Capturing Argument Relationships for Chinese Semantic Role Labeling

Lei Sha, Tingsong Jiang, Sujian Li, Baobao Chang, Zhifang Sui
Key Laboratory of Computational Linguistics, Ministry of Education
School of Electronics Engineering and Computer Science, Peking University
Collaborative Innovation Center for Language Ability, Xuzhou 221009 China
`shalei, tingsong, lisujian, chbb, szf@pku.edu.cn`

Abstract

In this paper, we capture the argument relationships for Chinese semantic role labeling task, and improve the task's performance with the help of argument relationships. We split the relationship between two candidate arguments into two categories: (1) Compatible arguments: if one candidate argument belongs to a given predicate, then the other is more likely to belong to the same predicate; (2) Incompatible arguments: if one candidate argument belongs to a given predicate, then the other is less likely to belong to the same predicate. However, previous works did not explicitly model argument relationships. We use a simple maximum entropy classifier to capture the two categories of argument relationships and test its performance on the Chinese Proposition Bank (CPB). The experiments show that argument relationships is effective in Chinese semantic role labeling task.

1 Introduction

Semantic Role Labeling (SRL) is defined as the task to recognize arguments for a given predicate and assign semantic role labels to them. Because of its ability to encode semantic information, there has been an increasing interest in SRL on many languages (Gildea and Jurafsky, 2002; Sun and Jurafsky, 2004). Figure 1 shows an example in Chinese Proposition Bank (CPB) (Xue and Palmer, 2003), which is a Chinese corpus annotated with semantic role labels.

Previous works of Chinese SRL include feature-based approaches and neural network based approaches. Feature-based approaches often extract a large number of handcrafted features from the sentence, and feed these features to statistical classifiers such as CRF, MaxEnt and SVM (Sun and Jurafsky, 2004; Xue, 2008; Ding and Chang, 2008; Ding and Chang, 2009; Sun, 2010). Neural network based approaches usually take Chinese SRL as sequence labeling task and use bidirectional recurrent neural network (RNN) with long-short-term memory (LSTM) to solve the problem (Wang et al., 2015).

However, both of the above two kinds of approaches identify each candidate argument separately without considering the relationship between arguments. We define two categories of argument relationships here: (1) **Compatible arguments**: if one candidate argument belongs to a given predicate, then the other is more likely to belong to the same predicate; (2) **Incompatible arguments**: if one candidate argument belongs to a given predicate, then the other is less likely to belong to the same predicate. For example, in Figure 1, the word "外商"(foreign businessman) and "企业"(entrepreneur) tend to be compatible arguments when the predicate word is "投资"(invest). On the other hand, "企业"(entrepreneur) and "规定"(rule) are not likely to belong to the same predicate "投资"(invest).

In this paper, we propose to use a quadratic optimization method to explicitly model the relationship between candidate arguments to improve the performance of Chinese SRL. We train a maximum entropy classifier, and then use the classifier to predict argument relationships between any two candidate arguments in a sentence. Experiments show that argument relationships can greatly improve the performance of Chinese SRL.

2011

Word:	保护 protect	外商 foreign businessman	投资 invest	企业 entrepreneur	合法 legal	权益 profit	六 six	项 item	规定 rule
POS:	VV	NN	VV	NN	JJ	NN	CD	M	NN
SRL:	O	S-ARG0	rel	S-arg1	O	O	O	O	O

Figure 1: A sentence with semantic roles labeled from CPB. "rel" represents the predicate, English translation: "Six rules to protect foreign businessman's legal profits when investing entrepreneurs"

2 Related Work

Semantic Role Labeling (SRL) task was first proposed by Gildea and Jurafsky (2002). Previous approaches on Chinese SRL can be classified into two categories: (1) feature-based approaches (2) neural network based approaches.

Among feature-based approaches, Sun and Jurafsky (2004) did the preliminary work on Chinese SRL without any large semantically annotated corpus and produced promising results. Xue and Palmer (2003) proposed Chinese Proposition Bank (CPB), which leads to more complete and systematic research on Chinese SRL (Xue and Palmer, 2005; Xue, 2008; Ding and Chang, 2009). Sun et al. (2009) extended the work of Chen et al. (2006), performed Chinese SRL with shallow parsing, which took partial parses as inputs. Yang and Zong (2014) proposed multi-predicate SRL, which showed improvements both on English and Chinese Proposition Bank.

Neural network based approaches are free of handcrafted features, Collobert and Weston (2008) proposed a convolutional neural network for SRL. Their approach achieved competitive performance on English SRL without requiring task specific feature. Wang et al. (2015) proposed a bidirectional LSTM-RNN for Chinese SRL.

However, most of the aforementioned approaches did not take the compatible arguments and incompatible arguments into account. Inspired by Sha et al. (2016), our approach model the two argument relationships explicitly to achieve a better performance on Chinese SRL.

3 Capturing the Relationship Between Arguments

We found that there are two typical relationships between candidate arguments: (1) Compatible arguments: if one candidate argument belongs to one event, then the other is more likely to belong to the same event; (2) incompatible arguments: if one candidate argument belongs to one event, then the other is less likely to belong to the same event.

We trained a maximum entropy classifier to predict the relationship between two candidate arguments. We choose the following features:

1. PREDICATE: the predicate in the current sentence

2. ARGUMENT DISTANCE: the distance between the two candidate arguments in the sentence

3. Whether the two candidate arguments occur on the same side of the predicate

4. PARENT DEPENDENCY DISTANCE: the distance between the two candidate arguments' parents in the dependency parse tree

5. PARENT POS: if the two candidate arguments share the same parent, take the common parent's POS tag as a feature

6. Whether the two candidate arguments occur on the same side of the common parent if the two candidate arguments share the same parent

All the words in one sentence except for the predicate are candidate arguments. The word pairs in the ground truth Chinese SRL annotation (training data) are extracted as training data. The training examples are generated as follows: For an candidate argument pair, if both of them are labeled as semantic roles, we take it as positive example. For each positive example, we randomly exchange one of the arguments with an irrelevant argument[1] to get a negative example.

---

[1] an irrelevant argument is in the same sentence with the predicate, but it is not labeled as semantic role
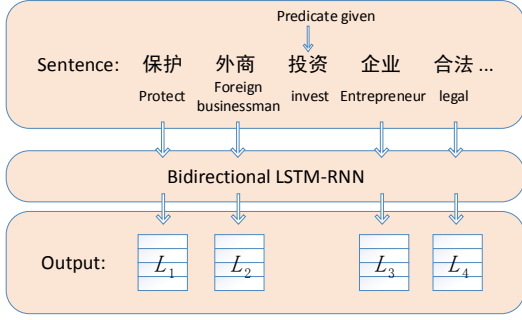
Figure 2: The bidirectional LSTM-RNN architecture.

The chinese sentences should be segmented to chinese words first. For a sentence with $n+1$ words, we denote $C \in \mathbb{R}^{n \times n}$ as the argument relationship matrix. In the testing procedure, the maximum entropy classifier is used to predict the relationship between argument i and argument j as C_{ij}.

When the output of the maximum entropy classifier is around 0.5, it is not easy to figure out whether it is the first relationship or the second, we call this kind of information "uncertain information"(unclear relationship). For a better performance, we **strengthen** the certain information and weaken the uncertain information. We transform the result of maximum entropy classifier as follows:

$$C(i,j) = \begin{cases} 1 & 0.8 < MaxEnt(i,j) \leq 1.0 \\ 0 & 0.2 \leq MaxEnt(i,j) \leq 0.8 \quad (1) \\ -1 & 0.0 \leq MaxEnt(i,j) < 0.2 \end{cases}$$

We set two thresholds, if the output of the maximum entropy classifier is larger than 0.8, we set $C_{i,j} = 1$ (compatible arguments), if the output is lower than 0.2, we set $C_{i,j} = -1$ (incompatible arguments), otherwise, we set $C_{i,j} = 0$ (unclear relationship). The threshold 0.8 and 0.2 are tuned by development set.

4 Quadratic Optimization Method (QOM)

4.1 Post-processing Module of Bidirectional LSTM-RNN

Our quadratic optimization method is a post-processing module of bidirectional LSTM-RNN(Wang et al., 2015). The simplified architecture of bidirectional LSTM-RNN is shown as Figure 2.

Each dimension of the output vector $L_i \in \mathbb{R}^{n_L}, i = 1 \cdots n$ corresponds to the score of a certain semantic role label. n_L represents the number of semantic role labels. Then we normalize L_i over semantic roles as Eq 2 shows.

$$\widetilde{L}_i = Normalize(L_i) \quad (2)$$

Each dimension of $\widetilde{L}_i$ represents the probability of a certain semantic role label.

Let $P_{Arg} \in \mathbb{R}^n$ be a probability vector, each dimension of which represents the probability that the current word has a semantic role as is shown in Eq 3. $P_{Role} \in \mathbb{R}^n$ is another probability vector, each dimension represents the probability of the most likely semantic role the current word may be labeled as is shown in Eq 4.

$$P_{Arg}(i) = \sum_j \widetilde{L}_i(j)[label(j) \neq \text{'0'}] \quad (3)$$

$$P_{Role}(i) = \max_j \widetilde{L}_i(j) \quad (4)$$

where $[\cdot]$ equals to 1 if the inner statement is true and 0 otherwise. $label(j) \neq \text{'0'}$ means the j-th word is not labeled with semantic role.

4.2 Quadratic Optimization

We use a n-dim vector X to represent the identification result of candidate arguments. Each entry of X is 0 or 1, 0 represents "noArg", 1 represents "arg". X can be assigned by maximizing $E(X)$ as defined by Eq 5.

$$X = \underset{X}{\operatorname{argmax}} E(X)$$
$$E(X) = \lambda_1 X^T C X + \lambda_2 X^T P_{arg} \quad (5)$$
$$+ (1 - \lambda_1 - \lambda_2) X^T P_{role}$$

Here, $X^T C X$ means to add up all the relationship value if the two arguments are identified. Hence, the more the identified arguments are related, the larger the value $X^T C X$ is. $X^T P_{arg}$ is the sum of all chosen arguments probability. $X^T P_{role}$ is the sum of all the classified roles' probability.

Eq 5 means that, while we should select the semantic role with a larger probability, the argument relationship evaluation should also as large as possible.

2013

We use Beam Search method (Algorithm 1) to search for the optimal assignment X. The hyperparameters λ_1 and λ_2 can be chosen according to development set.

Input: Argument relationship matrix: C
    the argument probabilities required by P_{sum}^{arg}
    the role probabilities required by P_{sum}^{role}
Data: K: Beam size
    n: Number of candidate arguments
Output: The best assignment X
Set beam $\mathcal{B} \leftarrow [\epsilon]$;
for $i \leftarrow 1 \cdots n$ **do**
    buf$\leftarrow \{z' \circ l | z' \in \mathcal{B}, l \in \{0,1\}\}$;
    $\mathcal{B} \leftarrow [\epsilon]$;
    while $j \leftarrow 1 \cdots K$ **do**
        $x_{best} = \text{argmax}_{x \in \text{buf}} E(x)$;
        $\mathcal{B} \leftarrow \mathcal{B} \cup \{x_{best}\}$;
        buf$\leftarrow$buf$-\{x_{best}\}$;
    end
end
Sort $\mathcal{B}$ descendingly according to $E(X)$;
return $\mathcal{B}[0]$;

Algorithm 1: Beam Search decoding algorithm for SRL. ∘ means to concatenate an element to the end of a vector.

5 Experiment

We conduct experiments to compare our model with previous landmark methods on the benchmark dataset CPB for Chinese SRL. We use Wang et al. (2015)'s model as baseline. The result reveals that our quadratic optimization method can further improve the result of bidirectional LSTM-RNN.

5.1 Experiment Settings

We conduct experiments on the standard benchmark dataset CPB 1.0[2]. We follow the same data setting as previous work (Xue, 2008; Sun et al., 2009), which divided the dataset into three parts: 648 files (from `chtb_081.fid` to `chtb_899.fid`) are used as the training set. The development set includes 40 files, from `chtb_041.fid` to `chtb_080.fid`. The test set includes 72 files,

Method	$F_1(\%)$
Xue (2008)	71.90
Collobert and Weston (2008)	74.05
Sun et al. (2009)	74.12
Yang and Zong (2014)	75.31
Wang et al. (2015)	77.21
QOM - stengthen	76.24
QOM - feature 4,5,6	77.52
QOM	77.69

Table 1: Results comparison on CPB dataset.

which are `chtb_001.fid` to `chtb_040.fid`, and `chtb_900.fid` to `chtb_931.fid`.

The training dataset of the argument relationship matrix contains 1.6M cases (736K positive and 864K negative) which are randomly generated according to the ground truth in the training documents. We use Stanford Parser[3] for dependency parsing.

We tuned the coefficients λ_1 and λ_2 of Eq 5 on the development set, and finally we set $\lambda_1 = 0.10$ and $\lambda_2 = 0.45$.

5.2 Chinese SRL Performance

Table 1 shows our SRL performance compared to previous landmark results. We can see that with quadratic optimization method as the postprocessing module, our approach (QOM) outperforms Wang et al. (2015) by a large margin (Wilcoxon Signed Rank Test, $p < 0.05$). We also did some ablation test, in Table 1, "QOM - stengthen" is the result when we do not strengthen the argument relationship matrix. We can see that the uncertain information is very harmful to the performance, which worsen the accuracy for about 1%. "QOM - feature 4,5,6" is the performance when we do not use the dependency features when capturing the argument relationships since Wang et al. (2015) didn't use any dependency feature. We can see that event without dependency feature, our method still can outperform Wang et al. (2015)'s result.

Figure 3 visualizes the candidate argument relationship matrix. From this graph, we captured the compatible arguments ("外商$_{\text{foreign businessman}}$" and "企业$_{\text{entrepreneur}}$"), incompatible arguments ("项$_{\text{item}}$" and "外商$_{\text{foreign businessman}}$"), ("规定$_{\text{rule}}$"
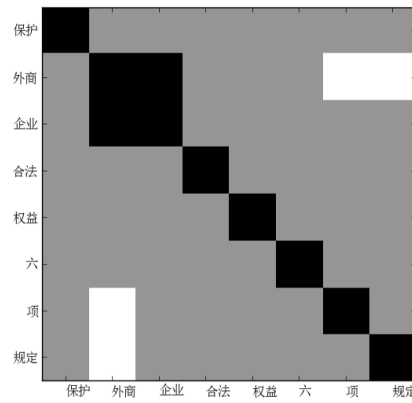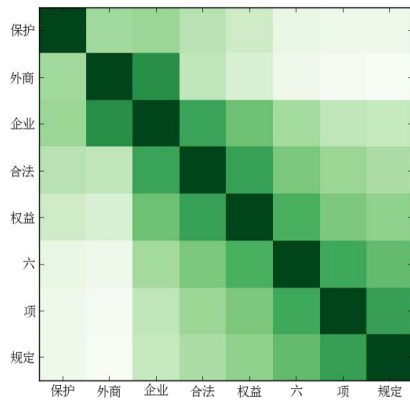
Chinese	保护	外商	企业	合法	权益	六	项	规定
English translation	protect	foreign businessman	entrepreneur	legal	profit	six	item	rule

Figure 3: The Visualization of argument relationship Matrix, Left is the origin matrix. Right is the strengthened matrix. In the origin matrix, we can directly see the argument relationship we captured (the darker green means stronger relationship, lighter green means weaker relationship). After strengthening, on the right, the words with strong relationship are classified as compatible arguments (the black squares), weak relationship are classified as incompatible arguments (the white squares). Others (the grey squares) are unclear relationship.

and "外商$_{\text{foreign businessman}}$"). Therefore, "外商$_{\text{foreign businessman}}$" and "企业$_{\text{entrepreneur}}$" should be the roles of "投资$_{\text{invest}}$" simultaneously , ("项$_{\text{item}}$", "外商$_{\text{foreign businessman}}$") and ("规定$_{\text{rule}}$", "外商$_{\text{foreign businessman}}$") should not be the roles of "投资$_{\text{invest}}$" simultaneously.

6 Conclusion

In this paper, we propose to use a quadratic optimization method based on two kinds of argument relationships to improve the performance of Chinese SRL. We first train a maximum entropy classifier to capture the compatible arguments and incompatible arguments. Then we use quadratic optimization to improve the result of bidirectional LSTM-RNN(Wang et al., 2015). The experiment has proved the effectiveness of our approach. This method can also be used in other probabilistic methods.

Acknowledgements

References

Wenliang Chen, Yujie Zhang, and Hitoshi Isahara. 2006. An empirical study of chinese chunking. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 97–104. Association for Computational Linguistics.

Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM.

Weiwei Ding and Baobao Chang. 2008. Improving chinese semantic role classification with hierarchical feature selection strategy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 324–333. Association for Computational Linguistics.

Weiwei Ding and Baobao Chang. 2009. Word based chinese semantic role labeling with semantic chunking. *International Journal of Computer Processing Of Languages*, 22(02n03):133–154.

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

Lei Sha, Jing Liu, Chin-Yew Lin, Sujian Li, Baobao Chang, and Zhifang Sui. 2016. Rbpb: Regularization-based pattern balancing method for event extraction.

In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1224–1234, Berlin, Germany, August. Association for Computational Linguistics.

Honglin Sun and Daniel Jurafsky. 2004. Shallow semantic parsing of chinese. In *Proceedings of NAACL-HLT*, volume 2004.

Weiwei Sun, Zhifang Sui, Meng Wang, and Xin Wang. 2009. Chinese semantic role labeling with shallow parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1475–1483. Association for Computational Linguistics.

Weiwei Sun. 2010. Improving chinese semantic role labeling with rich syntactic features. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 168–172. Association for Computational Linguistics.

Zhen Wang, Tingsong Jiang, Baobao Chang, and Zhifang Sui. 2015. Chinese semantic role labeling with bidirectional recurrent neural networks. In *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1626–1631.

Nianwen Xue and Martha Palmer. 2003. Annotating the propositions in the penn chinese treebank. In *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*, pages 47–54. Association for Computational Linguistics.

Nianwen Xue and Martha Palmer. 2005. Automatic semantic role labeling for chinese verbs. In *IJCAI*, volume 5, pages 1160–1165. Citeseer.

Nianwen Xue. 2008. Labeling chinese predicates with semantic roles. *Computational linguistics*, 34(2):225–255.

Haitong Yang and Chengqing Zong. 2014. Multi-predicate semantic role labeling. In *EMNLP*, pages 363–373.