Using WordNet to Extend FrameNet Coverage

Johansson, Richard; Nugues, Pierre

Published in:
LU-CS-TR: 2007-240

2007

Using WordNet to Extend FrameNet Coverage

Richard Johansson and **Pierre Nugues**
Department of Computer Science, Lund University, Sweden
{richard, pierre}@cs.lth.se

Abstract

We present two methods to address the problem of sparsity in the FrameNet lexical database. The first method is based on the idea that a word that belongs to a frame is "similar" to the other words in that frame. We measure the similarity using a WordNet-based variant of the Lesk metric. The second method uses the sequence of synsets in WordNet hypernym trees as feature vectors that can be used to train a classifier to determine whether a word belongs to a frame or not. The extended dictionary produced by the second method was used in a system for FrameNet-based semantic analysis and gave an improvement in recall. We believe that the methods are useful for bootstrapping FrameNets for new languages.

1 Introduction

Coverage is one of the main weaknesses of the current FrameNet lexical database; it lists only 10,197 lexical units, compared to 207,016 word–sense pairs in WordNet 3.0. This is an obstacle to fully automated frame-semantic analysis of unrestricted text.

This work addresses this weakness by using WordNet to bootstrap an extended dictionary. We report two approaches: first, a simple method that uses a similarity measure to find words that are related to the words in a given frame; second, a method based on classifiers for each frame that uses the synsets in the hypernym trees as features. The dictionary that results from the second method is three times as large as the original one, thus yielding an increased coverage for frame detection in open text.

Previous work that has used WordNet to extend FrameNet includes Burchardt et al. (2005), which applied a WSD system to tag FrameNet-annotated predicates with a WordNet sense. Hyponyms were then assumed to evoke the same frame. Shi and Mihalcea (2005) used VerbNet as a bridge between FrameNet and WordNet for verb targets, and their mapping was used by Honnibal and Hawker (2005) in a system that detected target words and assigned frames for verbs in open text.

1.1 Introduction to FrameNet and WordNet

FrameNet (Baker et al., 1998) is a medium-sized lexical database that lists descriptions of English words in Fillmore's paradigm of Frame Semantics (Fillmore, 1976). In this framework, the relations between predicates, or in FrameNet terminology, *target words*, and their arguments are described by means of *semantic frames*. A frame can intuitively be thought of as a template that defines a set of slots, *frame elements*, that represent parts of the conceptual structure and correspond to prototypical participants or properties. In Figure 1, the predicate *statements* and its arguments form a structure by means of the frame STATEMENT. Two of the slots of the frame are filled here: SPEAKER and TOPIC. The initial versions of FrameNet focused on describing situations and events, i.e. typically verbs and their nominalizations. Currently, however, FrameNet defines frames for a wider range of semantic relations, such as between nouns and their modifiers. The frames typically describe events, states, properties, or objects. Different senses for a word are represented in FrameNet by assigning different frames.

As usual in these cases, [both parties]$_{\text{SPEAKER}}$ agreed to make no further **statements** [on the matter]$_{\text{TOPIC}}$.

Figure 1: Example sentence from FrameNet.

WordNet (Fellbaum, 1998) is a large dictionary whose smallest unit is the *synset*, i.e. an equivalence class of word senses under the synonymy relation. The synsets are organized hierarchically using the is-a relation.

2 The Average Similarity Method

Our first approach to improving the coverage, the Average Similarity method, was based on the intuition that the words belonging to the same frame frame show a high degree of "relatedness." To find new lexical units, we look for lemmas that have a high average relatedness to the words in the frame according to some measure. The measure used in this work was a generalized version of the Lesk measure implemented in the WordNet::Similarity library (Pedersen et al., 2004). The Similarity package includes many measures, but only four of them can be used for words having different parts of speech: Hirst & St-Onge, Generalized Lesk, Gloss Vector, and Pairwise Gloss Vector. We used the Lesk measure because it was faster than the other measures. Small-scale experiments suggested that the other three measures would have resulted in similar or inferior performance.

For a given lemma l, we measured the relatedness $\mathrm{sim}_F(l)$ to a given frame F by averaging the maximal relatedness, in a given similarity measure sim, over each sense pair for each lemma λ listed in F:

$$\mathrm{sim}_F(l) = \frac{1}{|F|} \sum_{\lambda \in F} \max_{\substack{s \in \mathrm{senses}(l) \\ \sigma \in \mathrm{senses}(\lambda)}} \mathrm{sim}(s, \sigma)$$

If the average relatedness was above a given threshold, the word was assumed to belong to the frame.

For instance, for the word *careen*, the Lesk similarity to 50 randomly selected words in the SELF_MOTION frame ranged from 2 to 181, and the average was 43.08. For the word *drink*, which does not belong to SELF_MOTION, the similarity ranged from 1 to 45, and the average was 13.63. How the selection of the threshold affects precision and recall is shown in Section 4.1.

3 Hypernym Tree Classification

In the second method, Hypernym Tree Classification, we used machine learning to train a classifier for each frame, which decides whether a given word belongs to that frame or not. We designed a feature representation for each lemma in WordNet, which uses the sequence of unique identifiers ("synset offset") for each synset in its hypernym tree.

We experimented with three ways to construct the feature representation:

```
Sense 1 (1 example)
{01924882} stagger, reel, keel, lurch, swag, careen
   => {01904930} walk
     => {01835496} travel, go, move, locomote

Sense 2 (0 examples)
{01884974} careen, wobble, shift, tilt
   => {01831531} move


1924882:0.67 1904930:0.67 1835496:0.67
1884974:0.33 1831531:0.33
```

Figure 2: WordNet output for the word *careen*, and the resulting weighted feature vector

First sense only. In this representation, the synsets in the hypernym tree of the first sense was used.

All senses. Here, we used the synsets of all senses.

Weighted senses. In the final representation, all synset were used, but weighted with respect to their relative frequency in SemCor. We added 1 to every frequency count.

Figure 2 shows the WordNet output for the word *careen* and the corresponding sense-weighted feature representation.

Using these feature representations, we trained an SVM classifier for each frame that tells whether a lemma belongs to that frame or not. We used the LIBSVM library (Chang and Lin, 2001) to train the classifiers.

4 Evaluation

4.1 Precision and Recall for SELF_MOTION

To compare the two methods, we evaluated their respective performance on the SELF_MOTION frame. We selected a training set consisting of 2,835 lemmas, where 50 of these were listed in FrameNet as belonging to SELF_MOTION. As a test set, we used the remaining 87 positive and 4,846 negative examples. Both methods support precision/recall tuning: in the Average Similarity method, the threshold can be moved, and in the Hypernym Tree Classification method, we can set a threshold on the probability output from LIBSVM. Figure 3 shows a precision/recall plot for the two methods obtained by varying the thresholds.

The figures confirm the basic hypothesis that words in the same frame are generally more related,

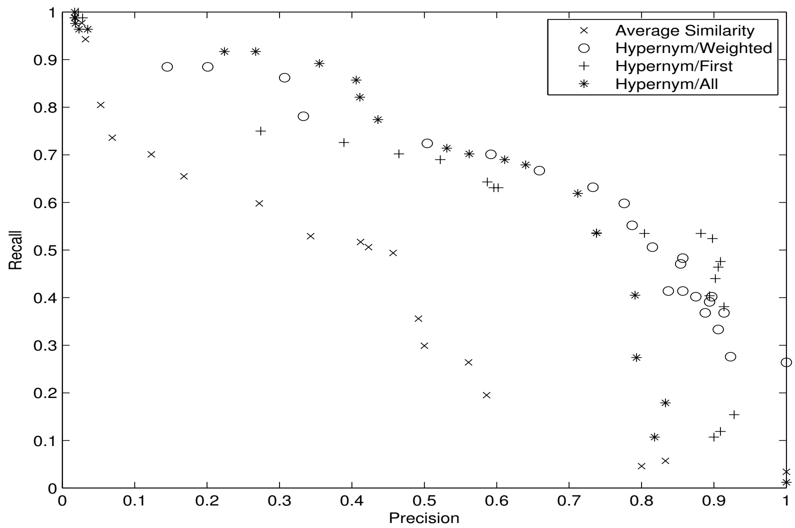

Figure 3: Precision/recall plot for the SELF_MOTION frame.

but the Average Similarity method is still not as precise as the Hypernym Tree Classification method, which is also much faster. Of the hypernym tree representation methods, the difference is small between first-sense and weighted-senses encodings, although the latter has higher recall in some ranges. The all-senses encoding generally has lower precision. We used the Hypernym Tree method with weighted-senses encoding in the remaining experiments.

4.2 All Frames

We also evaluated the performance for all frames. Using the Hypernym Tree Classification method with frequency-weighted feature vectors, we selected 7,000 noun, verb, and adjective lemmas in FrameNet as a training set and the remaining 1,175 as the test set – WordNet does not describe prepositions, and has no hypernym trees for adverbs. We set the threshold for LIBSVM's probability output to 50%. When evaluting on the test set, the system achieved a precision of 0.788 and a recall of 0.314. This can be compared to the result for from the previous section for the same threshold: precision 0.787 and recall 0.552.

4.3 Dictionary Inspection

By applying the hypernym tree classifiers on a list of lemmas, the FrameNet dictionary could be extended by 18,372 lexical units. If we assume a Zipf distribution and that the lexical units already in FrameNet are the most common ones, this would increase the coverage by up to 9%.

We roughly estimated the precision to 70% by manually inspecting 100 randomly selected words in the extended dictionary, which is consistent with the result in the previous section. The quality seems to be higher for those frames that correspond to one or a few WordNet synsets (and their subtrees). For instance, for the frame MEDICAL_CONDITION, we can add the complete subtree of the synset *pathological state*, resulting in 641 new lemmas referring to all sorts of diseases. In addition, the strategy also works well for motion verbs (which often exhibit complex patterns of polysemy): 137 lemmas could be added to the SELF_MOTION frame. Examples of frames with frequent errors are LEADERSHIP, which includes many insects (probably because the most frequent sense of *queen* is the queen insect), and FOOD, which included many chemical substances as well as inedible plants and animals.

4.4 Open Text

We used the extended dictionary in the Semeval-2007 task on Frame-semantic Structure Extraction (Baker, 2007). A part of the task was to find target words in open text and correctly assign them frames.

Our system (Johansson and Nugues, 2007) was evaluated on three short texts. In the test set, the new lexical units account for 53 out of the 808 target words our system detected (6.5% – this is roughly consistent with the 9% hypothesis in the previous section).

Table 1 shows the results for frame detection averaged over the three test texts. The table shows exact and approximate precision and recall, where the approximate results give partial credit to assigned frames that are closely related to the gold-standard frame. We see that the extended dictionary increases the recall – especially for the approximate case – while slightly lowering the precision.

Table 1: Results for frame detection.

	Original	**Extended**
Exact P	0.703	0.688
Exact R	0.504	0.528
Approx. P	0.767	0.758
Approx. R	0.550	0.581

5 Conclusion and Future Work

We have described two fully automatic methods to add new units to the FrameNet lexical database. The enlarged dictionary gave us increased recall in an experiment in detection of target words in open text. Both methods support tuning of precision versus recall, which makes it easy to adapt to applications: while most NLP applications will probably favor a high F-measure, other applications such as lexicographical tools may require a high precision.

While the simple method based on SVM classification worked better than those based on similarity measures, we think that the approaches could probably be merged, for instance by training a classifier that uses the similarity scores as features. Also, since the words in a frame may form disjoint clusters of related words, the similarity-based methods could try to measure the similarity to a subset of a frame rather than the complete frame. In addition to the WordNet-based similarity measures, distribution-based measures could possibly also be used.

More generally, we think that much could be done to link WordNet and FrameNet in a more explicit way, i.e. to add WordNet sense identifiers to FrameNet lexical units. The work of Shi and Mihalcea (2005) is an important first step, but so far only for verbs. Burchardt et al. (2005) used a WSD system to annotate FrameNet-annotated predicates with WordNet senses, but given the current state of the art in WSD, we think that this will not give very high-quality annotation. Possibly, we could try to find the senses that maximize internal relatedness in the frames, although this optimization problem is probably intractable.

We also think that the methods can be used in other languages. If there is a FrameNet with a set of seed examples for each frame, and if a WordNet or a similar electronic dictionary is available, both methods should be applicable without much effort.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLING-ACL'98.*

Collin Baker. 2007. SemEval task 19: Frame semantic structure extraction. In *Proceedings of SemEval-2007*, forthcoming.

Aljoscha Burchardt, Katrin Erk, and Anette Frank. 2005. A WordNet detour to FrameNet. In *Proceedings of the GLDV 2005 workshop GermaNet II*, Bonn, Germany.

Chih-Chung Chang and Chih-Jen Lin, 2001. *LIBSVM: a library for support vector machines.*

Christiane Fellbaum, editor. 1998. *WordNet: An electronic lexical database.* MIT Press.

Charles J. Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language*, 280:20–32.

Matthew Honnibal and Tobias Hawker. 2005. Identifying FrameNet frames for verbs from a real-text corpus. In *Australasian Language Technology Workshop 2005.*

Richard Johansson and Pierre Nugues. 2007. Semantic structure extraction using nonprojective dependency tress. In *Proceedings of SemEval-2007.* To appear.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity – measuring the relatedness of concepts. In *Proceedings of NAACL-04.*

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet, and WordNet for robust semantic parsing. In *Proceedings of CICLing 2005.*