5/12/12

Erk & Padó (2010)

Exemplar-Based Models for Word Meaning in Context

Discussion of papers on Distributional Semantics
- Isabel Keijer

Distributional models

- Distributional Hypothesis:
  words that occur in the same context tend to have similar meanings.
- Representing meaning as context:
  the meaning of a word is represented by a vector that records co-occurrence with context features.
- (At least) two possibilities:
  - Based on prototype theory
  - Based on exemplar theory

Prototype-based approach

- Prototype theory (quick recap):
  - prototype = summary representation
  - features that are usually found in the category members, weighted
  - “contradictory” features may be included with different weights
  - categorization criterion based on feature weights
  - no feature is required to be present
- Distributional model implementation:
  - one vector per category/term (what we’ve seen so far)

Exemplar models

- Exemplar theory (quick recap):
  - there is not just one representation that encompasses an entire concept
  - a concept is just the set of instances of that concept that one person remembers
  - to categorise new items, we weight them by how similar they are to the items in our memory
- Distributional model implementation:
  - memorize each seen instance of a category (/term);
  - perform categorization by comparing a new stimulus to each remembered exemplar vector

Why an exemplar-based approach?

- Polysemy:
  - (according to Erk & Padó) problematic for the prototype-based approach
    - because only one vector per category, all the different senses (and contexts) of the target are lumped together
  - solution to this problem: exemplar-based approach
    - only activate the relevant exemplars
    - there are different exemplar sets for each sense

Exemplar Activation

Model polysemy by activating relevant exemplars of a lemma *E* in a given sentence context *s*.

- *Activation* of a set *E* by exemplar *s*:
  $act(E,s) = \{e \in E \mid \text{sim}(e, s) > \theta(E, s)\}$
- Setting the threshold:
  - *k*NN activation: θ is set to the similarity of the *k*-th most similar exemplar
  - *q*-percentage activation: θ is set to the (100-*q*)-th percentile of the $sim(e, s)$ distribution

Task: paraphrasing

Predicting paraphrase felicity

- Given:
  - target lemma *T*;
  - in a particular sentential context *s*;
  - a list of potential paraphrases of *T*,
- Predict which of the paraphrases are appicable in *s*.

How? *actT* vs. *actP*

- We are given a target lemma *T*, plus its context *s*, and a list of potential paraphrases *P*.
- We have already computed a VSM based on the BNC, with instance-vectors grouped into sets of exemplars (senses).
- Activation approaches:
  - *actT*: activate the target, rank the paraphrases
  - *actP*: activate the paraphrase, using *s* as target word
  - No act. : ranks paraphrase candidates by the distance between their type vectors and the target's type vector

Evaluation

- Generalized Average Precision (GAP)

$$GAP(P,G) = \frac{1}{\sum_{j=1}^{m} I(y_j)\overline{y_j}} \sum_{i=1}^{n} I(x_i)\overline{x_i}$$

- The higher the number, the more correct predictions
- Used to evaluate other models, so can be used to accurately compare the results of different models

actT vs. *actP* - results

para-meter	*actT* kNN	*actT* perc.	*actP* kNN	*actP* perc.
10	36.1	35.5	36.5	**38.6**
20	**36.2**	35.2	36.2	37.9
30	36.1	35.3	35.8	37.8
40	36.0	35.3	35.8	37.7
50	35.9	35.1	35.9	37.5
60	36.0	35.0	36.1	37.5
70	35.9	34.8	36.1	37.5
80	36.0	34.7	36.0	37.4
90	35.9	34.5	35.9	37.3
no act.	34.6		35.7	
random BL	28.5			

Table 2: Activation of T or P individually on the full LexSub dataset (GAP evaluation)

Joint activation

P activation (%) $\Rightarrow$	10	20	30
T activation (kNN) $\Downarrow$			
5	**38.2**	38.1	38.1
10	37.6	37.8	37.7
20	37.3	37.4	37.3
40	37.2	37.2	36.1

Table 3: Joint activation of P and T on the full LexSub dataset (GAP evaluation)

Compared to other models

	Models		
	EP08	EP09	TDP09
EP08 dataset	27.4	NA	NA
EP09 dataset	NA	32.2	36.5
	actT	*actP*	*actTP*
EP08 dataset	36.5	38.0	**39.9**
EP09 dataset	39.1	**39.9**	39.6

Table 4: Comparison to other models on two subsets of LexSub (GAP evaluation)

EP08: Erk & Padó (2008)
EP09: Erk & Padó (2009)
TDP09: Thater et al. (2009)

Important points

- Distributional model of meaning
- Incorporating cognitive theory
- Prototype vs. Exemplar
- Polysemy
- Paraphrasing / Lexical substitution task