TSABCNN: Two-Stage Attention-Based Convolutional Neural Network for Frame Identification

Hongyan Zhao[1,2], Ru Li[1,3], Fei Duan[1], Zepeng Wu[1] and Shaoru Guo[1]

[1]School of Computer and Information Technology,
Shanxi University, Taiyuan, 030006, China
hongyanzhao@tyust.edu.cn, liru@sxu.edu.cn, fduan@sxu.edu.cn, 289977640@qq.com, guo-shaoru0928@163.com
[2]School of Computer Science & technology,
Taiyuan University of Science and Technology, Taiyuan, 030024, China;
[3]Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing, Shanxi University, Taiyuan, 030006, China

Abstract. As an essential sub-task of frame-semantic parsing, Frame Identification (FI) is a fundamentally important research topic in shallow semantic parsing. However, most existing work is based on sophisticated, hand-crafted features which might not be compatible with FI procedure. Besides that, they usually heavily rely on available natural language processing (NLP) toolkits and various lexical resources. Thus existing methods with hand-crafted features may not achieve satisfactory performance. In this paper, we propose a two-stage attention-based convolutional neural network (TSABCNN) to alleviate this problem and capture the most important context features for FI task. In order to dynamically adjust the weight of each feature, we build two levels of attention over instances at input layer and pooling layer respectively. Furthermore, the proposed model is an end-to-end learning framework which does not need any complicated NLP toolkits and feature engineering, and can be applied to any language. Experiments results on FrameNet and Chinese FrameNet (CFN) show the effectiveness of the proposed approach for the FI task.

Keywords: Frame Identification, FrameNet, Convolutional Neural Network.

1 Introduction

As the core task of Natural Language Processing (NLP), shallow semantic parsing abandons the complexity of deep components and relationships and has attracted great attention. In recent years, more semantic knowledge bases such as WordNet, PropBank, and HowNet have been built and widely used in the shallow semantic parsing task. Among these semantic knowledge engineering projects, FrameNet (Baker et al.,1998) is a rich linguistic resource containing considerable expert knowledge about lexical and predicate-argument semantics, and frame-semantic parsing has been proven to be an effective way that extracts a shallow semantic structure from text.

According to the theory of frame semantics (Fillmore,1982), one semantic frame represents an event or scenario, and possesses a set of targets (namely lexical units or predicate) that can evoke the semantic scenario and some frame elements (or semantic roles) that participate in the event (Hermann et al., 2015). Most work on frame-semantic parsing (Das et al., 2010; Das et al., 2014) has divided the task into two subtasks: (1) the first one is frame identification, which identifies the most suitable semantic frame for a given target in a sentence; (2) the second one is argument identification (or semantic role labeling), which performs semantic role labeling for the identified frame. However, current researches on frame-semantic parsing mostly focus on argument identification for given target and its frame (Carreas et al.,2008), skipping the frame identification step, which leads to the failure to automatically implement frame-semantic analysis task. This is also the main reason why frame-semantic parsing can't be widely used in many NLP tasks. We argue that the first subtask is an essential step in the frame-semantic parsing task. In this paper, we focus on the FI for given targets.

At present, FI task is treated as a multi classification task, and virtually all of the state-of-the-art approaches for this task are based on sophisticated, hand-crafted features, such as conditional random fields (CRF), support vector machine (SVM), maximum entropy (ME). In addition, extracting these features usually heavily rely on available NLP toolkits and various lexical resources, which might lead to the error propagation. Thus pre-existing methods with hand-crafted features may not achieve satisfactory performance.

In order to reduce the manual labor in feature extraction, recently, deep learning is used to learn features in many NLP tasks, and convolutional neural networks (CNN) have shown to be efficient to capture syntactic and semantic context features between words within a sentence for NLP tasks such as sentence modeling (Kalchbrenner et al., 2014), sentence classification (Kim, 2016) and relation classification (Zeng et al., 2014), event extraction(Chen et al., 2015), question answering (Li et al., 2015), short text ranking (Severyn et al., 2015), text chunks matching (Yin and Schütze, 2015). FI can also be considered as a sentence-classification task for a marked target. In this paper, we propose a novel attention-based conventional neural networks model for frame identification.

FI can also be considered as a sentence-classification task for a marked target . In this paper, we propose a two-stage attention-based convolutional neural network (TSABCNN) to alleviate this problem and capture the most important context features for FI task. In order to dynamically adjust the weight of each feature, we build two levels of attention mechanism over instances at input layer and pooling layer respectively. Furthermore, the proposed model is an end-to-end learning framework which does not need any complicated NLP toolkits and feature engineering, and can be applied to any language. Experiments results on FrameNet and Chinese FrameNet (CFN) show the effectiveness of the proposed approach for the FI task.

Our main contributions are: (1) we analyze the problem of exiting models on the task of FI, and propose an end-to-end FI method based on CNN, which does not need any complicated NLP toolkits and feature engineering; (2) In order to make the weight of important features bigger, we introduce a supervised attention based FI model on input layer and pooling layer respectively. (3) we improve the performance of FI and achieve better performance than the baselines.

2 Related Work

Since Gildea and Jurafsky (2002) pioneered semantic role labeling(SRL), particularly followed by the CONLL2004 (Carreras, 2004) and CONLL2005 (Carreras and Màrquez, 2005) treat SRL as a shared evaluation task, frame-semantic analysis has received a boost in attention. The FrameNet lexicon contains abundant linguistic information about lexical items and predicate-argument structures. In a frame-analyzed sentence, predicates evoking frames are known as targets, and a word or phrase filling a role is known as an argument. Figure 1 shows frame-semantic annotations for a sentence. In this figure, the target buy.V evokes the Commerce_buy frame. Buyer and Goods are some arguments (semantic roles) for this frame.

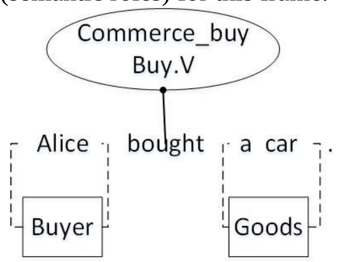

Figure 1. frame-semantic annotations for a sentence

Most early work on frame-semantic parsing used a supervised machine learning method. Fleischman, Kwon and Hovy (2003) used maximum entropy models to identify arguments and their roles for a given frame. Erk et. al. (2006) used the traditional word disambiguation method to conduct experiment on German FrameNet frame disambiguation. The LTH system presented by Johansson and Nugues (2007) achieved the best performance in the SemEval 2007 task of identifying frame. They adopted a series of SVMs to classify the frame for a given target, associating unseen lexical items to frames and identifying and classifying a word or phrase as various semantic roles. Adrian Bejan and Hathaway (2007) selected 556 ambiguous target words which can evoke two or more semantic frames and have more than five annotated sentences for each frame. They trained a multi-classifier for ambiguous targets from the FrameNet lexicon.

Recently, a tool called SEMAFOR was presented (Das et al., 2010), with a probabilistic models for FI that used a latent-variable log-linear model to capture frames for unseen targets. The feature set of this model leads to better performance on the SemEval 2007 data. The FrameNet project released a new version of annotating data in 2010. Das et al., using a two-stage statistical model (Das et al., 2014) on this dataset, improved their prior work and set the new state of the art. A few salient aspects of this updated version of SEMAFOR involved handling unseen targets by using a graph-based semi-supervised learning approach and a dual decomposition algorithm. Subsequently, Hermann et al. (2014) presented a novel model using distributed representations of the word context and dependency path for better FI, outperforming the aforementioned SEMAFOR.

Unfortunately, above work is mostly based on sophisticated, hand-crafted features. Besides that, they usually heavily rely on available natural language processing (NLP) toolkits and various lexical resources. Thus existing methods with hand-crafted features may not achieve satisfactory performance. Sample Heading (Third Level).Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

3 TSABCNN Model

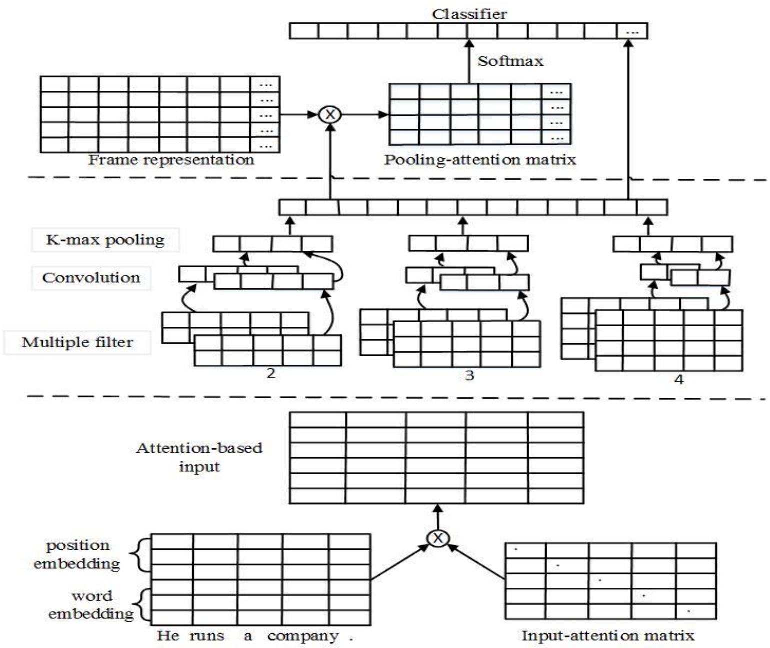

Figure 2. The architecture of TSABCNN .

The aim of FI is to choose a correct frame for the given target in a sentence. Because targets are usually ambiguous, a target may arouse multiple frames, which poses challenges for FI tasks. Traditional statistical machine learning method is based on sophisticated, hand-crafted features. In recent years, CNN has been proved to be able to learn more advanced contextual features, and has been applied to many NLP tasks(Chen et al., 2015; Li et al., 2015; Severyn et al., 2015; Yin and Schütze, 2015; Wenpeng Yin et.al., 2016). Inspired by Wenpeng Yin, we here propose a novel general two-stage attention-based convolution neural network model for FI. Our network architecture is shown as Figure 2. The input sentence is transformed into a vector by looking up word embedding. To acquire the word order, we here combine word embedding with relative position vector between the word and target. In input layer, an attention mechanism is used to acquire relevance of words and the target. In order to obtain context information such as n-gram, we use multiple kinds of filters with different region size (Zhang and Wallace, 2016), and next is a maximum pool layer. Another attention is used to determine more important n-gram features for FI. Finally, a $softmax$ classifier is used, and the highest scoring frame is regarded as the best frame. Our model will be further described in the remainder of this section.

3.1 Word Representation

Before entering the network, each word w_i is mapped into a real vector $v_i \epsilon R^d$ (d denote the word embedding dimension) by looking up the word embedding table V, which can be trained by word2vec (Mikolov et al., 2013) model. In addition, in order to embed the position information of a word in a sentence for FI, we introduce the relative distance between $i-th$ word and the target marked by a word position embedding P (Collobert et al.,2011). For example, given sentence in Figure 2, the relative distances of "He" and "company" to "runs" are -1 and 2 respectively. Every relative distance is randomly initialized a position vector p_i, and the dimensionality of the word position vector is q. Finally, combining the word embeddings v_i and the relative position embedding p_i, the word feature(WF) is represented as w_i^f. Thus, a given sentence or the sequence of n words can be encoded into a matrix as follows:

$$\boldsymbol{S} = [\boldsymbol{w_1^f}, \boldsymbol{w_2^f}, \ldots, \boldsymbol{w_n^f}] \boldsymbol{w_i^f} \in \boldsymbol{R^{(d+q)}} \tag{1}$$

3.2 Input-attention Mechanism

Attentive neural networks have been successfully applied to natural language process (NLP) tasks such as machine translations, question answering, relation extraction, and sentence pairs modeling (Bahdanau et al., 2014; Hermann et al., 2015; Zhou et al., 2016; Yin et al., 2016). In this subsection, we propose an input-attention mechanism for FI. Previous work has focused on the alignment of the input and output sequences, e.g. the alignment of input language and target language in machine translation. To make our model automatically learn the more useful features for the FI task, we propose a novel idea of applying the two-stage attention mechanism to heterogeneous objects, respectively input-attention and pooling-attention. The attention matrix on the input-layer intends to give higher weights to those words related to the target, and guide the convolution layer to learn more useful and higher level features. As is shown in Figure 2, the word "company" is very significant to FI, then the higher weight is allocated to it. Here, we define the input-layer Attention as a diagonal matrix A. It denote that the relevance of word w_i and target t in the given sentence. Formally, we define the diagonal attention matrix A as :

$$A_{ii} = f(w_i, t) \tag{2}$$

The function f can be computed by different ways. We here exploit $\frac{1}{0.0001+|w_i-t|}$ to initialize the matrix, where $|w_i - t|$ is Euclidean Distance of between $i-th$ word in a sentence and target word t, and A_{ii} is updated during thenetwork training process. The relevance degree of $i-th$ word in the sentence and its target word is defined as:

$$a_i = \frac{exp\,(A_{ii})}{\sum_{k=1}^{n} exp\,(A_{kk})} \tag{3}$$

with n being the length of the sentence. The attention-based input is represented as follows:

$$X = [a_1 w_1, a_2 w_2, \dots, a_n w_n] \tag{4}$$

subsequently, X is fed into convolution layer of the model.

3.3 CNN Architecture with multiple-size filters

FI is a complicated task. In prior work, a variety of context features (e.g. n-gram features) were extracted to solve this problem. Apparently, it is difficult to obtain these information only by word features. In order to enable the model to learn advanced features, following Collobert and Weston (2008), we regard the sentence matrix as a special matrix, and perform a convolution operation on it by multi-filters of different sizes. As columns represent discrete units (namely word or phrase) and a sentence has its inherent sequential structure, it is meaningful to use filters with the same height as the WF dimension. Therefore, we only choose the width of filters, namely adjusting the number of words jointed. In the subsequent section of the paper, the region size of the filter refers only to its width.

Given a filer with a region size m, it is a weight matrix W_f including $m \cdot k$ (k is *the* dimension of a word feature) parameters to be trained. The attention-based input matrix of the sentence X is fed into convolution layer. We use a sub-matrix $X[:i,:j]$ to represent the mapping from sentence $i\text{-}th$ column to sentence $j\text{-}th$ column. Input matrix X is folded by the above mentioned filters, and at the same time wide convolution method is used, and phrase-level features are generated. More formally, a feature map h_i is generated from the window of words $X_{[:i,:i+m-1]}$ as follows:

$$h_i = \sigma\left(W_f^T X_{[:i,:i+m-1]} + b_f\right) \tag{5}$$

where b_f is the bias of the convolution layer.

Through convolution layer, the output length is $n_{out} = n_{in} + m - 1$ (n_{in} is the input length of the convolution).

3.4 *K*-max-pooling and pooling-zttention

The higher-level phrase features are generated by the filter windows on convolution layer. Some features are very important features for the target task, but others are not relevant to the task. To assign greater weight to those important features, we hereby propose a novel attention-based pooling method to extract some crucial features for the FI task. We firstly use K-max pooling on every feature map, which helps to obtain a fixed size matrix, reduces dimensionality, and keeps important features and global information about position. Combined with all K-max features, the context of the target *t* is represented as:

$$S_p = [c_1^t, c_2^t, c_3^t, \dots, c_l^t] \tag{6}$$

with c_i^t being the i-**th** K-max feature vector, and l being the number of filters. Secondly, we use an pooling-attention strategy to determine the importance of the context features of target that encoded by convolutional kernel and $K-$ max pooling. We create a correlation matrix C that obtains relative connections between the context feature of target and frames embedding W_F,$C = S_p U W_F$, where U is a weight matrix to be learned by the network, and the frame embedding is represented by the mean of the word embedding of the frame name. Then we normalize the matrix C, and obtain pooling-attention matrix A^p as:

$$A_{i,j}^p = \frac{C_{i,j} - minvalue}{max\ value - minvalue} \tag{7}$$

Finally, to highlight important context features, we multiply this pooling-attention matrix with S_p to get S_o. It is denoted as follows:

$$S_o = A^p \cdot S_p \tag{8}$$

3.5 Regularization and classification

In order to overcome over fitting, following the works of Hinton et al. (2012) and Kim (2014), we execute a dropout regularization for S_o and produce the dropout vector S_d. What should be noted is that dropout is only performed during the training phase. The dropout vector S_d is input into a standard neural network by fully connected method, which use a weight matrix W_c as model parameters. Finally, a softmax layer is used to implement the classification and output the vector o. The vector o of i-th dimension represents the probability of the i-th frame classified, computed as follows:

$$p(f_i \vee x, \theta) = \frac{e^{o^i}}{\sum_{j=1}^{F} e^{o^j}} \tag{9}$$

Where θ is a set of all network parameter to be learned, and F is the frame number.

3.6 The Network Training

For all training examples $\left(x^{(i)}, y^{(i)}\right)$, we create the log-likelihood about the parameter θ as the objective function. It is denoted as follows:

$$L(\theta) = \sum_{i=1}^{N} l\, ogp\left(y^{(i)} \vee x^{(i)}, \theta\right) \tag{10}$$

With N is the number of all training examples, and we train θ by maximizing $L(\theta)$.

4 Experiments

In this section, we firstly introduce the datasets and evaluation metrics, experimental setting, and then we present our experiments and results obtained in FI.

4.1. Datasets and Evaluation Metrics

Datasets. We evaluate our model on English FrameNet (FN) and Chinese FrameNet (CFN) respectively. For English FrameNet, we use the FrameNet 1.5 release which is the full-text annotations and was used by Das et al. (2014). We use the same test data as Das et al., containing 4,458 targets. There are 19,582 targets in training data.

In experiment with Chinese FrameNet dataset. we select 25,000 annotated sentences in CFN exemplar sentences database as the training data set, which contained 1,567 targets and 180 frames. In order to compare with previous work in Chinese frame identification (CFI), we use two test sets. The first one, named tc1, consisted of 5000 sentences with marked targets that have not appeared in the training set. The second one, named tc2, used the same data set as Li et al. (2010) at the Coling 2010 Conference. This data set contains 7 different Chinese ambiguous targets. For each target, sentences were collected from Sogo Corpus and Contemporary Chinese Corpus of Beijing University and 940 sentences are selected for training data and 128 for test data.

Evaluation Metrics. We use the $\text{accuracy} = \frac{b}{r}$ to evaluate our model, where b is the number of correct frames identified, and r is the total number of frames identified.

4.2. Experimental Setting

In our experiment, word embedding is 100 dimensions, trained on Wikipedia by the skip-gram model (Mikolov et al., 2013). Position embedding is 50 dimension and initialized randomly. The filter matrixes on Convolution layer and the other weight matrixes are initialized randomly, following a Gaussian distribution. All biases are initialized to 0. Hyper-parameters are tuned on the development dataset. The results of the final values of hyper-parameter are shown in Table 1.

Table 1. Hyper-parameter setting.

description	Value
filter window size	2,3,4
filters number of each size	100
filter height	150
dropout rate	0.5
batch size	64
K-max	3
Initial learning rate	0.01

Table 2. The result of English FrameNet

Model	SEMAFOR Lexicon			Full Lexicon		
	All	Ambiguous	Unseen	All	Ambiguous	Unseen
Das et al.supervised	82.97	69.27	23.08	-	-	-
Das et al.best	83.60	69.19	42.67	-	-	-
LOG-LINEAR WORDS	84.53	70.55	27.27	87.33	70.55	-
LOG-LINEAR EMBEDDING	83.94	70.26	27.97	86.74	70.26	-
WSABIE EMBEDDING	86.49	73.39	46.15	88.41	73.10	-
CNN	86.13	80.35	70.46	86.54	80.43	70.66
CNN+multifilter	86.56	81.12	72.31	87.23	81.43	72.32
CNN+Input-att	86.45	81.45	73.67	86.55	81.53	73.85
CNN+Input-att+multifilter	87.25	82.03	72.64	87.75	82.47	72.13
CNN+Pooling-att	87.13	81.97	72.48	87.43	82.39	72.65
CNN+Pooling-att+multifilter	87.55	82.77	74.65	88.13	83.29	75.77
TSABCNN	**89.72**	**83.07**	**75.12**	**91.4**	**83.78**	**76.34**

4.3. Experimental results and analysis

To show the effectiveness of our proposed method, several state-of-the-art methods are selected as baseline for comparison on English FrameNet and CFN.

English FrameNet baseline are shown as follows:

Das et al.: A semi-supervised learning method was used to improve upon a supervised latent-variable log-linear model (Das et al., 2014).

Hermann et al.: Distributed representations of predicates and their syntactic context were used (Hermann et al., 2016).

Chinese FrameNet baseline are shown as follows:

Li et al: A tree-structed conditional random field model was used to solve Chinese FI based on Dependency Parsing (Li et al., 2010).

Zhao et al: BP neural network was used to learn the context features representation of a given target, and the selection of a frame for a given target (Zhao et al., 2016).

In addition, in order to analyze the effectiveness of each component of our neural network architecture, we use the ablation experiment on English FrameNet data set and CFN data set respectively, and compare with the above methods. The results are shown in Table 2, Table 3, Table 4.

Table 3. The comparision result with Zhao et al. of CFN

Model	All	Ambiguous	Unseen
Zhao et al	79.64	74.37	67.21
CNN	82.56	78.34	71.87
CNN+multifilter	83.34	78.95	72.25
CNN+Input-att	83.36	78.82	72.30
CNN+Input-att+multifilter	84.43	79.54	73.15
CNN+Pooling-att	84.13	79.34	72.10
CNN+Pooling-att+multifiter	85.3	80.34	73.41
TSABCNN	**86.8**	**81.76**	**73.97**

Table 4. The comparision result with Li et al..

Model	accuracy
Li et al.	81.46
CNN	83.57
CNN+multifilter	84.56
CNN+Input-att	84.58
CNN+Input-att+multifilter	85.65
CNN+Pooling-att	85.27
CNN+Pooling-att+multifiter	87.18
TSABCNN	**88.87**

Table 2 presents the accuracy for the state-of-the-art models on English FrameNet, and the comparision results with our proposed methods. Test data set is the SEMAFOR Lexicon and Full Lexicon. We present the results on all targets, ambiguous targets that evoke Multiple frames and unseen targets in the FrameNet lexicon or training data. As shown in Table 2, a high performance achieved on all targets and ambiguous word by traditional feature-based methods (Li et al., 2010), but a very low accuracy is obtained on unseen targets. Hermann et al (2016) used distributed representation of targets and their syntactic context performs better results on three test datasets than that of Das. For unseen targets, although the accuracy rate increased, it was far from being used. Simultaneously, we can see the performance is greatly improved on these datasets by a convolutional neural networks, especially on ambiguous targets and unseen targets, which show that deep learning can automatically produce some useful features. By ablation studies, multi filters and attention mechanism can also improve the accuracy of FI, which shows multi filters can learn diversity features and attention mechanism can make the important features get more weight. Consider all components, TSABCNN model proposed in this paper achieves the best performance, and strongly outperform other methods not only on all targets, but also ambiguous targets and unseen targets. On all target on FrameNet 1.5 release, our aproach achieves 91.4 accuracy, and beyond the best baseline nearly 4 point.

Table 3 shows the results on Chinese tc1 dataset. The results are consistent with the results in Table 2. From table 3, we can see accuracy on the CFN is slightly lower than that of FrameNet. It may be caused by the following factors: (1) Language differences. Chinese is more flexible and more ambiguous than English. (2) The scale of CFN corpus is much smaller than that of FN, and the neural network learning needs a large corpus support. Table 4 shows the results on Chinese tc2. On this dataset, CNN modelexceed that of Li et al. by 2.11 accuracy. It proves CNN can learn more useful context features for FI. While multi-filter, input-att, pooling-att are successively added to the model, Performance has a different degree of improvement. Finally, our proposed model, TSABCNN achieves 88.87 accuracy, and is higher 7.41 than the baseline.

5. Conclusion

In this paper, we propose a novel two-stage attention-based CNN model for FI. Our model utilizes CNN to automatically learn more important features for FI. We base attention mechanism to give higher weight to more important features. The experimental result shows that our model achieves better performance on three FI tasks than the baselines. Our model is not only suitable for the frame recognition of the all targets but also has a good effect on that of the unseen targets and the ambiguous targets. Furthermore, it is not limited to a language, so it is a general FI model. In the future work, we will focus on the research into joint identification of frames and arguments, and realize automatic labeling of semantic roles on FrameNet.

Acknowledgements. This work is supported by National Natural Science Foundation of China (No.61772324, No.61673248), and Shanxi Province Postgraduate Joint Training Base Talent Training Project (No.2018JD01, No.2018JD02).

References

1. Baker, C. F., Fillmore, C. J., Lowe, J. B.: The Berkeley FrameNet Project. Meeting of the Association for Computational Linguistics and, International Conference on Computational Linguistics,vol.47, pp.86-90(1998).
2. Fillmore, C. J.: Frame Semantics. Linguistics in the Morning Calm, pp.111-138(1982).

3. Hermann, K. M., Tomáš K., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P.: Teaching Machines to Read and Comprehend. In NIPS, pp.1693-1701(2015).
4. Das, D., Schneider, N., Chen, D., Smith, N. A.: Probabilistic frame-semantic parsing. Human Language Technologies: the 2010 Conference of the North American Chapter of the Association for Computational Linguistics, vol 40, pp. 948-956(2010).
5. Das, D., Chen, D., Schneider N., Smith, N. A.: Frame-semantic parsing. Computational Linguistics 40(1), 9-56(2014).
6. Carreras, X., Litkowski, K. C., Stevenson, S.: Semantic role labeling: an introduction to the special issue. Computational Linguistics 34(2):145-159(2008).
7. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A Convolutional Neural Network for Modelling Sentences. Eprint Arxiv 1.,(2014).
8. Kim, Y.: Convolutional Neural Networks for Sentence Classification. Eprint Arxiv(2014).
9. Zeng, D. J., Liu, K., Lai, S. W., Zhou, G. Y., Zhao, J.: Relation classification via convolutional deep neural network. Proceedings of COLING. Dublin, Ireland(2014).
10. Chen, Y., Xu, L., Liu, K., Zeng, D., Zhao, J.: Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. The Meeting of the Association for Computational Linguistics, pp. 167–176, Beijing, China (2015).
11. Dong, L., Wei, F., Zhou, M., Xu, K.: Question Answering over Freebase with Multi-Column Convolutional Neural Networks. Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, pp. 260-269, Beijing, China (2015).
12. Severyn, A., Moschitti, A.: Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. The International ACM SIGIR Conference , pp.373-382, ACM(2015).
13. Yin, W., Schütze, H.: MultiGranCNN: An Architecture for General Matching of Text Chunks on Multiple Levels of Granularity. Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, pp. 63-73, Beijing, China (2015).
14. Gildea, D., Jurafsky, D.: Automatic Labeling of Semantic Roles. Computational Linguistics 28(28), 245-288(2002).
15. Carreras X.: Introduction to the CoNLL-2004 shared task : Semantic role labeling. Proc. 8th Conference on Natural Language Learning (CoNLL), vol. 47, pp. 5-9, Boston, MA(2004).
16. Carreras, X., Màrquez, L.: Introduction to the CoNLL-2005 shared task: semantic role labeling. Proceedings of the 9th Conference on Computational Natural Language Learning(CoNLL), pp.152-164, ANN Arbor, MI(2005).
17. Fleischman, M., Kwon, N., Hovy, E.: Maximum entropy models for FrameNet classification. Conference on Empirical Methods in Natural Language Processing，pp.49-56. Sapporo(2003).
18. Erk, K.: Frame assignment as word sense disambiguation. In proceedings of the 6th International Workshop on Computational Semantic(IWCS-6).Proceedings(2005).
19. Erk, K.: Unknown word sense detection as outlier detection. Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Proceedings (2006).
20. Johansson, R., Nugues, P.: LTH: semantic structure extraction using nonprojective dependency trees. Proc of Semeval 13(4):227-230(2007).
21. Bejan, C. A., Hathaway, C.: UTD-SRL: a pipeline architecture for extracting frame semantic structures. International Workshop on Semantic Evaluations, vol. 8764, pp. 460-463 (2007).

22. Hermann, K. M., Das, D., Weston, J., Ganchev, K.: Semantic frame identification with distributed word representations. Meeting of the Association for Computational Linguistics,vol.1, pp.1448-1458(2014).
23. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. Computer Science(2014).
24. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research 12(1):2493-2537.(2011).
25. Yin, W. P., Schütze, H., Xiang, B., Zhou, B. W.: ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs. Action Editor: Brian Roark, vol. 4, pp. 259–272, (2016).
26. Zhang, Y., Wallace, B.: A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. Proceedings of the The 8th International Joint Conference on Natural Language Processing, pp. 253–263, Taipei, Taiwan, (2017).
27. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems, vol. 26, pp.3111-3119(2013).
28. Zhou, P., Shi, W., Tian, J., Qi, Z. Y., Li, B. C., Hao, H. W., Xu, B.: Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. Meeting of the Association for Computational Linguistics, pp. 207-212(2016).
29. Li, R., Liu, H. J., Li, S. H.: Chinese Frame Identification using T-CRF Model. COLING 2010, International Conference on Computational Linguistics, pp.674-682(2010).
30. Zhao, H. Y., Li, R., Zhang, S., Zhang, L. W.: Chinese frame identification with deep neural network. Journal of Chinese Information Processing 30(6):75-83(2016).