Bert-Based Denoising and Reconstructing Data of Distant Supervision for Relation Extraction

Tielin Shen, Daling Wang, Shi Feng, Yifei Zhang

School of Computer Science and Engineering, Northeastern University, Shenyang, China
shen_tielin@163.com
[wangdaling, fengshi, zhangyifei]@cse.neu.edu.cn

Abstract. Inter-Personal Relationship Extraction (IPRE) aims to find the relationships of people entity pairs for a set of given sentences containing the entity pairs. Distant supervision has been an important technique for data acquisition in relation extraction, but will inevitably generate a large number of noise data. In this paper, we regard IPRE as a classification problem, and propose an approach of Bert-based denoising and reconstructing dataset of distant supervision for IPRE. Based on Bag-Track (a label for a bag consisting of set of sentences) training set, we merge all sentences in one bag into a long sentence, so every bag is translated into a sentence. We build a binary classifier with Bert and class every sentence into "NA" or "Non NA" class. We remove "NA" class sentences which were tagged "Non NA" in initial Bag-Track training set for data denoising. Utilizing the denoised dataset, we further reconstruct bag (i.e. long sentence because of our translation) training set and build detail classifier innovatively for IPRE of Bag-Track and Sent-Track (a label for a sentence). The experiments in CCKS 2019 eval Task3 IPRE datasets show that our approach achieves better results than most methods.

Keywords: Data Denoising, Distant Supervision, Relation Extraction

1 Introduction

Relation extraction from texts is one of the hot research issues in nature language processing and knowledge graph, and inter-personal relationship extraction (IPRE) is an important research branch of the issue. It has important significance for building social network.

In many current researches, relation extraction is regarded as a machine learning problem, and deep learning technique has been applied widely. However, the relation extraction based on deep learning needs a large of labeled data which results in it restricted in some applications.

For solving the problem, Mintz et al. [1] proposed distant supervision in 2009. It is based on the assumption that, for an entity pair, if there is a kind of relation between them, all texts containing the entity pair can represent the relation.

For example:

Sentence: Qianlong was Yongzheng's son (乾隆是雍正的儿子).

Relationship: the relationship between Yongzheng and Qianlong was ***father*** (雍正与乾隆之间的关系是*父亲*).

According to the idea of distant supervision, all sentences containing the entity pair Yongzheng (雍正) and Qianlong (乾隆) can be added to training set as a positive case with the relation. Meantime, an entity pair was selected randomly and added to the training set as a negative case without the relation. Using the method, all sentences containing the entity pair constitute a bag. Applying this method, a large of training data can be obtained.

Distant supervision can rapidly expand training, but it will generate many error labels because it tags all sentences containing an entity pair as the same relation label. Such an example can be shown as follows:

Sentence: YongZheng and QianLong were the two emperors of the Qing Dynasty (雍正和乾隆是清朝的两个皇帝).

Relationship: the relationship between Yongzheng and Qianlong was ***father*** (雍正与乾隆之间的关系是*父亲*).

This sentence has also been added to the training set as a positive case, but in fact it does not reflect the father-son relationship between Yongzheng and Qianlong. It is a false positive sample.

CCKS 2019 eval Task3 IPRE [2] is such a distant supervision dataset. The task is divided into Bag-Track and Sent-Track task. The former task is extracting relation in bag level, and the latter is extracting relation in sentence level, which means two sentences may have different relation labels even they contain the same entity pair. However, only a Bag-Track training set is provided for the two tasks. This is also common in many applications.

Google published Bert [3] in 2018, which is a language presentation model. By fine-tuning, the model achieved good results in Bag-Track. Inspired by this process, we first transform a bag into a long sentence by merging the sentences in a bag, and build a binary classifier with Bert and class every sentence into "NA" or "Non NA" class. We remove the bags (sentences) with "NA" class tagged "Non NA" in initial Bag-Track training set for data denoising. Utilizing the denoised dataset, we further reconstruct Bag-Tract training set and build a detail classifier innovatively for IPRE of Bag-Track and Sent-Track. By the two-stage classification (binary classification and detail classification), we extract relationship between entity pair.

To sum up, the main contributions of this paper are as follows:

1) We design a method for denoising data of distant supervision binary classification based on Bert. Especially, we transform a bag into a long sentence by sentence merging, so bag training set can be used to build sentence classifier.

2) We reconstruct Bag-Track training set by adding entity name in sentences for fitting Bert, and build a Bert detail classifier. Because of our transformation, the sentence detail classification can be implemented by means of bag classification.

3) Combining above two-stage classification, our detail classifier not only achieves good IPRE result in Bag-Track, but further obtains good IPRE result in Sent-Track.

The rest of this paper is organized as follows. Section 2 introduces the related work with our task. Section 3 gives our approach for implementing IPRE. Section 4 shows our experiments. We conclude our work in Section 6.

2 Related Work

Mintz et al. [1] first introduced the distant supervision into the relational extraction task, and acquired a large number of training samples, which effectively solved the shortcomings of supervised relational extraction method labeling training corpus, which consumed a lot of manpower and material resources.

For distant supervision, Zeng et al. [4] proposed a method to process distant supervision data set by piecewise pooling and multi-instance learning. The remaining sentences in bag were discarded and the relationship between sentences with the highest confidence in bag was regarded as the classification result of the whole bag. Lin et al. [5] proposed an attention mechanism for distant supervision data set in 2016, which filters noise data without losing too much information. However, the calculation process of attention weight in the testing process is too complex to interpret. Jiang [6] considered more about the fact that some entities have multiple relations at the same time. He maximized the pooling of every sentence in bag. In 2017, Feng et al. [7] used the idea of memory networks to improve attention mechanisms, taking into account dependencies between entities. Ji et al. [8] introduced additional knowledge base information in relation extraction, which further enhanced the effect of lword embedding. Qin et al. [9] used reinforcement learning to denoise the data set generated by distant supervision in 2018, thus improving the performance of the next classifier, whose reinforcement reward comes from the performance change of the classifier. Feng et al. [10] also introduced the idea of reinforcement learning into relational extraction of distant supervision data set. The difference is that reinforcement rewards come from predictive probability.

Our work combines Bert with training set bag to train the binary classifier similar to [3], and finds out the false positive sentences in the training set, so as to achieve the purpose of noise reduction in the training set. Unlike the above work, our denoising work is independent of bag classifier and should be regarded as a step of data pre-processing.

3 Our Approach

Our approach includes three parts. The first part is data denoising by binary classification for obtaining a clear training set. The second part is reconstructing the training set for building a detail classifier. And the third part is implementing Bag-Track and Sent-Track IPRE task by using the built classifier. The first two parts are in training stage, and the third part is in test stage. The framework is shown in Fig.1.

3.1 Data Denoising Based on Binary Classifying

As mentioned above, there are a large of noises in distant supervision data sets. According to multi-instance learning, if a bag is tagged as “NA”, all sentences in the bag will be “NA”. So we think the cases of “NA” class are correct, and the noises will exist in not “NA” (having detail relation) class cases. Our denoising will process the

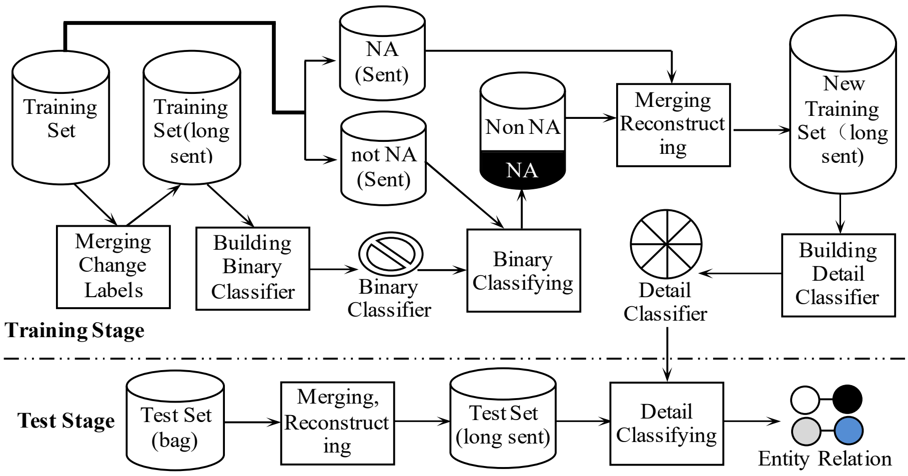

Fig. 1. Framework of Our Approach

data. We merge all the sentences in each bag and then change the labels of all bags to NA and non-NA. Then, we build a Bert-based binary classifier with Changed data set. By merging, we transform a bag into a sentence, so the binary classifier fits to both bag and sentence. Meantime, we classify all sentences with detail class label into "NA" or "Non NA" class by our binary classifier. Because these sentences are tagged as detail relation of their bag, for the cases classified into "NA", we regard them as false positive and remove them from training set.

An example of our data denoising process is shown in Fig.2. In Fig.2, the lift 9 sentences fed to middle Bert for binary classifying. After classifying and denoising, the result is the right 6 sentences. Because all sentences in a bag contain the same entity pair, the merging is reasonable. The experiments in Section 4 will show that here merging sentences and data denoising can reduce false positive cases in training set and improve classification results effectively.

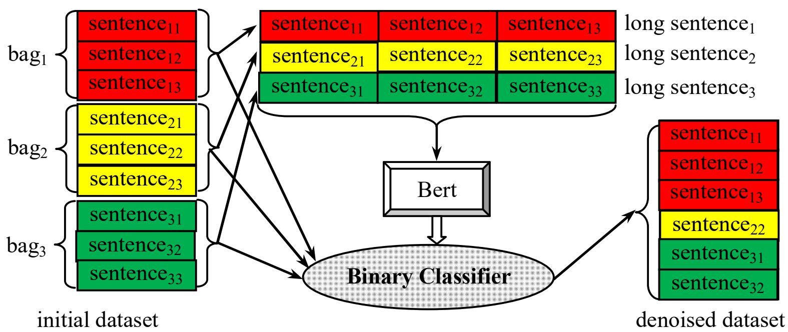

Fig. 2. An Example of Data Denoising Process

3.2 Reconstructing Training Set for Building Detail Classifier

The detail classifier used for final classification still adopts Bert. As we know, when Bert is used for text classification, the maximum sentence length acceptable to the model is 510. After training set is denoised, the size of bag is reduced. Therefore, when we classify the bag, the training set and test set are still in the form of long sentences, which is, the results of the sequential connection of all sentences in the bag.

Bert's Chinese pre-training model encodes sentences according to words and extracts features. In pre-training, 15% of the words input to Bert will be replaced by [MASK] token, and then the model predicts the hidden words based on the context of other words not covered by mask in the sequence. If we enter directly merged long sentences into Bert, a model may not be the name of the entity as a whole because of the word coding and random mask, so we fill in the name of the entity pair and a colon in front of every long sentence. From the point of view of interpretability, model can also understand, and the long sentence after the colon is used to describe the entity right before the colon. The reconstructing form of a sentence is “Entity 1, Entity 2: sentence”.

The reconstructed training set includes NA class cases from initial training set and Non NA class cases from denoised dataset which are tagged with their labels in initial training set.

Right side of training stage in Fig.1 is for the process. Moreover, Fig.3 shows the process of reconstructing training set. After denoising in Fig.2 we merge each bag and fill in the entity pair name in every merged long sentence and obtain the 3 long sentences with entity pair which can fit to Bert.

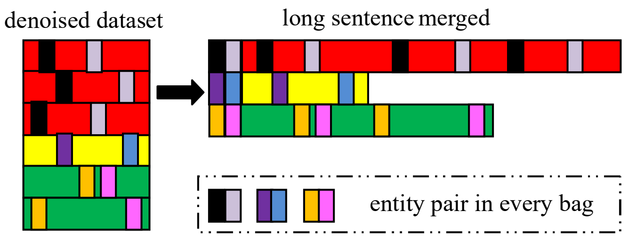

Fig. 3. An Example of Reconstructing Training Set

We take an example in Bag-Track training set of CCKS 2019 eval Task3 IPRE [2] as follows for explaining Fig.3.

Raw bag:

Sentence_1: 纪昀和刘墉更有着不解之缘 (Ji Yun and Liu Yong have an indissoluble bond).

Sentence_2: 而后来纪昀被发配的案件，又恰是刘墉负责 (Later, it was Liu Yong who was responsible for Ji Yun's case).

Sentence_3: 刘墉比纪昀年长4岁 (Liu Yong is 4 years older than Ji Yun).

Reconstructed long sentence:

刘墉，纪昀：纪昀和刘墉更有着不解之缘。而后来纪昀被发配的案件，又恰是刘墉负责。刘墉比纪昀年长4岁。(Liu Yong, Ji Yun: Ji Yun and Liu Yong have an indissoluble bond. Later, it was Liu Yong who was responsible for Ji Yun's case. Liu Yong is 4 years older than Ji Yun.).

Based on the training set, we further build a detail classifier for classifying new bags and sentences to extract relation between entities.

3.3 Classifying for IPRE Task in Bag-Track and Sent-Track

After the building detail classifier, we can use it to classify the bags and sentences. The test stage in Fig.1 is for the process.

Because we have transformed every bag into a sentence, our binary classifier can be appropriate for bag and sentence classification. For bag classification, we can use above detail classifier directly after merging and reconstructing every bag.

For sentence classification, we classify every sentence by means of the class of bag containing the sentence. Fig.4. shows an example of sentence classification. Now, in test set, bag_1 contains $sentence_{11}$, $sentence_{12}$ and $sentence_{13;}$ bag_2 contains $sentence_{21}$, $sentence_{22}$ and $sentence_{23;}$ bag_3 contains $sentence_{31}$, $sentence_{32}$ and $sentence_{33}$. We first classify every sentence using the binary classifier, and the result is Non NA, NA, Non NA, NA, Non NA, Non NA, Non NA, NA, and NA respectively. It means $sentence_{11}$, $sentence_{13}$, $sentence_{22}$, $sentence_{23}$, and $sentence_{31}$ (they are all be classified into Non NA class) can present the relation of entity pair. Then we reconstruct the 3 bags and use the detail classifier to classify the bags. We suppose that the result is "father", "NA", and "mother" (red, white, green color in Fig.4). Combing the two classifiers, we further classify sentences. For bag_1, $sentence_{11}$ and $sentence_{13}$ correspond to the class label of bag_1 (father), so the label of $sentence_{11}$ and $sentence_{13}$ is tagged as "father" (red color), and $sentence_{12}$ is tagged as "NA" (white color). For bag_2, $sentence_{22}$ and $sentence_{23}$ can present the relation of entity pair, but the result of bag_2 detail classified is "NA", so $sentence_{21}$, $sentence_{22}$, and $sentence_{23}$ are classified into "NA". Bag_3 is similar to bag_1, whose $sentence_{32}$ and $sentence_{33}$ are "NA" based on the binary classification result for sentence, and $sentence_{31}$ is "mother" based on the detail classification result for bag.

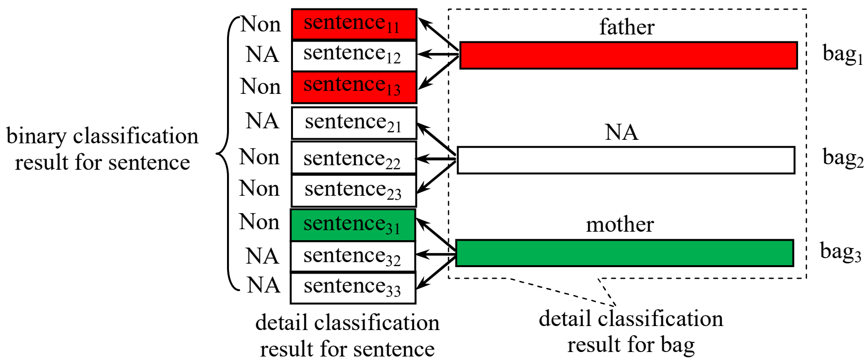

Fig. 4. An Example of IPRE Task in Bag-Track and Sent-Track

Finally, we describe the rule of detail classifying a sentence based on bag class is:
If *one sentence is binary classified into NA*
Then *its class label is NA*
Else *its class label is the result detail classified of bag containing it*

The experiments in Section 4 show that our process of IPRE task in Bag-Track and Sent-Track achieves good $F1$ values.

4 Experiment

In this section, we conduct some experiments for proving the advantage of our approach by comparing our experimental results with related work.

4.1 Data Set

We used the public data set released by CCKS 2019 eval Task3 [2] Inter-Personal Relationship Extraction (IPRE) data sets to conduct our experiments. This data set is generated by using distant supervision method on entity pair which appears frequently in Chinese Baidu Baike. The task includes Bag-Track and Sent-Track, but only Bag-Track has training set, i.e. every bag has a class label. In this case, we have to classify sentences by means of bag training set to build the classifier for Sent-Track task. The dataset has 35 types of relations, among which the training set has 37948 bags, including 287351 sentences. In particular, the not "NA" class has 2948 bags, including 38501 sentences. There are 5416 bags in the test set, including 38417 sentences. Among them, there are 300 bags of not "NA" class, including 2498 sentences.

4.2 Metrics

We use $F1$ metric for evaluation. $F1$ value is calculated with Formula (1).

$$P = \frac{N_r}{N_{sys}} \quad R = \frac{N_r}{N_{std}} \quad F1 = \frac{2PR}{P+R} \tag{1}$$

For Sent-Track, N_r is the sentence number classified correctly and classified as Non-NA by given algorithm, N_{sys} is the sentence number classified as Non-NA by given algorithm, and N_{std} is the sentence number tagged as Non-NA in given dataset.

For Bag-Track, N_r is the bag number classified correctly and classified as Non-NA by given algorithm, N_{sys} is the bag number classified as Non-NA by given algorithm, and N_{std} is the bag number tagged as Non-NA in given dataset.

4.3 Experimental Result of Detail Classification

Here we evaluate our method in Section 3.2, i.e. reconstructing data set and building a detail classifier. In this experiment, using the built detail classifier, we classify Bag-Track test set. We compare our algorithm with CNN+ATT [5], PCNN+ATT [4], Bert without Reconstructing (not filling in entity pair name). Our algorithm is Bert with Reconstructing (filling in entity pair name). The result is shown in Table 1.

Table 1. The Classification Result Comparison

Algorithm	$F1$ value
CNN+ATT	0.3217
PCNN+ATT	0.2804
Bert without Reconstructing	0.3158
Bert with Reconstructing	**0.5233**

Obviously, our Bert with Reconstructing achieves better result than other methods. We analyze that the reason should be there are more than one entity pairs in one long sentence with different relations and will affect the classification results for Bert without Reconstructing. We reconstruct dataset by filling in entity pair name, the purpose is to "tell" the classifier that sentences should be classified based on their entity pairs' relation.

4.4 Experimental Result of Denoising

We implement classification on denoised data set for evaluating the quality of data set denoising result. After test set denoising with Section 3.1 method, we classify test set with the detail classifier built from the denoised data set. We compare our method with CNN+ATT [5] in raw test set and denoised test set. The result is shown in Table 2. We can see that by denoising dataset, and we can achieve better classification result than the one in raw dataset. As mentioned above, distant supervision will generate many error labels. Our denoising is for solving the problem by removing thus training cases with false positive.

Table 2. The Classification Result Comparison in Raw and Denoised Dataset

Algorithm	Raw Dataset	**Denoised Dataset**	Denoising Ratio
CNN+ATT	0.3217	**0.3421**	0.2
Bert with Reconstructing	0.5233	**0.5497**	0.4

4.5 Experimental Results of Bag-Track and Sent-Track IPRE Task

According to Section 3.3 method, we extract inter-personal relationship of every bag in Bag-Track test set. The result shows our algorithm ranks third out of 80 teams. The result is shown in Table 3.

In Table 3 and Table 4, the results come from CCKS 2019 eval Task3 website (Bag-Track: https://biendata.com/competition/ccks_2019_ipre/final-leaderboard/, and Sent-Track: https://biendata.com/competition/ccks_2019_ipre/leaderboard/). Here NEU_DM1 is the result submitted by us.

Here, we need to clarify a problem. Our experiments in Section 4.3 and 4.4 are implemented in the test set given in Section 4.1, it is released by CCKS 2019, but the experiments in Section 4.5 are implemented in another test set, which is online and we can not perform our current denoising method in the test set for comparison with other teams' results. For the results in Table 3 and Table 4, our work only applied merging and reconstructing process without denoising. For this reason, we only compare our method denoising and without denoising in published test set. Table 2 shows that better result can be achieved by our data denoising.

Table 3. Result of Bag-Track IPRE Task

Team	$F1$ value
LEKG	0.63030
格物致知	0.62162
NEU_DM1	**0.57459**
4~80 teams	4~80 results

Table 4. Result of Sent-Track IPRE Task

Team	$F1$ value
格物致知	0.54279
LEKG	0.48427
NEU_DM1	**0.46200**
4~140 teams	4~140 results

5 Conclusion

For Inter-Personal Relationship Extraction (IPRE) in CCKS 2019 eval Task3, we propose a Bert-based denoising and reconstructing dataset of distant supervision method. By the sentence merging and binary classification, we implement dataset denoising. By the data reconstructing, we build a detail classifier, classifying bags and sentences for relation extraction from text. Especially, we innovatively combine sentence binary classification and bag detail classification to implement sentence detail classification. Furthermore, by data denoising, we obtain better result than the one of our submitted.

Acknowledgments

The work was supported by the National Key R&D Program of China under grant 2018YFB1004700, and National Natural Science Foundation of China (61772122, 61872074).

References

1. Mintz M, Bills S, Snow R, Jurafsky D: Distant supervision for relation extraction without labeled data. ACL/IJCNLP 2009: 1003-1011
2. Wang H, He Z, Ma J, Chen W, Zhang M: IPRE: a Dataset for Inter-Personal Relationship Extraction. CoRR abs/1907.12801 (2019)
3. Devlin J, Chang M, Lee K, Toutanova K: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT (1) 2019: 4171-4186
4. Zeng D, Liu K, Chen Y, Zhao J: Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. EMNLP 2015: 1753-1762
5. Lin Y, Shen S, Liu Z, Luan H, Sun M: Neural Relation Extraction with Selective Attention over Instances. ACL (1) 2016
6. Jiang X, Wang Q, Li P, Wang B: Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks. COLING 2016: 1471-1480
7. Feng X, Guo J, Qin B, Liu T,Liu Y: Effective Deep Memory Networks for Distant Supervised Relation Extraction. IJCAI 2017: 4002-4008
8. Ji G, Liu K, He S, Zhao J: Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions. AAAI 2017: 3060-3066
9. Qin P, Xu W, Wang W: Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning. ACL (1) 2018: 2137-2147
10. Feng J, Huang M, Zhao L, Yang Y, Zhu X: Reinforcement Learning for Relation Classification From Noisy Data. AAAI 2018: 5779-5786