Distinguishing Paradigmatic Semantic Relations: Human Ratings and Distributional Similarity

Sabine Schulte im Walde
schulte@ims.uni-stuttgart.de

Motivation

Paradigmatic semantic relations

- Central in organisation of mental lexicon (Miller & Fellbaum, 1991; Murphy, 2003): synonymy, antonymy, hypernymy, (co-)hyponymy
- Provide a structure for the lexical concepts that words express.
- Natural relation structure differs across word classes:
  - hypernymy → noun lexicon; minor for verbs;unnatural for adjectives
  - antonymy → adjective lexicon
  - hypernymy, antonymy, synonymy, entailment → verb lexicon

Distributional vector space models

- Rely on the distributional hypothesis (Harris, 1954; Firth, 1957).
- Model meaning and “similarity” of target words (Turney & Pantel, 2010).
- Paradigmatic relations are difficult to distinguish:
  The boy/girl/person loves/hates the cat.

Distributional Models

Cosine similarity

- Paradigmatic semantic relation pairs are expected to be close in word space.
- Vector space: 20-word co-occurrence in web corpus DECCOW14AX, weighted by frequency vs. local mutual information (Evert, 2006; Schäfer & Bildhauer, 2012)
- Distributional similarity: cosine of vector angle

Automatic classification

- Series of classification experiments
- Features: window co-occurrence vs. lexico-syntactic patterns
- Vector representations of relation word pairs:
  - Window-COS: cosine scores between word pairs
  - Window-DIFF: difference vector for word pair
  - Window-PROD: vector product for word pair
- Pattern: linear word sequences between related words
- Corpus: SdeWaC (Faaß & Eckart, 2013)
- Nearest-centroid classifier (also known as Rocchio classifier)
  - Use training pairs to initiate relation class centroids.
  - Assign test pairs to nearest class centroid.
  - Evaluation: 5-fold cross-validation, precision values.

Human Ratings

Target–response paradigmatic relation pairs

- Targets: Random choice of 99 WordNet targets per word class: nouns, verbs, adjectives (Scheibe & Schulte im Walde, 2014), balanced for
  - frequency class (low; mid; high)
  - polysemy class (monosemous; two senses; >2 senses)
  - size of semantic class

- Experiments:
  - generation (5,745/8,910 pair types/tokens)
  - rating (1,684 pair types; scale: 0–5)

Experiment 1: Generation of relation pairs (examples)

<table>
<thead>
<tr>
<th>ANTI SYN HYP</th>
<th>ANTI SYN HYP</th>
<th>ANTI SYN HYP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VERB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADJ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiment 2: Rating of relation pairs (examples)

<table>
<thead>
<tr>
<th>Target Pair</th>
<th>Generation</th>
<th>ANTI SYN HYP</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VERB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADJ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distinction between relation pairs

How well do experiment participants distinguish between paradigmatic relations? → differences in mean ratings across relation pairs

References