Tensor-based Models of Natural Language Semantics

Dimitri Kartsaklis

Joint work with *M. Sadrzadeh* (QMUL), *B. Coecke* (Oxford) and *R. Piedeleu* (Oxford)

School of Electronic Engineering and Computer Science

Queen Mary University of London

Workshop on Tensors, their Decomposition, and Applications

QMUL, 16th August 2016

In a nutshell

- Tensor-based models of meaning aim to unify two orthogonal semantic paradigms:
  - The type-logical compositional approach of formal semantics
  - The quantitative perspective of vector space models of meaning
- Useful in every NLP task: sentence similarity, paraphrase detection, sentiment analysis, machine translation etc.

In this talk:

I provide an introduction to the field by presenting the mathematical foundations, discussing important extensions and recent work, and touching implementation issues and practical applications.

Outline

Computers and meaning

Computational linguistics is the scientific and engineering discipline concerned with understanding written and spoken language from a computational perspective.

—Stanford Encyclopedia of Philosophy[1]

[1] http://plato.stanford.edu

The meaning of words

Distributional hypothesis

Words that occur in similar contexts have similar meanings [Harris, 1958].

The functional interplay of philosophy and	?	should, as a minimum, guarantee...
...and among works of dystopian	?	fiction...
The rapid advance in	?	today suggests...
...calculus, which are more popular in	?	-oriented schools.
But because	?	is based on mathematics...
...the value of opinions formed in	?	as well as in the religions...
...if	?	can discover the laws of human nature....
...is an art, not an exact	?	.
...factors shaping the future of our civilization:	?	and religion.
...certainty which every new discovery in	?	either replaces or reshapes.
...if the new technology of computer	?	is to grow significantly
He got a	?	scholarship to Yale.
...frightened by the powers of destruction	?	has given...
...but there is also specialization in	?	and technology...

The meaning of words

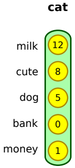
Distributional hypothesis

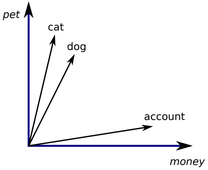
Words that occur in similar contexts have similar meanings [Harris, 1958].

The functional interplay of philosophy and	**science**	should, as a minimum, guarantee...
...and among works of dystopian	**science**	fiction...
The rapid advance in	**science**	today suggests...
...calculus, which are more popular in	**science**	-oriented schools.
But because	**science**	is based on mathematics...
...the value of opinions formed in	**science**	as well as in the religions...
...if	**science**	can discover the laws of human nature....
...is an art, not an exact	**science**	.
...factors shaping the future of our civilization:	**science**	and religion.
...certainty which every new discovery in	**science**	either replaces or reshapes.
...if the new technology of computer	**science**	is to grow significantly
He got a	**science**	scholarship to Yale.
...frightened by the powers of destruction	**science**	has given...
...but there is also specialization in	**science**	and technology...

Distributional models of meaning

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



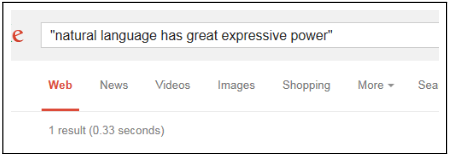


- Semantic relatedness is usually based on cosine similarity:

$$\text{sim}(\overrightarrow{v}, \overrightarrow{u}) = \cos \theta_{\overrightarrow{v}, \overrightarrow{u}} = \frac{\langle \overrightarrow{v} \cdot \overrightarrow{u} \rangle}{\|\overrightarrow{v}\| \|\overrightarrow{u}\|}$$

Moving to phrases and sentences

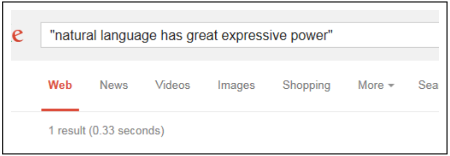
- We would like to generalize this idea to phrases and sentences
- However, it's not clear how
- There are practical problems—there is not enough data:



- But even if we had a very large corpus, what the context of a sentence would be?

Moving to phrases and sentences

- We would like to generalize this idea to phrases and sentences
- However, it's not clear how
- There are practical problems—there is not enough data:



- But even if we had a very large corpus, what the context of a sentence would be?

A solution:

For a sentence $w_1 w_2 \dots w_n$, find a function f such that:

$$\overrightarrow{s} = f(\overrightarrow{w_1}, \overrightarrow{w_2}, \dots, \overrightarrow{w_n})$$

Outline

Quantizing the grammar

Coecke, Sadrzadeh and Clark (2010):

Pregroup grammars are structurally homomorphic with the category of finite-dimensional vector spaces and linear maps (both share compact closure)

- In abstract terms, there exists a structure-preserving passage from grammar to meaning:

$$\mathcal{F} : \text{Grammar} \rightarrow \text{Meaning}$$

- The meaning of a sentence $w_1 w_2 \ldots w_n$ with grammatical derivation α is defined as:

$$\overrightarrow{w_1 w_2 \ldots w_n} := \mathcal{F}(\alpha)(\overrightarrow{w_1} \otimes \overrightarrow{w_2} \otimes \ldots \otimes \overrightarrow{w_n})$$

Pregroup grammars

A pregroup grammar $P(\Sigma, \mathcal{B})$ is a relation that assigns grammatical types from a pregroup algebra freely generated over a set of atomic types $\mathcal{B}$ to words of a vocabulary Σ.

- A pregroup algebra is a partially ordered monoid, where each element p has a left and a right adjoint such that:

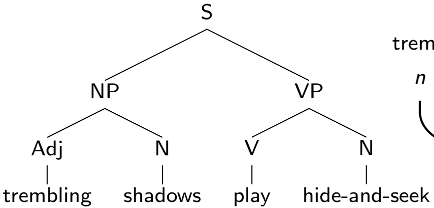
$$p \cdot p^r \leq 1 \leq p^r \cdot p \qquad p^l \cdot p \leq 1 \leq p \cdot p^l$$

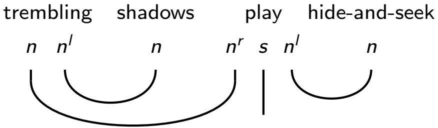
- Elements of the pregroup are basic (atomic) grammatical types, e.g. $\mathcal{B} = \{n, s\}$.
- Atomic grammatical types can be combined to form types of higher order (e.g. $n \cdot n^l$ or $n^r \cdot s \cdot n^l$)
- A sentence $w_1 w_2 \ldots w_n$ (with word w_i to be of type t_i) is grammatical whenever:

$$t_1 \cdot t_2 \cdot \ldots \cdot t_n \leq s$$

Pregroup derivation: example

$$p \cdot p^r \leq 1 \leq p^r \cdot p \qquad p^l \cdot p \leq 1 \leq p \cdot p^l$$





$$\begin{aligned} n \cdot n^l \cdot n \cdot n^r \cdot s \cdot n^l \cdot n &\leq n \cdot 1 \cdot n^r \cdot s \cdot 1 \\ &= n \cdot n^r \cdot s \\ &\leq 1 \cdot s \\ &= s \end{aligned}$$

Compact closed categories

- A monoidal category $(\mathcal{C}, \otimes, I)$ is compact closed when every object has a left and a right adjoint, for which the following morphisms exist:

$$A \otimes A^r \xrightarrow{\epsilon^r} I \xrightarrow{\eta^r} A^r \otimes A \qquad\qquad A^l \otimes A \xrightarrow{\epsilon^l} I \xrightarrow{\eta^l} A \otimes A^l$$

- Pregroup grammars are CCCs, with ϵ and η maps corresponding to the partial orders
- **FdVect**, the category of finite-dimensional vector spaces and linear maps, is a also a (symmetric) CCC:
    - ϵ maps correspond to inner product
    - η maps to identity maps and multiples of those

A functor from syntax to semantics

We define a strongly monoidal functor $\mathcal{F}$ such that:

$$\mathcal{F} : P(\Sigma, \mathcal{B}) \rightarrow \mathbf{FdVect}$$

$$\begin{aligned}
\mathcal{F}(p) &= P \quad \forall p \in \mathcal{B} \\
\mathcal{F}(1) &= \mathbb{R} \\
\mathcal{F}(p \cdot q) &= \mathcal{F}(p) \otimes \mathcal{F}(q) \\
\mathcal{F}(p^r) = \mathcal{F}(p^l) &= \mathcal{F}(p) \\
\mathcal{F}(p \leq q) &= \mathcal{F}(p) \rightarrow \mathcal{F}(q) \\
\mathcal{F}(\epsilon^r) = \mathcal{F}(\epsilon^l) &= \text{inner product in } \mathbf{FdVect} \\
\mathcal{F}(\eta^r) = \mathcal{F}(\eta^l) &= \text{identity maps in } \mathbf{FdVect}
\end{aligned}$$

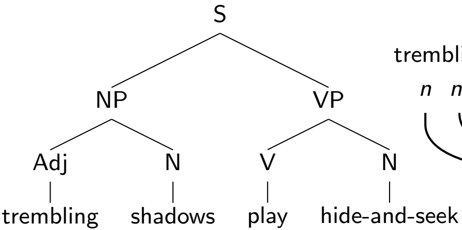
[Kartsaklis, Sadrzadeh, Pulman and Coecke, 2016]

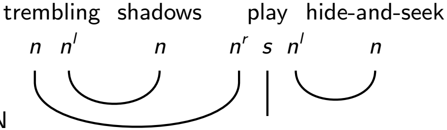
A multi-linear model

The grammatical type of a word defines the vector space in which the word lives:

- Nouns are vectors in N;
- adjectives are linear maps $N \rightarrow N$, i.e elements in $N \otimes N$;
- intransitive verbs are linear maps $N \rightarrow S$, i.e. elements in $N \otimes S$;
- transitive verbs are bi-linear maps $N \otimes N \rightarrow S$, i.e. elements of $N \otimes S \otimes N$;

- The composition operation is tensor contraction, i.e. elimination of matching dimensions by application of inner product.





Type reduction morphism:

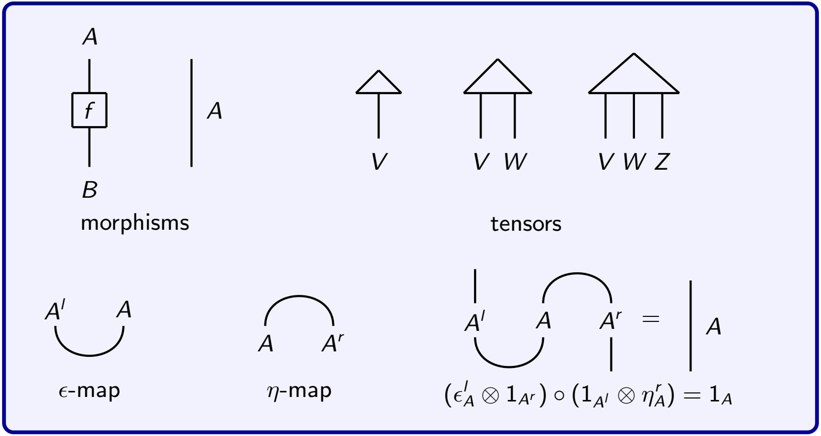
$$(\epsilon_n^r \cdot 1_s) \circ (1_n \cdot \epsilon_n^l \cdot 1_{n^r \cdot s} \cdot \epsilon_n^l) : n \cdot n^l \cdot n \cdot n^r \cdot s \cdot n^l \cdot n \to s$$

$$\mathcal{F}\left[(\epsilon_n^r \cdot 1_s) \circ (1_n \cdot \epsilon_n^l \cdot 1_{n^r \cdot s} \cdot \epsilon_n^l)\right] \left(\overline{trembling} \otimes \overrightarrow{shadows} \otimes \overline{play} \otimes \overrightarrow{hide\text{-}and\text{-}seek}\right) =$$

$$(\epsilon_N \otimes 1_S) \circ (1_N \otimes \epsilon_N \otimes 1_{N \otimes S} \otimes \epsilon_N) \left(\overline{trembling} \otimes \overrightarrow{shadows} \otimes \overline{play} \otimes \overrightarrow{hide\text{-}and\text{-}seek}\right) =$$

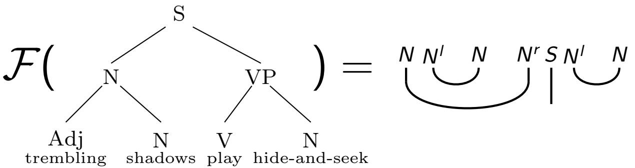
$$\overline{trembling} \times \overrightarrow{shadows} \times \overline{play} \times \overrightarrow{hide\text{-}and\text{-}seek}$$

$\overrightarrow{shadows}, \overrightarrow{hide\text{-}and\text{-}seek} \in N$ $\quad \overline{trembling} \in N \otimes N$ $\quad \overline{play} \in N \otimes S \otimes N$

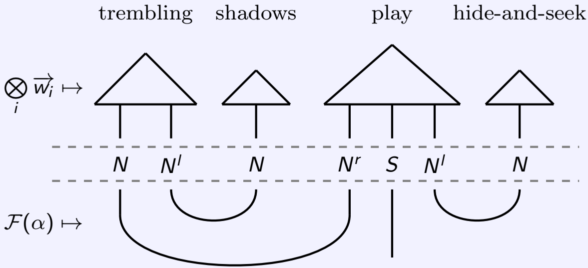


- Vectors and tensors are states: $\overrightarrow{v} : I \rightarrow V$, $\overline{w} : I \rightarrow V \otimes V$ and so on.

Graphical language: example



$$\mathcal{F}(\alpha)(\overrightarrow{trembling} \otimes \overrightarrow{shadows} \otimes \overrightarrow{play} \otimes \overrightarrow{hide\text{-}and\text{-}seek})$$



Outline

Extensional approach

Grefenstette and Sadrzadeh (2011); Kartsaklis and Sadrzadeh (2016):

A relational word is defined as the set of its arguments:

$$[\![red]\!] = \{car, door, dress, ink, \cdots\}$$

To give this linear-algebraically:

$$\overline{adj} = \sum_i \overrightarrow{noun_i} \otimes \overrightarrow{noun_i}$$

- When composing the adjective with a new noun n', we get:

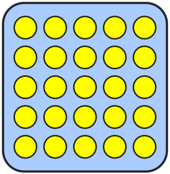
$$\overline{adj} \times \overrightarrow{n}' = \sum_i \langle \overrightarrow{noun_i}, \overrightarrow{n}' \rangle \overrightarrow{noun_i}$$

Statistical approach

Baroni and Zamparelli (2010):

Create holistic distributional vectors for whole compounds (as if they were words) and use them to train a linear regression model.

red



$\times \overrightarrow{\text{car}} = \overrightarrow{\text{red car}}$

$\times \overrightarrow{\text{door}} = \overrightarrow{\text{red door}}$

$\times \overrightarrow{\text{dress}} = \overrightarrow{\text{red dress}}$

$\times \overrightarrow{\text{ink}} = \overrightarrow{\text{red ink}}$

$$\hat{\overline{adj}} = \arg\min_{\overline{adj}} \left[\frac{1}{2m} \sum_i (\overline{adj} \times \overrightarrow{noun_i} - \overrightarrow{adj\ noun_i})^2 \right]$$

Decomposition of tensors

- 3rd-order tensors for transitive verbs (and 4th-order for ditransitive verbs) pose a challenge
- We can reduce the number of parameters by applying canonical polyadic decomposition:

$$\overline{verb} = \sum_{r=1}^{R} \mathbf{P}_r \otimes \mathbf{Q}_r \otimes \mathbf{R}_r$$

$$\mathbf{P} \in \mathbb{R}^{R \times S} \quad , \quad \mathbf{Q} \in \mathbb{R}^{R \times N} \quad , \quad \mathbf{R} \in \mathbb{R}^{R \times N}$$

- Keep R sufficiently small with regard to S and N
- Learn **P**, **Q** and **R** by multi-linear regression

$$\overrightarrow{svo} = f(\overrightarrow{s}, \overrightarrow{o}) := \mathbf{P}^{\mathsf{T}}(\mathbf{Q}\overrightarrow{s} \odot \mathbf{R}\overrightarrow{o})$$

$$L = \frac{1}{2m} \sum_{i=1}^{m} ||f(\overrightarrow{s_i}, \overrightarrow{o_i}) - \overrightarrow{t_i}||^2$$

[Fried, Polajnar, Clark (2015)]

Outline

Functional words

- Certain classes of words, such as determiners, relative pronouns, prepositions, or coordinators occur in almost every possible context.
- Thus, they are considered semantically vacuous from a distributional perspective and most often they are simply ignored.

> In the tensor-based setting, these special words can be modelled by exploiting additional mathematical structures, such as Frobenius algebras and bialgebras.

Frobenius algebras in **FdVect**

- Given a symmetric CCC $(\mathcal{C}, \otimes, I)$, an object $X \in \mathcal{C}$ has a Frobenius structure on it if there exist morphisms:

  $$\Delta : X \to X \otimes X \; , \; \iota : X \to I \qquad \text{and} \qquad \mu : X \otimes X \to X \; , \; \zeta : I \to X$$

  conforming to the Frobenius condition:

  $$(\mu \otimes 1_X) \circ (1_X \otimes \Delta) = \Delta \circ \mu = (1_X \otimes \mu) \circ (\Delta \otimes 1_X)$$

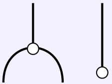
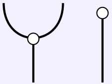
- In **FdVect**, any vector space V with a fixed basis $\{\overrightarrow{v_i}\}_i$ has a commutative special Frobenius algebra over it [Coecke and Pavlovic, 2006]:

  $$\Delta : \overrightarrow{v_i} \mapsto \overrightarrow{v_i} \otimes \overrightarrow{v_i} \qquad \mu : \overrightarrow{v_i} \otimes \overrightarrow{v_i} \mapsto \overrightarrow{v_i}$$

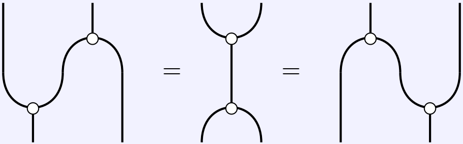
- It can be seen as copying and merging of the basis.

Graphical representation

- Frobenius maps:

$(\Delta, \iota) =$  $(\mu, \zeta) =$ 

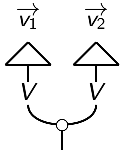
- Frobenius condition:



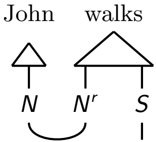
Merging (1/2)

- In **FdVect**, the merging μ-map becomes element-wise vector multiplication:

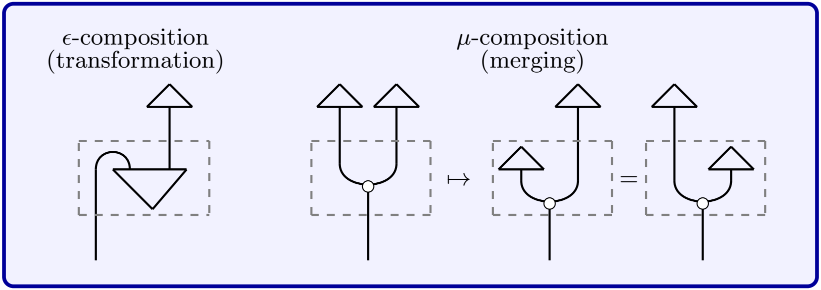
$$\mu(\overrightarrow{v_1} \otimes \overrightarrow{v_2}) = \overrightarrow{v_1} \odot \overrightarrow{v_2} =$$



- An alternative form of composition between operands of the same order; both of them contribute equally to the final result
- Different from standard ϵ-composition, which has a transformational effect. An intransitive verb, for example, is a map $N \rightarrow S$ that transforms a noun into a sentence:



Merging (2/2)

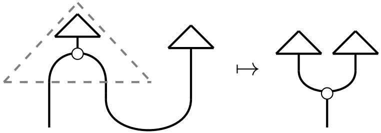


Applications of merging in linguistics:

- Noun modification by relative clauses [Sadrzadeh et al., MoL 2013]
- Modelling intonation at sentence level [Kartsaklis and Sadrzadeh, MoL 2015]
- Modelling non-compositional compounds (e.g. 'pet-fish') [Coecke and Lewis, QI 2015]
- Modelling coordination [Kartsaklis (2016)]

Copying

- In **FdVect**, the Δ-map converts vectors to diagonal matrices
- It can be seen as duplication of information; a single wire is split in two; i.e. a maximally entangled state
- A form of type-raising (converts an atomic type to a function) [Kartsaklis et al., COLING 2012]:



- A means of syntactic movement; the same word can efficiently interact with different parts of the sentence [Sadrzadeh et al., MoL 2013]

Coordination and Frobenius maps

Coordination

The grammatical connection of two or more words, phrases, or clauses to give them equal emphasis and importance. The connected elements, or conjuncts, behave as one.

Merging and copying are the key processes of coordination:

$$context\ c_1\ conj\ c_2 \mapsto [context\ c_1]\ conj\ [context\ c_2]$$

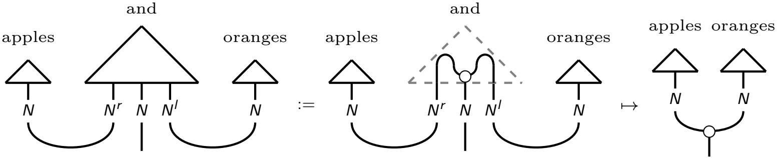
(1) Mary studies [philosophy] and [history] $\models$
[Mary studies philosophy] and [Mary studies history]

(2) John [sleeps] and [snores] $\models$
[John sleeps] and [John snores]

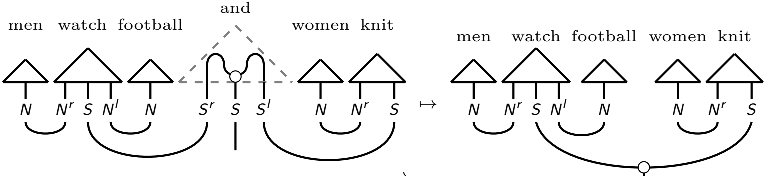
Coordinating atomic types

Coordination morphism:

$$\overline{conj}_X : I \xrightarrow{\eta^r_X \otimes \eta^l_X} X^r \otimes X \otimes X \otimes X^l \xrightarrow{1_{X^r} \otimes \mu_X \otimes 1_{X^l}} X^r \otimes X \otimes X^l$$



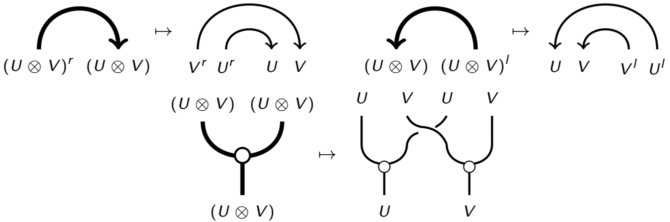
$$(\epsilon^r_N \otimes 1_N \otimes \epsilon^l_N) \circ (\overrightarrow{apples} \otimes \overline{conj}_N \otimes \overrightarrow{oranges}) = \mu(\overrightarrow{apples} \otimes \overrightarrow{oranges}) = \overrightarrow{apples} \odot \overrightarrow{oranges}$$



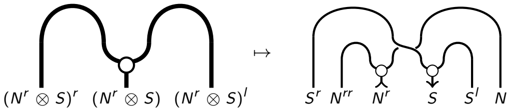
$$(\overrightarrow{men}^{\mathsf{T}} \times \overline{watch} \times \overrightarrow{football}) \odot (\overrightarrow{women}^{\mathsf{T}} \times \overline{knit})$$

Coordinating compound types

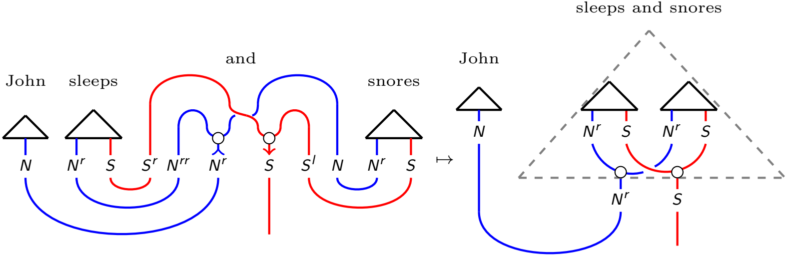
- Lifting the maps to compound objects gives:



- For the case of a verb phrase, we get:



Coordinating verb phrases



1. The subject of the coordinate structure ('John') is copied at the N^r input of the coordinator;
2. the first branch interacts with verb 'sleeps' and the second one with verb 'snores'; and
3. the S wires of the two verbs that carry the individual results are merged together with μ-composition.

$$\overrightarrow{John}^{\mathsf{T}} \times (\overline{sleep} \odot \overline{snore})$$

($\odot$ here denotes the Hadamard product between matrices)

Outline

Word vectors as quantum states

- We take words to be "quantum systems", and word vectors specific states of these systems:

$$|w\rangle = c_1|k_1\rangle + c_2|k_2\rangle + \ldots + c_n|k_n\rangle$$

- Each element of the ONB $\{|k_i\rangle\}_i$ is essentially an atomic symbol:

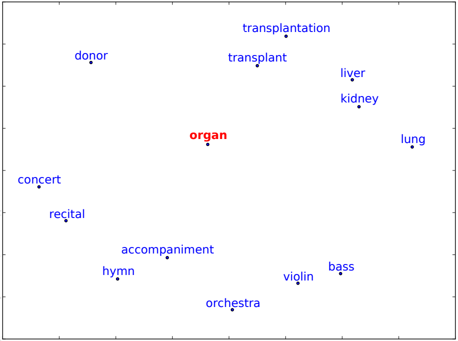
$$|cat\rangle = 12|milk'\rangle + 8|cute'\rangle + \ldots + 0|bank'\rangle$$

- In other words, a word vector is a probability distribution over atomic symbols
- $|w\rangle$ is a pure state: when word w is seen alone, it is like co-occurring with all the basis words with strengths denoted by the various coefficients.

Lexical ambiguity

We distinguish between two types of lexical ambiguity:

- In cases of homonymy (*organ*, *bank*, *vessel* etc.), due to some historical accident the same word is used to describe two (or more) completely unrelated concepts.



- Polysemy relates to subtle deviations between the different senses of the same word.

Encoding homonymy with mixed states

- Ideally, every disjoint meaning of a homonymous word must be represented by a distinct pure state:

$$|bank_{fin}\rangle = a_1|k_1\rangle + a_2|k_2\rangle + \ldots + a_n|k_n\rangle$$
$$|bank_{riv}\rangle = b_1|k_1\rangle + b_2|k_2\rangle + \ldots + b_n|k_n\rangle$$

- $\{a_i\}_i \neq \{b_i\}_i$, since the financial sense and the river sense are expected to be seen in drastically different contexts
- So we have two distinct states describing the same system
- We cannot be certain under which state our system may be found – we only know that the former state is more probable than the latter
- In other words, the system is better described by a probabilistic mixture of pure states, i.e. a mixed state.

Density operators

- Mathematically, a mixed state is represented by a density operator:

$$\rho(w) = \sum_i p_i |s_i\rangle\langle s_i|$$

- For example:

$$\rho(bank) = 0.80|bank_{fin}\rangle\langle bank_{fin}| + 0.20|bank_{riv}\rangle\langle bank_{riv}|$$

- A density operator is a probability distribution over vectors.

Properties of a density operator ρ

- Positive semi-definite: $\langle v|\rho|v\rangle \geq 0 \;\; \forall v \in \mathcal{H}$
- Of trace one: $\text{Tr}(\rho) = 1$
- Self-adjoint: $\rho = \rho^\dagger$

Complete positivity: The CPM construction

In order to apply the new formulation on the categorical model of Coecke et al. we need:

- to replace word vectors with density operators
- to replace linear maps with completely positive linear maps, i.e. maps that send positive operators to positive operators while respecting the monoidal structure.

Selinger (2007):

Any dagger compact closed category is associated with a category in which the objects are the objects of the original category, but the maps are completely positive maps.

From vectors to density operators

- The passage from a grammar to distributional meaning is defined according to the following composition:

$$P(\Sigma, \mathcal{B}) \xrightarrow{\mathcal{F}} \textbf{FdHilb} \xrightarrow{\mathcal{L}} \textbf{CPM(FdHilb)}$$

- The meaning of a sentence $w_1 w_2 \ldots w_n$ with grammatical derivation α becomes:

$$\mathcal{L}(\mathcal{F}(\alpha))\,(\rho(w_1) \otimes_{\text{CPM}} \rho(w_2) \otimes_{\text{CPM}} \ldots \otimes_{\text{CPM}} \rho(w_n))$$

- Composition takes this form:

$$\text{Subject-intransitive verb: } \rho_{IN} = \text{Tr}_N(\rho(v) \circ (\rho(s) \otimes 1_S))$$
$$\text{Adjective-noun: } \rho_{AN} = \text{Tr}_N(\rho(adj) \circ (1_N \otimes \rho(n)))$$
$$\text{Subj-trans. verb-Obj: } \rho_{TS} = \text{Tr}_{N,N}(\rho(v) \circ (\rho(s) \otimes 1_S \otimes \rho(o)))$$

[Kartsaklis DPhil thesis (2015)]

[Piedeleu, Kartsaklis, Coecke, Sadrzadeh (2015)]

Entropy as degree of ambiguity

Von Neumann entropy:

For a d.m. ρ with eigen-decomposition $\sum_i e_i|n_i\rangle\langle n_i|$:

$$S(\rho) = -\mathrm{Tr}(\rho \ln \rho) = -\sum_i e_i \ln e_i$$

- Von Neumann entropy shows how ambiguity evolves from words to compounds
- **Disambiguation = purification:** Entropy of *'vessel'* is 0.25, but entropy of *'vessel that sails'* is 0.01 (i.e. almost a pure state).

Outline

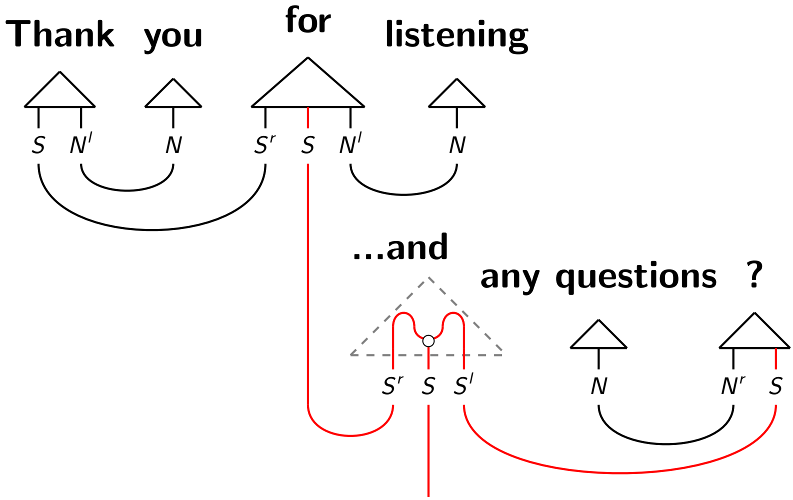
Conclusions and future work

Main points:

- Tensor-based models of meaning provide a linguistically motivated procedure for computing the meaning of phrases and sentences.
- Words of relational nature, such as verbs and adjectives, become (multi-)linear maps acting on noun vectors.
- A test-bed for studying compositional aspects of language at a deeper level.

Future work:

- The application of a logic remains an open problem.
- The density-operator formulation opens various new possibilities to be explored in the future.
- A large-scale evaluation on unconstrained text is remain to be done.


Thank you
for
listing
S N^l N S^r S N^l N
...and
any questions ?
S^r S S^l N N^r S

References I

Baroni, M. and Zamparelli, R. (2010).
Nouns are Vectors, Adjectives are Matrices.
In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Coecke, B. and Lewis, M. (2016).
Quantum Interaction: 9th International Conference, QI 2015, Filzbach, Switzerland, July 15-17, 2015, Revised Selected Papers, chapter A Compositional Explanation of the 'Pet Fish' Phenomenon, pages 179–192.
Springer International Publishing, Cham.

Coecke, B. and Pavlovic, D. (2006).
Quantum measurements without sums.
arXiv preprint quant-ph/0608035.

Coecke, B., Sadrzadeh, M., and Clark, S. (2010).
Mathematical Foundations for a Compositional Distributional Model of Meaning. Lambek Festschrift.
Linguistic Analysis, 36:345–384.

Fried, D., Polajnar, T., and Clark, S. (2015).
Low-rank tensors for verbs in compositional distributional semantics.
In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 731–736, Beijing, China. Association for Computational Linguistics.

Grefenstette, E. and Sadrzadeh, M. (2011).
Experimental support for a categorical compositional distributional model of meaning.
In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404, Edinburgh, Scotland, UK. Association for Computational Linguistics.

References II

Kartsaklis, D. (2015).
Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras.
PhD thesis, University of Oxford.

Kartsaklis, D. (2016).
Coordination in Categorical Compositional Distributional Semantics.
In *Proceedings of the 2016 Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science*. EPTCS.
To appear.

Kartsaklis, D. and Sadrzadeh, M. (2014).
A study of entanglement in a categorical framework of natural language.
In Coecke, B., Hasuo, I., and Panangaden, P., editors, Proceedings 11th workshop on *Quantum Physics and Logic,* Kyoto, Japan, 4-6th June 2014, volume 172 of *Electronic Proceedings in Theoretical Computer Science*, pages 249–261. Open Publishing Association.

Kartsaklis, D. and Sadrzadeh, M. (2016).
A compositional distributional inclusion hypothesis.
Submitted.

Kartsaklis, D., Sadrzadeh, M., and Pulman, S. (2012).
A unified sentence space for categorical distributional-compositional semantics: Theory and experiments.
In *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012): Posters*, pages 549–558, Mumbai, India. The COLING 2012 Organizing Committee.

Kartsaklis, D., Sadrzadeh, M., Pulman, S., and Coecke, B. (2015).
Reasoning about meaning in natural language with compact closed categories and Frobenius algebras.
In Chubb, J., Eskandarian, A., and Harizanov, V., editors, *Logic and Algebraic Structures in Quantum Computing and Information*, Association for Symbolic Logic Lecture Notes in Logic. Cambridge University Press.

References III

Piedeleu, R., Kartsaklis, D., Coecke, B., and Sadrzadeh, M. (2015).
Open system categorical quantum semantics in natural language processing.
In *Proceedings of the 6th Confenrence on Algebra and Coalgebra in Computer Science (CALCO)*, Nijmegen, Netherlands.

Sadrzadeh, M., Clark, S., and Coecke, B. (2013).
The Frobenius anatomy of word meanings I: subject and object relative pronouns.
Journal of Logic and Computation, Advance Access.

Sadrzadeh, M., Clark, S., and Coecke, B. (2014).
The Frobenius anatomy of word meanings II: Possessive relative pronouns.
Journal of Logic and Computation.

Selinger, P. (2007).
Dagger compact closed categories and completely positive maps.
Electronic Notes in Theoretical Computer Science, 170:139–163.