A Study of Entanglement in a Categorical Framework of Natural Language

Dimitri Kartsaklis[1] Mehrnoosh Sadrzadeh[2]

[1]Department of Computer Science
University of Oxford

[2]School of Electronic Engineering
and Computer Science
Queen Mary University of London

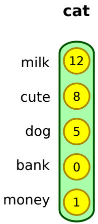
QPL 2014, June 4-6

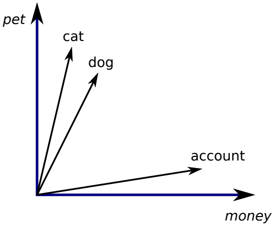
The necessity of compositionality

Distributional hypothesis

The meaning of a word is determined by its context (Harris, 1954)

- A word is a *vector* of co-occurrence statistics with every other word in the vocabulary:





- **Not enough data** to do the same for phrases or sentences, (e.g. '*coursework meets deadline*','*script lack information*' appear 1 time in a corpus of 100m sentences).

A categorical framework for composition

A solution

Use the grammar rules to *compose* the vectors of the words in a sentence into a sentence vector.

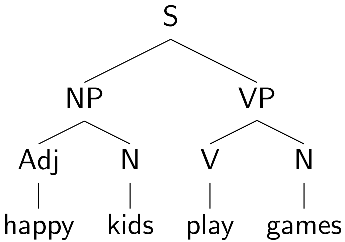
- Both a pregroup grammar and the category of finite-dimensional vector spaces and linear maps over a field share a compact closed structure
- We can then define a strongly monoidal functor $\mathcal{F}$ such that:

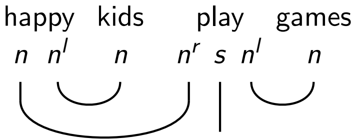
$$\mathcal{F} : \mathbf{Preg}_F \to \mathbf{FVect}_W \tag{1}$$

- The meaning of a sentence $w_1 w_2 \dots w_n$ with type reduction α is given as:

$$\mathcal{F}(\alpha)(\overrightarrow{w_1} \otimes \overrightarrow{w_2} \otimes \dots \otimes \overrightarrow{w_n}) \tag{2}$$

An example





Type reduction:

$$(\epsilon_n^r \otimes 1_s) \circ (1_n \otimes \epsilon_n^l \otimes 1_{n^r \cdot s} \otimes \epsilon_n^l)$$

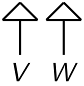
$$\mathcal{F}\left[(\epsilon_n^r \otimes 1_s) \circ (1_n \otimes \epsilon_n^l \otimes 1_{n^r \cdot s} \otimes \epsilon_n^l)\right]\left(\overline{happy} \otimes \overrightarrow{kids} \otimes \overline{play} \otimes \overrightarrow{games}\right) =$$

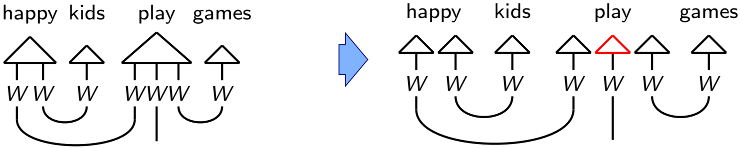
$$(\epsilon_W \otimes 1_W) \circ (1_W \otimes \epsilon_W \otimes 1_{W \otimes W} \otimes \epsilon_W)\left(\overline{happy} \otimes \overrightarrow{kids} \otimes \overline{play} \otimes \overrightarrow{games}\right) =$$

$$(\overline{happy} \times \overrightarrow{kids})^{\mathsf{T}} \times \overline{play} \times \overrightarrow{games}$$

Entanglement in linguistics

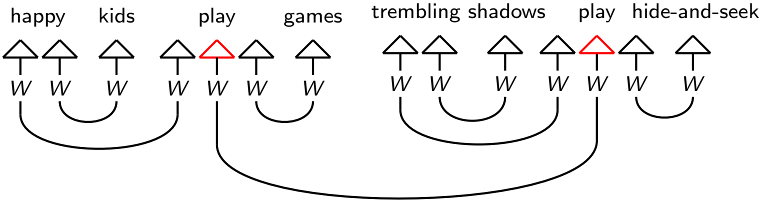
Entangled tensor: 

Separable tensor: 



Euclidean: $\langle \overrightarrow{happy}^{(r)} | \overrightarrow{kids} \rangle \langle \overrightarrow{happy}^{(l)} | \overrightarrow{play}^{(l)} \rangle \langle \overrightarrow{play}^{(r)} | \overrightarrow{games} \rangle \overrightarrow{play}^{(m)}$

Cosine: $\overrightarrow{play}^{(m)}$



Concrete models for verb tensors (1/2)

- A transitive verb should live in $W^{\otimes 3}$, but tensors of order higher than 2 are difficult to create and manipulate

A workaround:

Start with a matrix, then inflate this to tensors of higher order using Frobenius algebras

$$\overline{verb} = \sum_i (\overrightarrow{subject_i} \otimes \overrightarrow{object_i}) \tag{3}$$

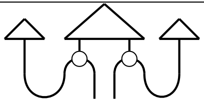
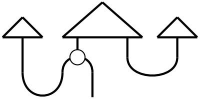
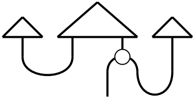
- Compare with the following separable version:

$$\overline{verb} = \left(\sum_i \overrightarrow{subject_i}\right) \otimes \left(\sum_i \overrightarrow{object_i}\right) \tag{4}$$

- ... and the rank-1 approximation of $\overline{verb}$:

$$\overline{verb}_{R1} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^{\mathsf{T}} \quad \text{for } \overline{verb} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}} \tag{5}$$

Concrete models for verb tensors (2/2)

Model	Diagram	Formula
Relational		$\overline{s} = (subj \otimes obj) \odot \overline{verb}$
Copy-subj		$\overrightarrow{s} = \overrightarrow{subj} \odot (\overline{verb} \times \overrightarrow{obj})$
Copy-obj		$\overrightarrow{s} = \overrightarrow{obj} \odot (\overline{verb}^{\mathsf{T}} \times \overrightarrow{subj})$

We further combine **Copy-subj** and **Copy-obj** as follows:

- Frobenius additive: $CopySubj + CopyObj$
- Frobenius multiplicative: $CopySubj \odot CopyObj$
- Frobenius tensored: $CopySubj \otimes CopyObj$

Detecting sentence similarity (1/2)

The task

Compare the similarity of transitive sentences by composing vectors and measuring the cosine distance between them. Evaluate the results against human judgements.

Dataset 1: *Same subjects/objects, semantically related verbs*

Model	ρ **with cos**	ρ **with Eucl.**
Verbs only	0.329	0.138
Additive	0.234	0.142
Multiplicative	0.095	0.024
Relational	0.400	0.149
Rank-1 approx. of relational	0.402	0.149
Separable	0.401	0.090
Copy-subject	0.379	0.115
Copy-object	0.381	0.094
Frobenius additive	**0.405**	0.125
Frobenius multiplicative	0.338	0.034
Frobenius tensored	**0.415**	0.010
Human agreement	0.60	

Detecting sentence similarity (2/2)

Dataset 2: *Different subjects, objects and verbs*

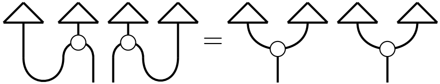
Model	ρ **with cos**	ρ **with Eucl.**
Verbs only	0.449	0.392
Additive	0.581	0.542
Multiplicative	0.287	0.109
Relational	0.334	0.173
Rank-1 approx. of relational	0.333	0.175
Separable	0.332	0.105
Copy-subject	0.427	0.096
Copy-object	0.198	0.144
Frobenius additive	0.428	0.117
Frobenius multiplicative	0.302	0.041
Frobenius tensored	0.332	0.042
Human agreement	0.66	

Simplifications on the models

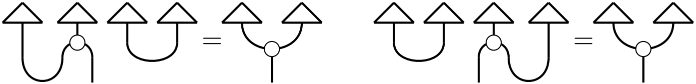
Conclusions from experimental work

1. Verb matrices created as $\sum_i (subj_i \otimes obj_i)$ are *essentially separable*[1](too much linear dependence between vectors?)
2. The only level of entanglement in the inflated verb tensors is provided by the Frobenius operators

This introduces a number of simplifications in the models:



$$\overline{s} = (\overrightarrow{subj} \odot \overrightarrow{verb}^{(l)}) \otimes (\overrightarrow{verb}^{(r)} \odot \overrightarrow{obj})$$



$$\overrightarrow{s} = (\overrightarrow{subj} \odot \overrightarrow{verb}^{(l)}) + (\overrightarrow{verb}^{(r)} \odot \overrightarrow{obj})$$

[1] Average cos similarity of verbs with their rank-1 approximations: 0.99

Using linear regression

- For a given verb, collect all $\langle \overrightarrow{obj_i}, \overrightarrow{play\ obj_i} \rangle$ pairs (e.g. the vector of 'flute' paired with the holistic vector of 'play flute', and so on)
- Learn a matrix for the verb by minimizing the quantity:

$$\frac{1}{2m}\left(\sum_i \overline{verb} \times \overrightarrow{object}_i - \overrightarrow{verb\ object_i}\right)^2 \tag{6}$$

- Cosine similarity between the verb matrices and their rank-1 approximations: **0.48**
- Same concept can be applied to Frobenius additive model:

$$\frac{1}{2m}\left(\sum_i (\overline{verb} \times \overrightarrow{obj}_i \odot \overrightarrow{subj_i} + \overline{verb}^{\mathsf{T}} \times \overrightarrow{subj_i} \odot \overrightarrow{obj_i}) - \overrightarrow{subj\ verb\ obj_i}\right)^2 \tag{7}$$

Work in progress...

Conclusion

- A preliminary study on entanglement aspects of tensor-based compositional models
- A number of concrete implementations of the Coecke-Sadrzadeh-Clark categorical framework have been proved problematic from an entanglement perspective
- However, in all cases the involvement of Frobenius algebras in the creation of verb tensors equips the fragmented compositional structure with flow
- The separability problem is not present for verb tensors constructed by gradient optimization techniques
- Corpus-based methods, such as the "Frobenius additive" model, are still viable and "easy" alternatives for creating verb tensors

Conclusion

- A preliminary study on entanglement aspects of tensor-based compositional models
- A number of concrete implementations of the Coecke-Sadrzadeh-Clark categorical framework have been proved problematic from an entanglement perspective
- However, in all cases the involvement of Frobenius algebras in the creation of verb tensors equips the fragmented compositional structure with flow
- The separability problem is not present for verb tensors constructed by gradient optimization techniques
- Corpus-based methods, such as the "Frobenius additive" model, are still viable and "easy" alternatives for creating verb tensors

Thank you!