Táto teória je konzistentná, ľahko môžeme pomocou prirodzenej dedukcie odvodiť rôzne logické vyplývania, napr.

$$\Phi\models(p\Rightarrow r\wedge s),\quad \Phi\models(p\wedge q\wedge r\wedge s)$$

Táto teória Φ je rozšírená o nové poznatky z $\Psi=\{q\Rightarrow\neg p\wedge r,\neg s\}$, dostaneme novú rozšírenú teóriu

$$\Phi'=\{p,q,q\Rightarrow p\wedge r,r\Rightarrow s,q\Rightarrow\neg p\wedge r,\neg s\}$$

ktorá je evidentne nekonzistentná, o čom sa môžeme ľahko presvedčiť pomocou prirodzenej dedukcie

1.	p	1. predpoklad z Φ'
2.	q	2. predpoklad z Φ'
3.	$q\Rightarrow p\wedge r$	3. predpoklad z Φ'
4.	$r\Rightarrow s$	4. predpoklad z Φ'
5.	$q\Rightarrow\neg p\wedge r$	5. predpoklad z Φ'
6.	$\neg s$	6. predpoklad z Φ'
7.	$\neg p\wedge r$	modus ponens na 2. a 5.
8.	$\neg p$	eliminácia konjunkcie na 7.
9.	$p\wedge\neg p$	introdukcia konjunkcie na 1. a 8.

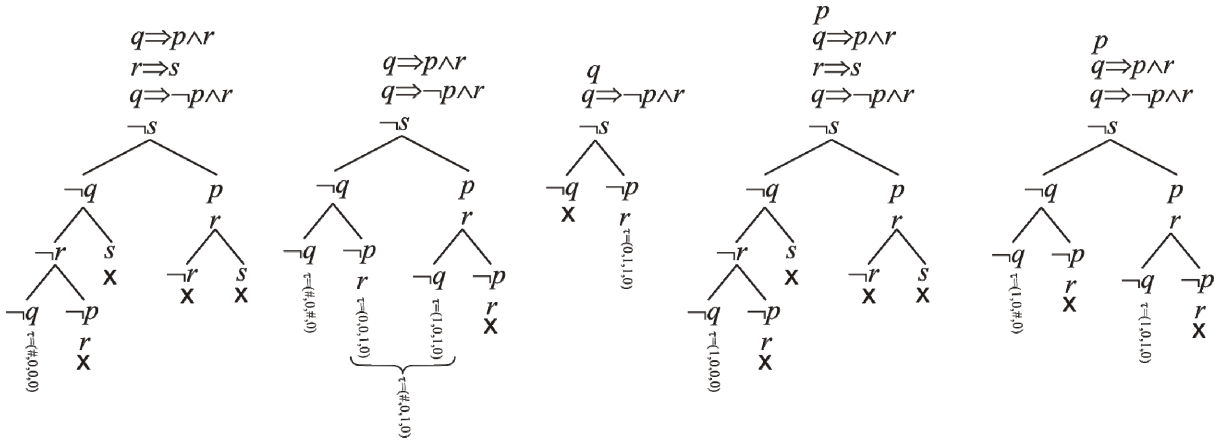
Týmto sme dokázali, že $\Phi'\vdash p\wedge\neg p$, t. j. rozšírená teória ne nekonzistentná.

Pravdivostná tabuľka formúl z rozšírenej teórie Φ' je znázornená tabuľkou 3. V tejto tabuľke je „vysvietených“ päť riadkov, ktoré obsahujú nové formuly z množiny Ψ.

Poznamenajme, že len tieto riadky sa podieľajú na tvorbe alternatívnych „revidovaných“ teórií Φ''_i, pretože už v úvode tejto podkapitoly sme postulovali, že nové poznatky z množiny Ψ sa pri revízii ponechávajú, odstraňujú sa len poznatky z pôvodnej teórie Φ. Päť nových „revidovaných“ teórií je prezentovaných v tabuľke 4.

Tabuľka 3. Tabuľka pravdivostných hodnôt formúl z rozšírenej teórie Φ' z príkladu 5.

	premenné				*pôvodná teória* Φ				*nová teória* Ψ	
#	p	q	r	s	p	q	$q \Rightarrow p \wedge r$	$r \Rightarrow s$	$q \Rightarrow \neg p \wedge r$	$\neg s$
1	0	0	0	0	0	0	1	1	1	1
2	0	0	0	1	0	0	1	1	1	0
3	0	0	1	0	0	0	1	0	1	1
4	0	0	1	1	0	0	1	1	1	0
5	0	1	0	0	0	1	0	1	0	1
6	0	1	0	1	0	1	0	1	0	0
7	0	1	1	0	0	1	0	0	1	1
8	0	1	1	1	0	1	0	1	1	0
9	1	0	0	0	1	0	1	1	1	1
10	1	0	0	1	1	0	1	1	1	0
11	1	0	1	0	1	0	1	0	1	1
12	1	0	1	1	1	0	1	1	1	0
13	1	1	0	0	1	1	0	1	0	1
14	1	1	0	1	1	1	0	1	0	0
15	1	1	1	0	1	1	1	0	0	1
16	1	1	1	1	1	1	1	1	0	0



Obrázok 4. Hľadanie modelov výsledných teórií pomocou techniky sémantických tabiel, ktoré vznikli zjednotením množín Φ a Ψ a následným odstránením vybraných formúl $\Delta \subseteq \Phi$ tak, aby výsledná teória $\Phi'' = (\Phi \cup \Psi) - \Delta$ bola konzistentná

Poznamenajme, že pri konštrukcii formúl χ, ktoré tautologicky vyplývajú z revidovanej teórie Φ'', použili sme heuristiku symbolov #. Obrázok 5 obsahuje sémantické tablá pre funkcie z výslednej konzistentnej množiny Φ'' (pozri prvý stĺpec v tab. 4), pomocou otvorených vietiev týchto tabiel ľahko zostrojíme modely výsledných teórií Φ''.

Ako určiť množinu Δ tak, aby nová konzistetná teória $\Phi'' = (\Phi \cup \Psi) - \Delta$ mala maximálnu kardinalitu, t. j. množina Δ mala minimálnu kardinalitu. Tento výber môžeme ľahko realizovať pomocou tab. 3. V tejto tabuľke nájdeme taký riadok pre formuly z Ψ, ktorý obsahuje len jednotky a pre formuly z Φ obsahuje minimálny počet núl, takýmto riadkom v tab. 3 je riadok 9, potom $\Delta = \{q\}$ a nová konzistentná množina má tvar $\Phi'' = \{p, q \Rightarrow p \wedge r, r \Rightarrow s, q \Rightarrow \neg p \wedge r, \neg s\}$ (tento optimalizovaný výsledok je uvedený v tab 15.4 vo 4 riadku).

Tabuľka 4. Revízia teórie $\Phi'' = (\Phi \cup \Psi) - \Delta$

teória Φ''	odstránená teória Δ	$[\![\Phi'']\!]$
$\Phi''_1 = \{q \Rightarrow p \wedge r, r \Rightarrow s, q \Rightarrow \neg p \wedge r, \neg s\}$	$\Delta_1 = \{p, q\}$	$\tau_1 = (\#000)$
$\Phi''_2 = \{q \Rightarrow p \wedge r, q \Rightarrow \neg p \wedge r, \neg s\}$	$\Delta_2 = \{p, q, r \Rightarrow s\}$	$\tau_1 = (\#0\#0), \tau_2 = (\#010)$
$\Phi''_3 = \{q, q \Rightarrow \neg p \wedge r, \neg s\}$	$\Delta_3 = \{p, q \Rightarrow \neg p \wedge r, \neg s\}$	$\tau = (0110)$
$\Phi''_4 = \{p, q \Rightarrow p \wedge r, r \Rightarrow s, q \Rightarrow \neg p \wedge r, \neg s\}$	$\Delta_4 = \{q\}$	$\tau = (1110)$
$\Phi''_5 = \{p, q \Rightarrow p \wedge r, q \Rightarrow \neg p \wedge r, \neg s\}$	$\Delta_5 = \{q, r \Rightarrow s\}$	$\tau_1 = (10\#0), \tau_2 = (1010)$

9.5 Epistemická významnosť poznatkov

Pri revízii a kontrakcii musíme z množiny poznatkov *K* odstrániť také pôvodné poznatky, aby výsledná množina Φ' bola konzistetná a mala ďalšie požadované vlastnosti. Pretože v rámci výrokovej logiky nepoznáme kritéria významnosti jednotlivých výrokov z danej teórie Φ (alebo databázy poznatkov $K = Cn(\Phi)$), Gärdenfors [xx] fenomelogický zaviedol ako mimologický prostriedok koncepciu epistemickej významnosti výrokov pomocou relácie ′≤′

$\varphi \le \psi$ $=_{def}$ ′***epistemická významnosť'*** φ nie je väčšia ako epistemická významnosť ψ′ (12a)

ktorú môžeme chápať ako binárnu reláciu nad jazykom výrokovej logiky $L_{\mathcal{A}}$, kde výraz $\varphi \le \psi$ je definovaný pre ľubovoľnú dvojicu formúl $\varphi, \psi \in L_{\mathcal{A}}$. Pomocou tejto relácie „usporiadania“ môžeme definovať ďalšie dve odvodené relácie

$$(\varphi = \psi) =_{def} (\varphi \le \psi) \wedge (\psi \le \varphi) \quad (12b)$$

$$(\varphi < \psi) =_{def} (\varphi \le \psi) \wedge \neg(\varphi = \psi) \quad (12c)$$

Takto špecifikovaná relácia ′≤′ má silne kvalitatívny charakter, preto Gärdenfors [xx] postuloval päť podmienok, ktoré musia byť splnené:

Tranzitívnosť: $(\varphi \le \psi) \wedge (\psi \le \chi) \Rightarrow (\varphi \le \chi)$, (13a)

Konjunktívnosť: $(\varphi \le (\varphi \wedge \psi)) \vee (\psi \le (\varphi \wedge \psi))$, (13b)

Dominantnosť: $(\Phi \vdash \psi) \Rightarrow (\exists \varphi \in \Phi)(\varphi \le \psi)$, (13c)

Minimálnosť: $(\varphi \notin Cn(\Phi)) \equiv (\forall \psi \in Cn(\Phi))(\varphi \le \psi)$, (13d)

Maximálnosť: $(\forall \varphi)((\forall \psi)(\psi \le \varphi) \Rightarrow (\vdash \varphi))$ (13e)

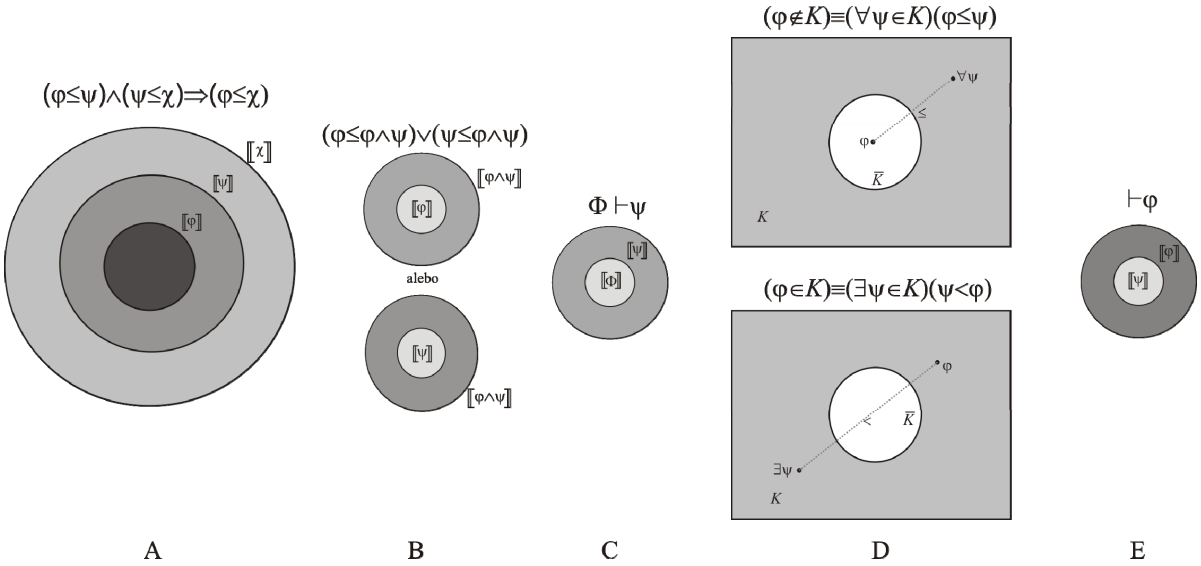
kde $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$. Pomocou prvých troch postulátov môžeme dokázať, že relácia epistemickej významnosti vyhovuje podmienke konektívnosti, $(\forall \varphi, \psi)(\varphi \le \psi \vee \psi \le \varphi)$. Z tretej podmienky (13c) vyplýva, že ak z teórie Φ logicky vyplýva formula ψ, potom musí existovať formula $\varphi \in \Phi$, ktorá nemá väčšiu epistemickú významnosť ako formula ψ. Význam tejto podmienky spočíva v tom, že ak chceme z teórie odstrániť nejakú formulu tak, aby z nej tautologicky nevyplývala formula ψ, tak odstránenie formuly $\varphi \in \Phi$ je postačujúce na dosiahnutie tejto podmienky. Na záver týchto poznámok o relácii epistemickej významnosti poznamenajme, že význam tejto relácie spočíva v ***heuristike, že ak máme odstrániť jeden výrok z dvoch možných kandidátov φ a ψ, odstránime ten, ktorý má menšiu epistemickú významnosť***.

Relácia epistemickej významnosti môže byť v rámci množinového formalizmu dostupných svetov definovaná podľa Groveho [xx] takto:

$$(\varphi \le \psi) =_{def} [\![\varphi]\!] \subseteq [\![\psi]\!] \tag{14}$$

Pre takto špecifikovanú relácie epistemickej významnosti môžeme dokázať podmienky (13).

Dôkaz (13a). Táto vlastnosť je priamym dôsledkom tranzitívnosti množinovej operácie inklúzie, $(A \subseteq B) \wedge (B \subseteq C) \Rightarrow (A \subseteq C)$.



Obrázok 9.5. Interpretácia jednotlivých postulátov (13) pomocou formalizmu dostupných svetov. (A) Postulát tranzitívnosti pre tri výrokové premenné φ, ψ, χ, ktoré spĺňanú podmienky φ≤ψ, ψ≤χ a φ≤χ. (B) Postulát konjunktívnosti, pre ľubovoľné φ, ψ platí φ≤(φ∧ψ) alebo ψ≤(φ∧ψ). (C) Postulát dominantnosti, ak φ⊢ψ, potom φ≤ψ. (D) Postulát minimálnosti, ak φ nepatrí do konzistentnej množiny *K*, potom pre každé ψ z tejto databázy *K* platí relácia φ≤ψ. (E) Postulát maximálnosti, ak pre dané φ a pre každé ψ platí ψ≤φ, potom φ je tautológia.

Dôkaz (13b). Nech $\varphi, \psi \in L_A$ dvojica takých výrokových formúl pre ktoré je platná binárna relácia epistemickej významnosti, t. j. medzi nimi platí vzťah $(\varphi \le \psi) \vee (\psi \le \varphi)$, čo môžeme pomocou definície (14) prepísať do množinovej relácie $([\![\varphi]\!] \subseteq [\![\psi]\!]) \vee ([\![\psi]\!] \subseteq [\![\varphi]\!])$. Tieto dve alternatívne inklúzie platia vtedy ak $([\![\varphi]\!] = [\![\varphi \wedge \psi]\!]) \vee ([\![\psi]\!] = [\![\varphi \wedge \psi]\!])$, čo môžeme prepísať do množinovej formuly

$$\big((\llbracket\varphi\rrbracket \subseteq \llbracket\varphi\wedge\psi\rrbracket)\wedge(\llbracket\varphi\wedge\psi\rrbracket \subseteq \llbracket\varphi\rrbracket)\big)\vee\big((\llbracket\psi\rrbracket \subseteq \llbracket\varphi\wedge\psi\rrbracket)\wedge(\llbracket\varphi\wedge\psi\rrbracket \subseteq \llbracket\psi\rrbracket)\big) \qquad (\clubsuit)$$

Poznamenajme, že druhá a štvrtá inklúzia platia automaticky pre každú dvojicu formúl φ,ψ (platia množinové zákony – tautológie $A\cap B\subseteq A$ a $A\cap B\subseteq B$. K tomu, aby bola pravdivá formula (♣), musí platiť jej prvá a tretia inklúzia

$$(\llbracket\varphi\rrbracket \subseteq \llbracket\varphi\wedge\psi\rrbracket)\vee(\llbracket\psi\rrbracket \subseteq \llbracket\varphi\wedge\psi\rrbracket),$$

Túto formulu môžeme prepísať do tvaru $\big(\varphi\le(\varphi\wedge\psi)\big)\vee\big(\psi\le(\varphi\wedge\psi)\big)$, čo bolo potrebné dokázať.

Dôkaz (13c). Z predpokladu, že ψ je tautologickým dôsledkom teórie Φ, vyplýva, že musí platiť podmienka $\llbracket\Phi\rrbracket\subseteq\llbracket\psi\rrbracket$, alebo $\llbracket\varphi_1\rrbracket\cap\ldots\cap\llbracket\varphi_n\rrbracket\subseteq\llbracket\psi\rrbracket$, alebo $\varphi_1\wedge\ldots\wedge\varphi_n\le\varphi$. Z predchádzajúcej vlastnosti (13b), že existuje také $\varphi_i\in\Phi$, že platí $\varphi_i\le\varphi_1\wedge\ldots\wedge\varphi_n$, alebo $\varphi_i\le\varphi$, čo bolo potrebné dokázať.

Dôkaz (13d). Korektný dôkaz tejto vlastnosti v rámci použitého množinového formalizmu neexistuje. Chceme ukázať, že ak $K=Cn(\Phi)$ je korektná databáza poznatkov, potom podmienka $\varphi\notin K$ je ekvivalentná s $(\forall\psi\in K)(\varphi\le\psi)$, t. j. $(\varphi\notin K)\equiv(\forall\psi\in K)(\varphi\le\psi)$. Pre jednoduchosť predpokladajme, že podmienka $\varphi\notin K$ implikuje, že formula φ je kontradikcia, $\llbracket\varphi\rrbracket=\varnothing$, potom pre každú formulu ψ platí $\varphi\le\psi$. V druhej časti dôkazu predpokladajme, že pre každú formulu ψ platí $\varphi\le\psi$, toto je možné jedine vtedy, aj formula φ je kontradikcia, čiže $\varphi\notin K$ (predpokladáme, že K je konzistentná databáza), čo bolo potrebné dokázať.

Dôkaz (13e). Predpokladajme, že pre danú formulu φ platí $(\forall\psi)(\psi\le\varphi)$, t. j. každá formula ψ je nanajvýš tak epistemicky významná ako formula φ, to je možné len vtedy, ak táto formula je tautológia, $\llbracket\varphi\rrbracket=W$, čo bolo potrebné dokázať.

Grafické interpretácie jednotlivých postulátov (13) sú znázornené na obrázku 15.4. Pristúpime teraz k jednoduchej „racionálnej“ interpretácii týchto jednotlivých postulátov:

(a) ***Podmienka tranzitívnosti***. Táto podmienka (13a) a jej grafické znázornenie na obr. 15.5, diagram A, odráža rozumný predpoklad tranzitívnosti relácie epistemickej významnosti. Grafická interpretácia tejto podmienky spočíva v troch kruhových oblastiach v množina W dostupných svetov, ktoré sú vložené do seba, takže tranzitivita automaticky platí (pozri obr. 15.5, diagram A).

(b) ***Podmienka konjunktívnosti***. Z (13b) vyplýva, že epistemická významnosť konjunkcie $\varphi\wedge\psi$ nie je menšia, ako epistemická významnosť jej komponent φ resp. ψ (pozri obr. 5, diagram B). Táto vlastnosť konjunktívnosti môže byť indukciou zovšeobecnená do tvaru: Nech $\Phi=\{\varphi_1,\ldots,\varphi_n\}$ je teória obsahujúca *n* formúl φ_i, potom platí

$$(\exists\varphi\in\Phi)(\varphi\le\varphi_1\wedge\ldots\wedge\varphi_n) \qquad (15)$$

(c) ***Podmienka dominantnosti.*** Ak platí (13c) , potom skutočnosť, že z teórie Φ logicky vyplýva ψ, $(\Phi\vdash\psi)$, implikuje, že epistemická významnosť niektorého predpokladu φ z teórie Φ je menšia ako epistemická významnosť dôsledku ψ (pozri obr. 5, diagram C).

(d) ***Podmienka minimálnosti.*** Podmienka minimálnosti nám hovorí, že ak formula φ nepatrí do databázy poznatkov *K,* potom táto skutočnosť je ekvivalentná podmienke, že každá formula ψ z tejto databázy nie je menšej epistemickej významnosti ako formula φ. Negovaná podmienka minimálnosti $\left(\left(\varphi \in Cn(\Phi)\right) \equiv \left(\exists \psi \in Cn(\Phi)\right)\left(\psi < \varphi\right)\right)$ hovorí, že vlastnosť, že formula φ patrí do databázy *K*, je ekvivalentná podmienke, že existuje v tejto databáze existuje taká formula ψ, ktorá má menšiu epistemickú významnosť ako formula φ. (pozri obr. 5, diagram D).

(e) ***Podmienka maximálnosti.*** Z (13e) vyplýva heuristika, ak pre dané φ platí, že jeho epistemická významnosť nie je menšia, ako epistemická významnosť ľubovoľného ψ, potom φ je tautológia. To znamená, že tautológie majú významné postavenie (maxím) na povrchu epistemickej významnosti, každé ψ má epistemickú významnosť menšiu alebo rovnú ako tautológia (pozri obr. 5, diagram E).

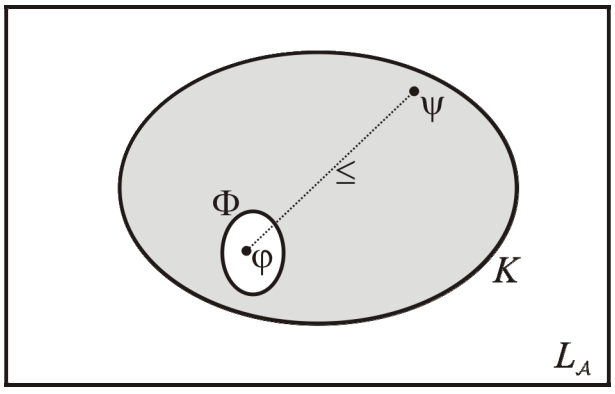
Veta 2.
Nech $K = Cn(\Phi) = \{\psi; \Phi \vdash \psi\}$ je databáza poznatkov priradená teórii Φ, táto databáza môže byť pretransformovaná do tvaru

$$K = \{\psi; (\exists \varphi \in \Phi)(\varphi \leq \psi)\} \tag{16}$$

Dôkaz tejto vety je pomerne jednoduchý, nech $K = \{\psi; \Phi \vdash \varphi\}$, potom platí sled rovností

$$\begin{aligned} K &= \{\psi; \Phi \vdash \psi\} = \{\psi; [\![\Phi]\!] \subseteq [\![\psi]\!]\} = \{\psi; [\![\varphi_1 \wedge \ldots \wedge \varphi_n]\!] \subseteq [\![\psi]\!]\} \\ &= \{\psi; (\exists \varphi \in \Phi)([\![\varphi]\!] \subseteq [\![\varphi_1 \wedge \ldots \wedge \varphi_n]\!] \subseteq [\![\psi]\!])\} = \{\psi; (\exists \varphi \in \Phi)([\![\varphi]\!] \subseteq [\![\psi]\!])\} \\ &= \{\psi; (\exists \varphi \in \Phi)(\varphi \leq \psi)\} \end{aligned}$$

Čo bolo potrebné dokázať. Diagramatická interpretácia tejto vety je znázornená na obrázku 6.



Obrázok 9.6. Diagramatická interpretácia vety 1.

Na záver tejto podkapitoly obrátime našu pozornosť na formuláciu kontrakcie teórie Φ, ktorá je pôvodne nekonzistentná, odstránime z nej vybranú podmnožinu $\Delta \subseteq \Phi$ ta, aby rezultujúca teória $\Phi' = \Phi - \Delta$ bola konzistentná a vyplývala z nej formula χ, t. j. $(\Phi - \Delta) \vdash \chi$. Takto špecifikovaná databáza poznatkov je podľa (7) špecifikovaná takto

$$K_{\Delta}^{(-)} =_{def} Cn(\Phi - \Delta) = \{\psi; (\Phi - \Delta) \vdash \psi\} \tag{17}$$

Použitím vety 1 dostaneme

$$K_{\Delta}^{(-)} = \left\{\psi; \left(\exists \varphi \in \left(\Phi - \Delta\right)\right)\left(\varphi \leq \psi\right)\right\} \tag{18}$$

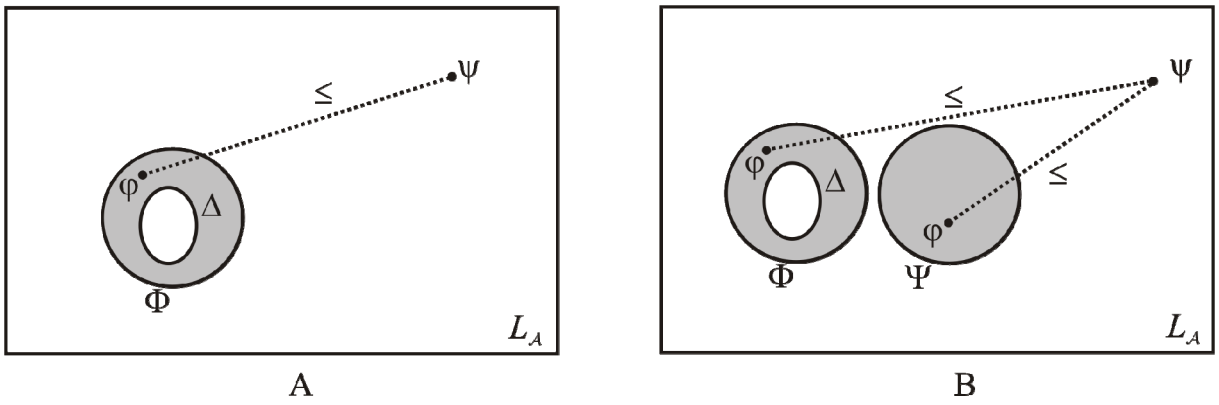
kde sme použili (16).

Úplne analogickým spôsobom môžeme študovať aj operáciu revízie, kde pôvodne konzistentná teória Φ bola rozšírená o nové poznatky teóriou Ψ, čím sa však stala nekonzistentnou, preto v nasledujúcom kroku musí z tejto teórie $\Phi' = \Phi \cup \Psi$ odstrániť niektoré poznatky z pôvodnej teórie Φ. Výsledkom tohto procesu je konzistentná teória $\Phi' = \left(\Phi \cup \Psi\right) - \Delta$, ktorá je priradená novej databáze poznatkov

$$K_{\Psi,\Delta}^{(+,-)} =_{def} Cn\left(\left(\Phi \cup \Psi\right) - \Delta\right) \tag{19}$$

Pomocou (16) ju môžeme prepísať do tvaru

$$K_{\Psi,\Delta}^{(+,-)} = \left\{\psi; \left(\exists \varphi \in \left(\left(\Phi \cup \Psi\right) - \Delta\right)\right)\left(\varphi \leq \psi\right)\right\} \tag{20}$$



Obrázok 9.7. (A) Znázornenie operácie kontrakcie pomocou formuly (18), kde pre každú formula $\psi \in K_{\Delta}^{(-)}$ existuje také $\varphi \in \Phi - \Delta$, že $\varphi \leq \psi$. (B) Znázornenie operácie revízie pomocou formuly (20), kde pre každé $\psi \in K_{\Psi,\Delta}^{(-,+)}$ existuje také $\varphi \in \left(\Phi \cup \Psi\right) - \Delta$, že $\varphi \leq \psi$.

9.6 Spájanie poznatkov

Predstavme si multiagentový systém, kde úlohou každého agenta – prieskumníka je zbierať poznatky o danom prostredí (napr. ako sú rozložené prekážky, potrava a ostatní agenti). Po určitom čase agent sa vráti do „centrály“ a odovzdá svoje poznatky agentovi - integrátorovi, ktorý má na starosti spájanie poznatkov získaných od rôznych agentov – prieskumníkov do jednej databázy poznatkov o danom prostredí. Môže sa stať, že poznatky získané dvoma rôznymi agentmi sú navzájom kontroverzné, t. j. agent – integrátor stojí pred úlohou rozhodnúť sa, ktoré poznatky budú/nebudú použité pri vytváraní spojenej databázy poznatkov.

Už z tohto jednoduchého ilustračného príkladu vyplýva, že problém spájania poznatkov do jednotného väčšieho celku – ***spojenej databázy poznatkov*** – patrí v umelej inteligencii medzi základné problémy distribuovanej inteligencie (napr. multiagentových systémov). Tento problém môžeme byť v rámci výrokovej (alebo aj predikátovej) logiky jednoducho formalizovaný, pričom na sémantickej úrovni môžeme špecifikovať „mimologické“ prostriedky ako riešiť nekonzistentnosti vyskytujúce sa pri spájaní nových poznatkov.

Príklad 6. Nech lekárske konzílium je zložené z troch lekárov A, B a C, cieľom konzília je určiť symptómy a diagnózu pacienta. Každý lekár si vytvoril svoju teóriu

$$\Phi_A = \{s_1, s_2, s_1 \Rightarrow d_1\} \tag{21a}$$

$$\Phi_B = \{s_1, s_1 \Rightarrow d_2, s_2 \Rightarrow d_3\} \tag{21b}$$

$$\Phi_C = \{s_3, s_3 \Rightarrow d_1\} \tag{21d}$$

V týchto troch teóriách sú identifikované tri symptómy s_1, s_2, s_3, tri diagnózy d_1, d_2, d_3, a štyri poznatky reprezentované implikáciami $s_1 \Rightarrow d_1$, $s_1 \Rightarrow d_2$, $s_2 \Rightarrow d_3$, $s_3 \Rightarrow d_1$. Lekári určili rôzne diagnózy d_1, d_2 a d_1, potom pri vytváraní spoločného záveru konzília môžeme vybrať ako záver konzília diagnózu d_1, pretože sa vyskytuje u prvého a tretieho lekára, iná diagnóza d_2 sa vyskytuje len u druhého lekára. Spoločná databáza môže mať tvar

$$\text{prvý spôsob spájania:} \quad \mathcal{M}\left(\tilde{\Phi} = \{\Phi_A, \Phi_B, \Phi_C\}, \chi\right) = \left\{\Phi_A, \Phi_C, \chi^{(1)}\right\} \tag{25a}$$

$$\text{druhý spôsob spájania:} \quad \mathcal{M}\left(\tilde{\Phi} = \{\Phi_A, \Phi_B, \Phi_C\}, \chi\right) = \left\{\Phi_B, \chi^{(2)}\right\} \tag{*}$$

Táto databáza poznatkov môže byť rozšírená ešte o ďalšie diagnózy a implikácie, avšak to by mohlo viesť k porušeniu zásady, že rezultujúca databáza obsahuje len jednu diagnózu.

Pri štúdiu problému spájania teórií z metateórie $\tilde{\Phi} = \{\Phi_1, \Phi_2, ..., \Phi_n\}$ do jednej spoločnej teórie $\mathcal{M}\left(\tilde{\Phi}\right)$ využijeme formalizmus z predchádzajúcej časti tejto kapitoly. Tento formalizmus rozšírime o nový parameter, ktorý špecifikuje tvar spojenej teórie $\mathcal{M}\left(\tilde{\Phi}\right)$, budeme ho realizovať pomocou výrokovej formuly χ, potom spojenú teóriu označíme $\mathcal{M}\left(\tilde{\Phi}, \chi\right)$. Tak napríklad, pre ilustračný príklad 15.6, výsledná (mergovaná) diagnóza je označená $\mathcal{M}\left(\tilde{\Phi} = \{\Phi_A, \Phi_B, \Phi_C\}, \chi\right)$, kde formula χ špecifikuje podmienku, že riešenie obsahuje len jednu diagnózu

$$\chi = \underbrace{\left(d_1 \wedge \neg d_2 \wedge \neg d_3\right)}_{\chi^{(1)}} \vee \underbrace{\left(\neg d_1 \wedge d_2 \wedge \neg d_3\right)}_{\chi^{(2)}} \vee \underbrace{\left(\neg d_1 \wedge \neg d_2 \wedge d_3\right)}_{\chi^{(3)}} = \chi^{(1)} \vee \chi^{(2)} \vee \chi^{(3)} \tag{**}$$

ktorá je pravdivá vtedy, ak práve jedna diagnóza je pravdivá a ostatné dve sú nepravdivé.

Príklad 7. Budeme pokračovať v riešení príkladu 6. K tomu, aby sme mohli vytvoriť spojenie $\mathcal{M}\left(\tilde{\Phi} = \{\Phi_A, \Phi_B, \Phi_C\}, \chi\right)$ musíme vedieť, či $\{\Phi_A, \Phi_B, \Phi_C\} \cup \{\chi\}$ je konzistentná teória alebo nie je. V prípade, že nie je konzistentná, potom musíme pristúpiť k revízii $\{\Phi_A, \Phi_B, \Phi_C\} \cup \{\chi\}$ a odstrániť minimálny počet členov, ktoré spôsobujú nekonzistentnosť. Model $[\![\{\Phi_A, \Phi_B, \Phi_C\} \cup \{\chi\}]\!]$ môžeme prepísať do tvaru

$$[\![\{\Phi_A, \Phi_B, \Phi_C\} \cup \{\chi\}]\!] = [\![\left\{\Phi_A, \chi^{(1)}\right\}]\!] \cap [\![\left\{\Phi_B, \chi^{(2)}\right\}]\!] \cap [\![\left\{\Phi_C, \chi^{(1)}\right\}]\!] \tag{22a}$$

kde sme využili tieto dve vlastnosti: $[\![X \cup Y]\!] = [\![X]\!] \cap [\![Y]\!]$ a $[\![X]\!] = [\![X]\!] \cap [\![X]\!]$. Podľa (21) je teória $\{\Phi_A, \Phi_B, \Phi_C\} \cup \{\chi\}$ konzistentná práve vtedy, ak dvojice $\left\{\Phi_A, \chi^{(1)}\right\}$, $\left\{\Phi_B, \chi^{(2)}\right\}$, $\left\{\Phi_C, \chi^{(1)}\right\}$ sú konzistentné (ktoré vznikli rozšírením pôvodných teórií Φ_X o ohraničujúcu podmienku χ. Konzistentnosť týchto dvojíc je študovaná na obr. 8 pomocou techniky sémantických tabiel. Teória $\{\varphi_1, \varphi_2, ..., \varphi_n\}$ je konzistentná práve vtedy, keď sémantické tablo zostrojene z jej elementov $\varphi_1, \varphi_2, ..., \varphi_n$ má otvorenú aspoň jednu vetvu, ak tablo má uzavreté všetky vetvy, potom daná teória je nekonzistentná. Modely dvojíc $\{\Phi_X, \chi\}$ majú tento tvar

$$[\![\{\Phi_A,\chi^{(1)}\}]\!]=\{(11\#100)\},\ [\![\{\Phi_B,\chi^{(2)}\}]\!]=\{(10\#010)\},\ [\![\{\Phi_C,\chi^{(1)}\}]\!]=\{(\#\#1100)\} \quad (22b)$$

Z týchto modelov vyplýva, že teórie- dvojice majú nasledujúce vlastnosti sémantického vyplývania

$$\{\Phi_A,\chi^{(1)}\}\models d_1,\ \{\Phi_B,\chi^{(2)}\}\models d_2,\ \{\Phi_C,\chi^{(1)}\}\models d_1 \quad (23)$$

To znamená, že prvú a tretiu dvojicu môžeme spojiťdo novej „spojenej“ teórie z ktorej sémanticky vyplýva d_1 (podobne, z druhej dvojice sémanticky vyplýva d_2)

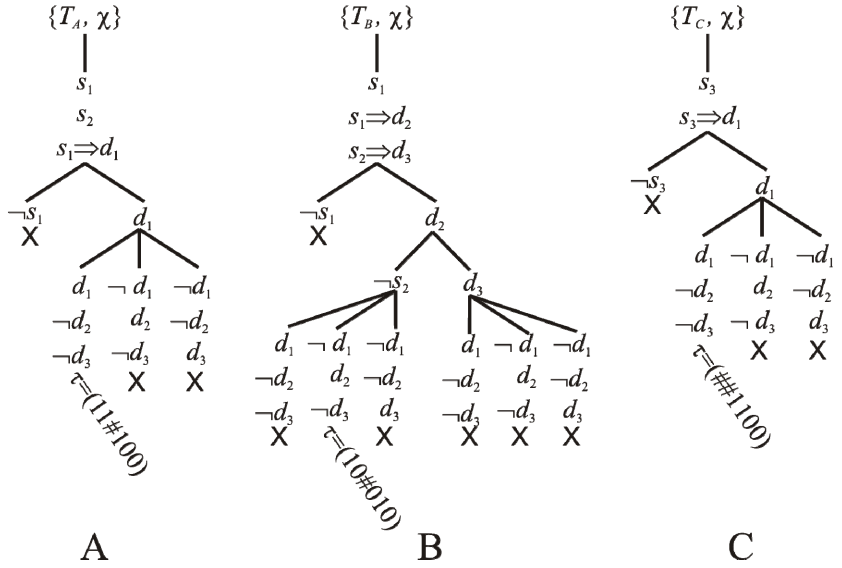
$$\{\Phi_A,\Phi_C,\chi^{(1)}\}\models d_1\ ,\ \{\Phi_B,\chi^{(2)}\}\models d_2 \quad (24)$$

Môžeme teda o dvoch výsledkoch procesu spájania pôvodnej teórie $\{\Phi_A,\Phi_B,\Phi_C\}\cup\{\chi\}$

prvý spôsob spájania: $$\mathcal{M}_1\left(\tilde{\Phi}=\{\Phi_A,\Phi_B,\Phi_C\},\chi\right)=\left\{\underbrace{\Phi_A,\Phi_C}_{\tilde{\Phi}_1},\chi^{(1)}\right\} \quad (25a)$$

druhý spôsob spájania: $$\mathcal{M}_2\left(\tilde{\Phi}=\{\Phi_A,\Phi_B,\Phi_C\},\chi\right)=\left\{\underbrace{\Phi_B}_{\tilde{\Phi}_2},\chi^{(2)}\right\} \quad (25b)$$

V oboch prípadoch pomocou procesu spájania vznikla konzistentná teória z ktorej sémanticky vyplýva d_1 resp. d_2.



Obrázok 9.8. Sémantické tablá troch teórií typu $\{T_X,\chi\}$, kde $X = A, B, C$. Každé sémantické tablo obsahuje práve jednu otvorenú vetvu, potom modely týchto teórií sú tvorené vektorom τ, ktorý sa nachádza pri otvorenej vetve a je tvorený literálmi nachádzajúcich sa na tejto vetve. .

Ako vyplýva z ilustračných príkladov 6-7 , proces spájania nezávislých teórií $\tilde{T}=\{T_1,T_2,...,T_n\}$ do jednej teórie, kde tento proces spájania je špecifikovaný výrokovou formulou (v tvare DNF formuly)

$$\chi=\chi^{(1)}\vee\chi^{(2)}\vee...\vee\chi^{(a)} \quad (26)$$

kde $\chi^{(i)}$ sú konjunktívne klazuly $\chi^{(i)}=l_1^{(i)}\wedge l_2^{(i)}\wedge...\wedge l_p^{(i)}$ obsahujúca literály $l_k^{(i)}$.

Budeme riešiť problém, ktoré časti metateória $\tilde{\Phi}$ spolu s reštrikčnou podmienkou spájania χ (pozri (26)) tvoria konzistentnú teóriu. Riešenie problému konzistentnosti $[\![\{\tilde{\Phi},\chi\}]\!]$ môžeme zovšeobecnením (21) prepísať do tvaru

$$[\![\{\tilde{\Phi},\chi\}]\!]=\bigcap_{i=1}^{n}\bigcap_{j=1}^{a}{}^{*}[\![\{\Phi_i,\chi^{(j)}\}]\!] \tag{27}$$

kde symbol $*$ na pravej strane znamená, že sú vylúčené členy, ktorých model je prázdny (nekonzistentné), $[\![\{\Phi_i,\chi^{(j)}\}]\!]=\varnothing$. Túto vylúčenie tvorí základ pre tvorbu spojenia $\mathcal{M}(\{\tilde{\Phi},\chi\})$,

konzistentné dvojice s rovnakou špecifikačnou klauzulou $\chi^{(j)}$ sú unifikované do jednej podmnožiny $\tilde{\Phi}_j=\{\Phi_{j_1},\Phi_{j_2},...\}\subseteq\tilde{T}$, potom platí

$$\mathcal{M}(\{\tilde{\Phi},\chi\})=\bigcup_{j=1}^{a}{}^{*}\{\tilde{\Phi}_j,\chi^{(j)}\} \tag{28}$$

túto formulu budeme nazývať ***kanonický tvar metateórie*** $\tilde{\Phi}$ (samozrejmé, tento kanonický tvar je definovaný vzhľadom ku klasifikačnej formuly χ). Opakovane upozorňujeme (pozri symbol $*$ na pravej strane (28)), ktorý znamená, že nekonzistentné členy, $[\![\{\tilde{\Phi}_j,\chi^{(j)}\}]\!]=\varnothing$, sú vynechané. Ak použijeme výsledky z príkladu 15.6, potom zjednotenie (28) má tento konkrétny tvar (pozri (25))

$$\mathcal{M}(\tilde{T}=\{\Phi_A,\Phi_B,\Phi_C\},\chi)=\{\Phi_A,\chi^{(1)}\}\cup\{\Phi_B,\chi^{(2)}\}\cup\{\Phi_C,\chi^{(1)}\}=\underbrace{\{\Phi_A,\Phi_C,\chi^{(1)}\}}_{\mathcal{M}_1}\cup\underbrace{\{\Phi_B,\chi^{(2)}\}}_{\mathcal{M}_2}$$

Na záver tejto kapitoly o spájaní poznatkov pristúpime k diskusii reálnej situácie, keď ku konzistentnej aktuálnej metateórii $\tilde{T}$ chceme pridať nové konzistentné poznatky, ktoré sú klasifikované pomocou klauzúl $\chi^{(i)}$ (pozri (26)). Nech pôvodná metateória má kanonický tvar (28)

$$\tilde{\Phi}=\bigcup_{j=1}^{a}{}^{*}\{\tilde{\Phi}_j,\chi^{(j)}\} \tag{29}$$

Nové poznatky reprezentované konzistentnou Φ_{new} sú taktiež reprezentované v kanonickom tvare (28)

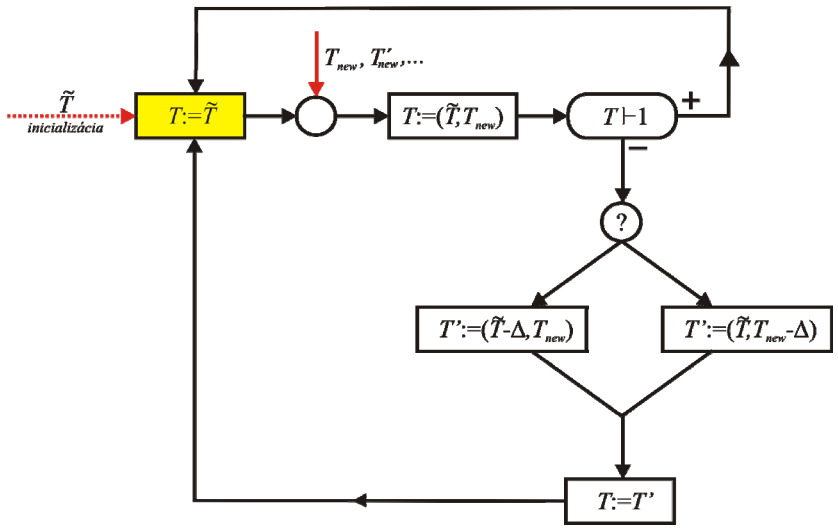
$$\Phi_{new}=\bigcup_{j=1}^{a}{}^{*}\{\Phi_{new,j},\chi^{(j)}\} \tag{30}$$

Jednoduchým zjednotením týchto dvoch teórií (29-30), dostaneme novú modifikovanú metateóriu

$$\tilde{\Phi}\cup\Phi_{new}=\bigcup_{j=1}^{a}{}^{*}\{\tilde{\Phi}_j\cup\Phi_{new,j},\chi^{(j)}\} \tag{31}$$

kde pre jednoduchosť pripúšťame, že teória $\tilde{\Phi}_j$ alebo $\Phi_{new,j}$ sú prázdne, v prípade, že obe teórie sú prázdne, potom podľa už použitej konvencie symbol ´$*$´ v (31) vynecháva tento index *j*. Stojíme pred úlohou ako zmodifikovať zjednotenie $\tilde{\Phi}_j\cup\Phi_{new,j}$ dvoch pôvodne konzistentných teórií tak, aby výsledná „spojená“ teória $\mathcal{M}(\{\tilde{\Phi}_j\cup\Phi_{new,j},\chi^{(j)}\})$ bola konzistentná. K tomuto účelu môžeme použiť techniku kontrakcie z kapitoly 15.3, kde z pôvodnej nekonzistentnej teórie Φ bola odstránená minimálna podmnožina Δ tak, aby výsledná teória $\Phi-\Delta$ bola konzistentná. V prípade nekonzistetnej teórie $\tilde{\Phi}_j\cup\Phi_{new,j}$ budeme používať tieto dve alternatívne možnosti: (1) $\Delta\subseteq\tilde{\Phi}_j$ alebo (2) $\Delta\subseteq\Phi_{new,j}$. V prvom prípade modifikujeme pôvodnú teóriu a zachovávame (preferujeme) novú teóriu , zatiaľ čo

v druhom prípade modifikujeme novú teóriu a zachovávame pôvodnú teóriu. V týchto prípadoch spojené teórie majú tvar $\mathcal{M}\left(\left\{\left(\tilde{\Phi}_j - \Delta\right) \cup \Phi_{new,j}, \chi^{(j)}\right\}\right)$ resp. $\mathcal{M}\left(\left\{\tilde{\Phi}_j \cup \left(\Phi_{new,j} - \Delta\right), \chi^{(j)}\right\}\right)$. Stojíme teda pred otázkou, ktoré poznatky preferujeme, či preferujeme pôvodné poznatky, potom používame druhý prístup založený na $\mathcal{M}\left(\left\{\tilde{\Phi}_j \cup \left(\Phi_{new,j} - \Delta\right), \chi^{(j)}\right\}\right)$, alebo preferujeme nové poznatky, potom používame formulu $\mathcal{M}\left(\left\{\left(\tilde{\Phi}_j - \Delta\right) \cup \Phi_{new,j}, \chi^{(j)}\right\}\right)$, pozri obr. 9.9.



Obrázok 9.9. Blokový diagram metódy spájania poznatkov. Metóda je inicializovaná počiatočným poznatkom $\tilde{\Phi}$, k aktuálnemu poznatku Φ sa pripojí nový poznatok Φ_{new}. Spojený poznatok Φ sa testuje, či je alebo nie je konzistentný. V prípade, že je konzistentný, potom sa vracia na začiatok procesu, prípoji sa k nemu novy poznatok Φ_{new}. V opačnom prípade, ak spojený poznatok Φ nie je konzistentný, potom na základe vnútornej špecifikácie túto nekonzistentnosť riešime dvoma nezávislými spôsobmi. Z pôvodnej teórie Φ odstránime buď jej minimálnu časť Δ, alebo z pripojeného poznatku Φ_{new} odstránime minimálnu časť Δ. V obidvoch prípadoch vytvoríme konzistetný pozntok Φ', vraciame sa na začiatok diagramu a vytvorený poznatok Φ' sa identifikuje s poznatkom Φ.

Príklad 8. Budeme pokračovať v riešení príkladu 7. Bolo ukázané, že pôvodné tri teórie (21a-c) môžu byť spojené do dvoch nových teórií (25a-b):

prvý spôsob spájania: $\mathcal{M}_1\left(\tilde{\Phi} = \left\{\Phi_A, \Phi_B, \Phi_C\right\}, \chi\right) = \left\{\Phi_A, \Phi_C, \chi^{(1)}\right\} = \left\{s_1, s_2, s_3, s_1 \Rightarrow d_1, s_3 \Rightarrow d_1\right\}$

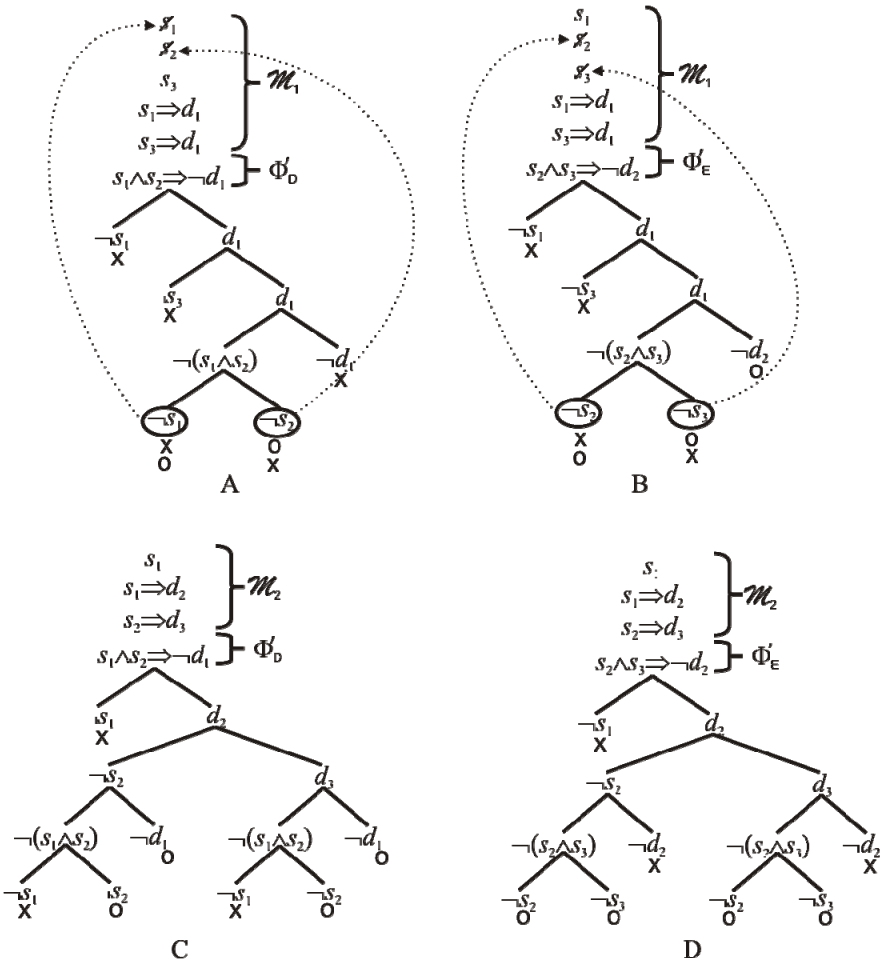
druhý spôsob spájania: $\mathcal{M}_2\left(\tilde{\Phi} = \left\{\Phi_A, \Phi_B, \Phi_C\right\}, \chi\right) = \left\{\Phi_B, \chi^{(2)}\right\} = \left\{s_1, s_1 \Rightarrow d_2, s_2 \Rightarrow d_3\right\}$

Tieto dve konzistentné teórie rozšírime o nové poznatky

$$\Phi'_D = \left\{s_1 \wedge s_2 \Rightarrow \neg d_1\right\} \tag{32a}$$

$$\Phi'_C = \left\{s_2 \wedge s_3 \Rightarrow \neg d_2\right\} \tag{32b}$$

Stojíme pred úlohou, ao pridať tieto dve teórie k už spojeným teóriam $\mathcal{M}_1\left(\tilde{\Phi} = \left\{\Phi_A, \Phi_B, \Phi_C\right\}, \chi\right)$ and $\mathcal{M}_2\left(\tilde{\Phi} = \left\{\Phi_A, \Phi_B, \Phi_C\right\}, \chi\right)$, tak, aby výsledné teórie boli konzistentné, pozri obr.15.10 .



Obrázok 9.10. Sémantické tablá pre 4 rôzne prípady spájania stávajúcich teórií $\mathcal{M}_1$ a $\mathcal{M}_2$ o nové teórie Φ'_D a Φ'_E špecifikované (32a-b). Diagram A znázorňuje prvú možnosť rozšírenia $\mathcal{M}_1$ o novú teóriu Φ'_D. V tomto prípade musíme z pôvodnej teórie $\mathcal{M}_1$ vylúčiť buď premennú s_1 alebo s_2 (nie súčasne obidve), potom sémantické tablo obsahuje práve jednu otvorenú vetvu, ktorá špecifikuje model novej spojenej teórie. Diagram B špecifikuje podobnú situáciu ako predchádzajúci diagram, k tomu, aby sme dostali spojením konzistentnú teóriu musíme vylúčiť buď premennú s_2 alebo s_3 (nie súčasne obidve). Diagramy C a D sú sémantické tablá pre spájanie teórie $\mathcal{M}_2$ s novou teóriou Φ'_D alebo Φ'_D, v tomto prípade nemúsime z pôvodných teórií odstrániť žiadnu premennú, sémantické table sú aj pre túto situáciu orvorené, čiže nové spojené teórie sú konzistentné.

Proces spájania pôvodných teórií $\mathcal{M}_1 = \{s_1, s_2, s_3, s_1 \Rightarrow d_1, s_3 \Rightarrow d_1\}$ a $\mathcal{M}_2 = \{s_1, s_1 \Rightarrow d_2, s_2 \Rightarrow d_3\}$ s novými teóriami Φ'_D a Φ'_E vedie podľa obr. 10 k týmto 4 možnostiam

$$\mathcal{M}_{1,new} = \{s_2, s_3, s_1 \Rightarrow d_1, s_3 \Rightarrow d_1, s_1 \wedge s_2 \Rightarrow \neg d_1\} \quad (33a)$$

$$\mathcal{M}_{2,new} = \{s_1, s_3, s_1 \Rightarrow d_1, s_3 \Rightarrow d_1, s_1 \wedge s_2 \Rightarrow \neg d_1\} \quad (33b)$$

$$\mathcal{M}_{3,new} = \{s_1, s_1 \Rightarrow d_2, s_2 \Rightarrow d_3, s_1 \wedge s_2 \Rightarrow \neg d_1\} \quad (33c)$$

$$\mathcal{M}_{4,new} = \{s_1, s_1 \Rightarrow d_2, s_2 \Rightarrow d_3, s_2 \wedge s_3 \Rightarrow \neg d_2\} \quad (33d)$$

Cvičenia

Cvičenie 1. Pomocou sémantických tabiel zistite konzistentnosť alebo nekonzistentnosť teórií, v prípade, že teória je nekonzistentná, zostrojte takú minimálnu podmnožinu $\Delta \subseteq \Phi$, aby množina $\Phi - \Delta$ bola konzistetná.

(a) $\Phi = \{p \Rightarrow q, q \wedge r, \neg p\}$

(b) $\Phi = \{p \vee q \Rightarrow p \wedge q, p \wedge q \Rightarrow p \vee q, \neg q\}$

(c) $\Phi = \{p \vee q, q \vee r, p \wedge \neg r, \neg q \wedge r, \neg p\}$

Cvičenie 2. Nech Φ je konzistentná teória (preverte), zostrojte pre túto teóriu takú minimálnu množinu Ψ, aby rozšírená teória $\Phi' = \Phi \cup \Psi$ bola konzistentná (preverte) a vzplýva z nej formula φ (preverte), pričom táto formula nevyplýva z pôvodnej teórie Φ (preverte), formálne $\Phi' \vdash \varphi$ a $\Phi \nvdash \varphi$.

(a) $\Phi = \{p \wedge q \Rightarrow r, \neg q \Rightarrow r\}$, $\varphi = \neg p \wedge q$

(b) $\Phi = \{\neg p, q \Rightarrow r, q \wedge r \Rightarrow p\}$, $\varphi = r$

(c) $\Phi = \{p \Rightarrow q, \neg p \Rightarrow \neg q, q \Rightarrow r, r\}$, $\varphi = p \equiv q$

Cvičenie 3. Nech Φ je nekonzistentná teória (preverte), zostrojte takú minimálnu množinu Δ, ktorej odstránením odstránt, e z tejto teórie minimálnu množinu Δ tak, aby výsledná teória $\Phi - \Delta$ bola konzistentná (preverte) a vyplývala z nej formula φ (preverte)

(a) $\Phi = \{p \equiv q, p, \neg q\}$, $\varphi = p$

(b) $\Phi = \{p \Rightarrow q, p \Rightarrow \neg q, p\}$, $\varphi = \neg q$

(c) $\Phi = \{p \Rightarrow q, q \Rightarrow r, p, \neg r\}$, $\varphi = q \wedge r$

Cvičenie 4. Nech Φ je konzistentná teória (preverte), táto teória je rozšírená teóriou Ψ tak, že rezultujúca teória $\Phi \cup \Psi$ je nekonzistentná (preverte). Zostrojte z tejto novej teórie minimálnu podmnožinu $\Delta \subseteq \Phi$ tak aby nová teória $(\Phi \cup \Psi) - \Delta$ bola konzistetná (preverte) a vyplývala z nej špecifikovaná formula φ (preverte), $(\Phi \cup \Psi) - \Delta \vdash \varphi$

(a) $\Phi = \{p \Rightarrow q, q \Rightarrow r\}$, $\Psi = \{p, \neg r\}$, $\varphi = q$

(b) $\Phi = \{p \Rightarrow q, p \Rightarrow \neg q\}$, $\Psi = \{p \wedge r\}$, $\varphi = q$

(c) $\Phi = \{p \Rightarrow q \wedge \neg r, q \Rightarrow r\}$, $\Psi = \{p\}$, $\varphi = \neg r$

Literatúra

[1] Alchourròn, C. E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, **50** (1985), 510-530.

[2] Gabbay, D.M., Rodrigues, O.T., Russo, A.: *Revision, Acceptability, and Context. Theoretical and Algorithmic Aspects.* Springer, Berlin, 2010

[3] Gärdenfors P.: *Knowledge in Flux: Modeling the Dynamics of Epistemic States,* The MIT Press, Bradford Books, Cambridge, MA, 1988.

[4] Gärdenfors, P. (ed.): *Belief Revision*. Cambridge University Press, Cambridge, 1992.

[5] Konieczny, S., Perez, R.: Logic Based Merging. *Journal of Philosophical Logic* **40**(2011), 239–270.

[6] Lin, J., Mendelzon, A.O.: Merging databases under constraints. *International Journal of Cooperative Informatikon Systems* **7**(1998), 55-76.

[7] Rott, S.: *Change, Choice and Inference*. Oxford University Press, Oxford, UK, 2001.

[8] Grove, A.: Two Modelling for Theory Change. *Journal of Philosophical Logic* **17** (1988), 157-170.

[9] Tarski, A.: *Logic, Semantics and Metamathematics*. Oxford University Press, Oxford (UK), 1956

10. prednáška

Kolektívna pamäť

10.1 Úvodné poznámky

Kolektívna pamäť, tak ako je chápaná v súčasnosti v sociálnych vedách (menovite v sociológii) [xx] je unikátnym typom pamäti, ktorý sa líši od obvyklých neurovedných typov pamätí [xx]. Kolektívna pamäť je distribuovaný systém jednotlivých mozgov/myslí/pamätí patriacich do vybranej sociálnej skupiny a ktorá je (pamäť) súčasťou jej kultúry [xx]. Aj napriek tomu, že podstatné časti kolektívnej pamäti sú „umiestnené“ v jednotlivých mozgoch, jej vznik a pretrvávanie sa považuje za neštandardnú kolektívnu vlastnosť sociálnej skupiny. Kolektívna pamäť patrí medzi základné špecifikácie sociálnej skupiny, ktorá je prenášané negenetickými (kultúrnymi) metódami v priebehu života jednotlivých indivíduí skupiny. Z toho pohľadu má koncepcia kolektívnej pamäti mnoho spoločného so zovšeobecneným a modifikovaným memetickým prístupom evolučného biológa R. Dawkinsa [xx] tak, ako je v súčasnosti chápaný [xx].

Kolektívna pamäť má tri súčasti: (1) poznatky, (2) atribúty a (3) procesy. Prvá súčasť – *poznatky* – aj napriek tomu, že sa môže v priebehu času meniť, patrí do základných konštitučných elementov kultúry sociálnej skupiny. Druhá súčasť, *atribúty*, tvorí špecifický holistický pohľad na históriu sociálnej skupiny a preto je často používaná ako jedna zo základných špecifikácii vybranej skupiny. Konečne, tretia súčasť, *procesy*, sú tvorené neustále prebiehajúcim dialógom medzi jednotlivcami skupiny alebo jej podskupinami, ktorý môže meniť v určitom rozsahu prvé dve súčasti kolektívnej pamäti.

Za neštandardnú súčasť kolektívnej pamäti môžeme pokladať aj E. Durkheimovo „*kolektívne vedomie*“ [xx] obsahujúce “sociálne skutočnosti” a ktoré chápal ako emergovaný kolektívny poznatok o postojoch, hodnotách a správaní príslušníkov daného sociálneho systému. Neskoršie K. Jung [xx] rozpracoval túto ideu tak, že je umiestnená v "hlbokých vrstvách" našej psychiky a obsahuje univerzálnu prvotnú pamäť alebo „archetypy“. Jung študoval tento pamäťový systém pomocou mýtov. Z pohľadu evolučnej psychológie, Jungovo kolektívne vedomie môže byť chápané tak, že vzniklo v priebehu evolúcie ľudského druhu a obsahuje "univerzálie" ľudských mentálnych schopností, ktoré ovplyvňujú sociálne správanie sa ľudí.

Na záver tohto prehľadu o koncepcii „kolektívnej pamäti“ obrátime našu pozornosť na aspekty získavania, revízie, pretrvávania a zániku kolektívnej pamäti. V klasickej (halbwachovskej) teórii kolektívnej pamäti sú tieto jej aspekty prezentované len v náznakoch bez vzťahu k umelej inteligencii. Moderné prístupy k revízii databáz poznatkov, k ich tvorbe a spájaniu sú rozpracované do veľkej hĺbky [xx] a stali sa integrálnou súčasťou súčasnej logiky a teórie vedy (budú prezentované v kapitole 3.3) .

Ako určitú kuriozitu uvediem našu skúsenosť s webovským prehliadačom. Zistili sme, že na Internete má výskyt pojmu "*kolektívna pamäť*" dve maxima, jedno maximum je v sociálnych vedách a druhé maximum je v umelej inteligencii. Môže nás toto zistenie prekvapiť, že v tak rôznych vedných oblastiach existuje spoločný odborný termín s veľmi vysokou frekvenciou používania? Reflexívnym vysvetlením tejto skutočnosti je názor, že sa jedná o dve úplne odlišné veci, ktoré spolu nesúvisia (až na ich verbálnu podobnosť). Avšak je potrebné na tomto mieste zdôrazniť, že súčasná umelá intligencia, hlavne zásluhou jej časti

tvorenou teóriou multiagentových systémov, sa stala schopnou vykonávať teoretické štúdie a počítačové simulácie rôznych javov a procesov prebiehajúcich v sociálnych systémoch a ktoré súvisia s koncepciou kolektívnej pamäti. Z tohto pohľadu sa teda nejedná o náhodu, že sa rovnaký pojem vyskytuje dvoch diametrálne odlišných vedných oblastiach, v informatike a v sociálnych vedách, ale jedná sa o dôsledok skutočnosti, že *koncepcia kolektívnej pamäti tak v sociálnych vedách ako aj v umelej inteligencii ma* ***spoločný*** *teoretický základ*.

Musíme zdôrazniť, že súčasne možnosti prieniku prístupov umelej inteligencie do sociálnych vied sú nepomerne väčšie než tie, aké boli pred 20-30 rokmi, keď prebiehal známy „experiment“ s použitím synergetickej fyzikálnej terminológie v sociálnych vedách [xx]. Moderný informatický odbor umelá inteligencia vypracoval teoretický aparát, ktorý je vhodný (a k tomu bol vlastne zostrojený) k štúdiu vzniku (hovoríme emergencie) sociálnych štruktúr v systémoch, ktoré obsahujú veľké množstvo podobných elementov – indivíduí - agentov, ktoré sa správajú "racionálne" na základe určitých pravidiel, ktoré sú im imanentne vlastné - môžeme povedať, že predpísané. Tento prístup, ktorý sa nazýva "*multiagentový systém*" [xx] v súčasnosti patrí medzi búrlivo sa rozvíjajúce oblasti umelej inteligencie, s množstvom efektívnych aplikácii v rôznych oblastiach nielen informatiky ale aj mimo nej. Možno konštatovať, že multiagentový prístup ovplyvnil charakter výskumu systémov, ktoré obsahujú množstvo podobných elementov – agentov a ktoré medzi sebou interagujú na základe jednoduchých pravidiel typu "ak..., potom...". Tieto štúdie multiagentových systémov patria do triedy počítačových simulácií, kde priebeh "adaptívneho" procesu sa dá dobre sledovať pomocou rôznych parametrov a počítačovej grafiky.

Koncom minulého storočia bolo publikovaných mnoho prác z multiagentových systémov, ktoré sa zaoberali štúdiom emergencie kooperácie, komunikácie, sociálnych štruktúr a pod. v týchto systémoch. Menovite zásluhou amerického politológa R. Axelroda [xx] bolo ukázané, že sa jedná o neobyčajné plodný smer rozvíjania interdisciplinárnej spolupráce medzi politológiou a umelou inteligenciou, kde „kooperácia“ medzi antagonistickými - sebeckými agentmi vzniká ako výsledok racionálnych dôvodov jej výhodnosti. To znamená, že rôzne *ideologické špekulatívne dôvody vzájomnej výhodnosti alebo nevýhodnosti spolupráce sú nahradené racionálnou argumentáciou založenej na multiagentovom prístupe*.

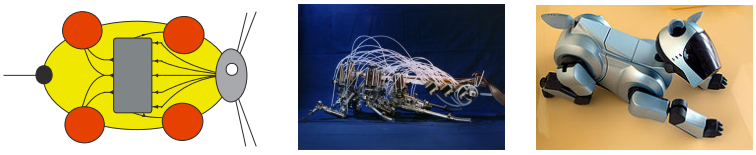
V ďalšej etape histórie multiagentových systémov (90. roky minulého storočia) bola jednoduchá forma intencionálnych agentov, ktorí vykonávajú reflexívne úzko špecifikované aktivity zovšeobecnená na sofistikovanejších agentov majúcich už *individuálnu myseľ* reprezentovanú informačnou databázou zloženou z užitočných poznatkov, pomocou ktorej sa agenti samostatne nereflexívne rozhodujú či vykonajú alebo nevykonajú nejakú aktivitu. Toto významné zovšeobecnenie multiagentového prístupu môže byť chápané, ako určitá špecifická verzia použitia memetického princípu, ktorý do evolučnej biológie už pred viac ako 30 rokmi zaviedol R. Dawkins v jeho slávnej knihe *The Selfish Gene* [xx], kde prezentoval svoje predstavy o radikálnom neodarvinizme. Dawkins v tejto knihe (v 7. kapitole) rozšíril pojem génu (alebo genotypu) o ďalšiu zložku, ktorá obsahuje informáciu, ktorá mení fitnes daného agenta (napr. informácie o prostredí, jeho regularitách alebo irregularitách, a pod). V tomto prístupe je evolučný princíp nahradený koevolučným, kde prebieha koevolúcia genotypu a „memetypu“, dochádza k spoločnej adaptácii oboch zložiek tohto rozšíreného 2-zložkového genotypu. Táto evolučná zmena memetypu je považovaná v memetike za príklad *evolučného vzniku kultúry*. Poznamenajme, že prvé snahy rozšíriť vlastnosti multiagentových systémov o individuálnu pamäť reprezentovanú databázou poznatkov, ktoré pomáhajú prežiť agentom v populácii (čiže ovplyvňuje ich fitnes), boli urobené [xx] autorom tejto state pred 10 rokmi pre aplikáciu multiagentového prístupu k formulácii memetiky.

Cieľom tejto práce je diskutovať koncepciu kolektívnej pamäti v rámci multiagentového prístupu, poukázať na skutočnosť, že tento prístup k existencii pamäti

v multiagentových systémoch je veľmi blízky ku koncepcii kolektívnej pamäti tak, ako sa používa v sociálnych vedách. Verím, že multiagentový prístup k štúdiu tak individuálnej ako aj kolektívnej pamäti poskytuje efektívny a jednoduchý interdisciplinárny formalizmus, ktorý sa pomerne ľahko dá aplikovať v sociálnych vedách.

10.2 Multiagentové systémy v umelej inteligencii

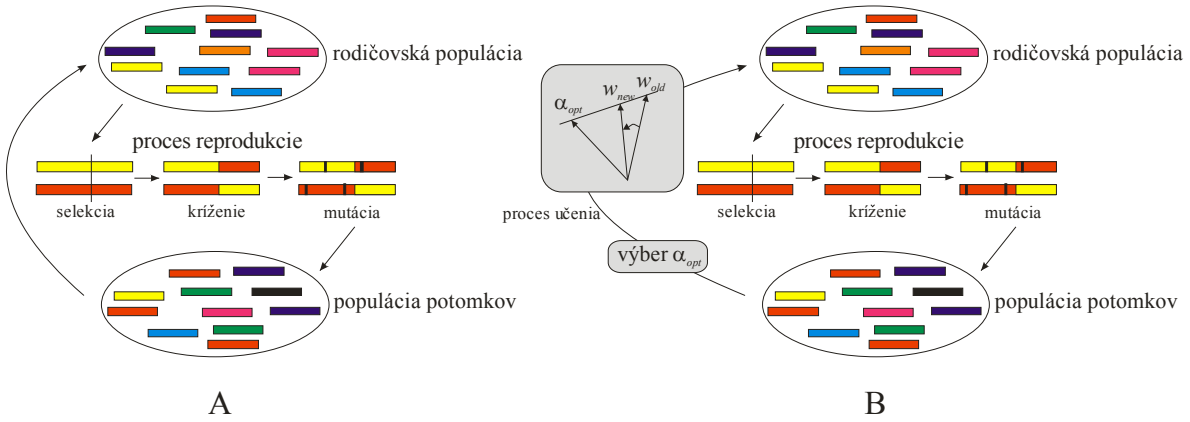
V priebehu posledných 20-30 rokov v umelej inteligencii sa stala veľmi populárnou koncepcia tzv. distribuovanej inteligencii založenej na myšlienke hmyzích spoločenstiev, kde obmedzené kognitívne schopnosti sú navzájom kombinované do kolektívnej inteligencie. Kolektívne správanie sa hmyzieho spoločenstva možno podľa moderného informatického jazyka chápať ako výsledok komplexných dynamických procesov vzájomnej interakcie jednoduchých indivíduí hmyzieho spoločenstva; hovoríme, že v priebehu časovej evolúcie tohto spoločenstva spontánne emerguje *zložitosť* (angl. complexity), ktorá pôvodne v študovanom hmyzom systéme nebola prítomná. Táto zložitosť vzniká ako dôsledok špecifických interakcií medzi hmyzom, akými sú napr. špecifická spolupráca a komunikácia, ktoré sú obvykle geneticky zafixované ako inštinkty. Samozrejme, tento prístup „kolektívnej inteligencie“ je pre štúdium mysle neaplikovateľný, avšak v umelej inteligencii tvorí širokú triedu metód nazývaných angl. „*natural computing*“ , ktoré tvoria významné a efektívne heuristiky pre riešenie zložitých (obvykle kombinatoriálnych) optimalizačných problémov, ktoré štandardnými prostriedkami aplikovanej matematiky sú len veľmi ťažko (ak vôbec) riešiteľné.



A **B** **C**

Obrázok 10.1. Rôzne reprezentácie agentov. (A) Hypotetický agent, ktorý obsahuje percepčný orgán (vidí), výstupy z tohto orgánu vstupujú do kognitívneho orgánu – mozgu, motorické výstupy z tohto mozgu riadia pohybové orgány agenta. (B) Mechanicko-elektronický agent – šváb, ktorý sa dokáže v prostredí pohybovať pomocou 6 nôh a optickým zariadením (zrakom) vníma okolie. Pohyb agenta je riadený malou neurónovou sieťou. (C) Japonská robotická hračka pes Aibo, dokáže sa pohybovať, vníma svoje okolie, vníma hlasové povely, atď.

Vo všeobecnosti, tieto a podobné systémy vyústili v informatike do *teórie multiagentových systémov* [xx], ktoré sú široko aplikovateľné na rôzne problémy umelej inteligencie, kde reprezentujú významnú triedu algoritmov pre riešenie zložitých problémov metódou postupného riešenia problému pomocou menších a podstatne jednoduchších podproblémov. Multiagentové systémy sa skladajú z malých jednotiek – agentov (pozri obr. 10.1), ktoré majú schopnosť (intenciu) spolupracovať s iným agentom, komunikovať s ním, vytvárať spoločné časo-priestorové štruktúry a pod. V nasledujúcej podkapitole budeme ilustrovať efektívnosť multiagentové prístupu na slávnom Axelrodovom [xx] príklade multiagentového prístupu kombinovaného s Darwinovou evolúciou pre štúdium emergencie kooperácie.

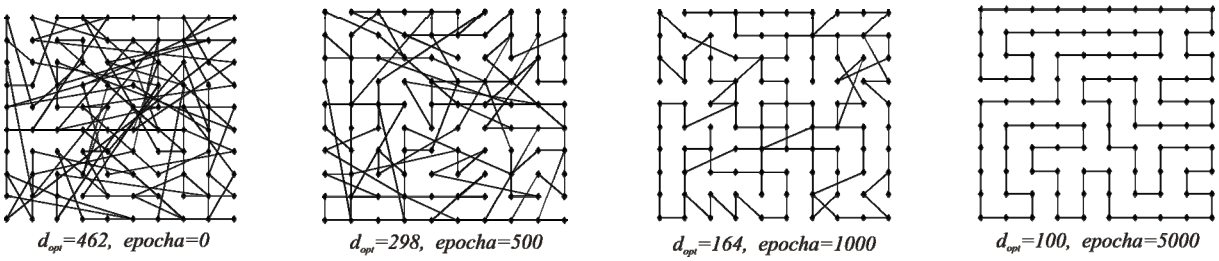


Obrázok 10.2. (A) Algoritmické vyjadrenie Darwinovej evolúcie (genetického algoritmu), kde operátor reprodukcie je rozložený na tri operácie: selekcie, kríženia a mutácie. Operáciou selekcie kvázináhodne (s pravdepodobnosťou úmernou fitnes replikátora) vyberáme z pôvodnej rodičovskej populácie dva replikátory; operáciou kríženia vytvoríme novú dvojicu replikátorov tak, že za bodom kríženia vymeníme medzi replikátormi zostávajúce časti; posledná tretia operácia mutácie spočíva v tom, že s malou pravdepodobnosťou zmeníme v náhodne vybraných polohách bity nových replikátorov; tento elementárny reprodukčný proces ukončíme tým, že dvojicu vytvorených replikátorov dáme do vytváranej populácie potomkov. Ak vytváraná populácia potomkov má rovnaký počet replikátorov, potom pôvodnú populáciu rodičov nahradíme populáciou potomkov a celý proces reprodukcie sa opakuje. (B) Jednoduchá modifikácia genetického algoritmu z diagramu A o proces "učenia", ktorý spočíva v tom, že na záver vytvárania populácie potomkov, vyberieme z nej optimálny replikátor, ktorý má v populácii maximálny fitnes. Pomocou tohto replikátora zmodifikujeme tzv. pravdepodobnostný vektor $\boldsymbol{w} = (w_1, w_2, ..., w_n)$ s cieľom, aby na záver evolučného procesu jeho komponenty w_i špecifikovali pravdepodobnosť toho, že v *i*-tej polohe replikátorov populácie bude obsiahnutý bit 1. Váhový vektor je inicializovaný pred zahájením evolúcie tak, že všetky jeho komponenty sa rovnajú 1/2, v jednotlivých epochách evolúcie je tento pravdepodobnostný vektor obnovovaný pomocou formule $w_{new} = (1-\lambda) w_{old} + \lambda \alpha_{opt}$, kde koeficient učenia λ je malé kladné číslo. Elementárne kroky učenia spočívajú v tom, že pôvodný vektor w_{old} sa "natáča" smerom k replikátoru α_{opt}, ktorý je v danej evolučnej epoche optimálnym replikátorom. Proces učenia (a aj proces evolúcie) je ukončený vtedy, keď všetky komponenty vektora $\boldsymbol{w}$ sú rovné 1 alebo 0. Môžeme povedať, že pravdepodobnostný vektor $\boldsymbol{w}$ tvorí "spoločnú pamäť" replikátorov z populácie. Jej význam spočíva v tom, že proces mutácie nie je úplne stochastický proces, ale berie v úvahu aj predchádzajúcu históriu evolúcie.

10.2.1 Evolučný prístup – evolučné algoritmy – genetické algoritmy

Evolučné algoritmy [xx] sú integrálnou súčasťou modernej umelej inteligencie (v tomto prípade nazývanou *počítačová inteligencia*), slúžia ako efektívny nástroj pre kvázistochastickú konštrukciu globálneho riešenia zložitých úloh, kde úplná stochastičnosť konštrukcie je nahradená heuristikou založenej na Darwinovom prirodzenom výbere. Evolučné algoritmy vychádzajú z paradigmy darvinovskej evolúcie, ktorá má univerzálny charakter [xx] a je v súčasnosti široko aplikovaná v rôznych oblastiach ľudského poznania (poznáme evolučnú ekológiu, evolučnú ekonomiku, evolučnú psychológiu, evolučnú lingvistiku, evolučnú epistemológiu, evolučnú výpočtovú vedu, evolučnú fyziku,...). Prvý, kto na túto univerzálnosť darvinovskej evolúcie upozornil, bol americký populačný genetik S. Wright. Ten v 30. rokoch minulého storočia charakterizoval *evolúciu ako optimalizáciu na povrchu fitnes funkcie* (fitness landscape), kde sa hľadá genotyp odpovedajúci globálnemu maximu. Z tohto pohľadu môžeme darvinovskú evolúciu charakterizovať ako univerzálnu paradigmu schopnú riešiť komplikované problémy vyskytujúce sa v širokom spektre vedných oblastí, od prírodných vied až po spoločenské vedy. V informatike našla táto zaujímavá idea svoj odraz už pred viac ako 30 rokmi, keď J. Holland navrhol genetické algoritmy (špeciálny prípad evolučných algoritmov), ktoré je možno chápať ako algoritmizáciu darvinovskej

evolúcie (pozri obr. 2). Ich použitie umožňuje riešiť mnohé problémy, ktoré sa donedávna zdali len ťažko (ak vôbec) riešiteľnými (pozri obr. 3).



Obrázok 10.3. Znázornenie efektívnosti genetického algoritmu, kde jednotlivé replikátory reprezentujú cesty obchodného cestujúceho na ortogonálnej mriežke 10×10 v rôznych epochách evolučného riešenia úlohy Obrázky možno interpretovať tak, že v priebehu evolúcie z chaosu emerguje vysoko regulárne riešenie, ktoré má optimálnu dĺžku cesty. (d_{opt}=100)

Použitie evolučných algoritmov v umelej inteligencii spolu s paradigmou multiagentových systémov otvára nové a neočakávané možnosti pre tvorbu stratégií a modelov (ktoré reprezentujú obvykle veľmi zložité problémy pre človeka), ich tvorba je prenechaná počítaču pomocou prístupu darvinovskej evolúcie „in silico“. Najvšeobecnejšia formulácia základnej idey univerzálneho darvinizmu je uskutočnená pomocou koncepcie ***darvinovského systému*** založeného na nasledujúcich dvoch postulátoch:

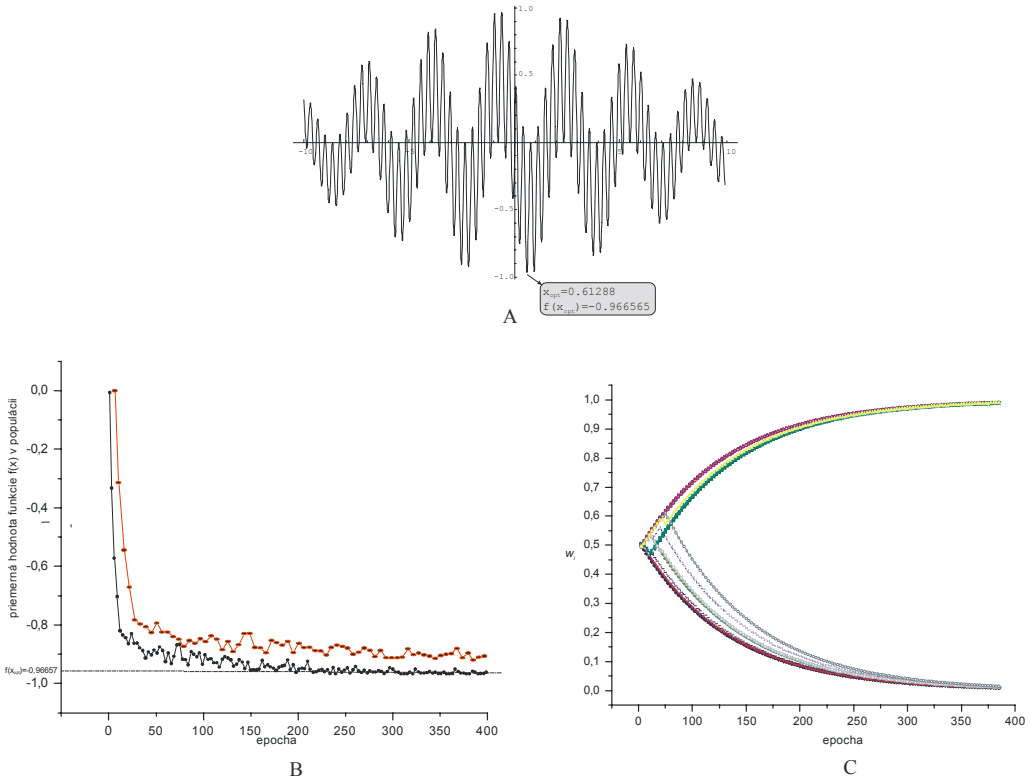
(1) *Darvinovský systém sa skladá z populácie replikátorov* – jedincov/objektov, ktoré za určitých vhodných podmienok sú schopné *replikácie* - rozmnožovania. Replikačný proces spočíva v „kopírovaní“ jedincov do populácie, pričom toto „kopírovanie“ sa uskutočňuje s *malými chybami*. Z tohto predpokladu nepresnej replikácie vyplýva, že *populácia nie je homogénna*, obsahuje jedincov – replikátory, ktoré vykazujú určitú malú variabilitu.

(2) Každý replikátor populácie je ohodnotený *fitnes*, toto hodnotenie vyjadruje schopnosť replikátora prežiť a úspešne vstupovať do replikačného procesu. *Prírodný výber* v darvinovskom systéme spočíva v tom, že jedinci populácie nie sú vyberaní do replikačného procesu náhodne, ale s *pravdepodobnosťou úmernou ich fitnes* (hovoríme, že výber sa deje kvázináhodne pomocou prirodzeného výberu).

Vo všeobecnosti, evolučné algoritmy (ktorých špeciálny typ sú genetické algoritmy) sú navrhnuté ako významná heuristika, ktorá je schopná poskytnúť kvalitné riešenie binárneho optimalizačného problému

$$f\left(\boldsymbol{x}_{opt}\right) = \min_{x\in\{0,1\}^n} f\left(\boldsymbol{x}\right) \Leftrightarrow \boldsymbol{x}_{opt} = arg \min_{x\in\{0,1\}^n} f\left(\boldsymbol{x}\right) \qquad (1)$$

v ktorom hľadáme globálne minimum funkcie $f(\boldsymbol{x})$ definovanej nad binárnymi reťazcami dĺžky n. Pretože počet binárnych vektorov dĺžky n je 2^n, zložitosť riešenia tohto problému rastie exponenciálne, t. j. čas potrebný na riešenie tohto problému rastie exponenciálne s rastom dimenzie n, $t_{výpočtu} \sim 2^n$. To znamená, že pre väčšie n (napr. n = 50 alebo n =100, tento problém nemáme šancu vyriešiť v rozumnom čase trvania našej civilizácie). Práve v tomto prípade vystupujú nastupujú evolučné algoritmy, ktoré na základe darvinovskej metafory evolúcie sú schopné nájsť riešenia tohto problému, ktoré aj keď sa nemusia rovnať globálnemu riešeniu $\boldsymbol{x}_{opt}$, ale je mu veľmi blízke a v mnohých prípadoch použiteľné ako globálne riešenie.



Obrázok 10.4. (A) Graf multimodálnej funkcie $f(x) = exp(-0.01x^2)sin(8x)cos(10x)$ na intervale $-10 \le x \le 10$, ktorá obsahuje 56 miním, globálne minimum je označené na diagrame. (B) Priebeh priemernej funkčnej hodnoty $f(x)$ v populácii replikátorov, optimálna funkčná hodnota je $f(x_{opt}) = 0.96657$. Dolná čierna čiara znázorňuje priebeh evolúcie s binárnymi replikátormi, kde je zahrnutá kolektívna pamäť pomocou pravdepodobnostného vektora $\boldsymbol{w}$. Horná svetlejšia čiara znázorňuje priebeh evolúcie bez zahrnutia kolektívnej pamäti. Z jej priebehu vyplýva že táto evolúcia bez kolektívnej pamäti nekonverguje k správnemu globálnemu minimu, ale k blízkej vyššej hodnoty. (C) Priebeh jednotlivých zložiek w_i pravdepodobnostného vektora $\boldsymbol{w}$, ktorý je inicializovaný hodnotami $w_i = 1/2$ pre všetky zložky. V priebehu evolúcie jednotlivé komponenty sa asymptoticky blížia k binárnym hodnotám 0 alebo 1. Vidíme, že asi po epoche 300 mutácia riadená zložkami pravdepodobnostného vektora vždy určí danú komponentu v správnej hodnote. Táto skutočnosť môže byť interpretovaná ako *kolektívna pamäť* pre evolúciu populácie replikátorov, ktorá riadi mutáciu pre reprodukčnom procese na základe predchádzajúcich epoch procesu.

Na záver tejto podkapitoly uvedieme jednoduchý ilustračný príklad rozšírenia genetického algoritmu o "*kolektívnu pamäť*", ktorá je realizovaná tzv. *pravdepodobnostným vektorom* $\boldsymbol{w} = (w_1, w_2, ..., w_n) \in \langle 0,1 \rangle^n$, ktorého jednotlivé komponenty $0 \le w_i \le 1$ určujú pravdepodobnosť toho, či pri mutácii replikátora $\boldsymbol{x} = (x_1, x_2, ..., x_n) \in \{0,1\}^n$ bude v polohe x_i binárna hodnota 1 alebo 0. V pôvodnom genetickom algoritme, proces mutácie je úplne náhodný, t. j. v replikátore v danej polohe pri mutácii sa bitové preklopenie $0 \leftrightarrow 1$ generuje náhodne (pozri obr. 2, diagram B):

(1) modifikovaný genetický algoritmus

$$x_i = \begin{cases} mutácia & (random \le P_{mut}) = \begin{cases} 1 & (random \le w_i) \\ 0 & (ináč) \end{cases} \\ x_i & (ináč) \end{cases} \qquad (2a)$$

(2) pôvodný genetický algoritmus

$$x_i = \begin{cases} 1 - x_i & (mutácia,\ ak\ random \le P_{mut}) \\ x_i & (inák) \end{cases} \tag{2b}$$

Kde *random* je náhodné číslo s rovnomernou pravdepodobnosťou z intervalu $\langle 0,1 \rangle$. Úloha, ktorá je riešená evolučným prístupom je hľadanie globálneho minima funkcie $f(x) = e^{-0.01x^2} sin(8x) cos(10x)$ na intervale $-10 \le x \le 10$. Táto funkcia má na danom intervale 56 miním, z ktorých len jedno je globálne, $f(x_{opt})$=0.966565, x_{opt}=0.61288 (pozri diagram A, obr.4).

Nech v danej epoche evolúcie populácia obsahuje replikátor $\boldsymbol{x}_{opt}$ s maximálnym fitnes, toto vybrané riešenie sa použije na modifikáciu pamäte algoritmu reprezentovanej pravdepodobnostnými vektorom $\boldsymbol{w}$ (pozri diagram B, obr. 2)

$$\boldsymbol{w} := (1 - \lambda)\boldsymbol{w} + \lambda \boldsymbol{x}_{opt} \tag{3}$$

kde λ je koeficient učenia, jeho hodnota je malé kladné číslo. Transformácia (3) môže byť interpretované ako „rotácia“ vektora $\boldsymbol{w}$ smerom k lokálne optimálnemu riešeniu $\boldsymbol{x}_{opt}$. Výsledky takto rozšíreného genetického algoritmu o kolektívnu pamäť (pomocou pravdepodobnostného vektora) sú znázornené na obr. 4, diagramy B a C. Z týchto numerických výsledkov jednoznačne vyplýva, že prístup kolektívnej pamäti použitý v genetických algoritmoch reprezentuje významný intenzifikačný faktor evolúcie, ktorý je založený na predchádzajúcich etapách evolúcie a riadi mutácie tak, aby neboli úplne náhodné, ale aby odrážali určité „historické“ skúsenosti agenta s riešením daného optimalizačného problému.

10.2.2 Problém kooperácie a sociálne dilemy

Ako vyplýva z teórie hier v ekonómii, aplikovanej tiež často na analýzu pretekov v **zbrojení medzi štátmi**, kooperácia sa môže vynoriť zo sebeckého záujmu, keď je výhodná pre obidvoch. Vo všeobecnosti, keď hľadáme racionálne dôvody, či pri našej interakcii s niekým iným (s našim obchodným alebo iným partnerom) pristúpiť na kooperáciu, alebo nie, naše rozhodnutie by malo byť založené na racionálnych argumentoch a nie na predsudkoch a ideologických dôvodoch.

Cieľom tejto podkapitoly je poukázať na populárny prístup k vysvetleniu vzniku kooperácie, ktorý je založený na simulačných výpočtoch pomocou hry nazývanej „***väzňova dilema***“ [3,4]. Scenár hry je tento: polícia zatkla dvoch páchateľov, avšak nemá dosť informácií k tomu, aby ich obvinila. Páchateľom, ktorí sú väznení separovane, polícia dá rovnakú ponuku: ak jeden obviní druhého (nespolupracuje) a druhý bude mlčať (spolupracuje), potom ten čo vypovedal (obvinil spoločníka), bude oslobodený (špecifika amerického súdnictva – svedok obžaloby) a ten čo mlčal (spolupracoval) dostane jednoročný trest. Ak budú obaja mlčať, obaja dostanú len 1-mesačný trest.. Ak budú obaja vypovedať (navzájom sa obvinia), dostanú 3-mesačný trest. Každý z obvinených sa musí rozhodnúť, či bude vypovedať proti svojmu spoločníkovi alebo bude mlčať; rozhodnutie každého z nich je neverejné.

Táto dilema ilustruje konflikt medzi individuálnou a skupinovou racionálnosťou. Skupina, v ktorej členovia preferujú svoje vlastné záujmy - nekooperujú, obvykle skončí zle, ako tá skupina, kde jej členovia uprednostňujú skupinové záujmy - kooperujú. Vo všeobecnosti, ak záujmy jednotlivcov reprezentujú záujmy celej skupiny, potom takáto skupina je úspešnejšia ako skupina, ktorej jednotliví členovia uprednostňujú vlastné záujmy a ciele. Tento problém bol navrhnutý a široko diskutovaný už počiatkom 50-tych rokov minulého storočia americkými ekonómami M. Floodom a M. Dresherom. Väzňova dilema sa stala často používaným prístupom k vysvetlenie vzniku kooperácie v rôznych sociálnych, ekonomických a iných systémoch, kde sa pomocou matematickej teórie hier a/alebo

počítačových simulácií hľadajú „racionálne dôvody“ k tomu, prečo a za akých okolností je výhodnejšie spolupracovať, než ako nespolupracovať.

K simulácii tejto hry použijeme multiagentový prístup, t. j. každý agent opakovane vstupuje do vzájomnej interakcie s iným agentom a podľa ich osobnej stratégie buď spolupracujú alebo nespolupracujú. Toto rozhodnutie, či spolupracovať alebo nespolupracovať, vykonávajú pomocou elementárneho kognitívneho orgánu - mysle, ktorý je riadený stratégiou daného agenta.

10.2.2.1 Základné princípy hry "väzňovej dilemy"

"Väzňova dilema" **(**VD) je hrou medzi dvoma hráčmi, kde každý z nich má výber z dvoch možností: spolupracovať (cooperate – *C*) alebo podvádzať (defect – *D*). Matica platieb tejto hry má tvar

$$f = \begin{pmatrix} (R,R) & (S,T) \\ (T,S) & (P,P) \end{pmatrix} \Leftrightarrow \begin{matrix} \begin{pmatrix} (3,3) & (0,5) \\ (5,0) & (1,1) \end{pmatrix} & \begin{matrix} C \\ D \end{matrix} \\ \begin{matrix} C & D \end{matrix} & \end{matrix} \tag{4}$$

Keď oba hráči spolupracujú, potom každý dostane „odmenu“ R=3 (reward), ktorá by mala byť väčšia ako bodový zisk „trest“ P=1 (punishment) získaný hráčmi, keď nekooperujú. Ale keď jeden hráč podvádza (t. j. nekooperuje, hrá *D*) zatiaľ čo druhý hráč spolupracuje (hrá C), potom podvádzajúci získa maximálnu výplatu bodov T=5 (temptation - pokušenia) ktorá je väčšia ako R, zatiaľ čo spolupracujúci dostáva minimálny počet bodov S=0 (sucker – ten, čo naletel) ktorá je menšia ako P. Pri takto zostavenej tabuľke platieb, je veľké pokušenie podvádzať, ak podvádzajú oba, tak získajú 1 bode, ale ak jeden hráč podvádza a druhý hráč spolupracuje, potom ten čo podvádza dostane 5 bodov, zatiaľ ten čo spolupracuje nedostane žiadny bod. V týchto platbách spočíva „dilema“ hry, či podvádzať alebo spolupracovať.

Uvažujme hráča – agenta a, jeho stratégia $s(a)$ je súbor pravidiel, ktoré použije pri výbere ťahu *C* alebo *D* (spolupráce alebo podvádzania) proti druhému hráčovi, na základe predošlej histórie hry. Ak by hra VD obsahovala len jeden krok, potom by samozrejme bolo výhodné podvádzať pri polovičnej šanci stretnúť spoluhráča, no keď sa hra s tým istým spoluväzňom opakujú a vopred nie je známy počet hier, metóda ako sa zachovať nie je jasná.

Nech sa hra opakuje t_{max}-krát (dôležitý parameter hry), na záver hry sa spočítajú platby jednotlivých hráčov. Stratégia je určená pomocou 3-bitového binárneho vektora (t.j. hráči si pamätajú predchádzajúci ťah oponenta) $\boldsymbol{s}=\{s_1, s_2, s_3\} \in \{0,1\}^3$, kde jednotlivé zložky (0 – kooperácie, 1 – nekooperácia) majú túto interpretáciu:

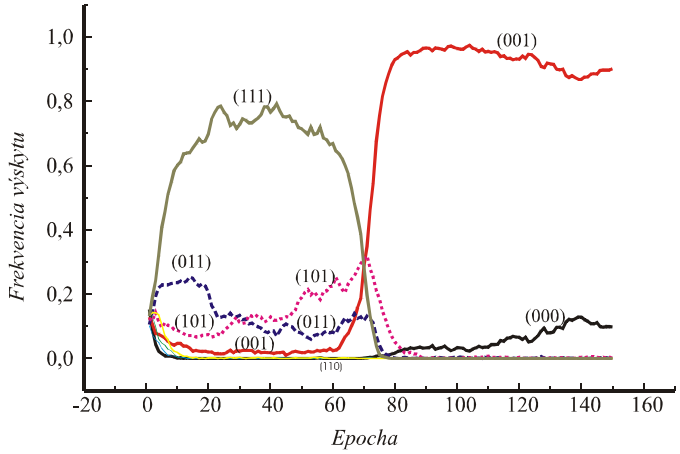
(1) s_1 určuje prvý ťah hráča: ak s_1=0, potom hráč kooperuje, v opačnom prípade, ak s_1=1, nekooperuje.
(2) s_2 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku kooperoval: ak s_2=0, potom hráč kooperuje, v opačnom prípade, ak s_2=1, nekooperuje.
(3) s_3 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku nekooperoval: ak s_3=0, potom hráč kooperuje, v opačnom prípade, ak s_3=1, nekooperuje.

Takáto stratégia vyžaduje len krátkodobú pamäť hráčov, požaduje len znalosť posledného kroku protivníka - druhého hráča. Celkový počet rôznych 1-stratégií je 2^3=8. Keby si väzni mali pamätať viac krokov do minulosti, s každým takým krokom by exponenciálne narastala zložitosť zápisu stratégie.

10.2.2.2 Simulácia emergencie stabilnej stratégie hry VD pomocou evolučného algoritmu

Pri plánovaní stratégie na základe predchádzajúcich skúseností s protihráčom sa často používa evolučný algoritmus. Táto aplikácia evolučného algoritmu prvýkrát vznikla zo spolupráce amerického politológa R. Axelroda [xx] s informatikom a tvorcom genetických algoritmov J. Hollandom a rozvíjala sa ďalej pre zložitejšie modely medzinárodnej bezpečnosti

odpovedajúce napr. na otázku, koľko z rozpočtu vydávať na obranu. V časovom priebehu je zaujímavé sledovať vývoj kooperatívnych a nekooperatívnych stratégií. Kooperatívne stratégie sa spočiatku ukázali menej výhodné, no v neskorších štádiách prevážili.



Obrázok 10.5. Výsledky evolučného algoritmu použitého pre emergenciu výslednej stratégie, pričom populácia stratégií bola inicializovaná náhodne. V priebehu evolúcie algoritmu spontánne emergovali dve „kooperatívne" stratégie (001) a (000), ktoré zahajujú hru kooperáciou a na kooperáciu súpera v predchádzajúcom ťahu odpovedajú kooperáciou v nasledujúcom ťahu. V dôsledku toho, že väčšinová stratégia (001) neprodukuje „podvod", menšinová stratégia (000) nemá príležitosť reagovať spoluprácou na podvod.

Numerické výsledky evolučného algoritmu sú znázornené na obr. 5. Ako výsledná stratégia emergujú dve podobné stratégie (001) a (000), ktoré obe zahajajú hru kooperáciou a na kooperáciu odpovedajú kooperáciou. Líšia sa v reakcii na podvod, prvá stratégia (001) reaguje na podvod podvodom, zatiaľ čo druhá stratégia (000) „hlúpo" reaguje na podvod kooperáciou. Stratégia (001) sa nazýva podľa Axelroda *tit-for-tat* (vo volnom preklade, ako ty mne, tak ja tebe). Menšinová stratégia „dobrák" (000) má šancu prežívať iba popri stratégii tit-for-tat.

Výsledky týchto simulačných výpočtov môžeme zosumarizovať tak, že evolučný návrh heuristík, ktoré optimalizujú dlhodobý zisk pri opakovanej interakcii medzi agentmi, viedol k návrhu pomerne jednoduchej stratégie – heuristiky *tit-for-tat*, ktorá má širšiu platnosť aj ako „morálna heuristika", návod pre dlhodobú opakujúca sa kooperatívnu interakciu medzi agentmi. Tento prístup bol zovšeobecnený aj pre štúdium etnických konfliktov [xx] medzi dvoma etnicky rôznymi populáciami. Získané výsledky racionalizujú koncepciu „kolektívnej viny", okolo ktorej v literatúre existuje množstvo navzájom protirečivých názorov, od zásadného odmietania až po jej schvaľovanie vo zvláštnych prípadoch. Pomocou simulačných výpočtov bolo dokázané, že princíp kolektívnej viny má fatálne následky na kooperáciu medzi rôznymi etnikami, ktorá nemôže vzniknúť, pretože príslušník inej etnickej skupiny je automaticky penalizovaný princípom kolektívnej viny.

Hra VD môže byť zovšeobecnená aj pre viac ako dvoch hráčov, potom sa obvykle nazýva „*tragédia spoločného*" (angl. *tragedy of the commons*), ktorú prvý krát formuloval G. Hardin v r. 1968 [xx]: "Predstavte si obecnú lúku prístupnú pre všetkých. Je možné očakávať, že každý užívateľ lúky sa pokúsi užívať túto lúku šetrným spôsobom tak, že bude na nej pásť len malý počet dobytka. Avšak, ako racionálny hospodár, každý užívateľ sa snaží maximalizovať svoj zisk. Isto si kladie otázku: "akú zmenu výnosu pre mňa znamená rozšírenie môjho stáda o jeden alebo viac kusov?"

Tento zaujímavý príklad sa dá modelovať tiež pomocou väzňovej dilemy, v tomto prípade namiesto dvoch väzňov súčasne interagujú traja alebo viacerí agenti - hráči. Podrobnou analýzou je možné dokázať, že v takto rozšírenej VD existuje stabilná stratégia,

ktorá hovorí, že ak v predchádzajúcom kroku aspoň jeden hráč nekooperoval, potom nekooperuj, ak všetci kooperovali, potom kooperuj aj ty. Táto "zničujúca" stratégia (ktorá vedie k devastácii spoločného majetku) znamená, že ak sa medzi spoločenstvom užívateľov vyskytne aspoň jeden nekooperujúci hráč, potom pre ostatných účastníkov dlhodobá kooperácia je nevýhodná.

Pretože vo väčších skupinách sa vždy nájde nekooperujúci jedinec, nastáva všeobecná nespolupráca - tragédia spoločného (lúky, u rybolovu v medzinárodných morských teritóriách, vyrubovanie lesov, údržba spoločného zavodňovacieho systému, znečisťovanie atmosféry Zeme, a pod.). Politológ R. Axelrod tvrdí [xx], že k tomu, aby v spoločnosti vznikla stabilná kooperácia, musia v nej existovať inštitúcie, ktoré kontrolujú - penalizujú nekooperáciu (t. j. represívne orgány typu polície).

Evolučné štúdium väzňovej dilemy sa dá použiť aj na modelovanie vnútro- a medzietnických vzťahov [xx]. U najjednoduchšieho modelu s dvoma etnikami medzi jedincami populácie *P* existujú tri typy interakcií: vnútroetnické interakcie v rámci etnickej skupiny *A*, vnútroetnické interakcie v rámci etnickej skupiny *B*, a konečne, medzietnické interakcie medzi jedincami z *A* a *B*. Postulujeme, že frekvencia medzietnickej interakcie je oveľa menšia ako frekvencia vnútroetnických interakcií v *A* alebo *B*. Jedinec je vybavený zafixovanou stratégiou väzenskej dilemy, dokáže rozpoznať etnickú skupinu druhého a jeho penalizáciu, čo je označenie nespolupracujúcich jedincov (alebo etnika) spolupracujúcimi jedincami. Evolúcia populácie prebieha nezávisle pre obidve etnické skupiny *A* a *B*. Fitnes jedincov je vypočítaný pomocou "turnaja", kde je predpísaný počet-krát náhodne vybraná dvojica agentov, ktorí spolu hrajú jeden ťah hry väzňovej dilemy. K štúdiu medzietnických interakcií boli použité dve rozdielne evolučné stratégie.

V prvej stratégii, ktorá sa nazýva *kolektívna vina*, je potrestaná celá etnická skupina za akt nespolupráce svojho člena voči členovi druhej etnickej skupiny. Zo simulačných výpočtov a teoretických úvah vyplýva, že tento typ stratégie nevedie k emergencii spolupráce medzi rozdielnymi etnickými skupinami, pokiaľ nezavedieme administrátora pokutujúceho akty nespolupráce. Keď etnické skupiny majú veľmi rozdielne veľkosti, princíp kolektívnej viny môže byť veľmi devastujúci pre menšiu etnickú skupinu, v ktorej môže zaniknúť aj vnútroetnická spolupráca.

Druhou použitou stratégiou pre podporu spolupráce je tzv. *osobná zodpovednosť*, kedy člen skupiny, ktorý nespolupracuje v rámci medzietnických interakcií, je potrestaný osobne v rámci svojej etnickej skupiny. To znamená, že na rozdiel od princípu kolektívnej viny, v tejto stratégii existuje len jeden typ "potrestania", členovia sú označení penalizáciou osobne. Simulačné výpočty spolu s teoretickou analýzou ukazujú, že v tomto prípade môže emergovať medzietnická spolupráca.

10.3 Agenti s mysľou [xx]

V predchádzajúcej časti tejto práce boli špecifikovaní najjednoduchší agenti, ktorých inteligencia bola ohraničená len na vykonávanie predpísaných – reflexívnych aktivít spolupracovať alebo nespolupracovať s nejakým iným agentom, vyhýbať sa prekážkam alebo zbierať potravu v blízkom okolí, a pod. V tejto kapitole popíšeme vlastnosti agentov, ktorý už majú „*elementárnu myseľ*", ktorá je tvorená (1) pamäťou (reprezentovanou množinou poznatkov, nazývanou v matematickej logike *teória*) a (2) inferenčným aparátom, pomocou ktorého je schopný agent vykonávať nad teóriou rozšírenou pozorovaním o nové poznatky jednoduché usudzovanie.

10.3.1 Pohľad umelej inteligencie a kognitívnej vedy na vzťah myseľ – mozog

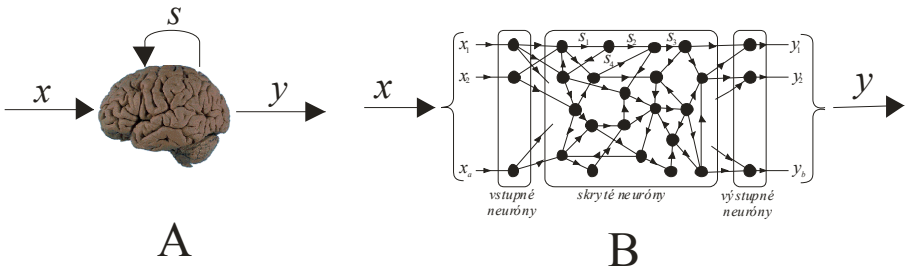
V prvej úvodnej časti tejto kapitoly poskytneme všeobecný pohľad umelej inteligencie na komplex mozog-myseľ ako na zariadenie, ktoré transformuje vstupné údaje ***x*** (produkované zrakom, sluchom, čuchom a pod.) na motorické impulzy ***y*** (pričom táto transformácia je závislá od vnútorného stavu ***s*** (pozri obr. 6, diagram A). Mozog môžeme interpretovať ako mohutný paralelný počítač realizovaný neurónovou sieťou, ktorý transformuje vstupnú informáciu ***x*** na výstupnú informáciu ***y***, pričom táto transformácia je ovplyvňovaná vnútorným stavom ***s***. (pozri obr. 6, diagram B) Táto neurovedná interpretácia mozgu na

mikroskopickej (neurálnej) úrovni neumožňuje priame štúdium vyšších kognitívnych aktivít (riešenie problémov, porozumenie ľudskej reči, a pod.). Nechcem týmto povedať, že je to principiálne nemožné, je to možné, ale je to veľmi ťažkopádne a komplikované. Je to podobné tomu, ako keby sme chceli študovať makroskopickú vlastnosť vody "*povrchové napätie*" metódami kvantovej mechaniky. Samozrejme, že túto veličinu môžeme študovať aj takto na "mikroskopickej" úrovni, ale je to tak numericky ako aj teoreticky náročné. Podstatne jednoduchší je fenomenologický prístup založený na makroskopickej termodynamike, kde dostaneme pre túto veličinu pomerne jednoduchú formulu obsahujúcu jednoduché experimentálne údaje. Podobná situácia existuje aj v teórii mozgu-mysle. Neurálny (konekcionistický alebo subsymbolický) prístup je vhodný na štúdium elementárnych kognitívnych aktivít (napr. prvotné spracovanie vizuálnej informácie zo sietnice oka). Vyššie kognitívne aktivity mozgu sú študované symbolickým alebo kognitivistickým prístupom založeného na predstave, *že ľudský mozog je počítač,* ktorý pracuje podľa týchto princípov (ktoré tvoria základ tzv. *symbolickej paradigmy*), ktorý

(1) transformuje symboly pomocou syntaktických pravidiel na iné symboly, pričom

(2) myšlienky sú symbolické reprezentácie implementované pomocou jazyka myslenia, a

(3) mentálne procesy sú kauzálne sekvencie symbolov generované syntaktickými pravidlami.



Obrázok 10.6. (A) Kybernetická interpretácia mozgu ako „zariadenia“, ktoré transformuje vstup ***x*** na výstup ***y***, pričom táto transformácia je ovplyvnená vnútorným stavom ***s*** . Touto špecifikáciou mozgu môžeme dostať dve rôzne odozvy $\boldsymbol{y}_1$ a $\boldsymbol{y}_2$ na rovnaký vstup ***x***. (B) Konekcionistický (neurálny) model mozgu pomocou neurónovej siete, ktorá obsahuje (1) vstupné neuróny (napr. percepčné neuróny retiny oka), (2) skryté neuróny, na ktorých prebieha transformácia vstupu na výstup a (3) výstupné neuróny (napr. neuróny riadiace motorické aktivity). Aktivity skrytých neurónov tvoria vnútorný stav neurónovej siete, rôzne počiatočné nastavenie týchto aktivít zapríčiňuje rôznu odozvu na rovnaké vstupné aktivity ***x***.

Použitie termínu „počítač“ obvykle evokuje predstavu sekvenčného počítača von neumannovskej architektúry (napr. personálne počítače majú túto architektúru), kde je možné striktne oddeliť hardware od software; kde na tom istom počítači – hardware môže byť vykonávaných nepreberné množstvo rozdielnych programov – softwarov. Pre tieto počítače existuje striktná dichotómia medzi počítačom a programom – t. j. medzi hardwarom a softwarom. Žiaľ, paradigma mysle ako počítača implikuje u mnohých ľudí predstavu, že je možné oddeliť mozog od mysle, ako dva „nezávislé“ fenomény, kde mozog hrá úlohu hardwaru, zatiaľ čo myseľ je software (vykonávaný na hardwaru – mozgu).

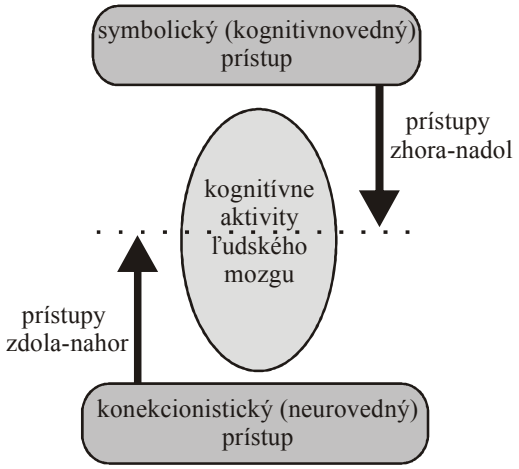
Obráťme našu pozornosť na moderný neurovedný prístup k chápaniu vzťahu medzi mozgom a mysľou [xx], ktorý je založený na neurovednom (hovoríme tiež o konekcionistickom alebo o subsymbolickom) poňatí tak mozgu, ako aj mysle. Základná predstava o mozgu (založená na experimentálnych neurovedných poznatkoch) je, že je tvorený z neurónov navzájom poprepájaných pomocou jednosmerných synaptických spojov (pozri obr. 6, diagram B). Možno konštatovať, *že schopnosť mozgu vykonávať nielen*

kognitívne aktivity, ale byť aj pamäťou, je plne zakódovaná do jeho architektúry. Na základe týchto neurovedných poznatkov bazálneho charakteru môžeme konštatovať, že počítačová paradigma ľudského mozgu sa musí formulovať tak, že mozog je paralelne distribuovaný počítač (obsahujúci mnoho miliárd neurónov, elementárnych procesorov, ktoré sú medzi sebou poprepájané do zložitej neurónovej siete). Program v tomto paralelnom počítači je priamo zabudovaný do architektúry neurónovej siete, t. j. ľudský mozog je jednoúčelový paralelný počítač reprezentovaný neurónovou sieťou, ktorý nie je možné preprogramovať bez zmeny jeho architektúry. Z týchto všeobecných úvah vyplýva, že myseľ s mozgom tvoria jeden integrálny celok; myseľ je v tomto prístupe možné chápať ako program vykonávaný mozgom, avšak tento program je špecifikovaný architektúrou distribuovanej neurónovej siete reprezentujúcej mozog. Mozog a myseľ tvoria dva rôzne pohľady na ten istý objektu:

(1) Keď hovoríme o mozgu, myslíme tým „hardwarovú“ štruktúru, biologicky realizovanú neurónmi a ich synaptickými spojmi (formálne reprezentovanú neurónovou sieťou), v opačnom prípade,

(2) keď hovoríme o mysli, myslíme tým kognitívne a iné aktivity mozgu, realizované výpočtami neurónovej siete reprezentujúcej mozog.

Aký je vzťah medzi neurovednou (konekcionistickou alebo subsymbolickou) distribuovanou reprezentáciou a symbolickou (kognitivistickou) lokalistickou reprezentáciou? Dichotomické chápanie týchto dvoch reprezentácií je dôsledkom ostrého rozlišovania medzi konekcionistickým a symbolickým prístupom, medzi tým, čo je explicitné a čo je implicitné. V konekcionistickej reprezentácii je explicitne určená len najnižšia úroveň (pri pohľade zdola nahor, pozri obr. 1) , t. j. obrazce aktívnych neurónov. O riešenie tohto pomerne zložitého problému sa prvý pokúsil Smolensky [xx], navrhol hierarchický konekcionistický model, pomocou ktorého sa pokúsil zosúladiť symbolizmus s konekcionizmom. Pri postupe zdola nahor v rámci tohto modelu nachádzame poznatkové atómy s veľkým stupňom implicitnosti, ktoré sa v určitom priblížení môžu interpretovať už ako symboly (pozri obr. 7).



Obrázok 10.7. Schematické znázornenie vzťahu medzi konekcionistickým a symbolickým prístupom k interpretácii kognitívnych aktivít ľudského mozgu. Interpretatívna sila tohto grafu je založená na existencie teorémy, ktorá pochádza od informatika a logika S. Kleene [xx]. Podľa tejto slávnej vety, každá neurónová sieť (konekcionistické výpočtové zariadenie) je ekvivalentná konečno-stavovému stroju (automatu, symbolickému výpočtovému zariadeniu). To znamená, idúc v grafe zdola-nahor, hľadáme také symbolické koreláty, ktoré integrujú aktivity jednotlivých neurónov. Ak ideme opačným smerom zhora-nadol, potom hľadáme konekcionistické koreláty pre symbolické výrazy. Môžeme teda konštatovať, že medzi konekcionistickým a symbolickým prístupom k štúdiu kognitívnych aktivít neexistuje exkluzívna disjunkcia, hlavným kritériom miery zahrnutia prvého alebo druhého prístupu je v pohodlnosti a jednoznačnosti štúdiu vybraných kognitívnych aktivít. V súčasnosti sa ustálil kompromis, že vyššie kognitívne aktivity sa obvykle študujú na symbolickej

úrovni (aj keď je možné použiť k ich štúdiu konekcionistický prístup), zatiaľ čo, nižšie kognitívne aktivity sa študujú pomocou konekcionistického formalizmu. Pre úplnosť, už len poznamenajme, že pred nedávnom publikoval izraelsko-anglický logik D. Gabbay [xx] so svojimi spolupracovníkmi dve úspešné knihy, ktoré sa zaoberajú neurónovými sieťami vykonávajúcimi logické usudzovanie.

Realistický názor na tieto dva prístupy je, že poskytujú dva alternatívne pohľady na ten istý problém, pričom symbolizmus je vhodný na interpretáciu vyšších kognitívnych aktivít ľudského mozgu, zatiaľ čo konekcionizmus interpretuje nižšie kognitívne aktivity (napr. vnímanie). Alternatívna formulácia tohto pohľadu je, že symbolizmus sa môže chápať ako prístup "zhora-nadol", ktorý interpretuje vyššie kognitívne aktivity pomocou symbolických prístupov známych z umelej inteligencie, pričom si musíme uvedomovať skutočnosť, že navrhovaný model musí mať aj určitú "konekcionistickú" plauzibilitu, že substrátom ľudského myslenia je mozog, ktorého architektúra je výlučne konekcionistická. Na druhej strane, konekcionistické prístupy k interpretácii kognitívnych aktivít ľudského mozgu sú založené na neurónových sieťach a predstavujú prístup "zdola-nahor". Pri aplikovaní konekcionistických metód k interpretácii vyšších kognitívnych aktivít ľudského mozgu je však potrebné zavádzať hypotetické bloky (moduly) vykonávajúce špeciálne aktivity, ktoré už majú veľmi blízko k blokovej štruktúre symbolického prístupu. V ideálnom prípade budeme očakávať, že sa tieto dva prístupy stretnú na polceste (vyznačenej prerušovanou horizontálnou čiarou na obr. 9), tak napr. konekcionistické prístupy budú poskytovať interpretáciu modulov používaných v symbolickom prístupe. Inak povedané, konekcionizmus poskytuje symbolickému prístupu "mikroskopickú" teóriu jeho fenomenologických pojmov, ktorá je v súlade so súčasnými predstavami o štruktúre a fyziológii ľudského mozgu.

Na záver tejto kapitoly povieme niekoľko poznámok o pamäti, o jej symbolickej a subsymbolickej realizácii. Najjednoduchší prístup k pamäti je symbolický, pamäť stotožňujeme s množinou poznatkov (teóriou) reprezentovaných prvorádovými formulami

$$T = \{\varphi_1, \varphi_2, ..., \varphi_n\} \tag{6}$$

Budeme implicitne predpokladať, že táto množina je konzistentná, t. j. z tejto teórie je vyvoditeľný buď záver χ alebo jeho negáciu $\neg\chi$ (ale nie oba, požívame logickú spojku „exkluzívnu disjunkciu“); v opačnom prípade by sme mohli odvodiť z teórie ľubovoľnú formulu, čo je nezmysel. Nad takto definovanou pamäťou môžeme vykonávať proces revízie (pozri kapitolu 3.3), ktorej cieľom je konštrukcia konzistentnej teórie.

Podstatne zložitejší problém je špecifikácia pamäti v subsymbolickom prístupe založenom na stotožnení mozgu s neurónovou sieťou. Pretože mozog je len a len neurónovou sieťou, *pamäť musí existovať* (bez nej nemôže mozog fungovať ako efektívne výpočtové zariadenie, ktoré na základe aktuálnych a minulých skúsenosti predpovedá blízku budúcnosť), potom pamäť musí byť nejakým spôsobom obsiahnutá aj v našom mozgu – neurónovej sieti. Na základe tohto poznatku môžeme usúdiť, že pamäť nie je lokalistická ale je distribuovaná nad neurónovou sieťou. O tejto skúsenosti svedčia mnohé experimentálne poznatky neurovedy; napríklad pri degeneratívnom ochorení mozgu (pri Alzheimerovom onemocnení, keď postupne nastáva zániku spojov medzi neurónmi, nedochádza k okamžitému zániky poznatkov ale k pomalému a postupnému, čo sa v neurovede vyjadruje angl. termínom „graceful degradation of memory”). Z týchto jednoduchých úvah založených na neurovedných poznatkoch bazálneho charakteru vyplýva, že v neurónovej sieti je zakódovaný nielen výpočtový charakter mozgu ale aj pamäť.

10.3.2 Symbolická špecifikácia mysle agenta

K špecifikácii mysle použijeme jednoduchý *symbolický prístup* založený na elementárnej prvorádovej logike [xx]. Postulujme, že v multiagentovom systéme $\mathcal{A} = \{A_1, A_2, ..., A_n\}$ každý agent $A \in \mathcal{A}$ obsahuje *teóriu* - množinu poznatkov $T_A = \{\varphi_1, \varphi_2, ...\}$ (kde $\varphi_1, \varphi_2, ...$ sú formule prvorádovej logiky, ktoré reprezentujú poznatky agenta) a *inferenčný aparát* reprezentovaný reláciou logického vyplývania $\vdash_A$. Inferenčný aparát môže byť reprezentovaný elementárnymi zákonmi r_1, r_2, r_3,... prirodzenej dedukcie [xx], tak napr. obsahuje pravidlá modus ponens a modus tollens, hypotetický sylogizmus, konkretizácia univerzálneho kvantifikátora a pod.

$$\vdash_A = \{r_1, r_2, r_3, ...\} = \left\{ \underbrace{\frac{p \Rightarrow q \quad p}{q}}_{r_1}, \underbrace{\frac{p \Rightarrow q \quad \neg q}{\neg p}}_{r_2}, \underbrace{\frac{p \Rightarrow q \quad q \Rightarrow r}{p \Rightarrow r}}_{r_3}, \underbrace{\frac{(\forall x) P(x)}{P(t)}}_{r_4} ... \right\} \tag{7a}$$

Použitím tohto inferenčného aparátu je agent schopný dedukovať na základe pozorovania $\Delta = \{\psi_1, \psi_2, ...\}$, kde $\psi_1, \psi_2, ...$ sú poznatky tohto pozorovania, nový poznatok χ

$$T_A \cup \Delta \vdash_A \chi \tag{7b}$$

Formálne, myseľ agenta *A* je špecifikovaná usporiadanou dvojicou $\mathcal{M}_A = (T_A, \vdash_A)$, t. j. myseľ je špecifikovaná pamäťou tvorenou množinou poznatkov T_A a inferenčnou reláciou $\vdash_A$. Môžeme povedať, že teória tvorená poznatkami reprezentuje individuálnu pamäť agenta a inferenčná relácia reprezentuje mozog, ktorý je schopný usudzovať.

10.3.3 Revízia poznatkov

Problém revízie poznatkov má vo filozofii a v logike dlhú tradíciu. Stal sa integrálnou súčasťou mnohých traktátov a monografií z filozofie poznania (epistemológie) a logiky od staroveku až po súčasnosť, ktoré obsahovali rozsiahle kapitoly, kde väčšinou na fenomenologicko-špekulatívnej úrovni sa formulovali zásady nášho myslenia, argumentácie a zmeny poznatkov vyplývajúcich zo zmien východiskových predpokladov, ich rozšírenia alebo čiastočnej falzifikácie. Táto zaujímavá problematika sa stala súčasťou aplikovanej matematickej logiky až koncom minulého storočia, kedy jej formálny aparát bol použitý na formuláciu teórie revízie poznatkov. Taktiež, problém revízie poznatkov sa stal v súčasnosti integrálnou časťou informatiky a umelej inteligencie a tým musel prejsť z roviny všeobecno–špekulatívnej k rovine formálno–exaktnej, ktorá môže slúžiť ako základ algoritmizácie problematiky revízie poznatkov na počítačoch.

Tak napríklad, ak chceme špecifikovať základné epistemické princípy vývinu vedy, dynamiku jej vývoja v čase, tento cieľ môže byť realizovaný na abstraktnej úrovni tak, že študujeme danú konzistentnú databázu poznatkov. Táto databáza je v čase postupne modifikovaná elementárnymi operáciami, akými sú dodanie nového poznatku a odstránenie pôvodného poznatku. V obidvoch prípadoch tieto operácie zmeny môžu ovplyvňovať ostatné poznatky, preto sa vykonáva ich revízia, aby sa odstránili prípadné nekonzistentosti. V počiatočnom období vzniku teórie revízie poznatkov (80. roky minulého storočia) základné idey boli formulované švédskym kognitívnym vedcom P. Gärdenforsom a dvojicou amerických logikov C. Alchourrrónom a D. Makinsomom, ktorí spoločne v r. 1985 publikovali významnú prácu [xx] (ktorá sa v odbornej literatúre označuje akronymom AGM počiatočných písmen ich priezvisk v abecednom poriadku jej autorov), v ktorej formulovali základné princípy, koncepcie a konštrukcie teórie revízie poznatkov.

Poznamenajme, že myseľ $\mathcal{M}_A$ nie je konštantnou – nemennou v priebehu existencie agenta A, ale môže sa meniť prostredníctvom procesu revízie, kde sa buď pridáva alebo uberá nejaký poznatok tak, aby výsledná pamäť bola konzistentná. Budeme rozlišovať tieto štyri spôsoby revízie databázy poznatkov T:

(1) ***Expanzia***, nové poznatky, reprezentované formulami z konzistentnej množiny $\tilde{T}'$ sú dodané do pôvodnej konzistentnej teórie T, za vzniku novej konzistentnej teórie $T' = T \cup \tilde{T}$, čo indukuje rozšírenie jej logických dôsledkov o nové poznatky, ktoré sú logickým dôsledkom rozšírenej teórie Φ'. Databáza poznatkov priradená rozšírenej teórii Φ' je označená

$$T_{\tilde{T}}^{(+)} =_{def} T \cup \tilde{T} \tag{8}$$

(2) ***Kontrakcia***, minimálna podmnožina $\Delta \subseteq T$ je odstránená z nekonzistentnej teórie T za vzniku maximálnej konzistentnej podmnožiny – teórie $T' = T - \Delta$. Databáza poznatkov priradená zúženej teórii T' je označená

$$T_{\Delta}^{(-)} =_{def} T - \Delta \tag{9}$$

(3) ***Revízia*** je kombináciou predchádzajúcich dvoch operácii expanzie a kontrakcie, pôvodne konzistentná teória T je expanziou rozšírená o nové poznatky z $\tilde{T}$, čím sa stane nekonzistentnou, z takto rozšírenej teórie je pomocou kontrakcie je odstránená podmnožina $\Delta \subseteq T \cup \tilde{T}$, výsledkom tohto procesu je konzistentná teória $T' = \left(T \cup \tilde{T}\right) - \Delta$. Pôvodná databáza T priradená modifikovanej rozšíreno – zúženej teórii T' je označená

$$T_{\tilde{T},\Delta}^{(+,-)} =_{def} \left(T \cup \tilde{T}\right) - \Delta \tag{10}$$

(4) ***Spájanie*** poznatkov môžeme špecifikovať takto: Predstavme si multiagentový systém, kde úlohou každého agenta – prieskumníka je zbierať poznatky o danom prostredí (napr. ako sú rozložené prekážky, potrava a ostatní agenti). Po určitom čase agent sa vráti do „centrály“ a odovzdá svoje poznatky agentovi - integrátorovi, ktorý má na starosti spájanie poznatkov získaných od rôznych agentov – prieskumníkov do jednej databázy poznatkov o danom prostredí. Môže sa stať, že poznatky získané dvoma rôznymi agentmi sú navzájom kontroverzné, t. j. agent – integrátor stojí pred úlohou rozhodnúť sa, ktoré poznatky budú/nebudú použité pri vytváraní spojenej databázy poznatkov. Spájanie poznatkov $T_1, T_2, ... T_n$ do jedného spoločného poznatku môže prebiehať tak, že všetky prijaté poznatky spojíme do jednej množiny $T_1 \cup T_2 \cup ... \cup T_n$, v prípade jej nekonzistentnosti hľadáme takú minimálnu množinu poznatkov, ktorú keď odstránime zo zjednotenej množiny, potom výsledná množina (teória) je konzistentná

$$merge\left(T_1, T_2, ... T_n\right) =_{def} \left(T_1 \cup T_2 \cup ... \cup T_n\right) - \Delta \tag{11a}$$

Existuje mnoho alternatívnych možností spájania rôznych databáz poznatkov do jednej konzistentnej databázy. Tak napr. spojenú databázu môžeme postupne vytvárať tak, že integrátor postupne pripája k už spojeným databázam novú práve prichádzajúcu databázu poznatkov a z výsledku tohto spojenia odstránime minimálnu podmnožinu z predchádzajúceho výsledku tak, aby výsledné aktuálne spojenie bolo konzistentné

$$merge\left(T_1, T_2, T_3, T_4 ...\right) =_{def} \left(\left(\left(\left(\left(T_1 \cup T_2\right) - \Delta_1\right) \cup T_3\right) - \Delta_{12}\right) \cup T_4\right) - \Delta_{123}\Big)... \tag{11b}$$

V prvom východiskovom kroku máme databázu T_1, túto rozšírime o databázu T_2, pričom zo vzniklého zjednotenia odstránime minimálnu podmnožinu $\Delta_1 \subseteq T_1$ tak aby sme zachovali konzistewntnosť výsledku. V druhom kroku pridáme databázu T_3 k predchádzajúcemu výsledku a odstránime minimálnu množinu $\Delta_{12} \subseteq \left(T_1 \cup T_2\right) - \Delta_1$, atď.

Problém spájania poznatkov do jednotného väčšieho celku – ***spojenej databázy poznatkov*** – patrí v umelej inteligencii medzi základné problémy distribuovanej inteligencie (napr. multiagentových systémov). Tento problém môžeme byť v rámci prvorádovej logiky jednoducho formalizovaný, pričom na sémantickej úrovni môžeme špecifikovať „mimologické“ prostriedky ako riešiť nekonzistentnosti vyskytujúce sa pri spájaní nových poznatkov.

10.3.4 Tri stupne zložitosti agentov

Budeme rozlišovať tri stupne zložitosti (alebo „evolúcie“) agentov, táto klasifikácia vychádza z ich vlastností, či obsahujú alebo neobsahujú pamäť, ak obsahujú pamäť, potom aký jej typ, či individuálnu alebo jej rozšírenie aj o kolektívnu pamäť (pozri obr. 8).

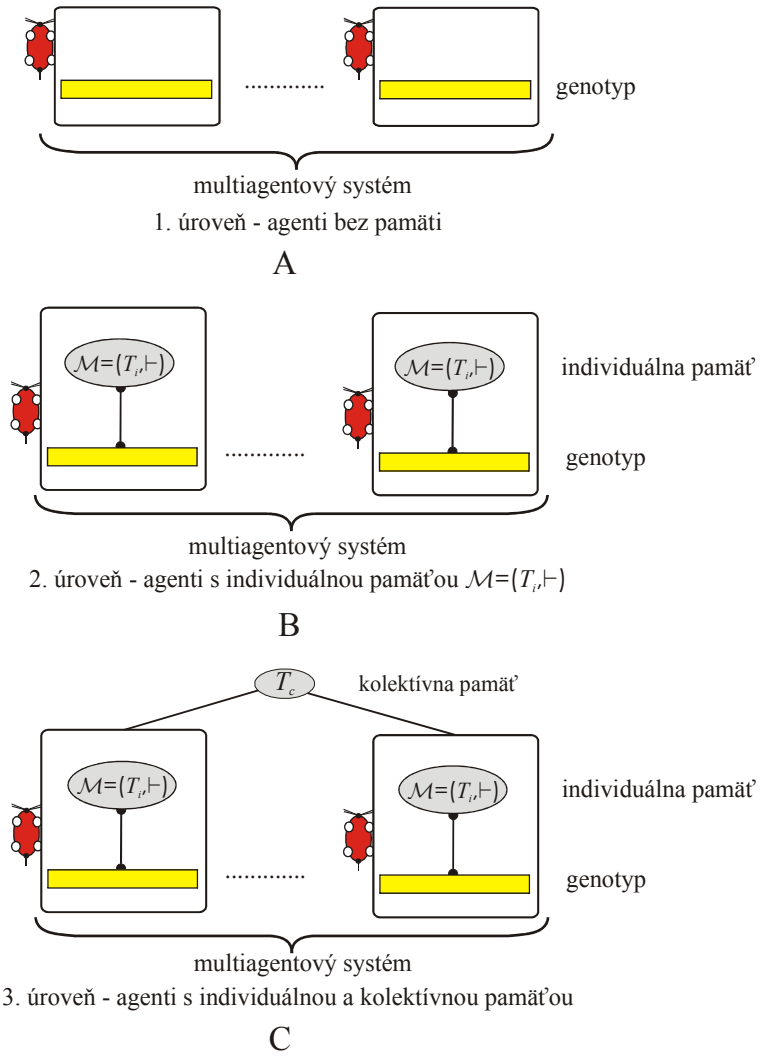
1. stupeň. Aktivity agenta sú *reflexívne*, t. j. nie sú založené na inferenčnom procese mysle. Tento prístup je v informatike reprezentovaný štandardnými evolučnými algoritmami, kde fitnes genotypu je priamo určený jeho zložením. Reflexívne aktivity agenta sú plne determinované geneticky, t. j. inštinktívne; fitnes agenta je plne určené jeho genotypom.

2. stupeň. Agenti populácie majú myseľ $\mathcal{M} = (T_i, \vdash)$, pomocou ktorej môžu riadiť jednoduché aktivity spočívajúce v efektívnejšej orientácii v prostredí, vyhodnocovaní blízkeho okolia, a pod. Obvykle, relácia logického vyplývania $\vdash$ je rovnaká pre všetkých agentov, môžeme konštatovať, že inferenčný aparát agenta je špecifikovaný genotypom. Individuálna pamäť agenta T_i môže mať memetický mechanizmus [xx] svojho pôvodu:

(a) horizontálnym prenosom (výchova potomkov) od rodiča,

(b) vertikálnym prenosom (od iného agenta),

(c) vlastným pozorovaním, t. j. individuálna pamäť je postupne rozširovaná o skúsenosti agenta.



Obrázok 10.8. Diagramatická ilustrácia troch stupňov zložitosti agentov (myseľ je na diagramoch reprezentovaná databázou poznatkov T a inferenčnou reláciou vyplývania $\vdash$). (A) Diagram znázorňuje prvý stupeň agenta, ktorý nemá pamäť, jeho aktivity v multiagentovom systéme sú plne špecifikovaná jeho genotypom, t. j. sú plne reflexívneho charakteru, t. j. fitnes genotype je plne určené len jeho zložením.. (B) Diagram znázorňuje druhý stupeň zložitosti agentov, ktoré už majú vlastnú individuálnu myseľ $\mathcal{M}=(T_i,\vdash)$, kde databáza T_i reprezentuje individuálnu pamäť agenta a relácia vyplývania $\vdash$ reprezentuje inferenčný aparát agenta. (C) Najzložitejší tretí stupeň zložitosti mysle, ktorý dostaneme z 2. úrovne, keď z jednotlivých individuálnych pamätí vytvoríme „spájaním“ individuálnu poznatkov kolektívnu pamäť T_i, ktorá je pre celý multiagentový systém rovnaká.

Na záver životného cyklu, keď agent vstupuje do reprodukčného procesu, svojmu potomkovi odovzdáva len svoj gén v ľahko zmutovanej forme, t. j. agent - potomok nededí priamo od rodičov pamäť (t. j. evolúcia je striktne darvinovská), ale dedí v ľahko zmutovanej forme jeho inferenčný aparát $\vdash$.

3. stupeň. Agenti vybavení mysľou $\mathcal{M}=(T_i,\vdash)$ majú taktiež *kolektívnu pamäť* T_c, ktorá je spoločná pre celý multiagentový systém. Najjednoduchší prístup k existencii kolektívnej pamäti je postulovanie tzv. „vyhľadávacích“ agentov, ktorý majú privilégium vstupovať do individuálnych pamätí jednotlivých agentov a vyhľadávať vhodnú informáciu všeobecného charakteru, ktorá bude tvoriť kolektívnu pamäť. Títo špecializovaní agenti pri tvorbe

kolektívnej pamäti majú tiež dôležitú úlohu, aby zabezpečili je *konzistentnosť*, t. j. nesmie obsahovať súčasne nejakú informáciu χ a jej negáciu $\neg\chi$ (poznamenajme, že v logike je problém vytvárania, revízie a spájania databáz poznatkov veľmi intenzívne študovaný [xx]). .

10.3.5 Symbolická špecifikácia kolektívnej pamäti a jej význam v umelej inteligencii

Táto kapitola je pokračovaním kapitoly 3.2, kde sme špecifikovali dvojzložkovú myseľ agenta A_i ako usporiadanú dvojicu $\mathcal{M}_i = (T_i, \vdash)$, kde T_i je pamäť *i*-teho agenta A_i a $\vdash$ je relácia logického vyplývania reprezentujúca schopnosť agenta deduktívneho usudzovania nad svojou pamäťou. Relácia vyplývania nie je indexovaná, to znamená, že predpokladáme rovnakú schopnosť usudzovať v celom multiagentovom systéme (nie je problém odstrániť tento unitársky predpoklad a postulovať, že agent A_i má svoju špecifickú reláciu vyplývania $\vdash_i$, ktorá sa líši od ostatných relácii nielen počtom ale aj typom deduktívnych pravidiel *r´*s).

Pamäť T_i každého agenta rozdelíme na dve časti T_i' a T_i''

$$T_i = T_i' \cup T_i'' \tag{12}$$

kde sme oddelili od pamäti T_i tie poznatky T_i'', ktoré sú dôležité aj pre ostatných agentov (napr. v pravom hornom rohu prostredia sa nachádza zariadenie na dobýjanie energie za malý počet kreditov alebo agent A_α je schopný vždy pomôcť s hľadaním riešenia pomocou vzájomného prepojenia $T_i \cup T_\alpha$ pamätí). Najjednoduchší spôsob ako zabezpečiť vznik kolektívnej pamäti je postulovať existenciu niekoľkých špeciálnych vyhľadávacích agentov, ktorí si od agenta A_i prevezmú „kolektívnu“ časť jeho pamäti T_i'', ktorá obsahuje poznatky tohto agenta o regularitách a vlastnostiach prostredia, ktoré môžu byť užitočné aj pre ostantých agentov. Potom kolektívna pamäť, ktorá je spoločná pre celý multiagentový systém môže byť vytváraná takto

$$T_{coll}^{(b+1)} = \left\langle T_{coll}^{(b)} \cup \left(\bigcup_i T_i'' \right) \right\rangle \tag{13}$$

kde symbol $\langle T \rangle$ znamená, že nad teóriou T bola vykonaná revízia, t. j. z T bola odstránená minimálna množina Δ tak, že $\langle T \rangle = T - \Delta$ je konzistentná teória.

Pri vytváraní kolektívnej pamäti T_{coll} môže byť využívaná dodatočná podmienka, ktorá preferuje len istý typ poznatkov vyberaných z lokálnych kolektívnych pamätí *T´*s. Týmto môžeme zabezpečiť, že kolektívna pamäť sa vytvára pomocou fixného kritéria, ktoré sa obvykle len veľmi vzácne mení v priebehu evolúcie multiagentového systému Hlavnou úlohou vytváranej kolektívnej pamäti je napomáhať agentom pri vyájomnej coyiálnej a kultúrnej interakcii, pri orientovaní sa v prostredí, pri pri tvorbe kroku, ktorým sa zmení aktuálny stav prostredia na nový (nasledujúci) stav prostredia, a pod. Jedna z hlavných úloh kolektívnej pamäti je preklenutie medzigeneračných informačných bariér, v kolektívnej pamäti sa uchováva podstatná informácia z histórie adaptívneho procesu multiagentového systému, ktorá môže podstatne uľahčiť hľadanie optimálneho kroku v množine alternatívnych možností, t. j. vybrať taký krok, ktorý priblíži systém k plánovanému cieľu.

Jednoduché ilustratívne príklad kolektívnej pamäti pre genetický algoritmus a genetické programovanie boli už diskutované v kapitole 2.1, ktorá sa zaoberala špecifikáciou evolučných algoritmov a ktoré spolu s teóriou multiagentových systémov tvoria efektívny prostriedok pre simuláciu sociálnych systémov. Bol ukázané, že koncepcia kolektívnej pamäti implementovaná do týchto dvoch evolučných algoritmov poskytuje podstatnú akceleráciu týchto evolučných prístupov, je založená na histórii algoritmu a umožňuje produkovať namiesto úplne náhodných mutácií tyv. cielené mutácie, ktoré vzplývajú y histórie algoritmu.

10.4. Záver

V práci bolo ukázané, že koncepcia kolektívnej pamäti, ktorá bola zavedená už viac ako pred polstoročím francúzskym sociológom M. Halbwachsom [xx] ako unikátny typ pamäti, ktorá je súčasťou kultúry, patrí medzi základné špecifikácie sociálnej skupiny a je prenášaná negentickými (kultúrnymi) metódami v reprodukčnom medzigeneračnom procese v multiagentovom systéme. V priebehu 90. rokov minulého storočia tento termín sa stal taktiež populárnym aj v informatickej disciplíne "umelá inteligencia", menovite v jej časti nazývanej "multiagentové systémy". Táto koncepcia sa v teórii multiagentových systémov využíva ako akceleračný prvok adaptácie systému, kolektívna pamäť zahŕňa skúsenosti a vedomosti agentov z histórie tohto procesu. Táto skutočnosť bola ilustrovaná jednoduchými príkladmi genetického algoritmu a genetického programovania, kde kolektívna pamäť je postupne inkrementálne budovaná a zavádza určitý determinizmus do mutácií v reprodukčnom proces založený na predchádzajúcich skúsenostiach adaptačného procesu.

Vychádzajúc zo súčasných výsledkov a trendov aplikácií multiagentových systémov v sociálnych vedách, môžeme formulovať hypotézu, že táto informatická disciplína tvorí pre sociálne vedy špecifický teoretický základ, ktorá je schopná poskytovať racionálne vysvetlenia mnohých sociálnych javov a procesov, pričom tieto vysvetlenia nie sú zaťažené zbytočným ideologickým balastom. Klasickým príkladom tejto skutočnosti sú Axelrodove výsledky [xx] vzniku kooperácie medzi sebeckými agentmi. Výsledky týchto simulačných výpočtov môžeme zosumarizovať tak, že evolučný návrh heuristík, ktoré optimalizujú dlhodobý zisk pri opakovanej interakcii medzi agentmi, vedie k návrhu pomerne jednoduchej stratégie – heuristiky tit-for-tat, ktorá má širšiu platnosť aj ako „racionálna a morálna heuristika“, návod pre dlhodobú opakujúca sa kooperatívnu interakciu medzi agentmi. Tento prístup bol zovšeobecnený aj pre štúdium etnických konfliktov [xx] medzi dvoma etnicky rôznymi populáciami. Získané výsledky racionalizujú koncepciu „kolektívnej viny“, okolo ktorej v literatúre existuje množstvo navzájom protirečivých názorov, od zásadného odmietania až po jej schvaľovanie vo zvláštnych prípadoch. Pomocou simulačných výpočtov bolo dokázané, že princíp kolektívnej viny má fatálne následky na kooperáciu medzi rôznymi etnikami, ktorá nemôže vzniknúť, pretože príslušník inej etnickej skupiny je automaticky penalizovaný princípom kolektívnej viny.

V sociológii a v humanitných vedách koncepcia „kolektívnej pamäti" sa čiastočne prekrýva s memetikou[1]. Podobne ako v kolektívnej pamäti, aj v memetike jej význam hlavne spočíva v tom, že informácia obsiahnutá v méme zvyšuje šancu agenta prežiť, napr, môže byť v ňom obsiahnutý popis prostredia, jeho nepravidelnosti, skryté vlastnosti a pod. Memetický prístup môže byť preformulovaný do podoby kolektívnej pamäti [xx] obsahujúcej „historické skúsenosti“ agentov z predchádzajúcej histórie adaptačného procesu.

Na záver tejto stati musíme konštatovať, že Halbwachbsova koncepcia poskytla teórii multiagentových systémov efektívny formálny nástroj k akcelerácii adaptačných procesov, ktorý umožňuje preklenúť medzigeneračnú informačnú bariéru jej prenosom do ďalších generácií kultúrnymi mechanizmami.

[1] Očistenej o zbytočné a zavádzajúce "biologické" predstavy evolučného biológa R. Dawkinsa, ktoré boli použité pri formulácii memetiky v jeho slávnej knihe The Selfish Gene [xx]. V súčasnosti pod memetikou [xx] rozumieme špeciálny multiagentový prístup, kde každý agent má pridružený mém, určitú informáciu, ktorá zvyšuje jeho fitnes. Pri reprodukčnom procese, táto informácia - mém sa prenáša na potomkov kultúrnym prenosom, t. j. učením. Adaptačný proces tohto systému sa realizuje pomocou kooperatívnej koevolúcie, kde existuje populácia génov a populácia mémov, spolu interagujú pri výpočtu ich fitnes.

Literatúra

Alchourròn, C. E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, **50** (1985), 510-530.

Assmann, J.: Communicative and Cultural MemoryManier. In Erll, A., Nünning, A., Young, S. B. (eds.): *Cultural Memory Studies. An International and Interdisciplinary Handbook.* Walter de Gruyter, Berlin, 2008, pp. 109–118.

Axelrod, R.: *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, NJ, 1997.

Axelrod, R.: *The Evolution of Cooperation*. Basic Books, New York, 1984.

Beim, A.: The Cognitive Aspects of Collective Memory. *Symbolic Interaction*, **30** (2007), pp. 7–26.

D´Avila Garcez, A. S. ; Lamb, L. C.; Gabbay, D. M.: *Neural-Symbolic Cognitive Reasoning*. Springer, Berlin 2009.

Dudai, Y.: *Memory From A to Z, Keywords, Concepts, and Beyond*. Oxford University Press, Oxford, UK, 2002, heslo "Collective memory", pp. 51-52.

Epstein, J. M., Axtell, R.: *Growing Artificial Sociedties. Social Sciences from the Bottom Up*. The MIT Press, Cambridge, MA, 1996.

Gabbay, D.M., Rodrigues, O.T., Russo, A.: *Revision, Acceptability, and Context. Theoretical and Algorithmic Aspects.* Springer, Berlin, 2010

Gärdenfors P.: *Knowledge in Flux: Modeling the Dynamics of Epistemic States,* MIT Press, Bradford Books, Cambridge, MA, 1988.

Gärdenfors, P. (ed.): *Belief Revision*. Cambridge University Press, Cambridge (UK), 1992.

Halbwachs, M.: *Kolektivní pamět*, Vydavatelství Slon, Praha, 2010.

Haynes, T.: Collective Memory Search: Exploiting an Information Center for Exploration. Department of Mathematics & Computer Science, The University of Tulsa, Tulsa, OK, 1998 (unpublished technical report).

Henry L. Roediger III, H. L:, Dudai, Y., Fitzpatrick, S.: *Science of Memory: Concepts.* Oxford University Press, New York, 2007.

Holland, J. H.: *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, MI, 1975 (Reprinted by The MIT Press, Cambridge, MA, 1992).

Ivanička, K.: *Synergetika a civilizácia.* Alfa, Bratislava, 1988.

Konieczny, S., Perez, R.: Logic Based Merging. *Journal of Philosophical Logic* **40** (2011), 239–270.

Krempaský, J. a kol.: *Synergetika,* Veda SAV, Bratislava, 1988.

Kvasnička, V. – Gál, E. – Pospíchal, J.: Memetika „in silico". In: J.Nosek (ed.): *Memy ve vědě a filosofii?*. Filosofia, Nakladatelství FU AV ČR, Praha, 2004. pp. 101-119.

Kvasnička, V. Replicator theory of coevolution of genes and memes. *Neural Network World* **13** (2003), 255-266.

Kvasnička, V., Pelikán, M., Pospíchal, J.: "Hill Climbing with Learning. An abstraction of Genetic Algorithm", *Neural Network World*, **5** (1996), 773-796.

Kvasnička, V., Pospíchal J.: Artificial Chemistry and Replicator Theory of Coevolution of Genes and Memes. *Collect. Czech. Chem. Commun.* **72** (2007) 223–251.

Kvasnička, V., Pospíchal, J.: A Multi-agent Study of Interethnic Cooperation. *Lecture notes in artificial intelligence*, vol. 2086, Springer 2001, pp. 415-435.

Kvasnička, V., Pospíchal, J.: Evolutionary study of interethnic cooperation. *Adv. Complex Systems* **2** (1999), 395-421.

Kvasnička, V., Pospíchal, J.: *Informatika pre sociálne vedy*. Vydavateľststvo Komenského university, Bratislava, 2006.

Kvasnička, V., Pospíchal, J.: *Matematická logika*. STU vydavateľstvo, Bratislava, 2006.

Kvasnička, V., Pospíchal. J., Tiňo, P.: *Evolučné algoritmy*. STU vydavateľstvo, Bratislava, 2000.

Lin, J., Mendelzon, A.O.: Merging databases under constraints. *International Journal of Cooperative Information Systems* **7** (1998), 55-76.

Luke, S.: *Essentials of metaheuristics*. Available for free at http://cs.gmu.edu/~sean/book/ metaheuristics, 2009.

Manier, D., Hirst, W.: A Cognitive Taxonomy of Collective Memories. In Erll, A., Nünning, A., Young, S. B. (eds.): *Cultural Memory Studies. An International and Interdisciplinary Handbook.* Walter de Gruyter, Berlin, 2008, pp. 253–262.

McCulloch, W. S., Pitts, W. H.: A Logical Calculus of the Ideas Immanent in nervous Activity. *Bulletin of Mathematical Biophysics* **5** (1943), 115 – 133.

Schenk, J.: *Samoorganizácia sociálnych systémov (Metodologické problémy synergetického prístupu v sociológii)*. IRIS, Bratislava, 1993.

Skyrms, B.: *Evolution of Social Contract*. Cambridge University Press, Cambridge, UK, 2010.

V. Kvasnička: Replicator theory of coevolution of genes and memes. 2nd Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence SAMI 2004, P.Sinčák, J. Fodor (eds.), ISBN 963715423X, pp. 7-30.

Wooldridge, M.: *An Introduction to Multiagent Systems.* J. Wiley & Sons, New York 2002.

Wooldridge, M.: *Reasoning about Rational Agents*. The MIT Press, Cambridge, MA, 2000.

11. prednáška

Kognitívna logika a nemonotónne usudzovanie

11.1 Úvodné poznámky

V kognitívnej vede (podobne ako aj v umelej inteligencii) logika má špecifické postavenie. Pomocou logiky môžeme algoritmizovať procesy usudzovania, ktoré prebiehajú v inteligentnom systéme (ľudskom alebo počítačovom, v ďalšom texte budeme používať neutrálny výraz „***agent***“ [xx]). Môžeme si položiť otázku, na akých princípoch je založený systém usudzovania (inferenčný systém) agenta. Pri optimistickom uvažovaní predpokladáme, že agent má k dispozícii kompletnú informáciu o danom probléme a kapacity jeho inferenčného systému (procesor, pamäť a pod.) sú také (v podstate neohraničené), že z týchto poznatkov je schopný vyvodiť každý relevantný poznatok o danom probléme. Pretože inferenčný systém pracuje s úplnou informáciou (s plnou bázou poznatkov), jeho teoretické princípy môžu byť založené na klasickej predikátovej logike alebo na špeciálnych neklasických logikách (napr. modálnej logiky). Hovoríme, že racionálne schopnosti tohto agenta sú ***neohraničené***, agent je ***omnipotentný*** (všemocný) a ***omniscientný*** (vševedúci).

Tieto klasické osvietenské ideály racionality[1], ktoré pretrvávajú v niektorých špeciálnych prípadoch až do súčasnosti, začali sa spochybňovať až v druhej polovici 20. storočia, ako dôsledok hlbšieho pochopenia princípov fungovania ľudskej mysle/mozgu. Rôzne psychologické experimenty jednoznačne poukazovali na kognitívnu ohraničenosť ľudskej mysle, ktorá má veľmi ďaleko od laplaceovskej univerzality a neohraničenosti. Je zásluhou kognitívnej psychológie [xx], keď pomocou experimentálnych techník, jednoznačne poukázala na ohraničené kognitívne schopnosti mysle. Z tohto pohľadu možno charakterizovať ľudskú myseľ ako nie veľmi dokonalý kognitívny aparát, ktorého impozantné výkony v zovšeobecňovaní a hľadaní riešení sú skôr dôsledkom veľkej systematickosti ľudí pri riešení problémov a v následnej zafixovanosti získaného riešenia v ľudskej pamäti a v „dedení“ týchto poznatkov pomocou kultúrnej vertikálnej a horizontálnej interakcie (výchova a vzdelávanie) medzi príslušníkmi ľudskej populácie[2].

Problém ohraničenej racionality bol prvýkrát explicitne formulovaný v 50. rokoch minulého storočia americkým ekonómom a psychológom E. H. Simonom [13] (ktorý je považovaný spolu so svojím blízkym spolupracovníkom A. Newellom za spoluzakladateľa umelej inteligencie). Renesancia tohto problému nastala pred nedávnom v r. 2002, keď bola udelená Nobelova cena za ekonómiu americkému kognitívnemu psychológovi Danielovi Kahnemanovi z University of Princeton, za vypracovanie integrovanej psychologickej metodiky pre potreby ekonomickej vedy ku skúmanie ľudského usudzovania a rozhodovania za prítomnosti neurčitosti. Tento významný psychológ tri desaťročia spolupracoval s Amosom Tverskym [5] (Stanford University, umrel v r. 1996), spolu sa zaslúžili o vznik *behaviorálnej ekonómie,* študovali problémy usudzovania, rozhodovania, výberu

[1] Tieto predstavy našli určité vyvrcholenie u francúzskeho osvieteneckého vedca a filozofa P. R. Laplaceho v jeho dikcii o superinteligentnej bytosti, ktorá pomocou matematických metód vidí svojim „zrakom“ všetky udalosti z minulosti, prítomnosti a aj budúcnosti, pre ňu neexistujú náhodné alebo neurčité udalosti.

[2] V súčasnosti je veľmi populárnou Dawkinsova memetika [1], ktorá rozvíja podobné názory.

a ohraničenej racionality v ekonomických vedách. Ich vedecké výsledky sú rovnako dôležité tak pre psychológiu, ako aj pre ekonómiu.

V súčasnosti ***paradigma agenta s ohraničenou racionalitou*** patrí medzi základné koncepcie umelej inteligencie a kognitívnej vedy. Pri formulovaní vlastností agenta už a-priori (realisticky) predpokladáme, že jeho kognitívny orgán je ohraničený (môžeme povedať, že je implementovaný konečnou neurónovou sieťou), jeho prístup k informácii je ohraničený tak kapacitne, ako aj časovo [xx]. Tento agent pri riešení dôležitých problémov jeho intencionálnych aktivít má k dispozícii len ohraničený čas. Môžeme si položiť otázku na akej logike je založený jeho inferenčný aparát, pomocou ktorého vyvodzuje nové poznatky z aktuálnej databázy poznatkov. Touto logikou nemôžu byť klasické monotónne logiky (napr. logika prvého rádu alebo modálna logika), pretože databáza poznatkov agenta je konečná a neúplná, klasické inferencie (napr. požitím metódy prirodzenej dedukcie) nad takýmto systémom budú obvykle rýchlo predčasne ukončené v dôsledku jej neúplnosti. Z týchto dôvodov musíme použiť iný typ logiky, ktorý umožňuje inferenciu aj nad neúplnou databázou poznatkov. V tejto práci popíšeme základné princípy nemonotónneho usudzovania, ktoré boli formulované na prelome 60. a 70. rokov minulého storočia. Pre tento typ približného usudzovania, ktorý sa veľmi blíži ideálom kognitívnej logiky, boli navrhnuté rôzne typy usudzovamia: defaultová logika, cirkumskripsia, autoepistemická logika a pod.

11.2 Monotónne a nemonotónne usudzovanie

Prv než pristúpime k špecifikácii nemonotónneho usudzovania, budeme špecifikovať jeho negáciu. ***Monotónne usudzovania*** je charakteristické tak pre klasickú logiku, ako aj pre väčšinu neklasických logík, ktoré sú striktne deduktívneho charakteru. Pristúpime k špecifikácii relácie – pojmu „dôsledok“, ktorý je v logike definovaný dvoma rôznymi prístupmi (pozri ref. xx) :

(1) ***Syntaktický prístup*** – nazývaný metóda dôkazu – je založený na deduktívnom dôkaze formuly ψ z predpokladov (teórie) $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$. Hovoríme, že formula ψ je ***logickým dôsledkom*** predpokladov z Φ, $\Phi \vdash \psi$, vtedy a len vtedy ak ψ deduktívne vyplýva z predpokladov Φ

$$(\Phi \vdash \psi) =_{def} \left(\underbrace{\{\varphi_1, \varphi_2, ..., \varphi_n\}}_{\Phi} \vdash \psi \right) \tag{11.1}$$

Teóriu Φ môžeme rozšíriť o ľubovoľnú formulu α tak, že aj z rozšírenej teórie $\Phi \cup \{\alpha\}$ formula ψ je deduktívnym dôsledkom

$$(\Phi \vdash \psi) \Rightarrow (\Phi \cup \{\alpha\} \vdash \psi) \tag{11.2a}$$

Túto implikáciu môžeme postupne pomocou indukcie rozšíriť do tvaru

$$(\Phi \subseteq \Phi') \Rightarrow ((\Phi \vdash \psi) \Rightarrow (\Phi' \vdash \psi)) \tag{11.2b}$$

Alebo v inom tvare, nech $\boldsymbol{S}_\Phi = \{\varphi; \Phi \vdash \varphi\}$ je množina formúl, ktoré sú logickým dôsledkom formúl z teórie ***T***, potom

$$(\Phi \subseteq \Phi') \Rightarrow (\boldsymbol{S}_\Phi \subseteq \boldsymbol{S}_{\Phi'}) \tag{11.2c}$$

(2) ***Sémantický prístup*** – nazývaný metóda modelu – je založený na predpoklade, že ak každý model teórie Φ je aj modelom formuly ψ, potom hovoríme, že formula ψ je ***sémantickým***

dôsledkom teórie Φ, $\Phi \models \psi$. Predpokladajme platnosť tejto formuly, potom pre každý model teórie Φ je formula ψ pravdivá. Formálne $\left(\forall \tau \in [\![\Phi]\!]\right)\left(val_\tau(\psi)=1\right)$, kde $[\![\Phi]\!]=\{\tau_1,\tau_2,...\}$ je množina obsahujúca všetky modely teórie Φ. Modely rozšírenej teórie $\Phi' = \Phi \cup \{\alpha\}$ (ktorá je konzistentná) tvoria neprázdnu podmnožinu množiny modelov pôvodnej teórie, $[\![\Phi']\!] \subseteq [\![\Phi]\!]$. Potom platí $\left(\forall \tau \in \Phi'\right)\left(val_\tau(\psi)=1\right)$, čiže formula ψ je tautologickým dôsledkom rozšírenej teórie $M_{\Phi'}$, alebo

$$\left(\Phi \models \psi\right) \Rightarrow \left(\Phi \cup \{\alpha\} \models \psi\right) \tag{11.3a}$$

podobne by sme mohli aj pre sémantický dôsledok zostrojiť analógie formúl (11.2b-c), ktoré majú tvar

$$\left(\Phi \subseteq \Phi'\right) \Rightarrow \left(\left(\Phi \models \psi\right) \Rightarrow \left(\Phi' \models \psi\right)\right) \tag{11.3b}$$

$$\left(\Phi \subseteq \Phi'\right) \Rightarrow \left(\tilde{\boldsymbol{S}}_\Phi \subseteq \tilde{\boldsymbol{S}}_{\Phi'}\right) \tag{11.3c}$$

kde $\tilde{\boldsymbol{S}}_T = \{\varphi; \Phi \models \varphi\}$; táto množina obsahuje všetky formuly, pre ktoré každý model z $[\![\Phi]\!]$ je aj modelom formuly φ.

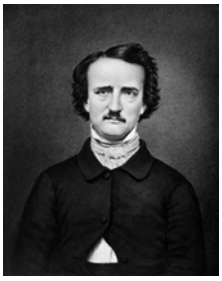
V klasickej logike a v mnohých neklasických logikách platí, že pojmy „logický dôsledok“ a „sémantický dôsledok“ sú ekvivalenmtné ($\left(\Phi \vdash \psi\right) \equiv \left(\Phi \models \psi\right)$ a $\boldsymbol{S}_T = \tilde{\boldsymbol{S}}_T$). Implikácie (11.2a-c) a/alebo (11.3a-c) sú základnou charakteristikou ***monotónnosti*** klasickej logiky ako aj mnohých neklasických logík. Táto terminológia pochádza z matematickej analýzy, kde napríklad funkcia $f(x)$ je monotónne rastúca vtedy a len vtedy ak platí implikácia $\left(x_1 < x_2\right) \Rightarrow \left(f\left(x_1\right) < f\left(x_2\right)\right)$.

Na prelome 60. a 70. rokov minulého storočia bol populárny názor, ***že klasická logika je pre umelú inteligenciu irelevantná***, pretože nie je schopná postihnúť nemonotónne usudzovanie. Pre ďalší rozvoj umelej inteligencie je tento typ usudzovania veľmi dôležitý v situáciách existencie neúplnej informácie, ktorý je schopný formalizovať usudzovanie aj pre prípad neúplnej informácie alebo konfliktných východiskových predpokladov. Jeden z prvých článkov, ktorý sa zaoberal nemonotónnym usudzovaním bola publikácia z r. 1972 od Erika Sandewala [xx]. Jeho prístup bol založený na pravidle

$$A \wedge \neg B \Rightarrow C$$

Podľa ktorého, C môže byť odvodené vtedy, ak A je odvodené a B nemôže byť odvodené. Samozrejme, takto jednoduché formulované pravidlá odvodenie s negatívnou vedľajšou podmienkou neboli všeobecne použiteľné pre rôzne situácie vyskytujúce sa v umelej inteligencii, preto sa hľadali ďalšie efektívnejšie postupy modifikácie klasickej logiky tak, aby bola schopná usudzovania pre neúplnú informáciu a konfliktné predpoklady.

Sherlock Holmes



August Dupin

Obrázok 11.1. Známe literárne postavy 19. storočia využívajúce deduktívnu metódu usudzovania.

Nemonotónna logika je taká logika, pre ktorú relácia dôsledku nevyhovuje podmienke monotónnosti. V tejto logike, rozšírenie teórie môže viesť k potlačeniu vzniku

záveru, ktorý predtým existoval, t. j. množina záverov sa nezväčšuje monotónne so zväčšovaním teórie. Je zaujímavé konštatovať, že klasické (monotónna) logika obvykle preceňuje silu a efektivitu svojho deduktívneho usudzovania. V tejto súvislosti môžeme uviesť známe literárne postavy, akými sú Sherlock Holmes anglického spisovateľa Conana Doyla alebo Auguste Dupin amerického spisovateľa Edgara Allana Poe, ktoré sa zaslúžili o preslávenie deduktívnej metódy usudzovania klasickej výrokovej logiky. Avšak pri hlbšej analýze týchto literárnych postáv zistíme, že ich usudzovanie nie je čisto deduktívne. V implicitnej forme často využívajú prijateľné pomocné predpoklady, ktoré sú získané pomocou nemonotónneho usudzovania, kde sa veľmi často pomocou ad-hoc predpokladov (defaultov) postuluje platnosť rôznych skutočností, ktoré použitím štandardnej monotónnej výrokovej logiky nie sú odvoditeľné. Tieto poznámky o úlohe deduktívneho usudzovania a nutnosti používať súčasne sním nemonotónne usudzovanie môžeme zovšeobecniť do všeobecného tvrdenia, že ľudské usudzovanie pri riešení rôznych situácií (menovite pochopenie komunikácie v prirodzenom jazyku) musí neustále používať nemonotónne usudzovanie.

Na ilustráciu dôležitosti nemonotónneho usudzovania uvedieme klasický príklad interpretácie rozprávky o malom dievčatku nazývanom *Červená Čiapočka* (angl. Little Red Riding Hood), ktorá sa rozpráva malým deťom vo veku 3 - 6 rokov. Týmto jednoduchým príkladom chceme zdôrazniť, že elementárne nemonotónne mechanizmy usudzovania musia byť vlastné už aj malým deťom, aby boli schopné správne pochopiť a interpretovať túto rozprávku.

Obrázok 11.2. Rôzne verzie zobrazenia Červenej Čiapočky v 19. storočí.

Nech malý poslucháč je práve v tom bode rozprávky, kde Vlk sa kontaktuje s Červenou Čiapočkou (ČČ), pozri obr. 11.2. V nasledujúcej časti rozprávky malý poslucháč získa túto informáciu: „*Vlk počul zvuky sekier blízkych drevorubačov a preto sa rozhodol počkať*“. Predpokladajme, že malý poslucháč spontánne (reflexívne) pochopí túto vetu. Avšak jej pochopenie vyžaduje celú postupnosť nemonotónnych inferencií, ktoré môžeme opísať takto (poznámky v zátvorke vyjadrujú základné vedomosti v pozadí, ktoré tvoria východiskové premisy pre inferenčný proces):

(1) Vlk sa priblíži k ČČ (aby niečo zjedol, čo nie je veľmi vzdialené);
(2) ČČ bude kričať (pretože deti kričia, keď sa k nim priblíži zlé zviera);
(3) drevorubači, keď budú počuť krik, budú vedieť, že dieťa je v nebezpečí (pretože krik dieťaťa znamená, že nie je bezpečí);
(4) drevorubači prídu k dieťaťu a pokúsia sa ČČ zachrániť pred Vlkom (ľudia ochraňujú deti, ktoré sú v nebezpečenstve, čo je však veľmi nebezpečné pre Vlka);
(5) preto drevorubači môžu napadnúť vlka (ochrana pre zvieraťom vyžaduje fyzickú silu...);
(6) Vlk sa rozhodol počkať (pretože zviera nechce byť napadnuté ľuďmi).

Z tohto jednoduchého ilustratívneho príkladu vyplýva, že už malé dieťa musí byť schopné uplatňovať nemonotónne usudzovanie, aby bolo schopné pochopiť príbeh rozprávky (alebo vo všeobecnosti, pochopiť prirodzený jazyk dospelých, ktorý už obsahuje implicitne skryté východiskové premisy a teórie).

Zrejme, keby ľudské bytosti boli omniscientné (vševedúce v zmysle Laplaceovej hypotetickej bytosti, ktorá riešením fyzikálnych pohybových rovníc vie si všetko vypočítať), tak by vlastne nemuseli vôbec deduktívne usudzovať , lebo by všetko poznali. Ale keďže nie sú omniscietné (majú len ohraničenú racionalitu [xx], o čom sa každodenne presvedčujeme pri riešení našich bežných problémov; napr. či ísť ráno do práce alebo nie, či použiť pritom auto alebo prostriedok verejnej dopravy, a pod.), musia pri usudzovaní používať ako „pomocné barličky" metódy nemonotóneho usudzovania, pomocou ktorých si „v dobrej viere" dotvárajú svoje vedomosti o svete v ktorom žijú. Pri charakterizovaní ľudských bytostí ako agentov s ohraničenou racionalitou [xx], ktorý využívajú heuristiky pri riešení problémov bežného života, patrí ako neoddeliteľná charakteristika aj používanie metód nemonotónneho usudzovania pri vytváraní potrebných poznatkov v rámci deduktívneho usudzovania.

Príklad 11.1. Predpokladajme, že generická pravda je „*každý vták lieta*" , ktorú prepíšeme do tvaru predikátovej logiky s univerzálnym kvantifikátorom

$$(\forall x)\big(vták(x) \Rightarrow lieta(x)\big) \tag{11.4a}$$

Ďalej predpokladajme, že objekt *Čvirik je vták*

$$vták\big(Čvirik\big) \tag{13:4b}$$

Tieto dva výroky tvoria teóriu

$$\Phi = \Big\{(\forall x)\big(vták(x) \Rightarrow lieta(x)\big), vták\big(Čvirik\big)\Big\} \tag{11.5}$$

Z tejto teórie vyplýva jednoduchý záver, $lieta\big(Čvirik\big)$, alebo

$$\Phi \vdash lieta\big(Čvirik\big) \tag{11.6}$$

Nech teória Φ je rozšírená o dodatočný predpoklad , že *Čviri k nelieta* (napr. preto, že Čvirik je vták kiwi), potom

$$\Phi \cup \Big\{\neg lieta\big(Čvirik\big)\Big\} \nvdash lieta\big(Čvirik\big) \tag{11.7}$$

Spolu s (11.6) tento výsledok môžeme písať ako implikáciu

$$\Big(\Phi \vdash lieta\big(Čvirik\big)\Big) \Rightarrow \Big(\Phi \cup \Big\{\neg lieta\big(Čvirik\big)\Big\} \nvdash lieta\big(Čvirik\big)\Big) \tag{11.8}$$

Tento výsledok je v protiklade s formulou (11.2a), t. j. daný logický systém nie je monotónny.

11.2 Teória uzavretého sveta

Nech $U_0 = \{a,b,c,...\} \subseteq U$ je množina objektov, ktoré majú vlastnosť *P*, potom platí $(\forall \mathrm{X} \in U_0) P(\mathrm{X})$. Predpoklad uzavretého sveta má dve formy: pozitívnu a negatívnu.

(1) *Pozitívna forma uzavretého sveta* (pozri diagram A, obrázok 11.3)

$$(\forall \mathrm{X} \in U_0) P(\mathrm{X}) \Rightarrow (\forall \mathrm{X} \in U - U_0) P(\mathrm{X}) \tag{11.9a}$$

t. j. ak pre každý objekt z množiny U_0 platí výrok *P*(X) , potom tento výrok aj pre každý objekt z *U*, ktorý je mimo množiny U_0. To znamená, že v zjednodušenom tvare pozitívna forma predpokladu uzavretého sveta má tvar

$$(\forall \mathrm{X} \in U_0) P(\mathrm{X}) \Rightarrow (\forall \mathrm{X} \in U) P(\mathrm{X}) \tag{11.9b}$$

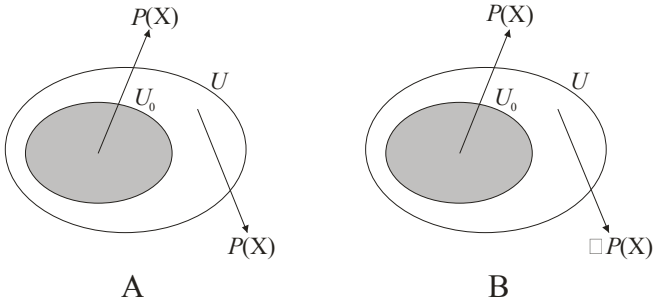
Čo nie je nič iné, ako jednoduché ***zovšeobecnenie*** pôvodného vzťahu $(\forall \mathrm{X} \in U_0) P(\mathrm{X})$. Alternatívna formulácia pozitívnej formy predpokladu uzavretého sveta je, ak nejaká vlastnosť P platí pre každý objekt z daného zoznamu, potom táto vlastnosť platí aj pre objekty, ktoré sa nenachádzajú v tomto zozname. Poznamenajme, že takto formulovanú pozitívnu formu predpokladu uzavretého sveta môžeme jednoducho falzifikovať tak, že nájdeme aspoň jeden objekt $\mathrm{X} \in U - U_0$, pre ktorý táto vlastnosť neplatí

$$(\exists \mathrm{X} \in U - U_0) \neg P(\mathrm{X})$$

(2) *Negatívna forma uzavretého sveta* (pozri diagram B, obrázok 11.3)

$$(\forall \mathrm{X} \in U_0) P(\mathrm{X}) \Rightarrow (\forall \mathrm{X} \in U - U_0) \neg P(\mathrm{X}) \qquad (11.10)$$

Verbálna formulácia negatívnej formy uzavretého sveta je, že ak pre každý objekt zo zoznamu U_0 platí vlastnosť P, potom pre objekty nepatriace do U_0 platí negácia $\neg P$.



Obrázok 11.3. (A) Znázornenie pozitívneho tvaru predpokladu uzavretého sveta – ak pre každý objekt X z U_0 platí vlastnosť P, potom táto vlastnosť platí aj mimo množinu U_0. (B) Znázornenie negatívneho tvaru predpokladu uzavretého sveta – ak pre každý objekt X z U_0 platí vlastnosť P, potom mimo túto množinu platí negácia vlastnosti $\neg P$.

Príklad 11.2. *Pozitívna forma* predpokladu uzavretého sveta môže byť ilustrovaná notoricky známy príkladom o labutiach, ktorý je lokalizovaný na Slovensko. Nech U_0 je množina labutí, ktoré sa vyskytujú pod Skleneným mostom v Pieštanoch a nech predikát P(X) znamená, že labuť X je biela. Potom formula $(\forall \mathrm{X} \in U_0) P(\mathrm{X})$ znamená, že každá labuť žijúca pod Skleneným mostom v Pieštanoch je biela. Použitím pozitívnej formy (optimistického) predpokladu uzavretého sveta (11.10) dostaneme, že aj každá labuť žijúca mimo Skleneného mosta v Piešťanoch je biela. Problém s touto (optimistickou) formou vlastnosti uzavretého sveta je v tom, že formula (11.10) nie je tautológia, čiže môžeme ju používať len ako optimistickú určitú heuristiku, ktorá nám umožní usudzovať pri nedostaku vstupnej informácie aj mimo danej množiny (zoznamu objektov) U_0. *Negatívna forma* predpokladu uzavretého sveta je ilustrovaná jednoduchým príkladom, že ak napr. nie som na zozname U_0 pozvaných hostí prezidentom republiky pri oslávach štátneho sviatku na bratislavskom hrade, ktorý je uverejnený na web stránke prezidentskej kancelárie, potom použitím (pesimistickej) negatívnej formy predpokladu uzavretého sveta je, že nie som pozvaný na danú oslavu.

Usudzovanie založené na predpoklade uzavretého sveta je efektívna heuristika, ktorá nám umožňuje prekonať nedostatok informácie a vyvodiť určité závery, ktoré sa však neskoršie nemusia ukázať ako pravdivé. Predpokladajme, že U_0 je zoznam letov nizkonákladovej leteckej spoločnosti *EastAir* v piatok a hľadáme let do Bruselu z bratislavského letiska, nech zoznam obsahuje tieto lety

U_0 = { Bratislava – Mníchov, Bratislava – Moskva, Mníchov – Brusel, ...}

Ak použijeme heuristiku negatívnej formy uzavretého sveta, potom ak v zozname nenájdeme priamy let Bratislava – Brusel, potom usudzujeme, že takýto spoj z bratislavského letiska

neexistuje. Avšak, ak vykonáme jednoduché usudzovanie nad databázou letov, zistím, že do Bruselu sa môžeme z Bratislavy dostať jedným prestupom cez Mníchov.

Môžeme si položiť otázku, kedy a za akých podmienok budeme používať buď pozitívnu alebo negatívnu formu uzavretého sveta? Tento problém výberu formy heuristiky uzavretého sveta môže byť vyriešený rôznymi spôsobmi. Popíšeme spôsob výberu heuristiky uzavretého sveta, ktorý sa nám zdá konzistentný s metódou uzavretého sveta. Kritérium výberu formy je náš odhad, či množina U_0 má rozšírenie U, $U_0 \subset U$, ktoré obsahuje objekty nenachádzajúce sa v v pôvodnej množine U_0. Ak takéto rozšírenie neexistuje, potom použijeme negatívnu formu uzavretého sveta; v opačnom prípade, ak takáto množina existuje, potom použijeme pozitívnu formu uzavretého sveta. To znamená, že v prvom prípade ($U_0 = U$) použijeme pesimistický prístup k metóde uzavretého sveta, t. j. objekty, ktoré sa nenachádzajú v U_0, $\mathrm{X} \notin U_0$ spĺňajú vlastnosť $\neg P(\mathrm{X})$. V druhom prípade ($U_0 \subset U$, alebo $U - U_0 \neq \varnothing$) použijeme optimistický prístup k metóde uzavretého sveta, t. j. objekty nachádzajúce sa v $U - U_0$ spĺňajú vlastnosť P. Vykonali sme vlastne zovšeobecnenie platnosti vlastnosti P taktiež pre objekty, ktoré sa nenachádzajú v pôvodnej množine U_0, čo môžeme formálne vyjadriť takto

$$(U_0 \subseteq U) \Rightarrow \begin{cases} (U_0 = U) \Rightarrow (\mathrm{X} \notin U_0 \Rightarrow \neg P(\mathrm{X})) \\ (U_0 \subset U) \Rightarrow (\mathrm{X} \notin U_0 \Rightarrow P(\mathrm{X})) \end{cases} \tag{11.11}$$

11.3 Defaultová logika

Defaultová logika patrí medzi tie typy neklasického usudzovania, ktoré sú často využívané v bežnom živote. Ak použijeme predpoklad uzavretého sveta, z neúplnej informácie môžeme pomocou defaultovej logiky odvodiť konzistentný záver, ktorý nie je v konflikte s východiskovými predpokladmi. Táto logika bola vytvorená kanadským informatikom a logikom Raymondom Reiterom [xx]. Kľúčové postavenie v tomto prístupe má pojem „konzistentný“, jeho význam je rovnaký ako v teórii modelov klasickej logiky. Reiterov prínos spočíva v tom, že navrhol formálny systém pre manipuláciu s defaultovými pravidlami, kde rôzne množiny formúl – poznatkov (ktoré Reiter nazýva extenziami) ovplyvňujú konzistentnosť možného záveru.

Príklad 11.3. Vráťme sa k predchádzajúcemu príkladu 11.1, pomocou defaultovho prístupu tento príklad môžeme prepísať do tvaru

$$\delta = \frac{vták(x) : K\, lieta(x)}{lieta(x)} \tag{11.12}$$

x je indivíduová premenná z univerza vtákov $U = \{a,b,...,u\}$. Toto defaultové pravidlo intuitívne znamená, že každý vták obvykle (táto modalita je vyjadrená písmenom K) lieta, čiže lieta aj vták x, ak v našej v aktuálnej databáze poznatkov nemáme opačnú informáciu. Nech Φ je teória (t. j. databáza poznatkov), ktorá obsahuje známe skutočnosti (pozri obr. 11.4)

$$\Phi = \{vták(holub), vtak(kiwi), lieta(holub) \wedge lieta(vrabec) \wedge lieta(vrana)\} \tag{11.13}$$

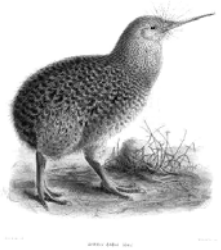
Uvažujme dva prípady hodnoty tejto premennej $x = holub$ a $x = kiwi$:

(1) $x = holub$, potom výrok *vták*(*holub*) je sémantickým dôsledkom teórie Φ, $\Phi \models vták(holub)$. Podmienka konzistentnosti *lieta*(x) pre indivíduovú konštantu $x = holub$

je splnená v teórii Φ (t. j. naše aktuálne poznatky poznajú skutočnosť, že holub lieta), potom môžeme pristúpiť k formulácii záveru: *lieta*(*holub*) .

(2) *vták*(*kiwi*), potom výrok *vták*(*kiwi*) je sémantickým dôsledkom teórie Φ, $\Phi \models vták(kiwi)$. Podmienka konzistentnosti[3] výroku *lieta*(*x*) pre indivíduovú konštantu *x* = *kiwi* nie je splnená v teórii Φ, čiže nemôžeme pristúpiť k formulácii záveru: *lieta*(*kiwi*). V tomto prípade použijeme pre odvodenie záveru ***negatívny princíp uzavretého sveta***, pretože teória Φ neobsahuje položku *lieta*(*kiwi*), potom môžeme predpokladať, že platí $\neg$*lieta*(*kiwi*).

Týmto sme ukázali, že aplikácia defaultového pravidla (11.12) na teóriu Φ špecifikovanej (11.13) produkuje pre *x* = *holub* záver *lieta*(*holub*). Avšak pre *vták*(*kiwi*) defaultové pravidlo (11.12) je neaplikovateľné, podmienka konzistentnosti *lieta*(*kiwi*) s danou teóriou Φ nie je splnená. V tomto prípade, môžeme dospieť k záveru pomocou princípu uzavretého sveta, pretože *lieta*(*kiwi*) nie je obsiahnutá v teórii, potom môžeme predpokladať, že platí $\neg$*lieta*(*kiwi*), t. j. vták kiwi nelieta.



Obrázok 11.4. Nelietajúci vták kiwi (lat. *Apteryx haastii*) žijúci na Novom Zélande

Základné definície a pravidlá

Pravidlo usudzovania δ v defaultovej logike má tvar

$$\delta = \frac{A : KB_1, KB_2, ..., KB_n}{C} \tag{11.14}$$

kde A, B_1, B_2, ..., B_n, C sú formuly výrokovej logiky alebo predikátovej logiky bez voľných indivíduových premenných (každá premenná je viazaná kvantifikátorom). V tejto notácii, A je *predpoklad* (angl. prerequiste), KB_1, KB_2, ..., KB_n sú *podmienky konzistentnosti* (podľa J. Šefránka [xx] *stráže*, angl. *justifications*) a C je *dôsledok* (angl. consequent). Občas sa bude používať nasledujúce označenie

$$pre(\delta) = A, just(\delta) = \{KB_1, KB_2, ..., KB_n\}, cons(\delta) = C$$

Defaultové pravidlo (11.14) sa môže z typografických dôsledkov písať v jednom riadku

$$\delta = A : KB_1, KB_2, ..., KB_n / C$$

Pristúpime k špecifikácii aplikácii defaultového pravidla (11.14) na teóriu $\Phi = \{\varphi_1, \varphi_1, ..., \varphi_p\}$, ktorá obsahuje p uzavretých formúl predikátovej logiky (každá premenná v predikáte je viazaná kvantifikátorom). K tomu, aby pravidlo (11.14) bolo aplikovateľné na teóriu Φ, musia byť splnené[4] tieto tri podmienky:

---

[3] Konjunkciu $lieta(holub) \wedge lieta(vrabec) \wedge lieta(vrana)$ formálne môžeme chápať ako konečný zoznam vtákov, ktorí lietajú. Pretože, vták kiwi sa v tomto zozname nenachádza, tak pomocou negatívneho princípu uzavretého sveta dostaneme záver, že $\neg lieta(kiwi)$, t. j. vták kiwi nelieta.

[4] Nech $\Phi = \{\varphi_1, \varphi_2, ...\}$ je teória a ψ je formula, hovoríme, že táto formula je konzistentná s danou teóriou, čo zapisujeme $\Phi \| \psi$, vtedy a len vtedy, ak existuje taký neprázdny spoločný model $[\![\Phi \cup \{\psi\}]\!] = [\![\Phi]\!] \cap [\![\psi]\!] \neq \varnothing$

- $\Phi \models A$, t. j. formula $pre(\delta) = A$ je tautologickým dôsledkom teórie Φ,
- $(\forall i)(\Phi \parallel B_i)$, každá stráž (podmienka konzistentnosti) $B_i \in just(\delta)$ je konzistentná s teóriou Φ,
- $\Phi \models C$, záver je sémantický dôsledkom teórie Φ.

Potom, ak tieto tri podmienky sú splnené, teória Φ sa ***rozšíri*** o dôsledok $cons(\delta) = C$, $\Phi \leftarrow \Phi \cup \{C\}$

$$\frac{\delta \text{ aplikované na } \boldsymbol{T}}{\Phi \leftarrow \Phi \cup \{cons(\delta)\}} \tag{11.15a}$$

V opačnom prípade, ak z predchádzajúcich troch podmienok nie je splnené aspoň jedna, teória Φ zostáva nezmenená

$$\frac{\delta \text{ aplikované na } \Phi}{\Phi \leftarrow \Phi \cup \varnothing} \tag{11.15b}$$

Niekedy sa podmienka konzistentnosti formuly ψ s teóriou Φ, $\Phi \parallel \psi$, zapisuje alternatívne pomocou symbolu tautologického vyplývania

$$(\Phi \| \psi) =_{def} (\Phi \not\models \neg\psi) \tag{11.16}$$

Dôkaz tejto formuly vykonáme v dvoch krokoch:

(1) Dokážeme implikáciu $(\Phi \| \psi) \Rightarrow (\Phi \not\models \neg\psi)\theta$. Predpokladajme, že platí ľavá strana implikácie $(\Phi \| \psi)$, potom pre každú interpretáciu $\tau \in M_{ext}$ platí $(\forall \tau \in [\![\boldsymbol{T}]\!])(val_\tau(\varphi) = val_\tau(\psi) = 1)$, alebo $val_\tau(\neg\psi) = 0$. Potom však negácia $\neg\psi$ nie je sémantickým dôsledkom teórie Φ, $(\Phi \not\models \neg\psi)$, čo sme mali dokázať.

(2) Dokážeme sporom opačnú implikáciu $(\Phi \not\models \neg\psi) \Rightarrow (\Phi \| \psi)$. Predpokladajme, že platí $\Phi \models \neg\psi$. To znamená, že pre každý model $\tau \in M_{ext}$ a pre každú formulu $\varphi \in \Phi$ platí $val_\tau(\varphi) = val_\tau(\neg\psi) = 1 - val_\tau(\psi)$. Potom platí $val_\tau(\varphi) \neq val_\tau(\psi)$, t. j. formula ψ nie je konzistetná s teóriou Φ, $\Phi \nparallel \psi$, čo však protirečí pravej strane dokazovanej implikácie.

Vyššie uvedené podmienky aplikovateľnosti defaultového pravidla δ na teóriu Φ majú tento ekvivalentný tvar

- $\Phi \models A$,
- $(\forall i)(\Phi \not\models \neg B_i)$,
- $\Phi \models C$.

Definícia 11.1. Defaultová teória Φ_D je určená dvojicou

$$\Phi_D = (D, \Phi) \tag{11.17}$$

kde $D = \{\delta_1, \delta_2, ..., \delta_k\}$ je množina defaultových pravidiel (11.14) a $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_n\}$ je teória – množina formúl výrokovej logiky alebo predikátovej logiky bez voľných premenných.

---

teórie Φ rozšírenej o formulu ψ, t. j. $(\forall \tau \in [\![\Phi \cup \{\psi\}]\!])(\forall \varphi \in \Phi)(val_\tau(\varphi) = val_\tau(\psi) = 1)$, alebo $(\Phi \| \psi) =_{def} ([\![\Phi]\!] \cap [\![\psi]\!] \neq \varnothing)$. V opačnom prípade, ak formula ψ nie je konzistentná s teóriou Φ, potom $(\Phi \nparallel \psi) =_{def} ([\![\Phi]\!] \cap [\![\psi]\!] = \varnothing)$.

Pristúpime k popisu aplikácie defaultovej teórie Φ_D, táto **a**plikácia defaultovej teórie je špecifikovaná pseudokódom, ktorý vytvára z pôvodnej teórie Φ jeho rozšírenú verziu, ktorej konkrétny tvar závisí od poradia aplikácie defaultových pravidiel z množiny D.

Algoritmus 11.1.
for $\delta \in D$ **do**
if $(\Phi \models pre(\delta)) \wedge (\Phi \| just(\delta)) \wedge (\Phi \models cons(\delta))$ **then** $\Phi := \Phi \cup \{cons(\delta)\}$;

Poznamenajme, že tento algoritmus nie je „deterministický“, výsledná rozšírená teória Φ závisí na poradí výberu defaultových pravidiel δ z množiny D, rôzne poradia tohto výberu generujú rôzne rozšírenia.

Príklad 11.4. Budeme študovať slávny príklad defaultovej logiky o americkom prezidentovi Nixonovi. Nech defaultová teória $\Phi_D = (D, \Phi)$ je špecifikovaná takto:

$$D = \{\delta_1, \delta_1\}, \delta_1 = \frac{Nixon\_kvaker : D\ pacifista}{pacifista}, \delta_1 = \frac{Nixon\_konzervativista : D\ antipacifista}{antipacifista}$$

$$T = \{Nixon\_kvaker, Nixon\_konzervativec\}$$

Táto defaultová teória má len dve rozšírenia

$$\underbrace{\frac{\delta_1 \text{ aplikované na } \Phi}{\Phi \cup \{pacifista\}}}_{\Phi_1} \text{ a } \underbrace{\frac{\delta_2 \text{ aplikované na } \Phi}{\Phi \cup \{antipacifista\}}}_{\Phi_2}$$

Ak na prvé (druhé) rozšírenie teórie opakovane použijeme defaultove pravidlo δ_1 (δ_2) výsledná teória sa nezmení. Zaujímavý je prípad, keď na teóriu Φ aplikujeme postupne defaulty δ_1 a δ_2 (alebo δ_2 a δ_1), v tomto prípade druhý default už nemôže byť aplikovaný na modifikovanú teóriu po aplikácii prvého defaultu, v tomto prípade druhá podmienka $(\forall i)(\Phi \| B_i)$ z (11.15b) nie je splnená.

Príklad 11.5. Budeme študovať ilustračný príklad v ktorom bude závisieť na špeciálnej konkretizácii indivíduových premenných defaultového pravidla

$$\delta = \frac{priateľ(x,y) \wedge priateľ(y,z) : K\ priateľ(x,z)}{priateľ(x,z)}$$

Toto pravidlo môžeme v hovorovej reči parafrázovať tak, že „priateľ môjho priateľa je aj mojím priateľom“. Nech teória Φ na ktorú bude toto pravidlo aplikované obsahuje tieto skutočnosti

$$\Phi = \{priateľ(Jana, Eva), priateľ(Eva, Mária), priateľ(Mária, Barbora)\}$$

Krok 1. Položme $x = Jana$, $y = Eva$, $z = Mária$, potom defaultové pravidlo má tvar

$$\delta = \frac{priateľ(Jana, Eva) \wedge priateľ(Eva, Mária) : K\ priateľ(Jana, Mária)}{priateľ(Jana, Mária)}$$

Toto konkrétizované defaultové pravidlo môže byť aplikované k teórii Φ, ktorá bude rozšírená o *priateľ*(*Jana*,*Mária*)

$$\Phi = \{priateľ(Jana, Eva), priateľ(Eva, Mária), priateľ(Mária, Barbora), priateľ(Jana, Mária)\}$$

Krok 2. Položme $x = Jana$, $y = Mária$, $z = Barbora$, potom defaultové pravidlo má tvar

$$\delta = \frac{priateľ\left(Jana, Mária\right) \wedge priateľ\left(Mária, Barbora\right) : K\ priateľ\left(Jana, Barbora\right)}{priateľ\left(Jana, Barbora\right)}$$

V tomto prípade je aktuálna teória ***T*** rozšírená o *priateľ*(*Jana*,*Barbora*). Poznamenaj, že toto rozšírenie teórie bolo umožnené prvým krokom, ktorom vznikla vlastnosť *priateľ*(*Jana*,*Mária*).

Tabuľka 11.1. Dvanásť defaultových pravidiel typu $\delta =: DB/C$, pre $B, C \in \{p, \neg p, q, \neg q\}$

$\delta_1 = \frac{: K\ p}{p}$		$\delta_2 = \frac{: K\ p}{q}$	$\delta_3 = \frac{: K\ p}{\neg q}$
	$\delta_4 = \frac{: K \neg p}{\neg p}$	$\delta_5 = \frac{: K \neg p}{q}$	$\delta_6 = \frac{: K \neg p}{\neg q}$
$\delta_7 = \frac{: K\ q}{p}$	$\delta_8 = \frac{: K\ q}{\neg p}$	$\delta_9 = \frac{: K\ q}{q}$	
$\delta_{10} = \frac{: K \neg q}{p}$	$\delta_{11} = \frac{: K \neg q}{\neg p}$		$\delta_{12} = \frac{: K \neg q}{\neg q}$

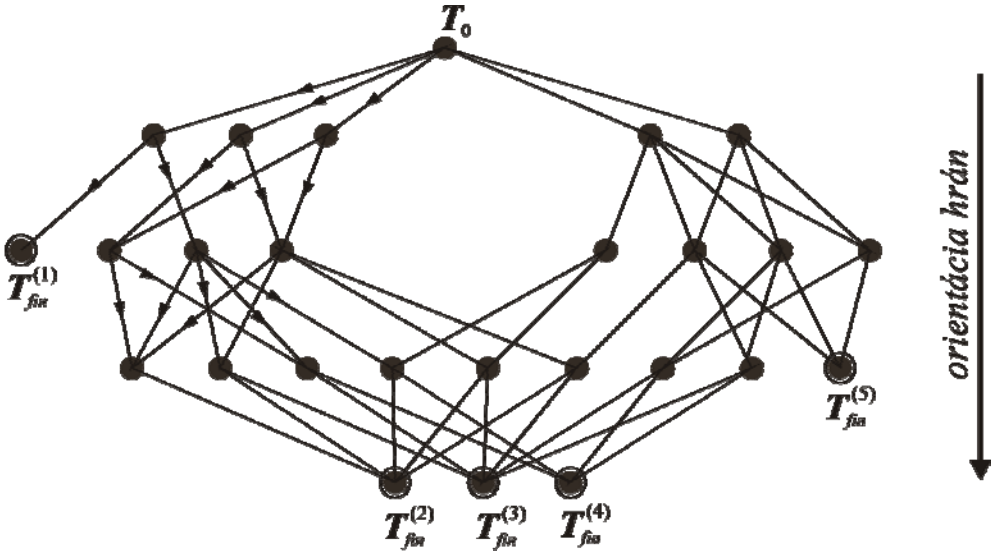
Príklad 11.6 Ukážeme postupnú aplikáciu jednoduchých defaultových pravidiel z tab. 11.1 bez predpokladu na prázdnu teóriu (t. j. $\Phi = \varnothing$). Postupným aplikovaním vyššie uvedených defaultových pravidiel zostrojíme Hasseho diagram znázornený na obr. 11.5. Pre danú defaultovú teóriu $\Phi_D = \left(D, \Phi\right)$ definujme ***postupnosť*** defaultov z D, $\Pi = \left(\delta_1, \delta_2, ..., \delta_k\right)$. Poznamenajme, že ak kardinalita $\left|D\right| = p$, potom existuje p! rôznych postupností Π defaultov. Pre danú postupnosť $\Pi = \left(\delta_1, \delta_2, ...\right)$ zostrojíme postupnosť rozšírení $\Phi_0 = \Phi, \Phi_1, \Phi_2, ..., \Phi_k$ pôvodnej teórie ***T***,

$$\frac{\delta_1 \text{ aplikované na } \Phi_0}{\Phi_1 = \Phi_0 \cup \left\{cons\left(\delta_1\right)\right\}}, \frac{\delta_2 \text{ aplikované na } \Phi_1}{\Phi_2 = \Phi_1 \cup \left\{cons\left(\delta_2\right)\right\}}, ..., \frac{\delta_k \text{ aplikované na } \Phi_{k-1}}{\Phi_k = \Phi_{k-1} \cup \left\{cons\left(\delta_k\right)\right\}} \qquad (11.18)$$

Týmto postupom dostávame sekvenciu teórií, ktoré vyhovujú dvom podmienkam:

(1) $\Phi_0 = \Phi \subseteq \Phi_1 \subseteq \Phi_2 \subseteq ... \subseteq \Phi_k$,

(2) Pre každú teóriu Φ_i (kde $1 \le i \le k$) platí, že ak na ňu aplikujeme defaultové pravidlo δ_j (kde $1 \le j \le i$, t. j. ktoré už raz bolo použité pri konštrukcii Φ_i), potom táto teória sa nemení, je invariantná vzhľadom k aplikácii tohto defaultového pravidla.



Obrázok 11.5. Hesseho diagram tvorby rozšírených teórií aplikáciou rôznych postupností defaultových pravidiel z defaultovej teórie Φ_D. Diagram obsahuje vrcholy (priradené rozšíreným teóriám) a hrany orientované zhora nadol (defaultové pravidla).Mimoriadne postavenie v tomto systéme majú tie rozšírené teórie Φ_{fin} , z ktorých nevychádza žiadna hrana (sú reprezentované bodom v krúžku). Tieto vrcholy sú priradené konečným stabilným teóriám, ktoré sú stabilné voči aplikácii ľubovoľného defaultového pravidla z Φ . Podobný ilustratívny príklad je ukázaný aj na obr. 11.4.

Vyššie uvedená vlastnosť (1) je evidentná. Platnosť (2) vyplýva z toho, že ak už raz bolo pravidlo δ_j aplikované, potom môžu nastať tieto dva prípady: (α) pravidlo δ_j vyhovuje podmienkam aplikovateľnosti na $\Phi_{i\text{-}1}$, potom však daná teória už obsahuje dôsledok cons(δ_j), čiže sa neprejaví v rozšírení teórie; (β) pravidlo δ_j je neaplikovateľné na $\Phi_{i\text{-}1}$, potom teória sa nemení. Z dokázanej druhej vlastnosti vyplýva nasledujúca veta.

Veta 11.1. Pre danú defaultovú teóriu $\Phi_D = (D,\Phi)$ a danú postupnosť defaultov z D, $\Pi = (\delta_1,\delta_2,...,\delta_k)$, rozšírená teória Φ_i je už nemenná (stabilná/invariantná) vzhľadom k aplikácií defaultových pravidiel z postupnosti Π.

Reiter definoval pomocou rozšírených teórií množiny, ktoré nazval ***extenzie***

$$\boldsymbol{E}^{(i)} = Th\left(\Phi_{fin}^{(i)}\right) \tag{11.19}$$

kde $Th(\boldsymbol{X}) = \{\varphi; \boldsymbol{X} \vDash \varphi\}$ je deduktívny uzáver teórie $\boldsymbol{X}$, t. j. deduktívny uzáver obsahuje všetky formuly, ktoré sémanticky vyplývajú z teórie $\boldsymbol{X}$. Z definície vyplýva, že pre danú defaultovú teórie $\Phi_D = (D,\Phi)$ existuje toľko rozličných extenzií, koľko existuje konečných rozšírení $\Phi_{fin}^{(i)}$. Takto definované extenzie môžeme chápať, ako konzistentné databázy poznatkov, ktoré boli odvodené z defaultovej teórie $\Phi_D = (D,\Phi)$. Určitou „výhodou“ extenzie je, že miesto pojmu „konzistentnosť“, ktorý používame v rámci teórií, môžeme používať množinovo-teoretický pojem, kde hovoríme, že daná formula je (nie je) ***elementom*** extenzie:

(1) $(\boldsymbol{E} \vDash \varphi) \equiv (\varphi \in \boldsymbol{E})$

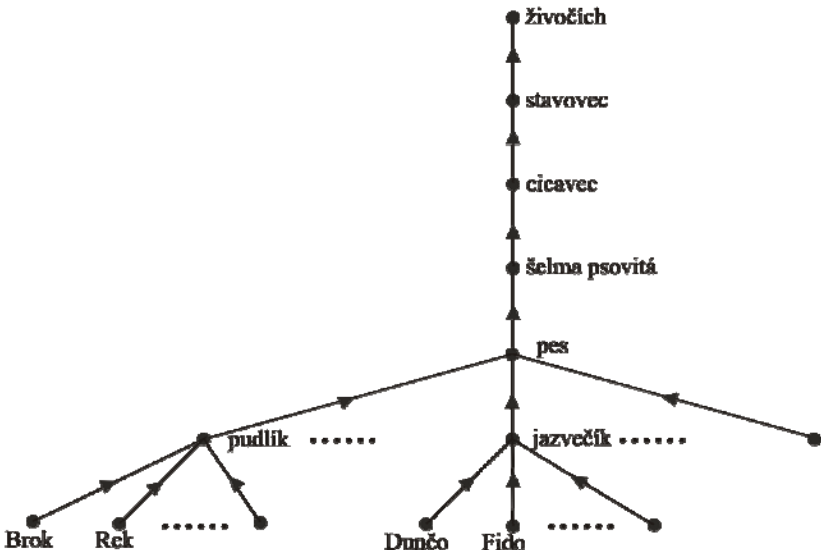
(2) $(\boldsymbol{E} \nvDash \varphi) \equiv (\varphi \notin \boldsymbol{E})$

Potom, napríklad podmienky použitia defaultového pravidla $\delta = A : K\,B_1, K\,B_2, ..., K\,B_n / C$ na extenzii $\boldsymbol{E} = Th(\Phi)$ majú tento tvar:

- predpoklad $pre(\delta)$ vyhovuje miesto podmienke $\Phi \vDash pre(\delta)$ novej podmienke $pre(\delta) \in \boldsymbol{E}$,
- stráže z $just(\delta) = \{B_1, B_2, ..., B_n\}$ vyhovujú miesto podmienke $(\forall i)(\Phi \parallel B_i)$ novej podmienke $(\forall i)(\neg B_i \notin \boldsymbol{E})$
- dôsledok $cons(\delta)$ vyhovuje miesto podmienke $\Phi \parallel con(\delta)$ novej podmienke $pre(\delta) \notin \boldsymbol{E}$.

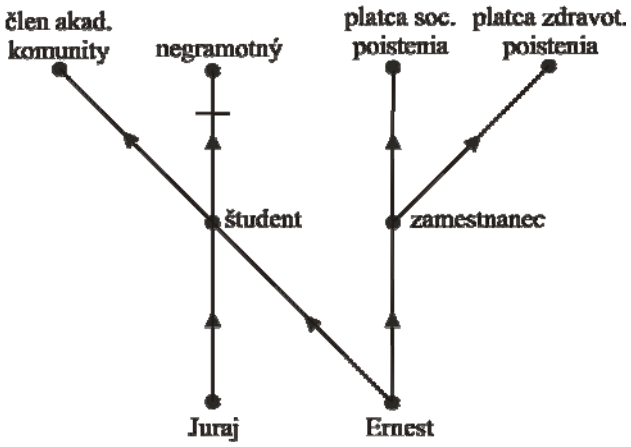
11.4 Usudzovanie s inheritáciou

Spôsob usudzovania s inheritáciou je známy od dôb Aristotela a často bol využívaný v stredovekej filozofii pri štúdiu sylogizmov. V umelej inteligencii tento prístup bol využitý v polovici 80. rokov minulého storočia americkým informatikom Daveom Touretzkym [xx], ktorý so svojimi kolegami vypracoval pomocou usudzovania s inheritáciou systém nemonotónneho usudzovania. Poznamenajme, že princíp inheritácie bol jeden z prvých efektívnych princípov objavených umelou inteligenciou pri formalizácii prirodzeného usudzovania.

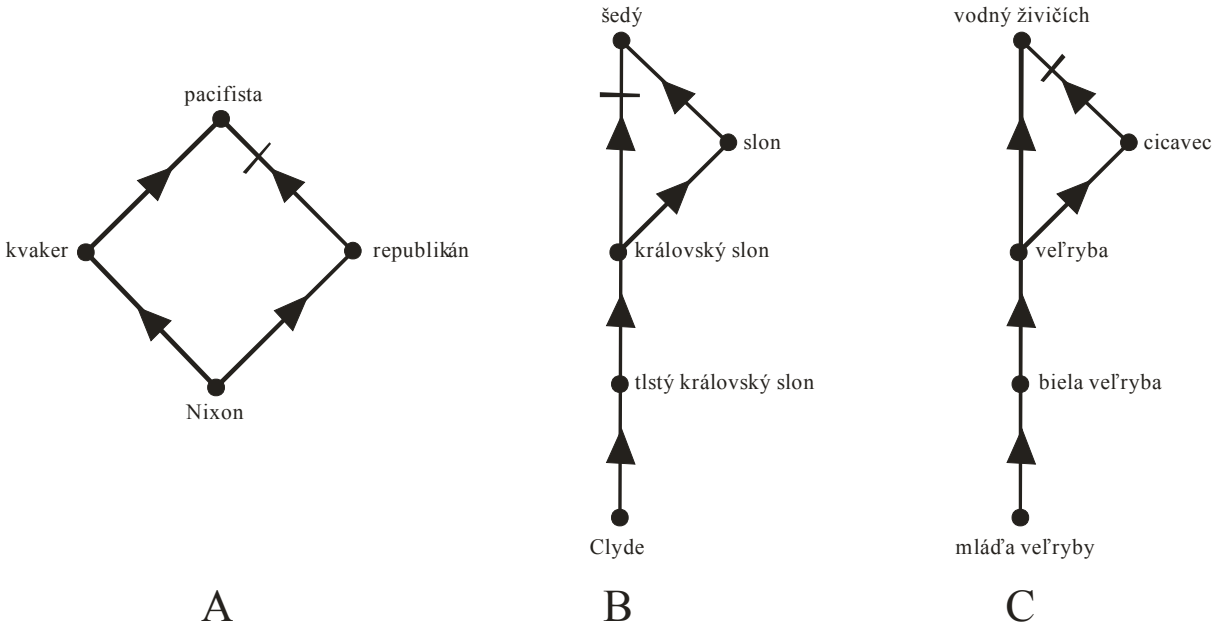


Obrázok 11.6. Inheritačná charakteristika objektov - psov s menom Brok, Rek,…, Dunčo, Fido, …, pričom psi Brok a Rek sú pudlíci, Dunčo a fido sú jazvečíci, pudlíci a jezvečíci sú psi, psi patria k šelmám psovitým, ktoré sú cicavce, tieto sú stavovce a konečne tieto sú živočíchy. Tento ***inheritačný strom*** živočíchov môže byť jednoducho rozšírený o ďalšie živočíchy, akými sú napr. vtáky (čo je veľká potrieda stavovovcov), alebo cicavce rozšírené o šelmy mačkovité, kopytníky, a pod. V rámci tohto ***inheritačného stromu*** môžeme ľahko odvodzovať pomocou ***princípu tranzitívnosti***, tak napr. Brok je nielen pudlík, ale je psom, šelmou psovitou, cicavcom, atď. Môžeme konštatovať, že naše vedomosti o okolitom svete môžeme názorne reprezentovať pomocou stromovej diagramamtickej štruktúry nazývanej ***inheritačný strom***, pričom nad daným inheritačným stromom môžeme vyvodzovať pomocou ***princípu tranzitívnosti*** ďalšie poznatky, ktoré nie sú v ňom explicitne obsiahnuté. Toto ododzovanie nových poznatkov môžeme vyjadriť pomocou existencie orintovanej cesty idúcej zdola-nahor, napr. medzi Rekom a cicavcom existuje orientovaná cesta, t. j. ´Rek je cicavec´. Medzi objektami pudlík a jazvečík neexistuje orientovaná cesta, t. j. medzi touto dvojicou neexistuje relácia inheritácie, ktorá by bola vyvoditeľná aplikáciou princípu tranzitívnosti na elementárne relácie z inheritačného stromu.

Svet v ktorom žijeme je hierarchicky organizovaný, v ktorom nejaká vec c z tohto sveta je „podvecou“ nejakej inej všeobecnejšej veci, atď. (pozri obr. 11.7).



Obrázok 11.7. Inheritačný strom, ktorý špecifikuje Juraja ako študenta a Ernesta ako študenta a zamestnaného. Tento strom obsahuje aj nový typ hrany, ktorá spája objekty ´študent´ a ´negramotný´ , t. j. študent nie je negramotný.



Obrázok 11.8. Tri inheritačné diagramy, ktoré budú podrobne diskutované v texte.

Formálna štruktúra „inheritačný strom“ z obr. 11.9 reprezentuje špecifický výsek našich vedomosti o okolitom svet, pričom nad týmto stromom môžeme vyvodzovať pomocou ***princípu tranzitívnosti*** ďalšie poznatky, ktoré nie sú v ňom explicitne obsiahnuté. Na obr. 11.10 je znázornený ´ďalší typ inheritačného stromu, ktorú obsahuje v jednom prípade aj tzv. negatívne vlastnosti, ktoré sa interpretujú pomocou negatívnej hrany.

Inheritačný strom môžeme definovať ako ***acyklický orientovaný graf*** $G=(V,E)$, kde ***množina vrcholov*** $V=\{v\}$ je priradené atomickým objektom nejakej databázy poznatkov, ktorá je reprezentovaná inheritačným strom. ***Množina hrán*** $E=\{e\}$ obsahuje usporiadané dvojice vrcholov z množiny V, výraz $e=(v,v')$ je orientovaná hrana, ktorá začína vo vrchole v a končí vo vrchole v', v grafe G je reprezentovaná orientovanou hranou (so šípkou) z v do

v´. Budeme rozlišovať dva typy hrán, pozitívne a negatívne; pozitívne (nagatívne) sú reprezentované jednoduchou hranou so šípkou (hranou so šípkou a s malou kolmou čiarou). Poznamenajme, že tieto negatívne hrany vchádzajú len do takých vrcholov, ktoré neobsahujú vychádzajúcu hranu. Na základe tejto podmienky, v inheritačnom strome existujú dva druhy orientovaných ciest: (1) kladné cesty – obsahujúce len kladné hrany, a (2) záporné cesty – končiace zápornou hranou.

Diagram A na obr. 11.11 poskytuje dve riešenia, ktoré sa navzájom vylučujú:

(1) ***Nixon je pacifista***,
(2) ***Nixon nie je pacifista***.

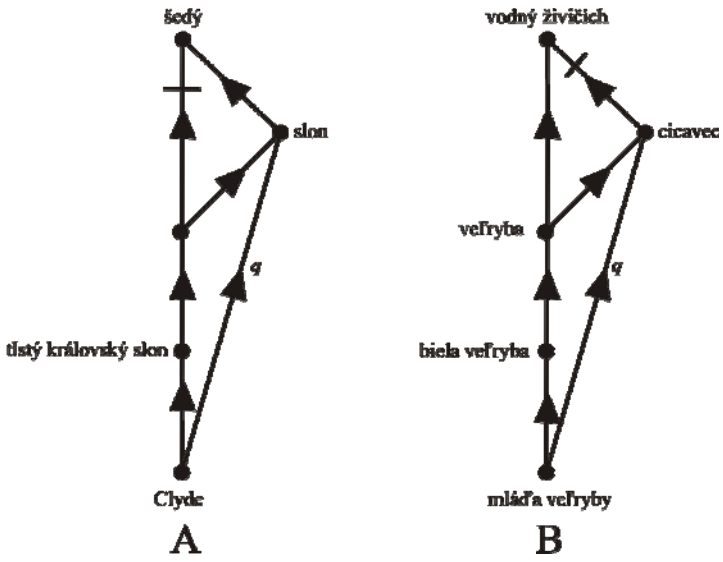
Obe riešenia sú rovnako dôležité, nemáme prostriedky na ich odlíšenie. Diagram B na obr. 11.10 poskytuje taktiež dve riešenia, ktoré sú odlišiteľné dĺžkou cesty z vrcholu Clyde do vrcholu šedý:

(1) Clyde je šedivý,
(2) ***Clyde nie je šedivý***.

Podobne ako v predchádzajúcom príklade založenom na diagrame A, tieto dve riešenia sú vo vzájomnej kontradikcii, t. j. stojíme pred problémom, ktorú vybrať ako viac pravdepodobne platnú pre naše logické úvahy nad poznatkami špecifikovanými inheritačným grafom B z obr. 11.10. Tento problém navrhol riešiť Dave Touretzky, jeden z hlavných tvorcov metódy usudzovania s inheritáciou pomocou ***princípu inheritácie***, podľa ktorého v našom „prirodzenom“ usudzovaní ***špecifická informácia prevráži menej špecifickú informáciu.*** Toto pravidlo môžeme dať do jednoduchej súvislosti so známym pravidlom Occamovej[5] britvy, podľa ktorého dôvody na vysvetlenie nejakej hypotézy majú také jednoduché, ako je to len možné, alebo o nič zložitejšie, ako je to potrebné (*Entia non sunt multiplicanda praeter necessitatem*). To znamená, že v našom usudzovaní použijeme argument špecifikovaný kratšou cestou (obsahuje len 3 hrany) pred argumentom, ktorý je špecifikovaný dlhšou cestou (obsahuje 4 hrany). Diagram C na obr. 11.10 poskytuje dve riešenia

(1) ***Mláďa veľryby je vodný živočích***,
(2) Mláďa veľryby nie je vodný živočích,

Ktoré sú taktiež dobre odlišiteľné dĺžkou cesty na danom inheritačnom grafe. Druhé riešenie je generované kratšou cestou dĺžky 3.



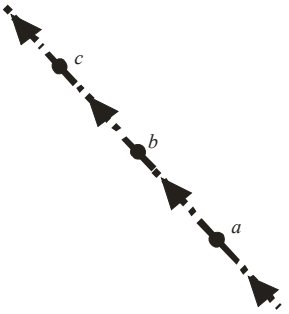
Obrázok 11.9. Modifikované inheritačné stromy z obr. 11.9 o redundantnú hranu označenú q.

[5] Wiliam z Occamu (1385–1349), anglický stredoveký filozof, ktorý sa zaslúžil o rozvoj logiky svojim dielom *Summa Logicae*, kde už skúmal trojhodnotovu logiku a negáciu konjunkcie alebo disjunkcie (t. j. analógie DeMorganových vzťahov). Je známy hlavne ako tvorca heuristiky Occamovej britvy, podľa ktorej dôvody na vysvetlenie nejakého javu majú byť tak jednoduché ako je to len možné.

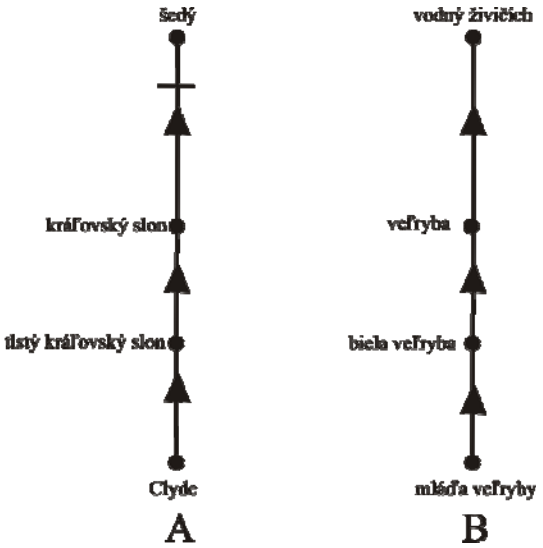
Žiaľ princíp minimálnej cesty nie je univerzálne použiteľný v prípade, že inheritačný strom obsahuje tzv. redundantné hrany (napr. hrana označená *q* na obr. 11.10). Táto nová hrana umožňuje tvorbu nových minimálnych ciest vychádzajúcich z vrcholu Clyde a končiacich vo vrchole Šedý, t. j. získané závery o farbe Clyde sú diametrálne odlišné od pôvodných záverov založených na diagrame B, obr. 11.10. Hranu *q* nazývame redundantnú preto, lebo explicitne reprezentuje tvrdenie, že Clyde je slon, čo z pôvodného inheritačného stromu vyplývalo implicitne pomocou principu tranzitivity. Preto na rieše nie tohto problému bol navrhnutý americkou informatičkou Lynn Stein [xx] sofistikovaný postup, v ktorom bol princíp minimálnej cesty nahradený princípom ***minimálnej inferenčnej vzdialenosti***: inferenčná vzdialenosť medzi vrcholmi *a* a *b* je menšia ako inferenčná vzdialenosti medzi vrcholmi a a c vtedy a len vtedy, ak existuje taká cesta z a do c cez vrchol b (pozri obr. 11.13)

$$d(a,b) < d(a,c) =_{def} \text{cesta z } a \text{ do } c \text{ cez vrchol } b \qquad (11.20)$$

kde symbol *d*(*x*,*y*) je inferenčná vzdialenosť medzi vrcholmi *x* a *y*.



Obrázok 11.10. Jednoduchá cesta z a do c cez vrchol b. Z tejto skutočnosti priamo vyplýva, že inferenčná vzdialenosť vrchola *a* k vrcholu *b* je menšia ako k vrchulu *c*.



Obrázok 11. 11. Redukované inheritačné stormy z obr. 11.10 o redundnatné hrany. Riešenie, ktoré vyplý z týchto stromov je pre diagram A ´Clyde nie je šedivý´ a pre diagram b ´mláďa verľryby je vodný živočích´.

Aplikujme princíp minimálnej inferenčnej vzdialenosti k inheritačným stromom z obr. 11.11, na ktoré bol princíp minimálnej vzdialenosti neaplikovateľný v dôsledku existencie redundantnej hrany *q*. Študujme diagram A zaoberajúci sa farbou slona. Naše úvahy sú uskutočnené na cestách inicializovaných vrcholom ´Clyde´ a ukončené vrcholom ´šedý´. Musíme sa rozhodnúť ktorú cestu budeme uvažovať, či tú, ktorá obsahuje vrchol ´tlstý kráľovský slon´ alebo vrchol ´slon´. Pomocou princípu minimálnej inferenčnej vzdialenosti

ukážeme, že najbližší vrchol k východiskovému vrcholu ´Clyde´ je vrchol ´tlstý kráľovský slon´. Analogickým spôsobom ukážeme aj pre diagram B na obr. 11.12, že najbližší vrchol k východiskovému vrcholu ´mláďa veľryby´ je vrchol ´biela veľryba´. Týmto jednoduchým spôsobom sa nám podarilo vylúčiť z našich úvah redundantné hrany, ktoré viedli k chybným neintuitívnym výsledkom. Rezultujúce inheritačné grafy sú znázornené na obr. 11.12 spolu aj s riešením.

Podrobná teória inferenčnej vzdialenosti je pomerne zložitá, obsahuje množstvo detailov a výnimiek. Pre účely tejto kapitoly domnievame sa však, že špecifikácia inferenčnej vzdialenosti dostatočne vystihuje jej základné princípy a je použiteľná pre interpretáciu jednoduchších inferenčných stromov.

11.5 Autoepistemická logika

Autoepistemická logika bola vytvorená americkým informatikom Robertom C. Mooreom [xx], ktorý navrhol špeciálny typ modálnej logiky, ktorá elegantne rieši mnohé vnútorné problémy nemonotónnej logiky. Navrhnutá logika môže byť alternatívne interpretovaná ako spôsob ideálneho racionálneho agenta vytvoriť množinu poznatkov na základe daných počiatočných poznatkov. V našom teoretickom prístupe budeme predpokladať, že jazyk modálnej logiky L_K je vytvorený nad danou množinou atomických výrokových premenných $P = \{p, q, ..., p', q', ...\}$ a množinou logických spojok $C = \{\neg, \wedge, \vee, \Rightarrow, \equiv, K\}$, ktorá vznikla jednoduchým rozšírením množiny logických spojok výrokovej logiky o modálnu spojku K, ktorú pre danú formulu $(K\varphi) \in L_K$ čítame *´agent vie, že platí* φ´. Pre zjednodušenie našich úvah zavedieme duálnu spojku $\tilde{K}$ k modálnej spojke K, $\tilde{K}\varphi =_{def} \neg K \neg \varphi$, čo čítame ako *´agent nevie, že neplatí* φ´, čo môžeme s trochou fantázie parafrázovať ako *´agent tuší, že platí* φ´

Príklad 11.7. Ukážeme aplikáciu modálnej logiky k jednoduchej simulácii nemonotónneho usudzovania. Majme pravidlo, že ak Dr. Johanides je univerzitný profesor, potom obvykle aj učí (výnimku z tohto pravidla môže byť situácia, že Dr. Johanides je dekanom fakulty). K tomu, aby sme vyjadrili toto pravidlo pomocou modálne logiky, zavedieme atomické výrokové premenné

p = ´Dr. Johanides je univerzitný profesor´
q = ´Dr. Johanides má v tomto akad. roku 2 prednášky´

Pravidlo vyjadríme pomocou formuly modálnej logiky

$$\left(p \wedge \tilde{K}q\right) \Rightarrow q \tag{11.21}$$

ktoré čítame *´ak Dr. Johanides je univerzitný profesor´* a *´tuším, že má v tomto akad. roku 2 prednášky´*, potom *´Dr. Johanides má v tomto akad. roku 2 prednášky´*. Pomocou formuly (11.21) racionálny agent zistí, že Dr. Johanides má prednášky (t. j. učí). Toto jednoduché pravidlo môžeme vyjadriť pomocou defaultu

$$\delta = \frac{p : \tilde{K}q}{q} \tag{11.22}$$

V defaultovej logike, ak položíme teóriu $\Phi = \{p\}$, potom defaultová teória $\left(\Phi = \{p\}, p : \tilde{K}q / q\right)$ poskytuje záver q.

V kapitole 8 bolo ukázané, že existuje mnoho modálnych logík, ktoré zásluhou Kripkeho relačnej sémantiky dostupných svetov sú jednoznačne klasifikované. V našich ďalších úvahách budeme používať modálnu logiku typu *S*5, pre ktorú relácia *R* je

ekvivalenciou (reflexívnou, symetrickou a tranzitívnou, co môže byť, napríklad, zabezpečené pomocou úplného orientovaného grafu).

Pre daný typ modálnej logiky *S5* môžeme definovať dva druhy relácií vyplývania: (1) reláciu logického vyplývania $\vdash$ a (2) reláciu sémantického vyplývania $\models$. Pretože modálna logika *S5* je korektná a úplná, tieto dve relácie sú ekvivalentné. Potom pre triedu Kripkeho modelov $\mathcal{C}=\{M,M',...\}$ a pre danú modálnu teóriu $\Phi=\{\varphi_1,\varphi_2,...,\varphi_3\}\subseteq L_K$ platí

$$\left(\psi\in Cn_{S5}(\Phi)\right)=_{def}\left(\left(\forall M\in\mathcal{C}\right)\left(M\models\Phi\right)\right)\wedge\left(\left(\forall M\in\mathcal{C}\right)\left(M\models\psi\right)\right)$$

kde Cn_{S5} je operátor logického vyplývania[6]. Táto formula vyjadruje skutočnosť, že pre každý Kripkeho model $M\in\mathcal{C}$ formula ψ je logickým dôsledkom teórie Φ vtedy a len vtedy, ak teória Φ a formula ψ majú rovnaký model *M*. Potom musí platiť, že formula $\varphi_1\wedge\varphi_2\wedge...\wedge\varphi_n\Rightarrow\psi$ je tautológia

$$\models\varphi_1\wedge\varphi_2\wedge...\wedge\varphi_n\Rightarrow\psi$$

V epistemickej logike sa rieši problém, akými podmienkami musí byť charakterizovaný racionálny agent, alebo presnejšie, akú modálnu logiku používa na tvorbu svojej ***poznatkovej databázy*** (***belief set***) *E* z teórie Φ. Podľa filozofa a logika Stalnakera [xx] modálna teória Φ sa stáva poznatkovou databázou vtedy *E*, ako vyhovuje týmto trom podmienkam:

(1) Teória *E* musí byť uzavretá vzhľadom k výrokovému operátoru dôsledku *Cn*[7]

$$Cn(E)\subseteq E \qquad (11.23a),$$

t. j. *E* obsahuje tak predpoklady ako aj ich logické dôsledky zostrojené pomocou pravidiel modus ponens a necesitácie.

(2) Poznatková databáza musí byť uzavretá vzhľadom k pozitívnej intropspekcie

$$\varphi\in E\Rightarrow K\varphi\in E \qquad (11.23b)$$

(3) Poznatková databáza musí byť uzavretá vzhľadom k negatívnej intropspekcie

$$\varphi\notin E\Rightarrow\neg K\varphi\in E \qquad (11.23c)$$

Stalnakerove postuláty sú v súčasnosti všeobecne akceptované ako definičné podmienky pre poznatkovú databázu. To znamená, že modálna teória *E* sa nazýva ***poznatková databáza***, ak vyhovuje podmienkam (11.23a-c). Poznamenajme, že poznatkové databázy majú veľmi bohatú a sofistikovanú teóriu [xx]. Moore definoval [xx] autoepistemickú logiku pomocou rovnosti

(11.24)

Ľavá strana (11.24) obsahuje taktiež poznatkovú databázu *E*, ktorá je týmto vzťahom vlastne definovaná, preto je niekedy podmienke nazývaná ***podmienka pevného bodu***. To však znamená, že konštrukcia poznatkovej databázy je možná len postupným „iteratívnym“ spôsobom, ktorý je ukončený vtedy, keď množina *E* je už stabilná – nemí sa.

Poznatková databáza má elegantnú sémantickú interpretáciu pomocou štruktúr možných svetov. Nech $\mathcal{T}=\{\tau,\tau',\tau'',...\}$ je množina obsahujúca všetky možne pravdivostné interpretácie výrokových premenných (poznamenajme, že ak máme *n* výrokových

[6] Operátor Cn_S je špecifikovaný dôkazom v danej modálnej logike *S*, t. j. vychádza z predpokladov, výrokových tautológií a modálnych axióm, ktoré sú použité pre špecifikáciu logiky *S*. Ako pravidlo usudzovania je použité pravidlo modus ponens a pravidlo necesitácie $\varphi/K\varphi$.

[7] Aplikácia výrokovému operátoru dôsledku *Cn* k formule modálnej logiky znamená, že formule $K\varphi$ sú chápané ako výrokové premenné.

premenných, potom $\mathcal{T} = \{0,1\}^a$ a kardinala $|\mathcal{T}| = 2^n$). Podmnožiny $Q \subseteq \mathcal{T}$ sú nazývané ***štruktúry možných svetov***, Kripkeho ***množina možných svetov*** je potom $W = \{Q, Q', Q'', ...\}$. Z každej štruktúry Q sa môžeme dostať do ľubovoľne inej štruktúryčiže môžeme povedať, že nad množinou W je definovaná univerzálna relácia dostupnosti, alebo $\Gamma(Q) = Q$.

Definícia 11.2. Skutočnosť, že formula φ je pravdivá vo svete $Q \in W$ pre danú interpretáciu $\tau \in Q$ vyjadríme takto $(Q,\tau) \vDash \varphi$, pričom rôzne formuly φ platí

(1) $(Q,\tau) \vDash p$, vtt ak $\tau(p) = 1$, t. j. pre danú interpretáciu je výroková premenný p pravdivá.

(2) $(Q,\tau) \vDash (\varphi \wedge \psi)$, vtt $(Q,\tau) \vDash \varphi$ a $(Q,\tau) \vDash \psi$,

(3) $(Q,\tau) \vDash (\varphi \vee \psi)$, vtt $(Q,\tau) \vDash \varphi$ alebo $(Q,\tau) \vDash \psi$,

(4) $(Q,\tau) \vDash (\varphi \Rightarrow \psi)$, vtt ak $(Q,\tau) \nvDash \varphi$ alebo $(Q,\tau) \vDash \psi$,

(5) $(Q,\tau) \vDash (\neg\varphi)$, vtt $(Q,\tau) \nvDash \varphi$,

(6) $(Q,\tau) \vDash (K\varphi)$, vtt $(\forall \tau' \in Q)((Q,\tau') \vDash \varphi)$,

(7) $(Q,\tau) \vDash (\tilde{K}\varphi)$, vtt $(\exists \tau' \in Q)((Q,\tau') \vDash \varphi)$.

Z tejto definície pravdivostných hodnôt formúl $K\varphi$ z jazyka L_K vyplýva, že táto pravdivostná hodnota nezávisí od pôvodnej interpretácie $\tau \in Q$, ale závisí len od danej štruktúry možných svetov Q, t. j. môžeme zaviesť zjednodušenie

$$((Q,\tau) \vDash K\varphi) =_{def} (Q \vDash K\varphi) \tag{11.25}$$

Teória vzhľadom k danej štruktúre možných svetov Q je definovaná ako množina modálnych formúl $K\varphi$, ktoré sú pravdivé v Q,

$$Th(Q) = \{\varphi; Q \vDash K\varphi\} \tag{11.26}$$

Veta 11.2.
Množina modálnych formúl $E \subseteq L_K$ je poznatková databáza vtedy a len vtedy, ak existuje taká štruktúra možného svetu $Q \subseteq \mathcal{T}$, že platí $E = Th(Q)$.

Cvičenie

Cvičenie 11.1. V tomto príklade máme tri súbory poznatkov
Juraj:
(1a) Juraj je námorník.
(1b) Námorník obvykle pije pivo.
(1c) Farár obvykle nepije pivo.
(1d) Pijak piva má obvykle nadváhu.
(1e) Námorník obvykle nemá nadváhu.

Poly:
(2a) Poly je austrálsky živočích.
(2b) Vtákopysk je obvykle cicavec.
(2c) Austrálske živočíchy obvykle nie sú cicavce.

(2d) Cicavce sa obvykle nerozmnožujú vajcami.
(2e) Vtákopysk obvykle sa rozmnožuje vajcami.

Fero:
(3a) Fero je republikán.
(3b) Kvakery sú obvykle pacifisti.
(3c) Republikáni obvykle nie sú pacifisti.
(3d) Republikáni sú obvykle militanti.
(3e) Pacifisti obvykle sú verejne aktívni.

Zostrojte:
(a) Pre každý súbor poznatkov zostrojte inheritačný strom.
(b) Ktoré sú ľahkovážné (kredulosné) extenzie stromu?
(c) Ktoré z nich sú preferované extenziu?
(d) Zostroj taký záver, ktorý je prijateľný pre ľahkovážneho riešiteľa ale je neprijateľný pre skeptického riešiteľa.
(e) Existujú také závery, ktoré sú pre skeptického riešiteľa a ideálneho skeptického riešiteľa neprijateľné?

Literatúra

[1] Brewka, G.: *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge University Press, Cambridge, UK, 1991.

[2] Gigerenzer, G., Selten, R., eds.: *The Bounded rationality*. *The Adaptive Tool Box.* MIT Press, Cambridge, MA, 2002.

[3] Ginsberg, M. L.: *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann Publishers, Los Altos, CA, 1987.

[4] Kvasnička, V., Pospíchal, J.: *Matematická logika*. Vydavateľstvo STU, Bratislava 2006.

[5] Marek, W., Truszczynski, M.: *Nonmonotonic Logic. Context_Dependent Reasoning.* Springer Verlag, Berlin, 1993.

[6] Reiter, R.: A Logic for Default Reasoning. *Artificial Intelligence* **13**(1980), 132.

[7] Rybár, J. (editor): *Sociálna inteligencia***.** Vydavateľstvo Európa, Bratislava, 2010.

[8] Sandewall, E.: An Approach to the Frame Problem, and its Implementation. In *Machine Intelligence* 7 (Meltzer, B., Michie, D., eds.). Edinburgh University Press, Edinburgh (UK), 1972, pp. 195-204.

[9] Šefránek, J.: *Inteligencia ako výpočet*. IRIS, Bratislava, 2000.

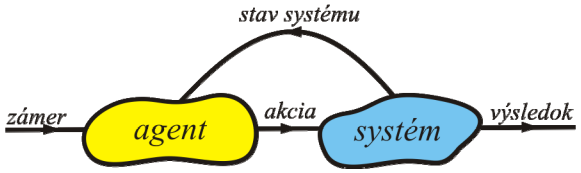
12. prednáška

Riadenia zložitých systémov a výpočtové modely okolia

12.1 Úvodné poznámky

Problém riadenia zložitých systémov (napr. banka, systém zásobovania veľkomesta, vysoká pec na výrobu železa, fabrika na výrobu automobílov, atomová elektráreň, neurónová sieť obsahujúca mnohotisíc neurónov,...), ktorých modely dobre nepoznáme (ak poznáme, tak len vo veľmi nedokonalej forme), sú tradičným problémom umelej inteligencie a kognitívnej vedy. Metódy modernej umelej inteligencie sú schopné poskytnúť nový pohľad na tento štandardný problém informatiky, poskytujú nielen jeho novú interpretáciu, ale umožňujú aj jeho implementáciu pomocou neurónových sietí. Z dôvodov neutrality našich záverov budeme používať termín „agent“ na označenie inteligentného jedinca s kognitívnym orgánom (mozgom), ktorého cieľom je riadiť zložitý systém. Kognitívny orgán agenta (informatický termín pre ľudský mozog) je schopný vykonávať zložité kognitívne aktivity, akými sú učenie, plánovanie, vyhodnocovanie, riešenie problémov, vnímanie okolia, a pod. Tieto schopnosti kognitívneho orgánu sa využívajú pri tvorbe vnútorných modelov riadeného systému, ktoré tvoria ústrednú zložku využívanú pri riadení zložitých systémov alebo pri rozhodovaní. V tejto kapitole budeme prezentovať ***model distančného učenia*** (angl. ***distal learning***), ktorý bol navrhnutý počiatkom 90. rokov americkými informatikmi a kognitívnimi vedcami, Michaleom Jordanom a Davidom Rumelhartom [xx], ktorí sa zaslúžili o rozvoj moderných konkcionistických metód a ich aplikácie v umelej inteligencii a kognitívnej vede. Distančné učenie slúži ako určitá alternatíva k učeniu s odmenou a trestom (pozri kap. 5), poskytuje riešenia, ktoré sú pre psychológov učenia akceptovateľnejšie ako vysledky poskytované učením s odmenou a trestom a taktiež môžu slúžiť ako priamy návod na implementáciu systémov riadenia zložitých systémov.

Zložitý systém je špecifikovaný vstupom, výstupom a vnútorným stavom. Agent prostredníctvom svojich orgánov vnímania rozpoznáva aktuálny stav systému a taktiež môže priamo ovplyvňovať vstup systému. *Cieľom riadenia systému (t. j. úloha pre agenta) je pre dané vnútorné stavy systému meniť vstup systému tak, aby boli dosiahnuté požadované hodnoty výstupu* (pozri obr. 12.1).

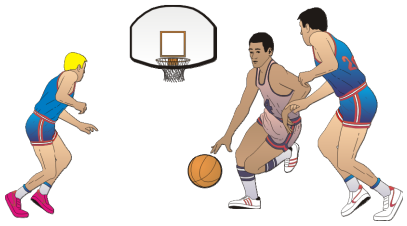


Obrázok 12.1 Komplex agent – systém. Agent transformuje vstupný zámer a stav systému na akciu, ktorá tvorí vstup pre systém. Odozvou systému na akciu je zmena jeho stavu a výsledok. Zámer a akcia sú pre agenta *blízke premenné*, ktoré je schopný bezprostredne kontrolovať, zatiaľ čo stav systému a výsledok sú *vzdialené premenné*, ktoré môže ovplyvňovať len nepriamo - prostredníctvom akcie.

Aká je úloha učiteľa (supervisora) v tomto všeobecnom prístupe k učeniu agenta interagujúceho so systémom? V štandardnom prístupe k učeniu agenta učiteľ vyhodnocuje výstupy agenta, ktoré je agent schopný ovládať svojím výstupom (blízkou premennou). V ďalej rozoberanom všeobecnom prístupe učiteľ vyhodnocuje až výsledok systému, ktorý agent môže ovládať len sprostredkovane pomocou akcií. Týmto dochádza k podstatnému posunu významu učenia. V štandardnom prístupe učiteľ hodnotí bezprostredný výstup agenta, ktorý je schopný agent ovládať svojou intenciou. V tomto všeobecnom prístupe učiteľ hodnotí vzdialený výstup systému, ktorý agent môže ovládať len sprostredkovane pomocou svojich akcií.

Niekoľko „psychologických“ poznámok o priebehu učenia sa v komplexe agent – systém: Proces učenia možno rozdeliť na dve etapy. V *prvej etape* si agent pomocou svojho kognitívneho orgánu vytvára tzv. ***dopredný model*** zložitého systému, na základe svojich akcií sa učí predvídať pre daný stav výstupy. V *druhej etape* agent adaptuje svoj kognitívny orgán (vzhľadom na dopredný model), aby mohol transformovať predvídané výstupy pomocou svojich akcií, vytvára si vo svojom kognitívnom orgáne tzv. ***inverzný model***, pomocou ktorého agent dokáže predvídať akcie na základe jeho intencie a požadovaných výstupov systému, ktoré sú simulované doprdným modelom.

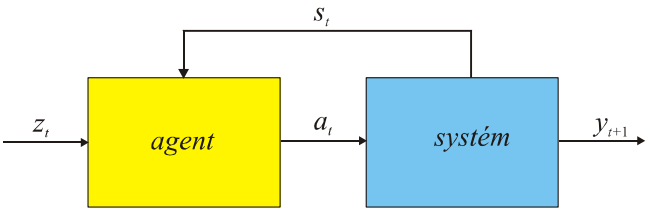
Naznačený dvojetapový model učenia, keď sa agent učí riadiť komplexný systém pomocou svojich akcií a zámerov, má dôležité postavenie v informatike a v kognitívnej vede. Môže byť užitočný na pochopenie reflexívnych aktivít človeka pri riadení takých komplexných zariadení, akým je napr. auto alebo lietadlo. Neobyčajne veľká rýchlosť, s akou človek reaguje pri riadení komplexného zariadenia, podporuje hypotézu, že človek sa učí riadiť systém v dvoch etapách. V prvej etape sa vytvára v mozgu dopredný model systému, ktorý dokáže s dobrou presnosťou predpovedať výstupy zariadenia na základe aktivít človeka. Tento dopredný model má veľký význam pre vznik reflexívneho riadenia celého zariadenia na základe zámeru a aktuálneho výstupu a stavu systému. V druhej etape učenia sa už človek nemusí neustále obracať na systém ako taký, jeho odozvy sú simulované dopredným modelom, predvída potrebné akcie ako odozvu na požadovaný výstup pomocou inverzného modelu. To umožňuje vznik reflexívneho (veľmi rýchleho) „modulu“ v mozgu na riadenie zložitých zariadení.



Obrázok 12.2 Jednoduchý ilustračný príklad dvojice agent – systém. Agent je reprezentovaný hráčom s loptou, zatiaľ čo systém je reprezentovaný ihriskom a polohou ostatných hráčov. Intenciou hráča – agenta je hodiť loptu do koša, výstupom systému sú audiovizuálne efekty sprevádzajúce hodenie lopty do koša.

Ako jednoduchý príklad takéhoto študovaného komplexu uvedieme hráča basketbalu (agent), ktorý na basketbalovom ihrisku obsadenom inými hráčmi (systém) má zámer hodiť loptu do koša (pozri obr. 12.2). Systém je charakterizovaný stavom, ktorý je reprezentovaný polohami iných hráčov, pričom výstupom zo systému – výsledokom, sú audiovizuálne efekty sprevádzajúce hodenie lopty do koša. Agent – hráč je schopný rekognoskovať stav systému, t.j. postavenia ostatných hráčov, túto skutočnosť svojím kognitívnym orgánom pretransformuje na príkazy pre svoje motorické centrá, pomocou ktorých vykoná sekvenciu elementárnych pohybov. To môže byť chápané ako akcia agenta, ktorá mení stav systému. Vyššie uvedené dvoj-etapové učenie sa agenta v komplexe so systémom znamená, že agent – hráč si v prvej etape vytvára model systému (basketbalovej hry), v ktorom je schopný viac-menej presne predpovedať dôsledky svojho pohybu – akcie na stav systému, ako naň

zareagujú ostatní hráči. V druhej etape už agent neustále neočakáva reakciu systému na svoju akciu, ale ***pomocou svojho vnútorného dopredného modelu adaptuje svoj kognitívny orgán (vzniká inverzný model) na stavu systému a svoj zámer transformuje na akciu***. Vo všeobecnosti môžeme očakávať dobré výsledky aj vtedy, keď vnútorný dopredný model systému nie je celkom presný, adaptáciou kognitívneho orgánu pri tvorbe inverzného modelu sa môžu odstrániť možné nedostatky tohto modelu.



Obrázok 12.3 Dvojica zložená z dvoch častí: z agenta, ktorý transformuje zámer a stav systému na akciu, a zo systému, ktorý transformuje akciu na výsledok (pre daný stav).

Predpokladajme, že zámer (z), akcia (a), stav systému (s) a výsledok (y) sú špecifikované pre každý časový krok t. Dvojica agent a systém je schematicky reprezentovaná na obr. 12.3. Tento obrázok je v podstate totožný s obrázkom 12.1, v tejto novej schematickej reprezentácii sú explicitne uvedené jednotlivé premenné komplexu a ich časové indexy (predpokladáme diskrétny čas). Ako už bolo poznamenané, cieľom agenta je riadiť výsledok y (výstup systému) pomocou svojich akcií a a pre daný stav systému s. Pretože výsledná premenná y je pre agenta „vzdialená“, nemôže ju bezprostredne ovplyvňovať, učenie sa dvojice agent – systém sa označuje ako ***učenie s učiteľom***. Tréningová množina pre tento prípad má tvar

$$A_{train} = \left\{ z_t \,/\, \boldsymbol{y}^*_{t+1} \,;\, t = 1,2,...,t_{max} \right\} \tag{12.1}$$

kde $\boldsymbol{y}^*_{t+1}$ je požadovaný výstup pre zámer z_t. V tomto prípade si môžeme predstaviť, že riadime auto a v každom časovom okamihu máme predpísané, akú rýchlosť máme dosiahnuť. Táto množina spolu so špecifikovaným počiatočným stavom systému s_1 tvorí základnú informáciu pre adaptácie s učiteľom (ktorý poskytuje požadované výstupy $\boldsymbol{y}^*$).

Poznamenajme, že ***konštrukcia tejto tréningovej množiny je pomerne problematická a reprezentuje netriviálny problém***. Existencia tréningovej množiny implikuje existenciu učiteľa, ktorý je schopný napr. rozfázovať pohyby hráča basketbalu na elementárne zložky a každú ohodnotiť. To je takmer nepredstaviteľné, že by tréner basketbalu učil hrať basketbal týmto spôsobom. Preto volí inú možnosť: vhodnými tréningovými metódami vedie hráčov k tomu, aby si v rámci svojho kognitívneho aparátu vygenerovali model hry, pomocou ktorého budú schopní vnútornou inferenciou vyvodiť dôsledky svojej akcie na systém, čo je podstatne rýchlejšie a hlavne efektívnejšie než to, keď sa neustále obracať svojimi percepčnými orgánmi na externý systém.

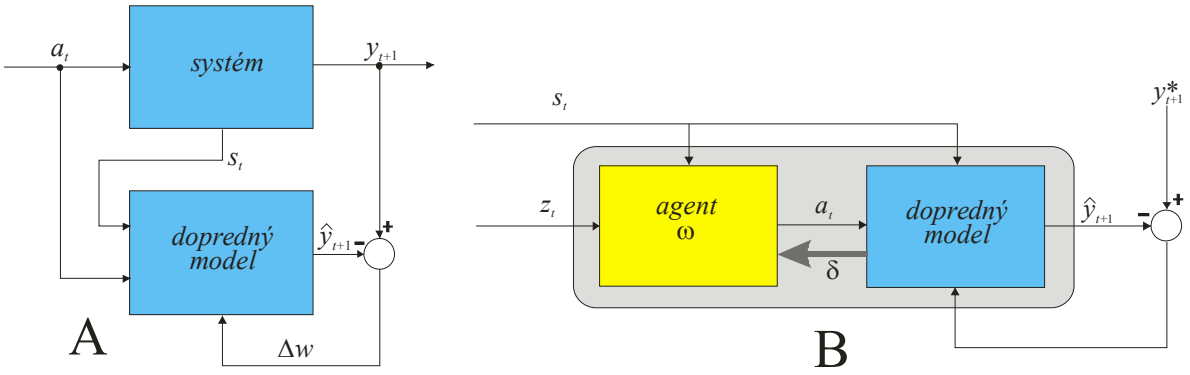
12.2 Dopredný model systému

Dopredný model (angl. forward model) systému je chápaný ako parametrická funkcia[1] G, ktorá zobrazuje akciu a_t a stav systému s_t do výsledku $\hat{y}_{t+1}$ (pozri obr. 12.4, diagram A).

$$\hat{y}_{t+1} = G\left(a_t, s_t; w\right) \tag{12.2}$$

[1] Parametrická funkcia $f(x,w)$ obsahuje okrem nezávisle premennej x aj parameter w, pomocou ktorého môžeme modifikovať vlastnosti-priebeh funkcie f.

Učenie sa tohto komplexu v podstate môže prebiehať štandardnou metódou s učiteľom, pričom požadované výstupné aktivity poskytuje priamo systém (t. j. pri tvorbe dopredného modelu nie je potrebná tréningová množína, požadované výstupné hodnoty poskytuje priamo systém). Treba však poznamenať, že konštrukcia dopredného modelu musí predchádzať učenie sa komplexu, pričom výsledný dopredný model nemusí byť perfektný. V teórii riadenia je zvykom dopredný model nazývať ***identifikácia systému***.



Obrázok 12.4 (A) Dopredný model systému je chápaný ako parametrická funkcia, ktorá zobrazuje akciu a stav do výsledku. (B) Modifikovaná sústava agent – systém (pozri obr. 12.3), kde systém je nahradený sústavou agent - dopredný model. V kognitívnom orgáne agenta sa vytvára ***model dopredného modelu*** tak, že dopredný model je zafixovaný, adaptuje sa len kognitívny orgán agenta (ktorého štruktúra je špecifikovaná parametrom **ω**) u ktorého sa postuluje určitá flexibilita tak, aby bol schopný adaptácie. Rozdiel medzi vypočítaným a požadovaným výsledkom sa nechá, prebehnúť" cez dopredný model, vyprodukovaný vektor „chýb" δ je vstupom do učenia sa agenta.

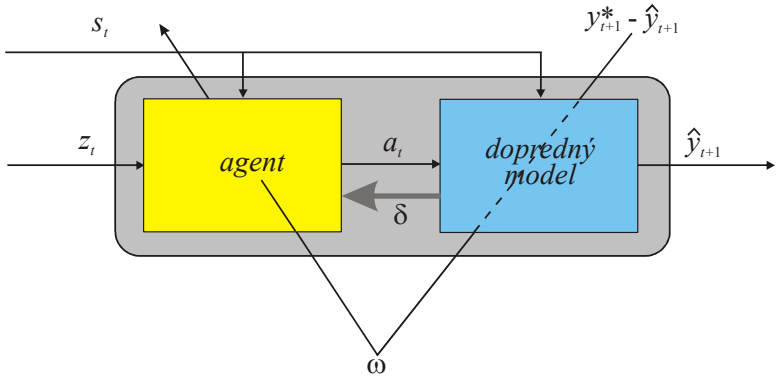
Obráťme teraz našu pozornosť na učenie sa komplexu zloženého z agenta (jeho kognitívneho orgánu určeného parametrami **ω**) a z dopredného modelu systému (pozri diamgram B, obr. 12.4). Hovoríme, že učením tohto komplexu vzniká v kognitívnom orgáne agenta ***model dopredného modelu***. Vstupom do komplexu je dvojica tvorená zámerom z a stavom s systému, výstupom je výsledok $\hat{y}$, účelová funkcia má v tomto prípade tento jednoduchý tvar

$$E(\boldsymbol{\omega}) = \frac{1}{2}\left(\hat{y} - y^*\right)^2 \tag{12.3}$$

Pomocou derivácie tejto účelovej funkcie podľa parametrov kognitívneho orgánu zisťujeme, ako veľmi sa má zmeniť akcia agenta. Táto zmena musí byť tiež úmerná „chybe" **δ** vygenerovanej dopredným modelom v bloku kognitívneho orgánu agenta (pozri obr. 12.4, diagram B). Schéma B na obr. 12.4 sa nazýva podľa Jordana a Rumelharta [xx] ***distančné učenie*** (angl. distal learning), jeho cieľom je mapovať zámery na požadované výsledky pomocou dát z tréningovej množiny. To znamená, že y^*_{t+1} sú požadované hodnoty pre zložený systém s distančným učením s dohľadom. Ľubovolný algoritmus učenia s učiteľom môže byť použitý pre adaptačný proces, avšak postulujeme, že (vytvorený v predchádzajúcom podkroku a reprezentovaný parametrickou funkciou G špecifikovanou formulou (12.2) , pozri diagram A, obr. 12.4) ***dopredný model je zafixovaný*** pri adaptácii zloženého systému. Touto podmienkou sa redukuje adaptačný proces na hľadanie optimálneho zobrazenia reprezentujúceho kognitívny orgán agenta. Štandardný prístup k učeniu zloženého systému je taký, že sa systematicky menia parametre **ω** kognitívneho orgánu agenta tak, aby sa minimalizovala chyba $y^*_{t+1} - \hat{y}_{t+1}$, pozri obr. 12.5. Tento prístup k adaptácii zloženého systému požaduje konštrukciu tréningovej množiny

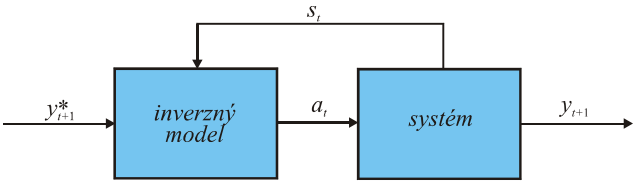
$$A_{train} = \left\{ \left(z_t, s_t\right) / y^*_{t+1} \right\} \tag{12.4}$$

čo môže reprezentovať netriviálny problém.



Obrázok 12.5. Znázornenie procesu ***dištančného učenia*** komplexu agent – dopredný model, ktorým sa vytvára ***model dopredného modelu*** v kognitívnom orgáne agenta. Obrazne môžeme povedať, že pomocou tohto komplexu „mapujeme“ dopredný model systému do kognitívneho orgánu agenta. Použitím tohto „modelu modelu“ agent vie rýchlo rozhodnúť (nemusí sa obracať na aktuálny systém) aký je požadovaný výstup $\hat{y}_{t+1}$ ako odozva na vstup z_t a s_t. Lomená čiara reprezentuje „ladenie“ (angl. tunning) kognitívneho orgánu agenta (realizovaného zmenou parametru **ω**) .

Na záver tejto podkapitoly niekoľko poznámok ku konštrukcii tréningovej množiny (12.4). Vrátime sa k úvodnému textu z podkapitoly 12.1, kde sme použili „športový“ ilustračný príklad hráča basketbalu. Pre tohto hráča sme načrtli dvojetapové učenia sa tajto hry. V prvej etape hráč sa za pomocou trénera – učiteľa učí základným operáciam tejto hry, ako napríklad hodiť loptu do koša, driblovať s loptou, a pod. Úlohou učiteľa – trénera je v tejto prvej etape vytvárať pre hráča „vhodné tréningové množiny“, pomocou ktorých sa učí tieto elementárne situácie hry, ktoré v druhej etape integruje do štandardnej performácie basketbalu. V tejto druhej etape už hráč reflexívne rieši elementárne situácie, ktoré sa „naučil“ v prvej etape pomocou trénera a ktoré ma zafixované vo svojom kognitívnom orgáne pomocou modelu dopredného modelu. Druhú etapu budeme špecifikovať v nasledujúcej kapitole pomocou inverzného modelu systému.



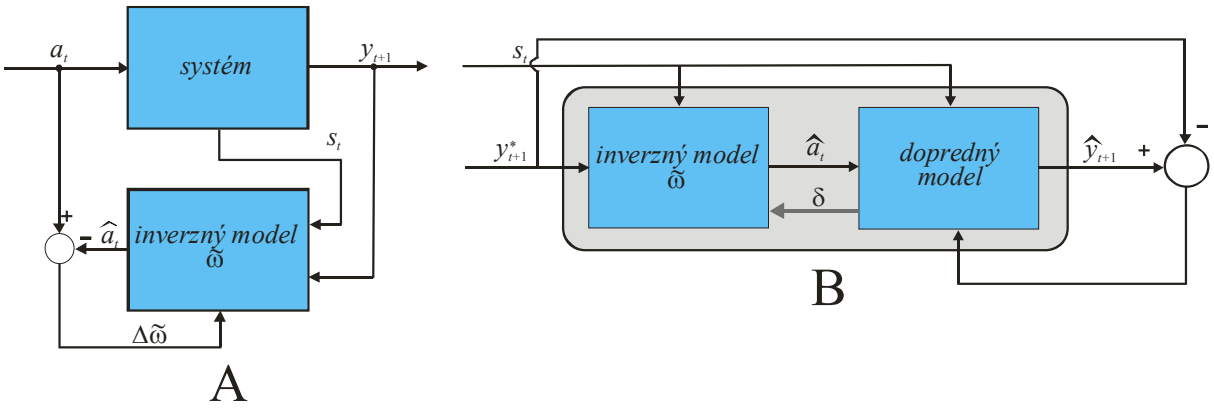
Obrázok 12.6 Sériové zapojenie inverzného modelu spolu so systémom, požadovaný výstup tohto zapojenia je identické zobrazenie, t. j. $y^*_{t+1} = y_{t+1}$.

12.3 Inverzný model systému

Inverzný model systému je taký produkuje akciu a_t ako odozvu na aktuálny stav systému s_t a požadovaný výsledok y^*_{t+1}. Tento model je definovaný tak, že poskytuje identické zobrazenie, ak je sériovo vložený pred systém (okolie), pozri obr. 12.6. V tejto zostave môže byť inverzný model chápaný ako nejaké riadiace zariadenie, ktoré kontroluje systémový blok; požadované výstupy zo systému tvoria vstupy do riadiaceho zariadenia špecifikovaného paramterom $\tilde{\omega}$, ktorého výstup je aktivita pre systém meniaca jeho stav

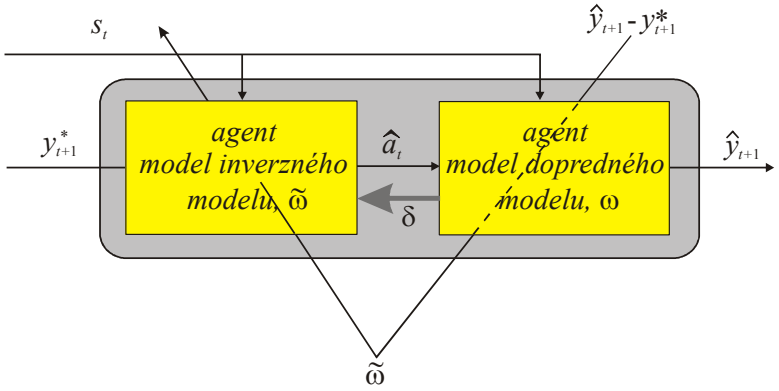
$$\hat{a}_t = F\left(y_{t+1}, s_t; \tilde{\omega}\right) \tag{12.5}$$

V tejto súvislosti si môžeme položiť otázku o jednoznačnosti určenia inverzného systému pomocou daného systému. Pretože systém ako taký netvorí vo všeobecnosti jedno-jednoznačné zobrazenie aktivity do stavu a výsledku, potom ***inverzné zobrazenie k nemu nemôže byť vo všeobecnosti určené jednoznačne***. Ďalší problém s inverzným modelom je, že jeho určenie pomocou diagramu A na obr. 12.7 nie je cieľovo orientované. To znamená, že nemusíme nájsť takú akciu, ktorá produkuje požadovaný výstup systému.



Obrázok 12.7 (A) Učenie sa inverzného modelu, pričom vstupy do modelu sú tvorené výstupmi systému. (B) Komplex tvorený z inverzného modelu a z dopredného modelu. V tomto komplexe dopredný model nemusí byť perfektný, učenie sa komplexu (t. j. inverzného modelu) prebieha tak, že parametre dopredného modelu sú zafixované, adaptácia prebieha len v rámci parametru $\tilde{\omega}$ inverzného modelu. Rozdiel medzi vypočítaným a požadovaným výsledkom sa nechá „prebehnúť" cez dopredný model, vyprodukovaný vektor „chýb" δ je vstupom do učenia sa inverzného modelu.

Učenie sa komplexu vytvoreného z inverzného a z dopredného modelu (pozri obr. 12.7, diagram B) sa môže vykonať podobne ako učenie sa komplexu obsahujúceho agenta a dopredný model v predchádzajúcej podkapitole (pozri obr. 12.4., diagram B).



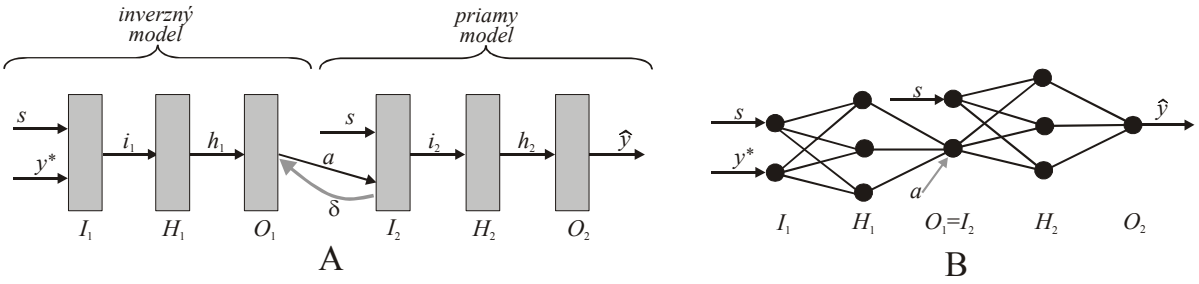
Obrázok 12.8. Tvorba ***modelu inverzného modelu*** v rámci kognitívneho orgánu agentu, štruktúra tohto modelu je popísaná parametrom $\tilde{\omega}$. Táto schéma je izomorfná so schémou znázornenej na obr. 12.7, diagram B, s tým rozdielom, že separátny dopredný model je teraz nahradený modelom dopredného modelu, ktorý je „umiestnený" v kognitívnom orgáne agenta Podobne, aj vytváraný model inverzného modelu je „umiestnený" v kognitívnom orgáne agenta.rejme, neurónové siete, ktoré realizujú tieto modely modelov sú separátné, pričom môžu mať spoločné prepojene výstupu jedne siete na vstup druhej siete.

Podobne, ako je umiestnený model dopredného modelu v kognitívnom orgáne agenta, aj inverzný model je umiestnený ako model inverzného modelu v kognitívnom orgáne agenta. Tieto dva modely nie sú izolované v kognitívnom orgáne, ale výstup modelu inverzného modelu je napojený na vstup modelu dopredného modelu. Tajto prepojený komplex vytvára

model modelu sveta, umožnuje agentovi rýchlo reagovať na zmenené situýciu externého sveta – systému.

12.4 Konekcionistický model

Komplex zobrazený na obr. 12.7, diagram B, je interpretovaný tak, že každý blok modelu (inverzný a dopredný) je reprezentovaný parametrickým zobrazením $F(\boldsymbol{\omega})$, resp. $G(\boldsymbol{w})$. V rámci konekcionistického prístupu tieto modely môžu byť implementované pomocou trojvrstvových neurónových sietí s dopredným šírením signálu (pozri obr. 12.9) Poznamenajme, že tieto jednoduché neurónové siete už majú vlastnosť univerzálneho aproximátora, pozri ref. [xx], učenie sa takejto neurónovej siete je štandardnou záležitosťou teórie neurónových sietí [xx].

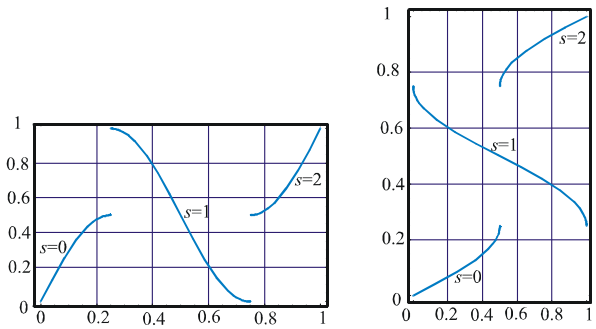


Obrázok 12.9 (A) Neurónová sieť zostrojená zo 6 vrstiev neurónov, pričom prvé tri vrstvy (I_1, H_1 a O_1) reprezentujú inverzný model, ďalšie tri vrstvy (I_2, H_2 a O_2) reprezentujú dopredný model (porovnaj s obr. 12.7, diagram B). Spoje medzi neurónmi zo susedných vrstiev I_1-H_1 , H_1-O_1, resp. I_2-H_2, H_2-O_2 sú realizované všetkými možnými spôsobmi. (B) Implementácia pomocou neurónovej siete obsahujúcej tri skryté vrstvy neurónov, na rozdiel od diagramu A sú vrstvy O_1 a I_2 stotožnené. V prvej etape učenia sa tohto komplexu sa adaptuje len pravá časť (zložená z vrstiev I_2, H_2 a 0_2) , pričom ako vstup do tejto podsiete je dvojica (a,s), požadovaný výstup je y^* . V druhej etape budeme predpokladať, že váhové koeficienty z dopredného modelu sú zafixované, adaptujú sa len váhové koeficienty z inverzného modelu.

Neurónová sieť obsahuje dve časti, prvá sieť je tvorená vrstvami I_1, H_1 a O_1, táto sieť zodpovedá inverznému modelu, druhá sieť, priradená doprednému modelu, je tvorená vrstvami I_2, H_2 a O_2. Učenie sa tohto komplexu dvoch neurónových sietí prebieha v dvoch etapách. V prvej etape učíme len samotnú neurónovú sieť zodpovedajúcu doprednému modelu. Po skončení učenia sa je adaptovaná neurónová sieť vložená do komplexu a potom nasleduje druhá etapa učenia sa. V tejto etape predpokladáme, že váhové koeficienty časti priradenej doprednému modelu sú konštantné, t. j. menia sa len váhové koeficienty z inverzného modelu.

Príklad 12.1. Majme „systémovú“ funkciu určenú vzťahom (pozri obr. 12.10, diagram A):

$$y = f_s(a) = \begin{cases} \frac{1}{2}\sin(2\pi a) & (s=0,\ a<1/4) \\ \frac{1}{2}\left(1+\sin(2\pi a)\right) & (s=1,\ 1/4<a<3/4) \\ 1+\frac{1}{2}\sin(2\pi a) & (s=2,\ a>3/4) \end{cases} \tag{12.5}$$



Obrázok 12.10. (A) Priebeh funkcie $f_s(a)$ pre tri rôzne hodnoty stavového parametra s. (B) Priebeh inverznej funkcie $f_s^{-1}(y)$ taktiež pre tri rôzne hodnoty stavového parametra s. Poznamenajme, že funkcie mimo svojich oborov definície majú nulovú funkčnú hodnotu.

Pre jednoduchosť budeme postulovať, že funkcie mimo svoj definičný obor sú nulové; tento predpoklad podstatne zjednoduší adaptáciu príslušnej časti neurónovej siete z jej implementácie znázornenej na obr. 12.10, diagram B.

Funkcia $f_s(a)$ má tri rôzne inverzné funkcie v závislosti od funkčnej hodnoty a stavu (pozri obr. 12.10, diagram B):

$$a = f_s^{-1}(y) = \begin{cases} \dfrac{1}{2\pi} arcsin(2y) & (s=0,\ y \le 1/2) \\ \dfrac{1}{2} - \dfrac{1}{2\pi} arcsin(2y-1) & (s=1,\ 0 \le y \le 1) \\ 1 + \dfrac{1}{2\pi} arcsin(1-2y) & (s=2,\ y > 1/2) \end{cases} \tag{12.6}$$

Naším cieľom bude zostrojiť tak dopredný model systému (ktorý transformuje akciu a a stav s na výsledok y), ako aj inverzný model (transformujúci stav s a výsledok y na akciu a). Ako už bolo poznamenané vyššie v tejto podkapitole, tento cieľ sa realizuje dvojetapovo. V prvej etape sa použije pravá časť komplexu na obr. 12.9, diagram B, na konštrukciu dopredného modelu obsahujúceho vrstvy I_2, H_2 a O_2, pričom tréningová množina je realizovaná „stochasticky“ takto:

$$s = random(3) \tag{12.7a}$$

$$a = \begin{cases} \dfrac{1}{4} random[0,1] & (s=0) \\ \dfrac{1}{4} + \dfrac{1}{2} random[0,1] & (s=1) \\ \dfrac{3}{4} + \dfrac{1}{4} random[0,1] & (s=2) \end{cases} \tag{12.7b}$$

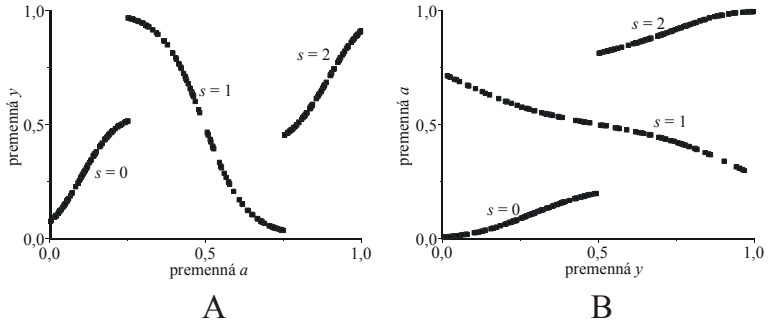
kde $random[0,1]$ je náhodný generátor čísel z intervalu $[0,1]$ s rovnomernou distribúciou, podobne $random(n)$ je náhodný generátor celých čísel z množiny $\{0,1,2,...,n-1\}$ s rovnomernou distribúciou. Požadované hodnoty y sa spočítajú pomocou vyššie špecifikovanej funkcie $f_s(a)$. Adaptačný proces neurónovej siete je po určitom počte krokov ukončený, potom je komplex v rámci druhej etapy použitý na konštrukciu inverzného modelu pomocou ľavej časti komplexu (I_1, H_1 a O_1), pričom váhové koeficienty dopredného modelu (pravej časti) sú nemenné. V tejto druhej etape adaptačný proces vyžaduje znalosť „chybového signálu“ δ,

ktorý je vytvorený štandardnou metódou spätného šírenia signálu cez pravú doprednú časť. Tréningová množina v druhej etape je taktiež realizovaná stochasticky:

$$s = random(3) \tag{12.8a}$$

$$y^* = \begin{cases} \frac{1}{2} random[0,1] & (s=0) \\ random[0,1] & (s=1) \\ \frac{1}{2} + \frac{1}{2} random[0,1] & (s=2) \end{cases} \tag{12.8b}$$

(v tejto druhej etape už nepotrebujeme poznať požadované hodnoty akcií *a*).Výsledkom adaptačného procesu v druhej etape bude neurónová sieť priradená inverznému modelu (vzhľadom na daný dopredný model z prvej etapy), ktorá realizuje viac-menej presne inverznú funkciu $f_s^{-1}(y)$, ktorú teraz nepotrebujeme poznať explicitne, bude sa používať len na vyhodnotenie presnosti vyprodukovanej inverznej neurónovej siete.



Obrázok 12.11. (A) Zobrazenie realizované neurónovou sieťou adaptovanou v prvej etape. Sieť reprezentuje dopredný model systému (ktorý bol realizovaný pomocou zobrazenia (12.2)). (B) Zobrazenie realizované neurónovou sieťou adaptovanou v druhej etape, reprezentujúcou inverzný model systému (realizovaný pomocou zobrazenia (12.5)).

Komplex inverzného a dopredného modelu (pozri obr. 12.7, diagram B) môže byť teraz chápaný ako zložená funkcia:

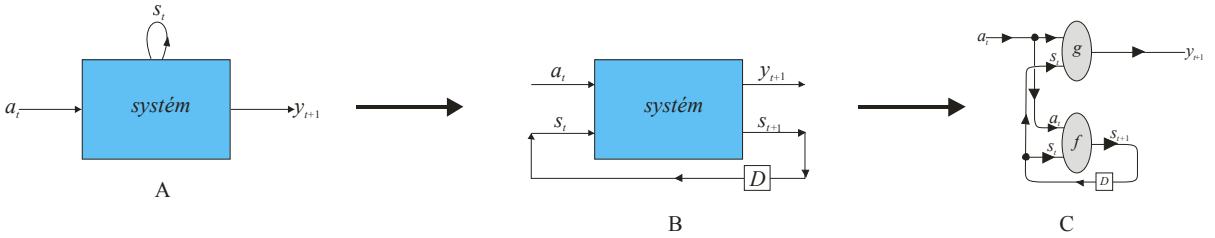
$$F_s(y) = f_s\left[f_s^{-1}(y)\right] = \begin{cases} y \quad pre\ y < 1/2, \quad 0 \quad pre\ y \geq 1/2 & (s=0) \\ y & (s=1) \\ y \quad pre\ y > 1/2, \quad 0 \quad pre\ y \leq 1/2 & (s=2) \end{cases} \tag{12.9}$$

To znamená, že v tomto konkrétnom prípade komplex znázornený na obr. 12.7 možno chápať ako jednotkové zobrazenie len v zúženom zmysle, a to na vhodných intervaloch. Táto skutočnosť je veľmi dôležitá pre správnu realizáciu adaptácie neurónovej siete tak v prvej, ako aj v druhej etape.

Získané numerické výsledky simulácie, realizované pomocou doprednej neurónovej siete obsahujúcej 15 skrytých neurónov, sú zosumarizované na obr. 12.11. Diagram A na tomto obrázku znázorňuje presnosť adaptovanej siete pre dopredný model, porovnaním s obr. 12.10, diagram A zistíme, že priebehy jednotlivých funkcií $f_s(a)$ sú veľmi podobné priebehom produkovaným neurónovou sieťou. Podobne diagram B na obr. 12.11 vyjadrujúci experimentálne priebehy inverzného modelu vypočítané príslušnou časťou neurónovej siete komplexu zodpovedá obr. 12.12, diagram B.

12.5 Alternatívne riešenie - spriahnutý model systému

V **podkapitole** 12.2 bol prezentovaný dopredný model systému. Ukážeme, že tento systém môže byť taktiež prezentovaný pomocou ***spriahnutého dopredného modelu***. Podľa obr. 12.3 systém môže byť reprezentovaný ako jednoduchý blok, pozri obr. 12.12.



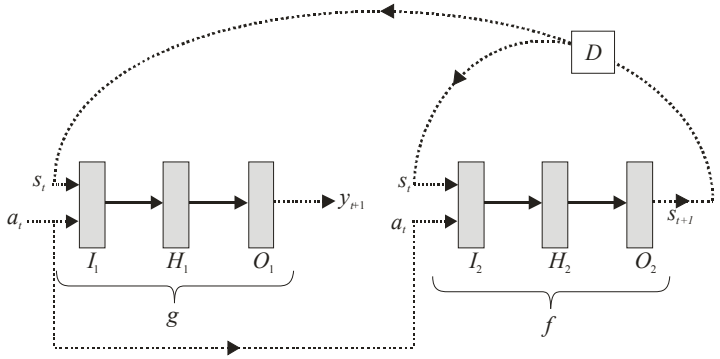
Obrázok 12.12. (A) Reprezentácia dopredného systému pomocou jedného bloku, ktorý na vstupe má akciu a_t , na výstupe má výsledok y_{t+1}, pričom transformácia vstupu na výstup je určená vnútorným stavom s_t. (B) Nech vstup dopredného systému je špecifikovaný dvojicou (s_t,y_t). Agent je schopný pretransformovať tento vstup na výstup (s_{t+1},y_{t+1}), formálne $(s_t,y_t) \rightarrow (s_{t+1},y_{t+1})$, blok D je oneskorovací element. (C) Vyjadrenie transformácie z diagramu B pomocou dvoch prechodových funkcií (12.10a-b),

V kapitole 1.2 bolo studované abstraktné výpočtové zariadenie nazývané konečnostavový stroj (alebo Mealyho automat), ktoré môže modelovať dopredný systém z obr. 12.12. Podobne ako v tomto prípade, aj teraz môžeme transformáciu $(s_t,y_t) \rightarrow (s_{t+1},y_{t+1})$ vyjadriť pomocou dvoch prechodových funkcií

$$s_{t+1} = f\left(a_t, s_t\right) \tag{12.10a}$$

$$y_{t+1} = g\left(a_t, s_t\right) \tag{12.10b}$$

Prvá formulula učuje následujúci vnútorný stav s_{t+1} pomocou prechodovej funkcie f s argumentami akcia a_t a stav s_t. Podobne, druhá formula (12.10b) určuje nasledujúci výsledok y_{t+1} pomocou prechodovej funkcie g s argumentami akcia a_t a stav s_t. Diagramatická reprezentácia týchto dvoch rovníc je znázornená na obr. 12.12, diagram C.

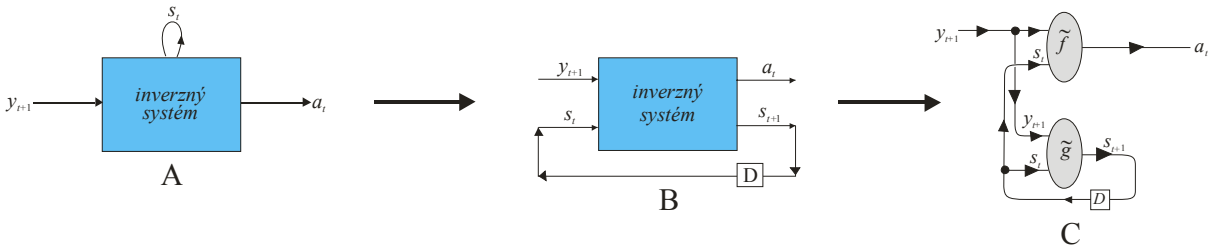


Obrázok 12.13. Vyjadrenie spriahnutých prechodových funkcií *f* a *g* pomocou dvoch 3-vrstvových dopredných neurónových sietí pre dopredný systém, efekt zpriahnutia týchto prechodových funkcií je reprezentovaný spätnými prerušovanými spojmi. Potom celková neurónová sieť je rekurentného charakteru, prerušované čiary reprezentujú spoje, ktoré vytvárajú orientované cykly. Hovoríme, že táto sieť je ***spriahnutým modelom dopredného systému***. Adaptačný proces tejto siete sa deje podľa zásad adaptácie rekurentných neurónových sietí (napr. metodou „back-propagation through time“, [xx]), pričom počiatočný stav s_1 je určený vopred. Výsledkom adaptačného procesu je postupnosť – trajektória $\left\{\left(s_t,a_t\right)/y_{t+1}; t=1,2,...,t_{\max}\right\}$

Podobným spôsobom môžeme diagramaticky reprezentovať aj inverzný systém, ktorého všeobecná bloková štruktúra je znázornená na obr. 12.14, diagramy A a B. Pomocou dvoch prechodových funkcií $\tilde{f}$ a $\tilde{g}$ vyjadríme výstupy z bloku inverzného systému

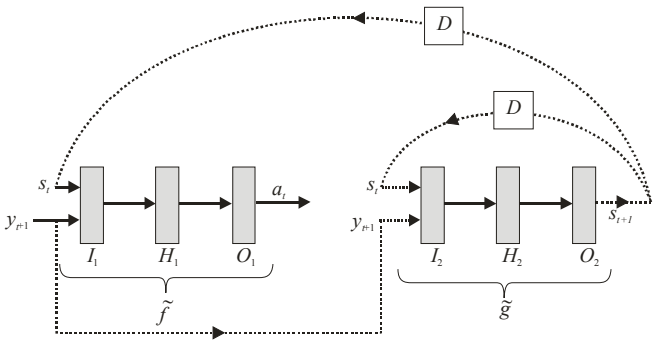
$$a_t = \tilde{f}\left(y_{t+1}, s_t\right) \tag{12.11a}$$

$$s_{t+1} = \tilde{g}\left(y_{t+1}, s_t\right) \tag{12.11b}$$



Obrázok 12.14. (A) Reprezentácia inverzného systému pomocou jedného bloku, ktorý na vstupe má výstup y_{t+1} , na výstupe je transformácia a_t, pričom transformácia vstupu na výstup je určená vnútorným stavom s_t. (B) Nech vstup inverzného systému je špecifikovaný dvojicou (y_{t+1},s_t), agent je schopný pretransformovať tento vstup na výstup (a_t,s_{t+1}), formálne $(y_{t+1},s_t) \rightarrow (a_t,s_{t+1})$. (C) Vyjadrenie transformácie z diagramu B pomocou dvoch prechodových funkcií (12.11a-b),

Príslušnú rekurentnú neurónovú sieť má tvar znázornený na obr. 12.15.



Obrázok 12.15. Diagram B znázorňuje rekurentnú neurónovú sieť, ktorú môžeme chápať ako ***spriahnutý model inverzného systém,*** ktorého základná štruktúra je reprezentovaná diagramom A. Spätné preušované čiary reprezentujú rekurentné spoje, ktoré vytvárajú orientované cykly. Adaptačný proces tejto siete bol špecifikovaný v legende k obr. 12.14.

Alternatívny prístup tejto podkapitoly ku konštrukcie dopredného a inverzného modelu, ktoré boli prezentované v podkapitolách 12.3 a 12.3 je založený na skutočnosti, že celkové prechodové funkcie $G\left(a_t, s_t; w\right)$ a $F\left(y_{t+1}, s_t; \tilde{\omega}\right)$ boli rozložené na spriahnuté elementárne prechodové funkcie f, g resp. $\tilde{f}$, $\tilde{g}$, ktoré sú špecifikované diagramami C obr. 12.12 resp. obr. 12.14. Ako určiť tieto elementárne prechodové funkcia pre študovaný systém. Za predpokladu, že vstupné, výstupné a vnútorné stavové veličiny sú prítupné k externému pozorovatelovi prostredníctvom tréningových množín, môžeme zostrojiť pre elementárne prechodové funkcie pomocou dopredných 3-vrstvových neurónových sietí. To znamená, že pre každý systém existujú modely (dopredný alebo inverzný) pomocou Minskeho výpočtového zariadenia nazývaného „konečnostavový stroj“ (alebo Mealyho automat).

Cvičenia

Cvičenie 12.1 Predstavte si hypotetickú situáciu, že ste generálnym riaditeľom fabriky na výrobu automobilov. Špecifikujte v tomto prípade výsledok systému, stav systému a pôsobenie akcie na zmenu týchto veličín. Aký je zámer agenta (generálneho riaditeľa) pri

riadení systému, pomocou čoho odhaduje svoje akcie, aby bol splnený jeho zámer. Opíšte význam dopredného modelu a aj inverzného modelu pre generálneho riaditeľa pri riadení systému.

Cičenie 12.2. aaa

Literatúra

[1] Johnson-Laird, P. N.: *Mental models*. Harvard University Press, Cambridge, MA, 1983.

[2] Jordan, M.I., Rumelhart, D.E.: Forward models: Supervised learning with a distal teacher. *Cognitive Science,* **16**(1992), 307-354.

[3] Minsky, M.: *Computation: Finite and infinite machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

[4] Ronco, E., Gawthrop, P.J.: Neural networks for modelling and control. *Technical Report,* csc97008, Centre for System and Control, Department of Mechanical Engineering, University of Glasgow, 1997.

[5] Stenning, K., van Lambalgen, M.: *Semantics and cognition: Semantic insight into the psychology of reasoning* . MIT Press, Cambridge, Ma, 2007.

13. prednáška

Symbolická teória Hofstadterovho systému *Copycat* pre tvorbu analógie

13.1 Úvodné poznámky

Analógia [8] je v kognitívnej vede a v umelej inteligencii špecifikovaná dvoma rôznymi spôsobmi:

(1) Reláciou medzi dvoma objektmi a_1 a a_2, pričom tieto objekty sú podobné inými objektom b_1 a b_2, $a_1 \sim b_1$ a $a_2 \sim b_2$, potom s určitou pravdepodobnosťou existuje analogická relácia aj medzi objektmi b_1 a b_2

$$\frac{\begin{array}{c} S(a_1, a_2) \\ (a_1 \sim b_1) \wedge (a_2 \sim b_2) \end{array}}{prob(S(b_1, b_2))} \tag{13.1}$$

Hovoríme, že doména relácie S bola rozšírená. Známy príklad tejto analógie je Ruthefordov planetárny model atómu. Relácia *S*(slnko, planéta) bola zavedená Kopernikom, podľa ktorej planéty obiehajú okolo slnka. Rutheford použil „podobnosti“ slnko~atómové jadro a planéta~elektrón k tomu, aby mohol rozšíriť doménu kopernikovskej relácie *S* aj pre model atómu, táto analógia Kopernikovho modelu sa nazýva Ruthefordov planetárny model atómu.

(2) Spôsobom usudzovania založeného na predpoklade, že ak dva objekty a_1 a a_2, sú si podobné vzhľadom k vybraným aspektom, potom tieto dva objekty sú si pravdepodobne podobné aj vzhľadom k iným aspektom:

$$\frac{\bigwedge_{i=1}^{n-1} S_i(a_1, a_2)}{prob(S_n(a_1, a_2))} \tag{13.2}$$

Tento tzv. funkcionálny typ analógie môžeme názorne ilustrovať pomocou odhadu ceny použitého auta, ktoré je dáme do relácií so známym typom auta, napr. podľa roku výroby, obsahu motoru, počtu najazdených kilometrov a pod. Ak tieto dve auta sú si podobné v mnohých aspektoch, tak môžeme predpokladať, že majú aj približne rovnakú cenu, t. j. cena použitého auta je s určitou pravdepodobnosťou rovná cene známeho typu auta .

Analogické usudzovanie patrí medzi ***induktívne módy*** usudzovania, t.j. jeho výsledkom je nová informácia, ktorá nebola pôvodne obsiahnutá vo východzích predpokladoch. Záver analogického usudzovania je platný len s určitou pravdepodobnosťou a je preto falzifikovateľný (K. Popper).

Hľadanie analógie medzi dvoma objektmi je operacionalizované pomocou zobrazenia čŕt jedného objektu na črty druhého objektu [1]. Čím viac takýchto zobrazení existuje, tým sú si skúmané objekty viac podobné, hovoríme, že medzi nimi existuje „analógia“. Tento prístup k formalizácii analógie môžeme ilustrovať pomocou jednoduchého príkladu, ktorý obsahuje 56 rôznych typov písmena A, pozri obr. 1. Ak si vyberieme dva ľubovolné typy z tejto množiny, potom musíme hľadať medzi nimi podobnosť – analógiu tak, že budeme porovnávať vybrané črty tak v prvom ako aj v druhom type. V prípade, že takýchto spoločných čŕt existuje veľa, potom vybrané typy sú navzájom silne analogické; v opačnom prípade, ak spoločných čŕt existuje len niekoľko, potom vybrané typy sú navzájom slabo analogické, alebo hovoríme, že len málo podobajú. Obidva tieto prípady môžu byť ilustrované vhodne vybranými dvojicami typov z obr. 1.

Obrázok 13.2. Rôzne typy písmena A [2], ktoré v niektorých prípadoch len veľmi ťažko možno identifikovať ako rovnaké písmeno. Pre každú dvojicu typov existuje homomorfné zobrazenie, ktoré priradí vybrané črty jedného písmena na črty druhého písmena. Ak počet týchto „priradení“ prekročí určitú prahovú hodnotu, môžeme povedať, že vybrané typy sú analogické, reprezentujú rovnaké písmeno A.

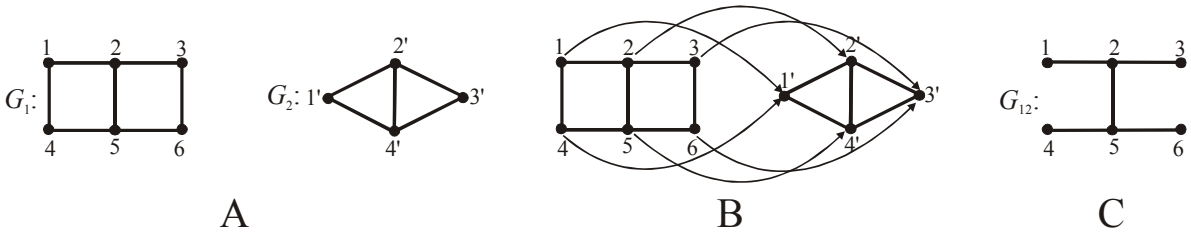
Popíšeme bližšie hľadanie analógie medzi dvoma objektmi O_1 a O_2, ktoré sú reprezentované grafmi G_1 a G_2. Konštrukcia týchto grafov môže byť urobené tak, že si zvolíme vybrané črty v oboch objektoch, tieto sú reprezentované vrcholmi v príslušných grafoch. Ak dve črty blízko súvisia (časovo, priestorovo alebo funkcionálne), potom ich vrcholy prepojíme hranou grafu. V umelej inteligencii [5] existuje mnoho rôznych metód ako reprezentovať objekty pomocou grafov (ktoré majú ohodnotené vrcholy a/alebo hrany). Hľadáme potom také zobrazenie množiny vrcholov jedného grafu na množinu vrcholov druhé grafu, ktoré zachováva maximálnu podobnosť medzi dvojicou

priradených vrcholov. Rozsah analógia medzi objektmi je priamoúmerný počtu týchto elementárnych podobností.

K formalizácii analógie medzi grafmi G_1 a G_2 použijeme koncepciu ***maximálneho spoločného podgrafu*** $G_{12} = (V_{12}, E_{12})$ pre ktorý platí (pozri obr. 13.2)

(1) spoločný graf G_{12} je súčasne podgrafom grafov G_1 a G_2, $G_{12} \subseteq G_1$ a $G_{12} \subseteq G_2$,

(2) spoločný podgraf G_{12} obsahuje maximálny počet hrán $|E_{12}| = max$,.



Obrázok 13:2. Diagram A znázorňuje dva grafy medzi ktorými neexistuje izomorfizmus (nie sú navzájom ekvivaletné, sú si navzájom len pdobné – analogické). Diagram B znázorňuje homomorfné zobrazenie pre konštrukciu analógie medzi dvoma grafmi G_1=(V_1,E_1) a G_2=(V_2,E_2) z diagramu A. Toto zobrazenie priraďuje vrcholy grafu G_1 na vrcholy grafu G_2, $f : V_1 \to V_2$. Pomocou tohto zobrazenia môžeme zostrojiť podobnosť medzi grafmi G_1 a G_2, ktorá spočíva v tom, že oba grafy majú dva cykly, ktoré sú susedné prostredníctvom jednej hrany.

Podobnosť medzi grafmi G_1 a G_2 vyjadríme pomocou maximálneho spoločného grafu takto

$$0 \le s(G_1, G_2) = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|} = \frac{|E_{12}|}{|E_1| + |E_2| - |E_{12}|} \le 1 \qquad (13:3)$$

kde $E_{12} = E_1 \cap E_2$ je množina hrán maximálneho spoločného podgrafu G_{12}, E_1 a E_2 sú množiny hrán grafov G_1 resp. G_2. Podobnosť $s(G_1, G_2) = 1$, ak grafy G_1 a G_2 sú izomorfné (rovnaké) , v opačnom prípade, ak tieto grafy sú rôzne, potom $0 < s(G_1, G_2) < 1$. Hľadanie maximálneho spoločného podgrafu je NP obtiažny problém, preto v kognitívnej vede používame ***rôzne heuristiky*** na jednoduchú konštrukciu tohto „zobrazenia“ .

13.2 Vzťah medzi analógiou a inteligenciou

Charles Spearman [7] je predchodca súčasného kognitívneho prístupu k inteligencii, pri jej špecifikácii rozlišoval tri etapy informačného procesu, ktorú budeme prezentovať spôsobom blízkym súčasnej umelej inteligencii a kognitívnej vede. Uvažujme jednoduché schéma, ktoré v prvom riadku obsahuje dva objekty A a B, medzi ktorými existuje relácia reprezentovaná symbolom ´:´. Cieľom riešenia tejto schémy je priradiť k objektu C nový objekt D na základe relácie medzi objektmi A a B.

$$\begin{array}{l} A \;:\; B \\ \underline{C \;:\; \quad} \\ \quad :\; D \end{array} \qquad (13.4)$$

Spearman rozlišoval tri etapy informačného procesu, ktoré je užitočné odlíšiť pri riešení tejto schémy a ktoré pokladal za elementárne akty kognitívneho procesu pri manifestácii kompetencie nazývanej inteligencia:

0. etapa – ***identifikácia*** objektov A a B, medzi ktorými existujú možné implicitné relácie znázornené symbolom ´:´,
1. etapa – ***inferencia*** explicitných relácií medzi časťami A a B,
2. etapa – ***aplikácia*** týchto explicitných relácií na objekt C tak, aby sme odvodili z nich taký objekt D, ktorý je v podobnej relácii s objektom C, ako je objekt B v relácii s objektom A.

Poznamenajme, že v mnohých prípadoch 0. etapa môže byť vynechaná, pretože riešiteľ problému (4) dostatočne dobre pozná štruktúru objektov A a B, preto môže hneď pristúpiť k 1. etape a potom k 2. etape.

Spearman takto operacionalizoval pojem inteligencia už v 20. rokoch minulého storočia, jeho prístup nesie všetky črty súčasného kognitívneho prístupu k inteligencii, v ktorom sa používajú modely, algoritmy a pod. k špecifikácii kognitívnych procesov prebiehajúcich v našej mysli. Ich hlavná výhoda spočíva v tom, že umožňujú nevrhnúť experimenty, pomocou ktorých môžeme experimentálne skúmať tieto procesy a navrhovať ich teoretické interpretácie.

Hofstadterov systém *Copycat* [3,4] môžeme chápať ako špeciálny prípad Spearmanovej schémy (4), ktorá je teraz zameraná na modelovanie riešenia problémov, ktoré sú založené na využití analógie medzi reťazcami znakov abecedy. Uvažujme nasledujúcu úlohu

$$\begin{array}{l} abc \Rightarrow abd \\ \underline{ijk \;\Rightarrow\; ?} \\ \Rightarrow \delta \end{array} \qquad (13.5)$$

Prvý riadok obsahuje reláciu dvoch objektov – reťazcov $abc \Rightarrow abd$, ktorá je interpretovaná tak, že ľavý reťazec je *zobrazený* na pravý reťazec. Riešiteľ v prvej etape inferencie si vytvára *model zobrazenia*, ktorý je v druhej etape aplikovaný na ľavý reťazec z druhého riadku $ijk \Rightarrow ?$ s cieľom vytvoriť pravý reťazec δ, t.j. nahradiť otáznik novým reťazcom, ktorý je v nejakom podobnom - analogickom vzťahu k reťazcu *ijk*, podobne ako je reťazec *abd* vo vzťahu k reťazcu *abc*. Môžeme teda povedať, že ku konštrukcii hľadaného reťazca δ využívame model relácie $abc \Rightarrow abd$, ktorá je založená na analógiách odpozorovaných pri analýze tejto relácie. V nasledujúcej tabuľke uvedieme niekoľko príkladov riešenia úlohy (13.5)

#	δ	Komentár
1	*ijl*	*k* je nahradené nasledujúcim *l*
2	*ijd*	*k* je nahradené *d*
3	*abd*	celý reťazec je nahradený reťazcom *abd*

Existuje množstvo ďalších riešení úlohy (13.3), ktoré sú založené na „sofistikovaných“ interpretáciach relácie $abc \Rightarrow abd$. Prvé riešenie obsahuje je založené na pozorovaní, že prvé dva znaky zostávajú nezmenené, zatiaľ čo tretí znak je zmenený na svojho nasledovníka. Druhé riešenie vychádza z podobného pozorovania, avšak v tomto prípade druhá časť pozorovania (tretí znak je zamenený za svojho nasledovníka) je

zamenená za menej abstraktnejšie (alebo menej všeobecnejšie) pozorovanie, že tretí znak je zamenený za písmeno *d*. Tretie riešenie *abd* bolo zostrojené na základe pozorovania, že ľavy reťazec *abc* bol nahradený reťazcom *abd*; toto riešenie je najjednoduchšie, nevyužíva žiadne vnútorné relácie medzi jednotlivými znakmi z reťazcov. Môžeme sa domnievať, že prvé riešenie *ijl* je najlepšie, využíva najviac vnútorných vzťahov medzi písmenami.

V evolučnej psychpológii sa predpokladá, že kognitívna kompetencia hľadať analógie medzi rôznymi objektmi a javmi okolitého sveta podstatne ovplyvnila nasledujúci vznik celej plejády ďalších dôležitých kognitívnych kompetencií týkajúcich sa schopnosti zovšeobecňovať pomocou indukcie a abdukcie [6]. Emergencia týchto kognitívnych kompetencií je dôležitým predpokladom u ľudí predvídať a plánovať aktivity v blízkom alebo aj vzdialenom časovom horizonte, čo mohlo mať zásadný význam pre evolučné presadenia sa *homo sapiens* ako nového perspektívneho druhu.

3. Analýza relácie $\alpha \Rightarrow \beta$

Majme dva reťazce α a β, ktoré sú vytvorené z malých písmen anglickej abecedy, ktoré môžu mať rôznu dĺžku (t.j. obsahujú rôzny počet písmen). Tieto dva reťazce sú vo vzájomnej relácii označenej $\alpha \Rightarrow \beta$, ktorej interpretácia je taká, že reťazec β je vytvorený z reťazca α pomocou elementárnych zobrazení – substitúcií, kde jednotlivé písmená sú nahradené inými písmenami . Našim cieľom je ukázať jednoduchý formálny spôsob analýzy tejto relácie $\alpha \Rightarrow \beta$, pričom výsledkom tejto analýzy bude množina elementárnych zobrazení, ktorá bude zovšeobecnená tak, aby bola aplikovateľná aj k inému reťazcu γ za vzniku nového reťazca δ

$$\begin{array}{l} \alpha \Rightarrow \beta \\ \underline{\gamma \Rightarrow ?} \\ \Rightarrow \delta \end{array} \qquad (13.6)$$

Cieľom tejto konštrukcie je, aby novovytvorený reťazec δ bol podobný/analogický reťazcu γ tak, ako je podobný reťazec β reťazcu α. Ináč povedané, konštrukcia je založená na podmienke, aby si relácie $\alpha \Rightarrow \beta$ a $\gamma \Rightarrow \delta$ boli podobné/analogické v tom, že spôsob kreovania pravého reťazca δ z ľavého reťazca γ bol založený na elementárnych transformáciach, ktoré boli použité v pôvodnej relácii $\alpha \Rightarrow \beta$.

Nech $A = \{a, b, ..., z\}$ je *abeceda* pozostávajúca zo všetkých 26 malých písmen anglickej abecedy. Pre jednotlivé písmená tejto abecedy existujú dve funkcie, nasledovník (***succ**essor*) a predchodca (***pred**ecessor*) (pozri obr. 13.3)

$$succ(a_i) = \begin{cases} a_{i+1} & (pre\ 1 \le i \le 25) \\ neexistuje & (pre\ i = 26) \end{cases} \qquad (13.7a)$$

$$pred(a_i) = \begin{cases} neexistuje & (pre\ i = 1) \\ a_{i-1} & (pre\ 2 \le i \le 26) \end{cases} \qquad (13.7b)$$

Obrázok 3. Znázornenie usporiadania abecedy, každé písmeno ξ (okrem *a*) má jedného predchodcu *pred*(ξ) a každé písmeno ζ (okrem *z*) jedného nasledovníka *succ*(ζ).

Pre reťazec $\boldsymbol{\alpha} = (\alpha_1\alpha_2...\alpha_n)$ zavedieme pojem ľavého krajného a pravého krajného písmena

$$ml(\boldsymbol{\alpha}) = \alpha_1 \quad \text{a} \quad mr(\boldsymbol{\alpha}) = \alpha_n \tag{13.8a}$$

Podobne, pre abecedu *A* zavedieme pojem prvého a posledného písmena

$$ml(A)=a \quad \text{a} \quad mr(A)=z \tag{13.8b}$$

Množina A^+ obsahuje všetky neprázdne reťazce vytvorené nad abecedou *A*

$$A^+ = \bigcup_{n\geq 1} A^n \tag{13.9}$$

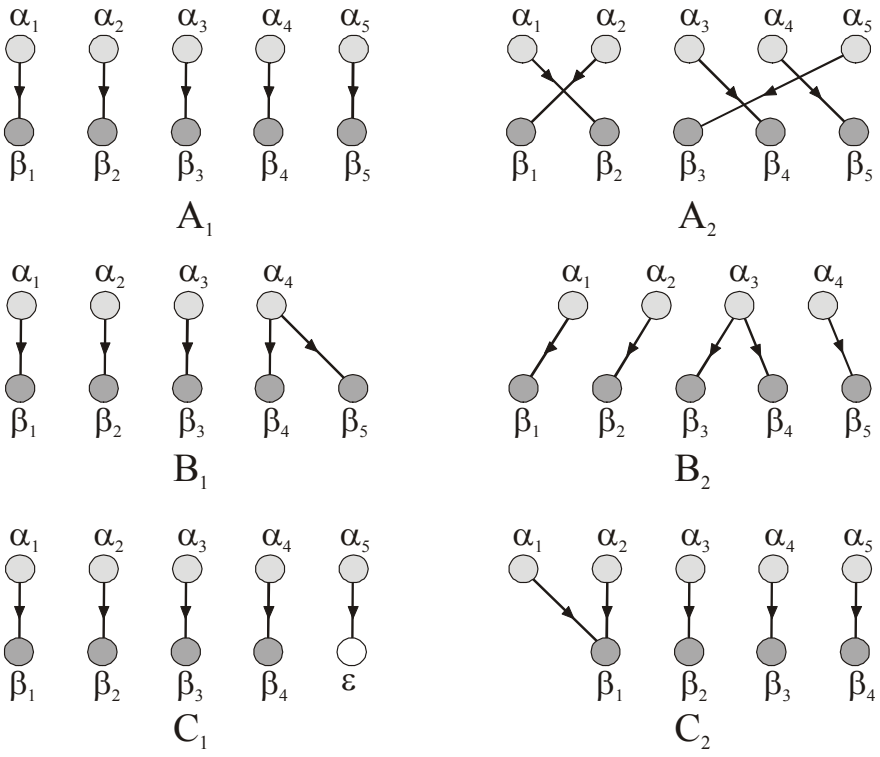
Uvažujme dva reťazce $\boldsymbol{\alpha},\boldsymbol{\beta} \in A^+$, ktorých dĺžka (počet znakov) je špecifikovaná $|\boldsymbol{\alpha}| = m$ a $|\boldsymbol{\beta}| = n$

$$\boldsymbol{\alpha} = (\alpha_1\alpha_2...\alpha_m) \text{ a } \boldsymbol{\beta} = (\beta_1\beta_2...\beta_n) \tag{13.10}$$

Pre tieto dva reťazce $\boldsymbol{\alpha},\boldsymbol{\beta} \in A^+$, ktoré sú v relácii $\boldsymbol{\alpha} \Rightarrow \boldsymbol{\beta}$, zavedieme zobrazenie $\boldsymbol{f} : \boldsymbol{\alpha} \to \boldsymbol{\beta}$, ktoré je špecifikované pomocou *m* elementárnych zobrazení takto (pozri obr. 4)

$$\begin{aligned} f^{(1)} &: \alpha_1 \to \beta_{p_1} \\ f^{(2)} &: \alpha_2 \to \beta_{p_2} \\ &..... \\ f^{(m)} &: \alpha_m \to \beta_{p_m} \end{aligned} \tag{13.11}$$

kde indexy $p_1, p_2,..., p_m \in \{0,1,2,...,n\}$. Index ´0´ je použitý pre elementárne zobrazenie s pravou stranou obsahujúcou prázdny znak ε (pozri diagram C_1 na obr. 13.4). Tento špeciálny typ elementárneho zobrazenia je obvykle použitý vtedy, keď $m > n$.



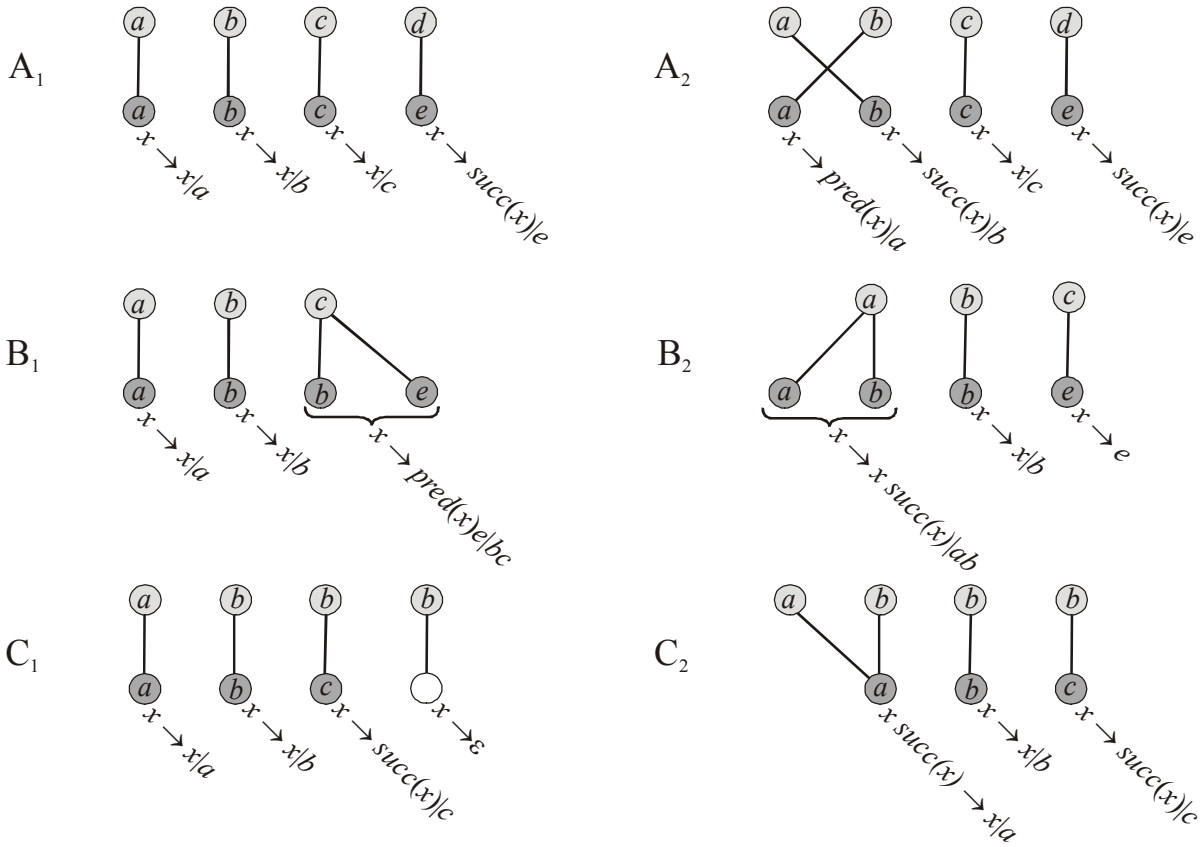
Obrázok 13.4. Tri rôzne prípady konštrukcie elementárnych zobrazení pre reláciu $\alpha \Rightarrow \beta$. (A) Zobrazenie pre reťazce, ktoré majú rovnakú dĺžku, $m=n$. (B) Zobrazenie pre reťazce s rôznou dĺžkou, kde $m<n$. (C) Zobrazenie pre reťazce s rôznou dĺžkou.

Elementárne zobrazenia (11) prepíšeme do zovšeobecneného tvaru pomocou nasledujúcej konvencie

$$f \;:\; \alpha \to \beta \Rightarrow \begin{cases} x \to x \,|\, \beta & (\text{ak } \alpha = \beta) \\ x \to pred(x) \,|\, \beta & (\text{ak } \alpha = pred(\beta)) \\ x \to succ(x) \,|\, \beta & (\text{ak } \alpha = succ(\beta)) \\ x \to \varepsilon & (\text{ak } \beta = \varepsilon) \\ x \to \beta & (\text{ináč}) \end{cases} \tag{13.12}$$

kde symbol x reprezentuje ľubovoľný znak z abecedy A. „Substitučné“ pravidlo $x \to y \,|\, \beta$ sa interpretuje tak, že symbol x v danej polohe v reťazci α môže byť alternatívne nahradený písmenom y alebo pôvodným písmenom β. Neskoršie ukážeme, že tieto dve alternatívy nie sú rovnocenné, zavedieme určitú ohodnotenie výsledného riešenie, ktoré usporiada rôzne riešenia tej istej úlohy podľa klesajúceho ohodnotenia; čím menšie ohodnotenie získa riešenie, tým viac ho možno pokladať jeho výskyt za menej pravdepodobný.

V prípade, že reťazce α a β obsahujú rôzny počet znakov, potom elementárne zobrazenia (13.9) musia byť trochu zložitejšie ako tie, ktoré sú uvedené v (13.10). Tieto špeciálne typy zobrazenia nebudeme formulovať na všeobecnej úrovni, pomocou ilustračných príkladov z obr. 13.5 ukážeme tvorbu elementárnych zobrazení pre rôzne prípady.



Obrázok 13.5. Konštrukcia množiny elementárnych zobrazení pre rôzne typy relácií $\alpha \Rightarrow \beta$.

Príklad 1 (Diagram A_1 na obr. 13.5). Pre reťazce $\alpha = (abcd)$, $\beta = (abce)$, ktoré majú rovnakú dĺžku $m = n = 4$, elementárne zobrazenia (10) majú tvar

$$f_1 : a \to a \Rightarrow x \to x \,|\, a$$
$$f_2 : b \to b \Rightarrow x \to x \,|\, b$$
$$f_3 : c \to c \Rightarrow x \to x \,|\, c$$
$$f_4 : d \to e \Rightarrow x \to succ(x) \,|\, e$$

Množina elementárnych zobrazení potom obsahuje štyri elementy

$$F(\alpha,\beta) = \{x \to x \,|\, a, x \to x \,|\, b, x \to x \,|\, c, x \to succ(x) \,|\, e\}$$
$$= \{x \to x \,|\, a \,|\, b \,|\, c, x \to succ(x) \,|\, e\}$$

Príklad 2 (Diagram A_2 na obr. 13.5). Pre rovnaké reťazce $\alpha = (abcd)$ a $\beta = (abce)$, aké boli použité v predchádzajúcom príklade, elementárne zobrazenia majú tvar

$$f_1 : a \to b \Rightarrow x \to succ(x) \,|\, b$$
$$f_2 : b \to a \Rightarrow x \to pred(x) \,|\, a$$
$$f_3 : c \to c \Rightarrow x \to x \,|\, c$$
$$f_4 : d \to e \Rightarrow x \to succ(x) \,|\, e$$

Množina elementárnych zobrazení potom obsahuje štyri elementy

$$F(\alpha,\beta) = \{x \to succ(x) \mid b,\ x \to pred(x) \mid a,\ x \to x \mid c,\ x \to succ(x) \mid e\}$$

Elementárne zobrazenie z príkladov 1 a 2 boli vytvorené pre rovnakú úlohu $(abcd) \Rightarrow (abde)$, avšak pri ich tvorbe boli použité rozdielne súbory indexov (p_1, p_2, p_3, p_4), v prvom prípade sme použili (1234) a v druhom prípade sme použili „transpozíciu“ (2134).

Príklad 3 (diagramy B_1 a B_2 na obr. 13.5), kde $\alpha = (abc)$ a $\beta = (abbe)$. Množina elementárnych zobrazení potom obsahuje len tieto štyri elementy

$$F_1(\alpha,\beta) = \{x \to x \mid a,\ x \to x \mid b,\ x \to pred(x)e \mid bc\}$$

$$F_2(\alpha,\beta) = \{x \to xsucc(x) \mid ab,\ x \to x \mid b,\ \ x \to e\}$$

Príklad 4 (diagramy C_1 a C_2 na obr. 13.5), kde $\alpha = (abbb)$ a $\beta = (abc)$. Množiny elementárnych zobrazení obsahuje tieto elementy

$$F_1(\alpha,\beta) = \{x \to x \mid a,\ x \to x \mid b,\ x \to succ(x) \mid c,\ x \to \varepsilon\}$$

$$F_2(\alpha,\beta) = \{xsucc(x) \to x \mid a,\ x \to x \mid b,\ x \to succ(x) \mid c\}$$

13.4 Aplikácia množiny elementárnych zobrazení k tvorbe reťazca δ

Majme množinu elementárnych zobrazení $F(\alpha,\beta)$, ktorej vlastnosti boli podrobne diskutované v predchádzajúcej časti tohto rukopisu, aplikujeme túto množinu na reťazec $\gamma = (\gamma_1\gamma_2...\gamma_{m'})$ s cieľom získať nový reťazec $\delta = (\delta_1\delta_2...\delta_{n'})$, pričom tieto reťazce sú v relácii $\gamma \Rightarrow \delta$, ktorá je analogická relácii $\alpha \Rightarrow \beta$. Formálne môžeme tvorbu reťazca vyjadriť takto

$$F(\alpha,\beta)(\gamma) = \delta \tag{13.13}$$

Táto formula sa interpretuje tak, že elementárne zobrazenia z množiny $F(\alpha,\beta)$ sa postupne aplikujú na elementy reťazca γ, pomocou tohto procesu vytvoríme nový reťazec δ, ktorý je pokladaný za riešenie danej úlohy systému Copycat. Zobrazenie F môžeme interpretovať takto

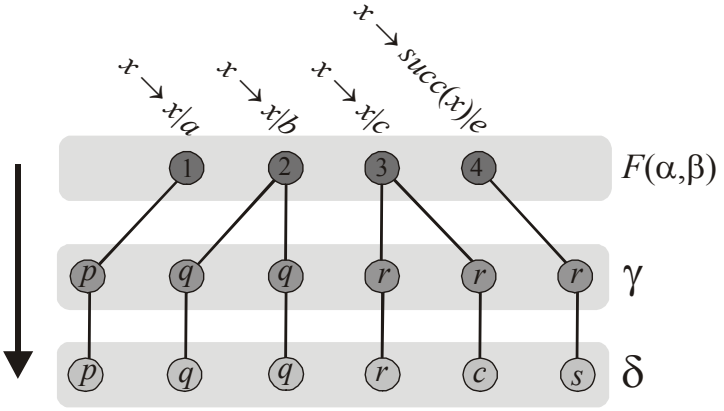
$$F : \underbrace{A^m \times A^n}_{\alpha \Rightarrow \beta} \times \underbrace{A^{m'}}_{\gamma} \to \underbrace{A^{n'}}_{\delta} \tag{13.14}$$

Ako už bolo naznačené, zobrazenie F nie je striktne deterministické, ale obsahuje aj určitú stochastickú zložku spočívajúcu v nasledujúcich dvoch aspektoch:

(1) Idúc zľava doprava po jednotlivých písmenách reťazca γ vyberáme z množiny $F(\alpha,\beta)$ postupne jednotlivé elementárne zobrazenie,

(2) tieto pravidlá aplikujeme na dané písmeno z reťazca γ, takto postupne vzniká reťazec δ.

Elementárne „substitučné“ pravidlá obvykle obsahujú dve alternatívne možnosti, napríklad pravidlo $x \to succ(x) \mid a$ obsahuje dve možnosti. Tieto dve rôzne možnosti spočívajú v tom, že dané písmeno *x* je buď nahradené jeho nasledovníkom *succ*(*x*) alebo písmenom *a*.



Obrázok 13.6. Diagramatická ilustrácia aplikácie množiny elementárnych transformácií $F(\alpha,\beta)$ k reťazcu $\gamma = pqqrrr$ za vzniku výsledného reťazca $\delta = pqqrcs$, ktorý reprezentuje jedno z možných riešení úlohy. Charakter aplikácie „funkcie“ $F(\alpha,\beta)$ k reťazcu γ má určité stochastické črty spočívajúce v tom, že výber elementárnej transformácie z $F(\alpha,\beta)$ môže byť v určitom malom rozsahu náhodný a taktiež aj v tom, že aj aplikácia vybranej elementárnej transformácie môže mať obvykle dve alternatívne možnosti, z ktorých vyberáme obvykle prvú možnosť.

Študujme schému z obr. 13.5, diagram A_1

$$\begin{array}{l} abcd \quad \Rightarrow abce \\ \underline{pqqrrr \Rightarrow \quad ?} \\ \Rightarrow \quad \delta \end{array} \qquad (13.15)$$

Cieľom riešenia tejto schémy je navrhnúť v treťom riadku na pravej strane (nahradiť otáznik) nový reťazec δ tak, aby vytvorená relácia $pqqrrr \Rightarrow \delta$ bola *analogická* k relácii $abcd \Rightarrow abce$. Tento cieľ uskutočníme pomocou nasledujúcich dvoch etáp:

1. etapa. Zostrojíme množinu elementárnych zobrazení $F(\alpha,\beta)$.

2. etapa. Elementárne transformácie aplikujeme postupným systematickým spôsobom (idúc zľava doprava) na reťazec $\gamma = pqqrrr$, týmto spôsobom vytvoríme hľadaný reťazec δ.

Poznamenajme, že obe etapy majú určitý stochastický charakter, existuje rôzne alternatívy tak pri tvorbe elementárnych zobrazení pomocou relácie $\alpha \Rightarrow \beta$, ako aj ich aplikácie k reťazcu γ.

Prvá etapa už bola pre danú reláciu $abcd \Rightarrow abce$ uskutočnená v príklade 1, viedla na nasledujúcu množinu elementárnych transformácií

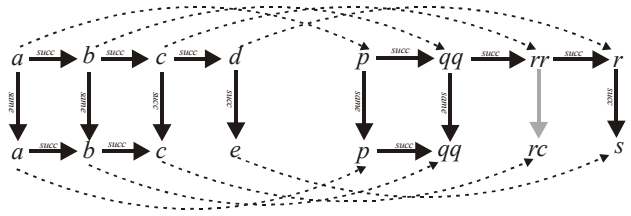
$$F(\alpha,\beta) = \{x \to x \mid a,\ x \to x \mid b,\ x \to x \mid c,\ x \to succ(x) \mid e\}$$

V druhej etape je táto množina aplikovaná k reťazcu $\gamma = pqqrrr$ za vzniku nového reťazca $\delta = pqqrcs$, ktorý môže byť chápaný ako výsledok riešenia schémy, pozri obr. 6, schéma má riešenie

$$\begin{array}{l} abcd \Rightarrow abce \\ \underline{pqqrrr \Rightarrow \ ?} \\ pqqrcs \end{array}$$

Musíme však poznamenať, že tento výsledok $\delta = pqqrcs$ je jeden z možných, existujú aj iné alternatívne možnosti, napr. $\delta = pqqcce$, ktorý dostaneme tak, že tretie a štvrté pravidlo z $F(\alpha,\beta)$ aplikujeme tak, že použijeme druhú „konštantnú“ alternatívnu možnosť z pravej strany.

Proces tvorby reťazca $\delta = pqqrcs$ z reťazca $\gamma = pqqrrr$ na základe relácie $abcd \Rightarrow abce$ je znázornený na obr. 13.7.



Obrázok 13.7. Grafické znázornenie riešenia (13.15), kde pomocou relácie $abcd \Rightarrow abce$ (ľavý diagram) zostrojíme z reťazca $\gamma = pqqrrr$ nový cieľový reťazec $\delta = pqqrcs$. Prerušované orientované čiary reprezentujú homorfizmus medzi ľavým a pravým diagramom, podrobnejšie vysvetlenie tohto grafu je uvedené v kapitole 13.5.

13.5 Diagramtická interpretácia zobrazenia a konštrukcia miery podobnosti novovytvoreného reťazca γ

Študujme úlohu

$$\begin{array}{l} abc \Rightarrow abd \\ \underline{ijk \ \Rightarrow \ ?} \\ \Rightarrow \delta \end{array} \qquad (13.16)$$

Množinu elementárnych zobrazení $F(abc,abd)$ obsahuje tieto elementárne transformácie

$$\begin{array}{l} f_1 : a \to a \Rightarrow x \to x \,|\, a \\ f_2 : b \to b \Rightarrow x \to x \,|\, b \\ f_3 : c \to d \Rightarrow x \to succ(x) \,|\, d \end{array} \qquad (13.17a)$$

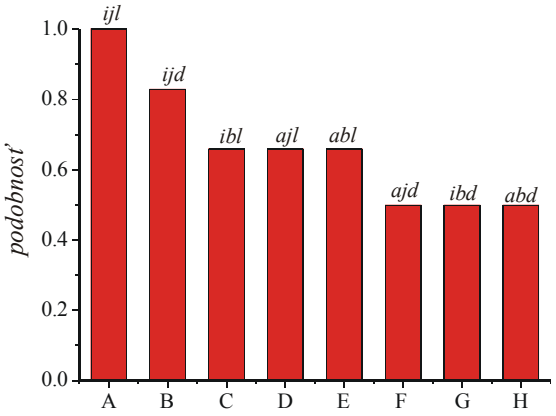
potom

$$F(abc,abd) = \left\{ \underbrace{x \to x \,|\, a}_{f_1}, \underbrace{x \to x \,|\, b}_{f_2}, \underbrace{x \to succ(x) \,|\, d}_{f_3} \right\} \qquad (132.17b)$$

V Tabuľke 1 máme osem rôznych spôsobov elementárnych zobrazení (transformácií) na reťazec *ijk*, kde tretí stĺpec obsahuje podobnosť medzi transformáciami $abc \Rightarrow abd$ a $ijk \Rightarrow \delta$. Grafické znázornenie týchto výsledkov je uvedené na obr. 7 a 8.

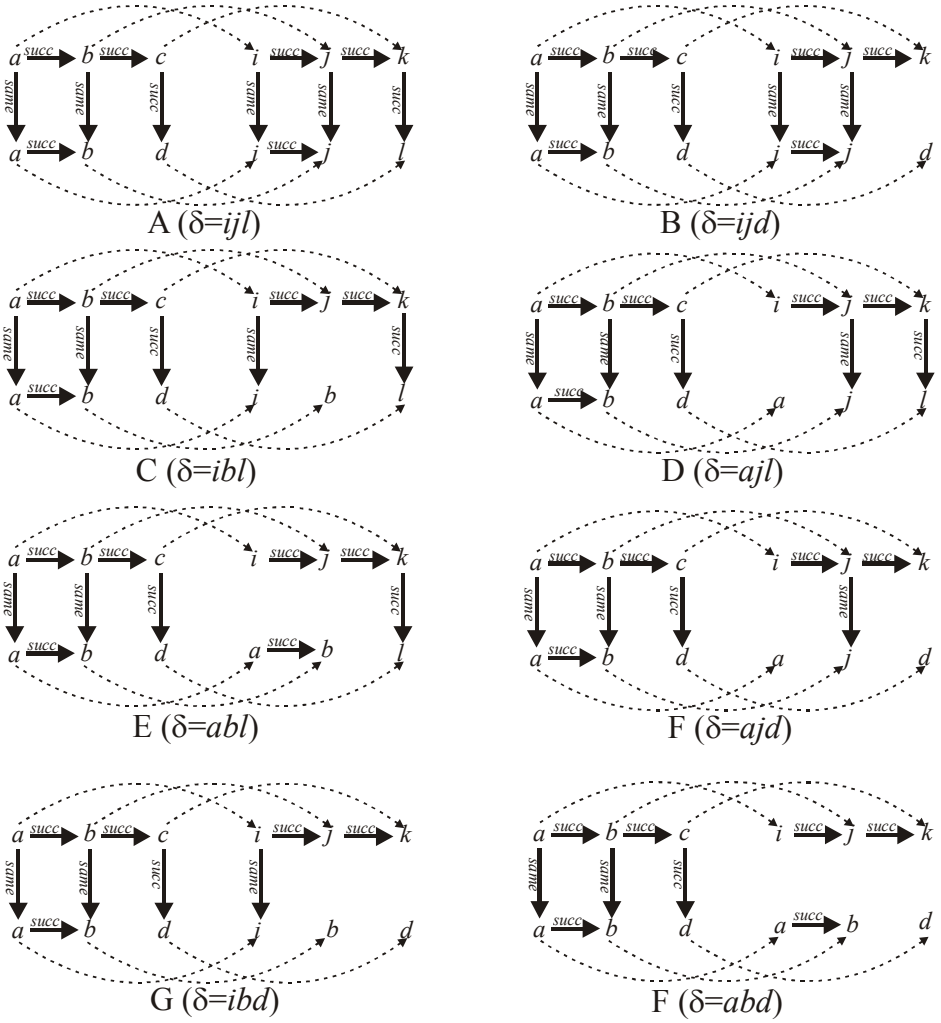
Tabuľka 13.1. Osem rôznych aplikácií množiny elementárnych zobrazení schémy (16) aplikovaných na reťazec *ijk*.

#	*riešenie*	*podobnosť*	*elementárne zobrazenia*
A	*ijl*	1	$f_1 : x \to x, f_2 : x \to x, f_3 : x \to succ(x)$
B	*ijd*	5/6	$f_1 : x \to x, f_2 : x \to x, f_3 : x \to d$
C	*ibl*	4/6	$f_1 : x \to x, f_2 : x \to b, f_3 : x \to succ(x)$
D	*ajl*	4/6	$f_1 : x \to a, f_2 : x \to x, f_3 : x \to succ(x)$
E	*abl*	4/6	$f_1 : x \to a, f_2 : x \to b, f_3 : x \to succ(x)$
F	*ajd*	3/6	$f_1 : x \to a, f_2 : x \to x, f_3 : x \to d$
G	*ibd*	3/6	$f_1 : x \to x, f_2 : x \to b, f_3 : x \to d$
H	*abd*	3/6	$f_1 : x \to a, f_2 : x \to b, f_3 : x \to d$



Obrázok 13.8. Histogram podobností pre osem rôznych možností $abc \Rightarrow abd$, $ijk \Rightarrow \delta$, kde reťazec δ je uvedený pre každú možnosť nad príslušným stĺpcom histogramu, pozri tab. 13.1.

Histogram znázornený na obr. 13.8 môže byť interpretovaný tak, že ukazuje relatívnu významnosť jednotlivých riešení, ktorá je založená na koncepcii maximálneho spoločného podgrafu. Najvýznamnejšie je riešenie δ = *ijl* , ktoré ja založené na maximálnej „podobnosti“ s východiskovou reláciou $abc \Rightarrow abd$. Ostatné riešenia (pozri riadky B – H v tab. 1) majú už menšiu relatívnu významnosť. Ďalšie riešenie je δ = *ijd,* ktoré vznikne tak, že tretie elementárne zobrazenie $f_3 : x \to x|d$ sa použije v druhej alternatívnej forme $x \to d$.



Obrázok 13.9. Osem rôznych spôsobov konštrukcie relácie $ijk \Rightarrow \delta$ na základe relácie $abc \Rightarrow abd$. Každá z týchto možností obsahuje dve časti: ľavá časť je priradená relácii $abc \Rightarrow abd$, zatiaľ čo, pravá časť obsahuje reláciu $ijk \Rightarrow \delta$, kde reťazec δ má konkrétny tvar. Prerušované orientované čiary medzi pravou a ľavou časťou znázorňuje zobrazenie homomorfizmu, pomocou ktorého zisťujeme maximálny spoločný podgraf medzi oboma časťami.

13.6 Záver

Ukázali sme, že Hofstadterov systém Copycat je možné formulovať taktiež aj na symbolickej úrovni. Navrhnutý prístup môže slúžiť ako model mentálnej reprezentácie Copycatu v našej mysli. Poskytuje určité vzájomné poradie riešení, čo môže byť aj predmetom experimentálneho štúdia Copycatu a vhodnosti navrhnutej symbolickej reprezentácie.

Literatúra

[1] Gentner, D.: Structure mapping : A theoretical framework for analogy. *Cognitive Science* **7** (1983), 155-170.

14. prednáška

Logický model hry Sudoku

14.1 Úvodné poznámky

V súčasnosti hra Sudoku patrí medzi populárne hry, ktorá sa v priebehu ostatných 10-15 rokov rozšírila po celom svete. Menovite, medzi seniormi je veľmi populárna, gerontopsychológovia ju odporúčajú ako jednoduchú prevencie proti Alzheimerovej nemoci. Riešenie tejto hry je pomerne jednoduché. Hra Sudoku je tak špecifikovaná, že má práve jedno riešenie, k nájdeniu tohto riešenia stačí schopnosť riešiteľa používať jednoduché zákony klasickej výrokovej logiky. Môžeme si položiť otázku, prečo je táto hra taká populárna? Jeden z hlavných dôvodov tejto obľúbenosti je skutočnosť, že k nájdeniu riešenia Sudoku postačuje papier a ceruzka a schopnosť logicky uvažovať. Hra je dostatočne zložitá k tomu, aby vyžadovala od riešiteľa seriózne mentálne úsilie a súčasne stačí len jednoducho „skanovať" riadky/stĺpce/podoblasti akými číslami sú - nie sú zaplnené.

Cieľom tejto práce je prezentovať nielen jednoduchý logický algoritmus pre riešenie hry Sudoku, ale aj pohľad súčasnej umelej inteligencie [10] a kognitívnej vedy [13] na tvorbu *mentálneho modelu*[1] tejto hry. Kognitívna veda sa už niekoľko desaťročí zaoberá problémom, ako ľudia používajú jednoduché logické zákony, riešia problémy, rozhodujú sa,... a aké sa pritom vyskytujú chyby. Inovatívnosť kognitivistického pohľadu na riešenieSudoku [5,6,12] spočíva v tom, že navrhuje mentálne modely kognitívnych aktivít človeka pri riešení jednoduchých logických úloh. Požaduje, aby tieto modely boli schopné postihnúť nielen správne riešenia ale aj interpretovať chyby pri konštrukcii riešenia. Mentálny model má určitú zložitosť, ktorá je premenlivá pri aplikovaní na rôzne problémy (napr. na rôzne kategorické sylogizmy) [5,6,11], pričom očakávame, že táto zložitosť modelu koreluje s výskytom chýb.

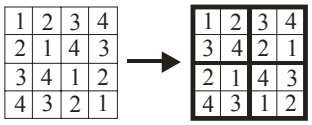
Navrhnutý mentálny hry Sudoku bude obsahovať hierarchicky usporiadané pravidlá, od jednoduchých pravidiel založených na definícii Sudoku, až po zložité netriviálne pravidlá, ktoré sú vyžadujú interakciu rôznych časti matice hry. Pomocou týchto pravidiel môžeme aj klasifikovať danú hru Sudoku podľa rôznych stupňov obťažnosti. Najnižšie stupne obťažnosti vyžadujú použitie len prvých najnižších pravidiel, avšak riešenie obťažnejších hier Sudoku vyžaduje obvykle už použitie pravidiel vyšších stupňov. Úspešné riešenie najobťažnejších hier Sudoku (nazývaných angl. „diabolical" alebo „fiendly") vyžadujú už použitie veľmi komplikovaných „kolektívnych" pravidiel, ktoré, pretože ich dobre nepoznáme, je jednoduchšie ich substituovať spätným prehľadávaním [10] (angl. back-track searching). V tejto práci sa budeme explicitne zaoberať len jednoduchšími pravidlami, ktoré riešia hru

[1] *Mentálny model* je psychologická reprezentácia reálnej alebo imaginárnej situácie, prvý krát bol postulovaný škótskym psychológom Kennethom Craikom v r. 1943 [1]. Predpokladal, že v našej mysli vznikajú modely skutočnosti (reálnej alebo imaginárnej), ktoré napomáhajú k pochopeniu udalostí a javov vznikajúcich v súvislosti s danou skutočnosťou. Podľa Johnsona – Lairda [5,6] tieto modely vznikajú v pracovnej pamäti našej mysle ako výsledok percepcie a/alebo imaginácie, snahy vysvetliť a pochopiť danú situáciu. Základnou črtou týchto modelov je, že sú štrukturálne podobné objektom, ktoré reprezentujú (asi tak, ako drevené modely molekúl v chémii ku skutočným molekulám).

Sudoku až po strednú a začínajúcu diabolskú úroveň obťažnosti (t. j. asi 95% publikovaných hier Sudoku v dennej tlači).

14.2 Špecifikácia hry Sudoku

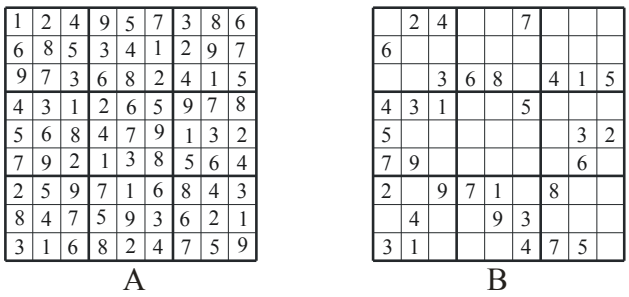
Sudoku [2,3,]je modifikácia klasickej úlohy známej už v staroveku, ktorú švajčiarsky matematik Leonhard Euler nazval ***latinský štvorec***. V tejto hre na ortogonálnej mriežke typu $n \times n$ v každom riadku a stĺpci musíme umiestniť prvých n prirodzených čísel $\{1, 2, ..., n\}$ tak, aby v každom stĺpci a riadku sa každé číslo vyskytovalo práve raz. Na obr. 1 je znázornený „latinský štvorec“ pre $n = 4$.. Táto jednoduchá hra, môže byť ľahko modifikovaná tak, že ortogonálna mriežka 4×4 je rozdelená na 4 podoblasti 2×2 tak, aby v každej podoblasti sa vyskytovali prirodzené čísla práve raz. To znamená, že pôvodná požiadavka, aby sa v každom stĺpci a riadku ortogonálnej mriežky prirodzené čísla vyskytovali práve raz (t. j. aby sa žiadne číslo neopakovalo dva- alebo viac-krát), bola rozšírená o ďalšiu podmienku výskytu čísel v podoblastiach (pozri obr. 1).



Obrázok 14.1. (A) Latinský štvorec typu 4×4. V každom riadku a stĺpci sa prirodzené číslo z {1, 2, 3, 4} práve raz. (B) Z latinského štvorca môžeme jednoducho zostrojiť hru Sudoku na ortogonálnej mriežka 4×4, ktorá je rozdelená na menšie podoblasti 2×2, potom sa zavedie dodatočná tretia podmienky, aby v každej podoblasti sa vyskytovalo prvých 4 prirodzených čísel práve raz.

Po týchto vstupných úvahách o hre Sudoku, pristúpime k špecifikácii jej ***finálnej pozície***:

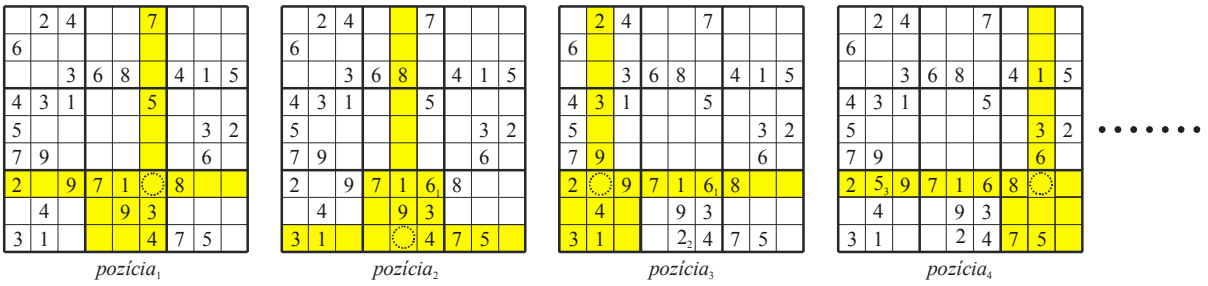
(1) Hra Sudoku je realizovaná na ortogonálnej mriežke typu 9×9, pričom táto mriežka je rozdelená na 9 podoblastí typu 3×3 (pozri obr. 2).
(2) Každá bunka z ortogonálnej mriežky je vyplnená číslom z množiny obsahujúcich prvých deväť prirodzených čísel {1,2,...,9}.
(3) V každom riadku/stĺpci/podoblasti sa čísla vyskytujú práve raz (pozri obr. 2).



Obrázok 14.2. (A) Finálna pozícia hry Sudoku je uskutočnená na ortogonálnej mriežke typu 9×9, ktorá je rozdelená na 9 podoblastí typu 3×3. Každá bunka mriežky obsadená nejakým číslom z množiny prvých 9 prirodzených čísel {1,2,…,9}, pričom tieto čísla sú rozmiestnené tak, že v každom riadku/stĺpci/podoblasti sa každé číslo z množiny vyskytuje práve raz. (B) Počiatočná pozícia vznikne z finálnej pozície tak, že odstránime vybrané čísla. Výber čísel nie je ľubovoľný, je ohraničený podmienkou jednoznačnosti, t. j. z počiatočnej pozície musí byť rekonštruovaná pozícia jednoznačne.

Počiatočná pozícia je vytvorená z finálnej pozície tak, že z vybraných buniek sú odstránené čísla, pozri obr. 14.2. Tvorba počiatočnej pozície je ohraničená podmienkou

jednoznačnej rekonštrukcie finálnej pozície z tejto neúplnej pozície. Táto ***podmienka jednoznačnosti*** je silne ohraničujúcou podmienkou pri tvorbe počiatočných pozícií. Vo všeobecnosti môže byť zabezpečená tak [9], že metódou pokusov a omylov vytvárame náhodne takú počiatočnú pozíciu, kde sme schopný metódou spätného prehľadávania vytvoriť len jednu pozíciu, ktorá má obsadené všetky bunky číslami.



Obrázok 14.3. Použitie jednoduchej „logickej“ metódy na riešenie počiatočnej pozície (z obr. 2, diagram B). V každom kroku riešenia sa snažíme vybrať takú prázdnu bunku pozície, ktorej číslo je jednoznačne určené okolím bunky. Určité problémy nastanú vtedy, ak taká prázdna bunka neexistuje, ktorá by mala jednoznačne určené číslo na základe svojho okolia, v tomto prístupe musíme pristúpiť k zložitejším pravidlám určenia prázdnej bunky s jednoznačne určeným číslom. Najhoršom prípade musíme pristúpiť k metóde spätného prehľadávania, kde pre danú bunku máme niekoľko alternatívnych hodnôt. Potom zložitosť hry Sudoku môže byť špecifikovaná zložitosťou stromu riešenia danej počiatočnej pozície. Počet vetvení obsiahnutý v strome riešenia je úmerný ***zložitosti*** počiatočnej pozície. V tomto konkrétnom prípadee hry Sudoku vo všetkých krokoch riešenia sme schopný určiť prázdnu bunku s jednoznačne špecifikovaným číslom. Preto môžeme povedať, že tento prípad hry Sudoku patrí do triedy jednoduchých hier.

Riešenie Sudoku spočíva v tom, že počiatočná pozícia je postupne doplnená na finálnu pozíciu tak, že všetky bunky sú obsadené číslami, ktoré vyhovujú podmienkam hry Sudoku. Tento proces „doplňovania“ počiatočnej pozície môže byť založený na rôznych stratégiách a heuristikách. V najjednoduchších počiatočných pozíciách s úspechom môže byť použitý prístup, ktorý je založený na postupnom výbere takých prázdnych buniek, ktorých obsadenie číslom je jednoznačne určené (pozri obr. 3). V tomto najjednoduchšom prípade obťažnosti hry Sudoku vlastne nepotrebujeme žiadne pravidlá na jej riešenie, ľubovoľná prázdna bunka má určené svojim okolím jednoznačne svoje obsadenie číslom z množiny {1,2,…,9}.

14.3 Pravidlá hry Sudoku

V tejto kapitole naformulujeme základné pravidlá hry Sudoku [2,7] tak, aby mohli slúžiť ako mentálny model tejto hry. To znamená, že tieto pravidlá musia byť dostatočne jednoduché a operabilné, aby boli použiteľné ako mentálny model pre čo najväčšiu množinu ľudí – užívateľov, bez nutnosti špeciálneho vzdelania a/alebo tréningu. Nulté pravidlo špecifikuje počiatočnú fázu riešenia hry Sudoku, keď prázdne bunky sú doplnené množinami čísel obsahujúcimi alternatívne možnosti, ktoré sú určené počiatočnou pozíciou.

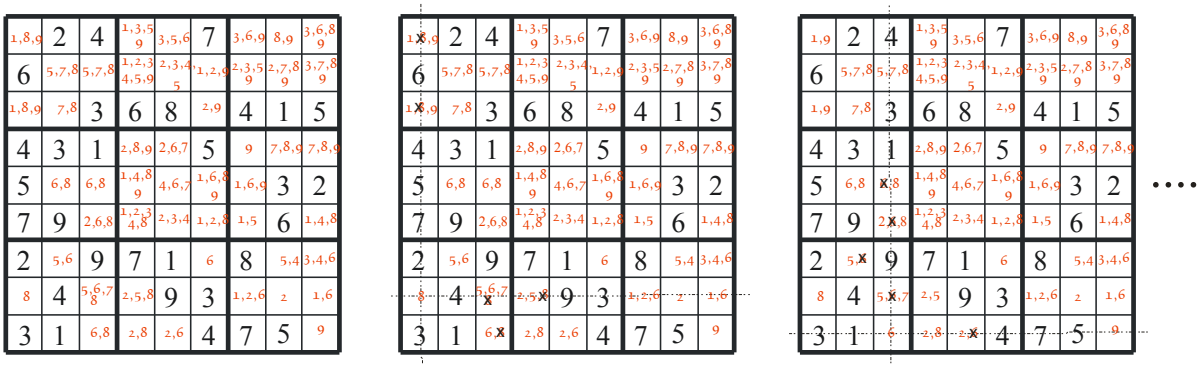
Pravidlo 0. Prázdne bunky v počiatočnej pozícii hry Sudoku vyplním množinami čísel, ktoré sú ***alternatívnymi možnosťami*** zaplnenia danej bunky. Nech bunka *i-j* je prázdna, potom táto bunka je obsadená množinou zostrojenou podľa formuly

$$\{...\}_{ij} = \{1,2,...,9\} - \left(\{r_i\} \cup \{c_j\} \cup \{Q\}\right)$$

kde $\{r_i\}$ je množina čísel – elementov, ktoré sú obsiahnuté v počiatočnej pozícii *i*-teho riadku, $\{c_j\}$ je množina čísel – elementov z *j*-teho stĺpca počiatočnej pozície a $\{Q\}$ je množina čísel – elementov z podoblasti obsahujúcej bunku *i-j* v počiatočnej pozícii

Ako ilustračný príklad aplikácie tohto pravidla ukážeme výpočet alternatívnych možností bunky 1-1 v počiatočnej pozícii znázornené diagramom A na obr. 4.

$$\{1,8,9\}_{11} = \{1,2,...,9\} - \left(\underbrace{\{2,4,7\}}_{\{r_1\}} \cup \underbrace{\{2,3,4,5,6,7,\}}_{\{c_1\}} \cup \underbrace{\{2,3,4,6\}}_{\{Q\}} \right)$$



A *D* *C*

Obrázok 14.4. Postupné riešenie hry Sudoku pre obťažnejšie úrovne hry. (A) Prázdne bunky sú zaplnené alternatívnymi číslami, ktoré sa v danej bunke môžu potenciálne vyskytovať. (B) Pôvodne prázdna bunke 8-1 je obsadená jedným číslom 8, v tomto prípade neexistujú alternatívne možnosti jej obsadenia. Preto toto číslo môžeme považovať za konečné obsadenie danej bunky. Skutočnosť, že bunka 8-1 je obsadená jednoznačne číslom 8, spôsobuje, že toto číslo sa nesmie vyskytovať v 1. stĺpci, 8. riadku a v pridruženej podoblasti tabuľky (ktoré sú vyznačené prerušovanou čiarou), preto v polohách 1-1, 1-3, 9-3 a 9-4 je alternatívna možnosť 8 zrušená prečiarknutím. (C) Tento proces zjednodušovania alternatívnych možností môžeme pre jednoduchšie hry Sudoku opakovať až do úplného vyriešenia hry Sudoku.

Aplikovaním pravidla 0 prázdne bunky sú zaplnené číselnými množinami, ktoré obsahujú alternatívne možnosti pre danú bunku. Najjednoduchšie pravidlo pre zaplnenie prázdnych buniek bude pravidlo, kde pre bunky, ktoré obsahujú množiny alternatívnych čísel s jednotkovou kardinalitou sú zaplnené práve týmto číslom.

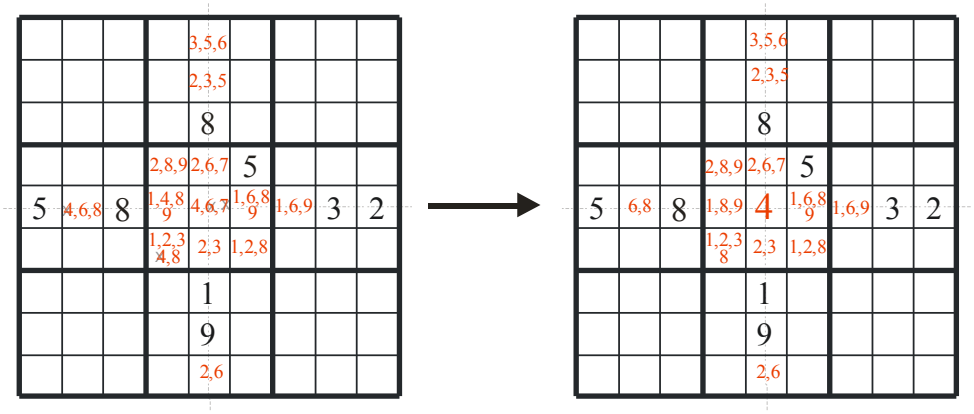
Pravidlo 1. Ak v polohe *i-j* máme jednoprvkovú množinu alternatívnych možností $\{\alpha\}_{ij}$, potom túto polohu zaplníme číslom α a v bunkách na *i*-tom riadku, *j*-tom stĺpci alebo v podoblasti obsahujúcej túto polohu odstránime alternatívnu možnosť α.

Toto jednoduché pravidlo je použiteľné na riešenie celej hry Sudoku len v jednoduchých hrách Sudoku, kde dochádza k efektu znižovania kardinality niektorých množín alternatívnych riešení v priebehu opakovaného použitia pravidla 1. Táto skutočnosť je názorne ilustrovaná na obr. 14.4, kde diagram A vyjadruje počiatočnú pozíciu hry, kde pôvodne prázdne bunky sú zaplnené množinami alternatívnych možností. Diagram B na tomto obrázku znázorňuje výber bunky 8-1, ktorá obsahuje len jednu alternatívnu číslicu 8, ktorá túto bunku aj zaplní. Ako dôsledok tohto aktu je skutočnosť, že v bunkách 1-1, 1-3, 9-3 a 9-4 môžeme z podmnožín alternatív odstrániť číslo 8. Diagram C znázorňuje opakovanú situáciu, keď nová vybraná poloha 9-3 s jednou alternatívou 6 slúži ako „stred“ pre odstránene štyroch alternatívnych možností čísla 6. Tento proces aplikácie pravidla 1 sa opakuje tak dlho, až sa toto pravidlo stane neaplikovateľné, t. j. neexistuje také polohy *i-j* s jednou alternatívou, ktoré by sa taktiež vyskytovali na *i*-tom riadku, *j*-tom stĺpci a v podoblasti, ktorá obsahuje túto polohu.

Pravidlo 2. Ak v polohe *i-j* máme viacprvkovú množinu alternatívnych možností $\{...,\alpha,...\}_{ij}$, pričom číslo α sa v iných bunkách na *i*-tom riadku, alebo *j*-tom stĺpci, alebo

v podoblasti obsahujúcej túto polohu nenachádza, potom danú polohu *i-j* obsadíme jednoprvkovou podmnožinou $\{\alpha\}_{ij}$, t. j. z pôvodnej množiny alternatívnych možností $\{...,\alpha,...\}_{ij}$ sme odstránili čísla všetky ostatné čísla rôzne od α.

Toto pravidlo je ilustrované na obr. 14.5.

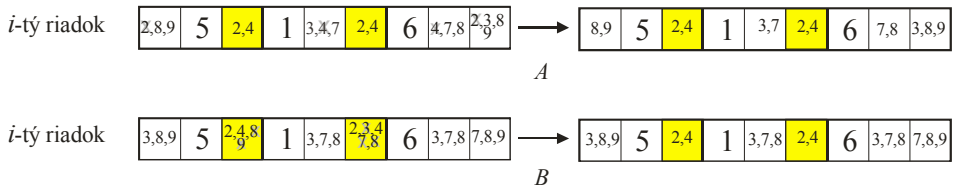


Obrázok 14.5. V ľavom diagrame je vybraná diagonálna bunka 5-5, ktorá obsahuje podmnožinu alternatívnych možností {4,6,7}. Pretože číslo 6 sa nevyskytuje v iných bunkách danej centrálnej podoblasti, z troch alternatívnych možnosti v polohe 5-5 vyberieme číslo 4, ostatné môžeme vylúčiť z množiny alternatívnych možností (pozri pravý diagram).

Týmto sme vyčerpali základné pravidlá pomocou ktorých zjednodušujeme počiatočnú pozíciu hry Sudoku doplnenú o alternatívne možnosti. Aplikovaním týchto pravidiel sme vždy dostali jednoduchšiu pozíciu, ktorá má menej alternatívnych možností a aspoň v jednej bunke alternatívne možnosti boli zredukované na jednu možnosť, ktorá je zafixované pre celú zostávajúcu hru. Ďalšie pravidlá, ktoré budú prezentované len zjednodušujú východiskovú pozíciu bez redukcie podmnožiny alternatívnych možnosti na jeden prvok.

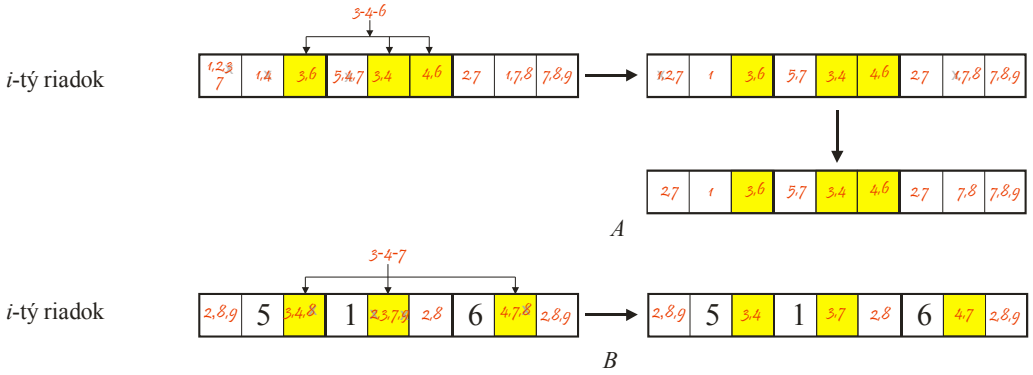
Pravidlo 3.
(***Explicitná dvojica***) Ak v *i*-tom stĺpci alebo v *j*-tom riadku alebo v danej podoblasti existujú také dve bunky, ktoré obsahujú rovnakú podmnožnu alternatívnych možnosti $\{\alpha,\beta\}$, potom čísla α,β môžu byť odstránené vo všetkých iných bunkách, ktoré ležia v danom riadku, stĺpci resp. podoblasti

(***Implicitná dvojica***)
Ak v *i*-tom stĺpci alebo v *j*-tom riadku alebo v danej podoblasti existujú také dve bunky, ktoré obsahujú alternatívne možnosti $\{A,\alpha,\beta,B\}$ a $\{A',\alpha,\beta,B'\}$, kde A, A', B, B' alternatívne možnosti rôzne od α,β, pričom v ostatných polohách v danom riadku, stĺpci resp. podoblasti sa nevyskytujú čísla α,β, potom obe uvedené podmnožiny alternatívnych možností môžu byť zjednodušené na $\{\alpha,\beta\}$.



Obrázok 14.6. (A) Znázornenie explicitnej dvojice , v riadku (alebo stĺpci alebo podoblasti) sú z ostatných buniek odstránené čísla α a β. (B) Znázornenie implicitnej dvojice, kde vybraná dvojica alternatívnych možností obsahuje spoločné čísla α a β, ktoré sa nevyskytujú v ostatných bunkách, potom tieto z týchto dvoch množín môžeme odstrániť prvky odlišné od α a β.

Pravidlo 3 je ilustrované na obr. 14.6. Musíme poznamenať, že toto pravidlo vo forme explicitnej dvojice môže viesť k zjednodušeniu pozície tak, že v niektorej polohe vznikajú alternatívne možnosti s jedným elementom, ktoré sú zafixované do konca riešenie Sudoku. Ak použijeme pravidlo implicitnej dvojice, dochádza k zjednodušeniu hry, kde vznikajú „čisté“ dvojice {α,β} v dvoch rôznych bunkách, ktoré pôvodne obsahovali aj iné irelevantné čísla. Modifikácia pravidla 3 pre explicitnú alebo implicitnú trojicu je znázornená na obr. 14.7. V konštrukcii pravidiel by sme mohli stále pokračovať, boli by stále väčšie a väčšie, pričom pravdepodobnosť ich výskytu by bola menšia a menšia, aj to len pre veľmi obťažné hry.



Obrázok 14.7. (A) Príklad pravidla explicitne trojice pre trojicu čísel 3-4-6, v prvom kroku v ostatných bunkách sú vylúčené čísla z tejto triády. Pretože vzniklo samostatné číslo 1, potom vo všetkých ostatných polohách toto číslo vylúčime. (B) Príklad pravidla implicitnej trojice čísel 3-4-7, pretože tieto čísla sa nachádzajú v 3, 5 a 8 bunke, môžeme v týchto bunkách odstrániť čísla rôzne od danej trojice.

Explicitná n-tica (pre $2 \le n \le 9$) je špecifikovaná [2] pomocou minimálnej množiny $\{k_1, k_2, ..., k_n\} \subseteq \{1, 2, ..., 9\}$ a buniek c_{ij} patriacich do vybraného riadku/stĺpca/oblasti, ktoré obsahujú alternatívne možnosti z podmnožiny $\{k_1, k_2, ..., k_n\}$, t. j. alternatívne možností z bunky c_{ij} sú $\{...\}_{ij} \subseteq \{k_1, k_2, ..., k_n\}$. Vlastnosť minimálnosti množiny $\{k_1, k_2, ..., k_n\}$ spočíva v tom, že nemôžeme z nej odstrániť žiadny element bez toho, aby nezanikla druhá vlastnosť explicitnej *n*-tice, ktorá požaduje, aby alternatívne množiny c bunkách (i_k, j_k) boli podmnožinou množiny $\{k_1, k_2, ..., k_n\}$. Formálne explicitna n-tica je špecifikovaná ako usporiadaná dvojica

$$\left(\underbrace{\{k_1, k_2, ..., k_n\}}_{\text{množina}}, \underbrace{\left\{ \{\ \}_{i_1 j_1}, \{\ \}_{i_2 j_2}, ..., \{\ \}_{i_n j_n} \right\}}_{\substack{\text{množina aletrnatívnych} \\ \text{možností pre vybrané} \\ \text{bunky}}} \right)$$

Pre lepšie pochopenie tejto abstraktnej definície uvedieme explicitnú dvojicu z obr. 14.6, diagram A

$$\left(\{2,4\}, \left\{ \{2,4\}_{i3}, \{2,4\}_{i6} \right\} \right)$$

Podobným spôsobom môžeme špecifikovať aj explicitnú trojicu z obr. 14.7, diagram A

$$\left(\{3,4,6\}, \left\{ \{3,6\}_{i3}, \{3,4\}_{i5}, \{4,6\}_{i6} \right\} \right)$$

Podľa pravidla 3 (pozri obr. 14.6 a 14.7, diagramy B) koncepcia explicitných *n*-tic môže byť zovšeobecnená na implicitné *n*-tice. Implicitná *n*-tica je špecifikovaná podobným spôsobom ako explicitná *n*-tica. Nech bunka (i_k,j_k), pre k=1, 2, ..., n, obsahuje alternatívne možnosti označené $\{\ \}_{i_k j_k} = [\alpha_k, \beta_k, ...]_{i_k j_k} \cup [\![A_k, B_k, ...]\!]_{i_k j_k}$, ktoré sú obsiahnuté v dvoch disjunktných podmnožinách $[...]_{i_k j_k}$ a $[\![...]\!]_{i_k j_k}$. Budeme postulovať, že alternatívne možnosti $\alpha_i, \beta_i, ...$ sa nachádzajú len v bunkách (i_k,j_k), pre k=1, 2, ..., n, potom implicitnú *n*-ticu môžeme zjednodušiť tak, že alternatívne možnosti z druhej podmnožiny $[\![...]\!]_{i_k j_k}$ sú odstránené, t. j.

$$\{\ \}_{i_k j_k} = [\alpha_k, \beta_k, ...]_{i_k j_k}$$

K tomu, aby sme ilustrovali túto relatívne abstraktnú koncepciu implicitnej *n*-tice obrátime našu pozornosť na ilustračné príklady z obr. 14.6 a 14.7, diagramy B. Ako prvý príklad budeme diskutovať obr. 14.6, diagram B

$$\{\ \}_{i3} = [2,4]_{i3} \cup [\![8,9]\!]_{i3}, \ \{\ \}_{i6} = [2,4]_{i6} \cup [\![3,7,8]\!]_{i6}$$

Kde alternatívne možnosti 2,4 sa nachádzajú len v bunkách (*i*,3) a (*i*,6), potom zanedbaním alternatívnych možností z $[\![...]\!]_{ij}$ dostaneme

$$\{\ \}_{i3} = [2,4]_{i3}, \ \{\ \}_{i6} = [2,4]_{i6}$$

To znamená, že v prípade implicitnej dvojice dochádza k podstatnému zjednodušeniu buniek (*i*,3) a (*i*,6), z nich sú odstránené nepodstatné alternatívne možnosti 8,9 resp. 3,7,8.

Podobným spôsobom môžeme skúmať aj implicitnú trojicu špecifikovanú obr. 14.7, diagram B, kde

$$\{\ \}_{i3} = [3,4]_{i3} \cup [\![8]\!]_{i3}, \ \{\ \}_{i5} = [3,7]_{i5} \cup [\![2,9]\!]_{i5}, \ \{\ \}_{i8} = [4,7]_{i8} \cup [\![8]\!]_{i8}$$

Podobným spôsobom, ako v predchádzajúcom ilustračnom príklade, táto implicitná trojica môže byť zjednodušená tak, že alternatívne možnosti z $[\![...]\!]_{ij}$ sa zanedbajú

$$\{\ \}_{i3} = [3,4]_{i3}, \ \{\ \}_{i5} = [3,7]_{i5}, \ \{\ \}_{i8} = [4,7]_{i8}$$

V tomto prípade aktuálne riešenie Sudoku bolo zjednodušené o 4 alternatívne možnosti.

14.4 Stratégia riešenia

Stratégia riešenia hry Sudoku obsahuje postupnosť týchto krokov:

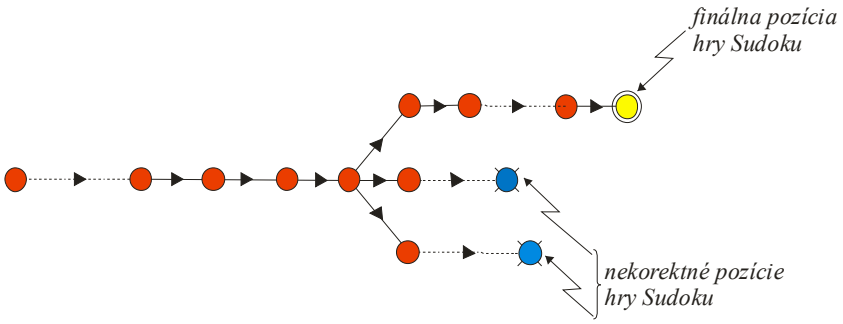
1. krok. Počiatočnú pozíciu doplníme v každej prázdnej bunke o alternatívne možnosti .
2. krok. Ak v alternatívnych možnostiach pre daný riadok, alebo stĺpec, alebo podoblasť sa

vyskytuje nejaké číslo práve raz, potom týmto číslom obsadíme danú bunku.
3. krok. Opakovane aplikujeme na pozíciu pravidlá 1-3 tak, aby sa pozícia neustále zjednodušovala buď zrušením (pravidlá 2 a 3) vybraných alternatívnych možností alebo obsadením vybranej bunky fixným číslom (pravidlo 1).
4. krok. V prípade, že pozícia obsahuje ešte alternatívne možnosti a nie je možné vykonať 3. krok (bod vetvenia, pozri obr. 14.8), náhodne vyberieme z aktuálnej pozície jednu alternatívnu možnosť a s ňou obsadíme danú bunku. Ak sa nám v ďalšom riešení od bodu vetvenia objaví nekorektná pozícia (v niektorom riadku/stĺpci/podoblasti sa vyskytujú dve rovnaké čísla), potom sa vrátime do bodu riešenia, opakovane náhodne vyberieme inú alternatívnu možnosť a opakovane pokračujeme v riešení.
5. krok. Ak sme dospeli v riešení do korektnej konečnej pozície, kde všetky pôvodne prázdne bunky sú obsadené číslami, hru ukončíme a získaná pozícia je riešením hry Sudoku.

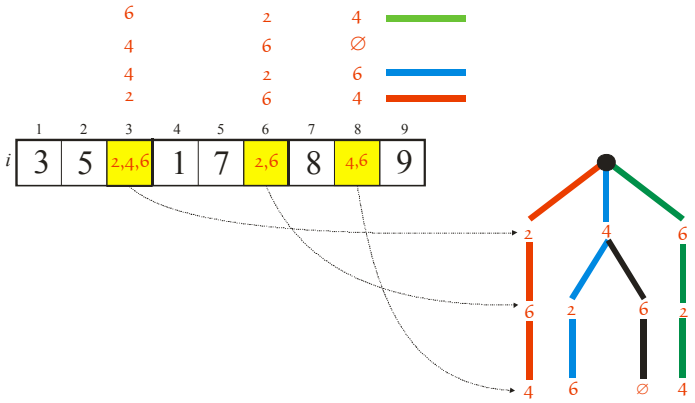
Pomocou tejto stratégie môžeme zaviesť klasifikáciu riešení hry Sudoku:

(a) **Jednoduché riešenie**, kde používame len pravidlá 0 alebo 1. t. j. každom kroku hry sa nám podarí urobiť také zjednodušenie hry, že sa zaplní prázdna pozícia číslom.
(b) **Stredne zložité riešenie**, kde sa okrem pravidiel 1 a 0 používajú taktiež pravidlá 2 a 3.Pre tento stupeň zložitosti hry taktiež dochádza monotónnemu zjednodušovaniu hry, ovšem už určitá časť krokov riešenie (okolo 10-15%) spočíva, že pre vybrané bunky aj keď nedochádza k ich zaplneniu fixovaným číslom, ale dochádza k celkovému znižovaniu počtu alternatívnych pozícií v aktuálnom riešení.
(c) **Veľmi zložité riešenie** (diabolské), kde sa aspoň raz musí použiť náhodný výber alternatívnej možnosti s možným návratom do bodu vetvenia, ak sme dostali nekorektnú pozíciu v ďalšom predlžovaní riešenia (pozri obr. 14.8).

Neexistujú pevné hranice medzi oblasťami zložitosti hier Sudoku, vyššie navrhnutá klasifikácia však odráža základné princípy zložitosti ich riešenia. Pre nás je však dôležité, že navrhnutý mentálny model hry Sudoku je dostatočne jednoduchý k tomu, aby slúžil ako návod k systematickú riešeniu úloh hier Sudoku a taktiež, ktorý taktiež odráža pomerne verne „experimentálne pozorovanie“ o zložitosti ich riešenia.



Obrázok 14.8. Znázornenie priebehu riešenia hry Sudoku, prvý vrchol zľava reprezentuje počiatočnú pozíciu kde prázdne bunky sú zaplnené alternatívnymi možnosťami. Riešenie „monotónne“ pokračuje až do pozície vetvenia (pozri obr. 9) v ktorej už nie je aplikovateľné žiadne pravidlo 1-3. V tomto bude náhodne z danej pozície vyberieme jednu alternatívnu možnosť a použijeme ju ako pokračovanie riešenie. Použitím pravidiel 1-2 predlžujeme riešenie. V prípade, že sa nám objaví nekorektná pozícia (t. j. v danom stĺpci/riadku/podoblasti existujú dve rovnaké čísla), riešenie ukončíme a vrátime sa do bodu riešenie a náhodne vyberieme ďalšiu alternatívnu možnosť a opakovanie predlžujeme riešenie novej pozície. Ak sa nám podarí ukončiť riešenie v korektnej konečnej pozícii, kde všetky alternatívne možnosti sú substituované číslami, potom sme vyriešili počiatočnú pozíciu hry Sudoku tak, že sme zaplnili korektným spôsobom všetky prázdne bunky. Z definičnej podmienky existencie jediného riešenia hry Sudoku vyplýva, že takto získané korektné riešenie je správne riešenie danej hry Sudoku.



Obrázok 14.9. Znázornenie vetvenia pozície hry Sudoku, ktorá v *i*-tom riadku obsahuje trojicu 2-4-6. Predpokladajme, že aktuálna pozícia hry Sudoku je taká, že nemôžeme zjednodušiť tieto alternatívne nožnosti bunkách (*i*,3), (*i*,6) a (*i*,8) , avšak chceme túto trojicu použiť k aktuálnemu riešeniu hry. V pravo dole je znázornený strom riešenia, ktorý jednotlivým bunkám priraďuje jednotlivé čísla považované v nasledujúcom riešení za konštantné. Z tohto stromu riešení vyplýva, že existujú len tri možnosti výberu, ktoré môžu byť v riešení odlíšené farebne. Vyberieme si tú možnosť, ktorá nám umožňuje získať korektné riešenie hry Sudoku. V prípade, že aktuálne riešenie zostrojí nekorektnú situáciu, potom toto riešenie opustíme a použijeme ďalšiu alternatívu zo stromu riešenia.

Závery

Z našich predchádzajúcich úvah vyplýva, že riešenie hry Sudoku môže byť interpretované ako problém deduktívneho usudzovania v rámci klasickej výrokovej logiky (pozri Gabbayovu učebnicu logiky [4], príklad 1.4.11). Táto skutočnosť vyplýva z faktu, že Sudoku má a-priori práve jedno riešenie, ktoré je jednoznačne určené každou aktuálnou korektnou pozíciou vznikajúcou v priebehu riešenia hry Sudoku. Priekopnícky charakter v tomto smere má práca Leeho a kol., kde Sudoku bolo analyzované pomocou pojmového aparátu kognitívnej vedy [8], pričom bola použitá koncepcia mentálneho modelu [5,6,12] k simulovaniu zložitosti (vzhľadom k riešiteľom) hry Sudoku. Títo autori odmietli používanie metódy spätného prehľadávania pre riešenie hry Sudoku, vychádzajú z pracovnej hypotézy, že v každom kroku existuje pravidlo, ktoré zjednodušuje danú pozíciu. Aj keď je pravdepodobne táto hypotéza pravdivá, v záverečných etapách riešenia hry Sudoku môže byť proces nájdenia tohto pravidla neobyčajne komplikovaná. Podľa našich skúseností, podstatne jednoduchšie je aplikovať pre vybranú *n*–ticu alternatívnych možností jej rozklad pomocou spätného prehľadávania (pozri obr. 9); aktuálny rozklad sa ukáže ako nekorektný, keď dostaneme v riadku/stĺpci/podoblasti dve rovnaké čísla, potom musíme vyskúšať inú možnosť rozkladu danej *n*–tice. Z tohto pohľadu, spätné prehľadávanie, aj keď nemusí skracovať celkový čas riešenia, v podstate veľmi zjednodušuje hľadanie vhodnej stratégiu v aktuálnom kroku riešenia hry Sudoku. Treba si uvedomit, že menovite v záverečnej fáze riešenia, hľadať v každom kroku optimálne pravidlo typu 0 – 3, môže byť pomerne zložitý proces.

Cieľom tejto práce bolo navrhnúť mentálny model Sudoku, ktorého výsledky korelujú s experimentálne známymi poznatkami o zložitosti riešenia hry Sudoku. Hry s najnižšou zložitosťou (ľahké) sú charakterizované tým, že sa používajú výlučne len pravidlá 0 a 1, t. j. v každom kroku existuje aspoň jedna bunka, ktorá je obsadená jednou alternatívnou možnosťou. Stredná zložitosť je charakterizovaná využívaním pravidiel 2 a 3, ktoré sú založené na existencii dvojíc alebo trojíc alternatívnych možností, pomocou ktorých sme schopný zjednodušovať aktuálnu pozíciu tak, že vybrané alternatívne možnosti zjednodušujú pozíciu tak, že vybrané iné alternatívne možnosti môžu byť zrušené. Najťažšia zložitosť (diabolská) hry Sudoku už vyžaduje v závere riešenia náhodný výber z vybraných aktuálnych

možností. V tomto prípade môže nastať situácia existencie nekorektnej pozície a preto sa musíme vrátiť k bodu riešenia, kde bol použitý náhodný vyber a zvoliť si inú možnosť obsadenia vybraných buniek číslami. Delenie hier Sudoku na základe používania pravidiel riešenia dobre koreluje s experimentálnou skutočnosťou, kde za najzložitejšie hry sa považujú tie, pre ktoré v určitej fáze hry je potrebné vykonať náhodný výber riešení. Naopak, najjednoduchšie sú tie hry, kde vystačíme s aplikáciou pravidiel 0 a 1, čiže monotónne zjednodušujeme počiatočnú pozíciu.

Predmetom výskumu kognitívnej vedy v tejto oblasti môže byť štúdium štýlu použitia jednotlivých pravidiel, ako sa hľadajú nové možnosti pre zjednodušenie aktuálnej pozície pomocou pravidiel. Táto skutočnosť otvára nové možnosti výskumu v kognitívnej vede v oblasti riešenia problémov, vytvárať mentálne modely používania základného mentálneho modelu, t. j. hľadať pravidlá – regularity pre používania pravidiel 0 – 3.

Poznámka 1: V kapitole 3 boli špecifikované základné princípy evolučných algoritmov, ktoré možno charakterizovať ako stochastické optimalizačné algoritmy akcelerované heuristikou darwinovského prirodzeného výberu. Tieto algoritmy sú aplikovateľné aj na riešenie Sudoku. V dôsledku ich stochastičnosti, zapĺňanie prázdnych buniek v matici sa deje kvázi-náhodne. Ich efektivita riešenia už najjednoduchších Sudoku nie je veľká. Tým, ako bolo ukázané v predchádzajúcej časti tejto publikácie, že obsadzovanie prázdnych buniek je deterministický problém riadený jednoduchými logickými úvahami, použitie evolučných algoritmov (aj keď v princípu možné) nie je vhodným prístupom k riešeniu Sudoku. Ďalší problém s použitím evolučných algoritmov pre riešenie Sudoku spočíva v tom, že požadované riešenie je reprezentované globálnym minimom, iné – suboptimálne riešenia sú nezaujímavé, pričom evolučné algoritmy svoju efektívnosť manifestujú schopnosťou rýchlej konštrukcie práve suboptimálneho riešenia.

Poznámka 2: Chceme upozorniť seniorov informatickej komunity na skutočnosť, že riešenie Sudoku je v podstate deterministický problém riadený jednoduchými logickými úvahami. Metóda riešenia Sudoku prezentovaná v tejto práci má jednoduchý algoritmický charakter a preto je zaujímavá aj pre seniorov ako vhodné „precvičovanie“ kognitívneho aparátu – mozgu (pre ktorý platí to, čo platí pre každý orgán, jeho „neprecvičovanie“ vedie k jeho atrofii).

Literatúra

[1] Craik, K.: *The Nature of Explanation.* Cambridge University Press, Cambridge, 1943.

[2] Crook, J.F.: A Pencil-and-Paper Algorithm for Solving Sudoku Puzzles. *Notices of the AMS* **56** (2009), 460–468.

[3] Delahaye, J.-P.: The sciennce behind Sudoku. *Scientific American.* **294** (2006), 82–87.

[4] Gabbay, D.M.: *Logic for Artificial Intelligence & Information Technology*. King's College Publications , London, 2007.

[5] Johnson-Laird, P.N.: *Mental Models*. Harvard University Press, Cambridge MA, 1983.

[6] P.N. Johnson-Laird,: Mental Models, Deductive Reasoning, and the Brain. In *The Cognitive Neuroscience*, ed. by Gazzaniga, M.S. The MIT Press, Cambridge, MA, 1996.

[7] Kvasnička, V., Pospíchal, J.: Darvinovská evolúcia ako algoritmus. *Československý časopis pro fyziku*, **59** (2009), 360-367.

[8] Lee, N.Y.L, Goodwin, G.P., Johnson-Laird, P.N.: The psychological puzzle of Sudoku. *Thinking & Reasoning* **14** (2008), 342–364.

[9] Mach, M.: Riešenie problémov s ohraničeniami. In *Umelá inteligencia a kognitívna veda II.*, eds. Kvasnicka, V., Pospíchal, J., Návrat, P., Lacko, P., Trebatický, P.: Vydavateľstvo STU, Bratislava, 2010, 291–318.

[10] Návrat, P.: *Umelá inteligencia*. Vydavateľstvo STU, Bratislava, 2006.
[11] Simonis, H.: Sudoku as a Constraint Problem. In Hnich, B., Prosser, P., Smith, B. (eds.): *Proceedings 4th Int. Works. Modelling and Reformulating Constraint Satisfaction Problems*, CP2005, October 1 - 5, 2005, Barcelona, Spain (pp. 13–27).
[12] Stenning, K., van Lambalgen, M.: *Human Reasoning and Cognitive Science*. MIT Press, Cambridge, MA, 2008.
[13] Wilsom, R.A., Kei, F. C.: *The MIT Encyclopedia of the cognitive science*. MIT Press, Cambridge, MA, 1999.

15. prednáška

Teória ohraničenej racionality

15.1 Úvod

Jeden z hlavných cieľov tejto kapitoly je poskytnúť čitateľovi pohľad umelej inteligencie a kognitívnej vedy na problém ľudskej racionality, ktorý patrí medzi základné koncepcie našej anticko-židovsko-kresťanskej civilizácie. Aristoteles v *Etike* špecifikoval ľudské bytosti ako ***racionálne zvieratá***, pretože majú telo s fyzickými potrebami a nutnosťou prijímať potravu a sú racionálnymi bytosťami, pretože majú dušu, ktorej časťou je pamäť a schopnosť myslieť a usudzovať. Cieľom tejto bytosti je dosiahnuť šťastie tak, že napĺňa zmysel svojej existencie. Podobne, ako zmyslom existencie noža je, aby krájal a zmyslom existencie očí je, aby videli, tak zmyslom ľudskej existencie – racionálneho zvieraťa je, aby používalo svoju schopnosť usudzovať, t. j. správať sa racionálne pri vyhodnocovaní svojich životných skúseností. V tomto momente sa dostávame ku kľúčovému momentu gréckej filozofie, ktorá považovala za jeden z podstatných prejavov racionality ľudských bytostí schopnosť usudzovať pomocou zákonov logiky, ktorej základy boli taktiež vytvorené v starom Grécku. Aristotelovská logika, menovite jeho teória sylogizmov, sa stala jednou zo základných súčastí tak starovekého ako aj novovekého vzdelania (napr. na talianskych univerzitách, prvých stredovekých univerzitách v Európe). Tento prístup pretrváva až do súčasnosti, keď nielen na teologických fakultách, ale aj na niektorých filozofických fakultách, jadrom výučby logiky je teória sylogizmov rozšírená o Vennove diagramy.

Pre stredovek a aj počiatok novoveku bola vzorom dokonalosti formálneho systému Euklidova geometria, ktorá je založená na niekoľkých východiskových predpokladoch – axiómach, z ktorých sme schopní vyvodiť pomocou logických zásad usudzovania netriviálne závery – teorémy. Tento systém bol pokladaný za vrchol racionálnosti, keď z niekoľkých poznatkov sme schopní vyvodiť nové netriviálne poznatky. Toto vyvodzovanie spočívalo len vo verbálnych úvahách, matematický formalizmus sa takmer nepoužíval. V 17. storočí zásluhou francúzskeho filozofa a matematika R. Descarta vznikla analytická geometria, kde v ortogonálnom súradnicovom systéme je bod v rovine špecifikovaný usporiadanou dvojicou čísel, priamka v rovine je špecifikovaná lineárnou rovnicou typu $y = kx + q$, a pod. Geometrické úvahy, ktoré boli v klasickom prístupe realizovateľné len použitím jednoduchých pomôcok (papier, ceruzka, pravítko, trojuholník a kružidlo), pomocou metód analytickej geometrie boli transformované na algebraické manipulácie s formulami. Vznik analytickej geometrie umožnil nemeckému filozofovi a matematikovi G. W. Leibnizovi zostrojiť infinitezimálny počet (teraz by sme povedali diferenciálny a integrálny počet, alebo jednoducho kalkulus), pomocou ktorého bol schopný analyticky zostrojiť dotyčnicu ku grafu funkcie.

Zo súčasného pohľadu možno konštatovať, že Leibniz taktiež zohral v logike podobnú úlohu, akú Descartes zohral v geometrii, kde nahradil geometrické konštrukcie matematickými manipuláciami s algebraickými výrazmi. Leibniz sa pokúsil vybudovať formálny systém, ktorý by nahradil verbálne metódy usudzovania manipuláciami s formulami. Postuloval formálny systém s dvoma časťami: (1) jazyk logiky *lingua*

characteristica, pomocou ktorého je možné reprezentovať každý výrok a (2) *calculus ratiocinator*, pomocou ktorého je možné uskutočňovať usudzovanie systematickým a matematicky presným spôsobom. Žiaľ, trvalo ešte ďalších 200 rokov než sa naplnila táto Leibnizova idea, keď v polovici 19 st. anglický matematici A. de Morgan a G. Boole zostrojili „kalkulus“ výrokovej logiky.

Pre Descartovho súčasníka anglického filozofa T. Hobbesa myslenie už bolo len špeciálny druh výpočtu. Táto hypotéza, ktorá v 17. storočí znela veľmi neobvykle ba až exoticky, bola až v súčasnosti akceptovaná v rámci kognitívnej vedy, kde má postavenie centrálnej paradigmy, a realizovaná pomocou umelej inteligencie.

V dobe vrcholného osvietenstva, na prelome 18. a 19. storočia, francúzsky astronóm, matematik a filozof P. R. Laplace vo svojich *Filozofických esejoch* postuluje superinteligentnú bytosť, ktorá je schopná nielen postihnúť všetky sily vládnuce v prírode, ale vykonať aj ich matematickú analýzu. Táto bytosť s istotou vidí svojim „zrakom“ všetky udalosti z minulosti, prítomnosti a aj budúcnosti, pre ňu neexistujú náhodné alebo neurčité udalosti. Ešte pred Laplaceom, anglický filozof John Locke dal do vzájomného kontrastu všemocného Boha s ľuďmi, ktorí žijú v tieni náhodných a nevysvetliteľných udalostí. Laplace sekularizoval tento vzájomný kontrast medzi Bohom a ľuďmi pomocou fiktívnej superinteligencie. Z perspektívy Boha a Laplaceovej superinteligencie, príroda je prísne deterministická (neexistujú náhodné udalosti), ale z perspektívy ľudí, príroda nie je plne poznateľná, je plná neurčitostí a záhad. Aj keď „všemocnosť“ Laplaceovej superinteligencie nie je dosiahnuteľná pre ľubovoľný reálny systém, jej duch a ideál neustále pretrváva vo víziách neohraničenej racionality formulovaných prostredníctvom rôznych matematicko-štatistických teórií (napr. teórie maximalizácie očakávanej užitočnosti), ktoré implicitne predpokladajú neohraničené výpočtové možnosti a prístupu k požadovanej informácii. V našom civilizačnom okruhu je racionalita ľudských bytostí hlboko kultúrne zakódovaná. Od staroveku, cez stredovek až do súčasnosti prostredníctvom renesancie, humanizmu a osvietenstva, predpoklad racionality ľudských bytostí je vždy ich integrálnou časťou, jednotlivé obdobia v našej histórii sa líšia len v posunoch ťažiska zmyslu našej racionality, či slúži pre pochopenie „božích zámerov“, alebo k civilnejším cieľom, akými sú napr. vyhnúť sa predátorovi.

Cieľom tejto kapitoly je venovať sa problému „*ohraničenej racionality*“ v kognitívnej vede [22,27] a v umelej inteligencii [21]. Ako tu už bolo uvedené, história tohto problému môže byť spätne vystopovaná až do čias osvietenstva. Až v druhej polovici 20. storočia sa začal spochybňovať postulát racionality (akoby neohraničenej) v dôsledku hlbšieho pochopenia princípov fungovania ľudskej mysle/mozgu. Rôzne psychologické experimenty jednoznačne poukazovali na kognitívnu ohraničenosť ľudskej mysle, ktorá má veľmi ďaleko od laplaceovskej univerzality a neohraničenosti. Je zásluhou práve kognitívnej psychológie [24,25], že pomocou experimentálnych techník jednoznačne poukázala na ohraničené kognitívne schopnosti mysle. Z tohto pohľadu možno charakterizovať ľudskú myseľ ako nie veľmi dokonalý kognitívny aparát, ktorého impozantné výkony v zovšeobecňovaní a hľadaní riešení sú skôr dôsledkom veľkej systematickosti ľudí pri riešení problémov a v následnej zafixovanosti získaného riešenia v ľudskej pamäti a v „dedení“ týchto poznatkov pomocou kultúrnej vertikálnej a horizontálnej interakcie (výchova a vzdelávanie) medzi príslušníkmi ľudskej populácie[1].

Problém ohraničenej racionality vstúpil do povedomia širokej pospolitosti ekonómov, kognitívnych vedcov, psychológov a informatikov len pred nedávnom, keď bola r. 2002 udelená Nobelova cena za ekonómiu americkému psychológovi Danielovi Kahnemanovi z University of Princeton, za vypracovanie integrovanej psychologickej metodiky pre potreby

[1] V súčasnosti je veľmi populárnou Dawkinsova memetika [7], ktorá rozvíja podobné názory.

ekonomickej vedy ku skúmaniu ľudského usudzovania a rozhodovania za prítomnosti neurčitosti. Tento významný kognitívny psychológ tri desaťročia spolupracoval s Amosom Tverskym (Stanford University, umrel v r. 1996), spolu sa zaslúžili o vznik *behaviorálnej ekonómie,* študovali problémy usudzovania, rozhodovania, výberu a ohraničenej racionality v ekonomických vedách. Ich vedecké výsledky sú rovnako dôležité tak pre psychológiu, ako aj pre ekonómiu.

Naše úvahy o ohraničenej racionalite sú založené na modernej paradigme multiagentového prístupu [13], ktorá v súčasnej umelej inteligencii [21] je chápaná ako abstraktná softwarová entita schopná v danom prostredí kognitívnych aktivít potrebných k naplneniu cieľov, ktoré agenti imanentne musia riešiť k naplneniu „zmyslu" svojej existencie. Agent v umelej inteligencii je virtuálny alebo reálny „robotický systém", ktorý sa pohybuje v danom prostredí a je schopný riešiť kognitívne problémy vynárajúce sa pred ním ako dôsledok dynamicky meniaceho sa prostredia.

V kognitívnej vede pod agentom rozumieme opäť virtuálny alebo reálny „živý systém" s kognitívnym orgánom implementovaným buď na symbolickej úrovni pomocou formálneho systému (napr. epistemickej logiky, [20]) alebo na subsymbolickej úrovni pomocou neurónovej siete [19]. Takto špecifikovaný kognitívny orgán má silne ohraničené schopnosti a možnosti, potom jeho kognitívne aktivity musia byť taktiež silne ohraničené. Používanie multiagentovej paradigmy tak v umelej inteligencii, ako aj v kognitívnej vede má unifikačný účinok na tieto už aj tak blízke oblasti. Jej významová neutrálnosť je výhodou pre obe vedy, čiastočne ich zjednocuje v tom, že umožňuje používať techniky a metódy, ktoré sú spoločné pre obe vedné oblasti.

15.2 Experimentálna behaviorálna ekonomika

Cieľom tejto kapitoly je prezentácia vybraných experimentov D. Kahnemana a A. Tverského [11,12], ktoré ich viedli k rozpracovaniu behaviorálnej ekonomiky a integrovanej psychologickej metodiky pre potreby skúmania ľudského usudzovania a rozhodovania za prítomnosti neurčitosti. Toto štúdium usudzovania, rozhodovania, výberu a ohraničenej racionality už viac ako polstoročie patrí medzi oblasti výskumu kognitívnej psychológie, v súčasnosti je táto problematika intenzívne študovaná v kognitívnej psychológii v úzkej spolupráci s umelou inteligenciou. K interpretácii experimentov sa používajú pomerne sofistikované teoretické prístupy a koncepcie, tak z teórie pravdepodobnosti, ako aj umelej inteligencie. Na rade psychologických experimentov bolo jednoznačne preukázané, že „homo economicus" sa správa pri rozhodovaní spôsobom, ktorý len veľmi ťažko možno označiť za racionálny. Kahneman a Tversky svojimi výsledkami podporili hypotézu o neexistencii racionálneho agenta množstvom experimentálnych pozorovaní, v ďalšej časti tejto kapitoly uvedieme ako príklad jeden typ ich experimentov.

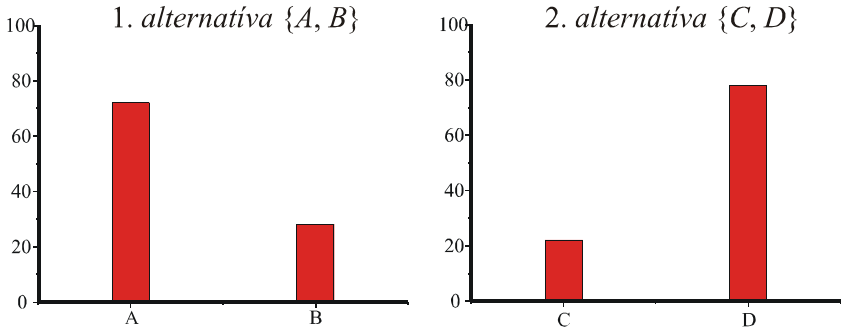
Rámcovanie (framing effects)
Experiment. V USA prepukla epidémia nebezpečnej choroby, na ktorú, podľa odhadu zomrie 600 ľudí. Dve rôzne skupiny respondentov boli požiadané, aby rozlíšili medzi dvoma skupinami dvoch rôznych programov boja proti nemoci. Prvá skupina bola požiadaná, aby rozlíšila medzi dvoma alternatívnymi programami s predpokladanými výsledkami:

A. 200 ľudí bude zachránených.
B. *prob*(bude zachránených 600 ľudí) = 1/3.

Iná skupina respondentov bola požiadaná, aby rozlíšili medzi dvoma inými alternatívnymi programami:

C. 400 ľudí zomrie.
D. *prob*(nikto nezomrie) = 1/3.

Výsledky. V 1. skupine 72% respondentov vybralo alternatívu A z možností $\{A, B\}$ a v druhej skupine 78% respondentov vybralo alternatívu D z možností $\{C, D\}$ (pozri obr. 1)



Obr. 1. Histogramy frekvencie výberu v 1. a 2. alternatíve. Aj keď tieto alternatívy sú ekvivalentné, v 1. alternatíve je preferovaný výber A, zatiaľ čo v 2. alternatíve je preferovaný výber D.

Možné vysvetlenie. Pre existenciu efektu rámcovania je dôležitá verbálna formulácia úlohy, ľudia preferujú pozitívnu formuláciu problému.

Kritika a obrana experimentov Kahnemana a Tverského
Základné tézy kritiky týchto experimentov spočívajú na zdôrazňovaní skutočnosti, že respondenti v experimentoch nemajú dostatočnú motiváciu, aby „dotiahli" experimenty do formy užitočnej pre ekonómiu, prijímaná odmena je malá a respondenti nie sú trénovaní k tomu, aby sa zaoberali riešenými problémami.

Odpoveď na túto kritiku spočíva na týchto skutočnostiach:

(1) Správanie sa respondentov v experimentoch je obvykle založené na dobrej intuícii pochádzajúcej z myšlienkových experimentov.

(2) Mnoho problémov reálneho života riešime za malú odmenu a mnoho našich každodenných rozhodnutí sa uskutočňuje v situáciách, ktoré sa neopakujú.

(3) Ak sú ľudia postavení pred vážne životné rozhodnutia, tieto sú často v silnom konflikte s racionalitou. V literatúre sa často citujú americkí univerzitní profesori, ktorí sa raz za rok musia rozhodnúť akým spôsobom umiestnia svoje investície v rámci federálneho dôchodkového fondu. Mnoho ekonómov označuje tieto rozhodnutia často za „iracionálne", ba až chaotické.

Experimentálne pozorovania Kahnemana, Tverského a ostatných (kognitívnych psychológov) sú často označované za *omyly*, ktoré sa neopakujú, ak si riešiteľ uvedomí ich existenciu. Ekonómia *nemusí študovať* správanie podnikateľov, ktorí sa domnievajú, že 1+1=3, alebo agentov, ktorých správanie je založené na procedúrach zjednodušenia pomocou rámcov. Odpoveď na tieto časté námietky na uvedené experimenty zo strany ekonómov podľa známeho teoretického ekonóma A. Rubinsteina [29] spočíva v tom, že odmieta ich neochotu vyvodzovať všeobecné závery zo správania respondentov v experimentoch, pretože ich označujú za omyl. Označenie tohto správania za omyl neznamená, že toto správanie je

nezaujímavé. Ak na trhu existuje mnoho podnikateľov, ktorí sa domnievajú, že 1+1=3, potom je tento omyl ekonomicky významný a pre teoretikov musí byť zaujímavý.

Skutočnosť, že správanie sa môže zmeniť, ak je podnikateľ informovaný, že jeho správanie je založené na omyle, je zaujímavé, ale nie je významné. V reálnom živote sa situácie „osvietenia" nevyskytujú, pretože agent – identifikátor chýb, ktorý je ochotný upozorniť podnikateľov na chyby, existuje len veľmi vzácne.

15.3 Teória ohraničenej racionality pomocou jednoduchých a rýchlych heuristík

Medzi hlavných proponentov teórie ohraničenej racionality v súčasnosti patrí Gerd Gigerenzer z berlínskeho *Centra pre adaptívne správanie a kogníciu MPI* , ktorý spolu so svojou skupinou v priebehu ostatných 10 rokov publikoval mnoho prác a kníh [9,10] o tejto problematike. Možno konštatovať, že toto Centrum v súčasnosti patrí medzi vedúce pracoviska kognitívnej vedy, ktoré sa zaoberajú výskumom ohraničenej racionality. Základná idea ich prístupu spočíva v postulovaní jednoduchých a rýchlych heuristík, ktoré sú štandardne používané ľuďmi pri riešení zložitých a zle definovaných rozhodovacích problémov.

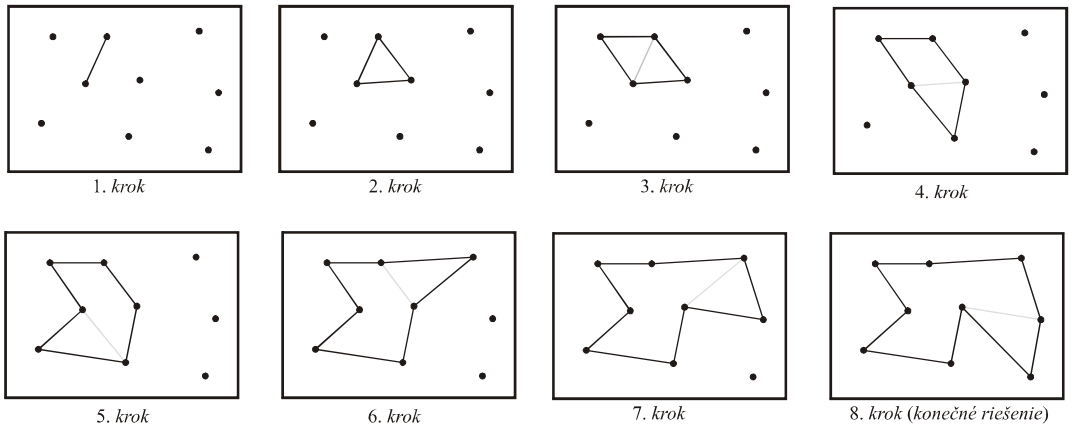
Ilustračný príklad takejto jednoduchej heuristiky, ktorý je často používaný G. Gigerenzerom a jeho spolupracovníkmi je úloha, ktoré mesto, či San Diego alebo San Antonio v USA má viac obyvateľov[2]? Ak nie ste Američan, potom s najväčšou pravdepodobnosťou odhadnete, že San Diego, pretože mnohí z nás vedia, že je to známe univerzitné mesto v Kalifornii, ktoré má nádherné pláže na pobreží Pacifiku. Ak ste Američan, potom máte problém sa rozhodnúť správne, s veľkou pravdepodobnosťou vyberiete náhodne jedno z miest. Použitá heuristika môže byť formulovaná prostredníctvom jednej vety: *Ak poznáte len jedno mesto, potom vyberte za mesto s väčšou populáciou to, ktoré poznáte; ak nepoznáte ani jedno, potom vyberte náhodne mesto s väčšou populáciou.* Ako budeme reagovať, keď poznáme obe mestá? Potom z pamäti si vyberieme dostupnú informáciu o týchto mestách a na základe nej sa rozhodneme, ktoré mesto má väčšiu populáciu. Ak takúto informáciu nemáme v pamäti, potom sa rozhodneme náhodne ako v predchádzajúcom prípade. Táto jednoduchá heuristika bola štatisticky testovaná na mnohých prípadoch nemeckých miest. Simulačné výpočty jednoznačne ukazujú, že napriek jej až extrémnej jednoduchosti je táto heuristika veľmi úspešná, poskytuje v priemere až 95% správnych odpovedí [9]. Taktiež sa porovnávala s inými podstatne zložitejšími a sofistikovanejšími heuristikami, ktoré poskytujú správnu odpoveď s približne rovnakou presnosťou.

Ako ďalší príklad jednoduchej heuristiky, ktorá je schopná rýchlo poskytnúť kvalitné suboptimálne riešenie zložitej kombinatoriálnej optimalizačnej úlohy obchodného cestujúceho uvedieme metódu postupného rozširovania uzavretej cesty pomocou najbližšieho vrcholu. Na obr. 2. je znázornený jednoduchý postup konštrukcie uzavretej cesty (Hamiltonovho cyklu) z počiatočného trojuholníka, ktorý je postupne rozširovaný o ďalšie vrcholy. Najznámejšou heuristikou pre riešenie úlohy obchodného cestujúceho sú evolučné algoritmy, kde použitím pomerne jednoduchých algoritmických prostriedkov sme schopní získať riešenie, ktoré aj pre niekoľko sto vrcholov - miest poskytuje s veľkou frekvenciou optimálne riešenie. Taktiež v rámci operačného výskumu bolo navrhnutých množstvo veľmi

[2] San Diego je v Kalifornii druhé najväčšie mesto s 1.25 mil. obyvateľov, ktoré je známe svojou University of California a plážami na pobreží Pacifiku. San Antonio je piate najväčšie mesto v Texase s 1.14 mil. obyvateľov, ktoré je pokladané za „pravé" hlavné mesto Texasu v dôsledku svojej histórie a významného kultúrneho a ekonomického postavenia. Takže rozhodnúť, ktoré z týchto dvoch miest je väčšie, predstavuje problém hlavne pre Američanov a nie pre cudzincov, ktorí v drvivej väčšine San Antonio nepoznajú a preto sa automaticky rozhodnú pre San Diego.

sofistikovaných heuristík, ktoré sú schopné poskytnúť kvalitné suboptimálne riešenie až pre niekoľko tisíc vrcholov – miest.

Gigerenzer a jeho spolupracovníci podrobne študovali mnoho takýchto jednoduchých heuristík, ukázali ich efektívnosť v rýchlosti hľadania optimálneho riešenia. Aj keď k svojmu „behu“ potrebujú o mnoho menej informácie a výpočtového času, ako štandardné „presné“ algoritmy, ich efektivita a jednoduchosť je pre ne typická. Môžeme si položiť otázku, v čom spočíva tajomstvo ich úspechu? Odpoveď na túto otázku spočíva v „*ekologickej racionalite*". Tieto heuristiky o „ekologickej racionalite“ sú založené na využívaní regularít a iných vlastností prostredia v ktorom daný problém existuje. Tak napríklad, vyššie zmienená heuristika pre nájdenie väčšieho mesta z dvojice miest využíva skutočnosť, že často systematicky ignorujeme mestá, ktoré sú malé a nevýznamné pred väčšími a významnejšími mestami.



Obr. 2. Použitie jednoduchej heuristiky pre konštrukciu cesty obchodného cestujúceho. Algoritmus je inicializovaný trojuholníkom, ktorý je zostrojený z najkratšej hrany doplnenej o najbližší vrchol. V ďalšom kroku nájdeme voľný vrchol x a hranu (y,z) tak, aby veličina $d(x,y)+d(x,z)-d(y,z)$ bola minimálna. Aktuálny cyklus rozšírime o hrany (x,y) a (x,z), pričom pôvodnú hranu (y,z) odstránime. Tento proces opakujeme tak dlho, až každý vrchol je zapojený do vytvoreného cyklu. Naše simulačné výpočty naznačujú, že takto vytvorená uzavretá cesta obchodného cestujúceho poskytuje kvalitné suboptimálne riešenie (ktoré v mnohých prípadoch je totožné s optimálnym riešením).

Výskumy Gigerenzera a jeho spolupracovníkov možno zosumarizovať [9,10] prostredníctvom hypotézy, že ***ľudský mozog bol formovaný v priebehu evolúcie tak, že nie je veľmi efektívnym výpočtovým zariadením, ale je schopný efektívne riešiť neustále sa vyskytujúce adaptívne a rýchlo sa meniace dynamické problémy, od riešenia ktorých často závisí existencia daného jedinca, pričom využíva k riešeniu týchto problémov jednoduché a rýchle heuristiky, ktoré sú nenáročné na čas, poznatky a výpočty***.

Možnosti súčasnej umelej inteligencie a kognitívnej vedy modelovať ohraničenú racionalitu

Integrálnou súčasťou modernej umelej inteligencie je *počítačová inteligencia* (často nazývaná aj ako *softcomputing*, ktorá obsahuje známu triádu moderných subsymbolických prístupov: *neurónové siete* [14], *evolučné algoritmy* [15] a fuzzy logiku [28]. Je potrebné poznamenať, že tieto tri subsymbolické oblasti umelej inteligencie sú založené na princípoch, ktoré sú veľmi blízke k teórii ohraničenej racionality. Neurónové siete sú vhodným prostriedkom na konštrukciu kognitívneho orgánu agentov, ktorý je schopný učenia v priebehu vykonávania svojich aktivít v danom prostredí a ktorý už svojou konštrukciou má ohraničené kognitívne možnosti a preto je vhodným prostriedkom pre štúdium ohraničenej racionality. Podobne,

evolučné algoritmy môžu slúžiť na konštrukcii stratégií v multiagentových systémoch. Klasický príklad tohto prístupu sú Axelrodove štúdie [2,3] kooperácie založené na väzenskej dileme, ktoré vyústili v zistení neočakávanej tit-for-tat stratégie. Fuzzy logika je schopná spracovať neúplnú informáciu (napr. verbálne formulovanú). Pomocou jej metód sme schopní zostrojiť rôzne typy riadiacich prvkov (kontrolérov) len pomocou vágne formulovaných pravidiel (napr. „*zabrzdi, ak sa približuješ k predchádzajúcemu autu*“). Tieto „kontroléry“ môžeme chápať ako jednoduché heuristiky na riadenie zložitých dynamických systémov, ktorých teóriu v mnohých prípadoch nepoznáme (napr. riadenie zložitého zariadenia, kde vzťah medzi vstupnými a výstupnými parametrami poznáme len ako „čiernu skrinku“). Počítačová inteligencia je v súčasnosti veľmi populárna v kruhoch „automatizérov“, kde poskytuje skoro univerzálny prostriedok pre riadenie zložitých systémov, pričom redukuje nutnosť využívať klasické matematické prístupy teórie riadenia na minimum.

Zo „symbolickej“ umelej inteligencie v tejto súvislosti musíme pripomenúť paradigmu „multiagentových systémov“ [13], ktorá sa v súčasnosti stala univerzálnym teoretickým prístupom k riešeniu zložitých problémov umelej inteligencie [21]. Mutiagentové prístupy umožňujú modelovať „sociálne“ javy medzi agentmi, akými sú napr. komunikácia, spolupráca, vznik hierarchických organizačných štruktúr medzi agentmi a pod. V súčasnosti existujú rozsiahle aplikácie teórie multiagentových systémov k simulácii sociálnych systémov a procesov v nich prebiehajúcich [4]. V kombinácii s evolučnými algoritmami (multiagentový systém je chápaný ako populácia agentov, ktoré sa reprodukujú, pričom prežívajú len agenti, ktorí majú vysoké fitnes) môžeme sledovať emergenciu nových vlastností, ktoré v multiagentovom systéme neboli pôvodne prítomné (napr. kooperáciu alebo koherentnú komunikáciu). Kognitívne aktivity agentov sú riadené tzv. kognitívnym orgánom agenta, ktorý môže byť implementovaný buď subsymbolickým/konekcionistickým spôsobom pomocou neurónovej siete, alebo symbolickým/kognitivistickým spôsobom pomocou vhodného logického systému.

Memetika

Pred viac ako 20 rokmi evolučný biológ Richard Dawkins [7] zaviedol pojem „*mém*“ ako informačnú a kultúrnu entitu, ktorá je analógiou génu a je schopná sa replikovať. *Memetika* (veda o mémoch) sa v súčasnosti stala veľmi kontroverznou oblasťou [5], od veľmi kritického a odmietavého postoja až po nekritické akceptovanie ako univerzálneho prístupu k vysvetleniu vzniku a šírenia kultúry. Použitím evolučných multiagentových systémov je možné vysvetľovať vznik kultúry ako jednoduchého spôsobu prekonania informačnej bariéry medzi generáciami. Vychádzame z predpokladu, že fitnes organizmov nie je určené výlučne len jeho genotypom, ktorý špecifikuje nielen telo ale aj štruktúru kognitívneho orgánu agenta, ale aj určitým „balíčkom“ informácie, ktorý obsahuje "kultúrne bohatstvo" predchádzajúcich generácií o tom, ako najlepšie prežiť v danom prostredí.

Pojem *mém* v tejto práci je operacionalizovaný ako informácia, ktorá môže agentovi pomôcť nastaviť jeho kognitívny orgán tak, aby efektívnejšie plnil intencionálne aktivity. V tomto rozšírenom prístupe fitnes genotypu už nie je určený len genotypom samotným, prípadne jeho blízkym okolím, ale aj daným mémom (alebo celým komplexom mémov*),* ktorý je pridružený genotypu. Voľne povedané, mémy môžeme chápať ako *kultúrne prostredie*, ktoré spolu s prirodzeným prostredím tvorí „životný priestor“ pre existenciu genotypov. Agent je reprezentovaný dvojicou ***w***/***m*** (ktorý budeme nazývať *m-genotyp*), kde ***w*** je genotyp špecifikujúci štruktúru jeho kognitívneho orgánu a ***m*** je mém. Kognitívny orgán agenta stojí pred úlohou, aby sa naučil (t.j. zabudoval do svojho kognitívneho orgánu) poznatky reprezentované mémom ***m***. V určitej etape každej generácie dochádza k šíreniu mémov v populácii, ktoré obrazne môže byť charakterizované ako „preskakovanie“

vybraného mému z agenta na iného agenta. Mém interaguje s genotypom nepriamo, prostredníctvom učenia kognitívneho orgánu agentov ovplyvňuje ohodnotenie génov – fitnes. Mémy reprezentujú určitú sumu informácie – vedomostí, ktoré dostane agent – potomok od svojho rodiča.

Aký majú význam mémy pre teóriu ohraničenej racionality? Jeden zo základných postulátov modernej kognitívnej psychológie je, že kognitívne schopnosti ľudskej mysle sú veľmi ohraničené, zďaleka nenapĺňajúce osvietenecké predstavy o našej racionalite. Aj napriek tejto skutočnosti, podľa ktorej ľudskú myseľ možno charakterizovať ako nie veľmi dokonalý kognitívny aparát, jej impozantné výkony v zovšeobecňovaní a hľadaní riešení sú skôr dôsledkom veľkej systematickosti ľudí pri riešení problémov a v následnej zafixovanosti získaného riešenia v ľudskej pamäti alebo pamäťových médiách a v „dedení“ týchto poznatkov pomocou kultúrnej vertikálnej a horizontálnej interakcie (výchova a vzdelávanie) medzi príslušníkmi ľudskej populácie. Mechanizmy tohto kultúrneho prenosu informácie sú študované memetikou [19], ktorá je z týchto dôvodov dôležitou časťou univerzálneho darwinizmu.

Epistemická logika a racionálni agenti

Subsymbolický prístup k implementácii kognitívneho orgánu znamená, že je realizovaný pomocou neurónovej siete, zatiaľ čo symbolický prístup znamená, že kognitívny orgán agenta je realizovaný pomocou formálneho systému, ktorý v mnohých prípadoch je ekvivalentný epistemickej logike[3]. Použitie epistemickej logiky k špecifikácii kognitívneho orgánu agenta *a-priori* zabezpečuje jeho racionálnosť (z pohľadu jej definície), ako kognitívnych aktivít, ktoré sú v súhlase s logikou. Klasická formulácia epistemickej logiky, ako jednoduchej modifikácie modálnej[4] logiky, produkuje *problém logickej omniscientnosti* (*vševedúcnosť*), na ktorý prvýkrát upozornil počiatkom 60. rokov fínsky logik a filozof Jaakko Hintikka a odvtedy je tento problém „spiritus movens“ ďalšieho rozvoja epistemických logík pre potreby teórie multiagentových systémov. Táto skutočnosť existencie potenciálnej logickej omniscientnosti je veľmi zaujímavá pre multiagentové systémy, poskytuje teoretikom neočakávanú možnosť ako modelovať omniscientnosť multiagentových systémov, ktorá môže byť chápaná ako realizácia koncepcie „neohraničenej racionality“ a taktiež aj jej redukciu na „ohraničenú racionalitu“. Týmto sa dostávame za rámec teórie multiagentových systémov, otvára sa možnosť pre umelú inteligenciu a kognitívnu vedu modelovať pomocou multiagentových systémov rôzne verzie racionality, menovite dôležitý prechod od neohraničenej k ohraničenej racionalite.

Obráťme našu pozornosť na špecifikáciu omniscientnosti epistemickej logiky. Skutočnosť, že epistemická logika je obvykle formulovaná pomocou modálnej logiky je v logike chápaná ako pozitívny fakt, ktorý umožňuje preniesť väčšinu teoretických výsledkov

[3] Epistemická logika [20] je určitý typ neklasickej modálnej logiky, ktorá je rozšírením klasického výrokového počtu o dve nové „epistemické“ spojky reprezentované znakmi □ a ◇. Nech φ je formula, potom □φ sa interpretuje „*agent vie, že platí* φ“, podobne, ◇φ sa interpretuje „*agent vie, že možno platí* φ“. Pre formuly, ktoré obsahujú tieto spojky už neplatí *princíp extenzionality*, podľa ktorého pravdivosť formuly je plne určená pravdivostnými hodnotami jej podformúl. Pre sémantickú interpretáciu formuly □φ sa využíva Kripkeho model možných svetov, podľa ktorého je táto formula pravdivá v kripkeovskom svete *w* vtedy a len vtedy, ak je pravdivá v každom svete *w´* dostupnom zo sveta *w*. Podobne, formula ◇φ je pravdivá vo svete *w* vtedy a len vtedy, ak existuje taký svet *w´,* dostupný zo sveta *w*, aby v ňom bola formula φ pravdivá. Medzi spojkami existuje ekvivalencia ◇φ ≡ ¬□¬φ, t. j. epistemickú spojku ◇ môžeme chápať ako špeciálny prípad spojky □. Kripkeovská sémantika poskytuje veľmi flexibilný a efektívny formálny nástroj umožňujúci rôzne jednoduché rozšírenia a modifikácie epistemickej logiky, ktoré zodpovedajú rôznym situáciám v multiagentových systémoch. Možno konštatovať, že v súčasnosti epistemické logiky patria medzi búrlivo sa rozvíjajúce oblasti umelej inteligencie a informatiky, tvoria teoretický základ symbolickej teórie multiagentových systémov.

[4] Modálna logika využíva formalizovaného pojmu možnosti a nutnosti.

modálnej logiky do epistemickej logiky. Žiaľ, tento „zisk“ nie je zadarmo, v teórii multiagentových systémov za túto skutočnosť draho platíme omniscientnosťou epistemickej logiky. To znamená, že kladieme nerealistické nároky na kognitívne orgány agentov, ktoré potom musia mať neohraničené schopnosti usudzovania.

Nech $\boldsymbol{T}$ je množina formúl (teória) epistemickej logiky a nech formula φ logicky vyplýva z tejto teórie, $\boldsymbol{T} \vdash \varphi$. Potom, ak agent pozná formule z teórie $\boldsymbol{T}$, pozná nielen formulu φ ale aj všetky ostatné formuly logicky vyplývajúce z $\boldsymbol{T}$. Opakovaním použitím tejto vlastnosti dostaneme, že agent pozná nielen logické dôsledky z $\boldsymbol{T}$, ale aj logické dôsledky z jej logických dôsledkov, atď. Z týchto jednoduchých úvah vyplýva, že ak kognitívny orgán agenta je implementovaný pomocou epistemickej logiky, potom implicitne predpokladáme, že tento orgán má neohraničené schopnosti usudzovania. Táto skutočnosť je v kontradikcii s našim pozorovaním, že reálni agenti majú kognitívne orgány len s ohraničenými možnosťami usudzovania. Agenti sú obvykle schopní odvodiť jednoduché formuly ako logické dôsledky teórie, ktorú si agent uvedomuje. Implementácia kognitívneho orgánu agenta pomocou epistemickej logiky má neželateľný vedľajší efekt spočívajúci v tom, že agenti nemôžu odlíšiť to čo explicitne (alebo vedome) poznajú, od toho, čo môžu implicitne (potenciálne) poznať.

Môžeme si položiť dôležitú otázku, ako modifikovať epistemickú logiku tak, aby logická omniscientnosť multiagentových systémov bola aj keď nie úplne, tak aspoň čiastočne potlačená? Naznačíme niekoľko možných prístupov k odstráneniu tohto problému:

(1) *Syntaktická metóda*, jej základné princípy spočívajú v skutočnosti, že špecifikácia axiomatického systému epistemickej logiky sa zoslabuje modifikáciou alebo priamo vylúčením niektorých axióm (obsahujúcich modálne spojky) a/alebo pravidiel odvodenia. Extrémny prípad tohto prístupe je, že množina povolených formúl-schém[5] je priamo taxatívne určená. Aj keď tento prístup je neobyčajne efektívny v odstránení logickej omniscientnosti, jeho nevýhodou je, že vytvorený epistemický logický neobsahuje formálne vlastnosti – regularity, ktoré by uľahčovali jeho teoretickú analýzu.

(2) *Neštandardné logiky*, v rámci ktorých je epistemická logika vybudovaná nie na klasickej výrokovej logike, ale na vybranej neštandardnej logike (napr. intuicionistickej logike alebo 4-hodnotovej logike podľa Belnapa). V týchto logikách je implikácia reinterpretovaná tak, aby sa nevyskytovali neželateľné problémy so štandardnou tzv. materiálnou implikáciou. Tak napríklad, v štandardnej logike je tautológiou formula $(p \wedge \neg p) \Rightarrow q$, podľa ktorej z nepravdy vyplýva ľubovoľný výrok; avšak v neštandardných logikách (napr. intuicionistickej) táto vlastnosť už neplatí. Logická omniscientnosť je v tomto prípade redukovaná na fakt, že agenti poznajú len tautológie neštandardnej logiky, ktorých môže byť podstatne menej ako klasickej výrokovej logiky.

(3) *Vigilantnosť* (uvedomelosť, ostražitosť). Tento prístup je modifikáciou Kripkeho sémantického prístupu k interpretácii epistemických formúl (pozri poznámku pod čiarou [7]). Potom formula $\Box\varphi$ pravdivá v kripkeovskom svete w musí byť nielen pravdivá v každom svete w' dostupnom zo sveta w, ale formula φ musí taktiež byť vo svete w vigilantnou, $\varphi \in V(w)$. Takto modifikovaná epistemická logika s vigilantnou sémantickou interpretáciou už neobsahuje omniscientnosť, táto je odstránená jednoduchou klasifikáciou formúl na vigilantné a nevigilatné vzhľadom k danému svetu w, $\varphi \in V(w)$ resp. $\varphi \notin V(w)$.

[5] Schéma je prevzatý pojem z výrokovej logiky, ktorý špecifikuje množinu formúl vzniknutých substitúciou z jednoduchej formuly epistemickej logiky. Tak napríklad, z tautológie $p \vee \neg p$ substitúciou vzniknú formuly tvoriace množinu $\{(p \Rightarrow q) \vee \neg(p \Rightarrow q), \Box p \vee \neg \Box p, \ldots\}$. Táto množina – schéma je označená symbolom $\varphi \vee \neg\varphi$, kde φ je metasymbol vyjadrujúci ľubovoľnú formulu epistemickej logiky.

15.4 Ilustračné príklady jednoduchých heuristík pre ohraničenú racionalitu.

Problém návštevnosti baru „El Farol“ v Santa Fe

V súčasnej informatickej literatúre zameranej na štúdium riadenia zložitých ekonomických systémov, pre ktoré neexistuje matematicky model (alebo ak existuje, tak je veľmi nedokonalý), je často skúmaný problém riadenia týchto systémov pomocou *stratégie*. Stratégia je súbor heuristických pravidiel, pomocou ktorých riadime zložitý systém na základe jeho predchádzajúcich stavov a výstupov. V tejto podkapitole budeme ilustrovať tento prístup pomocou úlohy nazývanej „*El Farol Bar*“, ktorú navrhol americký teoretický ekonóm W. B. Arthur [1,17] ako ilustráciu rozhodovania s obmedzenou racionalitou:

> Bar[6] “El Farol” v Santa Fe každý štvrtok večer uvádza írsku hudbu. Do baru sa zmestí asi 60 ľudí, pričom počet celkových záujemcov je okolo 100 ľudí. Na to, aby sa poslucháči štvrtkovému predstavenia vyhli návalu v baru, každý si hľadá svoju stratégiu (heuristiku), ako na základe predošlých predstavení odhadnúť návštevnosť aktuálneho predstavenia. Poslucháči medzi sebou nekomunikujú o tom, či navštívia, alebo nenavštívia bar.

Pokúsme sa túto úlohu formulovať presnejšie. Nech časový rad štvrtkovej návštevnosti baru El Farol má tvar: ..... $x_{t-k}, x_{t-k+1}, ..., x_{t-1}$; kde posledné číslo $x_{t\text{-}1}$ špecifikuje počet návštevníkov predstavenia ostatný štvrtok. Stojíme pred úlohou, ako odhadnúť dnešnú návštevnosť (x_t) pomocou tohto časového radu.

Predvídanie aktuálnej hodnoty časového radu na základe „okna do minulosti“, sa uskutočňuje pomocou stratégie špecifikovanej binárnym vektorom

$$\boldsymbol{s} = (s_1, s_2, ..., s_8) \in \{0,1\}^8$$

kde jej jednotlivé zložky sú špecifikované tabuľkou

#	*história*	#	*história*
s_1	(UUU)	s_5	(CUU)
s_2	(UUC)	s_6	(CUC)
s_3	(UCU)	s_7	(CCU)
s_4	(UCC)	s_8	(CCC)

Symbol C (U) znamená – *crowded – preplnený* (*uncrowded – nepreplnený*). Tak napríklad: história (UUC) je v čase t interpretovaná tak, že pre predchádzajúce tri časové okamžiky (soboty) platí:

$$x_{t-3} \le 60\,(U),\ x_{t-2} \le 60\,(U),\ x_{t-1} > 60\,(C)$$

Jednotlivé zložky binárneho vektora stratégie $\boldsymbol{s}$ sú opísané takto (pozri hore uvedenú tabuľku):

$s_i = 0 \Rightarrow$ na základe i-tej histórie agent ***nenavštívi*** predstavenie
$s_i = 1 \Rightarrow$ na základe i-tej histórie agent ***navštívi*** predstavenie

Na lepšie pochopenie tohto spôsobu kódovania stratégie rozoberme binárny vektor $\boldsymbol{s}$ = (1,1,0,1,1,0,0,1), jednotlivé zložky stratégie majú tento význam:

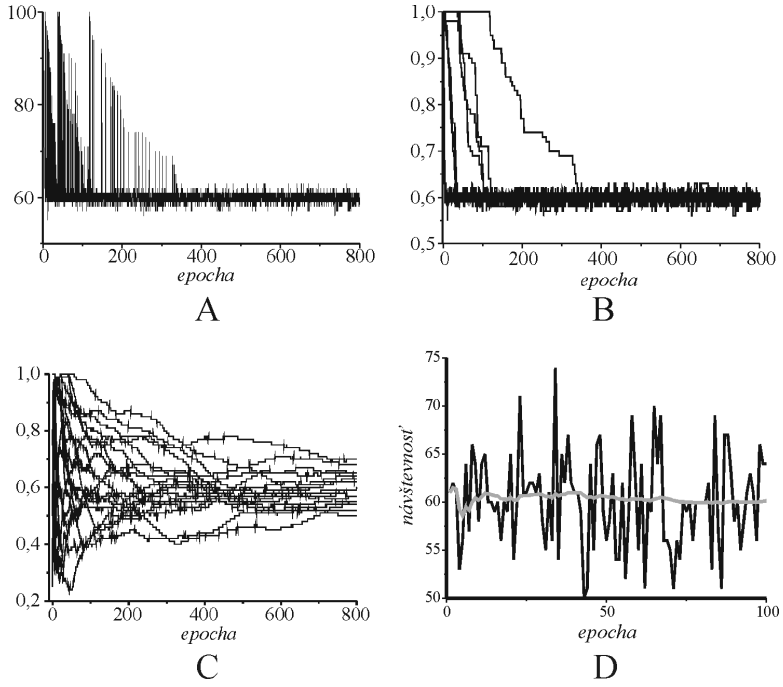
[6] Bar El Farol skutočne existuje v meste Santa Fe, štát Nové Mexiko, USA, dokonca existuje jeho domovská stránka na internete.

(1) $s_1 = 1$, história (UUU) $\Rightarrow$ agent navštívi bar,
(2) $s_2 = 1$, história (UUC) $\Rightarrow$ agent navštívi bar,
(3) $s_3 = 0$, história (UCU) $\Rightarrow$ agent nenavštívi bar,
(4) $s_4 = 1$, história (UCC) $\Rightarrow$ agent navštívi bar,
(5) $s_5 = 1$, históriu (CUU) $\Rightarrow$ agent navštívi bar,
(6) $s_6 = 0$, história (CUC) $\Rightarrow$ agent nenavštívi bar,
(7) $s_7 = 0$, história (CCU) $\Rightarrow$ agent nenavštívi bar,
(8) $s_8 = 1$, história (CCC) $\Rightarrow$ agent navštívi bar.

Nech $P_t = \{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{100}\}$ je populácia stratégií všetkých 100 agentov v epoche *t*. Táto množina nemôže byť homogénna (obsahujúca len jeden druh stratégie), ak by taká bola, potom *všetci* agenti by sa na základe predchádzajúcej histórie rozhodli buď ísť, alebo neisť v štvrtok podvečer do baru El Farol na predstavenie. Z týchto dôvodov každý agent musí mať svoju vlastnú stratégiu $\boldsymbol{s}$, podľa ktorej sa rozhoduje, či ísť, alebo neisť do baru. Jedna epocha v simulácii znamená štvrtkový podvečer, keď sa každý agent populácie musí rozhodnúť, či ísť alebo neisť na predstavenie do baru El Farol. Aby v populácii samovoľne vznikol stav, kde približne 60 agentov navštívi každú sobotu bar, musíme pripustiť proces učenia sa stratégií. Na záver každej epochy agenti vyhodnocujú svoju úspešnosť alebo neúspešnosť pomocou procesu učenia.

Tento proces spočíva v tom, že ak sa daný agent *rozhodol zle* na základe *i*-tej zložky svojej stratégie $\boldsymbol{s} = (..., s_i, ...)$ (t. j. podľa s_i mal ísť, ale bolo obsadené, alebo nemal ísť a bolo neobsadené), tak potom s malou pravdepodobnosťou ($P_{learn} = 0.01$) zmení túto zložku na jej komplementárnu hodnotu, $s_i - 1 \rightarrow s_i$. Epochy sa opakovali 800-krát, pričom bolo pozorované, že v populácii v priebehu asi 300 epoch samovoľne vzniknú také stratégie, že priemerná návštevnosť bola 60 ± 2, pozri obr. 3.

Ako interpretovať tieto výsledky? Problém baru El Farol je typický reprezentant takých stochastických problémov, ktoré nemajú deterministický model. Vo všeobecnosti: jedna z najjednoduchších stratégií je taká, že každý agent si v sobotu podvečer náhodne vygeneruje číslo s rovnomernou distribúciou $0 \le rand \le 1$, ak toto číslo je menšie ako 0.6, potom navštívi bar, v opačnom prípade, ak číslo je väčšie ako 0.6), potom bar nenavštívi. Samozrejme, dlhodobý priemer návštevností bude konvergovať k 0.6 (čo je očakávaný výsledok), avšak s veľkými fluktuáciami (okolo ± 10), pozri diagram D, obr. 3. Iný, „jemnejší“, prístup k riešeniu problému návštevnosti baru El Farol je ponechať agentom možnosť rozhodovať sa pomocou svojich stratégií, ktoré nie sú fixné – nemenné, ale sa menia na základe skúseností jednotlivých agentov.



Obrázok 3. (A) Priebeh návštevnosti baru El Farol pre jednotlivé epochy. Algoritmus bol iniciovaný tak, že každý agent mal stratégiu (11111111), t. j. za každých podmienok vo štvrtok navštívil bar. Proces učenia sa stratégií umožnil vznik novým stratégiám, ktoré už poskytovali priemernú návštevnosť okolo 60 hostí (čo je optimálny výsledok). (B) Frakcie používania jednotlivých histórií v stratégiách celej populácie. Z grafu vyplýva, že sa ustálil „rovnovážny" stav, kde každá história (UUU, ..., CCC) sa využíva približné na 60%. (C) Znázornenie priebehov frekvencie návštev jednotlivých agentov. Priemerná návštevnosť je približne 60%, avšak existujú agenti, ktorí na základe svojej stratégie navštevujú bar častejšie (zriedkavejšie). (D) Znázornenie priebehu návštevnosti baru, ak sa návštevníci rozhodujú tak, že si vygenerujú náhodné číslo $0 \le rand \le 1$, ak toto číslo spĺňa (nespĺňa) podmienku $rand \le 0.6$, tak bar navštívi (nenavštívi). Výsledkom tohto jednoduchého rozhodovacieho procesu je, že priemerná návštevnosť rýchlo konverguje k 60 %, avšak na rozdiel od rozhodovania so stratégiou (pozri diagramy A a B) sú v tomto prípade veľmi veľké fluktuácie.

Tento proces modifikácie stratégií na základe predchádzajúcich skúseností môžeme chápať ako proces učenia sa agentov, v ktorom sa snažia pomocou indukcie a abdukcie vytvárať hypotézy, ktoré stále lepšie a lepšie predpovedajú návštevnosť baru v aktuálnom štvrtkovom podvečere. Samozrejme, toto „vynáranie" sa (emergencia) optimálnych stratégií v spoločenstve všetkých 100 agentov je možné len vtedy, ak každý agent sa správa „racionálne", t. j. ak problém štvrtkovej návštevy baru El Farol rieši pomocou nejakého svojho vnútorného modelu, ktorý nie je pevne daný, ale mení sa v priebehu agentovej histórie na základe jeho úspešnosti. To znamená, že ide o veľmi zložitý dynamický riadiaci problém, v ktorom sa súčasne rieši nielen adaptácia vnútorného modelu každého agenta, ale aj prispôsobenie tohto modelu zmenám vnútorných modelov ostatných agentov (táto skutočnosť sa v umelej inteligencii nazýva „problém pohyblivého cieľa", kde cieľ rozhodovania nie je fixne daný, ale je premenlivý v čase). Problému riadenia zložitých stochastických systémov, kde každý agent musí riešiť problém rozhodovania v situáciách, ktoré nie sú jasne opísané a ktoré závisia aj od rozhodnutí ostatných agentov, patrí medzi extrémne zložité problémy, kde je minimálna šanca ich teoretického spracovania. Ako bolo ukázané na základe jednoduchých ilustračných výpočtov, agenti sa nesprávajú „racionálne" (t. j. nemaximalizujú svoje zisky), ale len s „ohraničenou racionálnosťou", keď si pomocou učenia vytvárajú stratégie (heuristiky) rozhodovania, ktoré s pomerne dobrou presnosťou riešia daný rozhodovací problém.

Problém kooperácie a sociálne dilemy

Ako vyplýva z teórie hier v ekonómii, aplikovanej tiež často na analýzu pretekov v **zbrojení medzi štátmi**, kooperácia sa môže vynoriť zo sebeckého záujmu, keď je výhodná pre obidvoch. Vo všeobecnosti, keď hľadáme racionálne dôvody, či pri našej interakcii s niekým iným (s našim obchodným alebo iným partnerom) pristúpiť na kooperáciu, alebo nie, naše rozhodnutie by malo byť založené na racionálnych argumentoch a nie na predsudkoch a ideologických dôvodoch.

Upriamme našu pozornosť na jeden z možných prístupov k vysvetleniu vzniku kooperácie, ktorý je založený na simulačných výpočtoch pomocou hry nazývanej „***väzňova dilema***“ [2,3,4,6]. “Dilema”, pred ktorou stoja obaja väzni, spočíva v tom, že nech sa druhý z nich rozhodne akokoľvek, vždy je lepšie nemlčať a priznať sa, než nepriznať svoju vinu, lebo keď sa potom prizná komplic, celá vina padne na mňa. Táto dilema ilustruje konflikt medzi individuálnou a skupinovou racionálnosťou. Skupina ktorej členovia preferujú svoj vlastný záujem obvykle skončí horšie, ako tá skupina, kde jej členovia uprednostňujú skupinové záujmy. Vo všeobecnosti, ak záujmy jednotlivcov reprezentujú záujmy celej skupiny, potom takáto skupina je úspešnejšia ako skupina, ktorej jednotliví členovia uprednostňujú vlastné záujmy a ciele. Tento problém bol navrhnutý a široko diskutovaný už počiatkom 50-tych rokov minulého storočia americkými ekonómami Merrilom Floodom a Melvinom Dresherom. Väzňova dilema sa stala často používaným prístupom k vysvetleniu vzniku kooperácie v rôznych sociálnych, ekonomických a iných systémoch, kde sa pomocou matematickej teórie hier a/alebo počítačových simulácií hľadajú „racionálne dôvody“ k tomu, prečo a za akých okolností je výhodnejšie spolupracovať, než ako nespolupracovať.

"Väzňova dilema" **(**VD) je hrou medzi dvoma hráčmi, kde každý z nich má výber z dvoch možností: spolupracovať (cooperate – *C*) alebo podvádzať (defect – *D*). Matica platieb tejto hry je daná takto

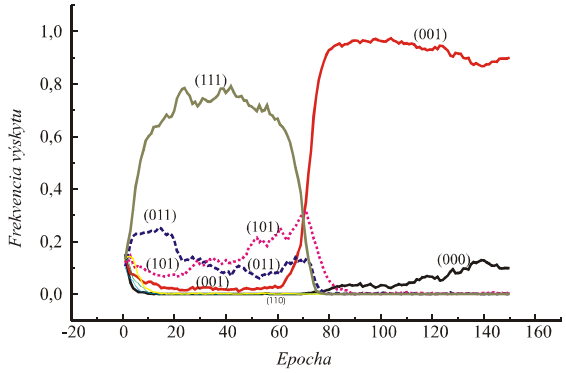
$$f = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \Leftrightarrow \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$$

Keď oba hráči spolupracujú, potom dostanú „odmenu“ *R* (reward), ktorá by mala byť väčšia ako bodový zisk „trest“ *P* (punishment) získaný obidvoma hráčmi, keď podvádzajú. Ale keď jeden hráč podvádza (hrá *D*) zatiaľ čo ten druhý spolupracuje (hrá C), potom podvádzajúci získa výplatu bodov *T* (temptation - pokušenia) ktorá je väčšia ako *R*, zatiaľ čo spolupracujúci dostáva počet bodov *S* (sucker – ten, čo naletel) ktorá je menšia ako *P*.

Uvažujme hráča – agenta *a*, jeho stratégia *s*(*a*) je súbor pravidiel, ktoré použije pri výbere ťahu *C* alebo *D* (spolupráce alebo podvádzania) proti druhému hráčovi, na základe predošlej histórie hry. Pri iba jednej iterácii by samozrejme bolo výhodné zradiť pri polovičnej šanci stretnúť spolupracujúceho, no keď sa iterácie s tým istým spoluväzňom opakujú a vopred nie je známy počet iterácií, metóda ako sa zachovať nie je jasná. Hra sa opakuje t_{max}-krát (dôležitý parameter hry), na záver hry sa spočítajú platby jednotlivých hráčov. Stratégia je určená pomocou 3-bitového vektora (t.j. hráči si pamätajú predchádzajúci ťah oponenta) $\boldsymbol{s}=\{s_1, s_2, s_3\} \in \{0,1\}^3$, kde jednotlivé zložky (0 – kooperácie, 1 – nekoooperácia) majú túto interpretáciu:

(1) s_1 určuje prvý ťah hráča: ak $s_1=0$, potom hráč kooperuje, v opačnom prípade, ak $s_1=1$, nekooperuje.
(2) s_2 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku kooperoval: ak $s_2=0$, potom hráč kooperuje, v opačnom prípade, ak $s_2=1$, nekooperuje.
(3) s_3 určuje ťah hráča v prípade, že druhý hráč v predchádzajúcom kroku nekooperoval: ak $s_3=0$, potom hráč kooperuje, v opačnom prípade, ak $s_3=1$, nekooperuje.

Takáto stratégia vyžaduje len krátkodobú pamäť hráčov, požaduje len znalosť posledného kroku protivníka - druhého hráča. Celkový počet rôznych 1-stratégií je 2^3=8. Keby si väzni mali pamätať viac krokov do minulosti, s každým takým krokom by exponenciálne narastala zložitosť zápisu stratégie.



Obrázok 4. Výsledky evolučného algoritmu použitého pre emergenciu výslednej stratégie, pričom populácia stratégií bola inicializovaná náhodne. V priebehu evolúcie algoritmu spontánne emergovali dve „kooperatívne" stratégie (001) a (000), ktoré zahajujú hru kooperáciou a na kooperáciu súpera v predchádzajúcom ťahu odpovedajú kooperáciou v nasledujúcom ťahu. V dôsledku toho, že väčšinová stratégia (001) neprodukuje „podvod", menšinová stratégia (000) nemá príležitosť reagovať spoluprácou na podvod.

Pri plánovaní stratégie na základe predchádzajúcich skúseností s protihráčom sa často používa evolučný algoritmus. Táto aplikácia evolučného algoritmu prvýkrát vznikla zo spolupráce amerického politológa Roberta Axelroda [2,3,4] s informatikom a tvorcom genetických algoritmov Johnom Hollandom a rozvíjala sa ďalej pre zložitejšie modely medzinárodnej bezpečnosti odpovedajúce napr. na otázku, koľko z rozpočtu vydávať na obranu. V časovom priebehu je zaujímavé sledovať vývoj kooperatívnych a nekooperatívnych stratégií. Kooperatívne stratégie sa spočiatku ukázali menej výhodné, no v neskorších štádiách prevážili.

Numerické výsledky evolučného algoritmu sú znázornené na obr. 4. Ako výsledná stratégia emergujú dve podobné stratégie (001) a (000), ktoré obe zahajajú hru kooperáciou a na kooperáciu odpovedajú kooperáciou. Líšia sa v reakcii na podvod, prvá stratégia (001) reaguje na podvod podvodom, zatiaľ čo druhá stratégia (000) „hlúpo" reaguje na podvod kooperáciou. Stratégia (001) sa nazýva podľa Axelroda tit-for-tat (vo volnom preklade, ako ty mne, tak ja tebe). Menšinová stratégia „dobrák" (000) má šancu prežívať iba popri stratégii tit-for-tat.

Výsledky týchto simulačných výsledkov môžeme zosumarizovať tak, že evolučný návrh heuristík, ktoré optimalizujú dlhodobý zisk pri opakovanej interakcii medzi agentmi, viedol k návrhu pomerne jednoduchej stratégie – heuristiky *tit-for-tat*, ktorá má širšiu platnosť aj ako „morálna heuristika", návod pre dlhodobú opakujúcu sa kooperatívnu interakciu medzi agentmi. Tento prístup bol zovšeobecnený aj pre štúdium etnických konfliktov [16] medzi dvoma etnicky rôznymi populáciami. Získané výsledky racionalizujú koncepciu „kolektívnej viny", okolo ktorej v literatúre existuje množstvo navzájom protirečivých názorov, od zásadného odmietania až po jej schvaľovanie vo zvláštnych prípadoch. Pomocou simulačných výpočtov bolo dokázané, že princíp kolektívnej viny má fatálne následky na kooperáciu medzi rôznymi etnikami, ktorá nemôže vzniknúť, pretože príslušník inej etnickej skupiny je automaticky penalizovaný princípom kolektívnej viny.

Na záver tejto kapitoly môžeme konštatovať, že moderné metódy umelej inteligencie sú schopné študovať vynáranie (emergenciu) jednoduchých heuristických stratégií [8], ktoré sú schopné efektívne riešiť zložité problémy. Táto skutočnosť otvára nové netušené možnosti

pre informatiku, ako pomocou počítačových simulácií študovať riešenie zložitých problémov rozhodovania a dynamických systémov. Ukazuje sa, že študované problémy sú približne rovnako kvalitne riešiteľné mnohými rôznymi heuristickými stratégiami, t. j. že existuje mnoho približných stratégií, ktoré sú schopné efektívne rieši daný zložitý problém. Táto nejednoznačnosť výberu približnej heuristiky má veľký význam pre rôzne multiagentové systémy v tom, že rôzni agenti môžu využívať rôzne heuristiky, čiže môže dochádzať k „sociálnym" javom, ktoré súvisia s problémom výberu vhodnej heuristiky a jej kultúrneho prenosu do nasledujúcej generácie agentov, a pod.

15.5 Otvorené problémy ohraničenej racionality – výzva pre umelú inteligenciu a kognitívnu vedu

Jedna zo základných paradigiem súčasnej umelej inteligencie (a taktiež aj kognitívnej vedy) je *multiagentový prístup,* ktorý obsahuje množinu agentov s ohraničeným kognitívnym orgánom, pričom ich prístup k výpočtovým prostriedkom a k informáciám je taktiež ohraničený. Z tejto špecifikácie agenta vyplýva, že jeho schopnosť používať matematicko-logické metódy pri rozhodovaní sú ohraničené, preto môžeme povedať, že je *ohraničene racionálny*.

1. Otázka ohraničenej racionality agentov v umelej inteligencii nepredstavuje problém, obrazne povedané, *agenti s ohraničenou veľkosťou pamäti môžu byť už z princípu len ohraničene racionálni*. Hovoriť o ich racionalite tak, že implicitne predpokladáme jej neohraničenosť patrí do sci-fi alebo fantasy literatúry.

2. V ekonómii je zrejme situácia iná, tam z historických dôvodov *dichotómia medzi racionalitou a ohraničenou racionalitou neustále pretrváva* a je neustálym námetom mnohých sofistikovaných a filozofujúco-špekulatívnych publikácií a prednášok o ohraničenej racionalite.

3. V kognitívne vede (menovite v kognitívnej psychológii) štúdium *používania jednoduchých heuristík* pri riešení zložitých problémov reprezentuje veľmi plodný smer (G. Gigerenzer v Berlíne na MPI for Human Development). Bolo ukázané, že používanie jednoduchých a rýchlych heuristík pri rozhodovaní a riešení problémov nie je vzácnym a ojedinelým javom, ale jedná sa dominantný prístup k rozhodovaniu, pričom používané heuristiky vykazujú evolučnú a kultúrnu zakotvenosť.

4. Moderná kognitívna psychológia nám dáva odpoveď na otázku, *prečo ľudia pri rozhodovaní sa správajú veľmi špecificky a „ohraničene"*, nie je to dôsledok „hlúposti" (nízkeho IQ, problému s pamäťou a pod.), manifestuje to našu „geneticky a kultúrne zakotvenú" schopnosť a tendenciu používať jednoduché heuristiky [24,25].

5. Môžeme diskutovať, či táto schopnosť a tendencia je pre „ľudského agenta" pozitívnou alebo negatívnou črtou. Súčasná kognitívna a evolučná psychológia poskytuje jednoznačnú odpoveď, že *táto schopnosť je „požehnaním" pre ľudský druh*, umožňuje rýchlu orientáciu v zložitom, tak materiálnom, ako aj sociálnom svete. Bez tejto vlastnosti by jeho šanca prežitia (fitnes) bola podstatne nižšia.

6. Pri zmenenej situácii (zmena prírodného prostredia, sociálnej štruktúry, a pod.) môže sa stať, že *niektoré používané heuristiky sa stanú „kontraproduktívne"*, brzdou ďalšieho rozvoja

multiagentového systému. Potom sa buď daná heuristika nahradí inou, v danej situácii vhodnejšou, alebo bude pre multiagentový systém predstavovať podstatnú bariéru jeho zachovania a rozvoja.

7. Pre umelú inteligenciu (a aj kognitívnu vedu) je veľkou výzvou *navrhovať mechanizmy vzniku týchto heuristík* použitím prostriedkov počítačovej inteligencie a sledovať ich kultúrny prenos a fixáciu pomocou memetických prístupov v systéme. Riešenie tohto problému spočíva hľadaní vhodnej stratégie pre rýchle a efektívne riešenie daného problému. Pri použití evolučných algoritmov je potrebné preto navrhnúť vhodné kódovanie stratégie, ktoré je dostatočne flexibilné k tomu, aby emergované riešenia boli dostatočne zaujímavé.

8. Môžeme si položiť záverečnú otázku, ako je možné, že u ľudí s tak nedokonalým a ohraničeným kognitívnym orgánom, ktorý pri riešení problémov používa skoro výlučne len jednoduché heuristické postupy, vznikla sofistikovaná veda? Jedná sa o veľmi zložitý problém, ktorého vysvetlenie nie je jednoduché. Vznik vedy (kognitívneho artefaktu) v populácii agentov, čo možno chápať ako emergenciu novej kvality, je asi umožnený tým, že v systéme existujú spoľahlivé memetické vertikálne prenosy poznatkov (mémov), ktoré zabezpečujú ich akumuláciu v priebehu mnohých životných cyklov agentov. Samozrejme, veda nie je lineárne ukladanie poznatkov, ktoré spolu nesúvisia, je reprezentovaná súborom navzájom previazaných konzistentných poznatkov. Preto musia existovať aj ďalšie „vyššie" kognitívne aktivity agentov nad touto sumou poznatkov, ktoré spočívajú napr. v odstraňovaní nekonzistentných alebo redundantných poznatkov, či v induktívnom zovšeobecňovaní a pod. V dôsledku ohraničenej racionality kognitívneho orgánu agentov je tento dvojitý proces pomalý a má inkrementálny charakter.

Problém ohraničenej racionality reprezentuje pre umelú inteligenciu a kognitívnu vedu zaujímavú a aktuálnu novú problematiku, ktorá doposiaľ v týchto oblastiach nebola študovaná. Štúdium vzniku heuristík pre riešenie zložitých úloh, problém ich klasifikácie podľa úspešnosti, spájanie rôznych heuristík do jednej novej heuristiky, a pod. reprezentujú veľmi zaujímavé problémy ležiace na rozhraní informatiky a sociálnych vied. Metódy počítačovej inteligencie (neurónové siete a evolučné algoritmy) a multiagentových systémov majú pri týchto štúdiách dôležité postavenie jednotiacej metodiky pre výskum a popis problematiky ohraničenej racionality.

Literatúra

[1] ARTHUR, W. B.: Cognition: The Black Box of Economics, In: *The Complexity Vision and the Teaching of Economics*. Ed. D. Colander. Edward Elgar Publishing, Northampton, MA, 2000.

[2] AXELROD, R. 1984: *The Evolution of Cooperation.* New York: Basic Books.

[3] AXELROD, R. 1997a: The Dissemination of Culture: A Model with Local Convergence and Global Polarization, In: *Journal of Conflict Resolution* 41, 203-226.

[4] AXELROD, R. 1997b: *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press.

[5] BLACKMORE, S. 1999: *The meme machin*e. Oxford: Oxford University Press (český preklad: *Teorie memů: Kultura a její evoluce*. Praha: Portál, 2000).

[6] CONTE, R., HEGSELMANN, R., TERNA, P., eds., 1997: *Simulating Social Phenomena.* Berlin: Springer Verlag.

[7] DAWKINS, R. 1976: *The selfish gene.* Oxford: Oxford University Press (český preklad: *Sobecký gen.* Praha: Mladá fronta, 1998).

[8] EPSTEIN, J., AXTELL, R. 1996: *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge, MA: MIT Press.

[9] GIGERENZER, G., SELTEN, R., eds., 2002: *The Bounded rationality: The Adaptive Tool Box.* Cambridge, MA: MIT Press.

[10] GIGERENZER, G., TODD, P.M., and the ABC Research Group, eds., 1999: *Simple Heuristics that Make Us Smart.* New York: Oxford University Press.

[11] KAHNEMAN, D., TVERSKY, A., eds., 2000: *Choices, values, and frames*. New York: Cambridge University Press.

[12] KAHNEMAN, D., SLOVIC, P. & TVERSKY, A., eds., 1982: *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

[13] KUBÍK, A. 2004: *Inteligentní agenty. Tvorba aplikačného software na bázi multiagentových systémů.* Brno: Computer Press.

[14] KVASNIČKA, V., BEŇUŠKOVÁ, Ľ., POSPÍCHAL, J., FARKAŠ, I., TIŇO. P., KRÁĽ, A. 1997: *Úvod do teórie neurónových sietí*. Bratislava: IRIS.

[15] KVASNIČKA, V., POSPÍCHAL, J. TIŇO. P. 2000: *Evolučné algoritmy*. Bratislava: Vydavateľstvo STU.

[16] KVASNIČKA, V., POSPÍCHAL, J. 1999: Evolutionary study of interethnic cooperation, In: *Advances in Complex Systems* 2, 395-421.

[17] KVASNIČKA, V., POSPÍCHAL, J. 2005: *Informatika pre sociálne vedy.* Bratislava: Vydavateľstvo UK.

[18] KVASNIČKA, V., POSPÍCHAL, J. 2002: Evolutionary Study of Interethnic Cooperation, In: *Modeling Complexity in Economic and Social Systems*. Ed. F. Schweitzer. New Jersey: World Scientific.

[19] KVASNIČKA, V., POSPÍCHAL, J. 1999: Simulation of Evolution of Dawkins Memes, In: *Evolution and Cognition* 5, 75-86.

[20] MEYER, J.-J. CH., VAN DER HOEK, W. 1995: *Epistemic Logic for AI and Computer Science*. Cambridge, UK: Cambridge University Press.

[21] NÁVRAT, P., kol. 2002: *Umelá inteligencia*. Bratislava: Vydavateľstvo STU (2006 nové vydanie).

[22] RYBÁR, J., BEŇUŠKOVÁ, Ľ., KVASNIČKA, V., editori, 2002: *Kognitívne vedy*. Bratislava: Kalligram.

[23] SIMON, H. A. 1957: *Models of Men*, New York: John Wiley and Sons.

[24] STERNBERG, R. J., ed., 2002: *Why Smart People Can Be So Stupid.* New Haven, CT: Yale University Press (slovenský preklad: *Prečo robia múdri ľudia hlúposti?* Bratislava: Ikar, 2004).

[25] STERNBERG, R. J. 1999: *Cognitive Psychology*. Orlando, FL: Harcourt (český preklad: *Kognitivní psychologie***.** Praha: Portál, 2002).

[26] SUTTON, R. S., BARTO, A. G. 1998: *Reinforcement Learning. An Introduction.* Cambridge, MA: MIT Press.

[27] THAGARD, P. 1996: *Mind. Introduction to Cognitive Science.* Cambridge, MA: MIT Press,(český preklad: *Úvod do kognitivní vědy. Mysl a myšlení.* Praha: Portál, 2001).

[28] KOLESÁROVÁ. A., KOVÁČOVÁ, M. 2004: *Fuzzy množiny a ich aplikácie.* Bratislava: Vydavateľstvo STU.

[29] RUBINSTEIN, A. 1998: *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.

[2] Hofstadter, D. R.: *Metamagical Themes*. Basic Books, New York, 1985.

[3] Hofstadter, D. R.: *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, New York. (1995).

[4] Mitchell, M.: *Analogy-Making as perception. A Computer Model.* The MIT Press, Cambridge, Massachusetts, 1993.

[5] Pavol Návrat a kol.:*Umelá inteligencia*, Vydavateľstvo STU, Bratislava 2002.

[6] Šefránek, J.: *Inteligencia ako výpočet*, IRIS, Bratislava, 2000.

[7] Spearman, C.: *The Nature of "Intelligence" and the Principles of Cognition.* Macmillan, London, 1923.

[8] Wilson, R. A., Keil, F. C.: *The MIT Encyclopedia of the Cognitive Sciences*, password *Analogy*, written by Gentner, D. The MIT Press, Cambridge, Massachusetts, 1999.