Research on syntactic dependency tree and Ontology constraint in remote Supervising relation extraction

Zhao Shuxu\textsuperscript{1,a}, Zhang yiguang\textsuperscript{1,b}, Wang xiaolong\textsuperscript{1,c}

\textsuperscript{1} School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China

\textsuperscript{a}small25300@163.com, \textsuperscript{b}kingzyguang@163.com, \textsuperscript{c}465006760@qq.com

Abstract. At present, research on remote-supervised relationship extraction fail to fully consider the position information among words in a sentence, and the extracted relationship may be inconsistent with the actual relationship. Therefore, this paper combines syntactic dependency tree and ontological constraint to carry out remote-supervised relationship extraction. The syntactic dependency tree is used to obtain the position weight of each word in the sentence and the domain ontology constraint is introduced to improve the accuracy of the extraction relationship. Experiments on Freebase+NYT data set show that this model can effectively reduce the noise interference of wrong labels, and the model improves the accuracy by 2\% compared with other reference models, thus better realizing relation extraction and laying a relevant foundation for the construction of high-precision knowledge map.

1. Introduction

As one of the main tasks in the process of knowledge map construction, relationship extraction aims to extract the relationship between two entities from unstructured text\cite{1}. Therefore, relation extraction has a wide application value in the fields of knowledge map construction, semantic analysis and knowledge question and answer.

In the supervised method, relationship extraction is solved as a multi-classification problem, it is mainly based on the methods of eigenvector and kernel function\cite{3}. Zhou and Guo Xiyue et al used support vector machine (SVM) as classifier to extract entity semantic relationship in 2014\cite{3}. The supervised relationship extraction method has a high accuracy rate, but it makes insufficient use of the context information. The relationship extraction method of unsupervised learning is based on the theory of distribution hypothesis. In 2004, Hasegawa\cite{6} et al. proposed to use the method of unsupervised learning to extract entity relationship in ACL meeting, but the relationship extracted by the unsupervised method does not have semantic information, which is difficult to be used to construct the knowledge map. In order to solve the problem of high cost of manual annotation corpus and difficult regularization of extraction relationship, the researchers put forward semi-supervised learning and open relation extraction methods, among which, Brin\cite{7} USES Boostrapping open method for entity naming identification. Craven\cite{8} et al. proposed the idea of semi-supervised learning for the first time in the process of extracting structured data from texts and establishing biological knowledge base.

In order to obtain a high-performance relational extraction model, remote monitoring\cite{10} based approach has begun to receive extensive attention. In 2009, Mintz\cite{9} et al. proposed the hypothesis that "if there is a relationship between two entities in the knowledge base, the unstructured sentences containing the two entities can represent the relationship". Based on the assumption of remote supervision, sentences in the external document library can be automatically labeled with relation according to the
existing knowledge map, so as to obtain a large number of training examples and greatly reduce the problem of manual labeling. Although this method overcomes the shortcoming of high cost of manual marking data in supervised method, it also brings the problem of new back-marking noise.

Hoffmann[11] et al. and Surdeanu[12] et al. proposed multi-instance learning and multi-instance and multi-tag methods respectively to alleviate the problem of mislabeling. Zeng[13] et al. used piecewise convolutional neural network to automatically extract features and combined the method of multi-instance learning to extract relationships. Lin[14][15] et al. introduced the attention mechanism on the basis of piecewise convolution magic network, and dynamically adjusted the instance weight in each package. Ji[16] et al. added entity pair description information to the attention mechanism to increase the attention weight to the target relationship. Che[17] et al. further optimized the model by combining the character-level attention mechanism with the sentiment-level attention mechanism. Based on the remote monitoring model, Li Yanjuan[20] added constraints of ontology domain to alleviate the impact of mislabeling more effectively.

2. Related work
The relationship extraction method based on remote monitoring model is mainly based on the hypothesis of Mintz[9]. As for the existing problems in relation extraction at present:

(1) Many existing researches mainly use the pre-trained word vector to give the representation of each word in the attribute and the position feature vector to splicing to get the input vector of the model. This representation methods only consider the sentence words with the relative distance between entities, on the one hand, missing words in the phrase sequence information, on the other hand, the current commonly used prediction models such as GloVe or word2vec training term vectors based on word co-occurrence information training within a certain window, thus lead to similar words from the close relationship between vector said, it is difficult to differentiate said.

(2) The existing models cannot accurately extract the features of the relationship descriptors for the sentences in which the entity pairs are close to each other.

(3) For a domain with multilevel complex layer structure entities, the knowledge map provides few relation instances, which makes it impossible to accurately judge whether the sentence is noise data or not, leading to the inconsistency between the extracted relation and the domain knowledge.

Therefore, based on the existing models, this paper introduces syntactic dependency tree[22] and ontology constraint layer[20] to improve the above problems. As for the relation attribute and the distance characteristic between each word in the sentence, this paper introduces syntactic dependency tree, fully considers the position information of words in the sentence and the grammatical dependence relation between words, and finally extracts the relation. At the same time, we add the constraint layer of ontology, which can provide a wealth of domain knowledge, including the stratification of entities and the relationship between the top-level entities. In this way, it can solve the problem that the final extraction relationship does not conform to the common sense, and make the extracted relationship more consistent with the common sense of the field.

3. Relationship extraction model based on syntactic dependency tree and ontology constraint and the proces
In this paper, the overall process of remote monitoring relationship extraction mod-el based on syntactic dependency tree and ontology constraint is shown in Figure.1
Because the assumption of remote supervision is too strong, it is easy to produce back-marking noise during the process of marking samples, which affects the effect of relation extraction. Therefore, in this paper, TreePCNNs(s)+OR model is proposed based on the senting-level attention model APCNNs(S)+OR[20], combined with the syntactic dependency tree and the relational constraint layer of ontology domain. As shown in Figure 2, the model is composed of three modules: feature engineering and feature extraction module, classifier module, and ontology constraint module.

Feature engineering and extraction module includes syntactic dependency tree, word vector representation and position vector representation layer, feature extraction module includes convolutional layer and segmental maximum pooling layer, and classifier module includes attention layer and classifier. On the characteristics of engineering module, in order to alleviate caused by the problem of expression from attributes of the relationship between linear decrease of the weight of related words, all the words in the input sentence word vectors, and use of syntactic dependency tree and combining grammar distance to generate the corresponding position of each word in the sentence weight combination, again with the position vector joining together to get the final model input vector. In the feature extraction module, PCNN (Piece-wise-CNN) was used to obtain the feature vector representation of sentences, and the sentence-level attention mechanism combining entity description was introduced into the classifier module, which effectively alleviated the impact of tag mislabeling on the model.

3.1. Feature engineering and feature extraction module
Do some preprocessing of the sentence. For example, the relative distance, grammatical distance and dependency relationship analysis are calculated, and the position weight of each word in the sentence is calculated according to the relation attribute and the corresponding relation attribute dependency clause of the sentence, and the input feature vector is constructed by using the word vector combined with the position weight and position vector. A word vector is a distributed representation of a word that maps each word in the text to a lower-dimensional vector. The position vector is a combination of the relative distances of the current word to entities and the eigenvector obtained by multiplying the word vector by the position weight and then splicing with the position vector is the final input vector.
3.2. word embedding

Given a set of sentences consisting of $S = \{w_1, w_2, w_3, \ldots, w_n\}$, use word2vec [18]. To vectorize the sentences using formula (1):

$$x_i = w_{word} \cdot V^i$$  (1)

Thereinto, $w_{word} \in R^{d \times m}$ is the vector matrix of the sentence, which is a fixed-size word list, $d^*$ is the dimension of the word vector, and $V^i$ is one-hot the representation of $w_i$. Finally, the vectorization representation of each word in the sentence is $V_s = \{x_1, x_2, x_3, \ldots, x_n\}$.

3.3. Dependency trees and distance information for relationship properties

3.3.1. Dependency subtrees of relational attributes.

In the process of relation extraction, each sentence may contain several different relation attributes and correspond to different entity pairs. For instance: “Zhang San graduated from Beijing University and worked in Huawei” contains two different relationship attributes “graduated” and “worked”, corresponding to different entities, “Zhang San”, “Beijing University” and “Huawei”.

![Sentence dependency tree structure](image)

As shown in figure 3, sentence $S$ contains two relationship properties: ($\alpha_1$, "graduated") and ($\alpha_2$, "worked"). Relationship attribute $\alpha_1$ is a word, and there is a modifier "from" with a limited part of speech. Attribute $\alpha_2$ is a word, and there is a modifier "in" with a limited part of speech. Obviously, relationship attribute modifiers general word, an entity attributes and attribute words together in relationship with word dependent child of a particular word for the root node in the tree, so the sentence dependency tree structure can be used to extract relationship attribute dependent subtree, and contains attributes and the relationships between the clauses [21]. Compared with the average value or linear attribute feature representation of normal word vectors of $S_{\alpha_1}$ and $S_{\alpha_2}$, the relational attribute clause contains not only the relational attribute word itself, but also the modifier information of the relational attribute and the position information between them.

3.3.2. Grammar distance

Using relative distance to depict the position relation between words and relation attribute words will make the weight of each word in the sentence centered on relation attribute words and decrease linearly to the left and right position weight. Therefore, in some cases, relative distance is difficult to fully and objectively reflect the impact of each word in the sentence on the relationship between attributes and positions. The distance between the word and the relation attribute is defined as the path length of the two in the sentence dependency tree, that is, the grammatical distance. In the dependency tree of a sentence, words are connected through the de-pendency relationship. Therefore, the more related words
are, the smaller their grammatical distance is, which is irrelevant to their actual position in the sentence. In other words, the grammatical distance can better reflect the correlation between words and relation attributes. The dependency tree syntax distance for relationship attributes $\alpha_1 = \text{"graduated"}$ and $\alpha_2 = \text{"worked"}$ is shown in Figure 4.

![Figure 4: Dependency tree syntax distance for relational attributes](image)

In the calculation of syntactic distance based on dependency tree, words closer to relational attribute have higher weight. If the relation attribute is a phrase, then the grammatical distance from each word in the sentence is calculated as represented by the last word in the phrase.

Each sentence $S$ and relationship attribute $\alpha$, the syntax distance vector $D_{s,\alpha} = (d_{1,\alpha}, d_{2,\alpha}, \ldots, d_{n,\alpha})$ is obtained based on the dependency tree calculation, where $d_{i,\alpha} \in R^n$ is the grammatical distance between word $w_i$ and relationship attribute $\alpha$. The weight of word $w_i$ in sentence $S$ is $l_{i,\alpha}$:

$$l_{i,\alpha} = 1 - \frac{d_{i,\alpha}}{2 \times d_{\max}} \quad (2)$$

Therein, $d_{\max}$ is the maximum value of $D_{s,\alpha}$. According to Equation (3): Position weight $l_i \in [0.5, 1]$. According to each relationship attribute $\alpha$, a weight vector $L_{\alpha} = (l_{1,\alpha}, l_{2,\alpha}, \ldots, l_{n,\alpha})$ based on grammatical distance with the same length as sentence $S$ can be calculated. Each element $l_{i,\alpha} \in R([i \leq i \leq |S|])$ is the position weight of word $w_i$ in the corresponding position in the sentence when predicting the specific relationship attribute $\alpha$. Finally, the word vector is multiplied by the position weight to get $V = V \times L_{\alpha}$, and then spliced with the position vector to get the final input vector.

### 3.4. position vector

In the task of relation extraction, the closer the words are to the entities, the more the relationship between the entities can be highlighted. In order to express the meaning of the sentence more accurately, the relative distance between each word and two entities in the sentence is spliced into the word vector representation of the word.

As shown in Figure 5, "China" and "world" are two entities. The distance from "is" to "China" is, and the distance from "world" is. The sign indicates that the word is located before and after the entity. The two position matrices are initialized randomly $P F_i (i = 1, 2)$, respectively represents the distance of each word to entity $e_1$ and entity $e_2$, then, convert the relative distance to a vector and put it in the matrix.

![Figure 5: Example of relative distance of position vector](image)
As shown in Figure 6, this paper spliced semantic vector and position vector together as the input vector of the model. If in sentence vectorization, the dimension of word vectorization and the dimension of position vectorization are, then the dimension of sentence vector is:

$$d^s = d^w + 2d^p$$  \hspace{1cm} (3)

![Figure 6 Model input vector](image.png)

### 3.5. PCNN Convolution layer

Let sentence $S = \{s_1, s_2, s_3, \ldots, s_n\}$ be the input, $s_i$ is the word vector for the $i$-th word, $S \in \mathbb{R}^{n \times k}$, $s_{i,j}$ is the vector dimension of the $i$-th word. Use $s_{i,j}$ to represent the sequence from the $j$-th word to the $i$-th word in the sentence.

A convolution operation is an operation between weight $W$ and the input matrix $S$. Eigenvector $c_j \in \mathbb{R}^{n \times (w-1)}$:

$$c_j = WS_{j-u+1:j}$$  \hspace{1cm} (4)

Where, variable $j$ ranges from 1 to $n \times w - 1$, and if $i < 1$ or $i > n$, goes out of the range and takes the value of 0.

In order to obtain more local features, multiple convolution kernels are used for convolution operation, as shown in Equation (5):

$$c_i = W_j S_{j-u+1:j}, \ 1 \leq i \leq m$$  \hspace{1cm} (5)

And you get the result matrix $C = \{c_1, c_2, \ldots, c_m\} \in \mathbb{R}^{m \times (n \times w - 1)}$

### 3.6. PCNN Maximum pooling of segments

Using segmental maximization, a sentence is divided into three paragraphs by given two entities, including the features between the two entities and the features before and after the entities. Segmental maximum pooling divides a sentence instance into three parts according to the position of head entity and tail entity $c_i = \{c_{i,1}, c_{i,2}, c_{i,3}\}$, Maximum pooling is performed for the vector obtained by each convolution kernel $p_{s,j} = \max(c_q), \ 1 \leq i \leq n, 1 \leq j \leq 2$. And you end up with your eigenvector $p_i = \{p_{i,1}, p_{i,2}, p_{i,3}\}$, Then concatenate all the sentence vectors to get the output vector $P = \{p_1, p_2, \ldots, p_n\}$.

### 3.7. Classifier module speed Sentence level attention mechanism combined with entity description

To ease the relationship extraction of remote monitor hypothesis is inevitably marked sample noise problem, this paper adopts the combined entity description layer sentence level attention mechanism of multi-instance learning 25, calculate all in-stances of the same package for (entity) (sentences) and predict tags (forecast), gives the relationship between vector relative degree higher sentence to higher weights.
The sentences that contain the same entity pair are first divided into a package (entity pair) $B_i$, each containing $q$ sentences. This paper introduces Luong attentional calculation formula to calculate the weight:

$$\alpha_i = \frac{\exp(sor_i)}{\sum_{j=1}^{q} \exp(sor_j)}$$  \hspace{1cm} (6)

$$sor_i = W^T_v (\tanh [p_i; v_{relation}]) + b_v$$  \hspace{1cm} (7)

Where $[x_1:x_q]$ represents the concatenation of $x_i$ and $x_q, 1 \leq i \leq q$, $p$ is the eigenvector obtained after piecewise maximum pooling, $v_{relation}$ is the hidden state of the sentence in the attention mechanism (the relation vector in the sentence), $W_v$ is the weight matrix. Finally, the weights of all instance sentences in the package are obtained $\alpha = [\alpha_2, \alpha_3, \ldots, \alpha_q]$, The characteristics of the entire package are then calculated as follows:

$$\bar{p} = \sum_{i=1}^{q} \alpha_i p_i$$  \hspace{1cm} (8)

e.g.: $v(“Beijing”) - v(“China”)$ $\approx v(“London”) - v(“Britain”).$

Select the vector difference of the entity pairs in the sentence $v_{relation} = v(e_1) - v(e_2)$.

Therefore, use the sum of the words vectors between entities to get the vector $v_{relation}$, In sentences with fewer words, the relational-related characteristics of each word can be effectively considered, thus assigning higher attention weights to sentences that actually describe relationships in packages (entity pairs).

3.8. Softmax

In this paper, classifier is used to calculate the confidence of each relationship:

$$o = W_i \bar{p} + b_s$$  \hspace{1cm} (9)

$o$ is the output, $W$ is the weight matrix. Let $B_i$ represent a package (entity pair), $\theta$ is the trainable parameter of the model. Then the conditional probability that the package (entity pair) is the $i$ ontology relation:

$$p(r_i|B_i; \theta) = \frac{\exp(o_i)}{\sum_{j=1}^{N} \exp(o_j)}$$  \hspace{1cm} (10)

The model loss function is shown in the formula:

$$Loss = \sum_{i=1}^{N} logp(r_i|B_i; \theta)$$  \hspace{1cm} (11)

3.9. Ontology constraint layer

An ontology is a special type of term set that provides a wealth of domain knowledge, including layers of entities and relationships between the topmost entities. However, the existing models do not make use of this part of information, resulting in the extraction of some relations do not accord with common sense.

In order to improve the precision rate of the final extracted relation of the model and make the final extracted relation conform to the basic knowledge of this field, ontology constraint layer is added in this
paper to make use of the levels provided by ontology and the relations among levels. Think of it as a constraint: let $T = (E_1, E_2, \ldots, E_n)$, $T$ is the concept set in the ontology, $E_i$ is the concept set, and the following formula is used for extraction:

$$p(r_i) = \begin{cases} p(r_i;\theta) & e_1 \in E_1, e_2 \in E_2, r_i \in R_{E_1,E_2} \\
0 & e_1 \in E_1, e_2 \in E_2, r_i \not\in R_{E_1,E_2} \end{cases}$$

(12)

If entity $e_1, e_2$ in the extracted statement satisfies $e_1 \in E_1, e_2 \in E_2, R_{E_1,E_2}$ represents the relational set of entity set $E_1, E_2$. In the case of $r_i \in R_{E_1,E_2}$, the relationship is retained after softmax layers. Otherwise change its confidence to 0.

Take the Freebase+NYT data set as an example to recognize the sentence "Nevada is a state in the Western, Mountain West, and Southwestern Regions of the United States.", the header entity is "Nevada", and the tail entity is "the United States". Among the candidate relationships obtained by relationship extraction model, the scores of "province" and "State" are relatively high. The meaning is also similar. However, the entity hierarchy given by the ontology shows the relationship "state" existing in the entity pair "[Nevada]" and "[The United States]", and the relationship does not exist in "province". Therefore, through the constraint layer of the ontology, the option "province" should be excluded in the final prediction.

4. Experimental results and analysis

4.1. Data sets and evaluation criteria

The data set Freebase+NYT[27] used in this paper was generated by Riedel in 2010 through heuristic alignment between entity pairs in Freebase knowledge base and New York Times Corpus (NYT). In recent years, it has been widely used in remote monitoring relationship extraction. As shown in Table 1, this paper takes 2017-2018 corpus aligned sentences as training data, and 2019 corpus aligned sentences as test data.

| Table 1 Freebase+NYT data set |
|-----------------------------|
| **The training set** | **The test set** | **value** |
| Sentence | 172448 | 522611 |
| Entity | 96678 | 281270 |
| Relationship | 1950 | 18252 |

In this paper, accuracy (precision, P), recall rate (remember, R) and F value are used to evaluate the performance of the model. The formula is as follows:

$$\text{precision} = \frac{\text{out\_right}}{\text{out\_all}}$$

(13)

$$\text{recall} = \frac{\text{out\_right}}{\text{test\_all}}$$

(14)

$$F\_\text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

(15)

Where out_right represents the number of relationships predicted to be correct in the output, out_all represents the total number of relationships in the output, and test_all represents the total number of relationships in the test set.
4.2. Data sets and evaluation criteria

In this paper, cross-validation is adopted to tune the model parameters. In the process of parameter setting, the training parameters are set as shown in Table 2:

| Model parameters          | Value  |
|---------------------------|--------|
| Vector dimension of words | $d^v = 50$ |
| Position vector dimension | $d^p = 5$ |
| Sentence vector dimension | $d_s = 230$ |
| Initial learning rate     | $\mu = 0.05$ |
| Dropout rate              | $p = 0.05$ |
| Initial value of weight   | $\lambda = 0.01$ |
| Batch_size size           | $b = 120$ |

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.3. Data sets and evaluation criteria

4.3.1. Model comparison analysis

The proposed TPCNNs (s) + the OR model, application of syntactic dependency tree get package in (entity) every word of the sentence and the word after the location of the weight vector multiplication, then stitching position vector input vector, then using PCNN to extract the feature vector of the sentence, introducing combining entity description sentence level attention mechanism, add constraint of entity information ontology layer, improve the model performance.

In order to verify the performance of the proposed relational extraction model TPCNNs(S)+OR, this paper compares it with the following representative remote monitoring models:

1. APCNNs(S)+OR[20]: PCNN(piecewise convolutional neural network) is used to extract feature vectors of sentences, and sentent-level attention mechanism and ontology constraint layer combined with entity description are introduced.

2. APCNNs(S)[20]: Use PCNN (piecewise convolutional neural network) to extract feature vectors of sentences and introduce sentence attention mechanism combining entity description.

3. APCNNs+OR[20]: The feature vector of PCNN (piecewise convolutional neural network) is used to introduce the sentent-level attention mechanism and ontology constraint layer.

4. PCNN+ONE[13]: Segmented convolution of shenlevel network, introducing multi-instance learning, but losing important information in packets (entity pairs).

5. APCNNs[23]: The method of sentence-level attention mechanism is introduced on the basis of PCNN, and entity information is added. However, the calculation of relationship hidden state vector in attention mechanism is not accurate enough.

6. RNN+Attention[28]: The cyclic neural network was used for feature extraction and attention mechanism was added. Although the temporal relation in the sentence was taken into account, the influence of the following words in the sentence on the preceding words was not considered.

7. BiRNN+Attention1. [28]: Bidirectional circulation neural network is used for feature extraction, and sentence bidirectional sequence information is used. However, problems of gradient disappearance and gradient explosion are easy to occur.
It can be seen from the PR curve in Figure 8 that, on the premise of P (accuracy) of 0.80, the R (recall rate) of the TPCNNs(S)+OR model proposed in this paper can reach 0.88, higher than the 0.85 of APCNNs(S)+OR model, which proves that the method proposed in this paper can enhance the coverage rate of entity relationship extraction. Under the condition of the same R (recall rate), the P (accuracy) obtained by the TPCNNs(S)+OR model proposed in this paper is also higher than that obtained by the existing representative methods, which indicates that the accuracy of the relation extraction model is improved by adding the position weight of sentences in the package (entity pair) and the constraint layer of ontology to the syntactic dependency tree in this paper.

It can be seen from FIG. 9 that the comparison between APCNN+Attention model and other models that introduce the Attention mechanism at the sentence level, such as BiRNN+Attention model, shows that the segmented convolutional neural network can better extract the feature information in the task of relation extraction.

The above 7 models were summarized and compared with the TPCNNs(S)+OR model proposed in this paper, and the accurate-recall rate curve was drawn, as shown in Figure 10:
In Figure 10, P (accuracy) and R (recall rate) corresponding to the maximum F value points of each curve are shown in Table 3. It can be seen from the table that TPCNNs(S)+OR has the highest F value of 86.62%. F is higher than the other models. Meanwhile, the accuracy and recall rates reached 84.76% and 88.53%, both higher than other models. To sum up, TPCNNs(S)+OR model has the highest extraction effect in Freebase+NYT data.

In the experimental comparison between APCNN+Attention model and APCNNs(S) model, it can be seen that the sentence-level Attention mechanism combined with entity description can obtain more information in each sentence in the package (entity pair), and the extraction effect of the model is better than that of the sentence-level Attention mechanism. In the comparison experiment between APCNNs(S) and APCNNs(S)+OR model, it can be seen that adding ontology constraint layer can effectively constrain the extracted results, make the extracted relationship more consistent with the facts, and improve the accuracy of the model.

Table 3 Experimental results of each model in Freebase+NYT data set

| Relational extraction method | Recall % | Precision % | F_score % |
|-----------------------------|----------|-------------|-----------|
| TPCNNs(S)+OR               | 84.76    | 88.53       | 86.62     |
| APCNNs(S)+OR               | 83.20    | 86.71       | 84.92     |
| APCNNs+OR                  | 81.35    | 84.67       | 82.98     |
| APCNNs(S)                  | 80.22    | 83.86       | 82.00     |
| APCNNs                     | 78.71    | 82.29       | 80.46     |
| PCNN                       | 73.43    | 80.22       | 76.18     |
| BiRNN+Attention            | 76.92    | 80.53       | 78.68     |
| RNN+Attention              | 75.02    | 83.11       | 78.85     |

5. conclusion
In this paper, a remote supervisory extraction model (TPCNNs(S)+OR) combining syntactic dependency tree and ontology constraint is proposed to solve the problem of echo noise caused by too strong assumption of remote supervision, the model can not only make full use of the relationship between ontology to the real constraint information, and fully considering the relationship between attributes in a sentence context information and the position of the words and the relationship between attribute information, effectively reduce the remote supervision and the assumption of noise problem, improve the model performance. Based on Freebase + NYT experiment data to build training data and testing data, the experimental results show that the proposed TPCNNs (s) + the OR model compared with the existing method of typical model of remote monitoring can be more effective in reducing the noise back problems as well as the interference of the words of information between two entities and relationship extraction model performance is better.

References
[1] Shaoyu, L., Bicheng,L., I.Zhigang,G., et al.: Review of Entity Relation Extracion[J]. Journal of information engineering university, (2016).
[2] Zhong-Hua,M., Chen-Hui,M., Zhi-Yuan,G., et al: Hidden feature extraction for unstructured agricultural environment based on supervised kernel locally linear embedding modeling[J], Advances in Manufacturing, (2018).
[3] Zhong-Hua,M., Chen-Hui,M., Zhi-Yuan,G., et al: Hidden feature extraction for unstructured agricultural environment based on supervised kernel locally linear embedding modeling[J], Advances in Manufacturing, (2018).
[4] Zhou, G., Su, J., Zhang, J., et al.: Exploring Various Knowledge in Relation Extraction[C], Acl, Meeting of the Association for Computational Linguistics, Conference, June, University of Michigan, USA. DBLP, (2005).

[5] Xiyue. Guo, TingTing. He, XiaoHua.Hu, QianJun.Chen: Chinese entity relationship extraction based on syntactic and semantic characteristics[J], Journal of Chinese Information Processing,28(06):183-189, (2014)

[6] Hasegawa, T., Sekine, S., Grishman, R., et al: Discovering Relations Among Named Entities from Large Corpora[M], Spain:416-423, Barcelona, DOI:10.3115/1218955.1219008.

[7] Brin, S.: Extracting Patterns and Relations from the World Wide Web[C], //International Workshop on the World Wide Web and Databases, Springer,Berlin,Heidelberg,(1998).

[8] Craven, M., Kumlien, J.: Constructing Biological Knowledge Bases by Extracting Information from Text Sources[C], International Conference on Intelligent Systems for Molecula-1ar Biology,(1999).

[9] Mintz, M., Bills, S., Snow R, et al.: Distant supervision for relation extraction without la-beled data[C], Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, 1003-1011, Association for Computational Linguistics, (2009)

[10] Craven, M., Kumlien, J.: Constructing biological knowledge bases by extracting information from text sources, 77-86, (1999).

[11] HOFFMANN, R., ZHANG, C., LING, X., et al: Knowledge-based weak supervision for information extraction of overlapping relations[C], Proceedings of the Association for Computational Linguistics:Human Language Technologies.Stroudsburg:Association for Computational Linguistics, 541-550, (2011).

[12] SURDEANUM., TIBSHIRANI,J., NALLAPATI,R., et al.: Multi-instance multi-label learning for relation extraction[C], Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natura Language Learning, 455-465,(2012).

[13] Zeng, D., Liu, K., Chen, Y., et al: Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C], Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing,(2015).

[14] Daojian,Zeng., Kang,Liu., Yubo,Chen., Jun,Zhao.: Distant supervision for relation extraction via piecwise convolutional neural networks[C], /D15-1203: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, September, 2015. Lisbon, Portugal: Association for Computational Linguistics, 1753-1762,(2015).

[15] Yankai,Lin., Shiqi,Shen., Zhiyuan,Liu., Huanbo,Luan, Maosong,Sun.: Neural relation extraction with selective attention over instances[C], P16-1200: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, August 2016, Berlin, Germany, Association for Computational Linguistics, 2124-2133,2016.

[16] Guoliang,Ji., Kang,Liu., Shizhu,He., Jun,Zhao.: Distant supervision for relation extraction with sentence-level attention and entity descriptions[C], AAAI-17: The Thirty-First AAAI Conference on Artificial Intelligence, Hilton San Francisco Union Square, February 4-10, 2017. San Francisco, California, USA, Palo Alto, 3060-3066,(2017)

[17] Che,J,L., Tang,L,W., Deng,S,J., et al.: Distant supervision Chinese relation extraction based on dual attention mechanism. Computer Engineering and Applications[J], Computer Engineering and Applications, 55(20): 107-113 (2019).

[18] Mikolov,T., Sutskever,I., Chen,Kai., et al.: Distributed representations of words and phrases and their compositionality[C], Proc of the 26th International Conference on Neural Information Processing Systems, [S. l.]:Curran Associates Inc., 3111-3119,(2013)

[19] Guoliang,Ji., Kang,Liu., Shizhu,He., Jun,Zhao.: Distant supervision for relation extraction with sentence-level attention and entity descriptions[C], The Thirty-First AAAI Conference on Artificial Intelligence, Hilton San Francisco Union Square, February 4-10, 3060-3066, San Francisco, California, USA, Palo Alto,(2017)
[20] Li Yanjuan, Zang Mingzhe, Liu Xiaoyan, Liu Yang, GUO Maozu: Remote Monitoring Relationship Extraction combining attention mechanism and ontology [J/OL], Computer Science and Exploration, :1-11, (2020)
[21] Li, Peifeng., Zhu, Qiaoming., Zhang, Wei.: A Dependency Tree Based Approach for Sentence-Level Sentiment Classification[J],(2011)
[22] Su Jin-tin, Ouyang Zhifan, YU Shanshan.: Sentence Attribute affective Classification based on dependency tree and Distance attention [J], Computer Research and Development, 56(08):1731-1745,(2019)
[23] Lin,Y., Shen,S., Liu,Z., et al.: Neural Relation Extraction with Selective Attention over Instances[C], Meeting of the Association for Computational Linguistics,(2016)
[24] Liu,J., Li,T., Xie,P., et al.: Urban big data fusion based on deep learning: An overview[J], Information Fusion, 53: 123-133,(2020)
[25] Ji,G., Liu,K., He,S., et al.: Distant supervision for relation extraction with sentence-level attention and entity descriptions[C], AAAI, 3060,(2017)
[26] Srivastava,N., Hinton,G., Krizhevsky,A., et al. :Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J], Journal of Machine Learning Research, 15(1):1929-1958,(2014).
[27] Riedel,S., Yao,L., Mccallum,A.: Modeling Relations and Their Mentions without Labeled Text[C], European Conference on Machine Learning & Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, (2010).
[28] Richard,Socher., Brody,Huval., Christopher,D.,Manning., Andrew,Y., Ng.: Semantic compositionality through recursive matrix-vector spaces[C], EMNLP-CoNLL’12: Proceeding and Computational Natural Language Learning, Stroudsburg, 1201-1211,PA,USA: Jun’ichi Tsujii, July 12-14,(2012)