MELB-YB: Preposition Sense Disambiguation Using Rich Semantic Features

Patrick Ye and Timothy Baldwin
Computer Science and Software Engineering
University of Melbourne, Australia
{jingy,tim}@csse.unimelb.edu.au

Abstract

This paper describes a maxent-based preposition sense disambiguation system entry to the preposition sense disambiguation task of the SemEval 2007. This system uses a wide variety of semantic and syntactic features to perform the disambiguation task and achieves a precision of 69.3% over the test data.

1 Introduction

Prepositional phrases (PPs) are both common and semantically varied in open English text. While the conventional view on prepositions from the computational linguistics community has been that they are semantically transient at best, and semantically-vacuous at worst, a robust account of the semantics of prepositions and disambiguation method can be helpful in a range of NLP tasks including machine translation, parsing (prepositional phrase attachment) and semantic role labelling (Durand, 1993; O'Hara and Wiebe, 2003; Ye and Baldwin, 2006a).

The SemEval 2007 preposition sense disambiguation task provides a common test bed for the evaluation of preposition sense disambiguation systems.

Our proposed method is maximum entropy based, and combines features developed in the context of preposition sense disambiguation for semantic role labelling (Ye and Baldwin, 2006a), and verb sense disambiguation (Ye and Baldwin, 2006b).

The remainder of this paper is structured as follows. We first discuss the pre-processing steps used in our system (Section 2), and outline the features our preposition disambiguation method uses (Section 3) and our parameter tuning method (Section 4). We then discuss and analyse the results of our method (Section 5) and conclude the paper (Section 6).

2 Pre-processing

The following list shows the pre-processing steps that our system goes through and the tools used:

Part of speech tagging SVMTool version 1.2 (Giménez and Màrquez, 2004).

Chunking An in-house chunker implemented with fnTBL, a transformation based learner (Ngai and Florian, 2001), and trained on the British National Corpus (BNC).[1]

Parsing Charniak's re-ranking parser, version August, 2006 (Charniak and Johnson, 2005).

Named entity extraction A statistical NER system described in Cohn et al. (2005).

Supersense tagging A WordNet-based supersense tagger (Ciaramita and Altun, 2006).

Semantic role labeling ASSERT version 1.4 (Pradhan et al., 2004).

3 Features

The disambiguation features used by our system can be divided into three categories: collocation features, syntactic features and semantic-role based features. We discuss each in turn below.

3.1 Collocation Features

The collocation features were inspired by the one-sense-per-collocation heuristic proposed by Yarowsky (1995). These features were designed to capture open class words that exhibit strong collocation properties with respect to the different senses of the target preposition. Details of the features in this category are listed below.

[1] This chunker is not exactly the same as Ngai and Florian's system, however it does use the default transformation templates supplied by fnTBL.

Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 241–244,
Prague, June 2007. ©2007 Association for Computational Linguistics

Bag of open class words The part-of-speech (POS) tags and lemmas of all the open class words that occur in the same sentence as the target preposition.

Bag of WordNet synsets The WordNet (Miller, 1993) synonym sets and their hypernyms of all the open class words that occur in the same sentence as the target preposition.

Bag of named entities Each named entity in the same sentence as the target preposition is treated as a separate feature.

Surrounding words These features are the combinations of the lemma, POS tag and relative position of the words surrounding the target preposition within a window of 7 words.

Surrounding super senses These features are the combinations of super-sense tag, POS tag and relative position of the words surrounding the target preposition within a window of 7 words.

3.2 Syntactic Features

The syntactic features were designed to capture both the flat and recursive syntactic properties of the target preposition. The flat syntactic features were derived from the surrounding POS tags and chunk tags of the target preposition; the recursive syntactic features were derived from the parse trees. The details of these feature are given below.

Surrounding POS tags These features are the combination of POS tag and relative position of the words surrounding the target preposition within a window of 7 words.

Surrounding chunk tags These features are the combination of IOB style chunk tag and relative position of the words surrounding the target preposition within a window of 5 words.

Surrounding chunk types Instead of using only the chunk tags themselves, we also extracted the actual chunk types (NP, VP, ADJP, etc) of the words surrounding the target preposition within a window of 5 words. Each chunk type is also combined with its relative position to the target preposition as a separate feature.

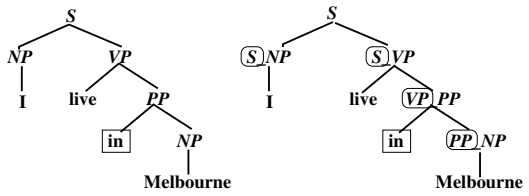

Figure 1: Parse tree examples

Parse tree features Given the position of the target preposition p in the parse tree, the basic form of the corresponding parse tree feature is just the list of nodes of p's siblings in the tree (the POS tags are treated as part of the terminal). For example, suppose the original parse tree for the sentence *I live in Melbourne* is the left tree in Figure 1, for the target preposition *in*, the basic form of the parse tree feature would be (1, NP). In order to gain more syntactic information, we further annotated each non-terminal of the parse tree with its parent node, and used the new non-terminals as our features. The right tree in Figure 1 shows the result of applying this annotation once to the original parse tree. Two levels of additional annotation were performed on the original parse trees in our feature extraction.

3.3 Semantic-Role Based Features

Finally, since prepositional phrases can often function as the temporal, location, and manner modifiers for verbs, we designed semantic-role-based features to specifically capture this type of verb-preposition semantic information. The details of these features are as follows:

Surrounding semantic role tags The semantic role tags of the words surrounding the target preposition within a window of 5 words are combined with their relative positions to the target preposition and treated as separate features. For example, consider the preposition *on* in the sentence *The man who stole my car on Sunday has apologised to me*, the semantic roles for the two verbs (*stole* and *apologised*) are shown in Table 1. The semantic roles for *stole* would generate the following features: (-5, I-A0), (-4, R-A0), (-3, TARGET), (-2, B-A1), (-1, I-A1), (0, B-AM-TMP), (1, I-AM-TMP), (2, O), (3, O), (4, O and (5, O).

Attached verbs This feature was designed to capture the verb-particle and verb-preposition-

	The	man	who	stole	my	car	on	Sunday	has	apologised	to	me
stole	B-A0	I-A0	R-A0	TARGET	B-A1	I-A1	B-AM-TMP	I-AM-TMP	O	O	O	O
apologised	B-A0	I-A0	I-A0	I-A0	I-A0	I-A0	I-A0	I-A0	O	TARGET	B-A2	I-A2

Table 1: Example semantic-role-labelled sentence

attachment relationships between verbs and prepositions. There are two situations in which a preposition p is deemed to be attached to a verb v: (1) p has a semantic role tag relative to v and this tag is a 'B' tag, (2) p has no semantic role tag relative to v, but the first token to the right of p has a 'B' tag relative to v. In the sentence shown in Table 1, *stole* would be considered as the governor of *on*.

Verb's relative position The lemma of each verb in the same sentence as the target preposition is combined with its relative position to the target preposition and treated as a separate feature. For example, the sentence shown in Table 1 would generate the two features: (-1, steal) and (1, apologize).

More detailed descriptions and examples for these features may be found in Ye and Baldwin (2006b).

4 Parameter Tuning

We used the ranking-based feature selection method from Ye and Baldwin (2006b) to select the most relevant feature based on our training data. This method works in two steps. Firstly, we calculated the information gain, gain ratio and Chi-squared statistics for each feature, and used these values to generate 3 sets of rankings for the features. We then summed up the individual ranks, and used the sums to create a set of final rankings for the features.

The feature selection process is based on 10-fold cross validation: we divided our training data into 10 pairs of training-test datasets; then for each fold, we extracted the top $N\%$ ranked features using our feature selection heuristic from the cv-training set (where N was set to values 5, 10, .., 100), and used these features to test the held-out test set. The best N as determined by the cross validation was then applied to the entire training data set.

Additionally, since we used a maximum entropy-based machine learning package,[2] it was important to determine the best Gaussian smoothing parameter g for the probability distribution. The tuning of g was incorporated into the cross validation process of feature selection.

Given the possible combinations of parameter tuning, we trained the following three classifiers for the preposition sense disambiguation task:

Non-tuned Using all the original features and 10.0 for the Gaussian smoothing parameter.

Smoothing-tuned Using all the original features but automatically tuned Gaussian smoothing parameter.

Fully-tuned Using both automatically tuned features and Gaussian smoothing parameter.

5 Results and Analysis

The overall precision (%) obtained by the three classifiers for the fine-grained senses are as follows:

Non-tuned	Smoothing-tuned	Fully tuned
67.9	68.0	69.3

The best overall results were achieved when both the features and the Gaussian smoothing parameters were automatically tuned, achieving a 1.4% absolute precision gain over the non-tuned system. However, such parameter tuning may not always be useful: the same tuning process was found to be detrimental in a Senseval-2 verb sense disambiguation task (Ye and Baldwin, 2006b). Consistent with the findings of Ye and Baldwin (2006b), the improvement caused by the tuning of the Gaussian smoothing parameter is only marginal compared with the improvement caused by the tuning of the features.

We also evaluated our features based on their categories and types. Collocation features performed the best among the three feature categories. Without any parameter tuning, the collocation-feature-only classifier achieved an overall precision of 67.4% on the test set; the semantic-role-feature-only classifier and the syntactic-feature-only classifier achieved precision of 46.9% and 50.5% respectively.

The best-performing individual features are the bag-of-words features and bag-of-synsets features.

[2] http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html

Feature type	Feature type % in top $N\%$ features			Overall % of the feature type
	10	20	30	
Bag of Words	13.46	13.43	12.94	13.37
Bag of Synsets	57.83	58.38	59.53	58.29
Verb's rel. positions	3.97	3.95	3.76	4.02
Surrounding POS tags	1.36	1.33	1.43	1.27

Table 2: Percentages of top-performing feature types in the top $N\%$ ranked features

On the test set, the bag-of-words-only classifier and the bag-of-synsets-only classifier achieved overall precision of 63.2% and 61.9% respectively.

We also analysed the top ranking features as calculated by our feature selection algorithm, as presented in Table 2. The results show the percentages of the top-performing feature types of each feature category in the top $N\%$ ranked features. It can be observed that none of the top-performing features seem to have a significantly disproportional representation in the top-ranked features. This indicates that the disambiguation power of a particular type of features is determined mostly by the number of features of that type.

On the other hand, the bag-of-words features appear to be the most effective, considering that they account for only 13.4% of the total features, but out-performed the bag-of-synsets features which account for nearly 60% of the total features.

It is also disappointing to see that the syntactic and semantic-role based features had little positive influence in the disambiguation process. However, this is perhaps caused by the sparseness of these features since they together only account for less than 10% of all the extracted features.

The overall finding from all this is that, similar to nouns and verbs, preposition sense is determined primarily by word context, and that syntactic and semantic role-based features play only a minor role.

6 Conclusions

In this paper, we have described a maximum entropy based preposition sense disambiguation system that uses a rich set of features. We have shown that this system performed well above the majority class baseline of 39.6% precision. Our analysis showed that the most important disambiguation features are collocation-based features. This indicates that the semantics of prepositions can be learnt mostly from their surrounding context, and not syntactic properties or verb-preposition semantics.

Acknowledgements

The research in this paper has been supported by the Australian Research Council through Discovery Project grant number DP0663879.

References

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 173–180, Ann Arbor, USA.

Massimiliano Ciaramita and Yasemin Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 594–602, Sydney, Australia.

Trevor Cohn, Andrew Smith, and Miles Osborne. 2005. Scaling conditional random fields using error-correcting codes. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 10–17, Ann Arbor, USA.

Jacques Durand. 1993. On the translation of prepositions in multilingual MT. In Frank Van Eynde, editor, *Linguistic Issues in Machine Translation*, pages 138–159. Pinter Publishers, London, UK.

Jesús Giménez and Lluís Màrquez. 2004. Svmtool: A general pos tagger generator based on support vector machines. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 43–46, Lisbon, Portugal.

George A. Miller. 1993. Wordnet: a lexical database for english. In *HLT '93: Proceedings of the workshop on Human Language Technology*, pages 409–409, Princeton, USA.

Grace Ngai and Radu Florian. 2001. Transformation-based learning in the fast lane. In *Proc. of the 2nd Annual Meeting of the North American Chapter of Association for Computational Linguistics (NAACL2001)*, pages 40–7, Pittsburgh, USA.

Tom O'Hara and Janyce Wiebe. 2003. Preposition semantic classification via Treebank and FrameNet. In *Proc. of the 7th Conference on Natural Language Learning (CoNLL-2003)*, pages 79–86, Edmonton, Canada.

Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2004. Support vector learning for semantic argument classification. *Machine Learning*, 60(1–3):11–39.

David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Meeting of the Association for Computational Linguistics*, pages 189–196, Cambridge, USA.

Patrick Ye and Timothy Baldwin. 2006a. Semantic role labeling of prepositional phrases. *ACM Transactions on Asian Language Information Processing (TALIP)*, 5(3):228–244.

Patrick Ye and Timothy Baldwin. 2006b. Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler. In *Proceedings of the Australasian Language Technology Workshop*, pages 141–148, Sydney, Australia.