Chinese Text Relation Extraction with Multi-instance Multi-label BLSTM Neural Networks

1st Liubo Ouyang Department of Software Engineering

Hunan University ChangSha, China oylb@hnu.edu.cn 2nd Hui Tang Department of Software Engineering Hunan University Changsha, China thlthl2525@hnu.edu.cn 3rd Guangyi Xiao Department of Software Engineering Hunan University Changsha, China gyxiao@hnu.edu.cn

Abstract-Recently, deep learning models have emerged as powerful tools for relation extraction. However, little work has been done on relation extraction for the Chinese language. One major challenge for relation extraction in Chinese texts is that Chinese sentences have no obvious word segmentation. This ambiguity increases the possibility of word segmentation errors. Another challenge is the lack of broad-scale Chinese text datasets. In this paper, we propose an attention-based multi-instance multi-label bidirectional long short-term memory network for distantly supervised Chinese relation extraction. Our model takes Chinese character embeddings and position embeddings as input without Chinese word segmentation errors. Then, the attention mechanism is used to extract richer Chinese character and sentence features. Finally, we handle the multi-label nature of relation extraction by using multi-label loss functions in the neural network classifier. Based on the idea of distant supervision, we constructed a new dataset for relation extraction in Chinese texts. Experiments on this dataset show that our method has achieved relatively high performance, and that the proposed network architecture is suitable for Chinese relation extraction. Furthermore, we also ran experiments on a popular English benchmark dataset, and the results show that our method is superior to some existing methods.

Index Terms—Chinese relation extraction, distant supervision, attention, bidirectional long short-trem memory network(BLSTM).

I. INTRODUCTION

Relation extraction is the detection and identification of semantic relations between natural language text entities [1]. The relation extraction problem can be formally described as follows. Given a sentence s and two entities, e1 and e2, in s, predict the relationship type r of e1 and e2 in the sentence s. The candidate set of r is a predefined relation set R, and the output is usually a triple (e1, e2, r) [2]. As one of the key tasks of natural language processing (NLP), relation extraction has high significance for many applications of NLP, such as question answering and knowledge graphs [3].

Distant supervision is proposed to automatically generate labeled training data by aligning knowledge bases and text for relation extraction [4]. The main idea of distant supervision is that if two entities have a relationship in the knowledge base, then all sentences containing the two entities will represent this relationship. Distant supervision solves

DOI reference number:10.18293/SEKE2019-106.

the problem of constructing labeled training data and saves human resources needed for manually labeling data. However, since a sentence containing two entities does not necessarily represent a corresponding relationship, a large amount of noise data is inevitably introduced. Therefore, distant supervision is modeled as a multi-instance multi-label classification problem [5]- [7]. In recent years, deep learning models have been applied in relation extraction [8]. The most commonly used models for relation extraction tasks include recursive neural networks (RecNN) [9], convolutional neural networks (CNN) [10], recurrent neural networks (RNN) [11]. Moreover, LSTM network is an improved variant of RNN that has been widely applied to natural language processing tasks and has achieved good performance [12]. Adding an attention mechanism to the model and weighting the data sequence can effectively improve sequence learning and boost the system performance [13] [14].

The research on relation extraction for Chinese texts is relatively less common than that for English ones [15]. This scarcity can be ascribed to three main difficulties of Chinese relation extraction. Firstly, there is no obvious separation between Chinese words; Chinese words are composed of characters whose combinations are highly complex and ambiguous. Secondly, current Chinese word segmentation systems still have considerable errors that introduce noise into the relation extraction task. Finally, there is a lack of a Chinese corpus for relation extraction.

In this paper, we propose an attention-based multi-instance multi-label bidirectional long short-term memory network (ATT+MIML+BLSTM) for distantly supervised Chinese texts relation extraction. It is sufficient to use raw Chinese sentences as input. The BLSTM and character-level attention modules are used to obtain the important semantic information in each sentence, and then the final sentence vector representation is obtained through sentence-level attention. Also, the multilabel loss function is used in the network model to deal with overlapping relations. In view of the lack of a Chinese dataset, and based on the distant supervision concept, we used CN-DBpedia of Fudan Knowledge Factory ¹ to obtain the identified entity pairs, and aligned them in the SogouCS2012

¹http://kw.fudan.edu.cn/

news text corpus ² to construct a large Chinese character relationship dataset. The model proposed in this paper is evaluated on this constructed dataset as well as on an English benchmark dataset. The results show that the proposed model is suitable for relation extraction in Chinese texts and achieves good performance.

The contributions of this paper are summarized as follows:

- Proposing a deep learning model for distantly supervised relation extraction in Chinese texts.
- Using Chinese character vectors as model inputs to avoid introducing Chinese word segmentation noise into the relation extraction process.
- Making full use of the information of each word in the sentence and all the sentence information with the same entity pair, and handling the overlapping relationships of entity pairs.
- Performing extensive experiments on Chinese and English benchmark datasets showing the suitability and good performance of our model for relation extraction in both languages.

II. RELATE WORK

Over the years, many methods for relation extraction have been proposed. These methods can be mainly divided into three categories: supervised, semi-supervised and unsupervised approaches. Supervised approaches treat relation extraction as a classification task and typically exhibit better performance in comparison to other approaches. However, supervised methods are very time consuming and labor intensive since they require a lot of labeled data. Distant supervision can solve this problem through automatic data labeling. However, distant supervision can cause false labeling problems. Riedel et al. [5]models distant supervision as a multi-instance singlelabel problem and selects for each entity pair. Furthermore, Hoffmann et al. [6] and Surdeanu et al. [7] noted that there might be multiple relations between entities and hence they cast relation extraction as a multi-instance multi-label learning problem.

Recently, with the emergence of deep learning, many scholars have begun to use deep learning to automatically learn features. In NLP, deep learning methods are mainly based on learning the distributed representation of each word, which is also called a word embedding [16]. Zeng et al. [17] proposed a convolutional neural network for relational classification. Furthermore, Zeng et al. [18] combined at least one multiinstance learning scheme with a neural network model to extract relationships using distant supervision. Zhou et al. [19] used a neural attention mechanism combined with a bidirectional long-term memory network (BLSTM) to obtain important semantic information in sentences. Lin et al. [14] showed that the establishment of a sentence-level attention mechanism for dynamically calculating the weights of multiple instances can be very effective in distantly supervised data. Jiang et al. [20] used the maximum pool across sentences to select features in different sentences, then aggregated the most important features into the vector representation of each entity pair. In addition, they addressed relation extraction as a multi-label problem. Jat et al. [21] combined multiple complementary models to improve relational extraction, and introduced a new distantly supervised dataset that eliminated the test data noise present in all previous benchmark datasets. Feng et al. [22] applied reinforcement learning to distantly supervised relationship extraction. Relevant side information from KB has been utilized for relation extraction [23] [24].

In the study of Chinese texts relation extraction, lexical and syntactic features are usually used to extract feature vectors, and the classifier SVM can be used for Chinese texts relation extraction [25]. The performance of Chinese relation extraction has been improved by clustering and pattern matching the feature vectors extracted by dependence relationship and parts of speech labeled in relational schema [26]. In the works of [27] and [28], all can reveal that performance of Chinese relation extraction based on kernel function has obtained significant improvement. However, The need for amounts of manually annotated corpora hinders the application of deep learning methods in Chinese texts relation extraction.

At present, relation extraction research and experiments are mainly focused on English. This paper focuses on distant supervision for relation extraction in Chinese texts. Based on the advantages of previous models, this paper proposes a multi-instance multi-label BLSTM model equipped with an attention mechanism. This model can obtain the rich semantic information of the Chinese sentences and handle overlapping relations.

III. MODEL

As shown in Fig.1, the proposed ATT+MIML+BLSTM model consists of three main parts:

- Sentence Representation: Enter raw characters of the input sentence into the BLSTM model, and use the character-level attention mechanism to weight the output of each time step of the BLSTM network. Merge the character-level features of each time step into a sentence-level feature vector.
- *Entity-pair Representation:* Use sentence-level attention to give different weights to different statements, implicitly discarding some noise statements.
- *Multi-label Classification:* Multi-label classification is performed by using multiple binary classifiers corresponding to a class of relationships.

A. Sentence Representation

Sentence-level features are designed to construct a distributed representation for each sentence. As shown in Fig.2, we first convert the Chinese characters in the sentence into real-valued vectors. The higher features of the sentence are then extracted by BLSTM. The final character-level attention weights the output of each time step of the BLSTM The character-level features of each time step are then merged into a sentencelevel feature vector.

²http://www.sogou.com/labs/



Fig. 1. The architecture of ATT+MIML+BLSTM used for distant supervised relation extraction.

Input Representation Layer. The input for each sentence consists of two embeddings:

Character Embedding: Chinese words are composed of characters, the combinations of which are very complicated. As well, there is no obvious separation between Chinese words. Chen et al. [29] observed experimentally that character features are more suitable for Chinese relation extraction tasks than word features. Based on this observation, we use character embeddings to avoid introducing word segmentation errors into the relation extraction process. The model takes as an input each raw Chinese character in a Chinese sentence s. Therefore, given a sentence s = $\{x_1, x_2, \dots, x_m\}$ consisting of m Chinese characters, we convert each Chinese character into a real-value vector by looking up the pre-trained embedded matrix $V \in \mathbb{R}^{d_w \times |V|}$ (where d_w is the dimension of the character embedding and |V| is vocabulary size).

Position Embedding: Position embeddings are used to represent the structural information of sentences, and two different dimensional position vectors are constructed by the relative distances between each Chinese character and the two entities e1 and e2.

We concatenate all the character embeddings and position embeddings to get a sequence of vectors $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ (where $\mathbf{w}_i \in \mathbb{R}^d$, $d = d_w + 2 \times d_p$ and d_p is the dimension of the position embedding) and set such a sequence as the model input.

BLSTM Layer. The LSTM model can effectively alleviate the long-distance dependence problem of RNN and CNN, and it has been improved and promoted recently by Graves [30]. In many problems, LSTM has achieved considerable success and has been widely used.

In LSTM, the forget gate f_t determines how much information is discarded in the cell state. The gate has three inputs: x_t is the input of the current time step, h_{t-1} is the output of the previous LSTM cell, and c_{t-1} is the memory of the previous cell.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$
(1)



Fig. 2. Detailed structure of the sentence representation.

where, W_{xf} , W_{hf} , W_{cf} , and b_f are weight matrices.

The input gate i_t determines what information will be updated and has the same input as the forget gate f_t :

$$i_t = \sigma (W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$
(2)

where , W_{xi} , W_{hi} , W_{ci} , and b_i are weight matrices.

Then create a new candidate value vector \tilde{c}_t that will be added to the state:

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{cc}c_{t-1} + b_c)$$
(3)

where , W_{xc} , W_{hc} , W_{cc} , and b_c are weight matrices. Update current cell status c_t :

$$c_t = i_t \tilde{c}_t + f_t c_{t-1} \tag{4}$$

Finally, how much information is controlled by output gate o_t should be entered into the next LSTM cell.

$$o_t = \sigma (W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_{t-1} + b_o)$$
(5)

$$h_t = o_t \tanh(c_t) \tag{6}$$

where , $W_{\rm xo}$, $W_{\rm ho}$, $W_{\rm co}$, and b_o are weight matrices.

As can be seen from Figure 2, the BLSTM layer contains forward and reverse LSTM networks. The output for the i^{th} Chinese character from the BLSTM layer is combined by the forward and reverse outputs by element-wise summation:

$$h_i = \left[\overrightarrow{h_i} \oplus \underbrace{h_i}_{\leftarrow}\right] \tag{7}$$

Character-level Attention Layer. Following the approach of [20], we use character-level attention to capture important information in sentences and improve the accuracy of sentence representation. The vector sequence $H = \{h_1, h_2, \dots, h_m\}$ output by the BLSTM layer is weighted to obtain the final representation of the sentence.

First apply a non-linear activation function on h_i :

$$N_i = \tanh(h_i) \tag{8}$$

Then calculate the weight of the Chinese characters in the sentence:

$$\alpha_i = softmax(W^T N_i) \tag{9}$$

where, W^T is the training parameter.

The vector representation r of the sentence is obtained by weighting the N_i :

$$r = \sum_{i=1}^{m} \alpha_i N_i \tag{10}$$

Finally, a non-linear activation function is applied to r to get the final representation of the sentence H^* .

$$H^* = \tanh(r) \tag{11}$$

B. Entity-pair Representation

Distant supervision produces a lot of noise or mislabeled data, and direct use of supervised methods to classify relationships is very ineffective. In order to solve the problem of mislabeling, this paper proposes the construction of sentence-level relational attention on multiple instances by using the selective attention mechanism to weigh sentence vectors and weaken the weight of noisy instances dynamically [14]. Suppose there is a set S containing n sentences of the same entity pair, and the set S is represented as having a real-valued vector, i.e. $S = \{H_1^*, H_2^*, \cdots, H_n^*\}$, where H_i^* is the representation of the sentence obtained in the subsection A.

First, calculate the degree to which a sentence H_i^* matches the corresponding relationship.

$$D_i = H_i^* \cdot A \cdot l \tag{12}$$

where A is a weighted diagonal matrix and l is the vector representing the relationship. Therefore, the size of D_i depends on the size of the mapping of H_i^* on l, and sentences that are more closely related to the entity relationship can achieve larger values. Then, we can get the weight β_i :

$$\beta_i = softmax(D_i) \tag{13}$$

Then for the sentence set S, it can be calculated as the weighted sum of all the sentences in the set:

$$S = \sum_{i} \beta_i H_i^* \tag{14}$$

C. Multi-label Classification

In this paper, we formalize distant supervision as a multiinstance multi-label learning problem. In this section, we will handle the overlapping relationship of entity pairs. The sentence set vector S obtained in the subsection B, and then get o through a layer of network:

$$o = MS + b \tag{15}$$

where M is the weight matrix of all relational vectors and b is a bias. Thus o represents the confidence score for each relation label.

Then, we use multiple binary functions to do multi-label classification. In particular, we calculate the probability of each relationship. The relationship label is considered accurate if the relationship probability exceeds a certain threshold.

$$p_i = sigmoid(o_i), i = \{1, 2, \cdots, k\}$$
 (16)

where k is the number of relation labels.

We set the binary label vector y to represent the set of true relationships between pairs of entities, where 1 represents a relation in the set and 0 otherwise. Finally, we use the cross entropy of the sigmoid function as the loss function:

$$loss = -\sum_{i=1}^{k} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$
(17)

where $y_i \in \{0, 1\}$ is the true value on label *i*.

We train the model in an end-to-end manner. We use Adam to optimize the loss function [31]. In the training phase, we used a dropout mechanism in the BLSTM layer to prevent overfitting [32]. For the testing phase, our method selects a relation with a probability of more than 0.5 as a predicted label.

IV. EXPERIMENTS

A. Experiments on Chinese Character Relationship Dataset

Building Chinese datasets. Using distant supervision, we can automatically construct a Chinese dataset, which avoids labor instensive of manually labeling data. The details of these steps are as follows.

Step 1: Obtain entity pairs with a defined relationship. The large-scale structured encyclopedia CN-DBpedia, developed and maintained by the Fudan University Knowledge Factory Laboratory, is the earliest and the largest open Chinese knowledge extraction system.We set up a list of seed names, use the free API provided by the Fudan Knowledge Factory to obtain the corresponding personal relationships, and then add the character entities not included in the entity list to the list, and iterate over and over again.

Step 2: Align the entity pairs with certain relationships with the Chinese text corpus. The Chinese text corpus uses SogouC-S from Sogou Lab, which is one of the most comprehensive Chinese text corpus resources. Pure text is obtained by data preprocessing through the corpus. The text corpus is aligned with the entity pairs of the first step to get the statements containing the entity pairs.

Finally, the dataset contains eight kinds of relationships (cooperation, friends, couples, parenthood, lovers, faculty-student, brother and sister, others). We divide the training set and the test set according to a ratio of 9:1. The training set contains 220,160 sentences and the test set contains 24,463 sentences.

Experimental settings and evaluation metrics. Chinese character embeddings are trained on the Chinese Wikipedia corpus by the word2vec tool³. We set the dimension of character embedding to 100 and the dimension of position embedding to 5. Position embeddings are randomly initialized with uniform distribution between [-1,1]. At BLSTM layer,

³https://code.google.com/p/word2vec/



Fig. 3. Performance comparison of our method with three baselines for Chinese dataset.

TABLE I Comparison of P@N results between our model and other models.

	PCNN	BGWA	ATT+CNN	ATT+MIML+BLSTM
P@100	0.87	0.91	0.95	0.97
P@200	0.845	0.885	0.93	0.955
P@300	0.817	0.85	0.923	0.946
Mean	0.844	0.882	0.934	0.957

the number of LSTM hidden units is set to 230. Meanwhile, we use a batch of 64 entity pairs, set learning rate to 0.001 and dropout keep probability to 0.5.

Following previous work [14] [18], We compare the performance of each model with the aggregate curves Precision/Recall(PR) curves and Precision@N(P@N).

We choose the following three deep learning models as baselines: (1) **PCNN**: A piecewise max-pooling over CNN based relation extraction model. (2) **ATT+CNN**: A CNN based model with sentence-level attention. (3) **BGWA**: A piecewise max-pooling over bidirectional gated recurrent unit based model with word-level attention. They are all superior to traditional methods and are also significant works for relation extraction.

Experimental results. The curves in Fig.3 show that our model has a relatively high accuracy and recall rate compared to other models when we perform Chinese relation extraction. In other words, under the same recall rate, the accuracy of Chinese relation extraction using ATT+MIML+BLSTM is higher than other models. We observed that the PR curve of the neural network method (ATT+CNN, BGWA, ATT+MIML+BLSTM) that introduces the attention mechanism is significantly higher than the ordinary neural network method PCNN. It can be seen that the attention mechanism can improve the performance of the model. In addition, our model is superior to the CNN model which also adopts sentence-level attention, which shows that BLSTM combined with word attention can obtain richer semantic information, and considering multi-label problem can



Fig. 4. Performance comparison of our method with five popular methods for English benchmark dataset.

TABLE II Statistics of the NYT10 dataset.

	Sentences	Entity Pairs	Relations
Training	522611	281270	53
Testing	172448	96678	53

improve model performance. Table I shows the results using the P @ N metric. Similar to Figure 4, our approach is superior to other methods overall. And when N is smaller, our method accuracy is marginally higher than others.

B. Experiments on the English Benchmark Dataset

In order to evaluate the performance of our model on other languages, we conducted a comparative experiment on the English dataset proposed by Riedel in 2010 [5]. The dataset is generated by the heuristic matching of Freebase and the New York Times(NYT), and is a popular benchmark dataset. As shown in Table II, the dataset contains 53 relationships (including "NA", indicating that there is no relationship between entity pairs). We used sentences from 2005 to 2006 as training data and sentences from 2007 as test data.

The parameter settings for the English experiments are like those of the Chinese ones, except that the input is word embedding, the word embedding dimension is 50, and the batch size is 50.

We compare our model with three traditional methods(Mintz [4], Multir [6] and MIML [7]) and two popular neural-based methods(ATT+CNN and BGWA). They are major works for English relation extraction based on distant supervision.

Fig.4 clearly shows that the PR curve of our model (AT-T+MIML+BLSTM) is above those of the other models, so our model is not only superior to traditional methods but also neural-based methods. Since the size of the Chinese dataset we constructed is smaller than the English benchmark dataset and the amount of noise data is different, there is a small error in the experimental results on the two datasets. However, we

can still conclude that our model is superior to other models, especially for Chinese relation extraction.

V. CONCLUSION

In this paper, we propose a multi-instance multi-label neural network model based on the attention mechanism for Chinese relation extraction. It can not only obtain the rich semantic information of the sentence, but also consider the multirelationship problem of the entity pair. Experiments on a Chinese dataset based on distant supervision prove that the neural network model combining attention mechanism and multi-label learning can achieve good results, and the proposed method has better performance than most existing methods on the benchmark English dataset. In the future, we hope to build a large-scale and standardized Chinese corpus, and further study how different loss functions affect the performance of the model.

REFERENCES

- [1] N. Bach and S. Badaskar, "A review of relation extraction," *Literature review for Language and Statistics II*, vol. 2, 2007.
- [2] I. Hendrickx, N. K. Su, Z. Kozareva, P. Nakov, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals," in *Workshop* on Semantic Evaluations: Recent Achievements & Future Directions, 2009.
- [3] Q. Zhang, M. Chen, and L. Liu, "A review on entity relation extraction," in 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), vol. 00, Dec. 2018, pp. 178–183. [Online].
- [4] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data," in *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.* Association for Computational Linguistics, 2009, pp. 1003–1011.
- [5] S. Riedel, L. Yao, and A. McCallum, "Modeling relations and their mentions without labeled text," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 148–163.
- [6] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proceedings of the 49th Annual Meeting* of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011, pp. 541–550.
- [7] M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning, "Multiinstance multi-label learning for relation extraction," in *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. Association for Computational Linguistics, 2012, pp. 455–465.
- [8] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [9] C. Goller and A. Kuchler, "Learning task-dependent distributed representations by backpropagation through structure," *Neural Networks*, vol. 1, pp. 347–352, 1996.
- [10] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [11] J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," *Machine learning*, vol. 7, no. 2-3, pp. 195–225, 1991.
- [12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

- [14] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural relation extraction with selective attention over instances," in *Proceedings of the* 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, 2016, pp. 2124–2133.
- [15] W. Zirui, M. Fang, and J. Libiao, "Review of chinese entity relation extraction," in 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), Aug 2017, pp. 633–637.
- [16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [17] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of COLING* 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014, pp. 2335–2344.
- [18] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," in *Proceedings* of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1753–1762.
- [19] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attentionbased bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 207–212.
- [20] X. Jiang, Q. Wang, P. Li, and B. Wang, "Relation extraction with multiinstance multi-label convolutional neural networks," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 1471–1480.
 [21] S. Jat, S. Khandelwal, and P. Talukdar, "Improving distantly supervised
- [21] S. Jat, S. Khandelwal, and P. Talukdar, "Improving distantly supervised relation extraction using word and entity based attention," 2018.
- [22] J. Feng, M. Huang, L. Zhao, Y. Yang, and X. Zhu, "Reinforcement learning for relation classification from noisy data," *CoRR*, vol. abs/1808.08013, 2018.
- [23] G. Ji, K. Liu, S. He, and J. Zhao, "Distant supervision for relation extraction with sentence-level attention and entity descriptions," pp. 3060–3066, 2017.
- [24] S. Vashishth, R. Joshi, S. S. Prayaga, C. Bhattacharyya, and P. Talukdar, "Reside: Improving distantly-supervised neural relation extraction using side information," 2018.
- [25] W. Che, T. Liu, and S. Li, "Automatic entity relation extraction," *Journal of Software*, vol. 19, no. 2, pp. 2–7, 2005.
- [26] L. Yu and K. Zhou, "A dynamic local path planning method for outdoor robot based on characteristics extraction of laser rangefinder and extended support vector machine," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 02, p. 1659004, 2016.
- [27] H. Zhang, S. Hou, and X. Xia, "A novel convolution kernel model for chinese relation extraction based on semantic feature and instances partition," in *Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on*, vol. 1. IEEE, 2012, pp. 411–414.
- [28] A. Yang, Y. Du, and Q. Meng, "Extracting personae interactive relation in chinese microblog based on an improved dependency trigram kernel," *DEStech Transactions on Engineering and Technology Research*, no. ICMITE2016, 2016.
- [29] Y. Chen, D. Zheng, and T. Zhao, "Chinese relation extraction based on deep belief nets," *Journal of Software*, vol. 23, no. 10, pp. 2572–2585, 2012.
- [30] A. Graves, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, 2013.
- [31] D. Kinga and J. B. Adam, "A method for stochastic optimization," in International Conference on Learning Representations (ICLR), vol. 5, 2015.
- [32] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.