# Research on Semantic Role Labeling Method

Bo Jiang(✉) and Yuqing Lan

School of Computer Science and Technology, Beihang University, Beijing 100191, China{jiangbo1,lanyuqing}@buaa.edu.cn

**Abstract.** Semantic role labeling task is a way of shallow semantic analysis. Its research results are of great significance for promoting Machine Translation [1], Question Answering [2], Human Robot Interaction [3] and other application systems. The goal of semantic role labeling is to recover the predicate-argument structure of a sentence, based on the sentences entered and the predicates specified in the sentence. Then mark the relationship between the predicate and the argument, such as time, place, the agent, the victim, and so on. This paper introduces the main research directions of semantic role labeling and the research status at home and abroad in recent years. And summarized a large number of research results based on statistical machine learning and deep neural networks. The main purpose is to analyze the method of semantic role labeling and its current status. Summarize the development trend of the future semantic role labeling.

**Keywords:** Semantic role labeling · Semantic analysis · Deep neural networks

## 1 Introduction

Natural language processing is an interdisciplinary subject that integrates multiple disciplines such as computer science and linguistics. It is one of the core research directions of artificial intelligence. It has a history of more than 70 years since the last century. The research on natural language understanding has made great breakthroughs through the efforts of many generations. Natural language understanding is broadly divided into three levels, lexical analysis, syntactic parsing and semantic analysis. Chinese lexical analysis mainly include Chinese word segmentation, named entity recognition and word sense disambiguation. The main task of syntactic parsing is to identify the syntactic structure of a sentence. And the semantic analysis task is to let the machine understand the meaning of natural language, which is divided into shallow semantic analysis and deep semantic analysis. The goal of deep semantic analysis is to understand the content of the entire sentence. Convert the sentence into a formal representation is the general approach. Its main methods are semantic analysis based on knowledge, supervised semantic analysis and semi-supervised or unsupervised semantic analysis. However, deep semantic analysis still faces many difficulties. For example, the mapping between ordinary text and relational predicates is more difficult to solve, and the results of semantic analysis for the open field are not satisfactory. Shallow semantic analysis proposes a simple solution, its goal is not to completely analyze the meaning of the whole sentence. It only specifies which components of the


X. Liu et al. (Eds.): ChinaCom 2018, LNICST 262, pp. 252–258, 2019.
https://doi.org/10.1007/978-3-030-06161-6_25

words or phrase are in the sentence, and then identifies the component of the words in the sentence. Compared with deep semantic analysis, it can perform semantic analysis more quickly and obtain higher correct rate results. Semantic role annotation (SRL) is a shallow semantic analysis technique. It can achieve higher efficiency and correct rate in semantic analysis if compared with deep semantic analysis. Semantic role labeling (SRL) is a shallow semantic analysis technique.

Firstly, this paper introduces the research content and importance of the semantic role labeling task. Then investigate the main research directions in recent years and the research status at home and abroad. Secondly, this paper analyzes the development of semantic role labeling tasks, and studies a lot of research results based on statistical machine learning and deep neural network. It is intended to analyze the processing methods of semantic role labeling tasks and their current developments. Finally, the paper summarizes the development trend of future semantic role labeling tasks.

## 2 Semantic Role Labeling

Semantic role labeling is an intermediate process of natural language processing. Its results can not be used to express the meaning of natural language, but it can help many application systems such as Machine Translation, Question Answering, Human Robot Interaction. The high accuracy rate of labeling results can improve the precision and recall rate of these systems, which is of great significance for performance improvement. Therefore, semantic role labeling has very high research value. The goal of semantic role labeling is to recover the predicate-argument structure of a sentence, based on the sentences entered and the predicates specified in the sentence. Then mark the relationship between the predicate and the argument, such as time, place, the agent, the victim, and so on.

The development of semantic role labeling, like most natural language processing tasks, has evolved from rule-based to statistical-based to deep-based learning. The early rule-based semantic role labeling requires people with certain linguistic knowledge or computer science to manually formulate language rules. The advantage of this method is that the rules are more flexible and easy to understand, but it is difficult to cover all language rules. High cost and inefficient processing results do not meet demand. After that, the statistical natural language processing becomes mainstream. Many statistical machine learning models can deal with the semantic role labeling problem, the sequence labeling problem is regarded as the classification problem. But it also has certain limitations in model optimization and performance. With the development of deep learning, which has swept the various fields of artificial intelligence, it also brought a new development direction for semantic roles labeling. Researchers used convolutional neural networks and LSTM neural network frameworks to deal with sequence labeling problems. In recent years, some research results have been achieved, and more and more people are beginning to shift their research direction to deep learning.

### 2.1 Semantic Role Labeling Based on Statistics

With the developing of statistics-based natural language processing technology, researchers built a large number of corpus base and corpus linguistics. It enables the rapid development of statistical natural language processing and greatly improves the correct rate of model processing results.

According to the three main methods of syntactic parsing: phrase structure parsing, shallow parsing, and dependency parsing. Semantic role labeling can be divided into semantic role labeling method based on phrase structure tree, semantic role labeling method based on shallow parsing result, and semantic role labeling method based on dependency parsing result. However, the basic processes of these methods are similar. After inputting the syntax analysis tree into the semantic role labeling system, the following four processes are performed: candidate arguments pruning, argument identification, argument labeling, and post-processing. The candidate argument pruning is to select a word or phrase that may be an argument from the leaf nodes of the input grammatical structure tree. Then delete the words that cannot be arguments and get the final set of arguments. Argument recognition is to judge the candidate argument and judge whether it is argument or not. It is a two-class problem. Argument labeling is to construct some features from the tree, and obtain a semantic role description for each argument through the multi-classification problem. Finally, get the result of the semantic role labeling through the post-processing process. The process is shown in the following and Fig. 1:

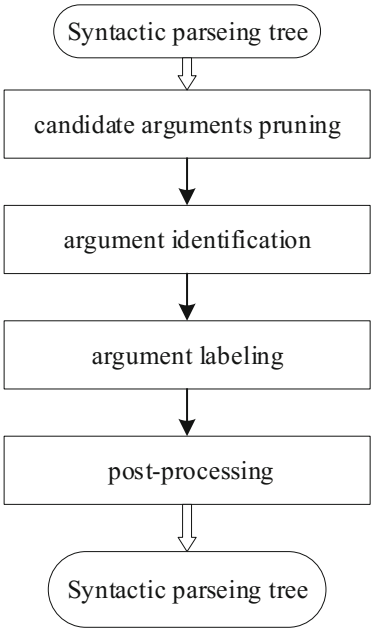


**Fig. 1.** Semantic role labeling process based on statistical machine learning

The commonly used machine learning models are Support Vector Machine (SVM), Maximum Entropy Model (ME), etc. It was first proposed by Gildea and Jurafsky (2002) [4] to use statistical machine learning models to deal with semantic role labeling tasks. They use supervised training methods to train models and training data is manually labeled. The model can achieve 65% precision and 61% recall rate when processing argument segmentation and semantic role labeling tasks. In the pre-segmented component statement, the semantic role tagging task can achieve an precision rate of 82%. Afterwards, a large number of researchers began there works on machine learning to handle semantic role labeling tasks. There are also scholars in China who have done research in this area. Liu Ting of Harbin Institute of Technology proposed a semantic role labeling system based on the maximum entropy classifier [5]. The system uses the maximum entropy classifier to identify and classify the semantic roles of predicates in sentences and obtains 75.49% and 75.60% of F1 values on the development set and test set respectively. He pointed out that various machine learning methods are mature in the feature engineering of Chinese semantic role labeling [6], and it is difficult to improve the performance of semantic role labeling by improving the machine learning model. Research rich features are more important for semantic role labeling. Li and Qian [7] of Suzhou University mainly studied the semantic role labeling of noun predicate. They further proposed the relevant feature set of noun predicate and obtained better STL processing performance.

However, the statistical role-based semantic role labeling method is highly dependent on the result of syntactic parsing. Its input is a parsing tree but syntactic parsing is a very difficult task in natural language processing. Its results generally do not have high accuracy. Further, semantic role labeling has been broken into several sub-problems. Even if each process has high accuracy, the final result may still be lower accuracy. Moreover, different machine learning models and optimization algorithms need to be selected for different processing tasks, which makes the labeling task complicated. Finally, the statistical-based semantic role labeling method has reached the technical bottleneck, and it is difficult to improve the performance through model optimization.

### 2.2 Semantic Role Labeling Based on Deep Learning

Semantic role labeling task based on deep learning is to use complex deep neural network to deal with semantic labeling problem. It uses deep learning methods to solve some problems, including traditional semantic role annotation dependent syntax analysis and needs to decompose the annotation task into multiple subtasks. It use an end-to-end manner to performs semantic role labeling in one step. This process takes an integrated approach and no longer relies on syntactic parsing, reducing the need for syntactic parsing and the risk of error accumulation.

Collobert [8] first proposed the application of deep learning to semantic role labeling tasks. They used a general convolutional neural network to treat semantic role labeling as a sequence labeling problem. The system uses the CNN network to obtain the context representation of the current word. The information used includes the current word, the distance between the current word and the predicate, the distance between the context word and the current word. The model uses a large amount of

unlabeled data for training. However, if you want to make the model have higher accuracy, you still need to use the results of syntactic parsing. And the performance of the system is lower than the traditional statistical machine learning based method.

Zhou and Xu [9] used the deep Bi-LSTM CNN as an end-to-end system. Only the original text information was used as input information and without using any grammatical knowledge in their research. The model implicitly captures the syntactic structure of the sentence. This strategy is superior to Collobert's practice and has achieved very good test results on the public test set, reaching an F1 value of 81.27%. However, this model has the problem of vanishing gradient and exploding gradient when the number of layers is high. Therefore, the model network layer can only be controlled within a range, which limits the improvement of network performance. The LSTM memory block is shown in Fig. 2:

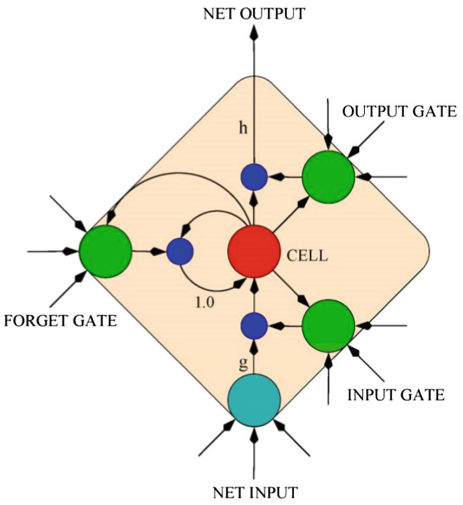


**Fig. 2.** LSTM memory block with a single cell [10].

He and Lee [11] used the deep highway bidirectional LSTMs with constrained decoding to significantly improve the performance of models. They also analyzed the advantages and disadvantages of the model in detail, and the 8-layer depth model finally achieved an F1 value of 83.4% on the test set. Their results show that the deep LSTM model has excellent performance in recovering long-distance semantic dependencies. But there are still obvious errors in this model. So there is still much room for improvement in semantic analysis techniques.

Wang and Liu [12] equipped the deep LSTM neural network with a novel "straight ladder unit" (EU), which can linearly connect the input and output of the unit. So the information can be transmitted between different layers. This design solved the problem that the number of model layers cannot be too high. The model does not require any additional feature input. It has achieved an 81.53% F1 value on the public test set

with improved performance. Table 1 compares the performance of the deep neural network model and the statistical machine learning model studied in recent years.

**Table 1.** Performance comparison of semantic role labeling models.

| | F/% | | |
|---|---|---|---|
| | Development | WSJ test | Brown test |
| Toutanova | 78.6 | 80.3 | 68.8 |
| Koomen | 77.35 | 79.44 | 67.75 |
| Pradhan | 78.34 | 78.63 | 68.44 |
| Collobert(w/parser) | 75.42 | – | – |
| Collobert(w/o parser) | 72.29 | – | – |
| Zhou | 79.6 | 82.8 | 69.4 |
| He | 82.7 | 84.6 | 73.6 |
| Wang | 81.26 | 82.95 | 72.61 |

## 3 Conclusion and Future Work

Semantic role labeling task is a way of shallow semantic analysis. Its research results are of great significance for promoting Machine Translation, Question Answering, Human Robot Interaction and other application systems. This paper introduces the semantic role labeling task and its main research directions at home and abroad in recent years. Then it summarizes a lot of research results based on statistical machine learning and deep neural network. It can be said that the research on semantic role labeling task has achieved very fruitful results, but there is still room for improvement.

Applying deep neural networks to natural language processing tasks is a research hotspot in currently because of its end-to-end characteristics and its reliance on syntactic analysis. The research work to be broken has several aspects:

(1) Construction of a deep neural network model. Among the deep learning models, it is found that the bidirectional LSTMs neural network has better performance in dealing with semantic role labeling. Its modeling idea conforms to the language generation process, and has great advantages in the processing of sequence labeling problems. However, there are many parameters in the model, and it is necessary to find a more suitable optimization method to build the model, so that the model has higher performance and better generalization ability.

(2) The construction of the corpus base. Corpus base and corpus linguistics are indispensable tools for natural language processing tasks. Its richness also plays a key role in semantic role labeling. The word vector representation trained by large-scale corpus can better help model processing label task.

(3) Chinese semantic role labeling. At present, the semantic role labeling system for training and testing is researched on English corpus, but the model for Chinese corpus is very few. Therefore, how to use the model to efficiently deal with Chinese semantic role labeling is also an important research direction.

## References

1. Knight, K., Luk, S.K.: Building a large-scale knowledge base for machine translation. Comput. Sci. 773–778 (1994)
2. Shen, D., Lapata, M.: Using semantic roles to improve question answering. EMNLP-CoNLL 2007. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 12–21, June 28–30, 2007, Prague, Czech Republic. DBLP (2007)
3. Bastianelli, E., Castellucci, G., Croce, D., et al.: Textual inference and meaning representation in human robot interaction. Newdesign.aclweb.org (2013)
4. Gildea, D., Jurafsky, D.: Automatic labeling of semantic roles. Comput. Linguist. **28**(28), 245–288 (2002)
5. Liu, T., Che, W., Li, S., et al.: Semantic role labeling system using maximum entropy classifier. In: Conference on Computational Natural Language Learning. Association for Computational Linguistics, pp. 189–192 (2005)
6. Liu, H., Che, W., Liu, T.: Feature engineering for Chinese semantic role labeling. J. Chin. Inf. Process. **21**(1), 79–84 (2007)
7. Li, J., Zhou, G., Zhu, M., et al.: Semantic role labeling of nominalized predicates in Chinese. J. Softw. **22**(8), 1725–1737 (2011)
8. Collobert, R., Weston, J., Karlen, M., et al.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(1), 2493–2537 (2011)
9. Zhou, J., Xu, W.: End-to-end learning of semantic role labeling using recurrent neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (vol. 1: Long Papers), vol. 1, pp. 1127–1137 (2015)
10. Graves, A., Liwicki, M., Fernã n S, et al.: A novel connectionist system for unconstrained handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 855–868 (2009)
11. He, L., Lee, K., Lewis, M., et al.: Deep semantic role labeling: what works and what's next. In: Meeting of the Association for Computational Linguistics, pp. 473–483 (2017)
12. Wang, M., Liu, Q.: Semantic role labeling using deep neural networks. J. Chin. Inf. Process. **32**(2) (2018)