Semantic Relations for the Recognition of Anaphoric Expressions with different head

Fabio Celli
CLIC-CIMeC, University of Trento
fabio.celli@unitn.it
June 5, 2010

Abstract

This technical report addresses the problem of how semantic relations can be used to improve the performance of anaphora resolution. In particular I took the mentions from coreference chains classified as false negatives by BART, a system for anaphora resolution, and I extracted the relations that hold between the two different mentions by means of pattern searching in google. The results show that it is possible to improve the performance of anaphora resolution by classifying false negative co-referring mentions with the same head using taxonomic and associative relations as features.

1 Introduction and Background

In anaphora resolution, that is the task of recognizing co-referring expressions automatically, there is the problem of co-referring mentions with different head. For example in (1) “the substance” co-refers to “crocidolite”, but they are different and this fact often generates errors.

(1) Crocidolite were used in cigarette filters and among 33 men who worked closely with the substance, 28 have died

In a previous work [1] I found that taxonomic relations such as Is-A and Coordinate are strongly associated to the presence of coreference while associative relations such as Part-Of and Member-Of are strongly associated to the absence of coreference. Thus the hypothesis is that: given a pair of expression, if I compare the frequency of that pair in a taxonomic and in an associative relation, I can retrieve a measure of the probability that the words in the pair are coreferring expressions.

In order to test if the words in the pair are related I choose to use patterns in google, since there is a long tradition in Is-A and Part-Of relation extraction via patterns, starting with Hearst 1992 [3], to Girju et al. 2006 [2], and there is also some evidence that conceptual information can be extracted from web using patterns (see Poesio and Almuhareb 2008, [6]).

In the next section I describe the experiment, and in section 3 I will discuss the results and draw some provisional conclusions.

1
2 Experiment

I took 20 random false negative examples pairs of coreferring expressions with different head from the output of BART, a system for anaphora resolution developed by Versley et al [9], and then I added other 20 pairs of random (non-coreferring) words extracted from Arrau. I tagged the false negative examples as coreferring expressions and the random pairs as non-coreferring expressions, then I run a classification task using two features: 1) frequency of the pair with the pattern “kind of”, and 2) frequency of the pair with the pattern “part of”, obtained from google APIs and normalized. The target class is coreferring/non-coreferring, so this is a binary classification task. I run the experiment in Weka (see [8]), using different learning algorithms: trees (see [7]), naive bayes (see [4]) and SVMs (Platt’s [5] sequential minimum optimization with a polynomial kernel). Evaluation is with a 10-fold cross validation. Results are reported in table 1 below: I obtained the same results adding two more patterns: “same

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>trees</td>
<td>0.753</td>
<td>0.750</td>
<td>0.749</td>
</tr>
<tr>
<td>bayes</td>
<td>0.751</td>
<td>0.675</td>
<td>0.648</td>
</tr>
<tr>
<td>SVMs</td>
<td>0.794</td>
<td>0.650</td>
<td>0.601</td>
</tr>
</tbody>
</table>

Table 1: Results of experiment 1.

of” and “different from”, thus using 4 features in total. A discussion follows in the next section.

3 Discussion and conclusion

The results show that it is possible to exploit semantic relations to discriminate coreferring from non-coreferring mentions. Trees proved to be the best algorithm for this kind of task (14 coreferring mentions correctly classified and 16 non-coreferring mentions correctly classified). SVMs proved to have more discriminative power in the classification of the non-coreferring examples but a terribly low recall on the coreferring examples (all 20 non-coreferring mentions correctly classified but only 6 coreferring mentions classified correctly). naive Bayes is along those lines. Changing the polynomial kernel’s exponent from 2 to 3 or 4 lowers the performance below an F1 of 0.5.

In conclusion it is possible to improve the performance of anaphora resolution by classifying false negative co-referring mentions with the same head by means of taxonomic and associative relations. The method adopted here, for instance using google queries, proved to be a suitable way to do the classification task. It has advantages and disadvantages: for example it requires a API key for a large number of queries, but it can be applied virtually to any language supported by google.

References


