IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019

Distantly Supervised Relation Extraction through a Trade-off Mechanism

Jun Ni*, Yu Liu*, Kai Wang*, Zhehuan Zhao*, Quan Z. Sheng†
*School of Software, Dalian University of Technology, Dalian 116620, Liaoning, China
†Department of Computing, Macquarie University, Sydney, NSW 2109, Australia
Email: nijun@mail.dlut.edu.cn, yuliu@dlut.edu.cn, kai_wang@mail.dlut.edu.cn
z.zhao@dlut.edu.cn, michael.sheng@mq.edu.au

***Abstract*—Distantly supervised relation extraction can label large amounts of unstructured text without human annotations for training. However, distant supervision inevitably accompanies with the wrong labeling problem, which can deteriorate the performance of relation extraction. What's more, the entity-pair information, which can enrich instance information, is still underutilized. In the light of these issues, we propose TMNN, a novel Neural Network framework with a Trade-off Mechanism, which combines the feature of text and entity pair on the sentence level to predict relations. Our proposed trade-off mechanism is a probability generation module to dynamically adjust the weights of text and corresponding entity pair for each sentence. Experimental results on a widely used dataset show that the proposed method reduces the noisy labels and achieves substantial improvement over the state-of-the-art methods.**

I. Introduction

Relation extraction is widely used in many tasks such as Knowledge Base Construction [1], [2], Information Retrieval [3], and Question Answering [4], [5]. It is usually defined as the task of predicting attributes and relations for entities in a sentence [5]–[7]. For example, (Apple, $/business/person/company$, Steve Jobs) is a relational fact in Knowledge Bases (KBs). A relation classifier is used for predicting the relation $/business/person/company$ from a given sentence like "Steve Jobs is the chief of Apple".

A major challenge in relation extraction is the lack of large-scale, real-world training data. Distant supervision [8] has been proposed to solve this problem because it can get large-scale labeled training data by aligning KBs to text corpus automatically. However, distant supervision always suffers from wrong labeling problem [9]–[11], such as false negative examples. False negative examples usually are caused by the relations of entity pairs missed from KBs and are labeled as "Not a relation" (NA). To alleviate the wrong labeling problem, significant research efforts have been made. Zeng et al. [12] proposed a multi-instance strategy to alleviate the influence of wrong labeling. This method assumes that at-least-one sentence can express the relation of entity pair and then only selects the highest confidence sentence for each entity pair to train. Lin et al. [13], [14] considered all the sentences with various attention strategies. The authors introduced sentence-level attention and multi-lingual attention with Convolutional Neural Networks (CNNs), which assign each sentence with different weights. Liu et al. [15] further exploited entity-pair representation and the confidence of the original label to obtain a new label, where the confidence is calculated by all related sentences for the entity pair [13].

As can be seen, most existing studies [12]–[14] reduce noises by leveraging raw sentence information. Only some research [15] has imported the corresponding entity pair as the only adjuster to correct those noisy instances for multi-instance methods. However, they ignore the fact that the same entity-pair information has different influence on different sentences in a bag. For example, the sentence "Obama is indeed from America" reflects "Nationality" relation between "Obama" and "America". However, it may be regarded as a NA sentence, because its structure is far from the "X of Y" structure which stands for the "Nationality" relation type in most positive instances. At this time, we correct the wrong label more dependent on entity-pair semantics, compared with these sentences containing the "X of Y" structure. But only one adjuster on the entity-pair level may not adapt to every relevant sentence. Therefore, considering that the entity pair like $\langle Obama, America\rangle$ having different meanings in different contexts may cause wrong labeling problem, we propose adding the entity-pair feature on the sentence level to enrich training information. In addition, we also design a trade-off mechanism to dynamically adjust the impact of the original feature and entity-pair feature for each sentence to alleviate the wrong labeling problem.

In this paper, we propose TMNN, a novel neural network framework for distantly supervised relation extraction through a trade-off mechanism. First, our model calculates the sentence representation and the entity-pair representation for each sentence respectively, where the sentence part is based on Bidirectional Gated Recurrent Unit (Bi-GRU) with word-level attention and the entity-pair part uses a full-connection encoder. Second, we design a novel trade-off mechanism to balance the weights of sentence feature and entity-pair feature for each sentence and combine these feature representations as the instance representation. Finally, the model adopts instance-level attention for each instance representation to predict relations. To evaluate the effectiveness of TMNN, we apply the model on a benchmark NYT-Freebase dataset [10]. Experimental results show that our proposed method achieves improvements over the state-of-the-art methods.

The rest of our paper is organized as follows. We outline

978-1-7281-2009-6/$31.00 ©2019 IEEE paper N-19163.pdf

related work about distantly supervised relation extraction in Section II. Section III gives a detailed description of TMNN model. Then, Section IV evaluates the performance of TMNN and analyzes the experimental results. Finally, we summarize this paper and present our future work in Section V.

II. Related work

Relation extraction has been one of the hot issues in Natural Language Processing (NLP). Traditional full-supervised learning methods are difficult to achieve high efficiency due to high cost of manual labeling. Thus, researchers have proposed many relation extraction methods, such as bootstrapping [16] and distant supervision [8]. Among them, distant supervision relies on KB to achieve annotation automatically and rapidly. Since it is convenient to label, the method is widely used. However, wrong labeling problem caused by alignment text corpus is difficult to solve [9]–[11].

To alleviate the wrong labeling problem, Bunescu and Mooney [17] connected weak supervision with multi-instance learning and extended it to relation extraction. Specifically, sentences as instances are grouped into a set of bags to label. Multi-instance learning considers the reliability of the label for each bag. If at-least-one instance in the bag is positive, a bag is labeled positive. Using these labels to train, relation extraction can learn a classification function that can predict the labels of bags in the testing data. However, feature-based relation extraction methods strongly rely on the features generated by NLP, which may cause error propagation.

With the rise of deep learning, Zeng et al. [12], [18] first combined at-least-one multi-instance learning with neural network model to extract relations on distantly supervised data, but the at-least-one method may lose a large amount of rich information containing in those neglected sentences. Different from the at-least-one method, attention strategy can help model to learn input by supplying different weights. A machine translation task [19] tried to apply attention firstly in the field of NLP and got great feedback. Inspired by this, Lin et al. [13] proposed sentence-level attention over multiple instances, which can utilize sentences information thoroughly.

Although these methods achieve great success, most of them only focus on text information, not taking the corresponding entities feature into consideration. Liu et al. [15] began to consider entity-pair information as a weight on the bag-level to correct those wrong labels, which is obtained by calculating the combination of sentences in the multi-instance bag. One concern is that an entity pair may play different roles in different sentences. After that, Lei et al. [20] considered sentence feature and the corresponding entity-sequence feature consisting of all mentioned entities, and train them respectively on the sentence level. Entity-sequence can gain robust superiority, but redundant entities information compared with entity-pair also may bring in noises.

In this study, we propose TMNN, a novel distantly supervised relation extraction method, which attempts to integrate entity-pair feature from triples with text information on the sentence level, to alleviate the wrong labeling problem. TMMN has two advantages. First, it combines text information and the corresponding entity-pair information into instance information without external sources and outperforms most of the state-of-the-art methods. Second, motivated by Pointer-generator Network [21], TMNN proposes a trade-off mechanism to dynamically adjust the influence of entity-pair representation and text representation on the sentence level.

III. Methodology

Our goal is to predict relation r for the entity pair $\langle e_1, e_2 \rangle$. In this section, we describe the procedure that handles the sentences in a bag to achieve our goal. This procedure includes four main parts: Input Representation, Bag Encoder, Trade-off Mechanism, Instance-level Attention and Optimization. Figure 1 illustrates the architecture of our approach.

A. Input Representation

As Figure 1 shows, given a set of sentences $S = \{s_1, s_2, s_3, \ldots, s_m\}$ and the target entity pair $\langle e_1, e_2 \rangle$, we adopt multi-instance learning to take a bag of sentences containing the same entity pair as input and compute the probabilities for each relation as output. For each training module, the bag will be split. The inputs of the model are raw words of the sentence s and the target entity pair $\langle e_1, e_2 \rangle$ from the bag. Then the sentence s, the head entity e_1 and the tail entity e_2 are mapped to vectors through an embedding layer respectively.

1) Sentence Embeddings:

a) ***Word Embeddings****:* For each sentence consisting of n words $s = \{x_1, x_2, x_3, \ldots, x_n\}$, the word x_i is the i^{th} word in the sentence, which can be mapped to a real-valued vector v_i from an embedding matrix $V_w \in R^{d_w \times |V|}$, where V is a fixed-sized vocabulary and d_w is the size of word embedding.

b) ***Position Embeddings****:* In relation extraction, the word nearby the target entity is usually informative, especially some trigger words. It may determine the relation for entity-pair. So we use position embeddings specified by entity pairs following Zeng et al. 's work [18]. Position embeddings request computing the relative distances of the current word to the head entity e_1 and the tail entity e_2. Then, we transform the relative distances to two entities into real-valued vectors p_i from a randomly initialized position embedding matrix $V_p \in R^{d_p \times |P|}$ separately, where d_p is the size of position embeddings and P is the fixed-size distance set.

Finally, we concatenate word vector v_i and position vector p_i as a sentence embedding S_i, $S_i \in R^{d_w + 2d_p}$, and then input S_i into the word-sequence encoder from Bag Encoder.

2) Entity-pair Embeddings: According to the property of word vector, vector representations between similar words are close to each other. Following Ji et al. [22] and Yang et al. [23], we believe entity-pair representation can help model to distinguish different categories of relational facts. Therefore, we encode e_1 and e_2 into vector representations looking up from the same embedding matrix V_w used in word embeddings. Finally, we calculate a relation representation V_r using point-wise multiplication for two entity vectors $\langle e_1, e_2 \rangle$.

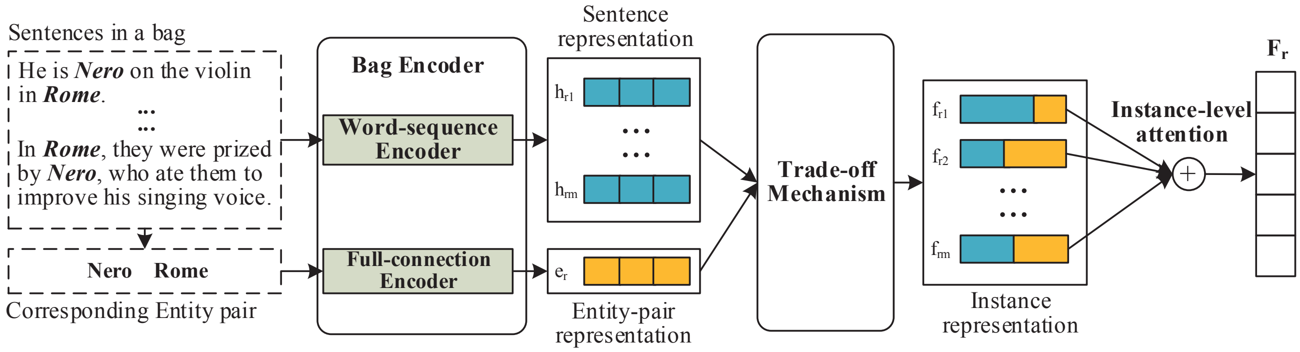

Fig. 1. The overview framework for TMNN.

B. Bag Encoder

Bag Encoder architecture contains a word-sequence encoder using Bidirectional Gated Recurrent Unit (Bi-GRU) with word-level attention and a full-connection encoder. Gated Recurrent Unit (GRU) is regarded as a simplified version of Long Short-Term Memory (LSTM) because GRU inherits the merits of gating mechanism with less parameters [24]. However, GRU networks process sequences in temporal order neglecting future text. Bi-GRU can exploit information both from the past and the future because the network contains two sub-networks for the left and right sequence context, which are forward and backward propagation respectively. Hence, we utilize Bi-GRU with word-level attention to get high level feature from each sentence. We concatenate each forward state h_i^f and backward state h_i^b for each word to encode patterns in the sentence s_i into a hidden representation of h_i. By calculation, the output of the i^{th} word is shown in the following equation:

$$h_i = [\overrightarrow{h_i^f} \oplus \overleftarrow{h_i^b}] \tag{1}$$

Here, we use element-wise sum to combine the forward and backward outputs.

Inspired by Zhou et al. [25], we use word-level attention to integrate word representations into sentence representation. We define a matrix $H \in R^{w_g \times |t|}$, which concatenates all output vectors $[h_1, h_2, h_3, \ldots, h_t]$ calculated by Bi-GRU in Equation (1), where w_g is the dimension of the word vector in Bi-GRU and t is the sentence length. The representation γ of the sentence is formed by a weighted sum of these output vectors as follows:

$$M = \tanh(H) \tag{2}$$

$$\alpha = softmax(w^T M) \tag{3}$$

$$\gamma = H\alpha^T \tag{4}$$

where w^T is a trained and transposed parameter vector. The dimension of w, α, γ is w_g, t, w_g respectively. We obtain the final sentence representation used for trade-off mechanism input from:

$$h_r = \tanh(\gamma) \tag{5}$$

To make full use of the known information, we employ full-connection layer to encode V_r into entity-pair representation e_r. We define a matrix $W_r \in R^{|d_w| \times g}$ as a training parameter matrix to encode entity pairs, where g is defined as the hidden cell size of Bi-GRU. The entity-pair representation e_r in a bag is generated as follows:

$$e_r = W_r V_r \tag{6}$$

C. Trade-off Mechanism

Considering the different effects of the sentence representation h_r and the corresponding entity-pair representation e_r for each sentence, we propose a trade-off mechanism, which adjusts the influence of each sentence information and the corresponding entity-pair information to generate instance representation f_r dynamically. The trade-off mechanism workflow is shown in Figure 2.

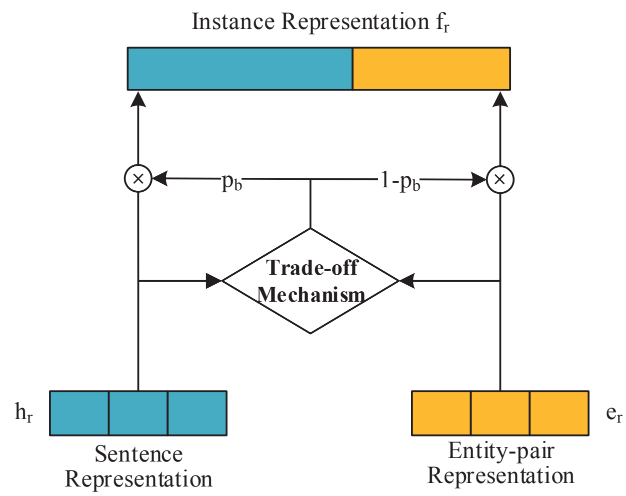

Fig. 2. The trade-off mechanism workflow.

In Figure 2, a trade-off probability $p_b \in [0, 1]$ for each time step is calculated from the sentence representation h_r and the entity-pair representation e_r as follows:

$$p_b = \sigma(w_s^T h_r + w_e^T e_r + b_{ptr}) \tag{7}$$

where vectors w_s, w_e and scalar b_{ptr} are learnable parameters and σ is the sigmoid function. Then, p_b can be regarded as a soft adjuster changing the confidence of the sentence representation h_r or the entity-pair representation e_r to determine relational facts. For each sentence, we denote the union of the sentence and the corresponding entity pair from source to correct the judgement of our model. We obtain the following instance information distribution f_r for each original data:

$$f_r = p_b e_r + (1 - p_b) h_r \tag{8}$$

Thus, if the syntax and semantic information in the sentence can clearly reflect the labeled relation, then p_b is close to zero and the sentence feature h_r is more helpful to judge relational facts. On the other hand, if the sentence feature is not obvious, our model will focus on entity-pair representation e_r.

D. Instance-level Attention and Optimization

As revealed by Lin et al. [13], each instance does not have the equal status in the bag. Therefore, it is necessary to use attention to neglect the noisy instance. Hence, we calculate high level extracted feature F_r using instance-level attention as follows:

$$F_r = \sum_{n=1}^{n} \alpha_i f_{ri} \tag{9}$$

$$\alpha_i = \frac{\exp(f_{ri} A r)}{\sum_{j=1}^{n} \exp(f_{rj} A r)} \tag{10}$$

where F_r is the relation representation of an instance, r is randomly initialized global relation vector which represents the classification of relations and A is a randomly initialized weighted matrix.

Finally, we import a softmax classifier to predict relation. Here, we use the following formula to calculate output conditional probability $P(r|S,\theta)$ for each relation:

$$y = W F_r + b \tag{11}$$

$$P(r|S,\theta) = \frac{\exp(y)}{\sum_{i=1}^{k} \exp(y_i)} \tag{12}$$

where W is the representation matrix of relations and $b \in R^m$ is a bias, k denotes the number of relational classifications. In addition, to optimize our model, we calculate loss using cross-entropy, adopt stochastic gradient descent (SGD) to find the optimal solution and prevent overfitting with $L2$ regularization and dropout.

IV. Experiments

A. Experimental Settings

In this paper, we conduct experiments on the widely used NYT-Freebase dataset [10]. This dataset contains 53 relations including a special relation NA which indicates there is no relation between head and tail entities. This dataset was generated by aligning Freebase relations with the New York Times corpus. Moreover, there are 522,611 sentences, 281,270 entity pairs, 18,252 relational facts in the training data and 172,448 sentences, 96,678 entity pairs, 1,959 relational facts in the testing data. The large size of the dataset can supply us more information to train. Similar to previous work [10]–[15], [18], [20], we evaluate our model in the held-out evaluation with Precision-Recall (P-R) curve and top-N precision (P@N) metrics. The evaluation is done by comparing the relations found in testing articles with those in Freebase.

Following previous work, we fine tune our model using validation on the training data. We use the word embeddings released by Lin et al. [13], which are trained on the NYT corpus using the word2vec tool. We select batch size among $\{50, 100, 150, 200\}$. For other parameters, we follow the settings used in the experiments of Lei at al. 's work [20]. Table I shows the detailed parameters used in the experiments.

B. Compared with Previous Methods

In this section, we compare our model with several feature-based and neural-based methods through held-out evaluation.

*a) **Feature-based Methods**:* Feature-based methods utilize NLP tools to discover syntactic or semantic information to infer results. Here, we choose three methods as follows:

- **Mintz** [8]: It is a traditional distantly supervised model with multiclass logistic regression.
- **MultiR** [11]: The model uses a probabilistic graphical model of multi-instance learning which can handle overlapping relations.
- **MIML** [9]: It is another probabilistic graphical model, but it is joint both multiple instances and multiple relations.

*b) **Neural-based Methods**:* Recently, most high performance methods are based on neural networks. We select the following four recent neural-based methods:

- **CNN+ATT** [13]: It is a sentence-level attention model over instance learning.
- **PCNN+ATT** [13]: The model is based on CNN+ATT, which achieves great performance with the piecewise max pooling.
- **PCNN+ATT+soft-label** [15]: This method adds a soft-label adjuster to PCNN+ATT, which takes both the confidence of distantly supervised labels and entity-pair representations into consideration.
- **CORD** [20]: It shows a cooperative denoising method for distantly supervised relation extraction based on Bi-GRU with two attentions and external rule knowledge, which considers all mentioned entities from the corpus as entity-sequence feature for each sentence.

Figure 3 draws the P-R curves for all compared methods on the whole test data, we can observe that: (1) As for all methods based on neural networks, they achieve higher coverage than the feature-based methods. This indicates that neural networks with multi-instance can express sentence information better than the methods using the human-designed feature. (2) Three methods drew by dashed lines have obvious improvement compared with others, especially when the recall is over 0.10, because they all take entity information into account. Note that, TMNN and PCNN+ATT+soft-label using

TABLE I
PARAMETER SETTINGS.

Word dimension d_w	Position dimension d_p	Batch size B	Cell size of GRU d_c	Learning rate λ	Dropout probability p
50	5	50	230	0.001	0.5

entity-pair information gain more superiority compared with entity-sequence embedding method in CORD. The reason may be that utilizing more entities information can bring in some extra noises. (3) TMNN outperforms other methods on most recall area. This demonstrates the effectiveness of TMNN. TMNN can express information fully by the trade-off mechanism combining text information and entity-pair feature on the sentence level. Although PCNN+ATT+soft-label also considers the entity-pair representation, the model only regards the entity-pair representation as a weighted combination of all sentences on the bag level. It might cause noises because the effect of entity pair is different for each sentence in the real world.

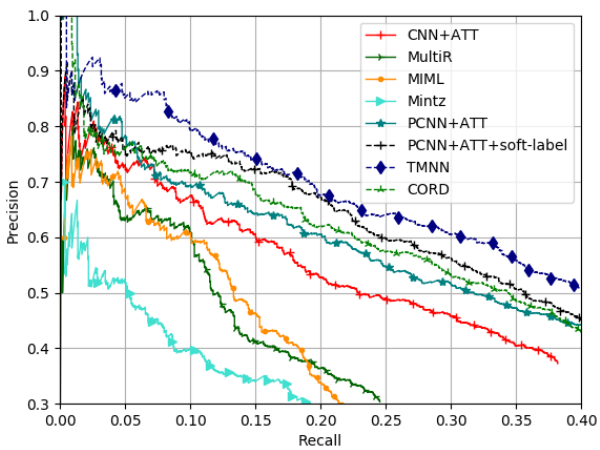

Fig. 3. P-R Curves comparing TMNN to previous methods.

C. Effect of Entity-pair Information and Trade-off Mechanism

We evaluate the effect of entity-pair information and the trade-off mechanism in our relation extraction method by comparing the performances before and after removing the corresponding feature or the mechanism. Among them, TMNN-TM is the model removing the trade-off mechanism by setting the fixed weighted parameter p_b as $\{0.1, 0.2, \ldots, 0.8, 0.9\}$ respectively. In Figure 4 and Table II, TMNN-TM only shows the fixed weighted parameter p_b as 0.5, because the others have similar results. TMNN-EPF is a Bi-GRU model combining word-level attention and sentence-level attention without the entity-pair feature and the trade-off mechanism. Here, we analyze their P-R curves from Figure 4 and their P@N results on the whole test data in Table II.

As Figure 4 shows, we can see that when the recall is over than 0.10, the performance of the TMNN-EPF method declines more quickly than that of the TMNN-TM/TMNN. Obviously, there is a gap between TMNN-EPF and methods considering entity-pair information in Figure 4. It proves that entity pairs in the dataset can supply more valuable information to help distantly supervised relation extraction denoising. Then, we compare two methods that involve both the sentences and the corresponding entity pairs. TMNN has higher coverage than TMNN-TM, which indicates that different entity pairs or sentences may have different optimal influence factors. Therefore, the descent of TMNN-TM is close to TMNN, but the TMNN method has robust superiority over the method with fixed weighted parameter p_b.

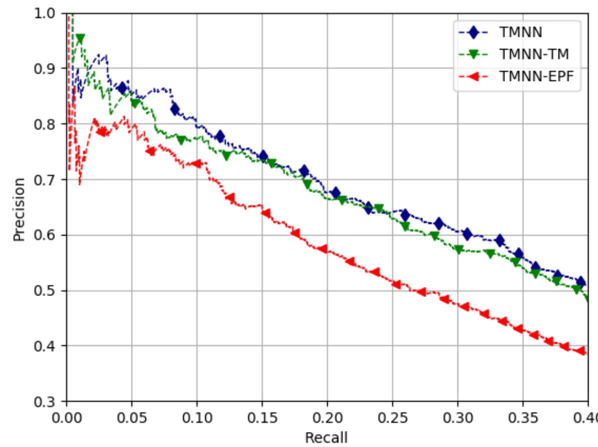

Fig. 4. P-R Curves comparing TMNN with models removing entity-pair feature or trade-off mechanism.

TABLE II
P@N ON THE WHOLE TEST DATA

Test Settings	P@100(%)	P@200(%)	p@300(%)	Average(%)
TMNN	**87.0**	**82.5**	**77.3**	**82.2**
TMNN-TM	**87.0**	78.5	74.0	79.8
TMNN-EPF	80.0	75.0	71.3	75.4

In Table II, we observe that TMNN gets the highest precision. Compared the first row with other rows, the trade-off mechanism with entity-pair feature can improve precision by 3%-7%. It further reflects that adjusting dynamically the trade-off probability according to different contexts may bring great benefits. In contrast, setting the fixed parameter limits the flexibility of the effect of text and entity pair, especially when there is more than one sentence in the bag.

TABLE III
P@N ON THE MORE-THAN-ONE-SENTENCE TESTING DATASETS.

Multi-Instance	Top N	CNN+ATT(%)	PCNN+ATT(%)	TMNN-EPF(%)	TMNN-TM(%)	TMNN(%)
One	100	76.2	73.3	70.0	83.0	**86.0**
	200	65.2	69.2	60.0	**77.0**	76.5
	300	60.8	60.8	54.0	67.3	**71.0**
	Average	67.4	67.8	61.3	75.7	**77.8**
Two	100	76.2	77.2	79.0	80.0	**88.0**
	200	65.7	71.6	65.0	**77.5**	77.0
	300	62.1	66.1	58.3	69.3	**71.3**
	Average	66.9	71.6	67.4	75.6	**78.7**
All	100	76.2	76.2	79.0	84.0	**92.0**
	200	68.6	73.1	72.0	77.0	**83.0**
	300	59.8	67.4	61.6	72.0	**75.6**
	Average	68.2	72.2	70.8	77.6	**83.5**

D. Effect of Sentence Number

To demonstrate that our method is effective to multi-instance learning, we select the entity pairs which have more than one sentence in the testing data. In Table III, the testing datasets are distinguished from the whole testing data. We select one, two, all sentences for each testing entity pair. Note that, for one or two sentences, we select sentences randomly, and we also use all the sentences for the training data. Following Lin et al. [13], we report the P@100, P@200, P@300 and the average of them for each model.

As shown in Table III, we can see that our models are similar to CNN+ATT and PCNN+ATT for multi-instance learning, which get higher precision on the average as the sentence number increases. TMNN-EPF as the basis of our model achieves comparable results with CNN+ATT and PCNN+ATT, which reveals the effectiveness of the Bi-GRU model with two attentions that can be the backbone of the TMNN framework. But as sentence number increases, their precision declines quickly. By contrast, TMNN-TM and TMNN achieve great improvement, because they correct wrong text information by the corresponding entity-pair feature for each instance. However, TMNN-TM regards entity-pair and text as the fixed effect ignoring the importance of feature selection. Hence, TMNN can have about 3%-6% improvements for the average results to TMNN-TM.

E. Case Study

We further pick two examples from the whole testing data to illustrate the effectiveness of our adjustable trade-off mechanism. In Table IV, we show two instances and the corresponding text weights and entity-pair weights, where the weights are learned by the trade-off mechanism. We highlight the entity pairs and larger trade-off weights in bold format.

From Table IV, we can find that different sentences have different degrees of dependence on text information and entity-pair information. The former example is related to the relation "Location contains". The expression "including" as a trigger can help machine predict relation accurately. So, the text weight is close to 1. Similarly, the next example has the keyword "chief" to express the relation "Business company". However, some obvious interference information such as "a member of", "co-founder", and additional person "jerry yang" can confuse the judgement of machine because of the position and semantics of these words in the sentence. At this time, entity-pair feature can correct the instance representation using a high entity-pair weight as 0.7433. In this way, our proposed trade-off mechanism can help reduce the wrong labeling problem for distant supervision effectively.

TABLE IV
TWO TESTING EXAMPLES USING TMNN.

Text Weight	Entity-pair Weight	Relation	Text
0.8825	0.1175	Location contains	There are many more in **Alaska**, including the 20,320-foot Denali, also known as **Mount McKinley**, the highest peak in north America.
0.2567	**0.7433**	Business company	**Eric_e._schmidt**, who is chief executive of **google** as well as a member of apple's board, and jerry yang, co-founder of yahoo, came on stage to endorse the new hand-held.

V. Conclusion and Future Work

In this paper, we introduce a neural framework for distantly supervised relation extraction through a trade-off mechanism. The framework can adjust the effects of text and entity-pair in the dataset by scaling weights dynamically to reduce wrong labeling problem. We show that considering entity-pair information on the sentence level can provide rich information for training. Besides, our proposed trade-off mechanism can help machine select valuable information adaptively to reduce noises from wrong labels. The experimental results on the NYT-Freebase dataset illustrate that our model outperforms those state-of-the-art feature-based methods and neural-based methods.

In the future, we plan to combine our method with knowledge graphs. For example, some representation learning methods like TransE [26] and ConvE [27], can provide more valuable information to reduce noises. To further improve its performance, we also plan to find a more suitable entity-pair encoder for our TMNN method.

REFERENCES

[1] Ce Zhang, Jaeho Shin, Christopher Ré, Michael J. Cafarella, and Feng Niu. Extracting databases from dark data with deepdive. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 847–859, 2016.

[2] Christopher De Sa, Alexander Ratner, Christopher Ré, Jaeho Shin, Feiran Wang, Sen Wu, and Ce Zhang. Incremental knowledge base construction using deepdive. *VLDB J.*, 26(1):81–105, 2017.

[3] Amina Kadry and Laura Dietz. Open relation extraction for support passage retrieval: Merit and open issues. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pages 1149–1152, 2017.

[4] Abdalghani Abujabal, Mohamed Yahya, Mirek Riedewald, and Gerhard Weikum. Automated template generation for question answering over knowledge graphs. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1191–1200, 2017.

[5] Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cícero Nogueira dos Santos, Bing Xiang, and Bowen Zhou. Improved neural relation detection for knowledge base question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 571–581, 2017.

[6] Razvan C. Bunescu and Raymond J. Mooney. Subsequence kernels for relation extraction. In *Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]*, pages 171–178, 2005.

[7] Yi Yao Huang and William Yang Wang. Deep residual learning for weakly-supervised relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1803–1807, 2017.

[8] Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data. In *ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2-7 August 2009, Singapore*, pages 1003–1011, 2009.

[9] Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. Multi-instance 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, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea*, pages 455–465, 2012.

[10] Sebastian Riedel, Limin Yao, and Andrew McCallum. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III*, pages 148–163, 2010.

[11] Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke S. Zettlemoyer, and Daniel S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 541–550, 2011.

[12] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 2335–2344, 2014.

[13] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*, 2016.

[14] Yankai Lin, Zhiyuan Liu, and Maosong Sun. Neural relation extraction with multi-lingual attention. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 34–43, 2017.

[15] Tianyu Liu, Kexiang Wang, Baobao Chang, and Zhifang Sui. A soft-label method for noise-tolerant distantly supervised relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1790–1795, 2017.

[16] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*, 2010.

[17] Razvan C. Bunescu and Raymond J. Mooney. Learning to extract relations from the web using minimal supervision. In *ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic*, 2007.

[18] Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 1753–1762, 2015.

[19] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.

[20] Kai Lei, Daoyuan Chen, Yaliang Li, Nan Du, Min Yang, Wei Fan, and Ying Shen. Cooperative denoising for distantly supervised relation extraction. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 426–436, 2018.

[21] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1073–1083, 2017.

[22] Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 3060–3066, 2017.

[23] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *CoRR*, abs/1412.6575, 2014.

[24] Pengda Qin, Weiran Xu, and Jun Guo. Designing an adaptive attention mechanism for relation classification. In *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017*, pages 4356–4362, 2017.

[25] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*, 2016.

[26] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2787–2795, 2013.

[27] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1811–1818, 2018.