

Holografická redukovaná reprezentácia v umelej inteligencii a kognitívnej vede

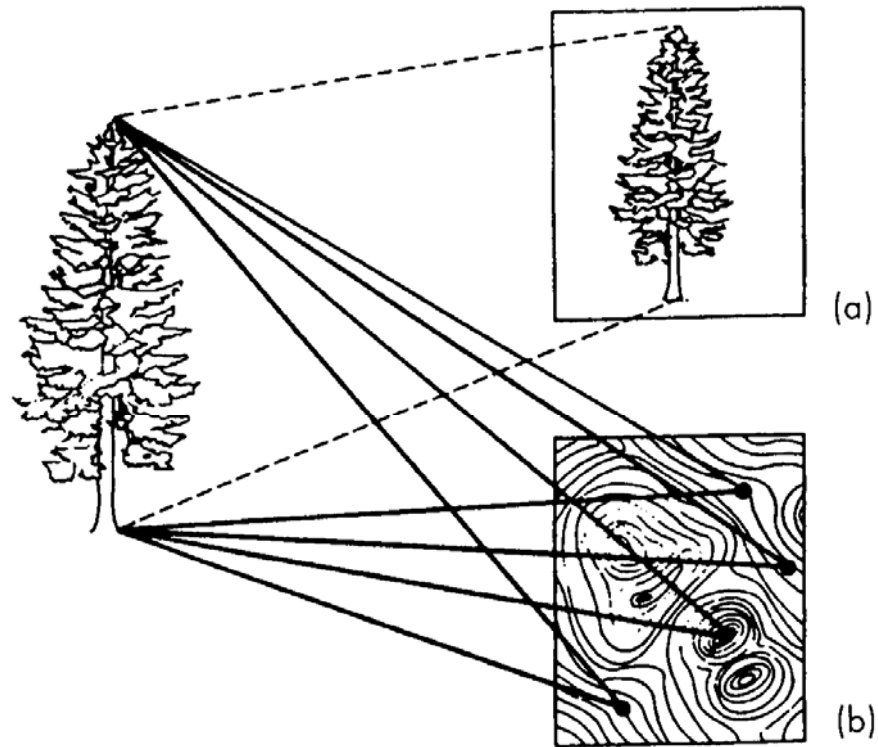
Vladimír Kvasnička

Ústav aplikovanej informatiky FIIT STU



November 2007

Hologram (Denis Gabor, 1948)



Hologram ako metafora pamäti

- P. J. van Heerden: A new Optical Method of Storing and Retrieving Informatoin. *Applied Optics* (1963).
- D. Gabor: Holographic model of temporal recall. *Nature* (1968).
- K. R. Pribram: The Neurophysiology of Remembering. *Scientific American* (1966).
- B. B. Murdock: Convolution and Correlation Perception and memory. *Perspectives of memory* (1979).
- D. Willshaw: Holography, Associative Memory and Inductive Generalization. *Parallel Models of Associative Memory* (1981)
- T. A. Plate: Holographic Reduced Representations. *IEEE Transactions on Neural Network* (1995).
- T. A. Plate: *Holographic Reduced Representation. Distributed Representation for Cognitive structures* (2003).

Základné postuláty holografického prístupu k pamäti

- Pamäť je v mozgu realizovaná distribuovaným spôsobom (je experimentálne zistené, že jednotlivé položky pamäti nemajú lokalistický charakter, lézie mozgu, Lashleyho experimenty na potkanoch)
- Graceful degradation of memory (pôvabná degradácia pamäti)
- *Distribuovaný charakter* pamäťových stôp v mozgu.
- Možnosť implementácie asociatívnej pamäti bez nutnosti „prehľadávania“

FTP anonymný server obsahuje niekoľko desiatok prác o HRR

`ftp://math.chtf.stuba.sk/pub/vlado/
DistributedRepresentation/`

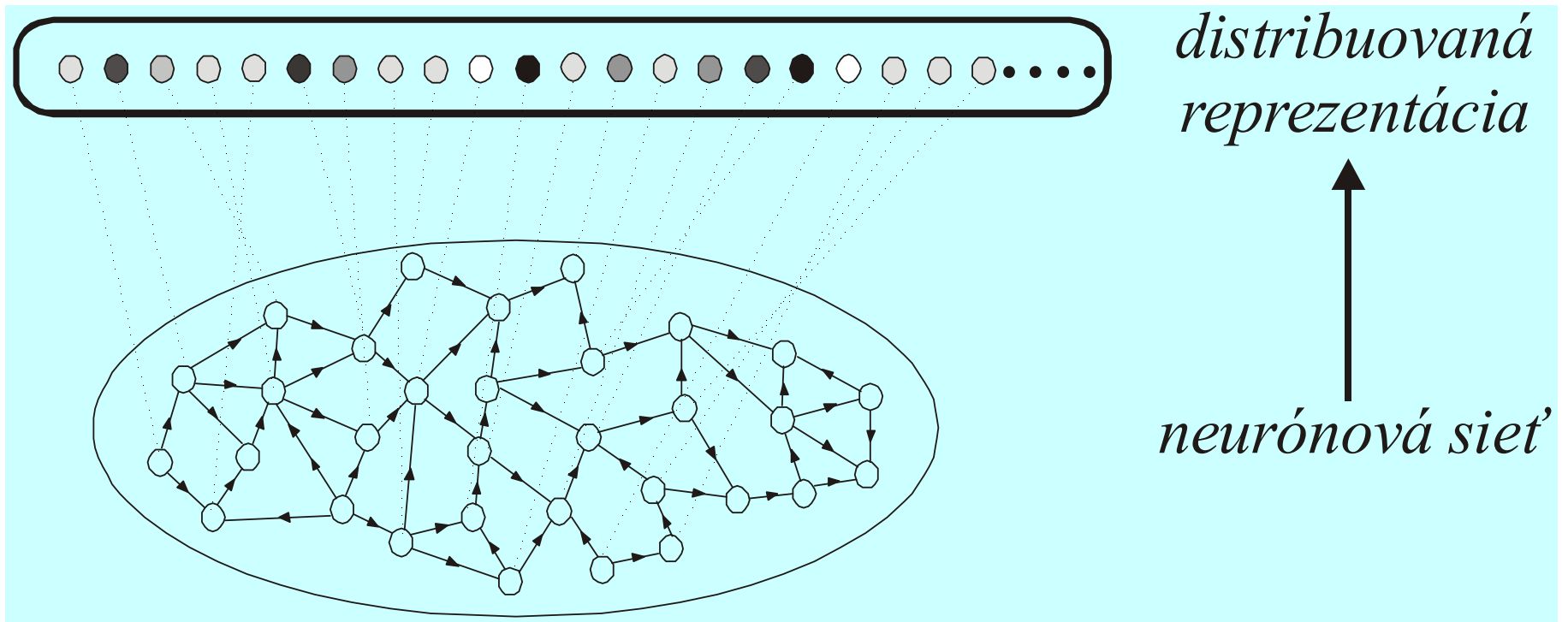
Čo je distribuovaná reprezentácia?

- **Distribuovaná reprezentácia** poznatkov (vedomostí, koncepcií,...) používa také kódovanie, ktoré je založené na neurónových sieťach - neurónové aktivity sú interpretované ako „*neurálalne koreláty*“ daných poznatkov.
- **Kognitívna aktivita** spočíva v časovom slede distribuovaných reprezentácií, pričom existujú pravidlá, ako meniť vybranú reprezentáciu na inú reprezentáciu.
- Tento prístup poskytuje **všeobecnú metódu**, ktorá kombinuje efektívnosť symbolického prístupu s
 - *robustnosťou,*
 - *učením a*
 - *schopnosťou generalizovať (indukciou)*subsymbolického prístupu.

Význam holografickej distribuovanej reprezentácie pre umelú inteligenciu a kognitívnu vedu

- Vzťah medzi *symbolickým* a *subsymbolickým* prístupom, alebo medzi *kognitivizmom* a *konekciónizmom*.
- Holografická distribuovaná reprezentácia ponúka určité *kompromisné riešenie* na polceste medzi kognitivizmom a konekciónizmom.
- Pre konekciónizmus je „endless“ problémom *interpretáciu koncepcie „symbolu“*, ktorá v kognitivistickom prístupe hrá centrálnu úlohu.
- Holografická distribuovaná reprezentácia symboly (koncepcie) vyjadruje pomocou mnohorozmerných „*konceptuálnych vektorov*“.
- Operácie nad symbolmi sú nahradené *numerickými operáciami* nad konceptuálnymi vektormi.
- Jednoduché modelovanie *asociačnej pamäti* (chunking effect)

Prechod od neurónovej siete k distribuovanej reprezentácii



Konceptuálne vektory

(1) n -rozmerny vektor

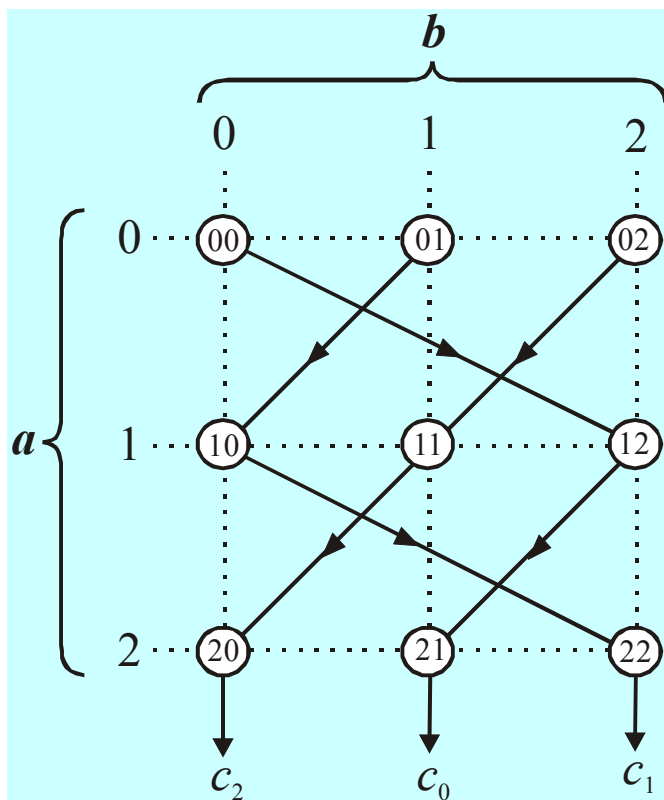
$$\mathbf{a} \in R^n \Rightarrow \mathbf{a} = (a_0, a_2, \dots, a_{n-1})$$

(2) náhodne generovaný

$$a_i = N\left(0, \frac{1}{n}\right) \quad \forall i \in \{0, 1, \dots, n-1\}$$

Cirkulárna konvolúcia

$$\otimes : R^n \times R^n \rightarrow R^n \quad c = a \otimes b$$



$$c_0 = a_0b_0 + a_1b_2 + a_2b_1$$

$$c_1 = a_0b_1 + a_1b_0 + a_2b_2$$

$$c_2 = a_0b_2 + a_1b_1 + a_2b_0$$

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

$$[k] = k \bmod n$$

Cirkulárna konvolúcia spĺňa tieto vlastnosti:

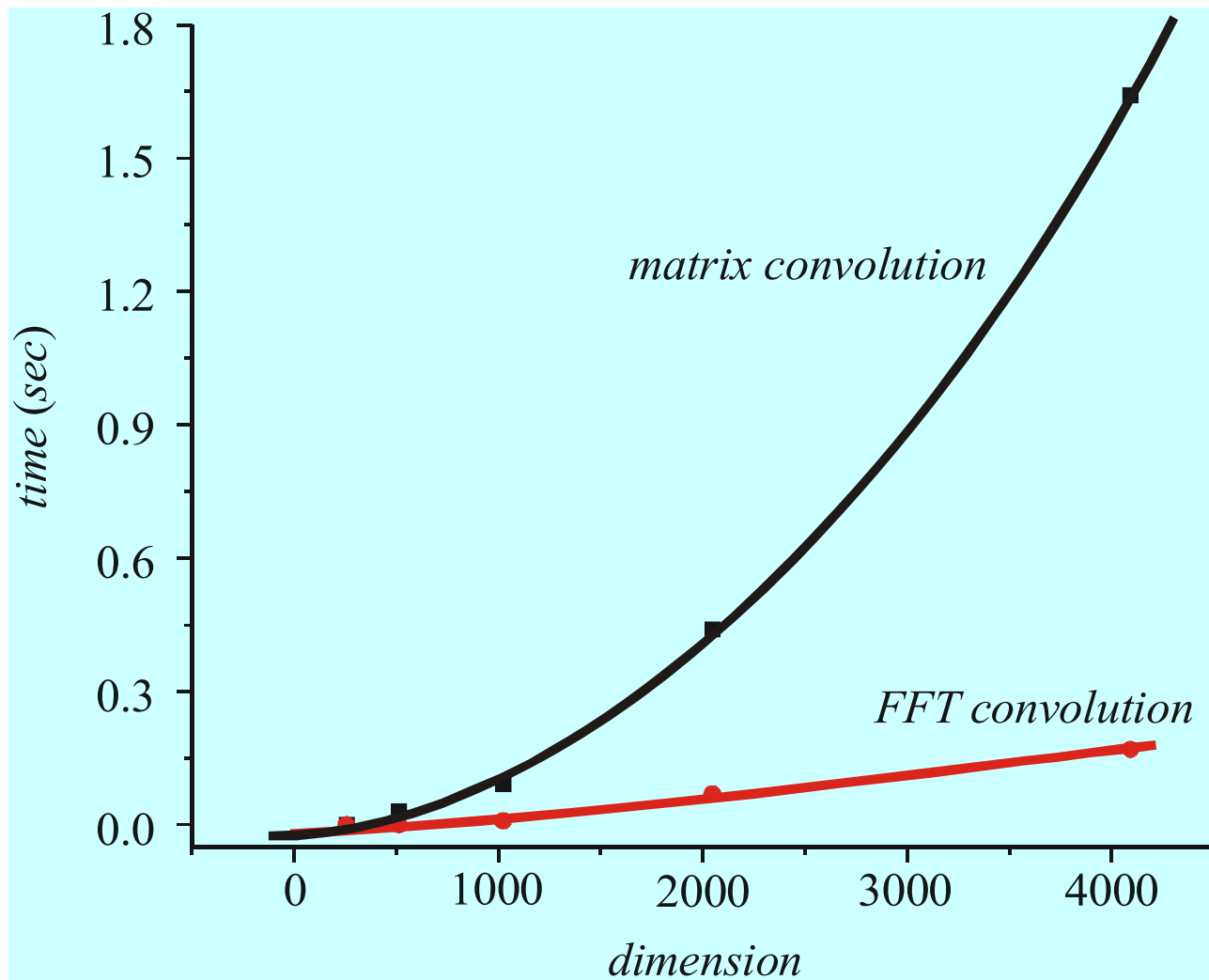
1. $\mathbf{a} \otimes \mathbf{b} = \mathbf{b} \otimes \mathbf{a}$
2. $(\mathbf{a} \otimes \mathbf{b}) \otimes \mathbf{c} = \mathbf{a} \otimes (\mathbf{b} \otimes \mathbf{c})$
3. $\mathbf{a} \otimes (\alpha \mathbf{b} + \beta \mathbf{c}) = \alpha (\mathbf{a} \otimes \mathbf{b}) + \beta (\mathbf{a} \otimes \mathbf{c})$
4. $\mathbf{1} \otimes \mathbf{a} = \mathbf{a} \quad (\mathbf{1} = (1, 0, \dots, 0))$
5. $\mathbf{a}^{-1} \otimes \mathbf{a} = \mathbf{1}$ (pre $|\text{circ}(\mathbf{a})| \neq 0$)

Jednoduchá implementácia konvolúcie vyžaduje výpočet n skalárnych súčinov vektora \mathbf{b} s vektorom \mathbf{a} , tj. táto naivná implementácia produkuje $o(n^2)$ algoritmus.

Implementácia konvolúcie pomocou rýchlej Fourierovej transformácie (FFT)

$$\mathbf{c} = \frac{1}{n} \mathbf{F}'((\mathbf{F}\mathbf{a}) \circ (\mathbf{F}\mathbf{b})) = \frac{1}{n} \mathbf{F}'(\text{diag}(\mathbf{F}\mathbf{a})(\mathbf{F}\mathbf{b}))$$

znižuje časovú náročnosť na $o(n \log(n))$.



Unárna operácia “involúcia”

$$(\)^* : R^n \rightarrow R^n$$

$$b = a^* = (a_{[0]}, a_{[-1]}, \dots, a_{[-n+2]}, a_{[-n+1]})$$

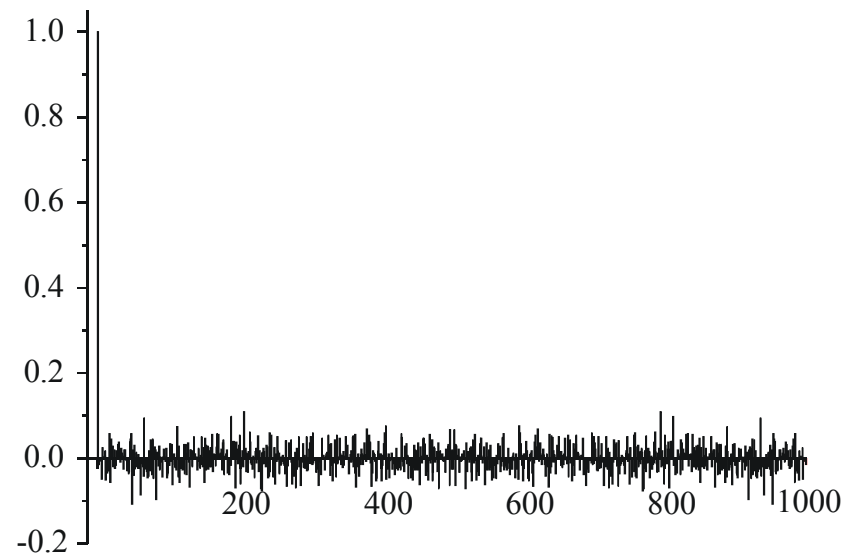
$$(a_0, a_1, a_2, \dots, a_{n-2}, a_{n-1})^* = (a_0, a_{n-1}, a_{n-2}, \dots, a_2, a_1)$$

Involúcia c^* je približne rovná inverznému vektoru c^{-1}

$$c^* \otimes c \approx \mathbf{1}$$

$$\begin{aligned}
(\mathbf{c}^* \otimes \mathbf{c})_i &= \sum_{k=0}^{n-1} c_k^* c_{[i-k]} = \sum_{k=0}^{n-1} c_{[-k]} c_{[i-k]} \\
&= \begin{cases} \mathbf{c} \cdot \mathbf{c} & (\text{for } i = 0) \\ \sum_{k=0}^{n-1} c_{[-k]} c_{[i-k]} & (\text{for } i > 0) \end{cases}
\end{aligned}$$

A histogram plot of single components of $\mathbf{c}^* \otimes \mathbf{c}$, where \mathbf{c} is randomly generated conceptual vector for $n=1000$.



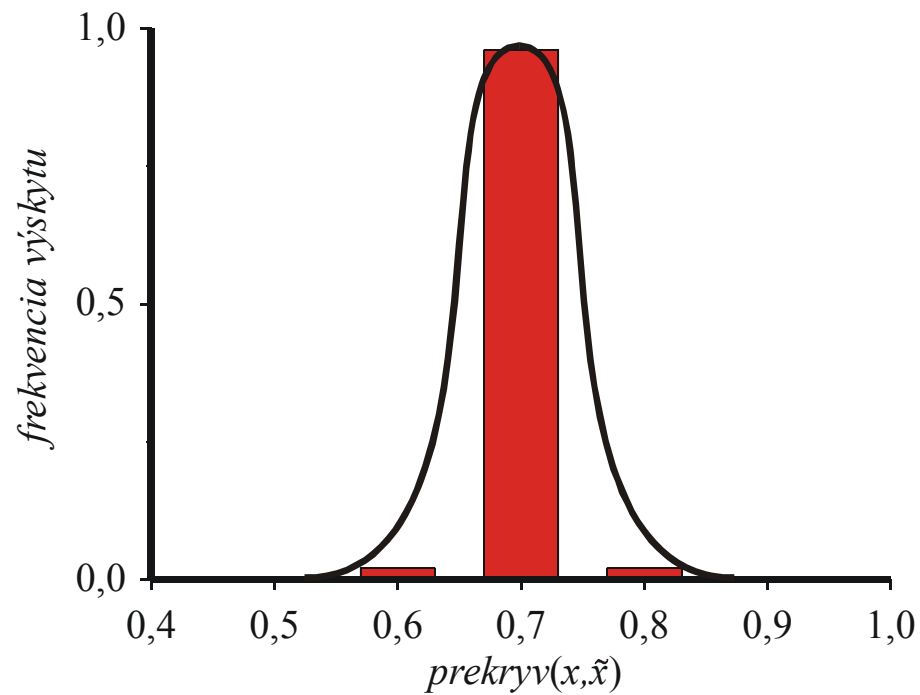
Rekonštrukcia \mathbf{x} z $\mathbf{c} \otimes \mathbf{x}$

$$\tilde{\mathbf{x}} = \mathbf{c}^* \otimes (\mathbf{c} \otimes \mathbf{x}) = (\mathbf{c}^* \otimes \mathbf{c}) \otimes \mathbf{x} \doteq \frac{1}{\mathbf{c} \cdot \mathbf{c}} \mathbf{1} \otimes \mathbf{x} = \frac{1}{\mathbf{c} \cdot \mathbf{c}} \mathbf{x}$$

pretože konvolúcia \mathbf{c}^* s vektorom $\mathbf{c} \otimes \mathbf{x}$ produkuje vektor $\tilde{\mathbf{x}}$, ktorý je blízky pôvodnému vektoru \mathbf{x} , $\tilde{\mathbf{x}} \approx \mathbf{x}$,

$$\frac{1}{(\mathbf{c} \cdot \mathbf{c})} \tilde{\mathbf{x}} = \begin{pmatrix} x_0 \\ x_1 + \eta_1 \\ \dots \\ x_{n-1} + \eta_{n-1} \end{pmatrix} = \mathbf{x} + \boldsymbol{\eta}$$

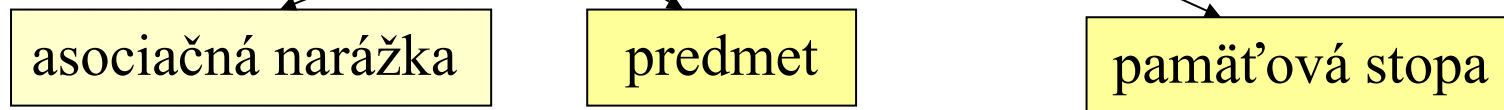
kde $\boldsymbol{\eta}$ je náhodný šum s normálnou distribúciou s nulovou strednou hodnotou a štandardnou odchýlkou podstatne menšou ako \mathbf{x} .



Histogram prekryvov medzi vektormi \tilde{x} a x (dimenzie $n=1000$) je s najväčšou frekvenciou okolo 0,7, z čoho vyplýva, že vektory \tilde{x} a x sú si podobné.

Asociatívna pamäť

$$A = \{c_i / x_i ; i = 1, 2, \dots, m\} \Leftrightarrow t = c_1 \otimes x_1 + \dots + c_m \otimes x_m$$



$$x = c^* \otimes t$$

$$prekryv = \max_{x' \in A} \frac{x' \cdot x}{|x'| |x|}$$

Proces „*upratovania*“ (clean-up) nájde pre „asociatívnu narážku“ (cue) vektor c priradený predmet x z množiny $A = \{c_i / x_i ; i = 1, 2, \dots, m\}$.

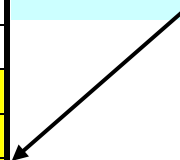
Ilustračný príklad

- 10 párov vektorov \mathbf{c} a \mathbf{x} (dĺžky $n=4000$) je náhodne generovaných pomocou $N(0,1/n)$,
- stopa t pre všetkých 10 párov je zostrojená,
- pre 20 pokusov aplikujeme s 50% pravdepodobnosťou buď narážku \mathbf{c} z A alebo použijeme náhodne generovaný vektor pomocou $N(0,1/n)$, potom $\tilde{\mathbf{x}} = \mathbf{c}^* \otimes t$ je vypočítaný,
- súhlas medzi $\tilde{\mathbf{x}}$ a \mathbf{x} z A je určený pomocou prekryvu

$$prekryv = \max_{\mathbf{x} \in A} \frac{\mathbf{x} \cdot \tilde{\mathbf{x}}}{\|\mathbf{x}\| \|\tilde{\mathbf{x}}\|}$$

#	prekryv	index narážky	index výstupu
1	0.311	6	6
2	0.047		
3	0.383	5	5
4	0.373	10	10
5	0.316	3	3
6	0.397	4	4
7	0.074		
8	0.065		
9	0.069		
10	0.039		
11	0.344	7	7
12	0.402	8	8
13	0.032		
14	0.073		
15	0.017		
16	0.004		
17	0.033		
18	0.056		
19	0.373	10	10
20	0.037		

hodnota prekryvu je menšia ako práhová hodnota



Alternatívna interpretácia asociatívnej pamäti – holografická transformácia

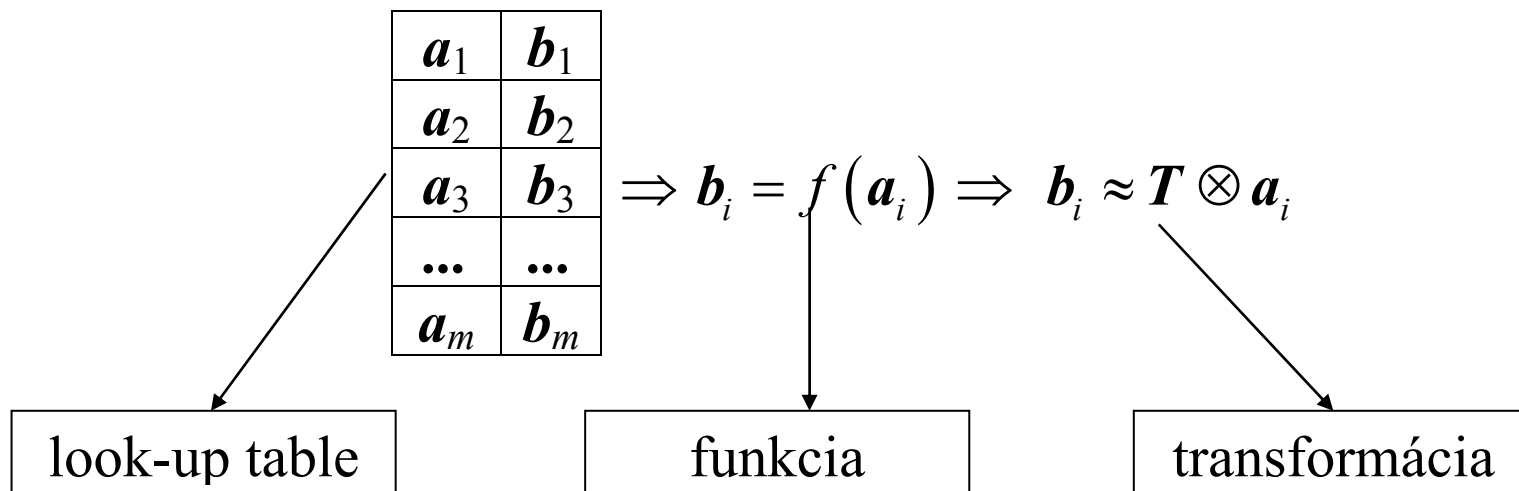
$$A_{train} = \{a_i / b_i ; i = 1, 2, \dots, m\}$$

Hladáme takú transformáciu, ktorá vyhovuje podmienke

$$b_i \approx a_i \otimes T \quad \forall i \in \{1, 2, \dots, m\}$$

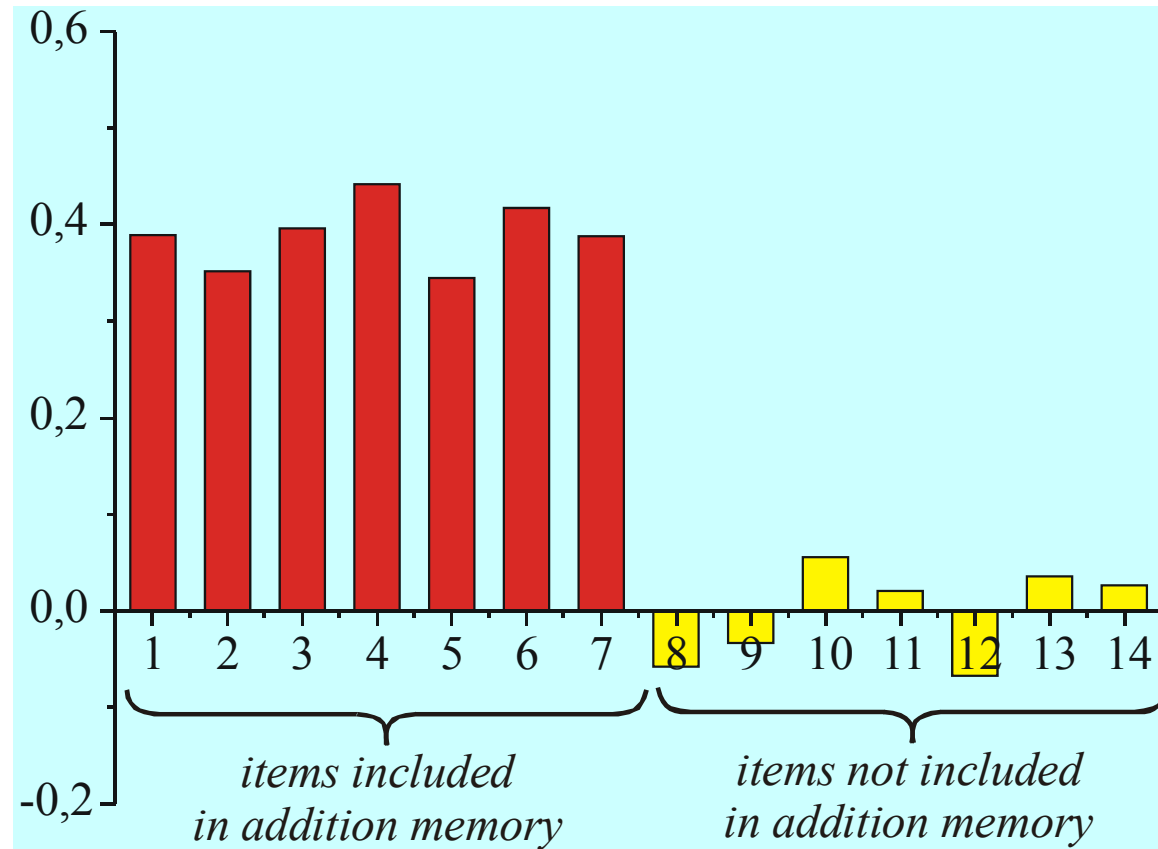
Približný tvar transformácie T je

$$T = \sum_{i=1}^m a_i^* \otimes b_i \Rightarrow a_i \otimes T \approx b_i \quad \forall i \in \{1, 2, \dots, m\}$$



Aditívna pamäť

$$t = \sum_{i=1}^m x_i$$



Aký je rozdiel medzi asociatívnou a aditívnou pamäťou

- Vyvolávanie pamäťovej položky pre *asociatívnu pamäť* je založená na použití narážky (cue). Asociatívna pamäť t obsahuje položku x , ktorá je asociovaná s narážkou c , Rozhodneme sa podľa veľkosti prekryvu

$$\text{prekryv}(x, \tilde{x} = c \otimes t) > \vartheta$$

- Vyvolávanie pamäťovej položky pre *aditívnu pamäť* je založená na použití samotnej položky. Aditívna pamäť t obsahuje položku x , ktorá je asociovaná s narážkou c , Rozhodneme sa podľa veľkosti prekryvu

$$\text{prekryv}(x, \tilde{x} = x \otimes t) > \vartheta$$

Postupnosť položiek (konceptí)

$$\textit{postupnosť} = a - b - c - d - e - f$$

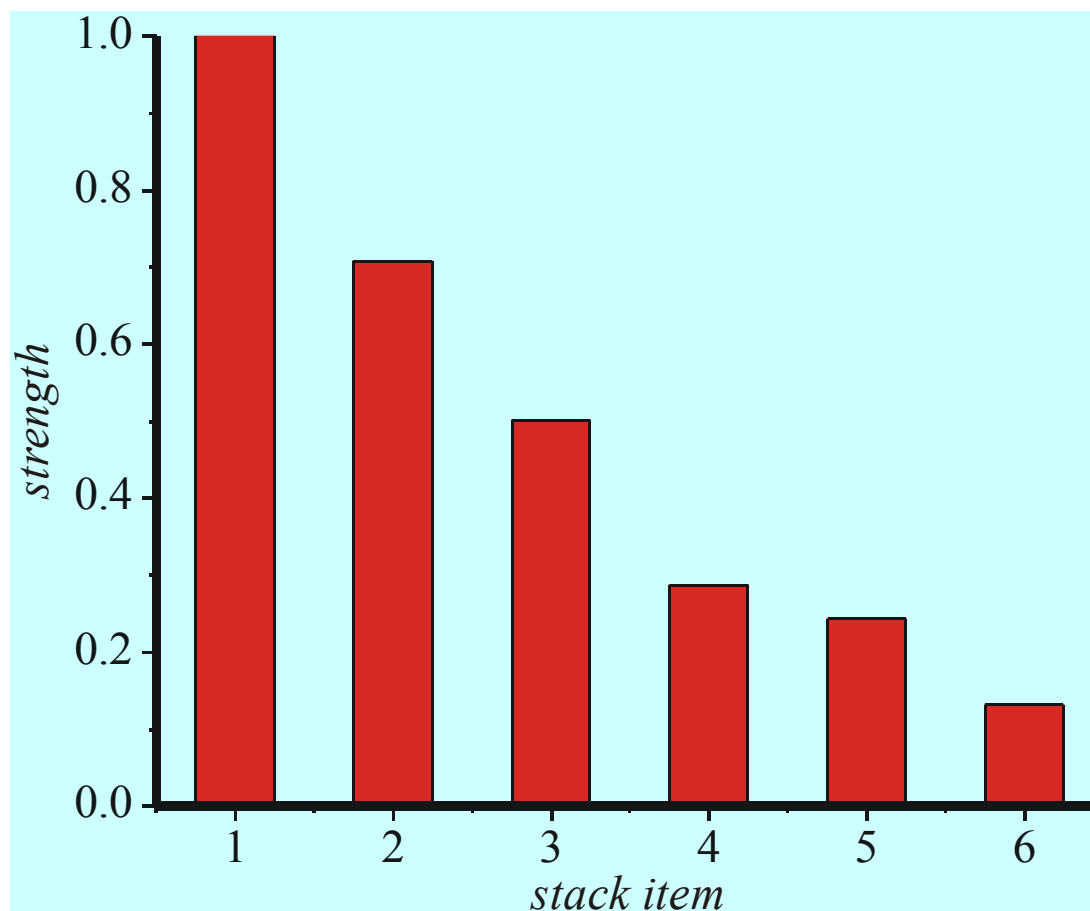
$$\begin{aligned} t_0 = & a + a \otimes b + a \otimes b \otimes c + \\ & a \otimes b \otimes c \otimes d + a \otimes b \otimes c \otimes d \otimes e + a \otimes b \otimes c \otimes d \otimes e \otimes f \end{aligned}$$

Rekonštrukcia

1. krok: $a = \textit{upratovanie}(t_0)$, $t_1 := t_0 - a$, $\tilde{t}_1 := a^* \otimes t_1$,
2. krok: $b = \textit{upratovanie}(\tilde{t}_1)$, $t_2 := t_1 - a \otimes b$, $\tilde{t}_2 := (a \otimes b)^* \otimes t_2$,
3. krok: $c = \textit{upratovanie}(\tilde{t}_2)$, $t_3 := t_2 - a \otimes b \otimes c$, $\tilde{t}_3 := (a \otimes b \otimes c)^* \otimes t_3$,

.....

Priebeh prekryvov pri rekonštrukcii pamäťového vektora t_0 postupnosti obsahujúcej 6 položiek ($n=4000$)



Pozorovanie

Rýchla degradácia rekonštrukcie pamäťového vektora

$$t_0 = a + a \otimes b + a \otimes b \otimes c + \dots$$

Približne 6 položiek je spoľahlivo rekonštruovateľných

Tento prístup môžeme pokladať za matematický model krátkodobej pamäti, ktorá je schopná si zapamätať približne 6-7 položiek (Miller, 1956).

G. A. Miller, Human memory and the storage of information. *IEEE Trans. Info. Theory.* **2** (1956), 129 - 137.

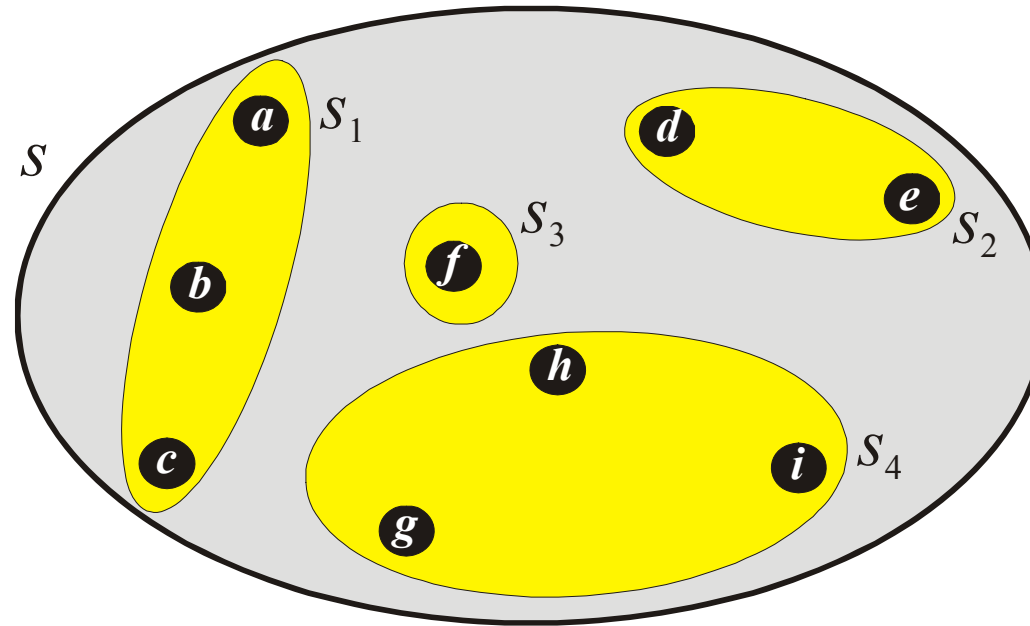
Agregovaná pamäť

Agregovaná pamäť pomáha prekonať problémy s rýchlo degradáciou rekonštrukcie pamäťového vektora postupnosti viac ako 6 položiek

$$S = S_1 \cup S_2 \cup S_3 \cup S_4 \cup \dots \quad (S_i \cup S_j = \emptyset, \text{ for } i \neq j)$$

Príklad

$$S = \{a, b, c, d, e, f, g, h\}$$
$$S_1 = \{a, b, c\}, S_2 = \{d, e\}, S_3 = \{f\}, \text{ and } S_4 = \{g, h\}$$

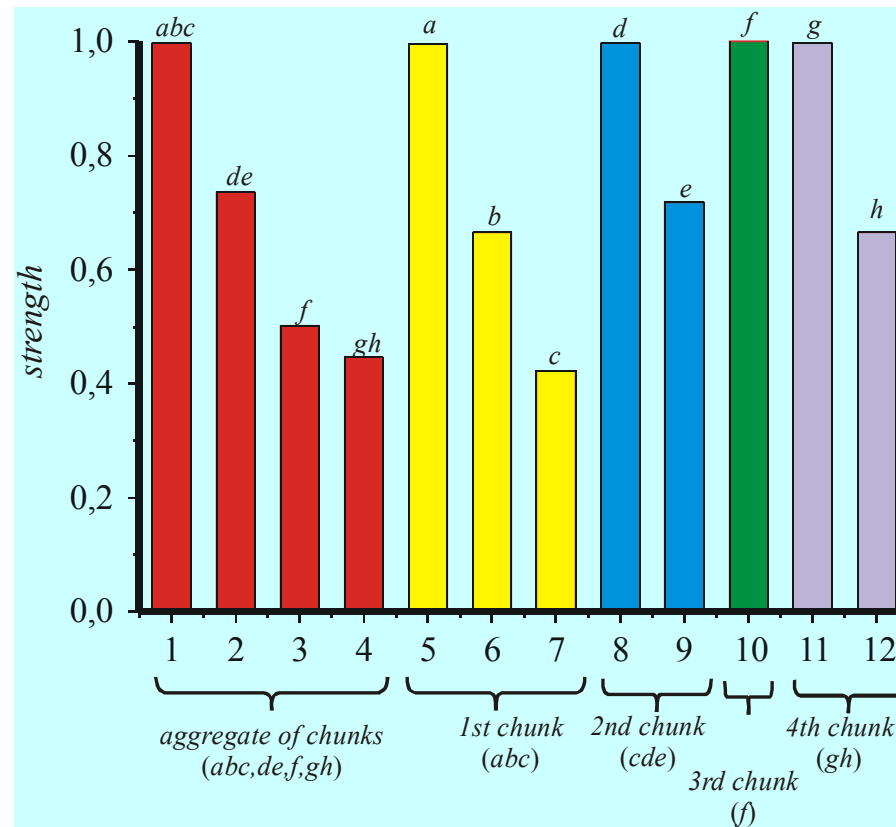


$$s_1 = a + a \otimes b + a \otimes b \otimes c, s_2 = d + d \otimes e, s_3 = f, s_4 = g + g \otimes h$$

$$t = s_1 + s_1 \otimes s_2 + s_1 \otimes s_2 \otimes s_3 + s_1 \otimes s_2 \otimes s_3 \otimes s_4$$

Rekonštrukcia je dvojkrokový proces:

1. krok – rekonštrukcia agregátov
2. krok – zistené agregáty sú ďalej rekonštruované na jednotlivé položky



Reprezentácia predikátov

Študujme predikát dvoch argumentov $P(x,y)$ špecifikovaný ako funkcia

$$\textit{function } P(x : \textit{variable}_1 ; y : \textit{variable}_2)$$

Táto funkcia je holograficky reprezentovaná pomocou asociatívneho pamäťového vektora takto

$$t = \textit{function} \otimes P + \textit{variable}_1 \otimes x + \textit{variable}_2 \otimes y$$

s narážkovými (cue) vektormi $\textit{function}$, $\textit{variable}_1$, a $\textit{variable}_2$. V mnohých prípadoch budeme používať zjednodušený tvar reprezentácie

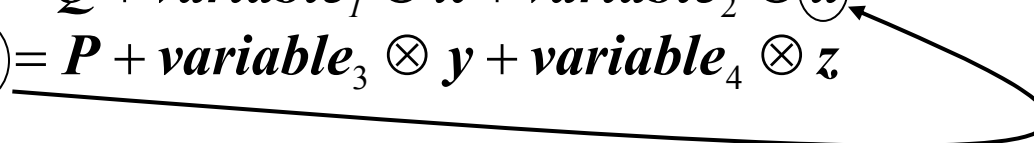
$$t' = P + \textit{variable}_1 \otimes x + \textit{variable}_2 \otimes y$$

Reprezentácia vnorených (nested) predikátov

Študujme príklad $Q(x, P(y, z))$, kde druhý argument predikátu Q je predikát $P(x, y)$.
Reprezentácia tohto vnoreného predikátu má tento tvar

function $Q(x : \text{variable}_1; u : \text{variable}_2)$

function $P(y : \text{variable}_3; z : \text{variable}_4)$

$$t_1 = Q + \text{variable}_1 \otimes x + \text{variable}_2 \otimes \textcircled{u}$$
$$\textcircled{t_2} = P + \text{variable}_3 \otimes y + \text{variable}_4 \otimes z$$


$$t_1 = Q + \text{variable}_1 \otimes x + \text{variable}_2 \otimes (P + \text{variable}_3 \otimes y + \text{variable}_4 \otimes z)$$
$$= Q + \text{variable}_1 \otimes x + \text{variable}_2 \otimes P +$$
$$\text{variable}_2 \otimes \text{variable}_3 \otimes y + \text{variable}_2 \otimes \text{variable}_4 \otimes z$$

Usudzovanie – *modus ponens* a *modus tollens*

$$\textit{modus ponens} : \frac{p \Rightarrow q}{q}, \quad \textit{modus tollens} : \frac{p \Rightarrow q}{\bar{p}}$$

Holografická reprezentácia implikácie $p \Rightarrow q$ má tvar

$$t_{p \Rightarrow q} = op \otimes \textit{impl} + \textit{var}_1 \otimes p + \textit{var}_2 \otimes q$$

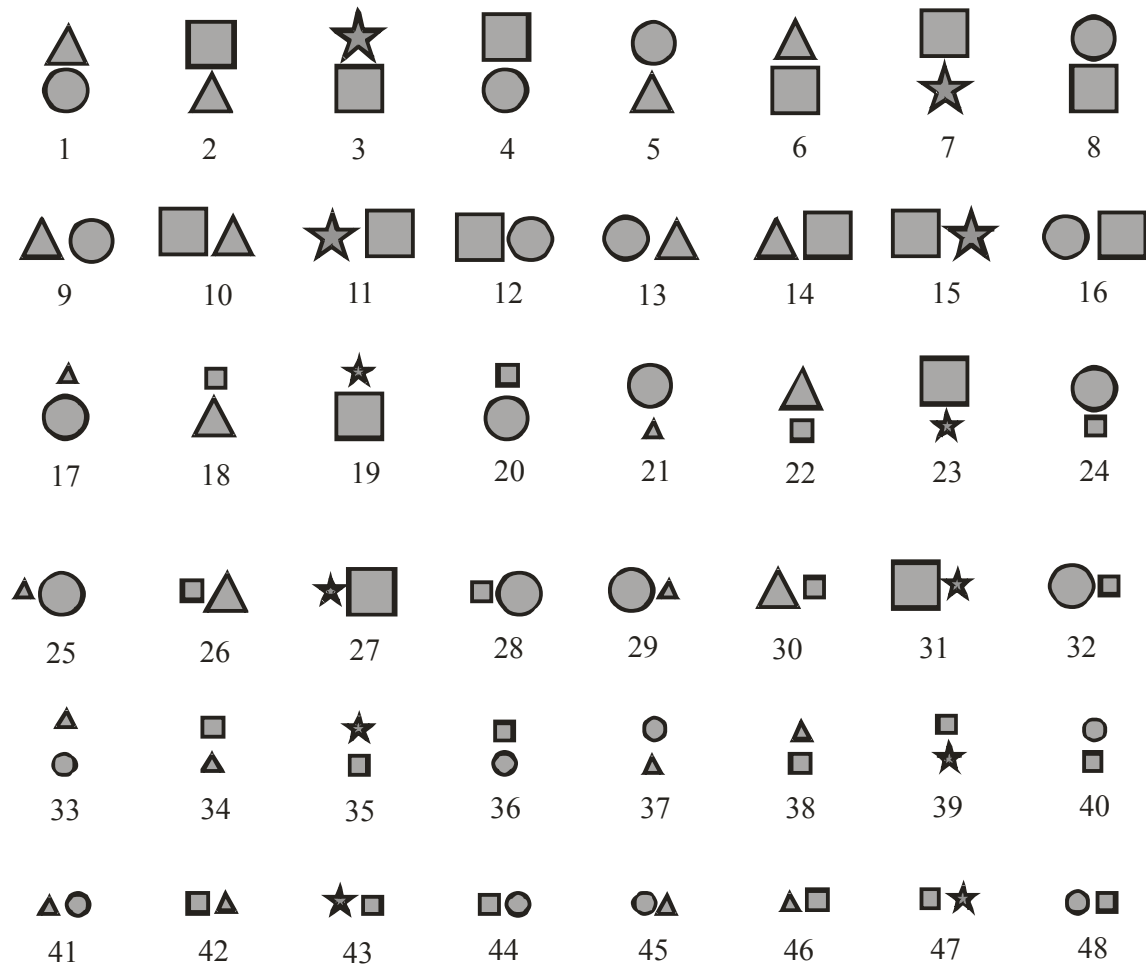
Pomocou transformačného vektora

$$T = \textit{var}_1^* \otimes p^* \otimes \bar{p}^* \otimes \bar{q} + \textit{var}_2^* \otimes q^* \otimes \bar{q}^* \otimes \bar{p}$$

zostrojíme holografickú reprezentáciu *modus ponens* a *modus tollens*

$$p \otimes T \otimes t_{p \Rightarrow q} \approx q \quad \text{a} \quad \bar{q} \otimes T \otimes t_{p \Rightarrow q} \approx \bar{p}$$

A similarity between geometric figures



Holografická reprezentácia jednotlivých prípadov

$$\mathbf{t}_{1,x,y} = \mathbf{ver} + \langle \mathbf{ver_var}_1 \otimes \mathbf{lg} \otimes \mathbf{x} + \mathbf{ver_var}_2 \otimes \mathbf{lg} \otimes \mathbf{y} \rangle$$

$$\mathbf{t}_{2,x,y} = \mathbf{hor} + \langle \mathbf{hor_var}_1 \otimes \mathbf{lg} \otimes \mathbf{x} + \mathbf{hor_var}_2 \otimes \mathbf{lg} \otimes \mathbf{y} \rangle$$

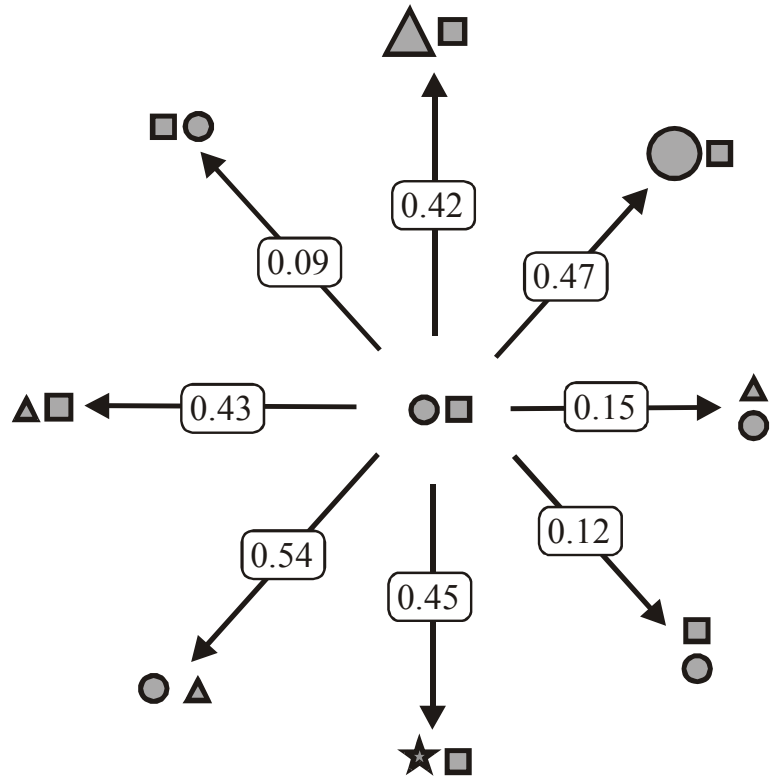
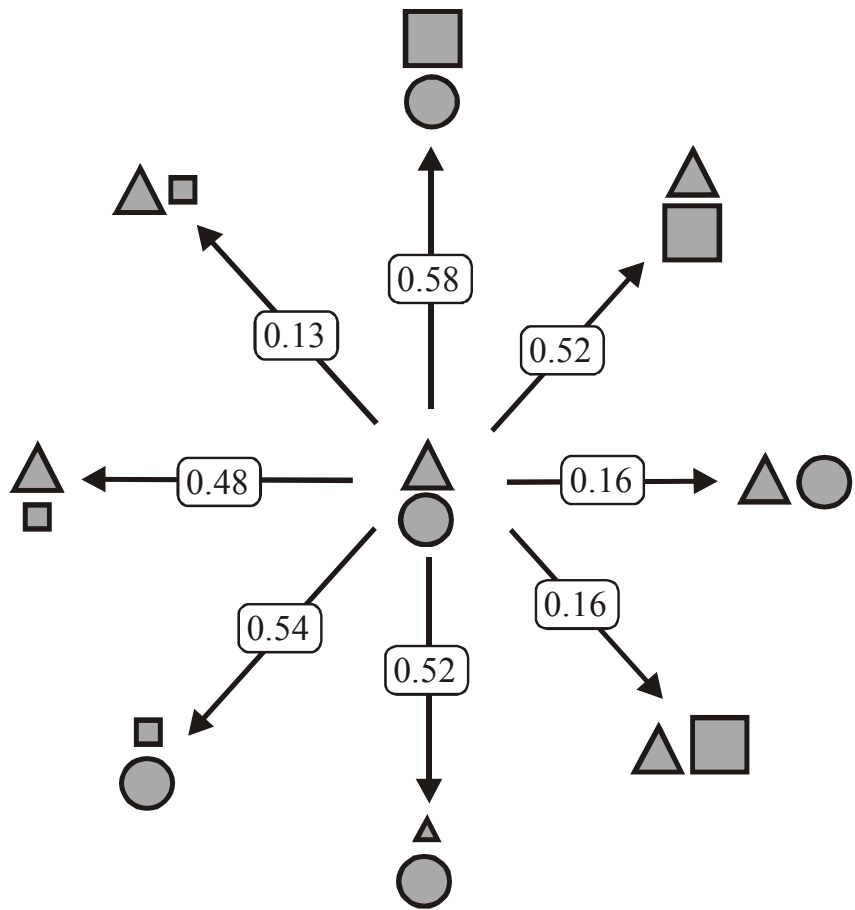
$$\mathbf{t}_{3,x,y} = \begin{cases} \mathbf{ver} + \langle \mathbf{ver_var}_1 \otimes \mathbf{lg} \otimes \mathbf{x} + \mathbf{ver_var}_2 \otimes \mathbf{sm} \otimes \mathbf{y} \rangle \\ \mathbf{ver} + \langle \mathbf{ver_var}_1 \otimes \mathbf{sm} \otimes \mathbf{x} + \mathbf{ver_var}_2 \otimes \mathbf{lg} \otimes \mathbf{y} \rangle \end{cases}$$

$$\mathbf{t}_{4,x,y} = \begin{cases} \mathbf{hor} + \langle \mathbf{hor_var}_1 \otimes \mathbf{lg} \otimes \mathbf{x} + \mathbf{hor_var}_2 \otimes \mathbf{sm} \otimes \mathbf{y} \rangle \\ \mathbf{hor} + \langle \mathbf{hor_var}_1 \otimes \mathbf{sm} \otimes \mathbf{x} + \mathbf{hor_var}_2 \otimes \mathbf{lg} \otimes \mathbf{y} \rangle \end{cases}$$

$$\mathbf{t}_{5,x,y} = \mathbf{ver} + \langle \mathbf{ver_var}_1 \otimes \mathbf{sm} \otimes \mathbf{x} + \mathbf{ver_var}_2 \otimes \mathbf{sm} \otimes \mathbf{y} \rangle$$

$$\mathbf{t}_{6,x,y} = \mathbf{hor} + \langle \mathbf{hor_var}_1 \otimes \mathbf{sm} \otimes \mathbf{x} + \mathbf{hor_var}_2 \otimes \mathbf{sm} \otimes \mathbf{y} \rangle$$

$$\mathit{similarity}(X, X') = \mathit{overlap}(\mathbf{t}, \mathbf{t}')$$



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.).
- Pomocou holografickej distribuovanej reprezentácie môžeme **modelovať rôzne typy asociatívnej pamäti**, ktoré sú reprezentované konceptuálnym vektorom a taktiež dekódovať pamäťový vektor.
- **Proces agregácie** vektorov umožňuje prekonať ťažkosti rýchlej degradácie rekognoskácie pamäťového vektora.

- Holografická redukovaná reprezentácia umožňuje hľadať podobnosť medzi dvoma štruktúrovanými konceptami pomocou jednoduchšej algebraickej operácie skalárneho súčinu medzi ich distribuovanými reprezentáciami.
- Táto skutočnosť môže byť veľmi užitočnou pre modelovanie procesov prehľadávania pamäti na základe podobnosti.
- Holografická redukovaná reprezentácia môže byť taktiež použitá k modelovaniu procesov usudzovania.

