Knowledge-free Learning of lexical hierarchical paradigmatic relations

Stefan Bordag
University of Leipzig
sbordag@informatik.uni-leipzig.de
April 29, 2007

Abstract
This paper represents an attempt at translating statistical methods and properties of rankings in a vector space model into an algorithm that extracts hierarchical paradigmatic relations such as hyperonymy and meronymy. The algorithm is based on simulating semantic similarity through co-occurrence statistics as well as observing symmetry in similarity rankings. A preliminary evaluation based on a manually created resource demonstrates a clear preference for hierarchically related words to be extracted.

1. Introduction
Automatic (in the sense of unsupervised and knowledge-free) extraction of hierarchical paradigmatic relations would be a useful alternative to using a thesaurus such as WordNet (Fellbaum, 1998) in situations where there is no such resource available and the quality of the results need not be perfect. At the very least it might serve as a fall-back method or as a rapid prototyping tool when creating a local ontology.

Observing statistically significant co-occurrences within a certain window around a target word token is a well-known method to obtain a representation of the general usage of the corresponding word type. The window may be syntactically filtered or represent a plain vector space model. In the present experiments, the Poisson approximation was used as a significance measure (Quasthoff & Wolff, 2002). Using such contextual representations it is also straightforward to compute a simulation of semantic similarity by comparing the co-occurrence vectors of words with each other. This combination of poisson co-occurrences and baseline similarity was found to produce the best results in earlier work (Bordag, 2007). For any input word this also allows to determine a ranking of words with descending semantic similarity.

The hypothesis underlying the present work is that contextual similarity of symmetrical paradigmatic relations results in symmetrical similarity rankings. For two cohyponyms $A$ and $B$ this means that one of the most contextually similar words to $A$ is $B$ and vice versa. Contrarily, if $B$ is a part of $A$ (or a hyponym), then $B$ is likely to be contextually similar to $A$, but not $A$ to $B$. This is because it is possible to replace a more specific word with a less specific one in most contexts, but not a general one with a specific one. For example, if the sentence I eat that apple. is given, it is possible to rephrase it to say I eat that fruit., but it is not possible to rephrase the first sentence from the second one without further situational knowledge.

2. Algorithm
In the experiments conducted, three methods were compared with each other $coocc$, $sim$ and $unsym$. The first, $coocc$ represents a direct ranking of the 100 most significant co-occurrences for each input word. The second method, $sim$, represents a ranking of 100 contextually most similar words (using the baseline measure that ranks according to matching non-zero entries in the vectors of the two words to be compared).

The third method is the actual algorithm described in this work. For a given input word it first obtains the ranking of most similar words and then reranks them according to the reversed similarity. The unsymmetric relevance $unsym_A(B)$ is the harmonic mean of the inverse relevance of $A$ for $B$ ($1 - rel_B(A)$) and the relevance of $B$ for $A$ ($rel_A(B)$): $unsym_A(B) = \frac{2 \times rel_A(B) \times (1 - rel_B(A))}{rel_A(B) + (1 - rel_B(A))}$.

According to the underlying hypotheses, and also as Table 1 showing a few examples supports, the first method should produce mostly syntagmatically related words, such as typical articles, adjectives, ad-
verbs, etc. The simulated semantic similarity, on the other hand, should produce mostly paradigmatically related words, such as cohyponyms or synonyms. Compared to that, the newly presented unsymSim produces indeed hierarchically related words to the input words, even though it is not possible to distinguish various subtypes of hierarchical relations such as hyperonymy from meronymy.

3. Results

Experiments were conducted on a 100 million words German corpus and the results were measured against the Annotation Project (Biemann, 2005). For each input word and for each method, the 100 most significant (or similar or asymmetrically similar) words were measured. The mean average precision (MAP) (the average inverse rank of each correct hit) that specializes on measuring ranking preferences and is often used in Information Retrieval was used to gauge whether any of the method produced rankings according to the hypotheses. For the unsym method additionally rankings split into frequent (the top 50K frequent words) and less frequent (50K to 100K words) are given, since it is possible to assume that higher frequent words are more likely to be higher in the corresponding hierarchy.

The measurements were performed by counting any related word according to the Annotation Project as relevant within the top 100 rankings of the computed words and averaging the obtained MAP score over all 100000 input words.

Table 1. Top 4 ranked results for the four examples using symCoocc and unsymSim.

<table>
<thead>
<tr>
<th>meth.</th>
<th>coocc</th>
<th>sim</th>
<th>unsym</th>
</tr>
</thead>
<tbody>
<tr>
<td>synt</td>
<td>8.01</td>
<td>3.83</td>
<td>2.34</td>
</tr>
<tr>
<td>para</td>
<td>8.18</td>
<td>12.07</td>
<td>4.24</td>
</tr>
<tr>
<td>h. para</td>
<td>3.56</td>
<td>7.81</td>
<td>4.35</td>
</tr>
<tr>
<td>total</td>
<td>16.07</td>
<td>18.48</td>
<td>8.35</td>
</tr>
</tbody>
</table>

Table 2. Measuring MAP in % for the top 100 ranked results for each input word against the Annotation Project.

The results show that there is a significant deviation of each method (in relation to the other two methods) to produce rankings according to the hypotheses formulated above. The splitting into the frequency ranges additionally shows that there in fact is a dependency between the ability of the unsym method to extract the hyporonym (or meronym) of a word and the word’s frequency.

4. Conclusions

Despite the difficulty to evaluate methods that extract semantic relations between words, it is clear that the present method unsym does rank hierarchically related words highest. The resulting figures are very low, which indicates also a low upper bound, but since they are based on several tens of thousands of cases, they are statistically very stable. It is likely that syntactic filtering or other relevant forms of additional linguistic knowledge (if available) improve the results of this algorithm. Nevertheless, as such the method is simple to implement and does not require any corpus preprocessing apart from sentence and word tokenization and still produces good results.

References


