A novel approach to mapping FrameNet lexical units to WordNet synsets

Sara Tonelli, Emanuele Pianta

Abstract

In this paper we present a novel approach to mapping FrameNet lexical units to WordNet synsets in order to automatically enrich the lexical unit set of a given frame. While the mapping approaches proposed in the past mainly rely on the semantic similarity between lexical units in a frame and lemmas in a synset, we exploit the definition of the lexical entries in FrameNet and the WordNet glosses to find the best candidate synset(s) for the mapping. Evaluation results are also reported and discussed.

1 FrameNet and the existing mapping approaches

The FrameNet database [1] is a lexical resource of English describing some prototypical situations, the *frames*, and the frame-evoking words or expressions associated with them, the *lexical units* (LU). Every frame corresponds to a scenario involving a set of participants, the *frame elements*, that are typically the syntactic dependents of the lexical units. The FrameNet resource is corpus-based, i.e. every lexical unit should be instantiated by at least one example sentence, even if at the moment the definition and annotation step is still incomplete for several LUs. FrameNet has proved to be useful in a number of NLP tasks, from textual entailment to question answering, but its coverage is still a major problem. In order to expand the resource, it would be a good solution to acquire lexical knowledge encoded in other existing resources and import it into the FrameNet database. WordNet [4], for instance, covers the majority of nouns, verbs, adjectives and adverbs in the English language, organized in synonym sets called *synsets*. Mapping FrameNet LUs to WordNet synsets would automatically increase the number of LUs per frame by importing all synonyms from the mapped synset, and would allow to exploit the semantic and lexical relations in WordNet to enrich the information encoded in FrameNet.

Proceedings of the 8th International Conference on Computational Semantics, pages 342–345,
Tilburg, January 2009. ©2009 International Conference on Computational Semantics

Several experiments have been carried out in this direction. Johansson and Nugues [5] created a feature representation for every WordNet lemma and used it to train an SVM classifier for each frame that tells whether a lemma belongs to the frame or not. Crespo and Buitelaar [3] carried out an automatic mapping of medical-oriented frames to WordNet synsets, trying to select synsets attached to a LU that were statistically significant in a given reference corpus. De Cao et al. [2] proposed a method to detect the set of suitable WordNet senses able to evoke the same frame by exploiting the hypernym hierarchies that capture the largest number of LUs in the frame. For all above mentioned approaches, a real evaluation on randomly selected frames is missing, and accuracy was mainly computed over the new lexical units obtained for a frame, not on a gold standard where one or more synsets are assigned to every lexical unit in a frame. Besides, it seems that the most common approach to carry out the mapping relies on some similarity measures that perform better on richer sets of lexical units.

2 The mapping algorithm

2.1 Motivation

We propose a mapping algorithm that is independent of the number of LUs in a frame and from the example sentences available. In fact, we believe that under real-usage conditions, the automatic expansion of LUs is typically required for frames with a smaller LU set, especially for those with only one element. In the FrameNet database (v. 1.3), 33 frames out of 720 are described only by one lexical unit, and 63 are described by two. Furthermore, almost 3000 lexical units are characterized only by the lexicographic definition and are not provided with example sentences. For this reason, we suggest an alternative approach that makes use of usually unexploited information collected in the FrameNet database, namely the *definition* associated with every lexical unit.

Since both WordNet glosses and FrameNet definitions are manually written by lexicographers, they usually show a high degree of similarity, and sometimes are even identical. For example, the definition of *thicken* in the *Change_of_consistency* frame is *"become thick or thicker"*, which is identical to the WordNet gloss of synset n. v#00300319. The *thicken* lemma occurs in three WordNet synsets, and in each of them it is the only lemma available, so no synonymy information could be exploited for the sense disambiguation.

2.2 The algorithm

We tried to devise a simple method to map a FrameNet Lexical Unit (LU) into one or more WordNet synsets. Given a LU L from a frame F, we first find the set of all synsets containing L (L candidate set, *LCandSet*). If *LCandSet* contains only one synset, this is assigned to L. Otherwise, we look for the synsets in *LCandSet* whose WN gloss has the highest similarity with the FrameNet definition of L. We tried two baseline similarity algorithms based respectively on stem overlap and on a modified version of the Levenshtein algorithm taking stems as comparison unit instead of characters. Stem overlap turned out to perform definitely better than Levehnstein. Then we tried to improve on simple stem overlap baseline by considering also the other LUs in F. To this extent, we calculate the set of all synsets linked to any LU in F (*FCandSet*). This is exploited in two ways. First, we boost the similarity score of the synsets in *LCandSet* with the largest number of links to other LUs in F (according to *FCandSet*). Secondly we assign to F the most common *WordNet Domain* in *FCandSet*, and then boost the similarity score of *LCandSet* synsets belonging to the most frequent *WordNet-Domain* in F. We discard any candidate synset with a similarity score below a MIN threshold; on the other side, we accept more than one candidate synset if they have a similarity score higher than a MAX threshold.

3 Evaluation

We created a gold standard by manually mapping 380 LUs belonging to as many frames to the corresponding WordNet synsets. Then, we divided our dataset into a development set of 100 LUs and a testset of 280 LUs. We tested the Levenshtein algorithm and the Stem Overlap algorithm (SO), then we evaluated the improvement in performance of the latter taking into account information about the most frequent domain (D) and the most frequent synsets (Syn). Results are reported in Table 1.

Table 1: Mapping evaluation

	Precision	Recall	F-measure
Levenshtein	0.50	0.49	0.49
Stem Overlap (SO)	0.66	0.56	0.61
SO+Domain (D)	0.66	0.57	0.61
SO+D+Syn	0.71	0.62	0.66

We carried out several tests to set the MIN and MAX threshold in order to get the best F-measure, reported in Table 1. As for precision, the best performance obtained with SO+D+Syn and a stricter MIN/MAX threshold scored 0.78 (recall 0.36, f-measure 0.49).

4 Conclusions

We proposed a new method to map FrameNet LUs to WordNet synsets by computing a similarity measure between LU definitions and WordNet glosses. To our knowledge, this is the only approach to the task based on this kind of similarity. The only comparable evaluation available is reported in [5], and shows that our results are promising. De Cao at al. [2] reported a better performance, particularly for recall, but evaluation of their mapping algorithm relied on a gold standard of 4 selected frames having at least 10 LUs and a given number of corpus instantiations.

In the future, we plan to improve the algorithm by shallow parsing the LU definitions and the WordNet glosses. Besides, we will exploit information extracted from the WordNet hierarchy. We also want to evaluate the effectiveness of the approach focusing on the new LUs to be included in the existing frames.

References

[1] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the 36th ACL Meeting and 17th ICCL Conference*. Morgan Kaufmann, 1998.

[2] Diego De Cao, Danilo Croce, Marco Pennacchiotti, and Roberto Basili. Combining Word Sense and Usage for modeling Frame Semantics. In *Proceedings of STEP 2008*, Venice, Italy, 2008.

[3] Mario Crespo and Paul Buitelaar. Domain-specific English-to-Spanish Translation of FrameNet. In *Proc. of LREC 2008*, Marrakech, 2008.

[4] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database.* MIT Press, 1998.

[5] R. Johansson and P. Nugues. Using WordNet to extend FrameNet coverage. In *Proc. of the Workshop on Building Frame-semantic Resources for Scandinavian and Baltic Languages, at NODALIDA*, Tartu, 2007.