Compositional Distributional Models for Simple Transitive Sentences

Phong Le

Based on
E. Grefenstette and M. Sadrzadeh, *Experimental Support for a Categorical Compositional Distributional Model of Meaning*, EMNLP 2011.

Outline

Formal Semantics

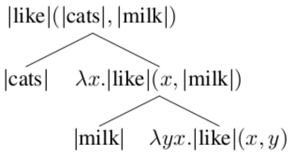
Principle of Compositionality

"The meaning of a whole is a function of the meanings of the parts and of the way they are syntactically combined."

- Formal semantics works well with this principle

$$\text{cats like milk :- } |like|(|cats|, |milk|)$$

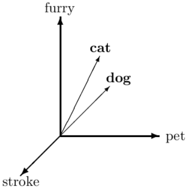
Syntactic Analysis	Semantic Interpretation
S $\rightarrow$ NP VP	$\|VP\|(\|NP\|)$
NP $\rightarrow$ cats, milk, etc.	$\|\text{cats}\|, \|\text{milk}\|, \ldots$
VP $\rightarrow$ Vt NP	$\|Vt\|(\|NP\|)$
Vt $\rightarrow$ like, hug, etc.	$\lambda yx.\|\text{like}\|(x, y), \ldots$



- However, this analysis can only deal with truth or falsity.

Distributional Semantics

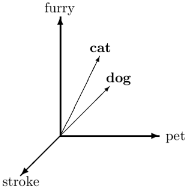
- $\overrightarrow{word} = (c_1^{word}, ..., c_n^{word}) = \sum_i c_i^{word} \overrightarrow{n_i}$
- Offers geometric means to reason about semantic similarity (e.g. *cosin*)



- However, the principle drawback is the non-compositional nature: grammatical structures are ignored

Distributional Semantics

- $\overrightarrow{word} = (c_1^{word}, ..., c_n^{word}) = \sum_i c_i^{word} \overrightarrow{n_i}$
- Offers geometric means to reason about semantic similarity (e.g. *cosin*)



- However, the principle drawback is the non-compositional nature: grammatical structures are ignored

Target

Given $\overrightarrow{sbj}, \overrightarrow{verb}, \overrightarrow{obj}$, calculate $\overrightarrow{\text{sbj verb obj}}$

Point-wise multiplication and tensor product

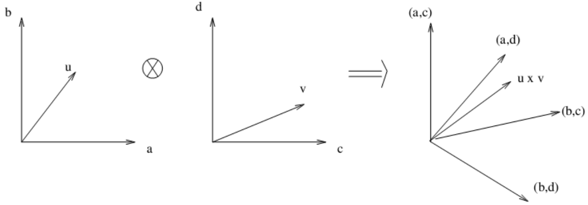
Given a vector space A with basis $\{\overrightarrow{n_i}\}_i$, and two vectors $\overrightarrow{u} = \sum_i c_i^a \overrightarrow{n_i}$ and $\overrightarrow{v} = \sum_i c_i^b \overrightarrow{n_i}$

- point-wise multiplication $\overrightarrow{u} \odot \overrightarrow{v} = \sum_i c_i^a c_i^b \overrightarrow{n_i}$

Point-wise multiplication and tensor product

Given a vector space A with basis $\{\overrightarrow{n_i}\}_i$, and two vectors $\overrightarrow{u} = \sum_i c_i^a \overrightarrow{n_i}$ and $\overrightarrow{v} = \sum_i c_i^b \overrightarrow{n_i}$

- point-wise multiplication $\overrightarrow{u} \odot \overrightarrow{v} = \sum_i c_i^a c_i^b \overrightarrow{n_i}$
- tensor product $\overrightarrow{u} \otimes \overrightarrow{v} = \sum_{i,j} c_i^a c_j^b (\overrightarrow{n_i} \otimes \overrightarrow{n_j})$



$$(u \otimes v)_{(a,d)} = u_a \cdot v_d$$

Outline

Naive Approaches (Mitchell and Lapata, 2008)

Given that $\overrightarrow{word_1}, \overrightarrow{word_2}$ in the same vector space

- Generally,

$$\overrightarrow{word_1\ word_2} = f(\overrightarrow{word_1}, \overrightarrow{word_2})$$

Naive Approaches (Mitchell and Lapata, 2008)

Given that $\overrightarrow{word_1}, \overrightarrow{word_2}$ in the same vector space

- Generally,

$$\overrightarrow{word_1\ word_2} = f(\overrightarrow{word_1}, \overrightarrow{word_2})$$

- Addition model

$$\overrightarrow{word_1\ word_2} = \overrightarrow{word_1} + \overrightarrow{word_2}$$

Naive Approaches (Mitchell and Lapata, 2008)

Given that $\overrightarrow{word_1}, \overrightarrow{word_2}$ in the same vector space

- Generally,

$$\overrightarrow{word_1\ word_2} = f(\overrightarrow{word_1}, \overrightarrow{word_2})$$

- Addition model

$$\overrightarrow{word_1\ word_2} = \overrightarrow{word_1} + \overrightarrow{word_2}$$

- Multiplication model

$$\overrightarrow{word_1\ word_2} = \overrightarrow{word_1} \odot \overrightarrow{word_2}$$

Naive Approaches (Mitchell and Lapata, 2008)

Given that $\overrightarrow{word_1}, \overrightarrow{word_2}$ in the same vector space

- Generally,
$$\overrightarrow{word_1\ word_2} = f(\overrightarrow{word_1}, \overrightarrow{word_2})$$
- Addition model
$$\overrightarrow{word_1\ word_2} = \overrightarrow{word_1} + \overrightarrow{word_2}$$
- Multiplication model
$$\overrightarrow{word_1\ word_2} = \overrightarrow{word_1} \odot \overrightarrow{word_2}$$
- Mixture model
$$\overrightarrow{word_1\ word_2} = \alpha\overrightarrow{word_1} + \beta\overrightarrow{word_2} + \gamma\overrightarrow{word_1} \odot \overrightarrow{word_2}$$

Drawbacks

- Addition and point-wise multiplication are commutative; hence,

$$\overrightarrow{\text{the dog bit the main}} = \overrightarrow{\text{the man bit the dog}}$$

- Those models ignore grammatical structures.

Outline

Categorical Compositional Distribution Model (Grefenstette and Sadrzadeh, 2011)

Vectors of different word types should live in different vector spaces

$$\text{cats} :- n \approx NP$$
$$\text{love} :- n^r s n^l \approx (S \setminus NP)/NP$$
$$\text{milk} :- n \approx NP$$

Categorical Compositional Distribution Model (Grefenstette and Sadrzadeh, 2011)

Vectors of different word types should live in different vector spaces

$$\text{cats} :- n \approx NP$$
$$\text{love} :- n^r s n^l \approx (S \backslash NP)/NP$$
$$\text{milk} :- n \approx NP$$

- a noun is represented by a vector in vector space N
- a transitive verb is represented by a relational matrix in $N \times N$, which is isomorphic with $N \otimes N$
- $\overrightarrow{\text{sbj verb obj}} = \overrightarrow{verb} \odot (\overrightarrow{sbj} \otimes \overrightarrow{obj})$

Noun vectors

Similar to vector spaces in traditional distributional semantics

i	$\overrightarrow{n_i}$	table	map	result	location
1	far	6.6	5.6	7	5.9
2	room	27	7.4	0.99	7.3
3	scientific	0	5.4	13	6.1
4	elect	0	0	4.2	0

Verb vectors

- Given that there are m transitive sentences with the verb *verb*

$$w_{1,1}\ verb\ w_{2,1}$$
$$\ldots$$
$$w_{1,m}\ verb\ w_{2,m}$$

we compute $\overrightarrow{verb} = \sum_k \overrightarrow{w_{1,k}} \otimes \overrightarrow{w_{2,k}}$

Verb vectors

- Given that there are m transitive sentences with the verb *verb*

$$w_{1,1}\ verb\ w_{2,1}$$
$$\ldots$$
$$w_{1,m}\ verb\ w_{2,m}$$

we compute $\overrightarrow{verb} = \sum_k \overrightarrow{w_{1,k}} \otimes \overrightarrow{w_{2,k}}$

- For example, if there are

$$s_1 = \text{table show result}$$

$$s_2 = \text{map show location}$$

then $\overrightarrow{show} = \overrightarrow{table} \otimes \overrightarrow{result} + \overrightarrow{map} \otimes \overrightarrow{location}$

	far	room	scientific	elect
far	79.24	47.41	119.96	27.72
room	232.66	80.75	396.14	113.2
scientific	32.94	31.86	32.94	0
elect	0	0	0	0

Outline

Experiment

- Evaluating how well compositional models disambiguate ambiguous words given the context of potentially disambiguating nouns.

Experiment

- Evaluating how well compositional models disambiguate ambiguous words given the context of potentially disambiguating nouns.
- For example, *show* could be
  - *express* (give expression to), or
  - *picture* (show in, or as in, a picture)

  but we can easily disambiguate it in
  - *table show result* $\approx$ *express*
  - *map show location* $\approx$ *picture*

  express, *picture* are called landmarks of *show*.

Dataset

- 200 entries with classification of 'LOW' or 'HIGH', e.g.

table show result	table express result	HIGH
map show location	map picture location	HIGH
table show result	table picture result	LOW
map show location	map express location	LOW

Dataset

- 200 entries with classification of 'LOW' or 'HIGH', e.g.

table show result	table express result	HIGH
map show location	map picture location	HIGH
table show result	table picture result	LOW
map show location	map express location	LOW

- 25 evaluators were asked to offer scores between 1 to 7 for each entry (UpperBound)

Dataset

- 200 entries with classification of 'LOW' or 'HIGH', e.g.

table show result	table express result	HIGH
map show location	map picture location	HIGH
table show result	table picture result	LOW
map show location	map express location	LOW

- 25 evaluators were asked to offer scores between 1 to 7 for each entry (UpperBound)

Evaluation methodology

Spearman's ρ against human judgements is used (higher is better)

Results

Model	High	Low	ρ
Baseline	0.47	0.44	0.16
Add	0.90	0.90	0.05
Multiply	0.67	0.59	0.17
Categorical	**0.73**	**0.72**	**0.21**
UpperBound	4.80	2.49	0.62

Outline

How does it work? An Intuitive Explanation

Recall

$$\overrightarrow{\text{sbj verb obj}} = \overrightarrow{verb} \odot \left(\overrightarrow{sbj} \otimes \overrightarrow{obj}\right)$$

$\overrightarrow{sbj} \otimes \overrightarrow{obj}$ shifts the meaning of *verb* toward the direction which it supports.

$$\begin{array}{ccc} \overrightarrow{express} \leftarrow & \overrightarrow{show} & \rightarrow \overrightarrow{picture} \\ \overrightarrow{express} \odot \left(\overrightarrow{table} \otimes \overrightarrow{result}\right) \leftarrow & \overrightarrow{show} \odot \left(\overrightarrow{table} \otimes \overrightarrow{result}\right) & \rightarrow \overrightarrow{picture} \odot \left(\overrightarrow{table} \otimes \overrightarrow{result}\right) \end{array}$$

How does it work? An Intuitive Explanation

Recall

$$\overrightarrow{\text{sbj verb obj}} = \overrightarrow{verb} \odot (\overrightarrow{sbj} \otimes \overrightarrow{obj})$$

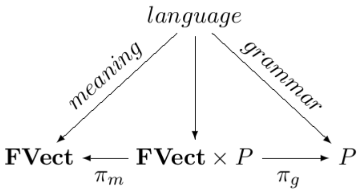
$\overrightarrow{sbj} \otimes \overrightarrow{obj}$ shifts the meaning of *verb* toward the direction which it supports.

$$\begin{array}{ccccc} \overrightarrow{express} \leftarrow & & \overrightarrow{show} & & \rightarrow \overrightarrow{picture} \\ \overrightarrow{express} \odot (\overrightarrow{table} \otimes \overrightarrow{result}) \leftarrow & & \overrightarrow{show} \odot (\overrightarrow{table} \otimes \overrightarrow{result}) & & \rightarrow \overrightarrow{picture} \odot (\overrightarrow{table} \otimes \overrightarrow{result}) \end{array}$$

How does it work? A Brief Mathematical Explanation

(see B. Coecke, M. Sadrzadeh, S. Clark, *Mathematical Foundation for a Compositional Distributional Model of Meaning*)

- Pregroup grammar P and the set of vector spaces **FVect** share the same high level math structure *compact closed category*.
- Therefore, we can work on the combined category $\mathbf{FVect} \times P$ in order to let grammar P drive compositionality in **FVect**



Outline

Summary

- The principle drawback of Distributional Methods is their non-compositional nature

Summary

- The principle drawback of Distributional Methods is their non-compositional nature
- Naive approaches use addition, point-wise multiplication, which are commutative and ignore grammatical structures

Summary

- The principle drawback of Distributional Methods is their non-compositional nature
- Naive approaches use addition, point-wise multiplication, which are commutative and ignore grammatical structures
- Categorical Compositional Distributional Model
  - represents different word types in different vector spaces
  - uses linear maps, which are driven by a pregroup grammar, to combine vectors
  - performed better than naive approaches
  - has solid math foundations