A Review on Entity Relation Extraction

Qianqian Zhang, Mengdong Chen and Lianzhong Liu
School of Computer Science and Engineering
BeihangUniversity
Beijing, China
zhangqianqian@buaa.edu.cn, cmd@buaa.edu.cn, lz_liu@buaa.edu.cn

***Abstract*—Because of large amounts of unstructured data generated on the Internet, entity relation extraction is believed to have high commercial value. Entity relation extraction is a case of information extraction and it is based on entity recognition. This paper firstly gives a brief overview of relation extraction. On the basis of reviewing the history of relation extraction, the research status of relation extraction is analyzed. Then the paper divides theses research into three categories: supervised machine learning methods, semi-supervised machine learning methods and unsupervised machine learning method, and toward to the deep learning direction.**

Keywords—internet; entity recognition;relation extraction;deep learning

I. INTRODUCTION

With the rapid development of information technology and the extensive application of network, the Internet has gradually become an indispensable part of people's life. But the explosive growth of the Internet makes it difficult for people to get useful information. In this context, information extraction technology emerges as the times require. The main purpose of information extraction is to extract the specified entities, relationships and events from natural language texts. Entity extraction is the basis of relation extraction and event extraction. Relation extraction is one of the key technologies of information extraction, and has been paid more and more attention in recent years.

Relationships between entities can be formally described as relational three tuples < Entity1, Relation, Entity2 >, in which Entiy1 and Entity2 are entity types, and Relation is relation description. Relation extraction is to extract the relational three tuple < Entity1, Relation, Entity2 > from natural language text so as to extract text information. Such as “Bill Gates is the founder of Microsoft”, we can extract the relationship between Bill Gates and Microsoft as the founder.

Relational extraction technology has practical value in many fields. In automatic question answering system, relation extraction automatically links related questions and answers. In the retrieval system, the implementation of relation extraction semantic retrieval function is possible. In the process of ontology learning, relation extraction can discover new relationships between entities to enrich the structure of ontology. In semantic web labeling tasks, relation extraction can automatically associate semantic web knowledge units.

This article introduces the history of relation extraction and research status. The next part treats the classification of relation extraction according to its principle. Finally, there are conclusions.

II. HISTORY AND RESEARCH STATUS

A. *The History of Entity Relation Extraction*

Information extraction technology has been studied for many years at home and abroad, and has made a mature development. The development of information extraction technology can not be separated from the convening of information understanding Conference (MUC). From 1987 to 1998, the MUC session was held for the seven time, and the text mining tasks including named entity recognition, template relation and plot template were defined. The task was introduced in 1998 MUC-7 2 meeting, the purpose is through filling the form template slots to extract entities between Location of, Employee of and Product of three kinds of relationship. One important activity of the conference is the evaluation of the information extraction system, which mainly considers two evaluation indexes: Recall , Precision and F_1:

$$\text{Precision} = \frac{C}{M} \tag{1}$$

$$\text{Recall} = \frac{C}{N} \tag{2}$$

$$F_1 = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{3}$$

Among them, C is the number of relational instances to be extracted correctly, M is the total number of extracted relational instances, and N is the number of relational instances in the standard result set.

In MUC-7, MUC was introduced by NIST ACE (Automatic Content Extraction) evaluation replaced. The ACE conference is designed to automatically extract the entities, relationships, and events in the news corpus. Relation extraction belongs to the relation detection and recognition (Relation Detection and Recognition, RDR) task of ACE conference definition. The ACE conference provides a corpus of relational extraction, and constructs a detailed entity relation type to refine the relation extraction task further.

Although the research of information extraction in China started late, it has already made some fruitful work in relation extraction. Ming et al. [1] at the center of China research Intel demonstrate a system they developed in ACL 2000, the role of the system is to select Chinese named entity and the relationship between these entities, the system uses memory based learning (MemoryBasedLearning, MBL) the relationship between rule acquisition algorithm is used to extract named entities and their relation. Che et al. [2] used 2004 ACE evaluation of training data as the experimental data, using the two feature vector machine learning algorithm based on Winnow and SVM of entity relation

978-1-5386-2628-3/17 $31.00 © 2017 IEEE
DOI 10.1109/ICMCCE.2017.14

extraction, and pointed out that in relation extraction, should focus on trying to find good features. Zhang et al. [3] first do relation extraction using composite kernel function method, they use convolution kernel (convolution parse tree kernel) and the physical characteristics of the kernel function to form a new compound function, experiments show that the composite kernel performance is better than any single kernel function. Liu et al. [4] add semantic information into the kernel of word sequence to form semantic sequence kernel function, and combine KNN machine learning algorithms to construct classifier to classify and label relation types.

B. Research Status of Entity Relation Extraction

In general, the problem of relation extraction is transformed into a classification problem, that is, first, lists all the entity pairs in a sentence, and then uses a classifier to decide which relationships we really need. Traditional systems treat entity relation extraction as a pipeline of two separated tasks, i.e., named entity recognition (NER) (Ferraro et al. [5], Dinu et al. [6]) and relation extraction.

Like the usual solution to classification problems, people initially used the knowledge base approach to solve the problem. Yes, this method requires experts to build a large knowledge base, which requires not only specialists with professional skills, but also a lot of work.

In order to overcome the shortcomings of the knowledge base method, people use the machine learning method to solve this problem later. The method does not require specialists with specialized skills to write a knowledge base. Only a person with a certain professional knowledge can make judgments about the relationship between any two entities and not the ones that we need. Then, as a training data, the classifier is constructed by using various learning methods.

The usual machine learning algorithms need to construct training data in the form of eigenvectors. Then various machine learning algorithms, such as support vector machines (SVM) [7] and Winnow [8], are used as learning machines to construct classifiers. This method is called eigenvector based learning algorithm. Then there appeared the Kernel based learning algorithm, which was first introduced in support vector machines (SVM), and later discovered that a variety of learning methods can be represented in the form of Kernel. They are also called Kernel based learning algorithms [9] [10]. Learning algorithm based on Kernel in Natural Language Processing applications does not need to construct the feature vector, the original form uses the string as the processing object. Zelenko et al. [11] and Culotta et al. [12] use Kernel method to solve the problem of relational extraction, and achieve better results. The methods they use need to make shallow parsing of the processing objects. One of the fatal drawbacks of Kernel, however, is that training and prediction are too slow to cope with large amounts of data. Recent studies show that end-to-end (joint) modeling of entity and relation is important for high performance (Miwa et al. [23]) since relations interact closely with entity information.

III. THE CLASSIFICATION OF ENTITY RELATION EXTRACTION METHODS

According to the artificial participation and benchmarking note depending on different corpora, relation extraction method based on machine learning can be divided into three categories: supervised machine learning methods, semi-supervised machine learning methods and unsupervised machine learning method, and toward to the deep learning direction.

A. Supervised Machine Learning

The supervised method studies the model from the training data set and predicts the relation types of the test data. The input space of a system is a natural statement, and the output space is a collection of predefined relationships. Because the vectors in the relation extraction task are unstructured natural language, it is necessary to formalize the language units at different levels of the text for making the machine recognize and predict. Different ways of processing sentences can be divided into 2 categories: eigenvector method and kernel function method.

1) Eigenvector Based Method: The Eigenvector based method extracts a series of features $[f_1, f_2, \dots f_n]$ from the context information, the part of speech, the syntax and so on to train a classifier (naive Bayes, support vector machines, maximum entropy, etc.) and then completed the task of relation extraction. The so-called eigenvector is a numerical representation of an instance. That is to say, an instance is transformed into the eigenvector x, where x^i is the i'th element of the N dimensional feature vector x. Based on the feature vector machine learning algorithm is given a set of training data for $(x^1, y^1), (x^2, y^2), \dots (x^n, y^n)$, of which two yuan for the classification problem of $y^i \in \{-1,1\}$, to learn a classification function f, so that for a given new feature vector x', f can classify them correctly, that is $f(x') = y'$. The classification function f is generally defined as a hyperplane determined by the weighted vector w, which is able to strictly separate the data labeled -1 and labeled +1. The weight vector w is solved by a variety of machine learning algorithms.

Kambhatla et al. [13] use the maximum entropy model to combine lexical, syntactic and semantic features to extract semantic relations. On the basis of the commonly used contextual features, Sun et al. [14] add long term correlation features, entity sequential features, sequential features among entities, and punctuation features between 2 entities. Then, a hybrid algorithm of Bias and perceptron is used to classify. Miao et al. [15] introduced the word feature, word segmentation feature and syntax feature into relational annotation, and used CRF (conditional random field) method to estimate the feature.

2) Kernel Based Method: Kernel based method is to use the kernel function to directly calculate the similarity between two instances to train the relation classification model. The core step is how to design a kernel $k(x, y)$ that calculates the similarity of two instances (x,y). The kernel function is a function of K, makes all of its domain $x, z \in X$, X is the input space .

$$\mathrm{k(x,z)} = (\emptyset(\mathrm{x}), \emptyset(\mathrm{z})) \tag{4}$$

The $\emptyset$ here is a mapping from the space X to the inner product feature space F.

Zelenko et al. [16] use the shallow syntactic parsing results, and then use the connection of the smallest common subtree entity relation instance representation, by calculating the kernel function between the two subtrees to traine SVM classifiers, has achieved good effect in relation extraction small news corpus. Kazuma et al. [17] present a recursive neural network (RNN) model that works on a syntactic tree. The model differs from previous RNN models in that the model allows for an explicit weighting of important phrases for the target task. Nguyen et al. [18] depart from traditional approaches with complicated feature engineering by introducing a convolutional neural network for relation extraction that automatically learns features from sentences and minimizes the dependence on external toolkits and resources. The model takes advantages of multiple window sizes for filters and pre-trained word embeddings as an initializer on a non-static architecture to improve the performance. Zayaraz et al. [19] propose a methodology to extract concept relations from unstructured text using a syntactic and semantic probability-based Naïve Bayes classifier. They propose an algorithm to iteratively extract a list of attributes and associations for the given seed concept from which the rough schema is conceptualized. A set of hand-coded dependency parsing pattern rules and a binary decision tree-based rule engine were developed for this purpose. Lee et al. [20] experimented with various strategies to incorporate argument ordering for ordering sensitive relations, showing that an efficient strategy is to fix the arguments ordering as appears on the text by introducing reverse relations.

Because supervised methods are based on well labeled training sets, good accuracy and recall rates can be obtained on the same type of test set. But at the same time, the supervised method is limited to the training corpus and can not recognize the semantic relation which is not included in the training corpus. Therefore, this method is not suitable for dealing with large scale corpus of open domain. The method of supervision is relatively mature, and it takes full advantage of all kinds of characteristic information in the sentence. It is worth to be used for reference.

B. Semi-Supervised Learning Method

Semi-supervised learning method is called weakly supervised learning method. Its basic idea is to establish a learning device using the assumption of the model, to label the unlabeled sample, which mainly solves the problem of learning to improve the generalization ability of the model in the label sample under the condition of insufficient. It can weaken the malpractice to a certain extent and optimize the learning effect. At present, it is usually used to solve the problems of classification and relation extraction in the field of knowledge extraction . Semi-supervised learning methods include support vector machines (SVM), bootstrap and remote monitoring methods. The model is shown in Fig. 1:

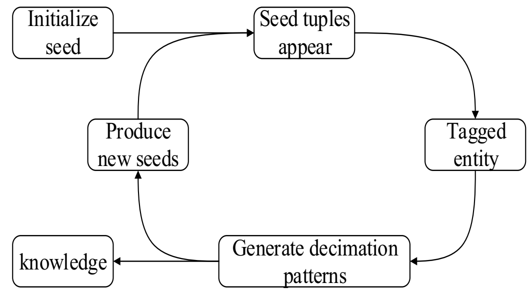

Fig.1. Semi-supervised learning method processing model

Hoffmann et al. [21] apply their model to learn extractors for NY Times text using weak supervision from Freebase. The model presents a novel approach for multi-instance learning with overlapping relations that combines a sentence-level extraction model with a simple, corpus-level component for aggregating the individual facts. Isabelle et al. [22] make relation extraction by jointly training the named entity classifier and the relation extractor using imitation learning which reduces structured prediction learning to classification learning. Miwa et al. [23] present a novel end-to-end neural model to extract entities and relations between them, their recurrent neural network based model captures both word sequence and dependency tree substructure information by stacking bidirectional treestructured LSTM-RNNs on bidirectional sequential LSTM-RNNs. This allows the model to jointly represent both entities and relations with shared parameters in a single model. Chandra et al. [24] propose a model, which is based on the end-to-end relation extraction model of Miwa et al. [23] with several enhancements such as semi-supervised learning via neural language models, character-level encoding, gazetteers extracted from existing knowledge bases, and model ensembles.

Semi-supervised machine learning method can effectively reduce the dependence of human participation and corpus annotation, and can be extended to large-scale text relational extraction task. It has been widely used. However, the bootstrap method has the problem of semantic drift [25] in the iterative process, which affects the accuracy of the extraction results.

C. Unsupervised Learning Method

Supervised and semi-supervised machine learning methods need to determine the type of relationship in advance, in fact, in large-scale corpus, people are often unable to predict all types of entity relations. Some researchers try to solve the problem by means of unsupervised machine learning based on the idea of clustering.

Unsupervised relational extraction was first proposed by Hasegawa et al. [26] at the ACL conference in 2004, and most of the subsequent approaches were improved on the basis of Hasegawa. With the previous hypothesis information, he uses the text information clustering between two entities to express the relation class. The result shows that the clustering method is very feasible in relation extraction. Shinyama et al. [30] proposed an unsupervised extraction method for multi-level clustering. First, they get the news text through the crawler, and then start the classification according to the origin of the article.

Then, according to the semantic structure of sentences, extract the basic pattern clustering entity in satisfy a series of constraints, these entities are mapped according to the basic model, the formation of secondary clustering, so each secondary cluster contains the relationship of entities with the same. Quan et al. [28] present an unsupervised method based on pattern clustering and sentence parsing to deal with biomedical relation extraction. Pattern clustering algorithm is based on Polynomial Kernel method, which identifies interaction words from unlabeled data; these interaction words are then used in relation extraction between entity pairs. Xu et al. [29] propose an unsupervised pre-training method based on the sequence-to-sequence model for deep relation extraction models. The pre-trained models need only half or even less training data to achieve equivalent performance as the same models without pre-training.

Unsupervised methods generally require large-scale corpora as support. By exploiting the redundancy of the corpus, the possible relational schema sets are mined, and then the relation names are determined. The deficiency of this method is that the relation name is difficult to describe accurately, and the recall rate of low frequency relation is low.

D. Deep Learning Method

The existing supervised methods for relational extraction have achieved good results, but they have relied heavily on the classification of natural language processing such as part-of-speech annotations, syntactic parsing, and natural language processing. Natural Language Processing tagging tools often contain a lot of errors. These errors will be amplified in relational extraction and affect the effect of relational extraction. Recently, many researchers have applied depth learning techniques to relational extraction.

Socher et al. [30] proposed the use of recurrent neural networks to solve the problem of relational extraction. Firstly, the sentence is parsed synthetically and then expressed as a learning vector for each node on the syntactic tree. Through the recursive neural network, it can start from the top of the sentence in the syntactic tree, and iterate through the syntactic structure of the sentence to finally get the sentence used for relational classification. The method can effectively consider the syntactic structure information of the sentence, but at the same time the method can not consider the position and sentence information of the two entities in the sentence well. The model is shown in Fig. 2:

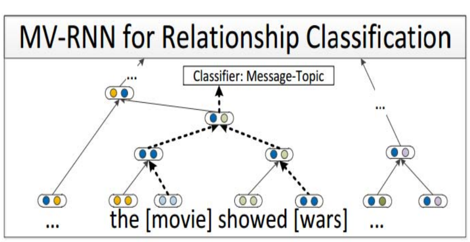

Fig.2. The MV-RNN learns vectors in the path connecting two words (dotted lines) to determine their semantic relationship. It takes into consideration a variable length sequence of various word types in that path.

Santos et al. [31]proposed a new convolution neural network for relational extraction, which uses a new loss function that can effectively improve the distinction between different relationship categories. The model is shown in Fig. 3:

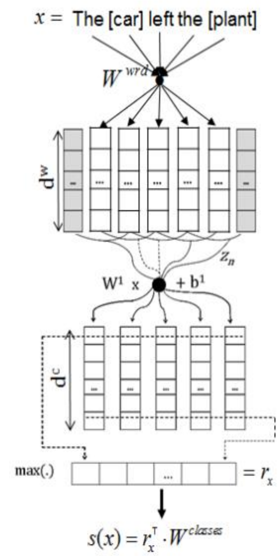

Fig.3. CR-CNN: a Neural Network for classifying by ranking.

Miwa et al. [23] proposed a relational extraction model based on end-to-end neural networks. The model uses both bidirectional LSTM and tree LSTM to model both entities and sentences. At present, the convolution neural network based on the relational extraction of the standard data set SemEval-2010 Task 8 to achieve the best results. Zeng et al. [34] attempts to extend the relational extraction model based on convolution neural networks to remote monitoring data. Which assumes that there is at least one sentence in each sentence of each entity that reflects the relationship of the entity pair, and proposes a new learning framework; the entity is the unit, and for each entity pair only the one that best reflects its relationship sentence. This method solves the application of the neural network relational extraction model in the remote monitoring data to a certain extent, and achieves the prediction effect of the model based on the feature-based relational extraction model in the NYT10 data set. However, the method still has some flaws: the model for each entity can only use a sentence to learn and predict, loss of information from a large number of other effective sentences. Lin et al. [35] proposed a neural network model based on the sentence-level attention mechanism to solve this problem, which can assign weights to each sentence according to a specific relationship, and obtain a higher weight through continuous learning. The noise of the sentence gets a smaller weight. Compared with the model, the method has been greatly improved. The model is shown in Fig. 4:

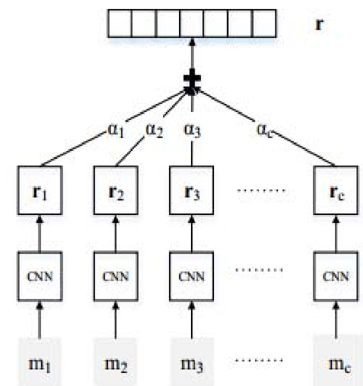

Fig.4. The architecture of sentence-level attention-based CNN, where m_i indicates the original sentence for an entity pair, α_i is the weight given by sentence-level attention.

The current neural network relation extraction is mainly used to pre set the set of relations. However, facing oriented domain relation extraction is still based on traditional methods such as template. Therefore, we need to explore how to introduce neural networks into open domain relation extraction and automatically discover new relationships and their facts. In addition, it is worthwhile to explore the existing neural network model for fast learning of new relations and examples [36].

IV. CONCLUSIONS

We have provided a very brief introduction to the entity relation extraction and its research status. Then the paper divides theses research into three categories: based on machine learning, based on natural language analysis and based on deep learning. Relation extraction technology has practical value in many fields, and with the continuous improvement of the entity relation extraction technology, its application areas will be growing.

REFERENCES

[1] Zhang Y, Zhou J F. A trainable method for extracting Chinese entity names and their relations[C]// The Workshop on Chinese Language Processing: Held in Conjunction with the, Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2000:66-72.

[2] Che W X, Liu T, Li S. Automatic entity relation extraction[J]. Journal of Chinese Information Processing, 2005.

[3] Zhang. M, Zhang. J, Su. J, et al. A Composite Kernel to Extract Relationsbetween Entities with both Flat and Structured Features[C]. 21st International Conferenceon Computational Linguistics and 44th Annual Meeting of the ACL, 2006: 825-832.

[4] Liu K, Fang L, Lei L, et al. Implementation of a Kernel-Based Chinese Relation Extraction System[J]. Journal of Computer Research & Development, 2007, 44(8):1406-1411.

[5] Qu L, Ferraro G, Zhou L, et al. Named Entity Recognition for Novel Types by Transfer Learning[J]. arXiv preprint arXiv:1610.09914, 2016.

[6] Ni J, Dinu G, Florian R. Weakly Supervised Cross-Lingual Named Entity Recognition via Effective Annotation and Representation Projection[J]. arXiv preprint arXiv:1707.02483, 2017.

[7] N.Cristianini and J.Shawe-Taylor:An Introduction to Support Vector Machines, Cambridge University Press(2000)[J]. Journal of Japanese Society for Artificial Intelligence, 2001, 16.

[8] T. Zhang. Regularized winnow methods. In Advances in Neural Information Processing Systems 13, pages 703-709, 2001.

[9] D. Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, 7, 1999.

[10] Lodhi H, Saunders C, Shawe-Taylor J, et al. Journal of Machine Learning Research 2 (2002) 419-444 Submitted 12/00; Published 2/02 Text Classification using String Kernels[J]. 2002.

[11] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[C]// Acl-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2002:71-78.

[12] Culotta A, Sorensen J. Dependency tree kernels for relation extraction[C]// Meeting of the Association for Computational Linguistics, 21-26 July, 2004, Barcelona, Spain. DBLP, 2004:423-429.

[13] Kambhatla, Nanda. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[J]. Acldemo Proceedings of the Acl on Interactive Poster & Demonstration Sessions, 2004.

[14] Sun X, Dong L. Feature-Based Approach to Chinese Term Relation Extraction[C]// International Conference on Signal Processing Systems. IEEE Computer Society, 2009:410-414.

[15] Miao Qingliang, Zhang Shu, Zhang Bo, et al. Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text[C]. 26th Pacific Asia Conference on Language, Information and Computation, 99-107.

[16] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[C]// Acl-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2002:71-78.

[17] Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. 2013. Simple customization of recursive neural networks for semantic relation classification. In EMNLP. pages 1372– 1376.

[18] Thien Huu Nguyen and Ralph Grishman. 2015. Relation extraction: Perspective from convolutional neural networks. In Proceedings of NAACL-HLT. pages 39–48.

[19] Godandapani Zayaraz and Suresh Kumara. 2015. Concept relation extraction using na¨ıve bayes classi- fier for ontology-based question answering systems. Journal of King Saud University-Computer and Information Sciences 27(1):13–24.

[20] Lee J Y, Dernoncourt F, Szolovits P. MIT at SemEval-2017 Task 10: Relation Extraction with Convolutional Neural Networks[J]. arXiv preprint arXiv:1704.01523, 2017.

[21] Hoffmann R, Zhang C, Ling X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 541-550.

[22] Augenstein I, Vlachos A, Maynard D. Extracting relations between non-standard entities using distant supervision and imitation learning[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2015: 747-757.

[23] Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using lstms on sequences and tree structures. In ACL.

[24] Ammar W, Peters M, Bhagavatula C, et al. The AI2 system at SemEval-2017 Task 10 (ScienceIE): semi-supervised end-to-end entity and relation extraction[C]//Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017: 592-596.

[25] Etzioni O, Cafarella M, Downey D, et al. Unsupervised named-entity extraction from the Web: An experimental study ☆[J]. Artificial Intelligence, 2005, 165(1):91-134.

[26] Riedel S, Yao L, Mccallum A. Modeling relations and their mentions without labeled text[C]// European Conference on Machine Learning and Knowledge Discovery in Databases. Springer-Verlag, 2010:148-163.

[27] Curran J R, Murphy T, Scholz B. Minimising semantic drift with Mutual Exclusion Bootstrapping[C]// 2007:172-180.

[28] Quan C, Wang M, Ren F. An unsupervised text mining method for relation extraction from biomedical literature[J]. PloS one, 2014, 9(7): e102039.

[29] Li Z, Qu L, Xu Q, et al. Unsupervised Pre-training With Seq2Seq Reconstruction Loss for Deep Relation Extraction

Models[C]//Proceedings of the Australasian Language Technology Association Workshop 2016. 2016: 54-64.

[30] Shinyama Y, Sekine S. Preemptive Information Extraction using Unrestricted Relation Discovery.[C]// Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 4-9, 2006, New York, New York, USA. DBLP, 2006.

[31] Rosenfeld B, Feldman R. Clustering for unsupervised relation identification[C]// Sixteenth ACM Conference on Conference on Information and Knowledge Management. ACM, 2007:411-418.

[32] Socher, Richard, et al. "Semantic compositionality through recursive matrix-vector spaces." Proceedings of EMNLP-CoNLL, 2012.

[33] Santos, Cicero Nogueira dos, Bing Xiang, and Bowen Zhou. "Classifying relations by ranking with convolutional neural networks." In Proceedings of ACL, 2015.

[34] Zeng D, Liu K, Chen Y, et al. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C]//Emnlp. 2015: 1753-1762.

[35] Lin Y, Shen S, Liu Z, et al. Neural Relation Extraction with Selective Attention over Instances[C]//ACL (1). 2016.

[36] Yankai Lin, Zhiyuan Liu. Relational extraction based on deep learning [J]. CIPS green work academic column third.