Sent for publication to Neural Network World, February 2004

HOLOGRAPHIC REDUCED REPRESENTATION IN ARTIFICIAL INTELLIGENCE AND COGNITIVE SCIENCE

Vladimír Kvasnička[*]

Abstract. Holographic reduced representation is based on a suitable distributive coding of structured information in conceptual vectors, which elements satisfy normal distribution $N(0,1/n)$. Existing applications of this approach concern various models of associative memory that exploit a simple algebraic operation of scalar product of distributed representations to measure an overlap between two structured concepts. We have described here a method that uses this representation to model a similarity between different concepts and an inference process based on the rules *modus ponens* and *modus tollens*.

Key words: *Holographic, reduced representation, distributed representation, convolution.*

1. Introduction

The goal of this paper is to highlight an alternative approach, which may overcome the gap between the symbolic and subsymbolic approaches in the description and interpretation of cognitive activities of the human brain [7-9]. We shall show, that the application of a distributed representation allows to integrate connectionism and cognitivism. The mental representations (symbols) are specified by distributed patterns of neural activities, while over these distributed patterns we can introduce formal algebraic operations, which not only allow to mathematically model cognitive operations, but also allow to simulate processes of storage and retrieval of information from memory.

[*] Vladimír Kvasnička, Center of Artificial Intelligence, Faculty of Computer Science and Information Technologies and Department of Mathematics, Faculty of Chemical and Food Technology, Slovak Technical University, 812 37 Bratislava, Slovakia, email: vladimir.kvasnicka@stuba.sk.

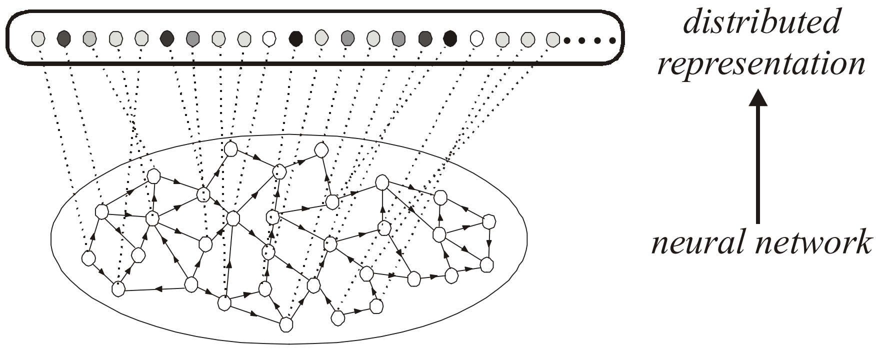

Figure 1. A visualization of the transition from neural network to distributed representation. The state of neural network in the time t is given by the activities of single neurons, which are determined by the activities in the previous time t-1 and by weight coefficients of single oriented connections. Using a certain abstraction, these activities can be ordered into one big one-dimensional array (vector) of real numbers (their size is determined by the level of gray of the corresponding component – neuron). In the distributed representation the architecture of the neural network is ignored, i.e. two randomly generated distributed representations must be understood as totally independent without mutual relations; their incidental connections derived from neural network are completely ignored. New unary and binary operations are introduced in the distributed representation, which enable to create new distributed representations from the original ones.

We shall study a nontraditional style of performing calculation by using distributed patterns. This approach is substantially different from classical numeric and symbolic computations and it is a suitable model tool for understanding of the global properties of neural networks. We shall demonstrate that such a „neurocomputing" is based on extensive randomly created patterns (represented by multidimensional vectors with random entries), see fig. 1.

This approach, which basic principles were formulated already at the end of sixties [2,4], was crowned by Tony Plate [5,6] in works on „*holographic reduced representation*" (HRR). Kanerva [3] in the middle of nineties proposed a certain alternative to HRR, which is based on randomly generated binary vectors. Our contribution to the development of HRR consists in its application to modeling of cognitive processes of reasoning by application of rules modus ponens a modus tollens.

2. A mathematical formulation of holographic representation

The aim of this chapter is a presentation of basic properties of a holographic representation, which was developed by Plate [5,6]. Its basic notion is a *conceptual vector*, which is represented by an *n*-dimensional vector $\boldsymbol{a} \in R^n \Rightarrow \boldsymbol{a} = (a_0, a_1, ..., a_{n-1})$. Components of this vector are random numbers with a standard normal distribution, $a_i = N(0,1/n)$, where $N(0,1/n)$ is a random number with a mean equal to 0 and a standard deviation $1/n$. A binary operation „*convolution*" is defined over conceptual vectors, which assigns to a couple of vectors a third vector, $\boldsymbol{c} = \boldsymbol{a} \otimes \boldsymbol{b}$, its entries are determined by

$$c_i = \sum_{j=0}^{n-1} a_j b_{[i-j]} \quad (i = 0,1,...,n-1) \tag{1}$$

where the index in the square brackets, [*k*], is defined using a *modulo n* operation. The operation of convolution is commutative, and associative. In general, almost for each conceptual vector $\boldsymbol{a}$ there exist an inverse vector $\boldsymbol{a}^{-1}$, $\boldsymbol{a}^{-1} \otimes \boldsymbol{a} = \mathbf{1}$. A unary operation over conceptual vectors is the so-called *involution*
$\boldsymbol{b} = \boldsymbol{a}^* = \left(a_{[0]}, a_{[-1]}, ..., a_{[-n+2]}, a_{[-n+1]}\right)$. It is possible to prove, that the involution $\boldsymbol{c}^*$ is approximately equal to an inverse vector $\boldsymbol{c}^{-1}$, $\boldsymbol{c}^* \otimes \boldsymbol{c} \approx \mathbf{1}$. According to this important property it is possible to decode the original information from the complex conceptual vectors. *Reconstruction of* $\boldsymbol{x}$ from $\boldsymbol{c}\otimes\boldsymbol{x}$ is based on the following property:

$$\tilde{\boldsymbol{x}} = \boldsymbol{c}^* \otimes (\boldsymbol{c} \otimes \boldsymbol{x}) = (\boldsymbol{c}^* \otimes \boldsymbol{c}) \otimes \boldsymbol{x} \approx \mathbf{1} \otimes \boldsymbol{x} = \boldsymbol{x} \tag{2}$$

An *overlap* of the resulting vector $\tilde{\boldsymbol{x}}$ with the original vector $\boldsymbol{x}$ is determined from a scalar product by

$$-1 \le overlap(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \frac{\boldsymbol{x} \cdot \tilde{\boldsymbol{x}}}{|\boldsymbol{x}||\tilde{\boldsymbol{x}}|} \le 1 \tag{3}$$

where the inequalities result directly from the Schwartz's inequality from linear algebra. The closer this value is to its maximum value, the more similar the vectors $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ are.

Let's turn our attention to the second possibility of the verificati on of the formula (2) with the application of the approach called the „*superposition memory*". Let us have a set containing *p*+*q* randomly generated conceptual vectors, $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_p, \boldsymbol{x}_{p+1}, ..., \boldsymbol{x}_{p+q}\}$. Using the first *p* vectors from X allows us to define a *memory vector* $\boldsymbol{t}$ as their sum

$$\boldsymbol{t} = \sum_{i=1}^{p} \boldsymbol{x}_i \tag{4}$$

The vector $\boldsymbol{t}$ represents a superposition memory, which by a simple additive way contains vector from the set *X*. The decision, whether some vector $\boldsymbol{x} \in X$

is contained in $\boldsymbol{t}$ must be based on the value of the overlap (3), $overlap(\boldsymbol{x},\boldsymbol{t})$. If this value is greater than a predefined threshold value, $overlap(\boldsymbol{x},\boldsymbol{t}) \geq \vartheta$, then the vector $\boldsymbol{x}$ is included in the superposition memory $\boldsymbol{t}$, in the opposite case, if $overlap(\boldsymbol{x},\boldsymbol{t}) < \vartheta$, then the vector $\boldsymbol{x}$ is not included in $\boldsymbol{t}$ (see fig. 2).

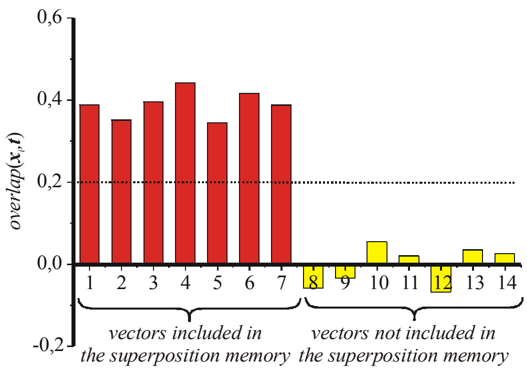

Figure 2. Illustration of the superposition memory for the first 7 vectors of the set *X*, which contains 14 randomly generated conceptual vectors of the dimension *n*=1024. The threshold value ϑ can be in this case set to 0.2.

In this simple illustrative example, the used method for a reconstruction of conceptual vectors, which can „appear“ in some other different complex conceptual vector (which can be the result of complicated previous calculations – transformations) is based on the concept of overlap between reconstructed and required vectors. The used method is called „*clean-up*“ [5,6] and it is specified as follows: Let us have a set of vectors $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ and some vector $\boldsymbol{t}$. We face the decision, whether the memory vector (trace) $\boldsymbol{t}$ contains a superposition component, which is similar (or which is not similar) to some vector from the set *X*. This problem can be solved by calculating the so called overlap (3), formally

$$\boldsymbol{x} \approx \boldsymbol{t} = \begin{cases} yes & \left(overlap(\boldsymbol{x},\boldsymbol{t}) \geq \vartheta\right) \\ no & \left(overlap(\boldsymbol{x},\boldsymbol{t}) < \vartheta\right) \end{cases} \tag{5}$$

where ϑ is a chosen *threshold value* of acceptance of the size of the overlap as the positive answer. The result of this cleaning-up process is a subset of vectors

$$X(\boldsymbol{t}) = \{\boldsymbol{x} \in X; \boldsymbol{x} \approx \boldsymbol{t}\} \subseteq X \tag{6}$$

We can put the question also in a rather different form, which is, whether the memory vector $\boldsymbol{t}$ is similar to any of the vectors from the set *X*? The answer to this more general question shall be decided from the maximum value of the overlap

$$overlap\left(\boldsymbol{t}, X\right) = \max_{x\in X} overlap\left(\boldsymbol{t}, \boldsymbol{x}\right) \tag{7}$$

Then we can rewrite (5) in the form

$$\boldsymbol{t} \in X = \begin{cases} yes & \left(overlap\left(\boldsymbol{t}, X\right) \geq \vartheta\right) \\ no & \left(overlap\left(\boldsymbol{t}, X\right) < \vartheta\right) \end{cases} \tag{8}$$

3. Associative memory

The construction of the associative memory belongs to the main results of the holographic reduced representation, which can be further generalized by so called chunking. Let us have a set of conceptual vectors $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ and a training set $A_{train} = \{\boldsymbol{c}_i / \boldsymbol{x}_i \,;\, i = 1, 2, ..., m\}$, which contains $m<n$ associated couples of conceptual vectors $\boldsymbol{c}_i / \boldsymbol{x}_i$ **,** where $\boldsymbol{c}_i$ is the *input* to the associative memory (*cue*) and $\boldsymbol{x}_i$ is the output from the memory. Let's create a memory vector $\boldsymbol{t}$ representing *the associative memory* created from the training set A_{train}

$$\boldsymbol{t} = \boldsymbol{c}_1 \otimes \boldsymbol{x}_1 + ... + \boldsymbol{c}_m \otimes \boldsymbol{x}_m = \sum_{i=1}^{m} \boldsymbol{c}_i \otimes \boldsymbol{x}_i \tag{9}$$

Let us suppose, that we know in advance only the inputs $\boldsymbol{c}_i$ to the associative memory, we do not know the possible outputs from the set $X_{train} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m\}$. The response of the associative memory to the input - clue $\boldsymbol{c}_i$ is determined by the process of „*cleaning-up*" represented by the formula (8). In the first step we shall calculate the vector $\tilde{\boldsymbol{x}}_i = \boldsymbol{c}_i^* \otimes \boldsymbol{t}$, then by a process based on the maximum value of the overlap we shall find whether $\tilde{\boldsymbol{x}}_i \approx \boldsymbol{x}_i \in X$

$$overlap\left(\tilde{\boldsymbol{x}}_i, X\right) = \max_{x\in X_{train}} overlap\left(\tilde{\boldsymbol{x}}_i, \boldsymbol{x}\right) \tag{10}$$

1st illustrative example

This example uses only the training set $A_{train} = \{\boldsymbol{c}_i / \boldsymbol{x}_i \,;\, i = 1, 2, ..., m\}$, which is randomly generated for m=8, while the dimension of conceptual vectors is n=1000. For each associated couple $\boldsymbol{c}_i / \boldsymbol{x}_i$ there are calculated $\boldsymbol{t}_i = \boldsymbol{c}_i \otimes \boldsymbol{x}_i$. The values of $overlap\left(\boldsymbol{c}_i^* \otimes \boldsymbol{t}_i, \boldsymbol{x}_j\right)$ are presented in the table

	1	2	3	4	5	6	7	8
1	0.717	-0.018	0.014	0.027	-0.014	-0.019	-0.024	0.013
2	-0.039	0.738	0.015	0.014	0.002	0.023	-0.005	-0.016
3	-0.027	-0.017	0.646	0.004	-0.115	-0.008	0.014	0.003
4	0.007	0.003	-0.018	0.687	-0.153	0.000	-0.005	0.001
5	-0.004	0.004	-0.018	-0.008	0.707	0.041	-0.033	-0.006
6	-0.014	0.025	-0.014	-0.013	-0.030	0.714	0.000	-0.005
7	0.029	0.008	-0.003	-0.003	0.012	0.009	0.707	0.018
8	-0.003	-0.010	0.008	-0.018	0.003	-0.026	0.008	0.589

It is evident from the table, that the overlaps are sufficiently great just for diagonal values, while the non-diagonal overlaps are smaller by an order of magnitude. We can therefore unambiguously decide from the overlap, whether $\boldsymbol{c}_i^* \otimes \boldsymbol{t}_i \approx \boldsymbol{x}_i$ is associated with the cue $\mathbf{c}_i$.

2nd illustrative example

In this illustrative example we shall use the training set $A_{train} = \{\boldsymbol{c}_i / \boldsymbol{x}_i\}$, generated for m=10 associated couples – vectors of dimension n=1000. This memory is represented by a memory vector $\boldsymbol{t} = \boldsymbol{c}_1 \otimes \boldsymbol{x}_1 + \ldots + \boldsymbol{c}_m \otimes \boldsymbol{x}_m$. The following table shows 20 experiments of „clean up“, where we used with a 50% probability as an associative entry a vector $\boldsymbol{c}_i$ from the training set or a randomly generated conceptual vector. The table contains maximal values of overlaps (7), by which we can unambiguously determine, whether the used input has an associated counterpart in the training set.

#	max. overlap	Input index	index of output with max.overlap
1	0.311	6	6
2	0.047	rand. gener.	nonexistent
3	0.383	5	5
4	0.373	10	10
5	0.316	3	3
6	0.397	4	4
7	0.074	rand. gener.	nonexistent
8	0.065	rand. gener.	nonexistent
9	0.069	rand. gener.	nonexistent
10	0.039	rand. gener.	nonexistent
11	0.344	7	7
12	0.402	8	8
13	0.032	rand. gener.	nonexistent
14	0.073	rand. gener.	nonexistent
15	0.017	rand. gener.	nonexistent

16	0.004	rand. gener.	Nonexistent
17	0.033	rand. gener.	nonexistent
18	0.056	rand. gener.	nonexistent
19	0.373	10	10
20	0.037	rand. gener.	nonexistent

It follows from the table, that the associative memory with the clean up process is unambiguously identifying, the values of a maximum overlap for conceptual vectors well specify the existence (or nonexistence) of corresponding associative outputs.

4. Coding of relations

Holographic reduced representation can also provide a suitable means for encoding relations (predicates). Let us study a binary relation $P(x,y)$, when the Pascal code is used, this relation is formally specified by the head

$$\mathit{function}\ P\left(x : \mathit{type}_1;\ y : \mathit{type}_2\right) : \mathit{type}_3 \tag{11}$$

The domains, over which the single arguments of the relation are defined, are specified by the types $type_1$ and $type_2$, respectively; similarly also the relation P itself is understood as a function, which domain of values is specified by the type $type_3$. In many cases the domain of variables and also the domain of the relation itself are equal to each other; therefore their specifications can be omitted, which substantially reduces the holographic representation of relations. The simplified form of relation (11) looks as $\mathit{function}\ P\left(x;\ y\right)$, where we know in advance the type of variables x, y, and also the type of the relation P itself. The holographic representation of the relation (11) can have the following form

$$\begin{gathered}\boldsymbol{t} = \boldsymbol{P} + \boldsymbol{variable}_1 + \boldsymbol{variable}_2 + \boldsymbol{P} \otimes \\ \left(\boldsymbol{type}_3 + \boldsymbol{variable}_1 \otimes \left(\boldsymbol{x} + \boldsymbol{type}_1\right) + \boldsymbol{variable}_2 \otimes \left(\boldsymbol{y} + \boldsymbol{type}_2\right)\right)\end{gathered} \tag{12}$$

Their decoding is carried out step-by-step. In the first step we use the clean up procedure to recognize the name (identifier) of the relation $\boldsymbol{P}$ and also the names (identifiers) of its variables $\boldsymbol{x}$ and $\boldsymbol{y}$. In the second step we identify the type $\boldsymbol{type}_3$ of the relation $\boldsymbol{P}$, in the last, third step we use previous results to identify variables $\boldsymbol{x}$, $\boldsymbol{y}$ and also their types $\boldsymbol{type}_1$ and $\boldsymbol{type}_2$. In many cases the representation of the relation $P(x,y)$ is satisfactory in the following simplified form

$$\boldsymbol{t} = \boldsymbol{P} + \boldsymbol{variable}_1 \otimes \boldsymbol{x} + \boldsymbol{variable}_2 \otimes \boldsymbol{y} \tag{13}$$

The chosen method of the holographic representation of relation can be easily generalized also for more complex (higher order) relations, where the variables are predicates as well, e.g. $P(x,Q(y,z))$, where the „inner" predicate Q is characterized by

$$\textit{function}\ Q(y:type_3;\,z:type_4):type_5 \tag{14}$$

In order to create a higher order relation $P(x,Q(y,z))$, we must presume a type compatibility of the second variable of the relation P and of the type of relation Q, i.e. $type_2$=$type_5$. In the simplified approach, where all the types are the same, it is not necessary to distinguish the types of single variables and the relations themselves. A simplified holographic representation of relation (14) has the following form

$$\boldsymbol{t'} = \boldsymbol{Q} + \boldsymbol{variable}_1 \otimes \boldsymbol{y} + \boldsymbol{variable}_2 \otimes \boldsymbol{z} \tag{15}$$

By exchanging the representation (15) for the variable $\boldsymbol{y}$ in the representation (13) we may construct "nested" representation of the higher order relation $P(x,Q(y,z))$.

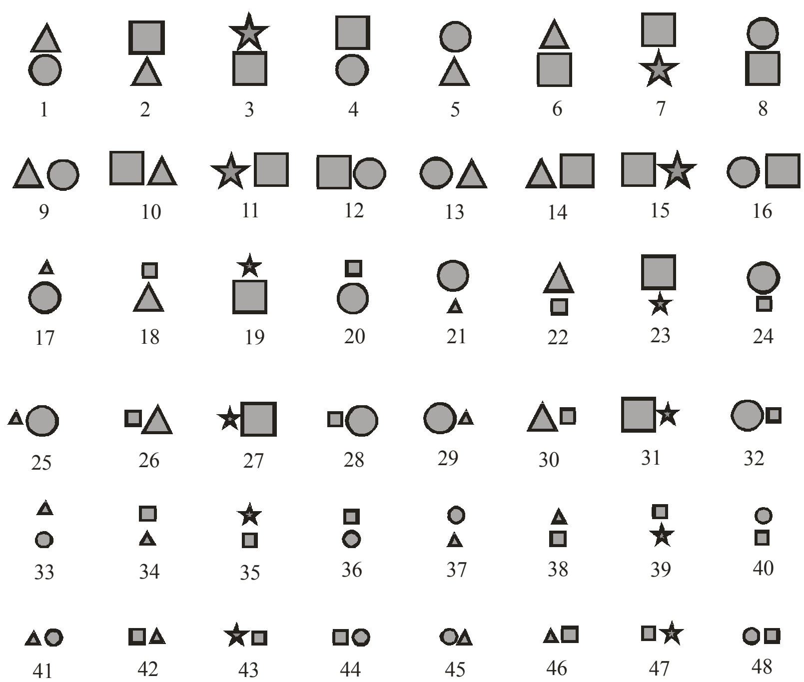

Figure 3. A set of 48 similar figures, which contain two objects, placed either next to each other, or above each other and the objects are either small or big. Every column contains a couple of similar objects, which differ only by their placement or size.

1st illustrative example –a similarity between geometric figures

In the figure 15 there are presented 48=6×8 geometric patterns, which contain either in horizontal or in vertical settings two objects, which moreover can be of two sizes, small and big. Let us mark holographic representations of corresponding atomic concepts as follows:

Objects: ***tr*** (triangle), ***sq*** (square), ***ci***(circle), ***st*** (star)
Unary relations: ***sm*** (small), ***lg*** (large)
Binary relations: ***hor*** (horizontal), ***ver*** (vertical)
Variables: $\boldsymbol{ver\_var_1}$ (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***)

Single figures from fig. 3 are characterized by relations given in the following table.

row	specification
1	***ver***(***lg***(***x***),***lg***(***y***))
2	***hor***(***lg***(***x***),***lg***(***y***))
3	***hor***(***sm***(***x***),***lg***(***y***)) and ***hor***(***lg***(***x***),***sm***(***y***))
4	***ver***(***sm***(***x***),***lg***(***y***)) and ***ver***(***lg***(***x***),***sm***(***y***))
5	***ver***(***sm***(***x***),***sm***(***y***))
6	***hor***(***sm***(***x***),***sm***(***y***))

Holographic representations of single cases from this table have the following form

$$\begin{aligned}
&\boldsymbol{t}_{1,\boldsymbol{x},\boldsymbol{y}} = \boldsymbol{ver} + \left\langle \boldsymbol{ver_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{ver_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \\
&\boldsymbol{t}_{2,\boldsymbol{x},\boldsymbol{y}} = \boldsymbol{hor} + \left\langle \boldsymbol{hor_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{hor_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \\
&\boldsymbol{t}_{3,\boldsymbol{x},\boldsymbol{y}} = \begin{cases} \boldsymbol{ver} + \left\langle \boldsymbol{ver_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{ver_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\ \boldsymbol{ver} + \left\langle \boldsymbol{ver_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{ver_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \end{cases} \\
&\boldsymbol{t}_{4,\boldsymbol{x},\boldsymbol{y}} = \begin{cases} \boldsymbol{hor} + \left\langle \boldsymbol{hor_var}_1 \otimes \boldsymbol{lg} \otimes \boldsymbol{x} + \boldsymbol{hor_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\ \boldsymbol{hor} + \left\langle \boldsymbol{hor_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{hor_var}_2 \otimes \boldsymbol{lg} \otimes \boldsymbol{y} \right\rangle \end{cases} \\
&\boldsymbol{t}_{5,\boldsymbol{x},\boldsymbol{y}} = \boldsymbol{ver} + \left\langle \boldsymbol{ver_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{ver_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle \\
&\boldsymbol{t}_{6,\boldsymbol{x},\boldsymbol{y}} = \boldsymbol{hor} + \left\langle \boldsymbol{hor_var}_1 \otimes \boldsymbol{sm} \otimes \boldsymbol{x} + \boldsymbol{hor_var}_2 \otimes \boldsymbol{sm} \otimes \boldsymbol{y} \right\rangle
\end{aligned} \tag{16}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are holographic representations of single objects (***tr***, ***sq***, ***ci***, ***st***) and the bracket $\langle \boldsymbol{u} \rangle$ indicates, that the vector $\boldsymbol{u}$ is normalized. The similarity between single figures is determined by the overlap of their holographic representations

$$similarity\left(X, X'\right) = overlap\left(\boldsymbol{t}, \boldsymbol{t}'\right) \tag{17}$$

The obtained results are shown in the fig. 4. The dominant feature controlling similarity value is the horizontal or vertical arrangement of objects. The overlap (i.e. also the similarity) between two figures, which have different arrangement is usually smaller than 0.1.

In general, holographic reduced representation allows fairly simple determination of similarity of objects specified by a predicate structure (19) or by its generalization through further nested predicates. This possibility opens new horizons on future developments in fundamental methods of search for similar objects or analogies, which are considered very difficult problems for artificial intelligence requiring special symbolic techniques [7].

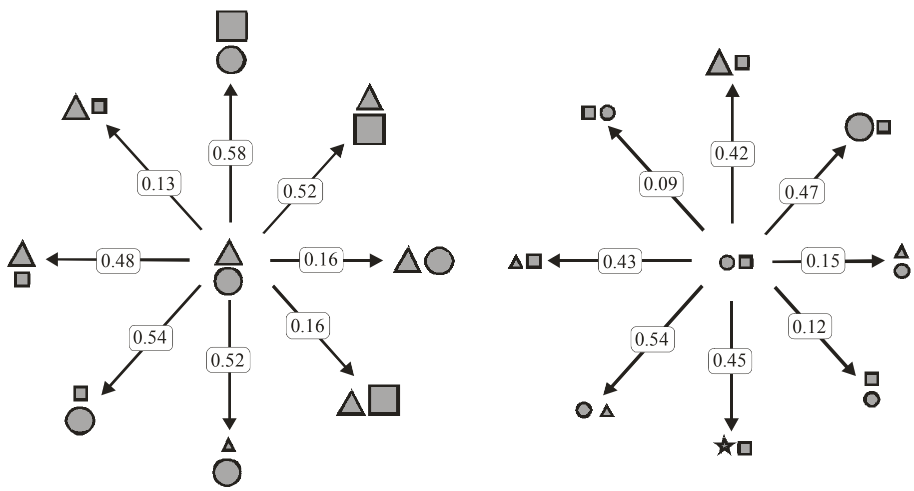

Figure 4. Illustrative presentation of similar figures for two chosen figures 1 and 48 (see fig. 3). Single arrows are marked by the overlap between the figures calculated by formula (50).

This simple illustrative example shows, that in the framework of the holographic distributed representation one can use (at least potentially) associative representations of the type (9), where associative cues correspond to numbers. It means that in this distributed approach there exists a possibility of associative simulation of an arbitrary function, which substantially increases the potential of the method to be used universally.

5. Reasoning by modus ponens and modus tollens

Simulation of reasoning processes (inference) belongs to the basic problems, which are repeatedly solved in artificial intelligence and cognitive science [7-10]. Fodor‘s critique [1] of connectionism (see ref. 9) was based precisely on the brash conclusion, that artificial neural networks are not able to simulate higher cognitive activities, which are purported to be an exclusive domain of the classical symbolic approach. This Fodor’s opinion was proved to be incorrect, further development of theory of neural networks showed, that connectionism is a universal computational tool, which does not have limits of applicability, it does not have domains of inapplicability, which would be forbidden to it. Of course, it can transpire, that in some domains its application is extremely cumbersome and exceedingly complicated, that there exist other approaches, which in the given domain provide substantially simpler and direct solution, than the one provided by neural networks.

In this chapter we shall show a possibility of representation of two basic modes of deductive reasoning of modal logic,

$$\begin{array}{l} p \Rightarrow q \\ \underline{p \qquad} \\ q \end{array} \quad \text{and} \quad \begin{array}{l} p \Rightarrow q \\ \underline{\overline{q} \qquad} \\ \overline{p} \end{array} \tag{18}$$

which are called *modus ponens* and *modus tollens*, respectively. These modes of reasoning are equivalent to the following tautologies of the predicate logic

$$\big((p \Rightarrow q) \wedge p\big) \Rightarrow q \tag{19a}$$

$$\big((p \Rightarrow q) \wedge \overline{q}\big) \Rightarrow \overline{p} \tag{19b}$$

Implication $'\Rightarrow'$ can be understood as a binary relation, which can be in holographic distribution represented like this

$$\boldsymbol{t}_{p\Rightarrow q} = \boldsymbol{impl} + \boldsymbol{var}_1 \otimes \boldsymbol{p} + \boldsymbol{var}_2 \otimes \boldsymbol{q} \tag{20}$$

which contains a sum of three parts, the first part specifies the type of relation (implication), the second and third parts specify the first (antecedent) resp. the second (consequent) variable of the relation of implication. This conceptual vector representing relation of implication can be transformed as follows

$$\tilde{\boldsymbol{t}}_{p\Rightarrow q} = \boldsymbol{t}_{p\Rightarrow q} \otimes \boldsymbol{T} \tag{21a}$$

where

$$\boldsymbol{T} = \boldsymbol{var}_1^* \otimes \boldsymbol{p}^* \otimes \boldsymbol{p}^* \otimes \boldsymbol{q} + \boldsymbol{var}_2^* \otimes \boldsymbol{q}^* \otimes \overline{\boldsymbol{q}}^* \otimes \overline{\boldsymbol{p}} \tag{21b}$$

The transformed representation of implication is represented by a sum of two associated couples

$$\tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \boldsymbol{p}^{*} \otimes \boldsymbol{q} + \overline{\boldsymbol{q}}^{*} \otimes \overline{\boldsymbol{p}} \tag{22}$$

which gives the holographic representation of the rules *modus ponens* and *modus tollens*

$$\boldsymbol{p} \otimes \tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \boldsymbol{q} \tag{23a}$$

$$\overline{\boldsymbol{q}} \otimes \tilde{\boldsymbol{t}}_{p\Rightarrow q} \approx \overline{\boldsymbol{p}} \tag{23b}$$

The first formula (23a) can be understood as a holographic representation of modus ponens (see (18) and (19a)), while the other formula is a holographic representation of modus tollens (see (18) and (19b)).

A similar result can be obtained also by an alternative approach, which is based on the disjunctive form of implication $(p \Rightarrow q) \equiv (\overline{p} \vee q)$. The distributed representation of implication in this alternative form can be expressed by

$$\boldsymbol{t}_{\overline{p}\vee q} = \boldsymbol{disj} + \boldsymbol{var}_1 \otimes \overline{\boldsymbol{p}} + \boldsymbol{var}_2 \otimes \boldsymbol{q} \tag{24}$$

By a transformation of this representation we can get (see (20))

$$\tilde{\boldsymbol{t}}_{\overline{p}\vee q} = \boldsymbol{t}_{\overline{p}\vee q} \otimes \boldsymbol{T} \approx \overline{\boldsymbol{p}} \otimes \overline{\boldsymbol{q}}^{*} + \boldsymbol{q} \otimes \boldsymbol{p}^{*} \tag{25}$$

where $\boldsymbol{T} = \boldsymbol{var}_1^{*}\overline{\boldsymbol{q}}^{*} + \boldsymbol{var}_2^{*}\boldsymbol{p}^{*}$. This transformation is much simpler than the one in the previous case (21b). The rules modus ponens and modus tollens are now realized by formulas similar to (23a-b). Moreover, we get also the following two „rules“

$$\boldsymbol{q}^{*} \otimes \tilde{\boldsymbol{t}}_{\overline{p}\vee q} \approx \boldsymbol{p}^{*} \tag{26a}$$

$$\overline{\boldsymbol{p}}^{*} \otimes \tilde{\boldsymbol{t}}_{\overline{p}\vee q} \approx \overline{\boldsymbol{q}}^{*} \tag{26b}$$

which remind us of the well known fallacies

$$\begin{array}{c} p \Rightarrow q \\ \underline{q \qquad} \\ p \end{array} \quad \text{and} \quad \begin{array}{c} p \Rightarrow q \\ \underline{\overline{p} \qquad} \\ \overline{q} \end{array} \tag{27}$$

that are known as „*affirming the consequent*“ and „*denying the antecedent*“, respectively. This fault is caused by the fact, that the transformed representations of implications $\tilde{\boldsymbol{t}}_{p\Rightarrow q}$ and $\tilde{\boldsymbol{t}}_{\overline{p}\vee q}$ are not identical, the representation $\tilde{\boldsymbol{t}}_{\overline{p}\vee q}$ leads to unexpected results (27), which represent erroneous modes of reasoning (which are however often used by people without knowledge of principles of logic).

6. Conclusions

Holographic reduced representation offers new unconventional solution to one of the basic problems of artificial intelligence and cognitive science, which is to find a suitable distributive coding of structured information (sequence of symbols, nested relational structures, etc.). The used distributed representation is based on a binary operation „convolution“ over a domain of n-dimensional randomly generated conceptual vectors, which elements satisfy normal distribution $N(0,1/n)$. Application of this distributed representation allows us to model various types of associative memory, which are represented by a conceptual vector and also to decode a memory vector, i.e. to determine the conceptual (atomic) vectors it is composed of. Such an analysis of the memory vector is carried out by a clean-up procedure that determines from the overlap of the vectors, which of the vectors is the most similar to the memory vector. Holographic reduced representation allows to measure similarity between two structured concepts by a simple algebraic operation of scalar product of their distributed representations. This fact can be very useful, when we want to model processes, which search through memory to find its similar (analogical) single components. In the last part of the paper we have demonstrated, that the holographic reduced representation may be used also to model an inference process based on the rules *modus ponens* and *modus tollens*.

Acknowledgments. This work was supported by Scientific Grant Agency of Slovak Republic under grants #1/0062/03 and #1/1047/04.

References

[1] Fodor, J.A., Pylyshyn, Z.W.: Connectionism and cognitive architecture: A critical analysis, *Cognition*, **28** (1988) 3-71.

[2] Gabor, D.: Holographic model for temporal recall, *Nature*, **217** (1968), 1288–1289.

[3] Kanerva, O.: 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.

[4] Murdock, B. B.: A theory for the storage and retrieval of item and associative information, *Psychological Review*, **89** (1982), 316–338.

[5] Plate, T. A.: *Distributed Representations and Nested Compositional Structure*. Department of Computer Science, University of Toronto. Ph.D. Thesis, 1994.

[6] Plate, T. A.: *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*, Stanford, CA: CSLI Publications, 2003.

[7] Russel, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*, New Jersey: Prentice Hall, 1995.

[8] 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.

[9] Smolensky, P.: On the proper treatment of connectionism, *The Behavioral and Brain Sciences,* **11** (1988), 1-74.

[10] Smolensky, P.: Tensor product variable binding and the representation of symbolic structures in connectionist networks, *Artificial Intelligence,* **46** (1990), 159-216.