EBERHARD KARLS UNIVERSITÄT TÜBINGEN

FACULTY OF HUMANITIES

Evaluating Semantic Composition of German Compounds

Corina Dima, Jianqiang Ma and Erhard Hinrichs

University of Tübingen, Department of Linguistics and SFB 833, Germany

Wer wurmt der Ohrwurm? An interdisciplinary, cross-lingual perspective on the role of constituents in multi-word expressions, DGfS 2017, 09.03.2017

Motivation

- vector space models of language (Mikolov et al., 2013; Pennington et al., 2014) create meaningful representations for the individual words in a language
- how to create meaningful, reusable representations for longer word sequences – in this work – for German compounds?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Motivation

- vector space models of language (Mikolov et al., 2013; Pennington et al., 2014) create meaningful representations for the individual words in a language
- how to create meaningful, reusable representations for longer word sequences – in this work – for German compounds?

Solution 1

Add compounds to the dictionary of the language model and directly learn representations for them.

[intractable due to the productivity of compounding]

Motivation

- vector space models of language (Mikolov et al., 2013; Pennington et al., 2014) create meaningful representations for the individual words in a language
- how to create meaningful, reusable representations for longer word sequences – in this work – for German compounds?

Solution 1

Add compounds to the dictionary of the language model and directly learn representations for them.

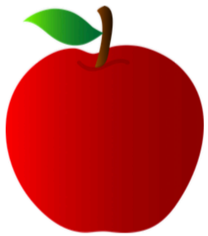
[intractable due to the productivity of compounding]

Solution 2

Use semantic composition to build the meaning of the compound starting from the meaning of individual words.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Semantic Composition

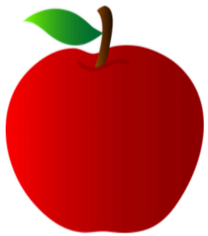


Apfel + Baum → Apfelbaum

0.3	0.1	0.7	1.3	0.2	u	0.5	0.9	0.1	0.4	1.2	v	0.2	1.0	0.6	0.7	1.1	w

$$f\left(\begin{bmatrix}0.3 & 0.1 & 0.7 & 1.3 & 0.2\end{bmatrix} u,\ \begin{bmatrix}0.5 & 0.9 & 0.1 & 0.4 & 1.2\end{bmatrix} v\right) = \begin{bmatrix}?? & ?? & ?? & ?? & ??\end{bmatrix} p$$

Semantic Composition



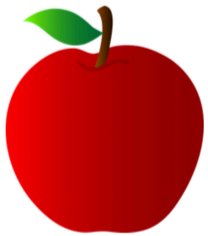
Apfel + Baum → Apfelbaum

$$\boxed{0.3}\boxed{0.1}\boxed{0.7}\boxed{1.3}\boxed{0.2}\;u \qquad \boxed{0.5}\boxed{0.9}\boxed{0.1}\boxed{0.4}\boxed{1.2}\;v \qquad \boxed{0.2}\boxed{1.0}\boxed{0.6}\boxed{0.7}\boxed{1.1}\;w$$

$$f\left(\boxed{0.3}\boxed{0.1}\boxed{0.7}\boxed{1.3}\boxed{0.2}\;u,\ \boxed{0.5}\boxed{0.9}\boxed{0.1}\boxed{0.4}\boxed{1.2}\;v\right) = \boxed{??}\boxed{??}\boxed{??}\boxed{??}\boxed{??}\;p$$

- learn a composition function **f** that combines the representations of the constituents *Apfel* and *Baum* into the representation of the compound *Apfelbaum*

Semantic Composition



Apfel + Baum → Apfelbaum

$$\boxed{0.3}\boxed{0.1}\boxed{0.7}\boxed{1.3}\boxed{0.2}\,u \qquad \boxed{0.5}\boxed{0.9}\boxed{0.1}\boxed{0.4}\boxed{1.2}\,v \qquad \boxed{0.2}\boxed{1.0}\boxed{0.6}\boxed{0.7}\boxed{1.1}\,w$$

$$f\left(\boxed{0.3}\boxed{0.1}\boxed{0.7}\boxed{1.3}\boxed{0.2}\,u,\ \boxed{0.5}\boxed{0.9}\boxed{0.1}\boxed{0.4}\boxed{1.2}\,v\right) = \boxed{??}\boxed{??}\boxed{??}\boxed{??}\boxed{??}\,p$$

- learn a composition function **f** that combines the representations of the constituents *Apfel* and *Baum* into the representation of the compound *Apfelbaum*
- the composed representation of *Apfelbaum* should be similar (cosine similarity) to its corpus-estimated representation

How to Choose the Composition Function?

Model	Formula
Mitchel & Lapata (2010) • vector addition, vector multiplication, etc.	$p = \lambda u + \beta v$ $p = u \odot v$
Baroni & Zamparelli (2010) • matrix for the adjective, vector for the noun	$p = \mathcal{U}v$
Zanzotto et al. (2010) • linear combination of vectors and matrices for both components	$p = M_1 u + M_2 v$
Socher et al. (2010) • global matrix to combine component vectors + nonlinearity	$p = g(W[u;v])$
Socher et al. (2012) • use a individual word matrix to modify each word before combining it though the global matrix + nonlinearity	$p = g(W[\mathcal{V}u; \mathcal{U}v])$

Empirically: Test All Models

Dataset

- 34497 compounds from the German wordnet, GermaNet, v9.0
- train-test-dev splits (70/20/10)
- with splitting information: immediate head and modifier for every compound (Henrich & Hinrichs, 2011)
- frequency filtered: modifier, head and compound with minimum frequency 500 in the support corpus

Empirically: Test All Models

Dataset

- 34497 compounds from the German wordnet, GermaNet, v9.0
- train-test-dev splits (70/20/10)
- with splitting information: immediate head and modifier for every compound (Henrich & Hinrichs, 2011)
- frequency filtered: modifier, head and compound with minimum frequency 500 in the support corpus

Word representations

- Trained 50, 100, 200 and 300 dimensional word representations using GloVe (Pennington et al., 2014)
- 10 billion words corpus from DECOW14AX (Schäfer, 2015); used 1 million word vocabulary (frequency min. 100)

Train Composition Models

- estimate the parameters of the composition functions using the training split of the dataset

  - start from corpus-induced representations for

    head, **modifier**, **compound**

  - apply the composition function => composed representation

    f(head, modifier) = compound

Train Composition Models

- estimate the parameters of the composition functions using the training split of the dataset
  - start from corpus-induced representations for

    head, modifier, compound

  - apply the composition function => composed representation

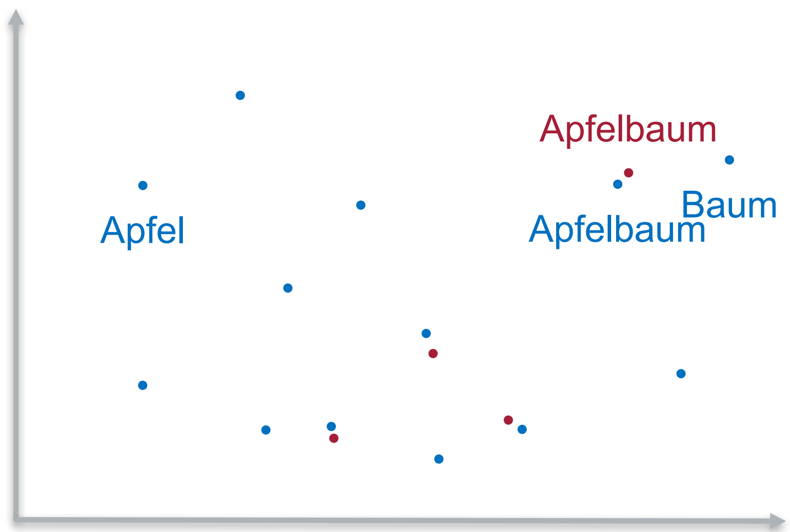
    f(head, modifier) = compound

- **objective function** for training: minimize the mean squared error between the composed and the corpus-induced compound representations

  compound ⇔ compound

Evaluate Composition Models

- **intuition**: a good composition model produces composed representations such that the corpus-observed representations of the **same compounds** are their nearest neighbors in the vector space



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Evaluate Composition Models (2)

- compute the **ranks** of the composed representations in the test set

- **rank computation**
  1. compute cosine distance between the composed representation (**compound**) and *all* the **corpus-induced vectors**
  2. sort, most similar first
  3. the rank is the position of the corresponding corpus-induced vector (**compound**) in the sorted list

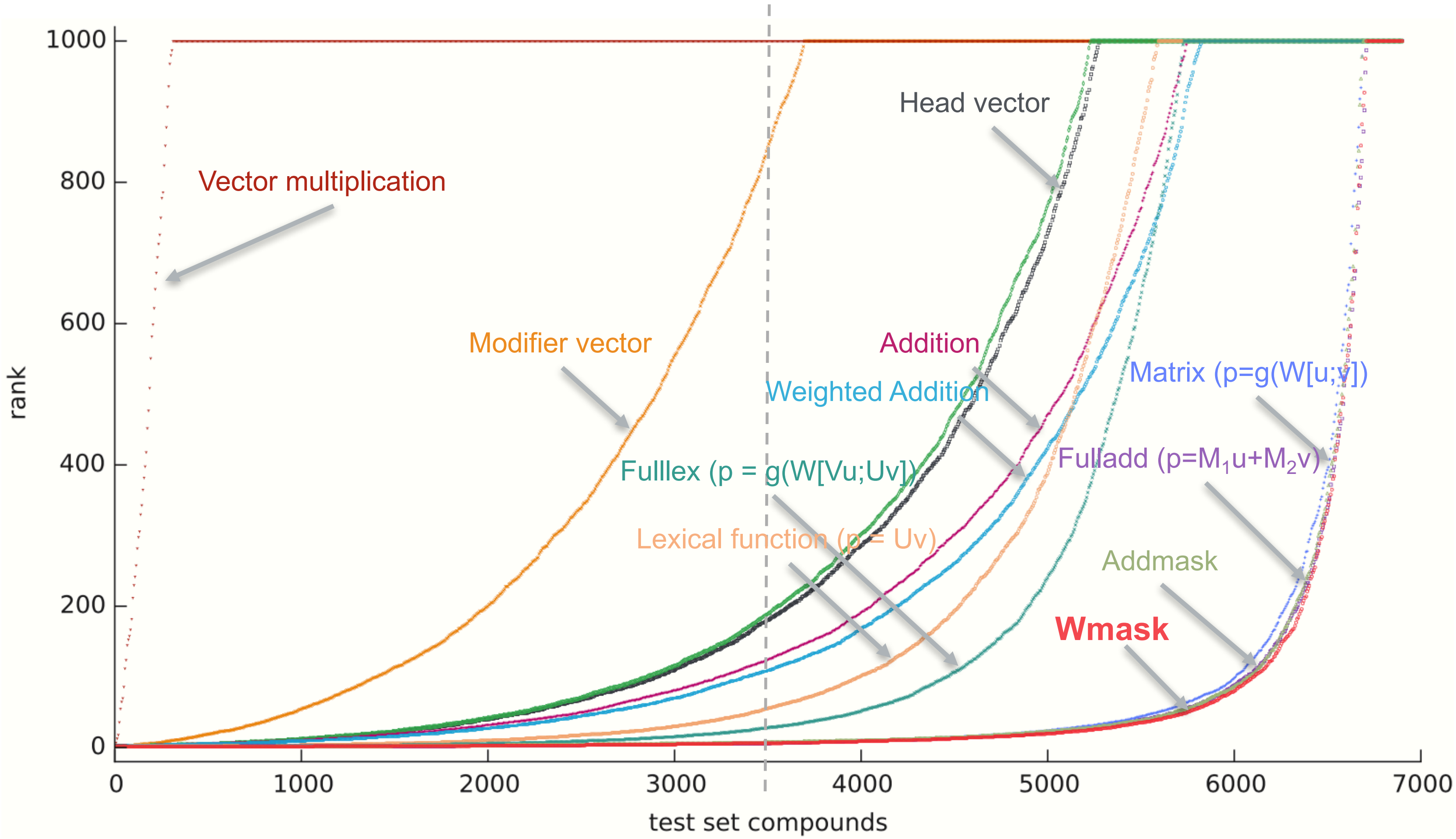
Evaluate Composition Models (2)

- compute the **ranks** of the composed representations in the test set

- **rank computation**
  1. compute cosine distance between the composed representation (**compound**) and *all* the **corpus-induced vectors**
  2. sort, most similar first
  3. the rank is the position of the corresponding corpus-induced vector (**compound**) in the sorted list

- **lower** rank is **better** ~ **composed representation** is closer neighbour to the **corpus-induced represention**

Evaluation Results



Composition with the Mask Models

- **masks**:1-dimensional vectors of the same size as the word vectors
- provide **position-dependent refinement** of the initial word vector

car factory ⇔ factory car

car => *car_as_modifier*, *car_as_head*
factory => *factory_as_modifier*, *factory_as_head*

Composition with the Mask Models

- **masks**:1-dimensional vectors of the same size as the word vectors
- provide **position-dependent refinement** of the initial word vector

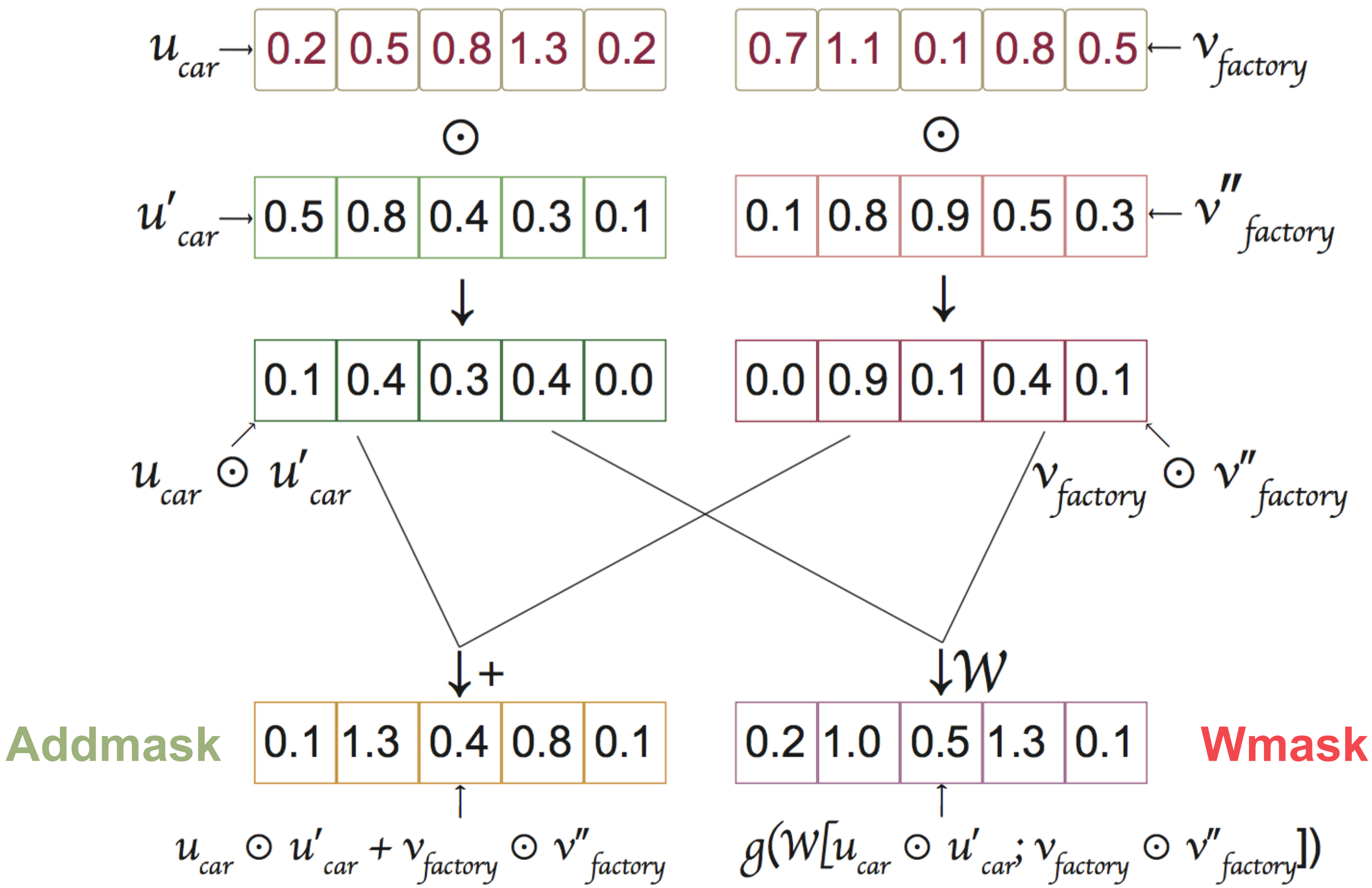
car factory ⇔ factory car

car => *car_as_modifier, car_as_head*

factory => *factory_as_modifier, factory_as_head*

- at composition time, the word vector is first multiplied with the corresponding mask vector
- train 2 vectors (one for the modifier position, one for head position) for each word

Composition with the Mask Models (2)



Wrap-up: Composition Models

- the best models create good composed representations (rank<=5) for 50% of the test data
- more details in:

  Dima, C. 2015. *Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds*. In Proceedings of EMNLP, pp. 17–21.

Wrap-up: Composition Models

- the best models create good composed representations (rank<=5) for 50% of the test data
- more details in:

  Dima, C. 2015. *Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds*. In Proceedings of EMNLP, pp. 17–21.

- how can they be improved?
  - try other models
  - get more training data

Wrap-up: Composition Models

- the best models create good composed representations (rank<=5) for 50% of the test data
- more details in:
  Dima, C. 2015. *Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds*. In Proceedings of EMNLP, pp. 17–21.

- how can they be improved?
  - try other models
  - get more training data

- take a closer look at their results for particular compound types – e.g. compare performance on **transparency-rated compounds**

Transparency-rated compound set

- dataset from Im Walde et al. (2013)
- 244 two-part noun-noun compounds (concrete, depictable)

head / **modifier**	**transparent**	**opaque**
transparent	**Ahornblatt** ‘maple leaf’	**Feuerzeug** ‘lighter’ lit. fire+stuff
opaque	**Fliegenpilz** ‘toadstool’ lit. fly+mushroom	**Löwenzahn** ‘dandelion’ lit. lion+tooth

Transparency-rated compound set: Mturk annotation

modifier \ head	transparent (7 ← 1)	opaque
transparent (7)	**Ahornblatt** ‘maple leaf’ whole: 6.03 modifier: 5.64 head: 5.71	**Feuerzeug** ‘lighter’ lit. fire+stuff whole: 4.58 modifier: 5.87 head: 1.90
opaque (1)	**Fliegenpilz** ‘toadstool’ lit. fly+mushroom whole: 2.00 modifier: 1.93 head: 6.55	**Löwenzahn** ‘dandelion’ lit. lion+tooth whole: 1.66 modifier: 2.10 head: 2.23

Transparency-rated compound set - average ranks

- used 219 compounds (intersection of transparency & compositionality datasets)

head / modifier	transparent 7 ← 3.5	opaque 3.5 ← 1
transparent (7 – 3.5)	144 compounds Average rank **50.6**	20 compounds Average rank **68.4**
opaque (3.5 – 1)	50 compounds Average rank **81.7**	5 compounds Average rank **635.8**

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Transparency-rated compound set - average ranks

- used 219 compounds (intersection of transparency & compositionality datasets)

head / modifier	**transparent** 7 ← 3.5	→ **opaque** 1
transparent 7 ↑ 3.5	Ahornblatt, rank **1** Schneemann, rank **15** lit. 'snow' + 'man' Regenbogen, rank **879** lit. 'rain' + 'arch','bow', 'arc',… (5) Average rank **50.6**	Feuerzeug, rank **10** Zahnseide, rank **117** lit. 'tooth' + 'silk' Average rank **68.4**
opaque 1	Fliegenpilz, rank **40** Flohmarkt, rank **424** lit. 'flea' + 'market' Average rank **81.7**	Löwenzahn, rank **1000** Nilpferd, rank **43** 'hippo', lit. 'Nile' + 'horse' Average rank **635.8**

Transparency-rated compound set - average ranks

- used 219 compounds (intersection of transparency & compositionality datasets)

head / modifier	**transparent** (7 ← 3.5)	**opaque** (3.5 ← 1)
transparent (7 ↑ 3.5)	composition works in the majority of cases	composition possible problem: **multisense** and **metaphoric** meaning of the **head**
opaque (3.5 ↑ 1)	composition possible problem: **multisense** and **metaphoric** meaning of the **modifier**	composition **impossible**: compound representation cannot be obtained compositionally

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Conclusion

- **composition models** create good representations for many compounds

Conclusion

- **composition models** create good representations for many compounds

- problem: **multisense** and **metaphoric** meaning of the head or modifier
  - solution ⇨ sense- & metaphor-aware word representations/ composition models

Conclusion

- **composition models** create good representations for many compounds
- problem: **multisense** and **metaphoric** meaning of the head or modifier
  - solution ⇨ sense- & metaphor-aware word representations/ composition models
- problem: **opaque compounds** - compound representation cannot be obtained compositionally
  - solution ⇨ identification of opaque compounds

Thank you!

- Contact

Corina Dima corina.dima@uni-tuebingen.de