A comparison of selectional preference models for automatic verb classification

Will Roberts and Markus Egg

Institut für Anglistik und Amerikanistik
Humboldt Universität zu Berlin

Sunday, 26 October, 2014
Outline

1. Introduction
2. Models
3. Results
Selectional preferences

- Predicates can select for their arguments:

  ? My aunt is a bachelor.  \hspace{1cm} (McCawley, 1968)

- We model verbs empirically:

  I eat  \hspace{1cm} meat
  bread
  fruit
  :
  newspaper

- Evaluate on an automatic verb classification task
- Baseline model clusters verbs based on \textit{subcategorisation}
We will want to record that this instance of *use* has:

- **Subject**: wir, *we* (pronoun, ignored)
- **Direct object**: Umfragedatum, *survey datum*
- **PP (für, *for*)**: Zweck, *purpose*

We also include indirect objects (datives)

A selectional preference model will map noun forms onto concept labels
Hypothesis

verb clustering score

only subcat: one concept containing all nouns

optimal concept granularity

lexical preferences: one concept per noun

effective SP model

ineffective SP model
The combination of syntactic argument types is assigned a subcategorisation frame (SCF) code:

\[ \text{benutzen} \rightarrow \text{nap:für.Acc} \]

A verb’s distribution over SCF codes is its subcategorisation preference.
Test set has 3 million verb instances

Gold standard: 168 verbs in 43 classes
Verb clustering

Verb dissimilarity is computed with the Jensen-Shannon divergence.

- **Verb:**
  - $p = 1$
  - $p = 0$

- **Corpus counts**
  - $scf_1$, $scf_2$, $scf_3$, $scf_4$, ..., $scf_{671}$, $scf_{672}$, $scf_{673}$
  - Corpus counts = discrete probability distribution = subcat prefs
Lexical preferences (LP)

Example

Wir *benutzen* Ihre Umfragendaten nicht für eigene Zwecke.
We *use* your survey data *not* for own purposes.
We will not use your survey responses for private purposes.

benutzen $\Rightarrow$ nap:für.Acc*dobj-Umfragedatum*prep-Zweck

- To control data sparsity, we employ a parameter $N$: number of nouns included in the lexical preferences model
  - Nouns with rank $\geq N$ are ignored (as if unseen)
Partition $N$ nouns into $M$ classes (equivalence relation)
Word space model (WSM)

- Built on lemmatised SdeWaC
- Features are the 50,000 most common words (minus stop words)
- Sentences as windows
- Feature weighting: t-test scheme
- Context selection zeroes out infrequent features in the model
- Use cosine similarity and spectral clustering to partition $N$ nouns into $M$ classes
GermaNet

- Granularity is controlled using depth, $d$
- Nouns can belong to more than one concept: soft clustering
Latent Dirichlet Allocation (LDA)

- Built with the same data used by the Sun/Korhonen model
- Each \(\langle\text{verb, grammatical relation}\rangle\) pair has a distribution \(\Phi\) over concepts
- Each concept \(z\) has a distribution \(\Theta\) over the \(N\) nouns
- Number of concepts \(M\) is 50 or 100
## Results

<table>
<thead>
<tr>
<th>SP model</th>
<th>Parameters</th>
<th>Granularity</th>
<th>$F$-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUN</td>
<td>10K nouns</td>
<td>1,000 noun classes</td>
<td>39.76</td>
</tr>
<tr>
<td>LDA (hard)</td>
<td>10K nouns</td>
<td>50 topics</td>
<td>39.09</td>
</tr>
<tr>
<td>LP</td>
<td>5K nouns</td>
<td></td>
<td>38.02</td>
</tr>
<tr>
<td>WSM</td>
<td>10K nouns</td>
<td>500 noun classes</td>
<td>36.92</td>
</tr>
<tr>
<td>LDA (soft)</td>
<td>10K nouns</td>
<td>50 topics</td>
<td>35.91</td>
</tr>
<tr>
<td>GermaNet</td>
<td>depth = 5</td>
<td>8,196 synsets</td>
<td>34.41</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td>33.47</td>
</tr>
</tbody>
</table>
Sparsity effects in LP

![Graph showing sparsity effects in LP](image-url)
Qualitative differences in noun partitions

<table>
<thead>
<tr>
<th>SUN</th>
<th>WSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$-score 39.76</td>
<td>$F$-score 36.92</td>
</tr>
<tr>
<td>syntagmatic information</td>
<td>paradigmatic information</td>
</tr>
<tr>
<td>synonym/co-hyponym structure</td>
<td>thematic structure</td>
</tr>
<tr>
<td>class size variance 37</td>
<td>class size variance 2800</td>
</tr>
<tr>
<td>semantically consistent</td>
<td>large classes inconsistent</td>
</tr>
</tbody>
</table>
Conclusions

1. Selectional preferences help automatic verb classification
2. Optimal concept granularity is relatively fine
   - Lexical preferences works very well if it is properly tuned
   - Classification of proper names is useful: given names, corporations, medications, etc.
3. Syntagmatic information works better than paradigmatic
Selectional preference models have been compared before
  - Almost always under a plausibility or pseudoword paradigm!
We are interested in semantic verb clustering
We evaluate several selectional preference models, comparing them using a manually constructed semantic verb classification

We show that modelling selectional preferences is beneficial for verb clustering, no matter which selectional preference model we choose

Other findings:
  - Capturing syntagmatic relations seems to work better than paradigmatic
  - A simple lexical preferences model performs very well; data sparsity does not seem to be more of a problem for this model than for others