The Syntax of Semantic Relations in Italian

Fabio Celli
CLIC, University of Trento
October 24, 2009

Abstract

The experiment reported in this paper is a feature engineering operation that explores the syntactic structures of three different semantic relation types with tree kernels. This is done in order to find whether syntax can be a good feature to improve semantic relation separability in classification tasks. The average accuracy obtained is 63.09%.

1 Introduction And related work

In a previous experiment [1] it was found that three semantic relation classes (“role”, “location”, “social”) yielded better results in a classification task with respect to the seven semantic relation classes in ACE 2004 (see [3]), since those three classes are easily separable. The three semantic relations are: Role (entity 1 is involved/included in entity 2), Location (entity 1 has a position with respect to entity 2), Social (entity 1 has some interaction with entity 2).

In another experiment ([2]) those relation classes were annotated on I-CAB, an Italian corpus, in order to test their performance on Italian. In this paper I want to test whether or not syntax could be a good feature for a further improvement in the separability of those three semantic relation classes in Italian. In this paper I will use tree kernels: a computational technique able to extract all the subtrees (tree fragments) from target tree structures and compare tree structures using subtrees, this technique already showed good performances in semantic role labeling tasks ([5]). The paper is structured as follows: in the next section there is the experiment, then a discussion follows in the third section.

2 Experiment

The data used for the experiment are taken from I-CAB, an Italian corpus previously annotated with named entities (person, organizations, locations and geo-political entities) and the three semantic relations, defined as information underlying between two named entities, described above (role, location, social). The Shortest Path-enclosed Trees (SPTs) between the named entities involved in a semantic relation were given to the system as inputs, since Zhang et Al ([7]) found that those context-free trees work better than the context-dependent ones. Thirty SPTs (18 training, 12 testing) were randomly sampled from I-CAB and manually parsed, other features, inputted as vectors, were named entity 1 type (it takes the four values described above) and named entity 2 type (that takes
the same four values as entity 1). The software used for the analysis is SVM-light-Tree Kernel (see [4] and [6]), the task was a binary classification, so three different training and test sets were created, one training and one test set for each semantic relation class (data are the same but each one has positive and negative examples changed). Results are reported in table 1 below:

<table>
<thead>
<tr>
<th>relation</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>role</td>
<td>57.89%</td>
</tr>
<tr>
<td>location</td>
<td>64.71%</td>
</tr>
<tr>
<td>social</td>
<td>66.67%</td>
</tr>
<tr>
<td>avg</td>
<td>63.09%</td>
</tr>
</tbody>
</table>

Table 1: results of experiment one

3 Discussion and Conclusion

Results reported in table 1 are not directly comparable to the state-of-the-art found in literature, because for this experiment has been used an Italian non-standard corpus. However results clearly show that Italian syntax is not a good feature for the extraction of the three semantic relations proposed here. In the future the experiment should be repeated using Context-Sensitive Trees (CSTs), which include the syntactic context before and after the SPTs, to see whether longer trees yield better performances.

References