Improving Semantic Role Labeling with Word Sense

Wanxiang Che, Ting Liu and Yongqiang Li
Research Center for Information Retrieval
MOE-Microsoft Key Laboratory of Natural Language Processing and Speech
School of Computer Science and Technology
Harbin Institute of Technology, China, 150001
`{car, tliu, yqli}@ir.hit.edu.cn`

Abstract

Semantic role labeling (SRL) not only needs lexical and syntactic information, but also needs word sense information. However, because of the lack of corpus annotated with both word senses and semantic roles, there is few research on using word sense for SRL. The release of OntoNotes provides an opportunity for us to study how to use word sense for SRL. In this paper, we present some novel word sense features for SRL and find that they can improve the performance significantly.

1 Introduction

Semantic role labeling (SRL) is a kind of shallow sentence-level semantic analysis and is becoming a hot task in natural language processing. SRL aims at identifying the relations between the predicates in a sentence and their associated arguments. At present, the main stream researches are focusing on feature engineering or combination of multiple results.

Word senses are important information for recognizing semantic roles. For example, if we know "cat" is an "agent" of the predicate "eat" in a sentence, we can guess that "dog" can also be an "agent" of "eat". Word sense has been successfully used in many natural language processing tasks, such as machine translation (Chan et al., 2007; Carpuat and Wu, 2007). CoNLL 2008 shared task (Surdeanu et al., 2008) first introduced the predicate classification task, which can be regarded as the predicate sense disambiguation. Meza-Ruiz and Riedel (2009) has shown that the predicate sense can improve the final SRL performance. However, there is few discussion about the concrete influence of all word senses, i.e. the words besides predicates. The major reason is lacking the corpus, which is both annotated with all word senses and semantic roles.

The release of OntoNotes corpus provides an opportunity for us to verify whether all word senses can help SRL. OntoNotes is a large corpus annotated with constituency trees (based on Penn Treebank), predicate argument structures (based on Penn PropBank) and word senses. It has been used in some natural language processing tasks, such as joint parsing and named entity recognition (Finkel and Manning, 2009) and word sense disambiguation (Zhong et al., 2008).

In this paper, we regard the word sense information as additional SRL features. We compare three categories of word sense features (subtree-word related sense, predicate sense, and sense path) and find that the subtree-word related sense feature is ineffective, however, the predicate sense and the sense path features can improve the SRL performance significantly.

2 Data Preparation

In our experiments, we use the OntoNotes Release 2.0[1] corpus (Hovy et al., 2006). The OntoNotes project leaders describe it as "a large, multilingual richly-annotated corpus constructed at 90% internanotator agreement." The corpus has been annotated with multiple levels of annotation, including constituency trees, predicate argument structure, word senses, co-reference, and named entities. For this work, we focus on the constituency trees, word senses, and predicate argument structures. The corpus has English and Chinese portions, and we just use the English portion, which has been split into seven sections: ABC, CNN, MNB, NBC, PRI, VOA, and WSJ. These sections represent a mix of speech and newswire data.

Because we used SRL system based on dependence syntactic trees, we convert the constituency

---

[1] http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2008T04

trees into dependence trees with an Constituent-to-Dependency Conversion Tool[2]. In addition, we also convert the OntoNotes sense of each polysemant into WordNet sense using sense inventory file provided by OntoNotes 2.0. For an OntoNotes sense with more than one WordNet sense, we simply use the foremost (more popular) one.

3 Semantic Role Labeling System

Our baseline is a state-of-the-art SRL system based on dependency syntactic tree (Che et al., 2009). A maximum entropy (Berger et al., 1996) classifier is used to predict the probabilities of a word in the sentence to be each semantic role. A virtual role "NULL" (presenting none of roles is assigned) is added to the roles set, so it does not need semantic role identification stage anymore. For a predicate, two classifiers (one for noun predicates, and the other for verb predicates) predict probabilities of each word in a sentence to be each semantic role (including virtual role "NULL"). The features used in this stage are listed in Table 1.

Feature	Description
FirstwordLemma	The lemma of the first word in a subtree
HeadwordLemma	The lemma of the head word in a subtree
HeadwordPOS	The POS of the head word in a subtree
LastwordLemma	The lemma of the last word in a subtree
POSPath	The POS path from a word to a predicate
PathLength	The length of a path
Position	The relative position of a word with a predicate
PredicateLemma	The lemma of a predicate
RelationPath	The dependency relation path from a word to a predicate

Table 1: Features that are used in SRL.

4 Word Sense for Semantic Role Labeling

From Table 1, we can see that there are lots of lemma or POS related features. However, the lemma feature is very sparse and may result in data sparseness

problem. As for the POS, it represents the syntactic information, but is not enough to distinguish different semantic roles. Therefore, we need a kind of new feature, which is general than the lemma and special than the POS.

The word sense just satisfies the requirement. Thus, we will add some new features related with word sense for SRL. Generally, the original features can be classified into three categories:

1. Subtree-word related: FirstwordLemma, LastwordLemma, HeadwordLemma, and HeadwordPOS

2. Predicate related: PredicateLemma

3. Word and predicate related: POSPath, RelationPath, PathLenght, and Position

Correspondingly, we add three categories of word sense features by replacing Lemma or POS into Sense, i.e.

1. Subtree-word related sense: FirstwordSense, LastwordSense, and HeadwordSense

2. Predicate related sense: PredicateSense

3. Word and predicate related sense: SensePath

Three strategies are designed to adopt these senses:

1. Lemma+Sense: It is the original word sense representation in OntoNotes, such as "dog.n.1". In fact, This is a specialization of the lemma.

2. Hypernym(n): It is the hypernym of a word sense, e.g. the hypernym of "dog.n.1" is "canine.n.1". The n means the level of the hypernym. With the increasing of n, the sense becomes more and more general. In theory, however, this strategy may result in inconsistent sense, e.g. word "dog" and "canine" have different hypernyms. The same problem occurs with Basic Concepts method (Izquierdo et al., 2007).

3. Root_Hyper(n): In order to extract more consistent sense, we use the hypernym of a word sense counting from the root of a sense tree, e.g. the root hypernym of "dog.n.1" is "entity.n.1". The n means the level of the root hypernym. With the increasing of n, the sense

becomes more and more special. Thus, word "dog" and "canine" have the same Root_Hyper: "entity", "physical_entity", and "object" with n = 1, 2, and 3 respectively.

5 Experiments

We will do our experiments on seven of the OntoNotes English datasets described in Section 2. For each dataset, we aimed for roughly a 60% train / 20% development / 20% test split. See Table 2 for the detailed statistics. In order to examine the influence of word senses in isolation, we use the human annotated POS, parse trees, and word senses provided by OntoNotes. The lemma of each word is extracted using WordNet tool.

	Training	Developing	Testing
ABC	669 (0001-0040)	163 (0041-0054)	138 (0057-0069)
CNN	1,691 (0001-0234)	964 (0235-0331)	1,146 (0333-0437)
MNB	381 (0001-0015)	130 (0016-0020)	125 (0021-0025)
NBC	351 (0001-0025)	129 (0026-0032)	86 (0033-0039)
PRI	1,205 (0001-0067)	384 (0068-0090)	387 (0091-0112)
VOA	1,238 (0001-0159)	325 (0160-0212)	331 (0213-0264)
WSJ	8,592 (0020-1446)	2,552 (1447-1705)	3,432 (1730-2454)
All	14,127	4,647	5,645

Table 2: Training, developing and testing set sizes for the seven datasets in sentences. The file ranges (in parenthesis) refer to the numbers within the names of the original OntoNotes files.

The baseline SRL system without sense information is trained with all the training corpus as described in Section 3. Its performance on the development data is F1 = 85.48%.

Table 3 shows the performance (F1) comparison on the development data among different sense extracting strategies with different feature categories. The numbers are the parameter n used in Hypernym and Root_Hyper strategies.

From Table 3, we can find that:

1. Both of the predicate sense feature and the sense path feature can improve the performance. For

		Subtree-word related sense	Predicate sense	Sense path
Lemma+Sense		85.34%	86.16%	85.69%
Hypernym(n)	1	85.41%	86.12%	85.74%
	2	85.48%	86.10%	85.74%
	3	85.38%	86.10%	85.69%
Root_Hyper(n)	1	85.35%	86.07%	85.96%
	2	85.45%	86.13%	85.86%
	3	85.46%	86.05%	85.91%

Table 3: The performance comparison on the development data among different sense extracting strategies with different feature categories.

the predicate sense feature, we arrive at the same conclusion with Meza-Ruiz and Riedel (2009). As for the sense path feature, it is more special than the POS, therefore, it can enhance the precision.

2. The subtree-word related sense is almost useless. The reason is that the original lemma and POS features have been able to describe the subtree-word related information. This kind of sense features is just reduplicate.

3. For different sense feature categories (columns), the performance is not very seriously affected by different sense extracting strategies (rows). That is to say, once the sense of a word is disambiguated, the sense expressing form is not important for SRL.

In order to further improve the performance, we add the predicate sense and the sense path features simultaneously. Here, we select the Lemma+Sense strategy for the predicate sense and the Root_Hyper(1) strategy for the sense path. The final performance achieves F1 = 86.44%, which is about 1% higher than the baseline (F1 = 85.48%).

Finally, we compare the baseline (without sense) result with the word sense result on the test data. In order to see the contribution of correct word senses, we introduce a simple sense determining strategy, which use the first (the most popular) WordNet sense for each word. The final detailed comparison results are listed in Table 4.

Averagely, both of the methods with the first sense and the correct sense can perform better than the baseline. However, the improvement of the method with the first sense is not significant (χ^2-test[3] with

---

[3]http://graphpad.com/quickcalcs/chisquared1.cfm

		Precision	Recall	F1
ABC	w/o sense	86.25	83.01	84.60
	first sense	84.91	81.71	83.28
	word sense	87.13	83.40	85.22
CNN	w/o sense	86.67	79.97	83.19
	first sense	86.94	80.73	83.72
	word sense	87.75	80.64	84.05
MNB	w/o sense	85.29	81.69	83.45
	first sense	85.04	81.85	83.41
	word sense	86.96	82.47	84.66
NBC	w/o sense	84.49	76.42	80.26
	first sense	84.53	76.63	80.38
	word sense	86.20	77.44	81.58
PRI	w/o sense	86.48	82.29	84.34
	first sense	86.82	83.10	84.92
	word sense	87.45	83.14	85.24
VOA	w/o sense	89.87	86.65	88.23
	first sense	90.01	86.60	88.27
	word sense	91.35	87.10	89.18
WSJ	w/o sense	88.38	82.93	85.57
	first sense	88.72	83.29	85.92
	word sense	89.25	84.00	86.54
Avg	w/o sense	87.85	82.46	85.07
	first sense	88.11	82.85	85.40
	word sense	88.84	83.37	86.02

Table 4: The testing performance comparison among the baseline without (w/o) sense information, the method with the first sense, and the method with the correct word sense.

$\rho < 0.01$). Especially, for some sections, such as ABC and MNB, it is harmful to the performance. In contrast, the correct word sense can improve the performance significantly (χ^2-test with $\rho < 0.01$)and consistently. These can further prove that the word sense can enhance the semantic role labeling.

6 Conclusion

This is the first effort to adopt the word sense features into semantic role labeling. Experiments show that the subtree-word related sense features are ineffective, but the predicate sense and the sense path features can improve the performance significantly. In the future, we will use an automatic word sense disambiguation (WSD) system to obtain word senses and study the function of WSD for SRL.

References

Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22.

Marine Carpuat and Dekai Wu. 2007. Improving statistical machine translation using word sense disambiguation. In *Proceedings of EMNLP/CoNLL-2007*, pages 61–72, Prague, Czech Republic, June.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of ACL-2007*, pages 33–40, Prague, Czech Republic, June.

Wanxiang Che, Zhenghua Li, Yongqiang Li, Yuhang Guo, Bing Qin, and Ting Liu. 2009. Multilingual dependency-based syntactic and semantic parsing. In *Proceedings of CoNLL-2009*, pages 49–54, Boulder, Colorado, June.

Jenny Rose Finkel and Christopher D. Manning. 2009. Joint parsing and named entity recognition. In *Proceedings of NAACL/HLT-2009*, pages 326–334, Boulder, Colorado, June.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. Ontonotes: The 90% solution. In *Proceedings of NAACL/HLT-2006*, pages 57–60, New York City, USA, June.

Rubén Izquierdo, Armando Suárez, and German Rigau. 2007. Exploring the automatic selection of basic level concepts. In *Proceedings of RANLP-2007*.

Ivan Meza-Ruiz and Sebastian Riedel. 2009. Jointly identifying predicates, arguments and senses using markov logic. In *Proceedings of NAACL/HLT-2009*, pages 155–163, Boulder, Colorado, June.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The conll 2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of CoNLL-2008*, pages 159–177, Manchester, England, August.

Zhi Zhong, Hwee Tou Ng, and Yee Seng Chan. 2008. Word sense disambiguation using OntoNotes: An empirical study. In *Proceedings of EMNLP-2008*, pages 1002–1010, Honolulu, Hawaii, October.