Microsoft Research
Faculty Summit
2014 15TH ANNUAL

Paul Smolensky
Cognitive Science Dept
Johns Hopkins Univ

Symbolic roles in vectorial computation

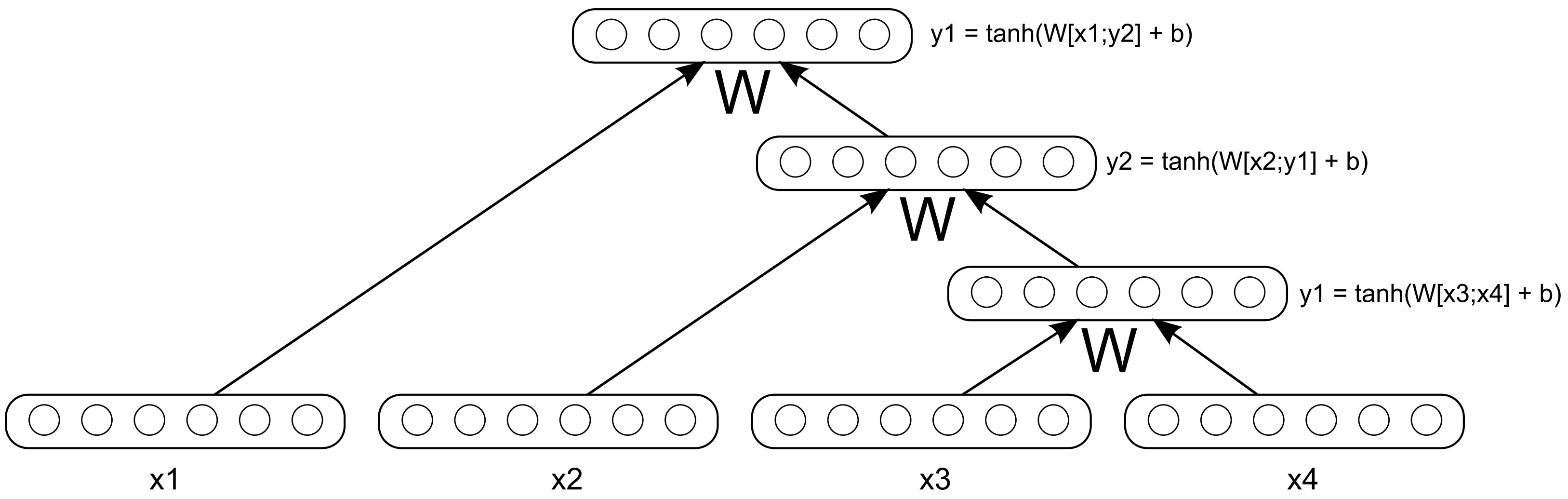
Microsoft

Vectorial encoding of symbolic *structure*

— in contrast to hybrid symbolic/vectorial representations

A single vector encodes (i) all the (vectorial) labels **and** (ii) the (discrete) structure in which they reside

Cognitive motivation:
vector ~ neural state



Socher, Manning & Ng 2010

Vectorial encoding of symbolic *structure*

TYPE: Decompose structure into roles $\{r_k\}$

Approach 1: Absolute position

[Approach 2: Contextual (~ *n*-gram)]

Each r_k is assigned a vector encoding $\mathbf{r}_k \in R$ (linearly indep.)

— designed or learned

Can be non-orthogonal:

- $\mathbf{r}_k \bullet \mathbf{r}_l \equiv \text{sim}(\mathbf{r}_k, \mathbf{r}_l) \neq 0$
- interaction of $\mathbf{r}_k$ & $\mathbf{r}_k$
- generalization

Vectorial encoding of symbolic *structure*

TYPE: Decompose structure into roles $\{r_k\}$

Approach 1: Absolute position

[Approach 2: Contextual (~ *n*-gram)]

Each r_k is assigned a vector encoding $\mathbf{r}_k \in R$ (linearly indep.)

— designed or learned

INSTANCE: Specific fillers for roles

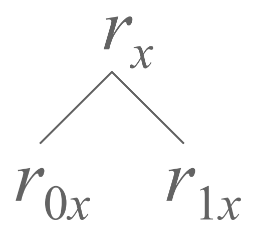
Let $\mathbf{f}_k \in F$ (linearly indep.) be the label in role r_k

— $\mathbf{f}_k$ may be a vector encoding of a symbol $\mathtt{f}_k \in A$

— designed or learned

ENCODING: $\mathbf{v} = \sum_k \mathbf{f}_k \otimes \mathbf{r}_k$

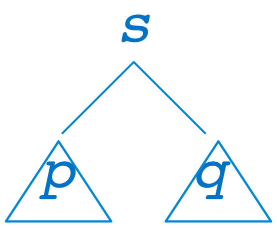
Can be recursive:



$\forall x \in \{0, 1\}^*$

$\mathbf{r}_{0x} = \mathbf{r}_0 \otimes \mathbf{r}_x$

$R = \bigoplus_d R^{(d)}$



$\mathbf{s} = \mathbf{p} \otimes \mathbf{r}_0 + \mathbf{q} \otimes \mathbf{r}_1$

⇒ exact computation of recursive functions

Size: linear in number of roles

Tensor Product Representations (TPRs: 1990)

Summary: TPRs

TYPE: Decompose structure into roles $\{r_k\}$

Each r_k is assigned a vector encoding $\mathbf{r}_k \in R$ (linearly indep.)

INSTANCE: Specific fillers for roles

Let $\mathbf{f}_k \in F$ (linearly indep.) be the label in role r_k

— $\mathbf{f}_k$ may be a vector encoding of a symbol $\mathtt{f}_k \in A$

ENCODING: $\mathbf{v} = \sum_k \mathbf{f}_k \otimes \mathbf{r}_k$

N.B.: Turns out to have important (indirect) implications for grammatical theory

Computability theory over TPRs

What symbolic functions can be computed over TPRs using neural computation?

The functions in the following classes are computable in a linear neural network:

$\mathcal{B}$ = base of in-place symbol mappings

$\mathcal{C}$ = closure under composition of
[tree-manipulating primitives $\cup$ $\mathcal{B}$]

$\mathcal{P}$ ~ "primitive recursive"

'Primitive recursive':

$\mathcal{C} \subset \mathcal{P}$

$g, h \in \mathcal{P} \Rightarrow f \in \mathcal{P}$ when

$$f(s) = \begin{cases} g(s) & \text{if } \mathbf{atom}(s) \\ h(f(\mathbf{ex}_0(s)), f(\mathbf{ex}_1(s))) & \text{otherwise} \end{cases}$$

Decoding TPRs

INSTANCE $\mathbf{v}$: Inner product

$\mathbf{f}_k = \mathbf{v} \cdot \mathbf{r}_k^{+}$ — given $\{\mathbf{r}_k\}$

SAMPLE $\{\mathbf{v}^{(\alpha)}\}$: Generative model

Hypothesis: $\{\mathbf{v}^{(\alpha)}\}$ is a collection of TPRs, each encoding an instance of a symbol structure of a single type

$\mathbf{v}^{(\alpha)} = \sum_k \mathbf{f}_k^{(\alpha)} \otimes \mathbf{r}_k$ — where $\mathbf{f}_k^{(\alpha)}$ encodes a symbol $\mathtt{f}_k^{(\alpha)}$

Learning algorithms: derived from generative model

TYPE: What are $\{\mathbf{r}_k\}$ and $\{\mathbf{f}_k\}$?

INSTANCE: For a given α,

which symbol $\mathtt{f}_k^{(\alpha)} \in A$ fills each role r_k?

APPLICATION: Decoding neuroimages of combinatorial stimuli (e.g., sentences, words).

Instance bindings $\{\mathtt{f}_k^{(\alpha)}/r_k\}$ of stimuli are known, so only need learn the TYPE encoding

Vectorial encoding of symbolic *structure*

TYPE: Decompose structure into roles $\{r_k\}$

Approach 1: Absolute position

Approach 2: Contextual (~ *n*-gram)

Each r_k is assigned a vector encoding $\mathbf{r}_k \in R$ (linearly indep.)

— designed or learned

INSTANCE: Specific fillers for roles

Let $\mathbf{f}_k \in F$ (linearly indep.) be the label in role r_k

— $\mathbf{f}_k$ may be a vector encoding of a symbol $\mathtt{f}_k \in A$

— designed or learned

ENCODING: $\mathbf{v} = \sum_k \mathbf{f}_k \otimes \mathbf{r}_k$

R(—, Y)

Role

If $\mathbf{X}$ fills this role, vector encoding is: $\mathbf{X} \otimes (\mathbf{R} \otimes \mathbf{Y})$

$\cong \mathbf{R} \otimes \mathbf{X} \otimes \mathbf{Y}$

(used in cognitive models)

Approach 1: [filler] ⊗ [position]

Approach 2: [filler_1] ⊗ [filler_2]

Few leaders are admired by George Bush $\xrightarrow{f}$ admire(George Bush, few leaders)

$$f(s) = \mathbf{cons}(\mathbf{ex}_1(\mathbf{ex}_0(\mathbf{ex}_1(s))), \mathrm{cons}(\mathrm{ex}_1(\mathrm{ex}_1(\mathrm{ex}_1(s))), \mathrm{ex}_0(s)))$$

$$\mathbb{W} = \mathbb{W}_{\mathrm{cons}_0}[\mathbb{W}_{\mathrm{ex}_1}\mathbb{W}_{\mathrm{ex}_0}\mathbb{W}_{\mathrm{ex}_1}] + \mathbb{W}_{\mathrm{cons}_1}[\mathbb{W}_{\mathrm{cons}_0}(\mathbb{W}_{\mathrm{ex}_1}\mathbb{W}_{\mathrm{ex}_1}\mathbb{W}_{\mathrm{ex}_1}) + \mathbb{W}_{\mathrm{cons}_1}(\mathbb{W}_{\mathrm{ex}_0})]$$