Vector-based Models of Semantic Composition

Jeff Mitchell and Mireia Lapata (2008)

Aina Garí Soler

Prof. Dr. Manfred Pinkal

November 2015

Selected Topics in Semantics and Discourse

EMLCT

UNIVERSITÄT DES SAARLANDES

Overview

- The problem of composition and previous attempts
- Composition models proposed
- Evaluation
  - Task design
  - Comparison of model's performance with human judgments

- Results
- Discussion

Composition

- Most vector-based models of word meaning are concerned with **words in isolation**

Composition

- Most vector-based models of word meaning are concerned with **words in isolation**

1 vector = 1 word

Composition

- Most vector-based models of word meaning are concerned with **words in isolation**

1 vector = 1 word

- But we want to model (compositional) **phrases or sentences** as well

Modeling composition: previous attempts

- Ignore word order (vector averaging: additive models)

a. *It was not the sales manager who hit the bottle that day, but the office worker with the serious drinking problem.*

b. *That day the office manager, who was drinking, hit the problem sales worker with a bottle, but it was not serious.*

(a = b)

Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

Expression of composition:

$$p = f(u, v, R, K)$$

Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

Expression of composition:

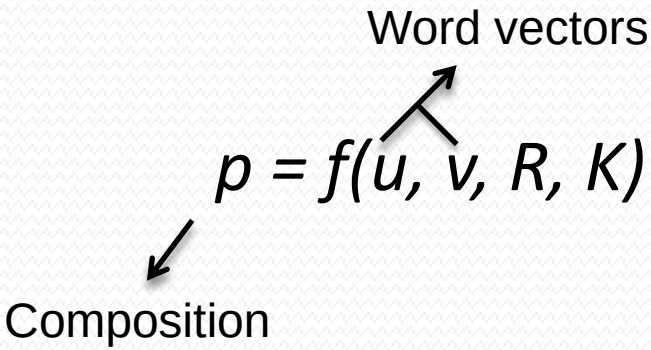
$$p = f(u, v, R, K)$$

Composition

Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

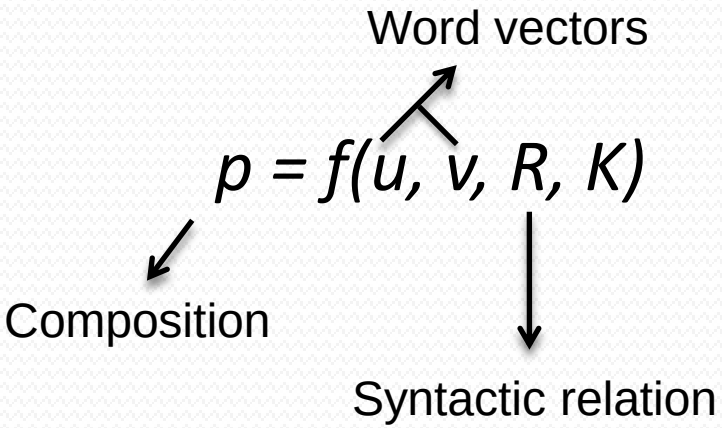
Expression of composition:



Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

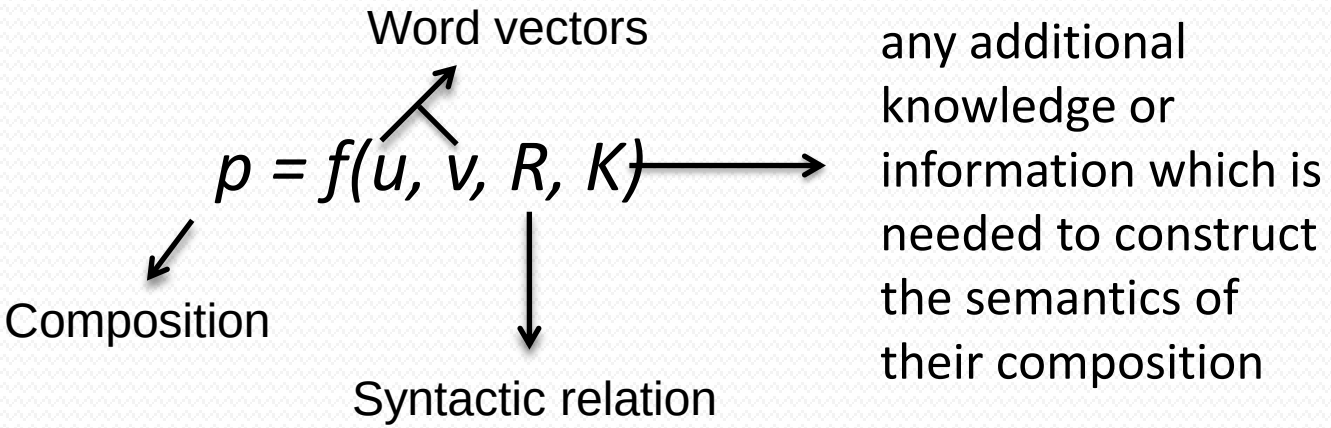
Expression of composition:



Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

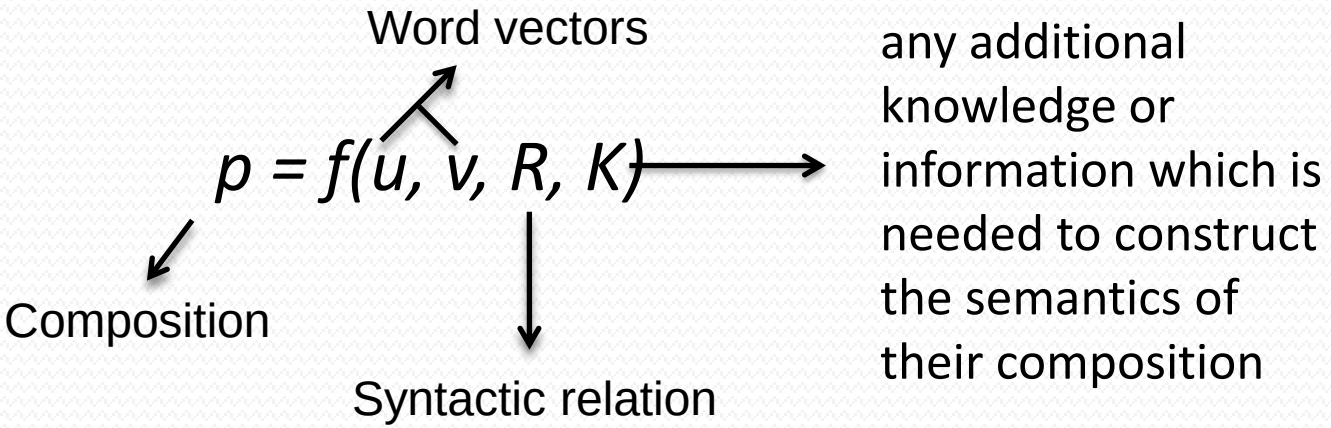
Expression of composition:



Composition Models

- Formulate semantic composition as a function of two vectors **u** and **v** obtained from a Distributional model.

Expression of composition:



- If we fix R and ignore K: ***p = f(u, v)***

Composition Models

Additive models (p = Au + Bv)

Simple addition:

$$p_i = u_i + v_i$$

	animal	stable	village	gallop	jokey
horse	0	6	2	10	4
run	1	8	4	4	0

Composition Models

Additive models (p = Au + Bv)

Simple addition:

$$p_i = u_i + v_i$$

	animal	stable	village	gallop	jokey
horse	0	6	2	10	4
run	1	8	4	4	0
(horse, run) =	(1,	14,	6,	14,	4)

Composition Models

Additive models (p = Au + Bv)

Simple addition:

$$p_i = u_i + v_i$$

Weighted addition:

$$p_i = \alpha u_i + \beta v_i \qquad (p_i = v_i)$$

Kintsch's model:

$$p = u + v + \sum n$$

Composition Models

Multiplicative models (p = Cuv)

$p_i = u_i \cdot v_i$

	animal	stable	village	gallop	jokey
horse	0	6	2	10	4
run	1	8	4	4	0

Composition Models

Multiplicative models (p = Cuv)

$p_i = u_i \cdot v_i$

	animal	stable	village	gallop	jokey
horse	0	6	2	10	4
run	1	8	4	4	0
(horse, run) =	(0,	48,	8,	40,	0)

Composition Models

Multiplicative models (p = Cuv)

$p_i = u_i \cdot v_i$

Combined (additive and multiplicative):

$pi = \alpha u_i + \beta v_i + \gamma u_i v_i$

Evaluation

- Sentence similarity task (Kintsch, 2001)

Evaluation

- Sentence similarity task (Kintsch, 2001)

  *The horse **ran***

  *The color **ran***

Evaluation

- Sentence similarity task (Kintsch, 2001)

*The horse **ran*** (gallop)

*The color **ran*** (dissolve)

Evaluation

- Sentence similarity task (Kintsch, 2001)

*The horse **ran*** (gallop)

*The color **ran*** (dissolve)

Evaluation

- Sentence similarity task (Kintsch, 2001)

*The horse **ran*** (gallop)

*The color **ran*** (dissolve)

Different subjects select <u>different senses</u> of a verb

Evaluation

- Sentence similarity task (Kintsch, 2001)

*The horse **ran*** (gallop)

*The color **ran*** (dissolve)

Different subjects select <u>different senses</u> of a verb

<u>Shift of position</u> in semantic space

Evaluation

- Sentence similarity task (Kintsch, 2001)

  p(horse, run) is closer to *gallop*

  p(color, run) is closer to *dissolve*

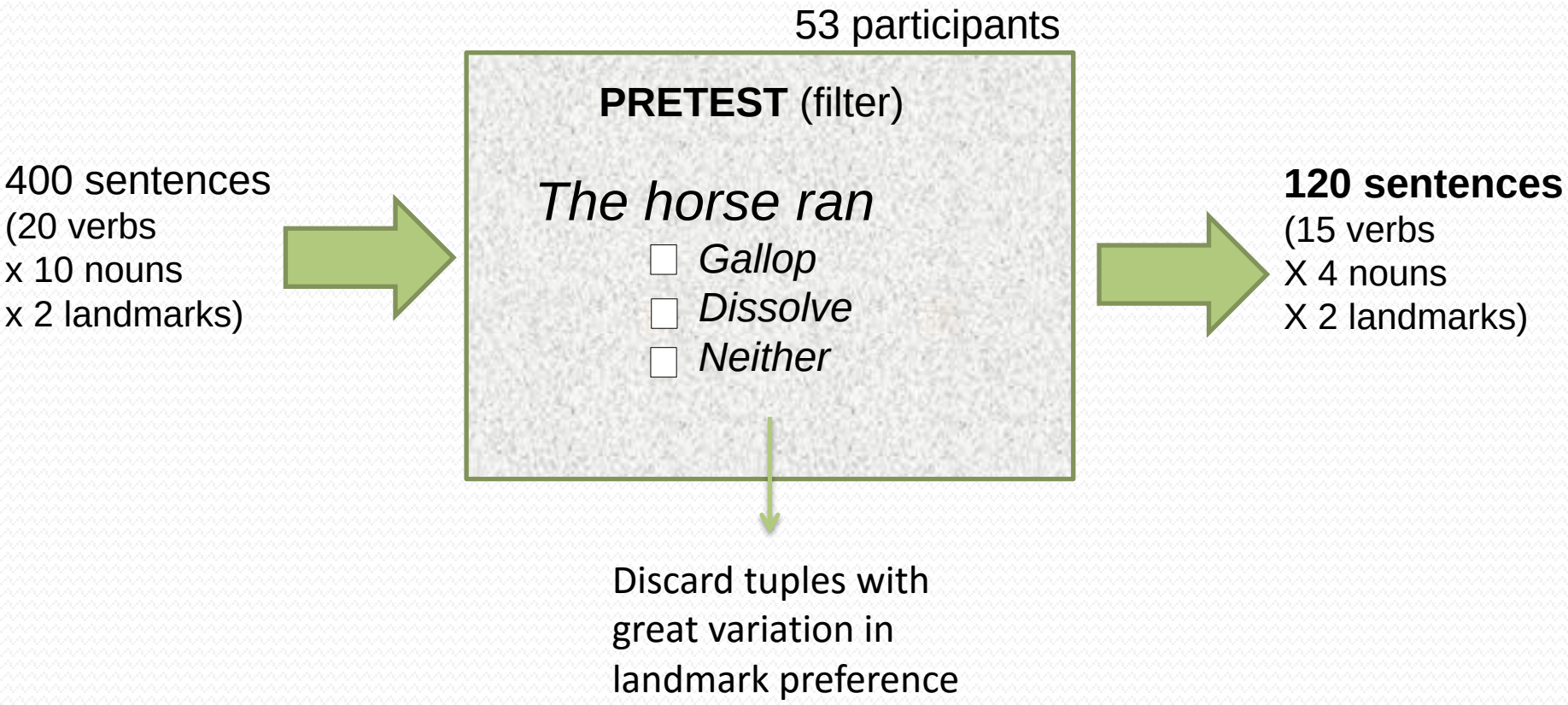
Evaluation

- Sentence similarity task: Materials and design
  - **subject + intransitive verb** structures (BNC)
  - 2 landmarks for each tuple (subject, verb), synonyms of different (incompatible) senses of the verb (WordNet)

    Landmarks of *(horse, run) = gallop, dissolve*

Evaluation

- Sentence similarity task: Materials and design



Evaluation: Procedure

The fire glowed
The fire burned

Evaluation: Procedure

The fire glowed
The fire burned

The fire glowed
The fire beamed

Evaluation: Procedure

The fire glowed
The fire burned

(HIGH) 

The fire glowed
The fire beamed

 (LOW)

Evaluation: Procedure

The fire glowed
The fire burned

(HIGH) 

The fire glowed
The fire beamed

 (LOW)

Rate similarity from 1 to 7

Evaluation: Procedure

The fire glowed
The fire burned

(HIGH) 

The fire glowed
The fire beamed

 (LOW)

Rate similarity from 1 to 7

OBJECTIVE: See what composition model yields results closer to human judgments in this task.

Evaluation

- Sentence similarity task: Analysis of similarity ratings

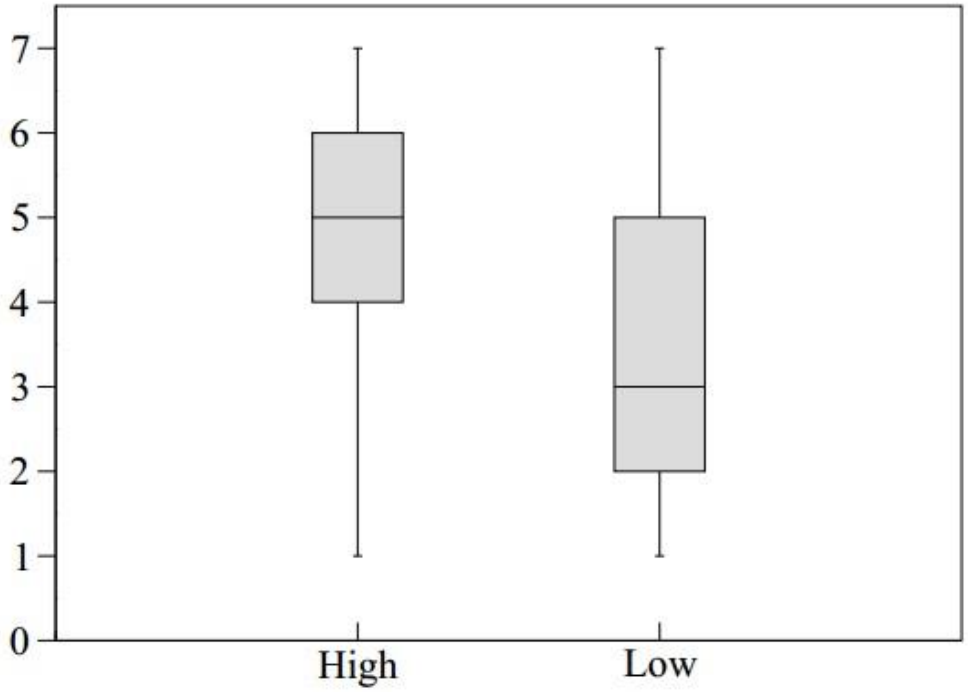


Figure 2: Distribution of elicited ratings for High and Low similarity items

Not an easy task: overlap in High and Low

But the difference is significant

Evaluation: Model Parameters

Optimize distributional parameters in a word-based semantic similarity task:

- Different contexts
- Variety of dimensions (from 50 to 500.000)
- Vector component definitions (PMI, Log likelihood ratio)
- Similarity measures (cosine, confusion probability)

Evaluation: Model Parameters

Best set of distributional parameters:

- context window of 5 words on either side of the target word
- cosine measure
- 2000 vector components (with content words)

Evaluation: Model Parameters

Adjusting concrete models' parameters

- Weights of
  - Weighted additive model: 80% verb, 20% noun
  - Combined model: 95% verb, 0% noun, 5% mult. (with held-out set)
- Neighbours in Kintsch's additive model: 1

Evaluation Methodology

- Estimate cosine similarity between sentences in each pair using the different models.

- Correlate all the individual human judgments with those of the models.

Evaluation Methodology

- Estimate cosine similarity between sentences in each pair using the different models.

- Correlate all the individual human judgments with those of the models.

EXPECTATION: Better models yield similarity scores like those observed in the human ratings and correlate better with them.

Results

Model	High	Low	ρ
NonComp	0.27	0.26	0.08**
Add	0.59	0.59	0.04*
WeightAdd	0.35	0.34	0.09**
Kintsch	0.47	0.45	0.09**
Multiply	0.42	0.28	0.17**
Combined	0.38	0.28	0.19**
UpperBound	4.94	3.25	0.40**

Table 2: Model means for High and Low similarity items and correlation coefficients with human judgments (*: $p < 0.05$, **: $p < 0.01$)

Results

Fail to distinguish between High and Low

Model	High	Low	ρ
NonComp	0.27	0.26	0.08**
Add	0.59	0.59	0.04*
WeightAdd	0.35	0.34	0.09**
Kintsch	0.47	0.45	0.09**
Multiply	0.42	0.28	0.17**
Combined	0.38	0.28	0.19**
UpperBound	4.94	3.25	0.40**

Table 2: Model means for High and Low similarity items and correlation coefficients with human judgments (*: $p < 0.05$, **: $p < 0.01$)

Results

Model	High	Low	ρ
NonComp	0.27	0.26	0.08**
Add	0.59	0.59	0.04*
WeightAdd	0.35	0.34	0.09**
Kintsch	0.47	0.45	0.09**
Multiply	0.42	0.28	0.17**
Combined	0.38	0.28	0.19**
UpperBound	4.94	3.25	0.40**

Significant distinction

Table 2: Model means for High and Low similarity items and correlation coefficients with human judgments (*: $p < 0.05$, **: $p < 0.01$)

Results

Model	High	Low	ρ
NonComp	0.27	0.26	0.08**
Add	0.59	0.59	0.04*
WeightAdd	0.35	0.34	0.09**
Kintsch	0.47	0.45	0.09**
Multiply	0.42	0.28	0.17**
Combined	0.38	0.28	0.19**
UpperBound	4.94	3.25	0.40**

Significant distinction →

Table 2: Model means for High and Low similarity items and correlation coefficients with human judgments (*: $p < 0.05$, **: $p < 0.01$)

Results

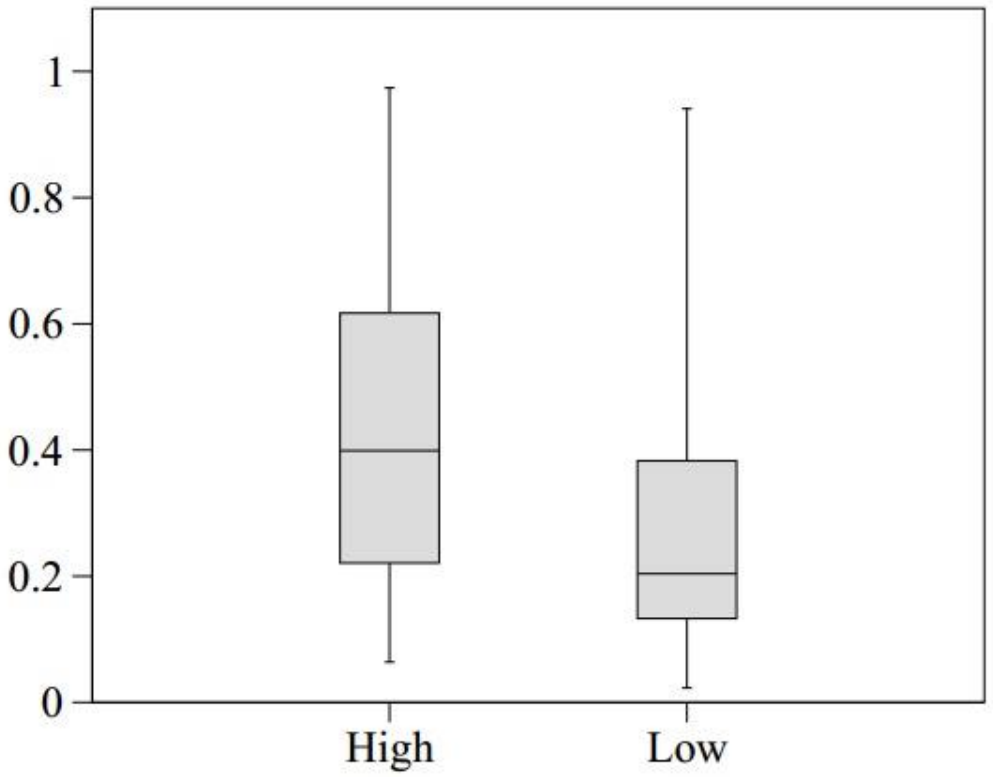


Figure 3: Distribution of predicted similarities for the vector multiplication model on High and Low similarity items

Discussion

- Models using multiplicative combinations are superior to additive models, at least for this task.
- Additive models had been used to modeling the gist of a document rather than the meaning of its sentences.

Discussion

- Additive models capture composition by considering all vector components.
- Multiplicative models consider a subset (non-zero components). The resulting vector is sparser but expresses the meaning more succinctly.

Discussion

- **Further research** is needed in
  - other structures (not only subj+verb)
  - other models

Discussion

- **Further research** is needed in
  - other structures (not only subj+verb)
  - other models

- In particular, the general class of multiplicative models appears to be a fruitful area to explore.

Thank you for your attention!

Any questions?