Computational Semantics and Pragmatics

Autumn 2012

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam

Summary of DSMs Features and Parameters

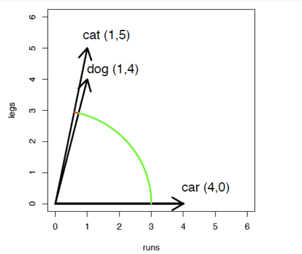
Usage-based perspective on meaning: word meaning depends, at least in part, on the contexts in which words are used.

How do we build contextual (i.e. distributional) meaning representations? Essentially, by quantifying with what kind of expressions words occur.

A distributional semantic model (DSM) is a co-occurrence matrix where rows correspond to vectors for our target terms and columns to contexts where the target terms appear (the vector dimensions).

	run	legs
dog	1	4
cat	1	5
car	4	0

A DSM allows us to measure the semantic similarity between words by comparing their vector representations.



Summary of DSMs Features and Parameters

Main parameters and steps in constructing a DSM:

- Target terms and contexts: raw text, lemmas, POS-tagged lemmas, only content words, dependency relations, patterns,...
- Corpus to be used and required pre-processing
- Context where to look for co-occurrence events: window of k words from the target, within sentence boundaries,...
- Build the matrix by extracting counts of co-occurrence events
- Possible logarithmic scaling of features
- Possible weighting of features to give more weight to less expected events
- Possible dimensionality reduction to compress the matrix
- Similarity measure [if model is evaluated on how well it captures semantic similarity]: cosine of angle between two vectors, ...

Evaluating DSMs

DSMs have mostly been evaluated on how well the model captures semantic similarity between words. Three classic examples:

Evaluating DSMs

DSMs have mostly been evaluated on how well the model captures semantic similarity between words. Three classic examples:

- Synonym identification: the TOEFEL task

Target: *levied*	Candidates: *imposed*, *believed*, *requested*, *correlated*

  * Humans: non-natives 86.75%; natives: 97.75%.
  * DSMs: Rapp's (2003) model trained on lemmatized BNC: 92.5%

Evaluating DSMs

DSMs have mostly been evaluated on how well the model captures semantic similarity between words. Three classic examples:

- Synonym identification: the TOEFEL task

  Target: *levied*   Candidates: *imposed*, *believed*, *requested*, *correlated*

  - Humans: non-natives 86.75%; natives: 97.75%.
  - DSMs: Rapp's (2003) model trained on lemmatized BNC: 92.5%

- Semantic similarity judgements
  - Dataset of noun pairs rated by humans on a 0-4 similarity scale
  - Padó and Lapata (2007): strong correlations ($r = 0.8$) between the distances in their dependency-based DSM and human judgements.

Evaluating DSMs

DSMs have mostly been evaluated on how well the model captures semantic similarity between words. Three classic examples:

- Synonym identification: the TOEFEL task

Target: *levied*	Candidates: *imposed*, *believed*, *requested*, *correlated*

  - Humans: non-natives 86.75%; natives: 97.75%.
  - DSMs: Rapp's (2003) model trained on lemmatized BNC: 92.5%

- Semantic similarity judgements
  - Dataset of noun pairs rated by humans on a 0-4 similarity scale
  - Padó and Lapata (2007): strong correlations ($r = 0.8$) between the distances in their dependency-based DSM and human judgements.

- Semantic priming: the word *pear* is recognized/accessed faster if it is heard/read after *apple*.
  - McDonald & Brew (2004), Padó & Lapata (2007): similarity between terms with a priming effect significantly higher ($p < .01$) than between those w/o effect.

Main Challenges for DSMs

Main Challenges for DSMs

- Beyond semantic similarity: DSMs are good at modeling similarity, but words may be similar in very different respects.
  - ∗ what about antonymy, or asymmetric relations such as hyponymy?
  - ∗ what about the systematic sense alternations of regular polysemy?

Main Challenges for DSMs

- Beyond semantic similarity: DSMs are good at modeling similarity, but words may be similar in very different respects.
  - what about antonymy, or asymmetric relations such as hyponymy?
  - what about the systematic sense alternations of regular polysemy?
- Compositionality: can DSMs explain how the meaning of a complex expression can be build up from the meaning of its parts?
  - what kind of vector-based operations could be appropriate?
  - can we combine vectors with the usual boolean operators?

Main Challenges for DSMs

- Beyond semantic similarity: DSMs are good at modeling similarity, but words may be similar in very different respects.
  - ∗ what about antonymy, or asymmetric relations such as hyponymy?
  - ∗ what about the systematic sense alternations of regular polysemy?
- Compositionality: can DSMs explain how the meaning of a complex expression can be build up from the meaning of its parts?
  - ∗ what kind of vector-based operations could be appropriate?
  - ∗ can we combine vectors with the usual boolean operators?
- Embodiment: meanings/concepts are not purely linguistic, they are *embodied*, grounded in perception.
  - ∗ is this an inherent limitation of DSMs? [*relevant for all challenges*]
  - ∗ can we combine distributional information from different modalities?

Main Challenges for DSMs

- Beyond semantic similarity: DSMs are good at modeling similarity, but words may be similar in very different respects.
  - ∗ what about antonymy, or asymmetric relations such as hyponymy?
  - ∗ what about the systematic sense alternations of regular polysemy?
- Compositionality: can DSMs explain how the meaning of a complex expression can be build up from the meaning of its parts?
  - ∗ what kind of vector-based operations could be appropriate?
  - ∗ can we combine vectors with the usual boolean operators?
- Embodiment: meanings/concepts are not purely linguistic, they are *embodied*, grounded in perception.
  - ∗ is this an inherent limitation of DSMs? [*relevant for all challenges*]
  - ∗ can we combine distributional information from different modalities?

New DSMs addressing these challenges will propose new techniques and new evaluation frameworks.

Presentations of Recent Papers

Most of the papers we will study during these two weeks try to move forward the field by addressing these challenges.

Week 6

- 5 Dec: Introduction to DSMs continued, plus two paper presentations
  - **[Isabel]** Erk & Pado (2010) Exemplar-Based Models for Word Meaning In Context, ACL.
  - **[Phong]** Grefenstette & Sadrzadeh (2011) Experimental support for a categorical compositional Distributional model of meaning, EMNLP.
- 7 Dec: More on Compositionality with DSMs
  - **[Sanne]** Guevara (2010) A Regression Model of Adjective-Noun Compositionality in Distributional Semantics, GEMS Workshop.
  - **[Swantje/Nadine]** Baroni & Zamparelli (2010) Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space, EMNLP.
  - **[Philip]** Socher, Huval, Manning, & Ng (2012) Semantic Compositionality Through Recursive Matrix-Vector Spaces, EMNLP.

Week 7

- 12 Dec: DSMs and word senses
  - **[Cecilia]** Erk and McCarthy (2009) Graded word sense assignment, EMNLP.
  - **[Max]** Boleda, Pado & Utt (2012) Regular polysemy: a distributional model, *SEM.
  - **[Ciyang]** Huang, Socher, Manning & Ng (2012) Improving Word Representations via Global Context and Multiple Word Prototypes, ACL.
- 14 Dec: Multi-modal DSMs
  - **[Jonathan]** Bruni, Tran & Baroni (2011) Distributional semantics from text and images, GEMS Workshop.
  - **[Jetze]** Bruni, Boleda, Baroni, & Tran (2012) Distributional semantics in technicolor, *SEM.
  - **[Gijs]** Silberer & Lapata (2012) Grounded Models of Semantic Representation, EMNLP-CoNNL.

Very good overview of current research in the area.