An Improved Semantic Role Labeling System for Chinese

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Third Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Abstract

We present an improved semantic role labeling system for Chinese parsed sentences. Our new system outperformed the previous reported one from two aspects: knowledge acquisition{utilization} and model design. As to the former, the semantic knowledge obtained from E-Hownet were utilized to solve the data sparseness issue; as to the latter, a combination of back-off models was proposed for semantic role classification. They enhanced the performance by 22.45% and 77.55%, respectively. Further performance gains through post-processing lead to an overall accuracy improvement from 92.71% to 94.81%.

1 Introduction

Over the last few years, syntactic parsing paradigms have enjoyed an admirable success, and this has paved the way for semantic parsing. Semantic role labeling (SRL), also known as shallow semantic parsing, is the task of semantically annotating natural language text. Conventionally, a syntactically parsed sentence is taken as input, and semantic arguments associated with predicate of the sentence are identified and classified to a particular semantic class. (Gildea and Jurafsky, 2002) were the first one to build an automatic SRL system, and since then, their ideas have been dominating the field. In their approach, they emphasized on selection of appropriate lexical and syntactical features for SRL, use of statistical classifiers and their combinations, and ways to handle the data sparseness issue. People have tried to build on that by augmenting and/or altering the feature set (Chen and Rambow, 2003; Xue, 2004), by experimenting with various classification approaches (Park and Rim, 2005; Tan et al., 2009), and by trying different ways to handle data sparseness concern (Zapirain et al., 2007; Lin et al., 2010).

In this study, we follow the traditions to enhance performance of an already reported SRL system for Chinese (You and Chen, 2004) by (1) enriching its feature set (2) using a semantic knowledgebase to address the data sparseness. However, we take a step further, and propose a different classification approach that is based on a combination of weighted simple probabilistic models. To show the effectiveness of our idea, we build a number of systems that are based on other well established classification approaches (e.g. NaiveBayes, Decision Trees, Maximum Entropy, Linear Interpolation), and compare the outcomes. The experimental results show that our strategy outperformed all other systems, and lead to a considerable improvement in accuracy of the previously reported system.

2 Experimental Material

2.1 Sinica Treebank

Sinica treebank (Chen et al., 2003) is a semantically annotated treebank for Chinese. Its version 3.0 contains 61,087 trees and 361,834 Chinese words. In addition to semantic relations of a verbal predicate, phrasal head-modifier relations have also been marked. There are 74 semantic roles including primary thematic roles (e.g. 'agent', 'theme', etc), secondary roles (e.g. 'location', 'time', 'place', etc), and noun-modification roles (e.g. 'quantifier', 'possessor' etc.). In this study, we have used this treebank as our training and testing data.

2.2 E-HowNet

E-HowNet(**?**) is a semantic lexical knowledgebase for Chinese. It is an entity-relation model that defines relationships between groups of words

(”entities”) based on their semantic attributes. The grouping, which is based on sense similarity, can be used to generalize lexical statistics. In this study, we have used E-HowNet to abate the effects of data sparseness on SRL by generalizing lexical statistics.

3 Previous SRL System

A semantic role labeling (SRL) system for Chinese was reported by (You and Chen, 2004). Sinica Treebank was used to train simple probabilistic models, which, together with a back-off strategy, were used for semantic role classification. As far feature set, the system relied on the following conventional lexical and syntactic features:

Lexical Features:
`h_word`: Head word
`h_pos`: Part of speech tag of the head word
`t_word`: Target word
`t_pos`: Part of speech tag of the target word

Syntactic Features:
`pt`: Phrase type
`position`: Position of the target word w.r.t verb

The overall accuracy of the system was reported to be 92.71%. Even though this accuracy can be considered on the higher side, there was room for improvement, especially in two subtasks: data sparseness handling and classification approach. We have considerably improved the performance of the system by better addressing the data sparseness issue (Section 4.1), and by proposing a different classification approach (Section 4.3).

4 Revised SRL System

4.1 Feature Set

Lexical Features:
`h_pos`: Part of speech tag of the head word
`t_pos`: Part of speech tag of the target word
`left_right_child_pos`: Part of speech tags of the immediate left and right siblings of a test node
`all_pos`: A set of part of speech tags of all nodes under a test node including the test node itself

Syntactic Features:
`pt`: phrase type
`passive`: A sentence-level boolean feature indicating weather the sentence, containing the test node, is passive or not
`position`: Position of the target word w.r.t verb

Semantic Features:
`semType_h_pos`: Semantic type of the head word extracted from E-HowNet
`semType_t_pos`: Semantic type of the target word extracted from E-HowNet
`all_semType`: A set of semantic types of all nodes of the tree, the test node is a child of

As pointed out by (Gildea and Jurafsky, 2002), lexical statistics, though very useful for semantic role labeling, often becomes a source of data sparseness. For this reason, we have generalized two lexical features (i.e. `h_word`, and `t_word`) by replacing them with more general features (i.e. `semType_h_word` and `semType_t_word`) respectively using Chinese semantic knowledgebase (E-HowNet). For our ten probabilistic models (see Section 4.2 for details about these models), Table 1 gives coverage and the corresponding accuracy statistics before and after adding these generalizations. As can be seen, after generalization, the coverage improved considerably for each feature combination, which ultimately resulted in better performance.

P_{f_c}	**without E-HowNet**		**with E-HowNet**	
	Coverage	Accuracy	Coverage	Accuracy
1	7.10%		12.68%	
2	52.06%		73.05%	
3	67.43%		82.68%	
4	74.56%		93.35%	
5	97.97%		97.97%	
6	28.60%		64.20%	
7	76.30%		94.23%	
8	82.97%		96.85%	
9	99.78%		99.78%	
10	90.09%		99.00%	

Table 1: Coverage & Accuracy Statistics

4.2 Probabilistic Models

Ten simple probabilistic models, each based on a particular feature combination, were build using the Sinical treebank labeled data. The probabilities were estimated using the following simple formula.

$$\begin{aligned} & P(r|constituent) \\ & \quad = \quad P(r|f_c) \\ & \quad = \quad \#(r, f_c)/\#f_c \end{aligned}$$

Where f_c represents a particular feature combination. A number of combinations were tried, and for the final system we used the following set of ten combinations.

```
{(semType_h_word,h_pos,
semType_t_word,t_pos,pt,position,
all_pos,passive,all_semType,
left_right_child_pos),
(h_pos,semType_t_word,t_pos,pt,
position,passive),
(semType_h_word,h_pos,t_pos,pt,
position,passive),
(semType_t_word,t_pos,pt,passive,
position),
(h_pos,t_pos,pt,position,passive),
(semType_h_word,semType_t_word,
pt,position,passive),
(semType_t_word,t_pos,pt,passive),
(semType_t_word,t_pos,passive),
(t_pos,pt,position,passive),
(semType_t_word)}
```

4.3 Classification Method

Notations

- Let F be the set of ten feature combinations (given in the previous section), and f_c be an element of F
- P be the set of ten corresponding probabilistic models, and P_{f_c} be an element of P that is based on feature combination f_c
- D_{f_c} be the probability distribution computed by P_{f_c}
- $M_{((p,r)|f_c)}$ be the most probable $(probability, role)$ pair in D_{f_c}
- W be a set of optimal weights, w be an element of W, and wp be a weighted probability (i.e. rank)

Algorithm

Our classification algorithm is as follows:

1. For a test candidate, extract values of all features (mentioned in section 2) using the parse tree and E-HowNet.
2. initialize $potential\_roles \leftarrow empty$
3. For each $f_c \in F$ do:
   - Find probability distribution D_{f_c} using the corresponding P_{f_c} model
   - Select $M_{((p,r)|f_c)}$ from D_{f_c}
   - Rank $M_{((p,r)|f_c)}$ by multiplying p with the corresponding w from W
   - Append (wp, r) to $potential\_roles$
4. Return the top ranked r from $potential\_roles$

4.4 An Example

Suppose we want to assign a semantic role to the circled node of the parse tree given in Figure 1. First, we have to extract the full set

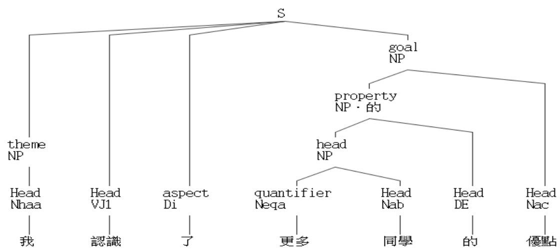

Figure 1: An example parsed sentence

of `(feature:value)` pairs for the target test node. The extracted set is given below:

```
{(semType_h_word:advantage),
(h_pos:Nac),(semType_t_word:human),
(t_pos:NP??),(pt:NP),(position:-1),
(all_pos:NP??-NP-Neqa-Nab-DE),
(passive:False),
(left_right_child_pos:empty-Nac),
(all_semType:
BecomeMore-human-tool-advantage)}
```

The probability distribution estimated by each of the ten probabilistic models is given in Table[1] 2. Next, from each distribution, we can select $M_{((p,r)|f_c)}$. The full list is given below:

```
[NULL,(0.5072,'possessor'),
(0.5,'property'),
(0.5035,'property'),
(0.8514,'property'),
(0.5,'property'),
(0.5035,'property'),
(0.5035,'property'),
(0.8598,'property'),
(0.1915,'agent')]
```

Note that a `NULL` is inserted for the models that do not have any probability distribution.

[1]Note that due to sparseness of the training data, their is no probability distribution for the feature combination of model 1.

P_{f_c}	**Probability Distribution** (D_{f_c})
1	
2	[(0.5072, 'possessor'), (0.4783, 'property'), 0.0145, 'quantifier')]
3	[(0.5, 'property'), (0.5, 'possessor')]
4	[(0.5035, 'property'), (0.4894, 'possessor'), (0.0071, 'quantifier')]
5	[(0.8514, 'property'), (0.1361, 'possessor'), (0.0083, 'quantifier'), (0.0042, 'apposition')]
6	[(0.5, 'property'), (0.5, 'possessor')]
7	[(0.5035, 'property'), (0.4894, 'possessor'), (0.0071, 'quantifier')]
8	[(0.5035, 'property'), (0.4894, 'possessor'), (0.0071, 'quantifier')]
9	[(0.8598, 'property'), (0.1240, 'possessor'), (0.0132, 'quantifier'), (0.0015, 'apposition'), (0.0011, 'predication'), (0.0004, 'frequency')]
10	[(0.1915, 'agent'), (0.1461, 'theme'), (0.1361, 'goal'), (0.1355, 'property'), (0.0984, 'range'), (0.0775, 'DUMMY'), (0.0540, 'possessor'), (0.0533, 'apposition'), (0.0414, 'experiencer'), (0.0267, 'DUMMY2'), (0.0230, 'DUMMY1'), (0.0073, 'topic'), (0.0024, 'location'), (0.0021, 'quantifier'), (0.0021, 'causer'), (0.0007, 'predication'), (0.0007, 'manner'), (0.0003, 'time'), (0.0003, 'particle'), (0.0002, 'complement'), (0.0002, 'comparison'), (0.0001, 'target'), (0.0001, 'source'), (0.0001, 'hypothesis'), (0.0001, 'companion')]

Table 2: Probability distributions for the test constituent

Finally, each role in this list is ranked by multiplying its probability to the corresponding weight from W. These weights encode the worth of a particular feature combination in determining the semantic role, and we have used genetic algorithms and a held out development data-set to find the optimal set (i.e. W). The observed optimal set after 100 generations is given below:

```
{(w1:0.9),(w2:1.0),(w3:0.9),
(w4:0.7),(w5:0.8),(w6:0.4),
(w7:0.5),(w8:0.4),(w9:0.5),
(w10:0.4)}
```

The final list of ranked roles is:

```
[NULL,(0.5072,'possessor'),
(0.45,'property'),
(0.505307,'property'),
(0.68112,'property'),
(0.2,'property'),
(0.25175,'property'),
(0.2014,'property'),
(0.4299,'property'),
(0.0766,'agent')]
```

The top ranked `(probability,role)` pair in this list is `(0.68112,'property')`, hence the role `'property'` will be assigned to the test node by our classification method.

5 Experiments and Evaluation

To evaluate the performance and to show the usefulness of our approach, we have build the following five semantic role labeling systems:

- **DT:** Based on Decision Tree classifier
- **NB:** Based on Naive Bays classifier
- **ME:** Based on Maximum Entropy classifier
- **LI:** Based on simple probabilistic models together with linear interpolation
- **WPM:** Based on our classification approach

All of these systems were tested using the same feature set (given in Section 4.1), and the same training and testing data (i.e. Sinica Treebank). Results of a 10-fold cross validation scheme are given in Table 3. From the results, we can see

System	**Accuracy**	**Precision**
DT	91.13%	
NB	92.54%	
ME	92.70%	
LI	94.32%	
WPM	94.58%	

Table 3: Evaluation results

that our system outperformed all other systems, and linear interpolation based system is the most competitive one.

To further enhance the performance of our system, we added a post-processing component to fix some of the obvious mistakes made by the probabilistic models. One such case is disambiguation between possessor/property roles. We used a heuristic rule that if the semantic type of a target word is `human`, it is more likely to be a `possessor` than a `property`. With few other such rules, the scores were improved from 94.58% to 94.81%.

6 Conclusion

We have presented an improved SRL system for Chinese. We have shown how by using a semantic knowledge-base to handle data sparseness, and by using a weighting strategy to rank the outcomes of simple probabilistic models, the performance of an existing system was enhanced. Together with gains through simple heuristic rules, the overall accuracy was improved from 92.71% to 94.81%.

References

John Chen and Owen Rambow. 2003. Use of deep linguistic features for the recognition and labeling of semantic arguments. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, pages 41–48, Stroudsburg, PA, USA. Association for Computational Linguistics.

Keh-Jiann Chen, Chi-Ching Luo, Ming-Chung Chang, Feng-Yi Chen, ChaoJan Chen, and Chu-Ren Huang. 2003. Sinica treebank: Design criteria, representational issues and implementation. *Anne Abeille (Ed.) Treebanks Building and Using Parsed Corpora. Language and Speech series. Dordrecht:Kluwer*, pages pp231–248.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Comput. Linguist.*, 28(3):245–288, September.

Xiaojun Lin, Meng Zhang, and Xihong Wu. 2010. Chinese semantic role labeling with hierarchical semantic knowledge. In *Proceedings of the 2010 International Conference on Electrical and Control Engineering*, ICECE '10, pages 583–586, Washington, DC, USA. IEEE Computer Society.

Kyung-Mi Park and Hae-Chang Rim. 2005. Maximum entropy based semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, CONLL '05, pages 209–212, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yongmei Tan, Xu Wang, and Yong Chen. 2009. Chinese semantic role labeling using crfs and svms. In *Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference on*, pages 1–5.

Nianwen Xue. 2004. Calibrating features for semantic role labeling. In *In Proceedings of EMNLP 2004*, pages 88–94.

Jia-Ming You and Keh-Jiann Chen. 2004. Automatic semantic role assignment for a tree structure. In Oliver Streiter and Qin Lu, editors, *ACL SIGHAN Workshop 2004*, pages 109–115, Barcelona, Spain, July. Association for Computational Linguistics.

Beñat Zapirain, Eneko Agirre, and Lluís Màrquez. 2007. Ubc-upc: Sequential srl using selectional preferences: An aproach with maximum entropy markov models. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 354–357, Stroudsburg, PA, USA. Association for Computational Linguistics.