Automatic Information Extraction for Financial Events by Integrating BiGRU and Attention Mechanism

Automatic Information Extraction for Financial Events by Integrating BiGRU and Attention Mechanism

Jiaheng Lu[1], Weirong Liu[1*]

[1]Central South University, School of Computer Science, and Engineering, 932 Lushan South Road, Changsha, Hunan 410083, China

[*]email: frat@csu.edu.cn

Abstract. In this paper, an information extraction method for financial events written in Chinese is proposed. The core entities of the causes and results, as well as the verbs and conditions are extracted from the financial events reported by websites. The method takes the original words and the part-of-speech of words as two inputs. BERT encoder is utilized to transform the original sentences to word-embedding vectors, which then are send to BiGRU to extract the sematic features. And a full-connected network is overlapped on BiGRU to reduce the impact of “covariate shift”. For the second input, the original sentences are cut by Chinese cut-word tool ‘jieba’ to get the part-of-speech of words, which are then transformed by self-attention mechanism to get global dependencies. The two outputs for word-embedding vectors and part-of-speech of words are combined and then decoded by CRF. Finally, the Viterbi algorithm is utilized to get the best sequences. The experiment results validate the effectiveness of the proposed method.

1. Introduction

Nowadays, with the development of information technology and social economy, more and more people get in touch with finance. Many chances related to financial news. However, prosperous financial information comes out on the Internet every day. The speed of dealing with financial reports has become a tough work on financial area. It’s hard to meet the real requirement with massive financial information which comes out on the websites manually every day. Thus, it is important to develop an automatic information extraction method to deal with the financial reports timely.

Natural language processing (NLP) is the branch of Artificial Intelligence that tries to enable computers understand human languages [1]. Traditional Information Extraction based on NLP is aimed to identify the entity like personal or place name in the sentences. But there is still unsatisfactory place in the area of analysing reasons and results in the financial events which are written by Chinese. In this paper, a novel method based on BiGRU and self-attention mechanism is proposed to identify the key words, including core nouns of the reasons, verbs and conditions of the reason, core nouns of the result, verbs and conditions of the results in the sentences of financial reports written by Chinese.

2. Method of information extraction

The model of the method is presented as Fig.1, which has two inputs. The left input handles the semantics and the right input handles the part-of-speech of words. Left input is a 1-dimensional array with 10000 entries. Here, 10000 means it can accommodate 10000 sentences. After the BERT encode, we get a 2-dimensional matrix with 100000 x 144. The first dimension also corresponds to the sentences. And the second dimension is the length of sentence. If a sentence is shorter than 144, we will pad it with "0" in the rear. And a sentence which is longer than 144 will be truncated. The right input is the 3-demensional tensor with 10000 x 144 x 29. The first two dimensions are same as the left input. The third dimension means 29 part-of-speech flags. [PAD], a~z, [CLS], [SEP] are utilized as flags to illustrate the part-of-speech of words. [PAD], [CLS], [SEP] means pad, start and separation of sentences. After the CRF decodes, the core words are output as an array with 10000 entries.

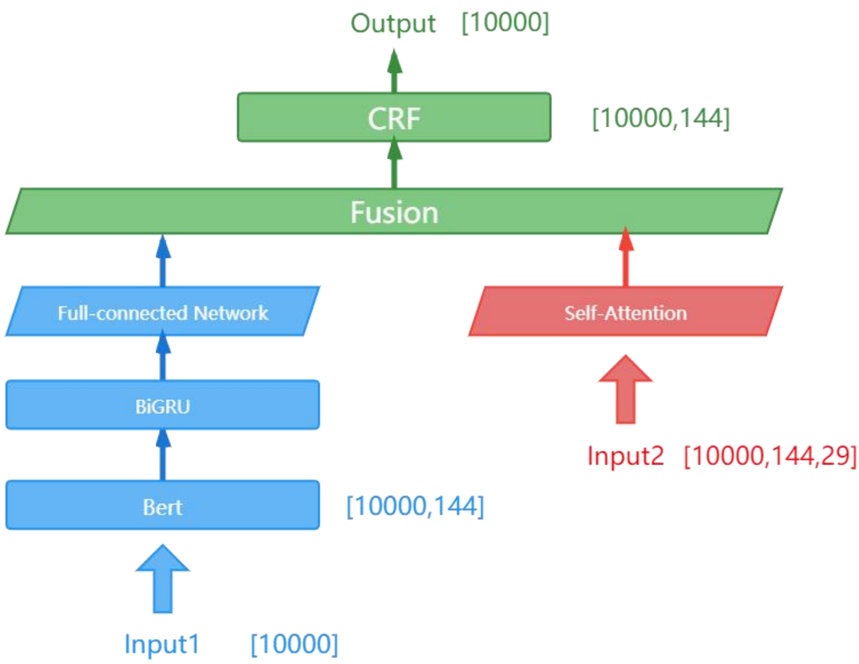


Fig. 1 The model of automatic information extraction.

2.1. Acquirement of training samples

In order to get more detailed and comprehensive sentences about finance, A RegEx based on web crawler developed by ourselves is used to get the financial information in the past decate. The information contains news over the world, financial report, social reviews. Most of the information comes from securities website and financial website. Then the key words are tagged manually, including core nouns of the reasons, verbs and conditions of the reasons, core nouns of the results, verbs and conditions of the results in the sentences.

2.2. Dealing with the semantics

2.2.1. Word encoder. Recently, Transformer [2] has achieved great results in various tasks in NLP field through attention mechanism. As for Chinese corpus, Transformer can also generate character level embeddings. And BERT [3] is based on the encoder of Transformer. After inputting one sentence, the BERT layer will encode the sentence by using different vectors to represent words, and the vectors also has the relationship of different words. The prepared model is named as 'bert-based-chinese' on Bert layer, and the words are transformed by Bert layer into word-embedding vector.

2.2.2. Sematic feature extraction. Then the word-embedding vectors are sent to the BiGRU layer. In this layer, the dependency relation between words is learned, and the semantic features of the context are extracted. Classical RNN model and LSTM [4] can be taken as the basic model to extract the sematic features. However, compared with classical RNN and LSTM, GRU can get similar result, but it runs faster because of the less calculation of tensors [5]. Because simple GRU cannot take full advantage of the context semantic information, the reversed GRU is introduced here to learn semantic information. The input sequence is propagated through forward and reverse network, then the two outputs of network are concatenated.

2.2.3. The feature fixing. In the process of training, especially after GRU layer, it's easy to appear the problem named " Internal Covariate Shift " , and appear exploding gradient. Thus, the layer of Full-connected network is added, and the input mean and variance of one layer of neurons are fixed to lower the influence of covariate shift.

2.3. Learning of the part of the and self-attention mechanism
This layer is aimed to catch the relationship between semantics and part-of-speech in the whole situation. Attention mechanism is the core of Transformer structure [2], as a part of which, self-attention calculation can draw global dependencies among different parts of an input sentence.

2.4. Combination of the learning and decode
In the Fusion layer, the vectors of original sentence and the vectors of part-of-speech are combined. Then the relationship between words and characteristics of the sentence can be learned better. The process of fusion is showed below:

$$F = W \cdot Concat(input_l, input_r) + B \quad (1)$$

After the fusion layer, Conditional Random Fields (CRF) is used to decode the sequence. 14 tags are set including beginning of the core noun of reason, end of the core noun of reason and so on. Then the Viterbi algorithm is utilized to get the best probability and the output the corresponding sequence. In the training process, the negative logarithmic likelihood of the maximum likelihood function is directly taken as the loss function.

3. Experiment results and analysis

3.1. Parameter setting

3.1.1. Choice of epochs. In the training, 80% data is used to train, and 20% is used to verify. The batch size between 16 and 128 are test thoroughly. The results shown that if the batch size is bigger than 64, then the accuracy will decrease a lot. Because when the batch size becomes bigger, loss of gradient and reverse propagation in each training cycle will decrease. Then the model cannot be trained fully. In the meantime, too big batch size leads to the increasement of randomness of data in each epoch. And the model cannot learn the needed features.

While if the batch size is smaller than 32, the rate of usage of BiGRU will be influenced, and the speed rate of training decrease greatly. Thus, the batch size is tested between 32 and 64. The figure 4 and figure 5 show the difference.

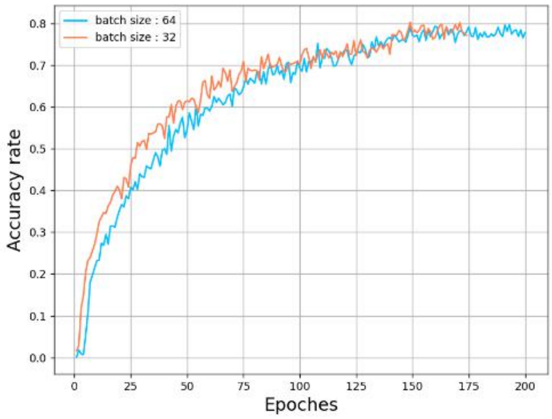


Fig. 2 Accuracy of information extraction on test set

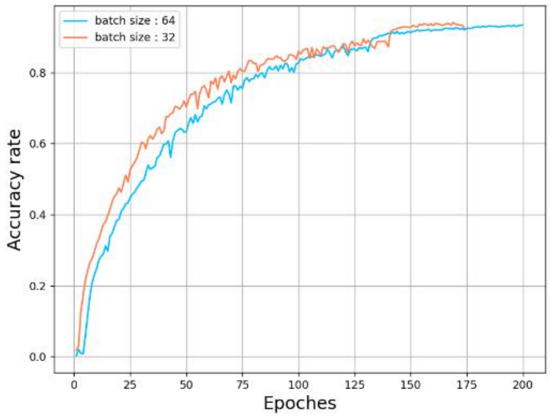


Fig. 3 Accuracy of information extraction on training set

It can be seen that accuracy changes little whether we choosing 32 or 64 as batch size. But choosing 32 as batch size can save a lot of time for training.

3.1.2. Avoidance of overfitting. The early stopping is employed to avoid overfitting. The training will stop when monitoring indicator stops improving. Original times of epochs is set as 200.

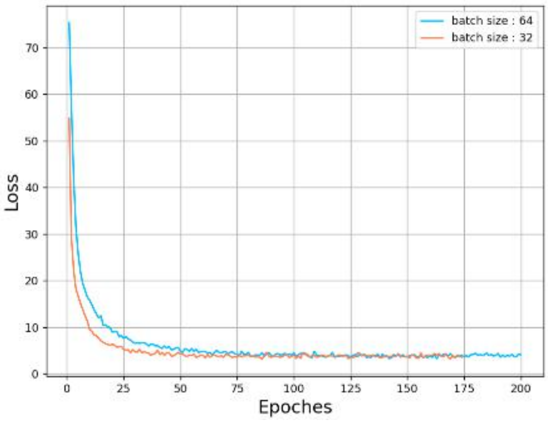


Fig. 4 Loss of valid set

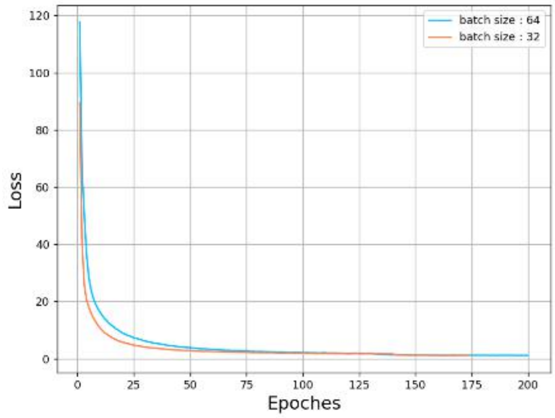


Fig. 5 Loss of training set

It can be seen when batch size is set as 32, the train will stop early and the calculation is saved, while the training loss is similar to that of batch size with 64. Thus, in the experiment, batch size is chosen as 32.

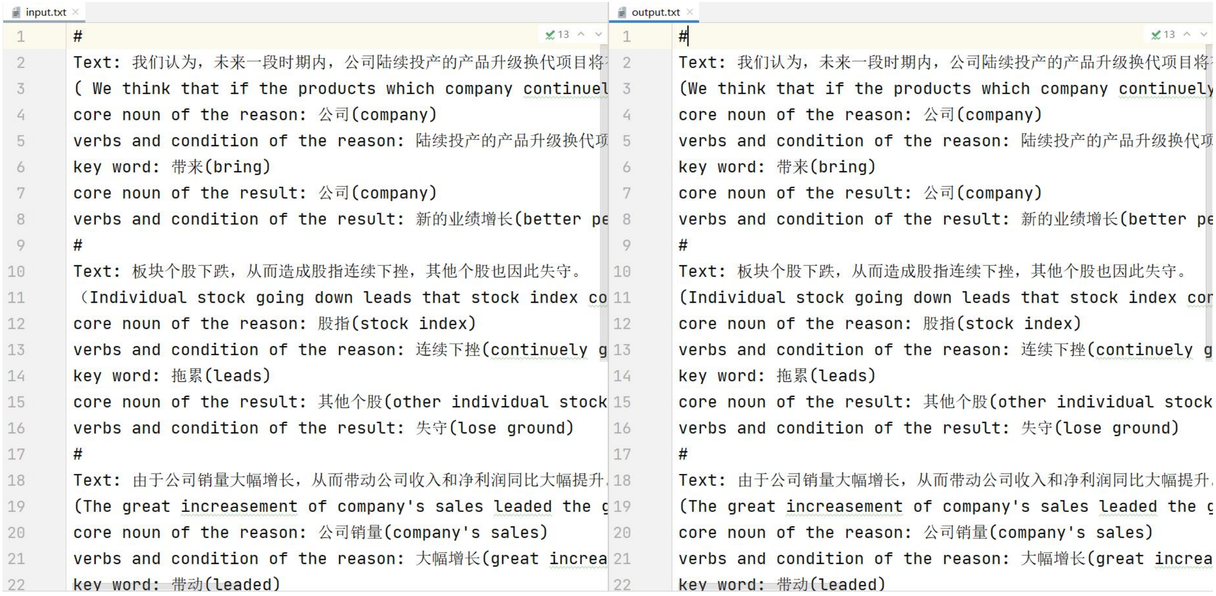
3.2. Experiment results

An example of information extraction for the financial reports is demonstrated in Fig. 7. In the figure, left is the model output; right is the expected label. A detailed example is offered firstly.

Tab. 1 The extract core words

Text:	石油工人罢工导致油价上涨	(That the oil workers struck led to oil price rising)
core noun of the reason:	石油工人	(the oil workers)
verbs and condition of the reason:	罢工	(struck)
key word:	导致	(led to)
core noun of the result:	油价	(oil price)
verbs and condition of the result:	上涨	(rising)

Journal of Physics: Conference Series **2171** (2022) 012001 doi:10.1088/1742-6596/2171/1/012001



```
input.txt
#
Text: 我们认为，未来一段时期内，公司陆续投产的产品升级换代项目将
( We think that if the products which company continuel
core noun of the reason: 公司(company)
verbs and condition of the reason: 陆续投产的产品升级换代项
key word: 带来(bring)
core noun of the result: 公司(company)
verbs and condition of the result: 新的业绩增长(better pe
#
Text: 板块个股下跌，从而造成股指连续下挫，其他个股也因此失守。
（Individual stock going down leads that stock index co
core noun of the reason: 股指(stock index)
verbs and condition of the reason: 连续下挫(continuely g
key word: 拖累(leads)
core noun of the result: 其他个股(other individual stock
verbs and condition of the result: 失守(lose ground)
#
Text: 由于公司销量大幅增长，从而带动公司收入和净利润同比大幅提升
(The great increasement of company's sales leaded the g
core noun of the reason: 公司销量(company's sales)
verbs and condition of the reason: 大幅增长(great increa
key word: 带动(leaded)
```

```
output.txt
#
Text: 我们认为，未来一段时期内，公司陆续投产的产品升级换代项目将
(We think that if the products which company continuely
core noun of the reason: 公司(company)
verbs and condition of the reason: 陆续投产的产品升级换代项
key word: 带来(bring)
core noun of the result: 公司(company)
verbs and condition of the result: 新的业绩增长(better pe
#
Text: 板块个股下跌，从而造成股指连续下挫，其他个股也因此失守。
(Individual stock going down leads that stock index cor
core noun of the reason: 股指(stock index)
verbs and condition of the reason: 连续下挫(continuely g
key word: 拖累(leads)
core noun of the result: 其他个股(other individual stock
verbs and condition of the result: 失守(lose ground)
#
Text: 由于公司销量大幅增长，从而带动公司收入和净利润同比大幅提升
(The great increasement of company's sales leaded the g
core noun of the reason: 公司销量(company's sales)
verbs and condition of the reason: 大幅增长(great increa
key word: 带动(leaded)
```

Fig. 6 Comparison of extract result and expected label

Tab. 2 Results of training

Batch_size	Accuracy rate of training set	Accuracy rate of test set	Loss of training set	Loss of test set
32	92.39%	77.18%	1.224	3.721
64	93.88%	77.81%	1.177	4.019

4. Conclusion

In this paper, an information extraction method is proposed for financial events written in Chinese which applies BERT to encode the sentences, BiGRU and self-attention mechanism to handle the sematic features and part-of-speech respectively. Then CRF and Viterbi algorithm is used to decode and obtain the best output. Experimental results demonstrate that the proposed method can achieve extraction performance with high accuracy.

References

[1] Al-Ghamdi S A, Khabti J and Al-Khalifa H S 2017 Exploring NLP web APIs for building Arabic systems *Twelfth International Conference on Digital Information Management (ICDIM)* pp 175-178

[2] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA)* pp 5998-6008

[3] Devlin J, Chang M W, Lee K and Toutanova K 2019 BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding arXiv1810. 04805v2[cs. CL]

[4] Hochreiter S and Schmidhuber J 1997 Long short- term memory *J. Neural computation* **9(8)** pp 1735-1780

[5] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation arXiv 1406.1078v3[cs. CL]