Improving Chinese Semantic Role Labeling with English Proposition Bank

Tianshi Li, Qi Li, and BaoBao Chang

Key Laboratory of Computational Linguistics, Ministry of Education, School of Electronics Engineering and Computer Science, Peking University,Beijing 100871, China
Collaborative Innovation Center for Language Ability, Xuzhou 221009, China
lts_417@hotmail.com
{qi.li,chbb}@pku.edu.cn

Abstract. Most researches to SRL focus on English. It is still a challenge to improve the SRL performance of other language. In this paper, we introduce a two-pass approach to do Chinese SRL with a Recurrent Neural Network (RNN) model. We use English Proposition Bank (EPB) to improve the performance of Chinese SRL. Experimental result shows a significant improvement over the state-of-the-art methods on Chinese Proposition Bank (CPB), which reaches 78.39% F1 score.

Keywords: Chinese semantic role labeling, two-pass approach, Recurrent Neural Network, English resource

1 Introduction

Semantic Role Labeling (SRL) aims to recognize the arguments for a given predicate and assign semantic roles to them. Figure 1 shows an example of SRL. *John* is the agent of the predicate *married* denoting that he is the man who got married, *his neighbor* is the patient of *married* denoting who John married. Both of the agent and patient are the sematic roles of *married*.

SRL can be formalized as a sequence labeling task, and we use IOBES tagging schema to tag the semantic roles. According to this tagging schema, argument identification consists of tagging all tokens of a sentence with IOBES tags (Inside, Outside, Begin, End, Single) relative to a given predicate. Figure 1 shows an example.

Most SRL approaches use supervised learning model and thus heavily rely on semantically annotated corpora. For the Chinese dataset English Proposition Bank (CPB) [1], it contains over 80,000 verb instances for 11,000 verb types. However, it is still not enough using only CPB to solve the whole Chinese SRL task. For the English standard benchmark dataset in English Proposition Bank (EPB) [2], it contains nearly 100 thousand annotated sentences. Given that manually annotating SRL corpus is labor-intensive and expensive, how to improve the monolingual SRL performance with merging different language resources is thus an important issue deserving to explore.

In this paper, we propose a simple but effective two-pass training approach based on recurrent neural network (RNN) with bidirectional long-short-term memory (LSTM), aiming at improving the Chinese SRL performance using English semantic role labeled

Chinese	约翰	娶	他的	邻居
English	John	married	his	neighbor
Role	(Agent)	Predicate	(Patient	)
	(ARG0)	REL	(ARG1	)
IOBES	S-ARG0	REL	B-ARG1	E-ARG1

Fig. 1. An Chinese sentence with semantic labels. REL denotes the given predicate.

Fig. 2. A Chinese-English parallel sentence pair with semantic labels and word alignment.

corpus. The main points are as follows: By representation learning, Chinese SRL corpus and English SRL corpus are mapped into an uniform semantic representation space. This makes it possible to merge the corpora of two languages and train a single SRL model across languages. On the basis of the cross language SRL model, we further train the SRL model specific to Chinese. Our approach requires neither parallel SRL corpus nor machine translation of the corpus. Experiments show that our approach outperforms current state-of-art systems for Chinese SRL task.

2 Related Work

SRL task was firstly proposed in the work of Jurafsky et al. (2002) [3] and a large body of work has been devoted to this task since then. Traditional SRL approaches normally use a lot of handcrafted features. Koomen et al. (2005) [8] get the best performance among all traditional approaches on English SRL task, they used different parse tree information with lots of traditional features. Most Chinese SRL work adopted similar strategies, although using a much smaller training corpus. Xue & Palmer (2005) [10] and Xue (2008) [11] stands for first through and systematic Chinese SRL research. Sun et al. (2009) [12] performed Chinese SRL with shallow parsing, which took partial parses as inputs. Yang and Zong (2014) [13] proposed multipredicate SRL, which showed improvements both on English and Chinese Proposition Bank. Recently, to reduce the heavy burden of feature engineering, deep learning models like CNNs and RNNs have been introduced into SRL task. Collobert and Weston (2008) [9] proposed a Convolutional Neural Network (CNN) on English. For Chinese SRL, Wang (2015) [14] used bidirectional LSTM and outperformed previous traditional models. Different from pre-

Chinese Roles	Common	Non-common
Type Numbers	11	7
Total Numbers	79748	1335
Example	ARG0	ARGM-CND
Meaning	agent	condition

Table 1. Statistics for Chinese common roles and non-common roles on CPB.

vious work, we focus on how to use English semantic role labeled corpora to improve the performance of Chinese SRL.

3 Two-pass Training Approach

3.1 Basic Idea

Generally, semantics is believed to be more language general than syntax. Especially in SRL corpus, many sematic roles are same or have similar meanings across languages. In this paper, we focus on Chinese and English SRL corpus. In Figure 2, although words and word order are different between Chinese and English, the semantic roles are the same. We call these language independent roles as common semantic roles, such as ARG0, and call the roles which only appear in the Chinese SRL corpus as non-common semantic roles, such as ARGM-CND[1]. Table 1 shows that over 60% types of Chinese roles can be found in English and the total number of common roles accounts for 98.35% on the whole CPB. Similar result applys for the English side, the total number of common roles account for 88.54% on whole EPB. Intuitively, adding EPB to train set is helpful for improving the performance of Chinese common role labeling.

Although the semantic similarity between Chinese and English, there is still a gap between the linguistic structure and syntax of Chinese and English corpus. To cross this gap, we propose to project these corpus into a same vector space by means of bilingual embedding. Progress in bilingual representation learning [4, 7] shows that words in different languages can be projected into the same vector space as distributed vector embeddings. Moreover, these word embeddings are shown to have the ability of capturing semantic coherence across languages [5, 6]. With bilingual embedding, words with similar meanings in different languages are projected into close position in the shared vector space.

With these considerations, we propose a two-pass training approach as follows: First, we merge and randomly shuffle CPB and EPB, keep the common semantic roles and remove the non-common semantic roles. We use bilingual parallel corpus to learn bilingual word embedding. Using these merged corpus and bilingual embedding, we train an RNN model for all common semantic roles. We ignore the Chinese non-common roles in first pass because these roles are rare and language-specific. Second, for CPB only, we learn both Chinese common roles and Chinese non-common roles together, using the same RNN model in the first pass. We utilize the parameters (including the

[1] Conditional clause is seen as a semantic role in the Chinese but not in the English.

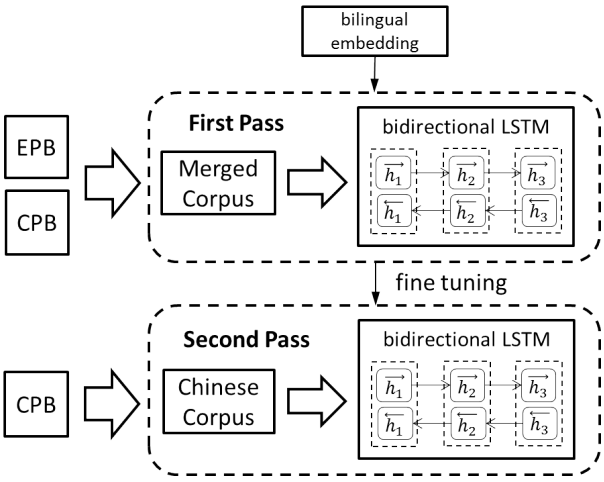

Fig. 3. The two-pass training model.

bilingual word embedding) we get in the first pass as the initialization of the RNN model in second pass. Our approach can be illustrated in Figure 3.

3.2 Bilingual Word Representation

In our approach, we utilize bilingual compositional vector model (BiCVM) [6] to learn our bilingual word embedding. BiCVM learns to assign similar embeddings to aligned sentences and dissimilar ones to sentence which are not aligned while not requiring word alignments. We chose BiCVM for the reason that SRL is a task based on the whole sentence and BiCVM can catch more semantic knowledge on sentence level across aligned parallel sentences between two languages.

3.3 Basic SRL Model

Given a sentence s, we first compute its representation sequence $\boldsymbol{z}$. Here $z_t = \sigma(W a_t)$ denotes the representation of the t-th word in s. a_t is the feature embedding of current word t which concatenates bilingual embeddings(word t, word $t-1$, word $t+1$ and the predicate), POS tag embeddings (word t, word $t-1$, word $t+1$) and distant feature(the distant from word t to the predicate). $W \in \mathbb{R}^{n_1 \times n_0}$, n_0 is the length of a_t, σ is the sigmoid function.

Then we use the bidirectional LSTM, the two LSTMs process the input sequence $\boldsymbol{z}$ from both forward and backward directions. We use the bidirectional LSTM because it is a sequence labeling model which can easily catch semantic information and works well in monolingual SRL task (Wang ,2015). We can compute LSTM layer at each word t as follows:

$$\widetilde{C}_t = \tanh(W_c z_t + U_c h_{t-1} + b_c) \tag{1}$$

$$g_j = \sigma(W_j z_t + U_j h_{t-1} + b_j) \tag{2}$$

$$C_t = g_i \odot \widetilde{C}_t + g_f \odot C_{t-1} \tag{3}$$

$$h_t = g_o \odot C_t \tag{4}$$

Where C_t is the memory cell of position t, $\widetilde{C}_t$ computes the candidate value for C_t, h_t is the output state of position t, $j \in \{i, f, o\}$, $g_i \backslash g_f \backslash g_o$ is input\forget\output gate of LSTM. $W_c, W_j \in \mathbb{R}^{n_2 \times n_1}$, $U_c, U_j \in \mathbb{R}^{n_2 \times n_2}$. $\odot$ indicates elementwise vector multiplication. For the t-th word, we get both a forward hidden state $\overrightarrow{h_t}$ and a backward hidden state $\overleftarrow{h_t}$ from the bidirectional LSTM. These hidden states are then concatenated together into a merged hidden state $hid_t = [\overrightarrow{h_t}^{\mathrm{T}}; \overleftarrow{h_t}^{\mathrm{T}}]^{\mathrm{T}}$.

At last, we put hid_t into a softmax layer to generate final output, each dimension of output corresponds to the score of a certain semantic role label in IOBES schema.

3.4 Training Criteria

Given training examples:

$$T = (x^{(i)}, y^{(i)}) \tag{5}$$

where $x^{(i)}$ denotes the i-th training sentence, $y^{(i)}$ is the correct sequence labels of $x^{(i)}$. $y_t^{(i)} = k$ means the t-th word has the k-th semantic role label in IOBES scheme. We can define the score of i-th sentence as follows:

$$s(x^{(i)}, y^{(i)}, \theta) = \sum_{t=1}^{N_i} o_{ty_t^{(i)}} \tag{6}$$

where N_i is the length of the i-th sentence, $o_{ty_t^{(i)}}$ is the value of the correct label in the output layer for the t-th word in the i-th sentence, θ is an ensemble of all the parameters in the whole network.

We use maximum log likelihood method to training all examples. For a single example, the log likelihood is:

$$\begin{aligned} \log p(y^{(i)}|x^{(i)}, \theta) &= \log \frac{\exp(s(x^{(i)}, y^{(i)}, \theta))}{\sum_{y'} \exp(s(x^{(i)}, y', \theta))} \\ &= s(x^{(i)}, y^{(i)}, \theta) - \log \sum_{y'} \exp(s(x^{(i)}, y', \theta)) \end{aligned} \tag{7}$$

where y' ranges from all the valid paths of tags.

The full log likelihood of the whole training corpus is as follows:

$$J_{MLE}(\theta) = \sum_i \log p(y^{(i)}|x^{(i)}, \theta) \tag{8}$$

We use stochastic gradient ascent in the experiments.

4 Experiments

4.1 Experiment Settings

To comparison with previous work, we conduct experiments on the standard benchmark dataset CPB, follow the same data setting as previous work [11, 14]. For English dataset, we use the training set of CoNLL-2005 dataset(based on EPB), the same data setting as Li and Chang (2015) [15]. For training the bilingual word embedding, we use PKU bilingual corpus[2].

Method	**F1(%)**
Xue (2008)	71.90
Yang and Zong (2014)	75.31
Sun et al. (2010)	76.46
Wang (2015)	77.59
Our approaches	
random (one-pass)	76.29
BiCVM (one-pass)	76.97
BiCVM+EPB (two-pass)	**78.39**

Table 2. Results comparison on CPB test set.

For the SRL model, the parameters are set as follows: the dimension of bilingual word embedding is 50; the dimension of POS tag embedding is 20; the dimension of distant feature embedding is 20; n_1 is 200; n_2 is 100; the number of bidirectional LSTM layer is 1; the learning rate in both first pass and second pass is 10^{-3}; the hyper-parameter λ in the objective function is 10^{-3}; Using early stop strategy to get the best result on development set, the training epochs in the first pass is set to 12, the training epochs in the second pass is set to 6.

4.2 SRL Results

Table 2 shows the Chinese SRL results on CPB. *one-pass* denotes training a bidirectional LSTM model on CPB for both Chinese common roles and non-common roles, *two-pass* denotes our two-pass approach described in Section 3.1. *random* denotes that the word embedding is randomized initialized, *BiCVM* in Table 2 denotes that we use the bilingual word embedding described in Section 3.2 instead of randomized initialization. We don't do experiment with monolingual word embedding because it is beyond our focus in this paper. *EPB* in Table 2 denotes that we use EPB in the first pass training during the two-pass approach.

Compared with randomized initialization of word embedding in one-pass approach, using bilingual embedding has a slight improvement. While compared with one-pass

[2] PKU bilingual corpus is developed by Peking University, it is a English-Chinese parallel corpus. It contains 807,500 aligned English-Chinese sentence pairs and is available by licensing.

training approach, our two-pass approach improves a lot and establishes a new state-of-the-art result in Chinese SRL with 78.39.

Figure 4 shows the effectiveness of EPB for both Chinese common roles and non-common roles. From Figure 4, EPB is helpful for most Chinese roles, especially none of common role's performance decreases after adding EPB. This proves that our strategy of using EPB is successful, our approach exactly capture the common semantic role information between Chinese and English. The performance on Chinese non-common roles is inconsistent, because these roles are language-specific and EPB can't definitely improve the performance of these roles.

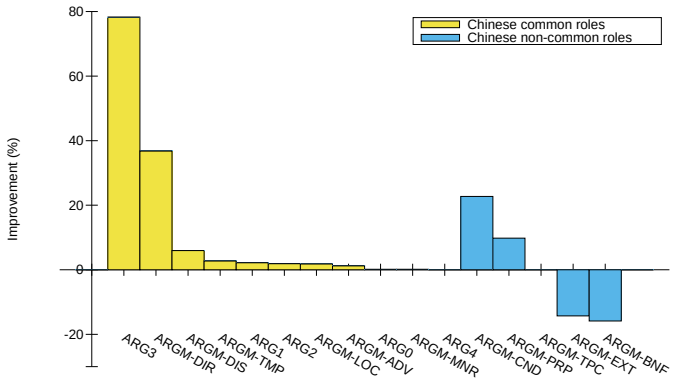

Fig. 4. Improvement of all Chinese semantic roles on CPB test set for two-pass approach compared with one-pass approach (randomized initialization). The 11 yellow columns denote to Chinese common roles in test set, the 5 blue columns denote to Chinese non-common roles in test set.

4.3 Translation Equivalent Regularizer

Furthermore, we try to make the embedding of the words, which has translation equivalent relation between Chinese and English, more closer in vector space in the first pass training. For each word t (either Chinese or English) in merging corpus, we define a translation equivalence regularizer which equals to $||x_t - \sum_j a_{tj} x_j||$. Here, x_t is the embedding of the word t, x_j is the embedding of the word j which has translation equivalent with word t, a_{tj} is the translation probability[3] from word t to word j. However, after adding the regularizer, the result gets 78.31 on CPB test set, doesn't outperform the best result we gets before. This is possibly because translation equivalence regularizer leads to overfitting in the first pass training.

[3] We use GIZA++ to get the translation probability from PKU bilingual corpus, GIZA++ can be download in http://code.google.com/p/giza-pp/downloads/list

5 Conclusion

In this paper, we introduce a two-pass approach with bidirectional LSTM, using EPB to improve the performance on Chinese SRL. Our approach doesn't need any parallel annotated SRL corpus, heavy job of feature engineering. And our approach can apply to other languages. Our approach achieves the state-of-the-art results on the Chinese SRL task. In future work, we plan to project different language sentences into same semantic space in a better way.

Acknowledgments. This work is supported by National Key Basic Research Program of China (2014CB340504) and National Natural Science Foundation of China (61273318).

References

1. Xue, Nianwen, and Martha Palmer.: Adding semantic roles to the Chinese Treebank. Natural Language Engineering, pp. 143–172. (2009)
2. Palmer, Martha, Daniel Gildea, and Paul Kingsbury.: The proposition bank: An annotated corpus of semantic roles. Computational Linguistics Journal. 71–106. (2005)
3. Daniel Gildea and Daniel Jurafsky.: Automatic Labeling of Semantic Roles. Computational Linguistics. 245–288 (2002)
4. Alexandre Klementiev, Ivan Titov and Binod Bhattarai.: Inducing Crosslingual Distributed Representations of Words. In: COLING. (2012)
5. Will Y. Zou, Richard Socher, Daniel Cer and Christopher D. Manning.: Bilingual Word Embeddings for Phrase-Based Machine Translation. In: the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1393–1398. (2013)
6. Karl Moritz Hermann and Phil Blunsom.: Multilingual models for compositional distributed semantics. In: ACL, pp. 58–68. (2014)
7. Sarath Chandar, Stanislas Lauly, Hugo Larochelle, Mitesh M. Khapra, Balaraman Ravindran, Vikas C. Raykar and Amrita Saha.: An autoencoder approach to learning bilingual word representations. In: NIPS, pp. 1853–1861. (2014)
8. Koomen P, Punyakanok V, Roth D and Yih W T: Generalized inference with multiple semantic role labeling systems. In: the Ninth Conference on Computational Natural Language Learning, pp. 181–184. Association for Computational Linguistics (2005)
9. Ronan Collobert and Jason Weston.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: the 25th international conference on Machine learning, pp. 160–167. ACM (2008)
10. Nianwen Xue and Martha Palmer.: Automatic semantic role labeling for chinese verbs. In: IJCAI, volume 5, pp. 1160–1165. Citeseer (2005)
11. Nianwen Xue.: Labeling chinese predicates with semantic roles. Computational linguistics. 225–255 (2008)
12. Weiwei Sun.: Semantics-Driven Shallow Parsing for Chinese Semantic Role Labeling. In: ACL, pp. 103–108. (2010)
13. Haitong Yang and Chengqing Zong.: Multipredicate semantic role labeling. In: the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 363–373. (2014)

14. Zhen Wang, Tingsong Jiang, Baobao Chang, Zhifang Sui.: Chinese semantic role labeling with bidirectional recurrent neural networks. In: the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1626–1631. (2015)
15. Tianshi Li and Baobao Chang.: Semantic Role Labeling Using Recursive Neural Network. In: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, pp. 66–76. (2015)