A Joint Financial Event Entity Extracting Method Based on Integrating of Sequence Labeling and Question Answering

Peiji Yang[1], Guanyu Fu[1], Xinyu Li[1], Dongfang Li[1], Jing Chen[1]
Qingcai Chen*, Yubin Qiu[2] and Yiqing Feng[1]

[1] Harbin Institute of Technology, Shenzhen 518055, China
[2] Shenzhen Tuling Robot Co., Ltd, Shenzhen 518000, China
*Corresponding Author:qingcai.chen@hit.edu.cn

Abstract. Event detection is one of the important tasks in the field of public opinion monitoring and financial risk control. The core of event detection lies in the entity extraction of a particular type of event. The CCKS committee sets up a task for event entity detection in the field of finance, which requires the extraction of the entity involved in a specific type of events in the text. To solve this problem, we develop a model which combines event entity extraction model(BERT−QA) based on question answering and named entity recognition model(BERT-CRF-NER) based on sequence labeling. Our best fusion model on the official test dataset achieves the F1-score of 0.8390 which ranks the first place in the competition.

Keywords: Question Answering, Name Entity Recognition, Sequence Labeling

1 Introduction

With the explosive growth of financial information, financial texts processing has become a very popular research area in Natural Language Processing(NLP). Financial Named Entity Recognition (NER), the fundamental task in financial texts processing, is a subtask of Information Extraction. The key of this task is to locate the entities mentioned in unstructured texts. Due to the huge amount of data and the complexity of texts, there are still many challenges in this task. To promote the progress of this task, the 2019 China conference on knowledge graph and semantic computing (CCKS-2019) organized a financial NER task to extract the company subject of a specific event type from the real Chinese news corpus.

The Earlier NER systems were based on handcrafted rules and explicit feature engineering such as machine learning[1]. Starting with Collobert et al. (2011), neural network NER systems become more and more popular[2]. With no need of handcrafted rules and better generalization ability, neural network NER models become more and more popular.

In this paper, we not only demonstrate the basic NER methods, but also innovatively introduce a Question Answering (QA) method to deal with the NER task. In the end, we combine the two methods and achieve the best results.

2 Method

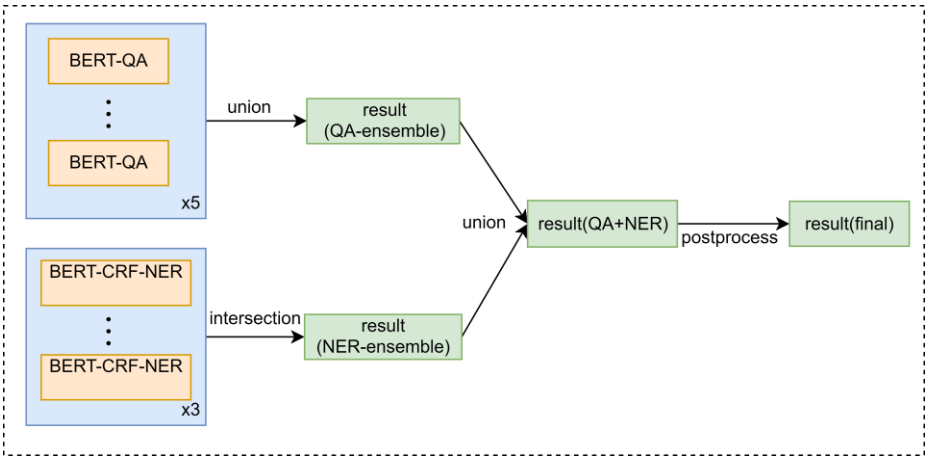

Fig. 1. The processing flow of our method

Our model, depicted in Figure 1, is a combination of QA and NER model. First, the data will enter two different modules in parallel, namely QA module composed of 5 BERT-QA models and NER module composed of 3 BERT-CRF-NER models. Then we fuse the results of these two modules based on their features. For BERT-QA, the extracted entities are accurate, so we preserve all of the model results to get more entity candidates. For BERT-CRF-NER, we only use the intersection of model results to improve precision. We merge the result of BERT-QA and BERT-CRF-NER as the final result. In the next section we will cover each of our models in detail.

2.1 Entity Extraction Based on Question Answering

The goal of our evaluation task is to extract specific business entities from texts. It seems to be a problem of classical Named Entities extraction, but additional event category information is given which means we need to determine which entity we should extract, since there are several different types of events in a piece of sample. Therefore, how to make good use of event type information has become the key problem. In this section, we convert this problem into an extracted question answering task. Let the event type information be the query and let event the initial text be the context. Then the target subject is the answer we need to extract[2].

2.1.1 Data Processing

As shown in Figure 2, we reconstruct the initial data into the SQuAD format.

```
{   "data": [
      {   "paragraphs": [
          {   "context": "9月30日，合肥市公安局庐阳分局再次通报了大志集团涉嫌非法集资的最
              新消息，通报称，经庐阳检察院批准，已依法对10名犯罪嫌疑人只是逮捕"
              "qas": [
                 {   "answer":[
                        {   "answer_start": 21,
                            "text": "大志集团"
                        } ],
                            "question": "事件高管负面",
                            "id": "102378"} ] } } ] } ,
}
```


Fig. 2. Sample of SQuAD format data

2.1.2 BERT-QA Model Architecture

The version of BERT we use is BERT-WWM[4], which is retrained on new Chinese dataset by Harbin Institute of Technology. The encoded result is then passed to the MLP layer to obtain the probabilities for each token to be the start and end positions. We sort the probabilities, take the most probable 20 tokens as the candidate positions of the start and end positions, and select all possible answers according to the length and positions of the answers. Afterwards, we calculate the sum of start probabilities and end probabilities as the final score for each result. The result which owns highest score is chosen as the final answer. The detail model structure is shown in Figure 3.

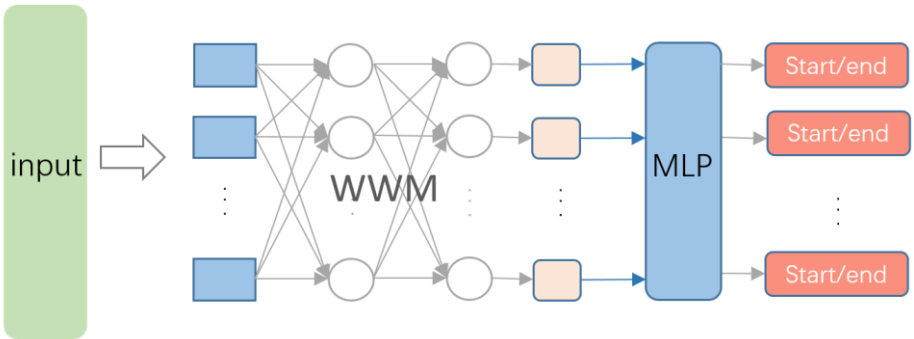

Fig. 3. BERT-QA model

2.1.3 Multi BERT-QA Ensemble

The target event in one sample may involve several entities, while a single QA model can only extract one entity. To solve this problem, we train five BERT-QA models with different parameters and merge the results of the five models with the union operation.

2.2 BERT-CRF Named Entity Recognition

In model fusion, the greater the difference of the models used for fusion, the higher the benefit the fusion model will get. Since we need to recognize all the correct event

entities in the event text, we build a NER model(BERT-CRF-NER) based on sequence labeling for fusion with the BERT-QA model.

2.2.1 Data Processing

Unlike traditional sequence tagging tasks, in this task, we need to recognize the event entity of the particular category event. So we decide to pass event class information to the model when building the data used for sequence labeling. Based on the powerful semantic understanding of BERT, we directly merge the event category with the event text[5]. The format of the merge is "__Category__Text". The merged text examples as Figure 4.

Ex1:__评级调整__4日增持: 乐视网(300104) 控股股东减持3524万股穆迪调降三甲塞浦路斯银行评级
Ex2:__涉嫌违法__P2P平台，一鼎金融，公司法人已被捕， 并立案

Fig. 4. Examples of spliced text

Then we build the tag sequence data for NER model. The example of the tag sequence for "__涉嫌违法__P2P平台，一鼎金融，公司法人已被捕，并立案" is shown in Figure 5.

_ \O _ \O 涉 \O 嫌 \O 违 \O 法 \O _ \O _ \O P \O 2 \O P \O 平 \O 台 \O ， \O 一 \B 鼎 \I 金 \I 融 \I ， \O 公 \O 司 \O 法 \O 人 \O 已 \O 被 \O 捕 \O ， \O 并 \O 立 \O 案 \O

Fig. 5. Examples of tag sequence

2.2.2 BERT-CRF-NER Model

For BERT method, the predictions are not conditioned on the surrounding predictions. A CRF layer has a state transition matrix as parameters[5]. With such a layer, the system can efficiently use past and future tags to predict the current tag. Therefore, we applied a CRF layer on the classification layer. We denote the output sequence after softmax layer as $H = [h_1, h_2, \dots h_n]$, then the predicted tag sequence $Z = [z_1, z_2, \dots z_n]$ is as follows:

$$Z = \underset{y}{\operatorname{argmax}} \frac{exp(score(H,y))}{\sum_{y'} exp(score(H,y'))} \tag{1}$$

where $score(H, y) = \sum_{t=1}^{n} E_{t,y_t} + \sum_{t=0}^{n-1} T_{y_t y_{t+1}}$, $E_{t,y_t} = w_{y_t}^{\mathrm{T}} h_t$ is the score of predicting tag y_t at the t_{th} time, and $T_{y_t y_{t+1}}$ is the score of transitioning from y_t to y_{t+1}[7]. And we add a masking constraint in the CRF layer, so that there will be no prediction that label result will shift from O to I. The model architecture of BERT-CRF-NER is shown in Figure 6.

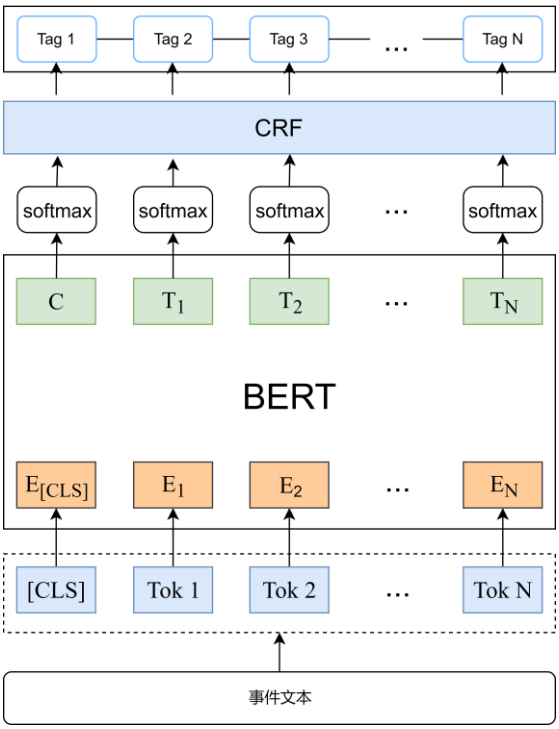

Fig. 6. The architecture of BERT-CRF-NER model

2.2.3 Multi BERT-CRF-NER Ensemble

Compared with BERT-QA model, BERT-CRF-NER model has the advantage of recognizing multiple entities in the event text instead of one. However, it will also recognize more error results. Therefore, we trained three BERT-CRF-NER models with different data distributions. In order to solve the low accuracy rate caused by the small probability of a single model recognizing the wrong results, we use the intersection method to fuse the results of three different BERT-CRF-NER models. In the following sections, we call this ensemble model as BERT-CRF-NER(intersection).

2.3 Model Fusion

BERT-QA(union) model and BERT-CRF-NER(intersection) model we use have different advantages in this task. BERT-QA model can more accurately recognize the entity of a particular category event, but it can only recognize a single result. BERT-CRF-NER can recognize all the entities in the event text, but it may get more error results. So we take the union of the results of these two models as the final results.

3 Experiment

3.1 Datasets

The CCKS 2019 event entity extraction for the field of finance task provides 17815 labeled corpus as training dataset. Among these labeled data, 14674 items are valid data, and the remaining 3141 event texts are labeled as empty. And the task also provides 3500 data as verification sets. 135519 event text with obfuscated data are provided as testing dataset.

3.2 Evaluation

The task uses the precision, recall and F1 to evaluate the recognition effect of the event entity. The precision of the event entity is calculated as

$$\mathrm{P} = \frac{\text{The number of correctly recognized entities}}{\text{Total number of the recognized entities}} \tag{2}$$

The recall of the event entity is calculated as

$$\mathrm{R} = \frac{\text{The number of correctly recognized entities}}{\text{Total number of the labeled entity}} \tag{3}$$

The F1 is calculated as

$$\mathrm{F1} = \frac{2 * P * R}{P + R} \tag{4}$$

3.3 Experimental Settings

The pretrained BERT model we use is BERT-WWM, which is retrained on new Chinese dataset by Harbin Institute of Technology.

3.3.1 Hyper-parameter of BERT-QA

(1) Max iteration: 3
(2) Learning rate: 3e-5
(3) Batch size: 16
(4) Max sequence length: 384

3.3.2 Hyper-parameter of BERT-CRF-NER

(1) Max iteration: 3
(2) Learning rate: 5e-6
(3) Dropout rate: 0.5
(4) Batch size: 16
(5) Max sequence length: 512

3.4 Experiment Result

In this section, we compare the BERT-QA(single), BERT-QA(union), BERT-CRF-NER(single), BERT-CRF-NER(inter) and the BERT-QA(union) + BERT-CRF-NER(inter) model. The comparative results are listed in Table 1.

Table 1. Results of various model on validation set.

Models	F1 Score
BERT-QA(single)	0.8921
BERT-QA(union)	0.9081
BERT-CRF-NER(single)	0.9088
BERT-CRF-NER(intersection)	0.9130
BERT-QA(union) + BERT-CRF-NER(intersection)	0.9260

The best result of our model BERT-QA(union) + BERT-CRF-NER(intersection) on official testing datasets is 0.8390 which rank the first place.

4 Conclusion

In this paper, we present a joint financial event entity extracting method based on integrating of sequence labelling and question answering. For different sub-model results, we use different strategies such as intersection and union to fuse the results.

The experimental results show that the fusion model has a higher recall rate, and the different fusion methods selected according to the characteristics of the model also ensure the accuracy of the results.

At last, our best submission achieves 83.90 points, which is the best score in the ranking list. In the future, we will focus on attempting different method to ensemble NER- QA and more effective extraction of financial Named Entities Recognition.

References

1. Nadeau, D., & Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3-26(2007).
2. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P.: Natural language processing (almost) from scratch. Journal of machine learning research, 12(Aug), 2493-2537(2011).
3. Devlin, J., Chang M, W., Lee, K., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805(2018).
4. Cui, Y., Che, W., Liu, T., et al.: Pre-Training with Whole Word Masking for Chinese BERT[J]. arXiv preprint arXiv:1906.08101(2019).
5. The Easiest way to use Bert, https://spaces.ac.cn/archives/6736
6. Huang, Z., Xu, W., & Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991(2015).
7. Chen, S., Huang, Y., Huang, X.: HITSZ-ICRC: A Report for SMM4H Shared Task 2019-Automatic Classification and Extraction of Adverse Effect Mentions in Tweets. In Proceedings of the Fourth Social Media Mining for Health Applications (# SMM4H) Workshop & Shared Task, 47-51(2019).