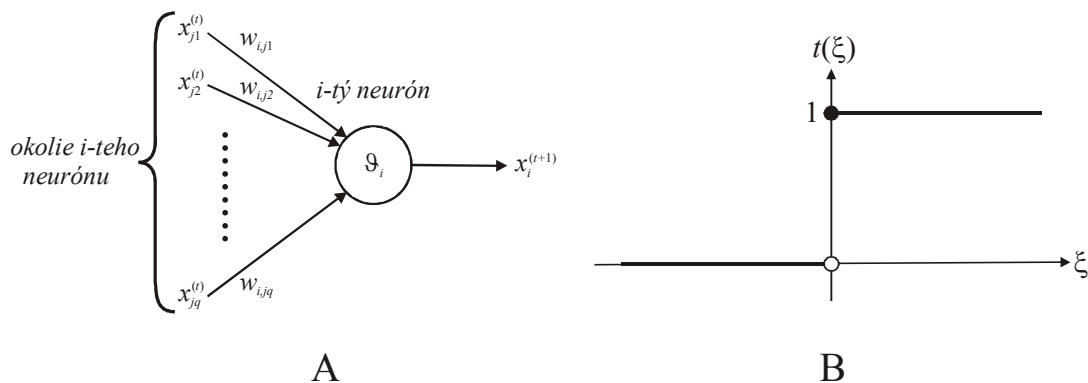


6 Hopfieldova sieť

6.1 Úvodné poznámky

Tento typ rekurentnej neurónovej siete bol navrhnutý americkým fyzikom Johnom J. Hopfieldom z Princetonskej univerzity v r. 1982 [xx]. Znamenal počiatok plodnej éry rozvoja konekcionizmu (neurónových sietí) s opätovným návratom záujmu informatikov k metafore umelých neurónových sietí, ktoré sa čoskoro aj zásluhou Hopfielda stali dominantnou oblasťou umelej inteligencie a zaznamenali prvé podstatné úspechy v interpretácii a modelovaní tak nižších ako aj vyšších kognitívnych aktivít. Metafora neurónových sietí sa stala dôležitou inžinierskou metódou pre riadenia zložitých systémov pre ktoré neexistuje matematický model (alebo je len veľmi ťažko zostrojiteľný).

Aj keď v súčasnosti Hopfieldova sieť nepatrí medzi často využívané metódy konekcionizmu, zohrala dôležitú historickú úlohu v rozvoji a formulovaní rôznych konekcionistických metód a myšlienok, akými sú napr. postupná pomalá degradácia schopností mozgu pri vymieraní neurónov a zániku spojov, charakteristika asociačnej pamäti pomocou pojmu minima účelovej funkcie, a pod. V tejto kapitole sa budeme zaoberať teóriou Hopfieldových sietí na veľmi zjednodušenej úrovni pomocou prahových neurónov McCullocha a Pittsa. Tento teoretický prístup, aj napriek svojej jednoduchosti, poskytuje v zásade všetky výsledky pôvodnej teórie navrhnutej už Hopfieldom [xx] pri riešení asociačných a optimalizačných úloh.



Obrázok 6.1. (A) Znáznorenie i -teho neurónu v sieti, ktorého predchodci (vytvárajúce in-okolie) daného neurónu vytvárajú množinu $\Gamma_{in}(i) = \{x_{i,j1}, x_{i,j2}, \dots, x_{i,jb}\}$. Ak potenciál tohto neurón (vážená suma aktivít vstupných neurónov z množiny $\Gamma_{in}(i)$ je väčší alebo rovný prahovému koeficientu ϑ_i , potom aktivita neurónu je jednotková (v opačnom prípade je nulová). (B) Grafické znázornenie priebehu prechodovej funkcie $t(\xi)$, pre neáporný argument je funkčná hodnota jednotková (v opačnom prípade nulová).

6.2 Špecifikácia siete

Predpokladajme, že neurónová sieť obsahuje n binárnych neurónov typu McCullocha a Pittsa. Ich binárna aktivita (nula alebo jednotka) je určená tak, že ak vážená suma vstupných aktivít je väčšia alebo rovná prahovému koeficientu, aktivita daného neurónu je jednotková; v opačnom prípade, ak vážená suma vstupných aktivít je menšia ako prahový koeficient, potom aktivita je nulová (pozri obr. 6.1)

$$x_i = t \left(\sum_j w_{ij} x_j + \vartheta_i \right) \quad (6.1a)$$

kde

$$t(\xi) = \begin{cases} 1 & (\xi \geq 0) \\ 0 & (\xi < 0) \end{cases} \quad (6.1b)$$

Postulujeme, že aktivity sú obnovované v diskrétnych časových krokoch 1, 2, ..., t, ... asynchrónnym spôsobom, t. j. v čase t náhodne vyberieme jeden neurón i, ktorého aktivita sa obnoví podľa (6.1a)

$$x_i^{(t+1)} = t \left(\sum_j w_{ij} x_j^{(t)} + \vartheta_i \right) \quad (6.2)$$

Zo špecifikácie asynchrónneho obnovovania aktivít neurónovej siete vyplýva stochastičnosť tohto procesu. Pre daný **stav siete** reprezentovaný vektorom aktivít $\mathbf{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \dots, x_n^{(t)})$ pomocou asynchrónneho obnovenia aktivít neurónov špecifikované indexom neurónu i, zostrojíme nový stav pomocou (6.2)

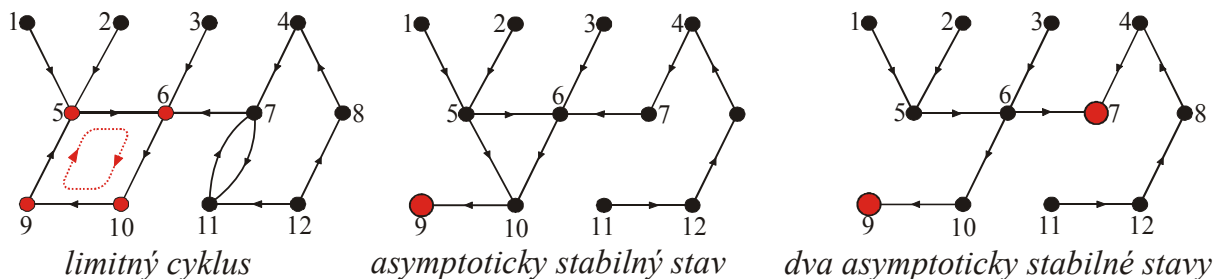
$$\mathbf{x}^{(t+1)} = f_i(\mathbf{x}^{(t)}) \quad (6.3)$$

To znamená, že pre každý stav $\mathbf{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \dots, x_n^{(t)})$ existuje n nových stavov v závislosti od toho, ktorý neurón bol náhodne vybraný na obnovovací proces

$$x_i^{(t)} = f_i(x_1^{(t)}, \dots, \boxed{x_i^{(t)}}, \dots, x_n^{(t)}) \quad (6.4)$$

kde vybraný neurón v stave $\mathbf{x}^{(t)}$ je zvýraznený štvorcom.

Nech priestor stavov neurónovej siete je označený $\mathcal{S} = \{0,1\}^n$, potom trajektória riešení s počiatočným stavom $\mathbf{x}^{(1)}$ je tvorená postupnosťou stavov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots, \mathbf{x}^{(t)}, \mathbf{x}^{(t+1)}, \dots$, pričom stav $\mathbf{x}^{(t)}$ vznikol z predchádzajúceho stavu $\mathbf{x}^{(t-1)}$ pomocou asynchrónneho obnovenia (6.3). Ak predpokladáme, že študovaná neurónová sieť je stabilná, potom na každej trajektórii existuje buď limitný cyklus alebo asymptoticky stabilný stav (pozri obr. 6.2).



Obrázok 6.2. Rôzne trajektórie dynamického systému obsahujúceho 12 stavov.

Definícia 6.1.

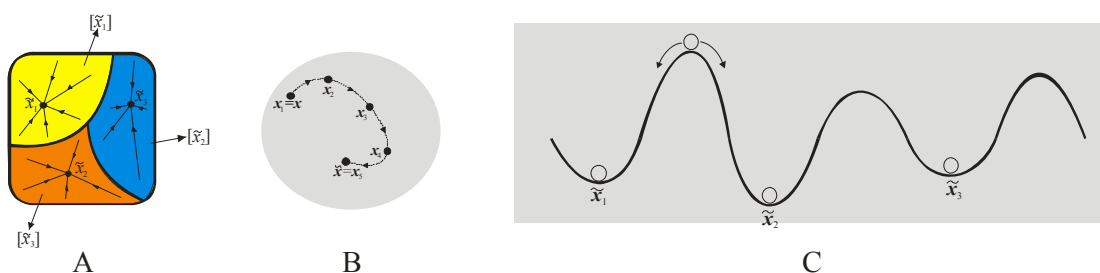
Hopfieldova sieť je špeciálny prípad neurónovej siete špecifikovanej vyššie, ktorá ma

- nulové diagonálne váhové koeficienty, $w_{ii} = 0$,
- symetrické nediagonálne váhové koeficienty, $w_{ij} = w_{ji}$, a
- aktivity neurónov sa obnovujú asynchrónne.

Pre takto definovanú Hopfieldovu sieť môžeme definovať tzv. účelovú funkciu energiu (ktorá hrá úlohu Ljapunovovej funkcie z kvalitatívnej teórie diferenciálnych rovníc [xx]). Ukazuje sa, že táto funkcia pozdĺž trajektórie stavov zostrojenej pomocou asynchrónneho obnovovania riešení tak, že akceptujeme len také nové stavy, ktorých energia je rovná alebo menšia predchádzajúcemu stavu, potom koncové stavy na trajektóriách musia byť také, že majú konštantnú energiu.

Nech energia Hopfieldovej siete pre binárny stav $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n) \in \{0, 1\}^n$ je definovaná takto

$$E(\mathbf{x}) = -\frac{1}{2} \sum_{\substack{j,k \\ (j \neq k)}} x_j w_{jk} x_k + \sum_{j=1}^n x_j \vartheta_j \quad (6.5)$$



Obrázok 6.3. (A) Rozklad priestoru stavov $S = \{0, 1\}^n$ na bazény príťažlivosti jednotlivých minim energií $E(\mathbf{x})$. Ak je trajektória zahájená v určitom bazéne príťažlivosti, potom celá trajektória leží v tomto bazéne. (B) Znárodnenie trajektórie obsahujúcej také susedné stavy, ktoré vznikli asynchrónnym obnovovacím procesom, každá trajektória buď končí v ostrom lokálnom minime alebo v niekoľkých stavoch navzájom prepojených s rovnakou hodnotou energie. (C) Funkcia energie má mnoho minim, z ktorých jedno môže byť charakterizované ako globálne minimum.

Veta 6.1.

Nech $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots, \mathbf{x}^{(t)}, \mathbf{x}^{(t+1)}, \dots$ je trajektória stavov, kde stav $\mathbf{x}^{(t)}$ vznikol z predchádzajúceho stavu $\mathbf{x}^{(t-1)}$ pomocou asynchrónneho obnovenia (6.3). Pre túto trajektóriu energia je nerastúca funkcia

$$E(\mathbf{x}) - E(\mathbf{x}') \geq 0 \quad (6.6)$$

kde stav $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ je obnovený na nový stav pre i -tu aktivitu na nový stav $\mathbf{x}' = (x_1, \dots, x'_i, \dots, x_n)$ tak, že

$$x'_i = x_i + \Delta x_i \quad (6.7a)$$

$$\Delta x_i = \begin{cases} 1 & (\text{pre } x_i = 0) \\ -1 & (\text{pre } x_i = 1) \end{cases} \quad (6.7b)$$

Pomocou jednoduchých algebraických úprav prepíšeme energiu (6.5) do tvaru

$$E(\mathbf{x}) - E(\mathbf{x}') = \Delta x_i \left(\sum_k w_{ik} x_k - \vartheta_i \right) \quad (6.8)$$

Pre rôzne znamienka vnútorného potenciálu i -teho neurónu $\xi_i = \sum_k w_{ik} x_k - \vartheta_i$ dostaneme tieto možné 4 prípady:

$$(1) x_i = 0 \rightarrow x'_i = 1 \quad \wedge \quad \xi_i \geq 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i \geq 0 \quad (6.9a)$$

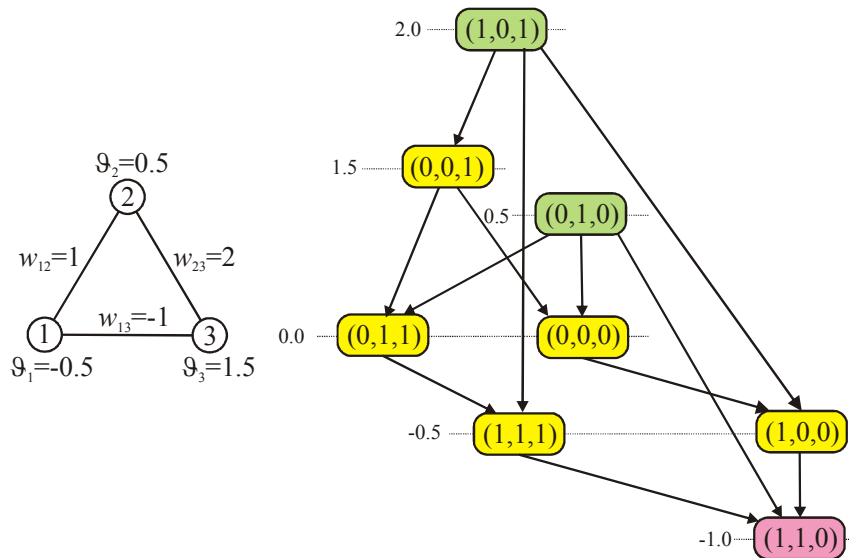
$$(2) x_i = 0 \rightarrow x'_i = 1 \quad \wedge \quad \xi_i < 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i < 0 \quad (6.9b)$$

$$(3) x_i = 1 \rightarrow x'_i = 0 \quad \wedge \quad \xi_i \geq 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i \leq 0 \quad (6.9c)$$

$$(4) x_i = 1 \rightarrow x'_i = 0 \quad \wedge \quad \xi_i < 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i > 0 \quad (6.9d)$$

Z týchto rôznych prípadov vyplýva, že súčin $\xi_i \Delta x_i$ môže byť kladný ako aj záporný (dokonca aj nulový). Z tohto priamočiario vyplýva, že ak pre daný stav \mathbf{x} neexistuje taký nový stav \mathbf{x}' s indexom $1 \leq i \leq n$, aby súčin $\xi_i \Delta x_i$ bol kladný, potom stav \mathbf{x} reprezentuje lokálne minimum. Toto minimum je ostré práve vtedy, ak pre každý „susedný“ stav \mathbf{x}_i súčin $\xi_i \Delta x_i > 0$. V prípade, že súčin $\xi_i \Delta x_i = 0$, potom pre susedný stav \mathbf{x}_i platí $E(\mathbf{x}) - E(\mathbf{x}') = 0$, t. j. oba stavy majú rovnakú hodnotu energie. Z týchto jednoduchých úvah vyplýva, že každá trajektória tvorená stavmi Hopfieldovej siete, pričom dva susedné stavy na trajektórii sú tvorené asynchrónnym obnovovacím procesom, končí v stave, ktorý je vzhľadom k funkcii energie v lokálnom minime, pozri obr. 6.3

$$\tau_{opt} = \arg \min_{\tau \in \{0,1\}^n} E_\tau \quad (6.10)$$



Obrázok 6.4. Ľavý diagram znázorňuje Hopfieldovu sieť z príkladu 6.2.1, pravý diagram je stavový diagram Hopfieldovej siete, spojené sú len také stavy, medzi ktorými je Hammingova vzdialenosť jednotková. Vidíme, že diagram obsahuje dva vstupné stavy, päť prechodných stavov a jeden výstupný stav, ktorý je priradený globálnemu minimu funkcie energie pre danú sieť.

6.2.1 Prvý ilustračný príklad Hopfieldovej siete

Nech Hopfieldova sieť obsahuje tri neuróny, pričom váhové a prahové koeficienty sú definované takto:

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 2 \\ -1 & 2 & 0 \end{pmatrix}, \quad \boldsymbol{\vartheta} = (-0.5, 0.5, 1.5)$$

Funkcia energie má potom tvar

$$E(\mathbf{x}) = -(x_1 x_2 - x_1 x_3 + 2x_2 x_3) + (-0.5x_1 + 0.5x_2 + 1.5x_3)$$

Stavový priestor tejto siete je generovaný binárnymi vektormi dĺžky tri, $S = \{0,1\}^3$. Funkčné hodnoty energie pre všetky stavy sú uvedené v tab. 6.1.

Tabuľka 6.1. Funkčné hodnoty energie pre stavy z $S = \{0,1\}^3$

stavový vektor	$E(x)$
x_1	(0,0,0) 0
x_2	(0,0,1) 1.5
x_3	(0,1,0) 0.5
x_4	(0,1,1) 0
x_5	(1,0,0) -0.5
x_6	(1,0,1) 2.0
x_7	(1,1,0) -1.0
x_8	(1,1,1) -0.5

Stavový diagram Hopfieldovej siete vytvoríme pomocou tab. 6.1 tak, že spojíme orientovanou čiarou len také dva stavy, medzi ktorými je jednotková Hammingova vzdialenosť. Stavy v diagrame môžeme charakterizovať na vstupné (zelené), výstupné (červené) a prechodné (žlté), pozri obr. 6.4.

6.2.2 Druhý ilustračný príklad Hopfieldovej siete

Nech Hopfieldova sieť obsahuje tri neuróny, pričom váhové a prahové koeficienty sú

$$W = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}, \quad \Theta = (0.5, 0.5, 0.5),$$

Funkcia energie má tvar

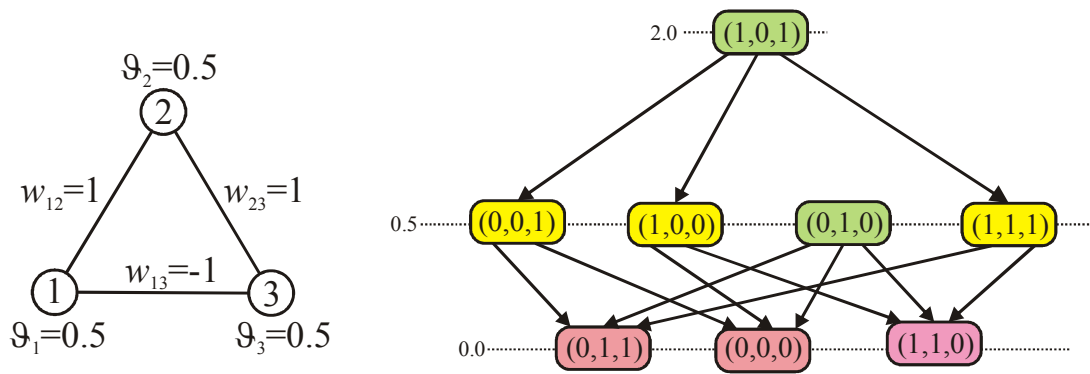
$$E(\mathbf{x}) = -(x_1x_2 - x_1x_3 + x_2x_3) + (0.5x_1 + 0.5x_2 + 0.5x_3)$$

Funkčné hodnoty energie sú uvedené v tab. 6.2.

Tabuľka 6.2. Funkčné hodnoty energie pre stavy z $S = \{0,1\}^3$

stavový vektor	$E(x)$
x_1	(0,0,0) 0
x_2	(0,0,1) 0.5
x_3	(0,1,0) 0.5
x_4	(0,1,1) 0
x_5	(1,0,0) 0.5
x_6	(1,0,1) 2.0
x_7	(1,1,0) 0
x_8	(1,1,1) 0.5

Stavový diagram Hopfieldovej siete vytvoríme pomocou tab. 6.2. Stavy v diagrame môžeme charakterizovať na vstupné (zelené), výstupné (červené) a prechodné (žlté), pozri obr. 6.5.



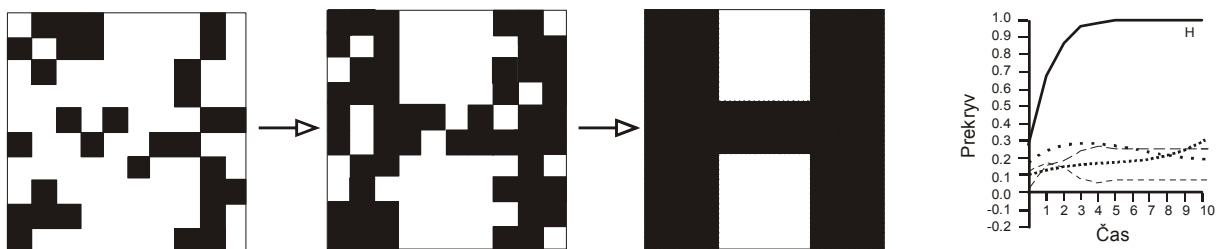
Obrázok 6.5. Ľavý diagram znázorňuje Hopfieldovu sieť z príkladu 6.2.2, pravý diagram je stavový diagram Hopfieldovej siete. Vidíme, že diagram obsahuje dva vstupné stavy, tri prechodné stavy a tri výstupné stavy, ktoré sú priradené lokálnym minimám funkcie energie pre danú sieť.

6.3. Učenie Hopfieldovej siete – asociatívna pamäť

Hopfieldova sieť môže fungovať ako asociatívna pamäť. Nech $\{ \mathbf{x}_k = (x_1^k, \dots, x_n^k); k = 1, 2, \dots, p \}$ je množina obsahujúca p binárnych obrazcov, potom

$$w_{ij} = \begin{cases} 0 & (i = j) \\ \sum_{k=1}^p x_i^k x_j^k & (i \neq j) \end{cases} \quad (6.11)$$

Pre takto špecifikované váhové koeficienty minimá energie Hopfieldovej siete sú priradené jednotlivým obrazcom z množiny (táto vlastnosť nie je niekedy splnená, existujú ešte tzv. fantómové obrazce, ktoré sú priradené ďalším, - falošným minimám), pozri obr. 6.6 prevzatý z kapitoly L. Beňuškovéj v [xx].



Obrázok 6.6. Ilustrácia činnosti Hopfieldovej autoasociatívnej pamäti. Sieť má 100 neurónov, 5 pamäťových vzorov (5 písmen) a deterministickú asynchrónnu dynamiku. Na vstupe je poškodené písmeno H, ktoré sieť vďaka svojej dynamike perfektne zrekonštruuje. Graf ilustruje vývoj prekryvov aktuálneho stavu siete s jednotlivými pamäťovými vzormi. Písmenom H je označená krivka odpovedajúca vývoju prekryvu pre písmeno H, ostatné krivky vyjadrujú vývoj prekryvu s ostatnými vzormi.

Tento jednoduchý ilustratívny príklad slúži v teórii neurónových sietí ako model asociatívnej pamäti v ľudskej mysli. Každý obrazec z asociatívnej pamäti je reprezentovaný jedným minimom funkcie energie. Ak narážka (angl. *cue*) na daný obrazec leží v jeho bazéne príťažlivosti, potom si v pamäti vybavíme správny obrazec. V opačnom prípade, ak narážka leží mimo bazénu príťažlivosti daného obrazca (hovoríme, že narážka je slabá – neurčitá), potom trajektória môže skončiť v inom minime než sme predpokladali. Tieto úvahy môžeme jednoducho interpretovať pomocou našich introspekčných skúseností, keď si chceme

spomenúť na nejaké meno. Ak narážka je dostatočne blízka danému menu, potom si správne meno okamžite vybavíme v našej pamäti, v prípade, že narážka je vzdialená od daného mena, obvykle si vybavíme nesprávne meno. Toto je základná vlastnosť asociačnej pamäti, kde jednotlivé pojmy predstavujú bazény príťažlivosti (pozri obr. 6.3, diagram A).

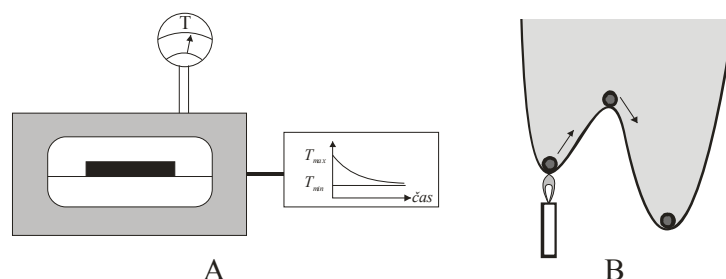
Ďalší dôležitý neurovedný problém, ktorý môže byť študovaný pomocou Hopfieldovej siete a úzko súvisiaci s predchádzajúcim problémom, je štúdium robustnosti a stability asociačných schopností ľudskej mysle vzhľadom k poklesu neurónov a/alebo synaptických spojov. Táto vlastnosť medzi základné dogmy konekcionizmu, podľa ktorej, asociačné vlastnosti neurónových sietí sú veľmi robustné a stabilné vzhľadom k stochastickému poklesu neurónov a/alebo synaptických spojov medzi nimi. V neurovede sa pokladá za experimentálne overený fakt, že schopnosti neurónovej siete fungovať ako asociačná pamäť je pomerne stabilná vzhľadom k stochastickému poklesu počtu neurónov a/alebo spojov. Táto vlastnosť robustnosti neurónových sietí platí pomerne do vysokého stupňa poklesu tak neurónov ako aj spojov medzi nimi, až po prekročení určitej prahovej hodnoty dochádza k dramatickému poklesu asociačných schopností.

6.3. Simulované žihanie a Hopfieldova sieť

V podkapitole 6.2 bolo ukázané, že Hopfieldova sieť má množstvo lokálnych miním (ktoré pri interpretácii siete ako autoasociatívnej pamäti sú priradené jednotlivým vyorom pamäti). Problém hľadania globálneho minima siete je integrálnou súčasťou niektorých aplikácií Hopfieldových sietí a preto sa budeme zaoberať týmto pomocou stochastickej optimalizačnej metódy nazývanej simulované žihanie [xx].

6.3.1 Simulované žihanie

Metóda simulovaného žihania [xx] (angl. simulated annealing - SA) patrí medzi tie stochastické optimalizačné algoritmy, ktoré, ako naznačuje už ich názov, majú základ vo fyzike (na rozdiel od iných stochastických optimalizačných algoritmov, ktoré majú svoj základ väčšinou v biológii). Algoritmus simulovaného žihania je založený na analógii medzi žiháním tuhých telies a optimalizačným problémom. Počiatkom 80-tych rokov Kirkpatrick, Gelatt a Vecchi [xx] a nezávisle Černý [xx] (z FMFI UK v Bratislave) dostali geniálny nápad, že problém hľadania globálneho minima sa môže realizovať podobným spôsobom ako žihanie tuhého telesa. Tento nový prístup k hľadaniu globálneho minima funkcie bol nazvaný *simulované žihanie*.



Obrázok 6.7. (A) Znáznornenie fyzikálnej realizácie žihania. Teleso sa vloží do pece (ľavý blok), ktorá je vyhriata na vysokú teplotu T_{max} . Teplota sa programovacím zariadením (pravý blok) pomaly znižuje na teplotu T_{min} . Týmto spôsobom sa odstránia štruktúrne defekty vyskytujúce sa v telese. (B) Použitie metafory simulovaného žihania k hľadaniu globálneho minima funkcií. Funkcia má dve lokálne minimá, k tomu, aby sme funkčnú hodnotu „vychýlili“ z ľavého lokálneho minima, zahrejeme systém a tým zvýšime pravdepodobnosť preskoku bariéry a možnosti argumentu dostať sa do pravého globálneho minima.

Žíhanie vo fyzike označuje taký proces, v ktorom je teleso umiestnené do pece vyhriatej na vysokú teplotu a postupným pomalým znižovaním teploty (pozri obr. 6.1) sa odstraňujú vnútorné defekty telesa. Pri vysokej teplote sú častice telesa (atómy alebo molekuly) náhodne usporiadané v priestore, takže teleso je roztopené. Potom sa teplota postupne znižuje, všetky častice telesa majú možnosť dostať sa do rovnovážnej polohy, čiže energia telesa sa znižuje. Budeme predpokladať, že proces ochladzovania je dostatočne pomalý, potom pre každú teplotu T je žíhané teleso v tepelnej rovnováhe, ktorá je opísaná boltzmannovským rozdelením pravdepodobnosti, že pri teplote T je teleso v stave i s energiou E_i

$$w_T(E_i) = \frac{1}{Q(T)} \exp\left(-\frac{E_i}{kT}\right) \quad (6.12a)$$

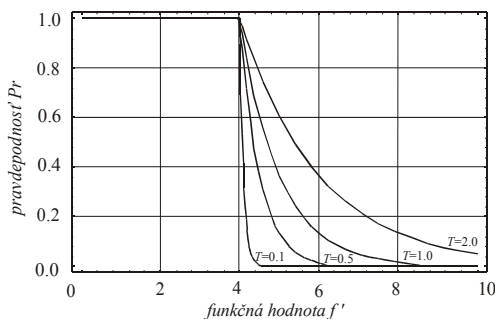
kde k je Boltzmannova konštanta a $Q(T)$ je normalizačný faktor nazývaný *partičná funkcia*

$$Q(T) = \sum_i \exp\left(-\frac{E_i}{kT}\right) \quad (6.12b)$$

kde sčítanie obsahuje všetky stavy i telesa. Ak teplota T klesá, potom boltzmannovská distribúcia uprednostňuje stavy s menšou energiou. V prípade, že sa teplota blíži k nule, stav s minimálnou energiou má nenulovú (jednotkovú) pravdepodobnosť. Žiaľ, ako je dobre známe, ak je ochladzovanie telesa príliš rýchle (telesu nie je umožnené získať tepelnú rovnováhu pre každú teplotu T), defekty v telese môžu "zamrznúť" za vzniku metastabilných štruktúr, ktoré sú vzdialené od mriežkovej štruktúry s najnižšou energiou.

Ako simulovať fyzikálnu evolúciu systému (napr. mnohočasticového kryštálu) smerom k tepelnej rovnováhe pre danú teplotu T ? Metropolis a spol. [xx] navrhli metódu Monte Carlo, ktorá simuluje evolúciu systému tak, že generuje postupnosť stavov systému nasledujúcim spôsobom: Nech je daný aktuálny stav systému (určený polohou častíc telesa), potom sa malá náhodná porucha generuje tak, že častice sú "jemne" posunuté. Poznamenajme, že táto porucha musí byť "symetrická", t.j. pravdepodobnosť toho, že malou poruchou sa stav A zmení na stav B, musí byť rovnaká ako pri zmene malou poruchou stavu B na stav A. Ak je rozdiel $\Delta E = E_{\text{perturbed}} - E_{\text{current}}$ medzi porušeným stavom a aktuálnym stavom negatívny ($E_{\text{perturbed}} < E_{\text{current}}$), potom proces pokračuje s novým porušeným stavom. V opačnom prípade, ak $\Delta E \geq 0$, pravdepodobnosť $Pr(\text{perturbed} \leftarrow \text{current})$ akceptovania porušeného stavu určuje exponenciála $\exp(-\Delta E / kT)$ (pozri obr. 6.2)

$$Pr(\text{perturbed} \leftarrow \text{current}) = \min(1, \exp(-\Delta E / kT)) \quad (6.12c)$$



Obrázok 6.8. Priebeh Metropolisovho kritéria $Pr = \min(1, \exp(-(f'-f)/T))$ pre rôzne teploty T , kde f je fixná funkčná hodnota ($f=4$) a f' je nezávislá premenná z intervalu $[0,10]$. Pre klesajúce hodnoty teploty T a pre $f' > f$ pravdepodobnosť $Pr \rightarrow 0$, ak $T \rightarrow 0$.

Toto pravidlo akceptovania porušeného stavu sa nazýva *Metropolisovo kritérium*. Podľa tohto kritéria aplikovaním veľkého počtu porúch a ich akceptovaním do ďalšieho procesu s pravdepodobnosťou (6.12) dostaneme systém v tepelnej rovnováhe, pričom distribúcia pravdepodobnosti rozloženia stavov sa asymptoticky blíži k boltzmannovskej distribúcii (6.11). Tento tvar metódy Monte Carlo sa v štatistickej fyzike nazýva *Metropolisov algoritmus* [5]. K tomu, aby sme formalizovali Metropolisov algoritmus (pozri algoritmus 6.1) zavedieme nasledujúci formalizmus (ktorý je vhodný pre aplikáciu simulovaného žihania na minimalizáciu funkcií): nech je stav systému určený *stavovou premennou* \mathbf{x} (vo všeobecnosti vektor obsahujúci mnoho nezávislých reálnych premenných) a energiou $f(\mathbf{x})$. Proces porušenia stavu \mathbf{x} na iný stav \mathbf{x}' je formálne reprezentovaný stochastickým operátorom, $\mathbf{x}' = O_{pertur}(\mathbf{x})$. Stochastický charakter tohto operátora spočíva v náhodnej zmene stavu \mathbf{x} na stav \mathbf{x}' .

Ako už bolo poznamenané, Metropolisov algoritmus [5] produkuje distribúciu pravdepodobnosti stavov, ktorá sa asymptoticky blíži k boltzmannovskej distribúcii (6.11)

$$w_T(x) = \frac{1}{Q(T)} \exp\left(-\frac{f(x)}{T}\right) \quad (6.13a)$$

$$Q(T) = \sum_x \exp\left(-\frac{f(x)}{T}\right) \quad (6.13b)$$

kde súčet obsahuje všetky stavy \mathbf{x} . Pre jednoduchosť sa Boltzmannova konštanta k teraz vynechala (pozri rovnice (6.11a-b), formálne bola táto konštanta zahrnutá do teploty T .

```

procedure Metropolis_algorithm(input:  $\mathbf{x}_{ini}$ ,  $k_{max}$ , T; output:  $\mathbf{x}_{out}$ );
begin k:=0;  $\mathbf{x}:=\mathbf{x}_{ini}$ ;
  while k< $k_{max}$  do
    begin k:=k+1;
       $\mathbf{x}' := O_{pertur}(\mathbf{x})$ ;
      Pr:=min(1, exp(-(f( $\mathbf{x}'$ ))-f( $\mathbf{x}$ ))/T));
      if random<Pr then  $\mathbf{x}:=\mathbf{x}'$ ;
    end;
     $\mathbf{x}_{out} := \mathbf{x}$ ;
end;

```

Algoritmus 6.1. Implementácia Metropolisovho algoritmu. Procedúra sa inicializuje tak, že počiatočný stav sa položí rovný vstupnému stavu \mathbf{x}_{ini} , opakuje sa k_{max} -krát (toto číslo musí byť dostatočne veľké, aby sa dosiahla tepelná rovnováha). Symbol O_{pertur} modifikuje aktuálny stav \mathbf{x} na \mathbf{x}' . Akceptovanie nového stavu sa rieši pomocou Metropolisovho kritéria realizovaného pre teplotu T . Po skončení Metropolisovho algoritmu je výstupným stavom \mathbf{x}_{out} posledný stav \mathbf{x} .

Metropolisov algoritmus sa môže použiť na počítačovú simuláciu žihania. V tomto prístupe sa simulované žihanie chápe ako postupnosť Metropolisových algoritmov realizovaných pre postupnosť vhodne sa znižujúcich teplôt, pričom výstupný stav \mathbf{x}_{out} z posledného Metropolisovho algoritmu slúži ako vstupný stav \mathbf{x}_{in} pre nasledujúci Metropolisov algoritmus. Táto procedúra sa inicializuje tak, že teplota T sa položí rovná maximálnej teplote T_{max} a Metropolisov algoritmus sa aplikuje tak dlho (k_{max} -krát, kde k_{max} je parameter Metropolisovho algoritmu), až sa dosiahne tepelná rovnováha. Potom sa teplota zníži (napr. pravidlom $T:=\alpha T$, kde $0 \ll \alpha < 1$) a znovu sa aplikuje Metropolisov algoritmus, pričom sa predpokladá, že sa dosiahne tepelná rovnováha. Algoritmus sa ukončí vtedy, keď teplota T dosiahne minimálnu hodnotu T_{min} , pre ktorú už žiadny stav s vyššou funkcionálnou hodnotou

nie je akceptovaný (skoro nulová pravdepodobnosť akceptovania). "Zamrznutý" stav \mathbf{x}_{out} sa potom chápe ako výsledné riešenie (pozri algoritmus 6.2).

```

procedure Simulated_annealing (input:  $T_{min}, T_{max}, k_{max}, \alpha$ ;
                                output:  $\mathbf{x}_{opt}$ );
begin  $\mathbf{x}_{ini}$  := náhodne vygenerovaný stavový vektor;
       $T := T_{max}$ ;
      while  $T > T_{min}$  do
        begin Metropolis_algorithm( $\mathbf{x}_{ini}, \mathbf{x}_{out}, k_{max}, T$ );
           $\mathbf{x}_{ini} := \mathbf{x}_{out}$ ;
           $T := \alpha * T$ ;
        end;
       $\mathbf{x}_{opt} := \mathbf{x}_{out}$ ;
end;

```

Algoritmus 6.2. Implementácia simulovaného žihania, vstupné parametre sú $T_{min}, T_{max}, k_{max}, \alpha$, výstupný parameter je \mathbf{x}_{opt} . Algoritmus sa inicializuje náhodným vygenerovaním počiatočného stavu \mathbf{x}_{ini} a maximálnou teplotou T_{max} . Cyklus **while** sa opakuje tak dlho, pokiaľ platí $T > T_{min}$, teplota T sa znižuje pomocou parametra α podľa vzťahu $T := \alpha * T$. Po ukončení cyklu **while** sa výsledný stav \mathbf{x}_{out} považuje za výsledné riešenie označené \mathbf{x}_{opt} .

Jeden zo základných problémov simulovaného žihania je špecifikácia teplôt T_{max}, T_{min} a pravidla pre znižovanie teploty T . V literatúre boli navrhnuté rôzne schémy, ako určiť tieto parametre simulovaného žihania [xx]. Pretože náplňou tejto kapitoly nie je podrobne študovať teóriu simulovaného žihania, budeme používať najjednoduchší prístup pre ochladzovanie (multiplikatívny prístup pomocou parametra α) a počiatočná teplota je zvolená tak, aby približne 50% porušených stavov bolo akceptovaných Metropolisovým algoritmom.

6.3.2 Konštrukcia stabilného stavu pomocou simulovaného žihania

V tejto podkapitole aplikujeme metódu simulovaného žihania k hľadaniu globálneho minima funkcie energie Hopfieldovej siete

$$\mathbf{x}_{opt} = \arg \min_{\mathbf{x}} E(\mathbf{x}) \quad (6.14)$$

Skombinujeme metódu asynchrónneho obnovovania stavov s metódou simulovaného žihania. Rozdiel energií pre dva stavy je definovaný v podkapitole 6.3.1 takto

$$\Delta E = E_{perturbed} - E_{current} = E(\mathbf{x}') - E(\mathbf{x}) = -\Delta x_i \left(\sum_k w_{ik} x_k - \vartheta_i \right) = -\Delta x_i \xi_i \quad (6.15)$$

Potom pravdepodobnosť akceptovania nového stavu je

$$Pr(\mathbf{x} \rightarrow \mathbf{x}') = \min(1, \exp(-\Delta E / kT)) = \min(1, \exp(\Delta x_i \xi_i / T)) \quad (6.16)$$

V prípade, že platí

$$x_i = 0 \rightarrow x'_i = 1 \quad \wedge \quad \xi_i < 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i < 0$$

alebo

$$x_i = 1 \rightarrow x'_i = 0 \quad \wedge \quad \xi_i \geq 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i \leq 0$$

Potom pravdepodobnosť akceptovania nového riešenia je

$$Pr(\mathbf{x} \rightarrow \mathbf{x}') = \exp(\Delta x_i \xi_i / T) \leq 1 \quad (6.17a)$$

V opačnom prípade, ak platí

$$x_i = 0 \rightarrow x'_i = 1 \quad \wedge \quad \xi_i \geq 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i \geq 0$$

alebo

$$x_i = 1 \rightarrow x'_i = 0 \quad \wedge \quad \xi_i < 0 \quad \Rightarrow \quad \Delta x_i \cdot \xi_i > 0$$

V tomto prípade je pravdepodobnosť

$$Pr(\mathbf{x} \rightarrow \mathbf{x}') = 1 \quad (6.17b)$$

Tieto dva výsledky (6.17a-b) môžeme spojiť do jednej formuly

$$Pr(\mathbf{x} \rightarrow \mathbf{x}') = \exp(\Delta x_i \xi_i / T) = \begin{cases} \leq 1 & (\text{pre } \Delta x_i \xi_i \leq 0) \\ > 1 & (\text{pre } \Delta x_i \xi_i \geq 0) \end{cases} \quad (6.18)$$

Akceptovanie nového stavu \mathbf{x}' Hopfieldovej siete sa vykoná s pravdepodobnosťou $Pr(\mathbf{x} \rightarrow \mathbf{x}')$ pomocou jednoduchšej podmienky

$$\text{random}[0,1] \leq Pr(\mathbf{x} \rightarrow \mathbf{x}') \quad (6.19)$$

kde $\text{random}[0,1]$ je náhodné číslo z intervalu $[0,1]$ s rovnomerným rozdelením pravdepodobnosti. V prípade, že $E(\mathbf{x}') < E(\mathbf{x})$, potom $Pr(\mathbf{x} \rightarrow \mathbf{x}') > 1$ a nové riešenie \mathbf{x}' sa akceptuje automaticky. V opačnom prípade, keď $E(\mathbf{x}') > E(\mathbf{x})$, potom nové riešenie \mathbf{x}' je akceptované s pravdepodobnosťou $Pr(\mathbf{x} \rightarrow \mathbf{x}') < 1$. Stupeň „malosti“ tejto pravdepodobnosti môžeme riadiť pomocou parametra „teploty“ T , pravdepodobnosť $Pr(\mathbf{x} \rightarrow \mathbf{x}')$ monotónne klesá k nule pre $T \rightarrow 0$ (t. j. systém ochladzujeme). Ak tento spôsob výpočtu pravdepodobnosti Pr akceptovania nového riešenia zavedieme do Metropolisovho algoritmu (pozri algoritmus 6.1), potom dostaneme modifikovaný Metropolisov algoritmus, ktorý je základom Boltzmannovho stroja, pozri algoritmus 6.3.

```

procedure Modified_Metropolis_algorithm(input:  $\mathbf{x}_{ini}$ ,  $k_{max}$ ,  $T$ ;
                                         output:  $\mathbf{x}_{out}$ );
begin  $k := 0$ ;  $\mathbf{x} := \mathbf{x}_{ini}$ ;
      while  $k < k_{max}$  do
        begin  $k := k + 1$ ;
           $i := 1 + \text{random}(n)$ ;
           $\xi = \sum_{j=1}^n w_{ij} x_j - \mathcal{G}_i$ ;  $Pr := \exp(-\xi/T)$ ;
          if  $\text{random}[0,1] < Pr$  then
            begin if  $(\xi_i \geq 0) \wedge (x_i = 0)$  then  $x_i := 1$  else
              if  $(\xi_i < 0) \wedge (x_i = 1)$  then  $x_i := 0$ ;
            end;
          end;
         $\mathbf{x}_{out} := \mathbf{x}$ ;
      end;

```

Algoritmus 6.3. Implementácia modifikovaného Metropolisovho algoritmu, kde pravdepodobnosť akceptovania nového riešenia je špecifikovaná formulou (6.18). Ak tento algoritmus použijeme v procedure `Simulated_annealing` (pozri algoritmus 6.2), potom dostaneme kombináciu Hopfieldovej siete so simulovaným žíhaním.

6.4 Optimalizačné úlohy

Ukazuje sa, že mnohé kombinatoriálne problémy je možné špecifikovať pomocou účelovej funkcie energie (6.5) pre Hopfieldovu sieť. V tejto podkapitole ukážeme dva kombinatoriálne problémy, a to problém nájdenia binárneho vektora a problém obchodného cestujúceho, pomocou ktorých ukážeme efektívnosť Hopfieldových sietí pre riešenie týchto problémov.

6.4.1 Hľadanie binárneho vektora

Nech $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \{0, 1\}^n$ je binárny vektor, ktorého tvar bude zakódovaný v Hopfieldovej sieti, našou úlohou bude nájsť tento binárny vektor. Najzložitejšia verzia tejto úlohy je taká, že vektor α je známy len *orákulu*, ktorému môžeme poslať dotazy, či daný binárny vektor x je totožný s α , dostávame len odpoveď áno/nie. Takto formulovaná úloha je *neriešiteľná* aj pre evolučné algoritmy. Ukážeme jednoduchý prípad tohto problému, keď pomocou znalosti vektora α zostrojíme funkciu energie

$$E(\mathbf{x}) = \left(\sum_{i=1}^n x_i - p \right)^2 + \sum_{i=1}^n (x_i - \alpha_i)^2 \quad (6.20a)$$

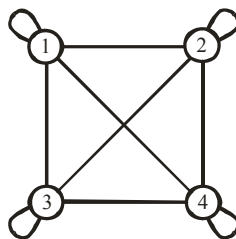
ktorú po jednoduchých úpravách prepíšeme do tvaru

$$E(\mathbf{x}) = -\frac{1}{2} \sum_{i \neq j} w_{ij} x_i x_j + \sum_i \vartheta_i x_i \quad (6.20b)$$

kde $w_{ij} = -2$ a $\vartheta_i = 2(1 - p - \alpha_i)$, parameter $p = |\alpha|$ špecifikuje počet '1' v α . Dosadením týchto „parametrov“ do (6.20b) a zanedbaním konštantného člena, dostaneme zjednodušený tvar energie

$$E(\mathbf{x}) = \sum_{i \neq j} x_i x_j + 2 \sum_i \alpha_i x_i = \sum_{i \neq j} x_i x_j + \alpha \mathbf{x}^T \quad (6.21)$$

kde symbol $\alpha \mathbf{x}^T$ je skalárny súčin riadkových vektorov α a \mathbf{x} . Grafické znázornenie Hopfieldovej siete pre $n = 4$ je ukázané na obr. 6.9.



Obrázok 6.9. Znázornenie Hopfieldovej siete pre hľadanie binárneho vektora x dimenzie $n = 4$.

6.4.2 Riešenie úlohy obchodného cestujúceho (TSP)

Nech G je úplný neorientovaný graf obsahujúci n vrcholov, pričom pre každú dvojicu vrcholov v_i a v_j poznáme „vzdialenosť“ d_{ij} medzi nimi. Úloha obchodného cestujúceho spočíva v tom, že hľadáme uzavretú hamiltonovskú cestu na grafe G , ktorá má minimálnu vzdialenosť. Nech $P = (p_1, p_2, \dots, p_n)$ je permutácia n objektov, ktorá nám špecifikuje uzavretú hamiltonovskú cestu na grafe

$$p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_{n-1} \rightarrow p_n \quad (6.22)$$

Túto schému môžeme interpretovať tak, že cesta je zahájená vo vrchole – meste p_1 , z ktorého ide do p_2 , potom pokračuje do mesta p_3 , atď., z posledného mesta p_n sa vracia do východiskového p_1 . Dĺžka tejto cesty P je

$$F(P) = d_{p_1, p_2} + d_{p_2, p_3} + \dots + d_{p_{n-1}, p_n} + d_{p_n, p_1} = \sum_{k=1}^n d_{p_k, p_{k+1}} \quad (6.23)$$

Úloha TSP spočíva v tom, že hľadáme optimálnu cestu – permutáciu P , ktorá minimalizuje dĺžku cesty

$$P_{opt} = \arg \min_{P \in S_n} F(P) \quad (6.24)$$

K tomu, aby sme preformulovali úlohu TSP pre Hopfieldovu sieť, permutácia P bude reprezentovaná permutačnou binárnou maticou $\mathbf{P} = (P_{ij})$, ktorá v každom stĺpci a v každom riadku obsahuje práve jeden '1' element a ostatné elementy sú '0'. Túto podmienku vyjadríme pomocou účelovej funkcie

$$G(\mathbf{P}) = \sum_i \left(\sum_j p_{ij} - 1 \right)^2 + \sum_j \left(\sum_i p_{ij} - 1 \right)^2 \quad (6.25)$$

Pomocou permutačnej matice môžeme vyjadriť aj funkciu vzdialenosti $F(P)$

$$F(\mathbf{P}) = \sum_{i,j=1}^n \left(\sum_{k=1}^{n-1} d_{ij} p_{i,k} p_{j,k+1} + d_{ij} p_{i,n} p_{j,1} \right) = \sum_{i,j,k=1}^n d_{ij} p_{i,k} p_{j,k+1} \quad (6.26)$$

Funkcia energie (Ljapunovova funkcia) má potom tvar

$$E(\mathbf{P}) = F(\mathbf{P}) + \gamma G(\mathbf{P}) \quad (6.27)$$

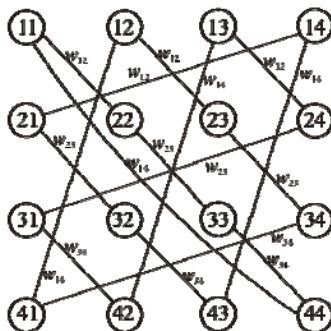
ktorá po jednoduchých úpravách môže byť prepísaná do tvaru

$$E(\mathbf{P}) = -\frac{1}{2} \sum_{ijk} (-2d_{ij} p_{i,k} p_{j,k+1}) + \gamma \left(\left(\sum_{j' i'} \sum_i p_{ij} p_{i'j'} - 2 \sum_{ij} p_{ij} + 1 \right) \left(\sum_{j' i'} \sum_j p_{ij} p_{i'j'} - 2 \sum_{ij} p_{ij} + 1 \right) \right) \quad (6.28)$$

Riadky a stĺpce permutačnej matice sú ortogonálne, čo vedie k zjednodušeniu funkcie energie

$$E(\mathbf{P}) = -\frac{1}{2} \sum_{ijk} (w_{ij} p_{i,k} p_{j,k+1}) \quad (6.29)$$

kde sme zanedbali konštantné členy. Hopfieldova neurónová sieť, ktorá reprezentuje TSP vyjadrený pomocou energie (6.29) je znázornená na obr. 6.10.



Obrázok 6.10. Hopfieldova sieť pre riešenie TSP s $n = 4$, obsahuje $4^2=16$ neurónov. Každý neurón je spojený s dvoma susednými neurónmi.

6.5 Hopfieldova sieť a penalizačná výroková logika

Cieľom tejto podkapitoly je ukázať neobvyklú aplikáciu Hopfieldovej siete [xx] pre hľadanie modelu penalizačnej výrokovej logiky [xx], ktorá je špecifikovaná takto: Nech $\Omega = \{p_1, p_2, \dots, p_a\}$ je množina (atomických) výrokových premenných, $|\Omega| = p$, a nech $\Phi = \{(\varphi_i, \alpha_i); i = 1, 2, \dots, a\}$ je teória, ktorá obsahuje dvojice (φ_i, p_i) , kde φ_i je výroková formula s premennými z Ω a α_i je kladná penalizačná konštanta. Nech $\mathcal{I} = \{\tau_1, \tau_2, \dots, \tau_b\}$ je množina interpretácií formúl z teórie Φ , interpretácia $\tau = (\tau_1, \tau_2, \dots, \tau_a) \in \{0, 1\}^a$ špecifikuje pravdivostné hodnoty výrokových premenných z množiny Ω ; potom pravdivostná hodnota i -tej premennej z Ω je určená veličinou $\tau_i \in \{0, 1\}$, čo zapisujeme $val_\tau(p_i) = \tau_i$. Toto pravdivostné zobrazenie môže byť zovšeobecnené aj pre neelementárne formuly (pozri obr. 6.11)

$$val_\tau(\varphi \wedge \psi) = val_\tau(\varphi) val_\tau(\psi), \quad (6.30a)$$

$$val_\tau(\neg\varphi) = 1 - val_\tau(\varphi) \quad (6.30b)$$

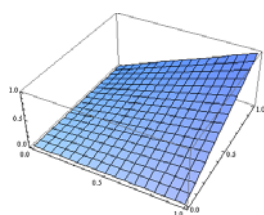
$$val_\tau(\varphi \vee \psi) = val_\tau(\neg(\neg\varphi \wedge \neg\psi)) = val_\tau(\varphi) + val_\tau(\psi) - val_\tau(\varphi) val_\tau(\psi) \quad (6.30c)$$

$$val_\tau(\varphi \Rightarrow \psi) = val_\tau(\neg\varphi \vee \psi) = 1 - val_\tau(\varphi) + val_\tau(\varphi) val_\tau(\psi) \quad (6.30d)$$

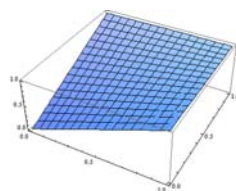
čo je umožnené tzv. extenzionálnym charakterom výrokových spojok, t. j. pravdivostná hodnota formuly závisí od pravdivostných hodnôt jej podformuly (poznajme, že táto skutočnosť nám umožňuje zistiť pravdivostné hodnoty formuly pomocou tabuľkovej metódy).

Sumárna penalizácia formúl z teórie Φ pre danú interpretáciu $\tau \in \mathcal{I}$ má tvar

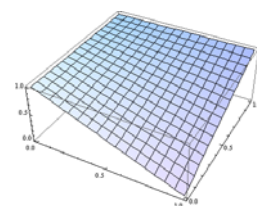
$$E(\tau) = \sum_{i=1}^a \alpha_i val_\tau(\neg\varphi_i) \quad (6.31)$$



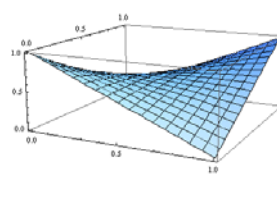
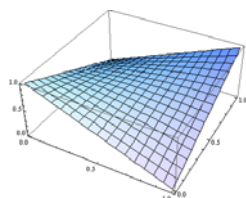
(A) $p \wedge q = p \cdot q$



(B) $p \vee q = p + q - p \cdot q$



(C) $p \Rightarrow q = 1 - p + p \cdot q$



(D) $p \equiv q = (p \Rightarrow q) \wedge (q \Rightarrow p) = 1 - p - q + p \cdot q (3 - p - q + pq)$

Obrázok 6.11. Grafická interpretácia logických spojok výrokovej logiky pre spojité pravdivostné ohodnotenie z intervalu $[0, 1]$.

V prípade, že množina $\mathcal{I} = \{\tau_1, \tau_2, \dots, \tau_b\}$ tvorí model teórie $\Phi = \{(\varphi_i, \alpha_i)\}$, potom sumárna penalizácia $E(\tau)$, pre každé $\tau \in \mathcal{I}$, je nulová. V penalizačnej výrokovej logike sa môže predpokladať, že teória nie je konzistentná, prípadne nekonzistentnosti sú penalizované. Optimálna interpretácia pre nekonzistentnú teóriu je špecifikovaná ako riešenie minimalizačného problému

$$\tau_{opt} = \arg \min_{\tau} E(\tau) \quad (6.32)$$

Tento optimalizačný problém sa bude riešiť pomocou Hopfieldovej neurónovej siete kombinovanej so simulovaným žiňaním (Boltzmannovým strojom)

6.5.1 Ilustračný príklad

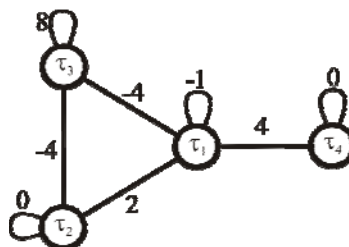
Nech teória Φ je definovaná pomocou tabuľky:

#	φ	α
1	m	1
2	$s \Rightarrow \neg m$	2
3	$c \Rightarrow s$	4
4	$c \Rightarrow m$	4
5	$v \Rightarrow \neg m$	4

Množina Ω obsahuje 4 výrokové premenné, $\{m, s, c, v\}$, potom interpretácia τ obsahuje 4 komponenty, ktoré špecifikujú pravdivostné hodnoty premenných z množiny Ω , $\tau = (m/\tau_1, s/\tau_2, c/\tau_3, v/\tau_4) = (\tau_1, \tau_2, \tau_3, \tau_4) \in \{0, 1\}^4$. Funkcia energie (6.31) má tvar

$$\begin{aligned} E(\tau) &= val_{\tau}(\neg m) + 2val_{\tau}(s \wedge m) + 4val_{\tau}(c \wedge \neg s) + 4val_{\tau}(c \wedge \neg m) + 4val_{\tau}(v \wedge m) \\ &= (1 - \tau_1) + 2\tau_1\tau_2 + 4\tau_3(1 - \tau_2) + 4\tau_3(1 - \tau_1) + 4\tau_4\tau_1 \\ &= \cancel{X} - \tau_1 + 8\tau_3 + 2\tau_1\tau_2 - 4\tau_1\tau_3 + 4\tau_1\tau_4 - 4\tau_2\tau_3 \end{aligned} \quad (6.32)$$

kde bol vynechaný konštantný člen ako nepodstatný pre hľadanie minima. Tejto energii je priradená Hopfieldova sieť znázornená na obr. 6.12.



Obrázok 6.12. Hopfieldova sieť priradená energii (6.32), sieť obsahuje 4 neuróny, 4 symetrické spoje a 4 slučky (z ktorých len dve majú nenulové ohodnotenie pomocou prahových koeficientov).

Stavový priestor Hopfieldovej siete z obr. 6.12 obsahuje $2^4 = 16$ stavov $\mathbf{x} = (x_1, x_2, x_3, x_4) \in \{0, 1\}^4$, ktoré sú spolu s hodnotami energie uvedené v tab. 6.3.

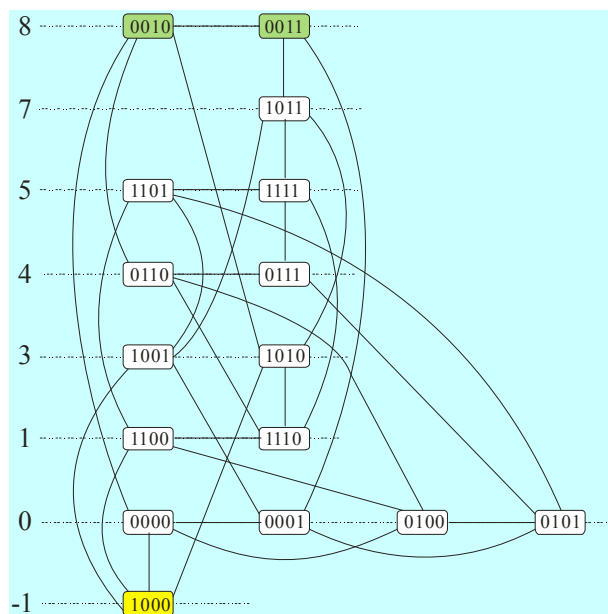
Tabuľka 6.3. Stavý Hopfieldovej siete z obr. 6.12 spolu s hodnotami energie

#	x	$E(x)$	#	x	$E(x)$
1	(0000)	0	9	(1000)	-1
2	(0001)	0	10	(1001)	3
3	(0010)	8	11	(1010)	3
4	(0011)	8	12	(1011)	7
5	(0100)	0	13	(1100)	1
6	(0101)	0	14	(1101)	5
7	(0110)	4	15	(1110)	1
8	(0111)	4	16	(1111)	5

Stavový diagram Hopfieldovej siete s energiou špecifikovanou (6.32) je znázornený na obr. 6.13. Vidíme, že diagram má dva východiskové stavy a jeden konečný stav, ostatné stavy sú prechodné. Energia má globálne minimum v stave (1000), t. j. model s minimálnou penalizáciou je $\tau_{opt} = (m/1, s/0, c/0, v/0)$, pozri tab. 6.4.

Tabuľka 6.4. Pravdivostné hodnoty formúl teórie pre τ_{opt}

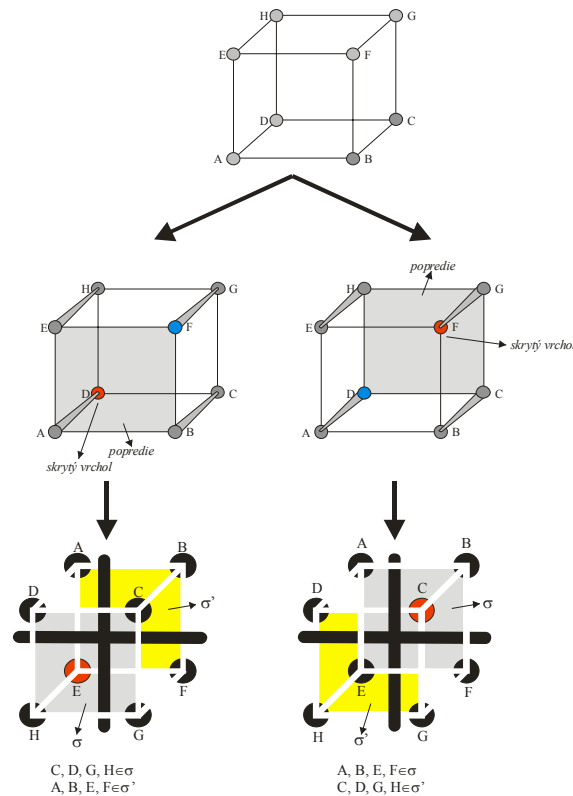
#	φ	$val_{\tau_{opt}}(\varphi)$
1	m	1
2	$s \Rightarrow \neg m$	1
3	$c \Rightarrow s$	1
4	$c \Rightarrow m$	1
5	$v \Rightarrow \neg m$	1



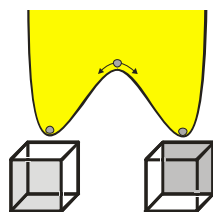
Obrázok 6.13. Stavový diagram Hopfieldovej siete s energiou (6.32).

6.6 Neckerova kocka

Neckerova kocka patrí medzi klasické príklady optickej ilúzie, kde sa nám jeden objekt (kocka) zdá byť v dvoch rôznych „konfiguráciách“, pričom vedomím sme schopní ovplyvňovať asi v 3 sekundových intervaloch formu, ktorú „vidíme“, pozri obr. 6.14. Túto vlastnosť ľudského mozgu neobyčajnej plasticity, schopnosť vedome meniť priestorovú formu Neckerov kocky, budeme interpretovať pomocou metafory Hopfieldovej siete. Táto sieť, ako bolo ukázané v predchádzajúcej časti tejto kapitoly, má zaujímavú vlastnosť v tom, že jej stavy môžu byť klasifikované ako stabilné a/alebo nestabilné, pričom asociačné vlastnosti siete sú špecifikované týmito stabilnými stavmi, ktoré v stavovom priestore interpretujeme ako stavy s minimálnou hodnotou účelovej funkcie – energie (pozri obr. 6.15).



Obrázok 6.14. Dve rôzne interpretácie „drôtového“ model Neckerovej kocky. V ľavej (pravej) interpretácii popredie kocky je realizované rovinou, ktorá obsahuje vrcholy A, B, E a F (C, D, G a H). V treťom riadku je znázornená Neckerova kocka pomocou pojmov známych z oblasti umelej inteligencie nazývanej „analýza scény“. Každá konfigurácia je jednoznačne klasifikovaná dvoma rovinami: prednou (σ) a zadnou (σ'). Taktiež na týchto diagramoch sú ukázané dve pevné telasa – tyče, ktoré prenikli do Neckerov kocky.



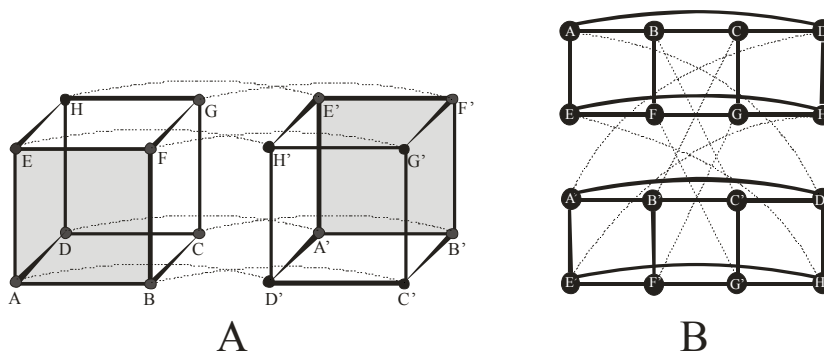
Obrázok 6.15. Interpretácia zmeny konfigurácie Neckerovej kocky pomocou metafory Hopfieldovej siete, ktorá obsahuje dve symetrické minima špecifikovaná ľavým/pravým priestorovým usporiadaním kocky. Ľudský mozog zrejme obsahuje mechanizmy (typ vedomia), pomocou ktorého môže meniť typ interpretácie Neckerovej kocky. Podľa metafory Hopfieldovej siete, tento jav môže byť interpretovaný pomocou dvoch susedných miním na funkcii energie v stavovom priestore.

Hopfieldova sieť, ktorá je schopná interpretovať problém dvoch konfigurácií Neckerovej kocky je znázornená na obr. 6.16. Obsahuje 16 neurónov, spoje sú znázornené buď pevnou čiarou (ohodnotenou váhovým koeficientom $w = 1$) alebo prerušovanou čiarou (ohodnotenou váhovým koeficientom $w = -2$). Poznamenajme, že tento typ Hopfieldovej siete má nulové prahové koeficienty, $\vartheta_X = 0$, pre $X = A, B, \dots, F, G, A', B', \dots, F', G'$. Účelová funkcia „energie“ má podľa (6.5) má tvar

$$E_{\tau} = - \sum_{(X,Y)} val_{\tau}(x \wedge y) - \sum_{(X',Y')} val_{\tau}(x' \wedge y') + 2 \sum_{X,Y'} val_{\tau}(x \wedge y') \quad (6.33)$$

kde prvá a druhá sumácia obsahuje „excitačné“ hrany z ľavej resp. pravej kocky, posledná tretia sumácia obsahuje zmiešané „inhibičné“ hrany medzi pravou a ľavou kockou (reprezentované prerušovanými čiarami na obr. 6.16)). Riešenie Hopfieldovej siete pre tento prípad spočíva v hľadaní optimálnych vektorov interpretácie τ_{opt} (pozri (6.32))

$$\tau_{opt} = \arg \min_{\tau \in \{0,1\}^{16}} E_{\tau} \quad (6.34)$$



Obrázok 6.16. Znázornenie Hopfieldovej siete pre interpretáciu javu zmeny konfigurácie Neckerovej kocky.

Pomocou (6.30) prepíšeme do štandardného „algebraického“ tvaru energiu (6.33)

$$E_{\tau} = - \sum_{(X,Y)} x \cdot y - \sum_{(X',Y')} x' \cdot y' + 2 \sum_{X,Y'} x \cdot y' \quad (6.35)$$

Riešením takto špecifikovanej Hopfieldovej siete získame optimálne stavy $\tau_{opt} \in \{0,1\}^{16}$, ktoré minimalizujú energiu (6.35) (pozri obr. 6.15)

$$\tau_{opt}^{(1)} = \left(\underbrace{1,1,\dots,1,1}_{8\text{-krát}}, \underbrace{0,0,\dots,0,0}_{8\text{-krát}} \right), \tau_{opt}^{(2)} = \left(\underbrace{0,0,\dots,0,0}_{8\text{-krát}}, \underbrace{1,1,\dots,1,1}_{8\text{-krát}} \right) \quad (6.36)$$

kde $E_{\tau_{opt}^{(1)}} = E_{\tau_{opt}^{(2)}} = -12$. Z týchto jednoduchých úvah vyplýva, že Hopfieldova sieť je schopná „rozpoznať“ jednotlivé konfigurácie Neckerovej kocky, ktorým sú priradené dve globálne minimá na povrchu účelovej funkcie energie. Tento výsledok sa pokladá v neurovede už od počiatku 80. rokov minulého storočia [xx] za významný prínos teórie neurónových sietí k pochopeniu vlastností ľudského mozgu rozpoznávania dvoch konfigurácií Neckerovej kocky, pričom sme schopní vedome meniť jednu konfiguráciu na druhú a naopak.

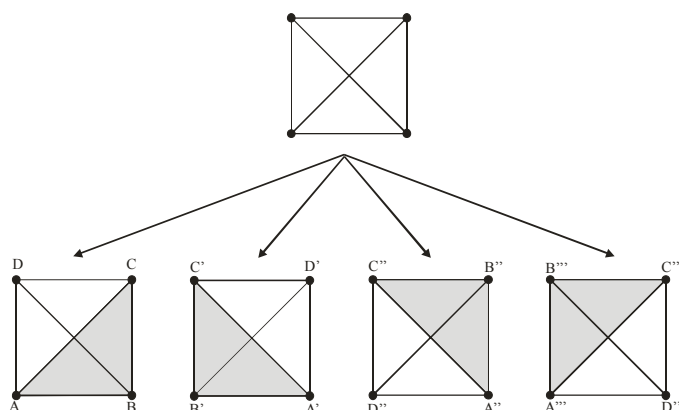
6.6.1 Ilustratívny príklad štvorstena

Ukážeme jednoduchý príklad štvorstena, kde sú rozlíšené štyri priestorové konfigurácie, pozri obr. 6.17. Analogickým spôsobom ako pre Neckerovu kocku, aj pre štvorsten môže byť zostrojená Hopfieldova sieť, ktorá pomocou rezultujúcich stabilných stavov bude rozlišovať

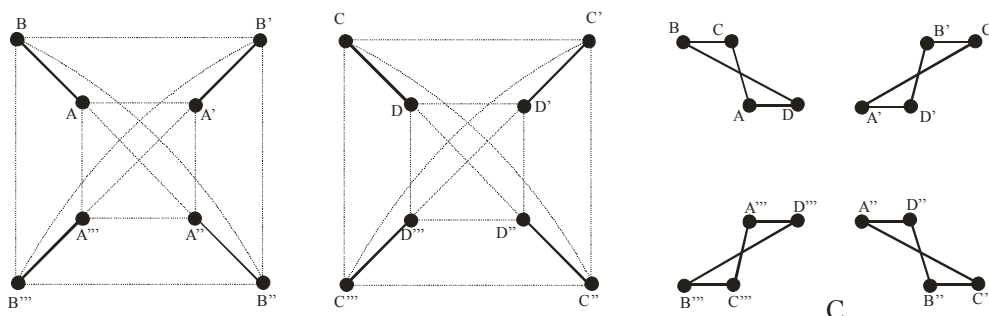
priestorové konfigurácie štvorstena. Nech účelová funkcia energie vyjadrená pomocou výrokovej logiky má tvar (pozri obr. 6.18)

$$\begin{aligned}
 E_{\tau} = & -\sum_{X,Y} val_{\tau}(x \wedge y) - \sum_{X',Y'} val_{\tau}(x' \wedge y') - \sum_{X'',Y''} val_{\tau}(x'' \wedge y'') - \sum_{X''',Y'''} val_{\tau}(x''' \wedge y''') \\
 & + 2\sum_{X,Y'} val_{\tau}(x \wedge y') + 2\sum_{X,Y''} val_{\tau}(x \wedge y'') + 2\sum_{X,Y'''} val_{\tau}(x \wedge y''') \\
 & + 2\sum_{X',Y''} val_{\tau}(x' \wedge y'') + 2\sum_{X',Y'''} val_{\tau}(x' \wedge y''') + 2\sum_{X'',Y'''} val_{\tau}(x'' \wedge y''')
 \end{aligned} \tag{6.37}$$

Kde prvá (druhá, tretia a štvrtá) sumácia obsahuje excitačné hrany prvého (druhého, tretieho a štvrtého) štvorstena. Ďalšie sumácie obsahujú inhibičné hrany pre vybranú dvojicu (1-2, 1-3, 1-4, 2-3, 2-4 a 3-4) štvorstenov.



Obrázok 6.17. Príklad štvorstena, ktorý obsahuje štyri vrcholy a kde môžu byť rozlíšené štyri priestorové konfigurácie.



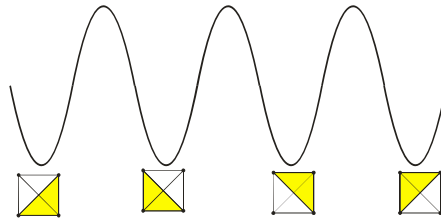
Obrázok 6.18. Konštrukcia Hopfieldovej siete pre rozlíšenie priestorových konfigurácií štvorstena, ktorá obsahuje obsahuje 16 vrcholov, ktoré sú rozdelené po štvorciach, pričom každá reprezentuje možnú konfiguráciu štvorstena. Diagrami A a B určujú hornú resp. dolnú časť štvorstenov, diagram C znázorňuje excitačné spoje medzi hornou a dolnou časťou štvorstenu. Hrany siete reprezentované pevnými čiarami sú ohodnotené váhovými koeficientami $w = 1$ (excitačné hrany), a hrany reprezentované prerušovanými čiarami medzi dvojicou vrcholov z rôznych štvorstenov sú ohodnotené váhovým koeficientom $w = -2$ (inhibičné hrany).

Použitím (6.30a) prepíšeme účelovú funkciu energie do konečného (algebraického) tvaru

$$\begin{aligned}
 E_{\tau} = & -\sum_{X,Y} x \cdot y - \sum_{X',Y'} x' \cdot y' - \sum_{X'',Y''} x'' \cdot y'' - \sum_{X''',Y'''} x''' \cdot y''' \\
 & + 2\sum_{X,Y'} x \cdot y' + 2\sum_{X,Y''} x \cdot y'' + 2\sum_{X,Y'''} x \cdot y''' \\
 & + 2\sum_{X',Y''} x' \cdot y'' + 2\sum_{X',Y'''} x' \cdot y''' + 2\sum_{X'',Y'''} x'' \cdot y'''
 \end{aligned} \tag{6.38}$$

Riešením takto špecifikovanej Hopfieldovej siete získame štyri optimálne stavy s rovnakou hodnotou energie rovnou $E = -6$ (pozri obr. 6.19)

$$\begin{aligned} \tau_{opt}^{(1)} &= (1111, 0000, 0000, 0000), \tau_{opt}^{(2)} = (0000, 1111, 0000, 0000) \\ \tau_{opt}^{(3)} &= (0000, 0000, 1111, 0000), \tau_{opt}^{(4)} = (0000, 0000, 0000, 1111) \end{aligned} \quad (6.39)$$



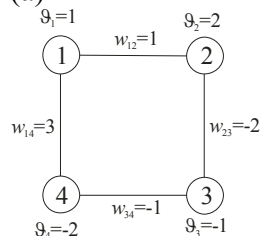
Obrázok 6.19. 4-minimový povrchu účelovej funkcie energie (6.38), pomocou ktorej interpretujeme percepciu priestorovej interpretácie štvorstena pomocou štyroch rôznych konfigurácií.

Sumarizujúc výsledky tohto jednoduchého ilustračného príkladu štvorstena, podobne ako pre neckerovu kocku, dostávame aj pre štvorsten výsledok, že jednoduchá Hopfieldova sieť je schopná interpretovať štyri rôzne priestorové percepcie štvorstena pomocou povrchu účelovej funkcie energie, na ktorej existujú štyri globálne minima, ktoré sú priradené jednotlivým priestorovým konfiguráciám štvorstena.

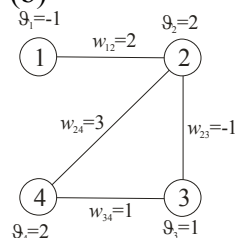
Úlohy

Úloha 6.1. Zostrojte tabuľku (pozri Tab. 6.1) hodnôt energie pre všetky možné stavy Hopfieldovej siete, ktorá obsahuje štyri neuróny. Pomocou tejto tabuľky zostrojte stavový diagram siete (pozri obr. 6.4). Hopfieldova sieť je špecifikovaná dvoma spôsobmi

(a)



(b)

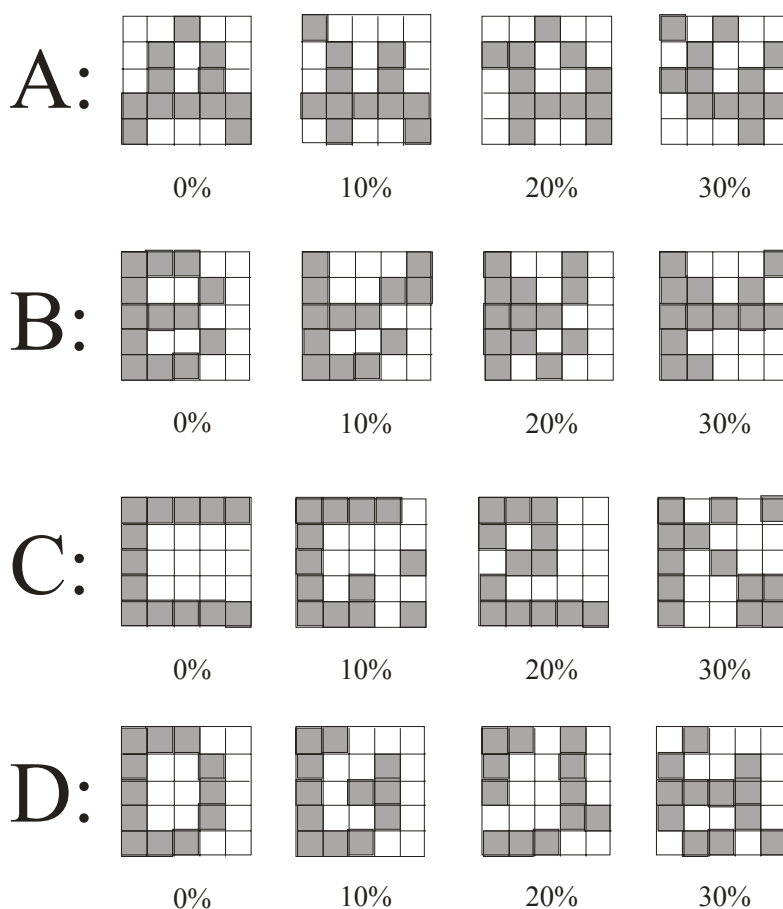


Úloha 6.2. Implementujte program pre Hopfieldovu sieť s asynchrónnym obnovovacím spôsobom aktivít, pričom váhové a prahové koeficienty a počiatkový vektor aktivít sú špecifikované pomocou vstupu. Problém nájdenia stavu s minimálnou hodnotou energie riešte pomocou heuristiky horolezeckého algoritmu (pozri knižku „Evolučné algoritmy“ [xx]).

Vykonajte ilustračné výpočty, pričom získané výsledky ilustrujte graficky a diskutujte vlastnosti siete pre zvolené ilustračné príklady.

Úloha 6.3. Podobne ako v predchádzajúcom príklade 6.2, vlen s tým rozdielom, že stratégia pre nájdenia stavu s minimálnou hodnotou energie je simulované žihanie (pozri kapitolu 6.3.1).

Úloha 6.4. Pomocou programu z úlohy 6.2 alebo 6.3 implementujte program pre Hopfieldovu sieť, po mocou ktorého ste schopný nájsť stabilný stav siete. Váhové koeficienty a prahové koeficienty nech sú zadané ako vstup programu. Pomocou tohto programu vykonaj ilustračné výpočty pre rekognoskáciu štyroch písmen abecedy A, B, C a D, ktoré sú zadané na ortogónálnej štvorcovej mriežke 5×5 a sú reprezentované binárnymi vektormi dĺžky 25 (pozri priložený obrázok).



Podľa formuly (6.11) zostrojte maticu váhových koeficientov pre Hopfieldovu sieť, ktorá v tomto prípade implementuje jednoduchú asociačnú pamäť. Ako narážku (počiatočný stavový vektor) používajte vektory, ktoré sú zadané v 2-4 stĺpci pomocou stochastického zašumenia pôvodného binárneho vektora z prvého stĺpca tak, že ho porušíte 10%, 20%, a 30% šumom. Pre každú takto zostrojenú narážku zostrojte trajektóriu pomocou asynchrónneho obnovovania. Zistite, pre aký stupeň zašumenia pôvodného binárneho vektora reprezentujúceho písmeno, Hopfieldova sieť ako stabilný stav poskytuje pôvodné písmeno. Znázornite výsledky graficky.

Úloha 6.5. Táto úloha je podobná predchádzajúcej úlohe 6.4, v tomto príklade sa bude študovať vplyv zániku neurónov a/alebo spojov medzi neurónmi na schopnosti nájsť korektnú asociáciu písmena k danému vstupu. Diskutujte túto zaujímavú vlastnosť Hopfieldovej siete pre rôzne stupne „zániku“ neurónov a spojov, vykonajte grafickú ilustráciu výsledkov a diskutujte získané výsledky.

Úloha 6.6. Implementujte program pre Hopfieldovu sieť, ktorá je schopná hľadať binárny vektor dĺžky 100 (pozri kapitolu 6.4.1), pričom si zvolíte svoju stratégiu (napr. horolezecký algoritmus, simulované žihania, a pod.) pre hľadanie stavu s minimálnou hodnotou energie. Získané výsledky ilustrujte graficky a diskutujte.

Úloha 6.7. Podobná úloha ako 6.6, v tomto prípade je potrebné implementovať neurónovú sieť pre riešenie úlohy obchodné cestujúceho (pozri kapitolu 6.4.2). Nech táto úloha je špecifikovaná na ortogonálnej mriežke 10×10 , pričom vzdialenosť sa počíta pomocou L_1 (Hammingovej) normy. Získané výsledky ilustrujte graficky pomocou znázornenia ciest obchodného cestujúceho v jednotlivých etapách algoritmu (hovoríme, že z chaotickej cesty emerguje optimálna cesta – poriadok).

Úloha 6.8. Implementujte program pre Hopfieldovu sieť, ktorá je schopná riešiť úlohu nájdenia optimálneho modelu pre penalizačnú logiku (pozri kapitolu 6.5). Penalizačná logika, Nixonov diamant

Úloha 6.9. Implementujte program pre Hopfieldovu sieť, ktorá je schopná riešiť úlohu Neckerovej kocky (pozri kapitolu 6.6). Získané výsledky ilustrujte graficky a diskutujte.

Úloha 6.10. Implementujte program pre Hopfieldovu sieť, ktorá je schopná riešiť úlohu štvorstena (pozri kapitolu 6.6.1). Získané výsledky ilustrujte graficky a diskutujte