A Multi-Classifier Based 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

The performance of a semantic role labeling (SRL) system depends on a number of factors including the feature set being used, syntactic representation of the sentence being labeled, and the underlying classification approach. Previously, a number of studies have been devoted to the selection of appropriate features, and to the effect of syntactic representation on semantic role labeling. In this paper, we study the effect of classification approach on SRL by comparing the performance of a number of systems based on different classification models (i.e. NaiveBays, Decision Trees, Maximum Entropy), and a system based on linear interpolation probability distribution model. We also propose a slightly different classification approach which is based on weighted simple probabilistic models. Experimental results show that, for the task of SRL, our approach is useful and outperforms those already existing classification approaches. Using our new approach, together with more general features to handle data sparseness, and a number of other heuristic rules, we report an improvement from 92.71% to 94.8% in the accuracy of Sinca parsers's semantic role labeling component.

1 Introduction

Conventionally, almost all automatic semantic role labeling (SRL) systems take a pre-parsed sentence as input, extract a set of useful lexical and syntactical features from it, and then use a particular classification approach to label each semantic constituent of the sentence. Since semantic relationships are predictable from syntactic realizations (Gildea and Jurafsky, 2002), and the lexical features are valuable while predicting the constituent's semantic role, the selection of an appropriate feature set, including both syntactic and lexical features, becomes crucial. With (Gildea and Jurafsky, 2002) being the pioneer, a number of other studies (Xue, 2004; Pradhan et al., 2004; Sun and Jurafsky, 2004; Xue, 2008) have highlighted the importance of different features, and tried to come-up with the best possible feature set. Similarly, the importance and influence of syntactic representation on SRL has been reported by a number of studies (Johansson and Nugues, 2008; Gildea and Hockenmaier, 2003; Pradhan et al., 2008; Swanson and Gordon, 2006).

However, very little work has been reported on the effect of chosen classification approach on SRL. Normally, the researchers concentrate most on the selection of useful features for SRL, on the syntactic representation, and use a particular classification technique in the classification phase of SRL. (Gildea and Jurafsky, 2002), however, have used equal linear interpolation, EM linear interpolation, geometric mean, and a backoff strategy to compare the performance of their system. Other than this, there is no detailed study on the effect of chosen classification approach on SRL.

In this study, we outline the influence of classification approach on SRL. We have build a number of SRL systems, which are based on different classification approaches. In all systems, the input has the same syntactic representation and the same feature set in being used. First, they are trained on the same training data, then they are tested on the same testing data. Their performance is different from each other, owing much to the difference of their classification approach. Based on these findings, we propose a slightly different classification model, which is based on weighted simple probabilistic models. These models use different features and the probabilities are calculated from se-

mantically labeled Sinical treebank (**?**). We then use a weighting strategy to rank their outputs, and then select the top ranked role to be assigned to a constituent. We use genetic algorithms for finding optimal weights. ....

The rest of this paper is organized as: ....

2 Experiments

3 Evaluation

4 Discussion

5 Future Work

References

Gildea, D. and Hockenmaier, J. (2003). Identifying semantic roles using combinatory categorial grammar.

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

Johansson, R. and Nugues, P. (2008). The effect of syntactic representation on semantic role labeling.

Pradhan, S., Ward, W., Hacioglu, K., and Martin, J. H. (2004). Shallow semantic parsing using support vector machines.

Pradhan, S. S., Ward, W., and Martin, J. H. (2008). Towards robust semantic role labeling. *Comput. Linguist.*, 34(2):289–310.

Sun, H. and Jurafsky, D. (2004). Shallow semantc parsing of chinese. In Susan Dumais, D. M. and Roukos, S., editors, *HLT-NAACL 2004: Main Proceedings*, pages 249–256, Boston, Massachusetts, USA. Association for Computational Linguistics.

Swanson, R. and Gordon, A. S. (2006). A comparison of alternative parse tree paths for labeling semantic roles. In *In Proceedings of COLING/ACL-2006*.

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

Xue, N. (2008). Labeling chinese predicates with semantic roles. *Comput. Linguist.*, 34(2):225–255.