SEBERTNets: Sequence Enhanced BERT Networks for Event Entity Extraction Tasks Oriented to the Finance Field

Congqing He[1], Xiangyu Zhu[1], Yuquan Le[1], and Yuzhong Liu[1] Jianhong Yin[1]

JD Digital
`hecongqing@hotmail.com`

Abstract. Event extraction lies at the cores of investment analysis and asset management in the financial field, and thus has received much attention. The 2019 China conference on knowledge graph and semantic computing (CCKS) challenge sets up a evaluation competition for event entity extraction task oriented to the finance field. In this task, we mainly focus on how to extract the event entity accurately, and recall all the corresponding event entity effectively. In this paper, we propose a novel model, **S**equence **E**nhanced **BERT Net**works (SEBERTNets for short), which can inherit the advantages of the BERT,and while capturing sequence semantic information. In addition, motivated by recommendation system, we propose Hybrid Sequence Enhanced BERT Networks (HSE-BERTNets for short), which uses a multi-channel recall method to recall all the corresponding event entity. The experimental results show that, the F1 score of SEBERTNets is 0.905 in the first stage, and the F1 score of HSEBERTNets is 0.934 in the first stage, which demonstarate the effectiveness of our methods.

Keywords: Event Extraction · BERT · Finance.

1   Introduction

Event Extraction, a challenging task Information Extraction, aims at detecting and typing events, and extracting different entity from texts. Event extraction lies at the cores of investment analysis and asset management in the financial field, and thus has received much attention. However, most of research is rarely involved in the finance field. For this purpose, the 2019 China conference on knowledge graph and semantic computing (CCKS) challenge sets up a evaluation competition for event entity extraction task oriented to the finance field. The goal of the evaluation is to extract the event entity according to a given real news corpus and event type.

To this end, Liu et al. [5] proposed to exploit argument information explicitly for event detection via supervised attention mechanisms. There are still several questions.(1) How to extract the event entity accurately and effectively? (2) How to recall all the corresponding event entity effectively when there exists multiple event entity in a text? Motivated by these, we propose a novel model, Sequence

Enhanced BERT Networks, which can inherit the advantages of the BERT [1], and while capturing sequence semantic information. In addition, motivated by recommendation system, we propose Hybrid Sequence Enhanced BERT Networks , which uses a multi-channel recall method to recall all the corresponding event entity.

Our mainly contributions are shown as follows:

(1) We propose a novel model, Sequence Enhanced BERT Networks, which can inherit the advantages of the BERT,and while capturing sequence semantic information.

(2) We propose Hybrid Sequence Enhanced BERT Networks , which uses a multi-channel recall method to recall all the corresponding event entity.

(3) The experimental results show that, the F1 score of SEBERTNets is 0.905 in the first stage, and the F1 score of HSEBERTNets is 0.934 in the first stage, which demonstarate the effectiveness of our methods.

2   Our Methods

Figure 1 describes the architecture of SEBERTNets, which primarily involves the following four components: (i) Input layer, (ii) BERT layer, (iii) Sequence layer, (iv) Output layer.

2.1   Input Layer

The description of text can be seen as a char sequence x $= \{x_1, \cdots, x_n\}$, and the corresponding event type can be seen as a char sequence t $= \{t_1, \cdots, t_m\}$. We first remove irregular punctuation, special text, etc, and contact them together. The description of text x and the corresponding event type t are then encoded as input.

2.2   BERT Layer

Inspired by Devlin et al [1], BERT obtains new state-of-the-art results on lots of natural language processing tasks. We introduce BERT as the layer of SEBERT-Nets, to produce sequence semantic representation, and capture global semantic representation.

$$l_1, l_2, \cdots, l_{n+m} = \mathbf{BERT}(x_1, \cdots, x_n, t_1, \cdots, t_m) \tag{1}$$

Where $l_i \in \mathbb{R}^d$, [1] and i represents $i - th$ element.

---

[1] In our experiment, d=768.

Start End

| C | T_1 | $\cdots$ | T_n | $T_{[sep]}$ | T'_1 | $\cdots$ | T'_M |

BERT

| $E_{[CLS]}$ | $E_{Tok\ 1}$ | $\cdots$ | $E_{Tok\ n}$ | $E_{[SEP]}$ | $E_{Tok\ 1}$ | $\cdots$ | $E_{Tok\ M}$ |

| [CLS] | Tok 1 | $\cdots$ | Tok n | [SEP] | Tok 1 | $\cdots$ | Tok M |

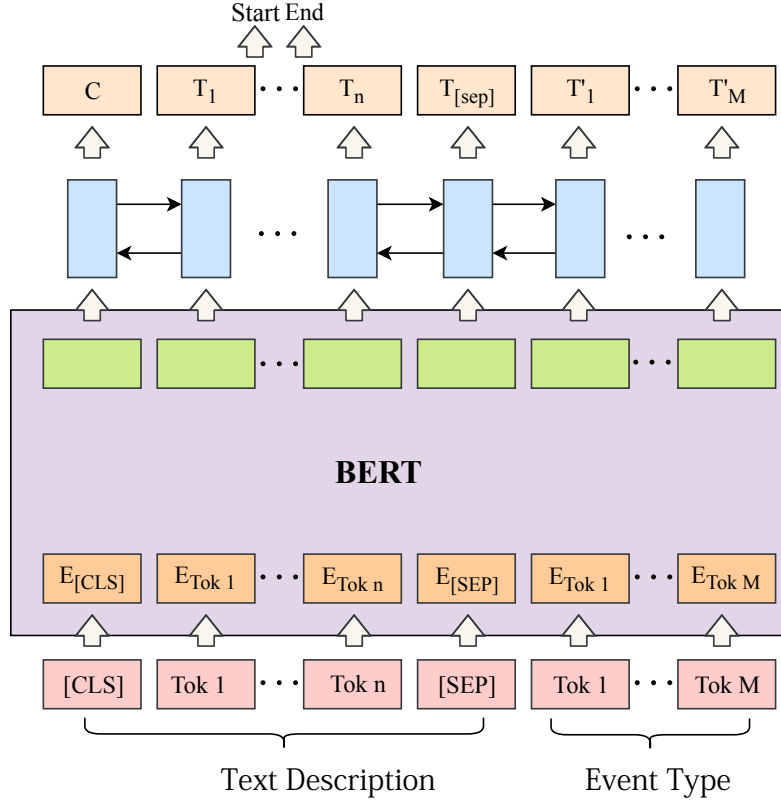Text Description   Event Type

Fig. 1. The architecture of SEBERTNets.

2.3   Sequence Layer

Since the BERT layer only captures the position information of the text through the position vector, then the SEBERTNets introduces the Sequence layer, which captures the sequence semantic information of the text, and enhances the semantics of the text sequence based on the BERT layer. In addition, the Sequence layer introduces a mask operation. Since the text is usually indefinitely long, the introduction of a mask can effectively alleviate the bias caused by the text filling.

$$f_t = \sigma(W_f l_t + U_f h_{t-1} + b_f), \tag{2}$$
$$i_t = \sigma(W_i l_t + U_i h_{t-1} + b_i), \tag{3}$$
$$o_t = \sigma(W_o l_t + U_o h_{t-1} + b_o), \tag{4}$$
$$\widetilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c), \tag{5}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t, \tag{6}$$
$$h_t = o_t \odot tanh(c_t). \tag{7}$$

Where l_t represents the input of the recurrent layer at time step t. f_t, i_t and o_t means forget gate, input gate, and output gate respectively. $\odot$ denotes element-wise multiplication of two vectors. To consider forward and backward contextual representation of the text, we use BiLSTM instead of LSTM as fact encoder in advance. The BiLSTM generates $h_{t\{t=1,\cdots,T\}}$ by concatenating a forward LSTM and a backward LSTM.

2.4   Output Layer

The BERT layer and the Sequence layer can capture the global semantic information and sequence semantic information of the text. Based on the text and event type, the SEBERTNets model predicts the beginning and end of the text of the event body. Finally, it is decoded by the start position and the end position to obtain the corresponding event body.

2.5   Optimization

Although Adam [4]   performed well at the beginning of model training, S-GDStochastic gradient descent (SGD) [6]   was superior to the Adam method in the later stages of training. Therefore, we introduce a new method, SWAT-S [3] , which first starts training with the Adam method and switches to SGD when appropriate. Verification on the real data set of the evaluation game, this method can achieve better performance than the Adam optimization method.

3   Experiments

3.1   Dataset

CCKS 2019 evaluation task consists of 17,815 training sets, 3,500 validation sets (preliminary test datasets) and 135,519 test sets (rematch test datasets). Each dataset includes a given text description, event type, and event entity respectively.

3.2   Evaluation

Following [2] , we use F1 score to evaluate the performance of our methods. The F1 score is computed as follows:

$$F1 = \frac{2PR}{P + R} \tag{8}$$

Where P means that number of correctly identified event entities divided by the total number of identified event entities, R represents that number of correctly identified event entities divided by the total number of annotated event entities.

3.3   Hyper-parameter Setting

Since all the text descriptions and event types have been employed for char segmentation and set each text maximum length to 140. We use character-based Chinese BERT model [2] and character-based Chinese BERT-wwm model [3] to training SEBERTNets and variant SEBERTNets models. In addition, GRU is adopted to Sequence layer and the hidden unit is set to 200. The batch size is set to 32 for SEBERTNets and variant SEBERTNets models.

4   Results

4.1   Main Results

Table 1. F1 score of models on preliminary test datasets.

Number of entities	Top 1	Top 2	Top 3	Top 4	Top 5
BERT	0.871	0.887	0.893	0.901	0.902
SEBERTNets	0.905	0.915	0.918	0.921	0.923
HSEBERTNets	**0.914**	**0.922**	**0.926**	**0.931**	**0.934**

In this section, we conduct experiments on preliminary test datasets to demonstrate the effectiveness of the proposed approach. Table 1 shows the performance of our methods on preliminary test datasets.

Overall, we can find the SEBERTNets model outperforms BERT model with a significant margin on preliminary test datasets. SEBERTNets model obtains 2.5% average absolutely considerable improvements on preliminary test datasets, which demonstrates the effectiveness of SEBERTNets model for event entity extraction.

---

[2] https://storage.googleapis.com/bert_models/2018_11_03/chinese_L-12_
H-768_A-12.zip

[3] https://drive.google.com/open?id=1RoTQsXp2hkQ1gSRVylRIJfQxJUgkfJMW

Additionally, we propose HSEBERTNets model to recall all the corresponding event entity. Compared to the SEBERTNets model, HSEBERTNets model achieves better performance, especially, in multiple event entities. The results demonstrates the effectiveness of HSEBERTNets model for multiple event entity extraction.

4.2   Visualization

"员工协助地方政府空气数据造假 先河环保遭深交所问询住维资金: "漏洞""黑洞"如何解","财务造假"

"上海一中院对快鹿集团、东虹桥小贷公司、东虹桥担保公司分别以集资诈骗罪判处罚金人民币十五亿元、二亿元、二亿元；对黄家骝、韦炎平以集资诈骗罪判处无期徒刑，并处罚金；对徐琪以集资诈骗罪、非法吸收公众存款罪两罪并罚判处有期徒刑十三年，并处罚金；对周萌萌等其余 12 名被告人以集资诈骗罪分别判处有期徒刑十五年至九年不等的刑罚，并处罚金","涉嫌非法集资"

"邦信资产抛售彭州民生村镇银行股权中航三鑫(002163)、中航重机(600765)控股股东将变更","实控人股东变更"

Fig. 2. Visualization of BERT layer.

In this section, we select several representative cases to give an intuitive illustration of how the BERT layer help to promote the performance of event entity extraction. As shown in Figure 4.2, red color means that the correct entities extracted by SEBERTNets model, and yellow color means that the incorrect entites extracted by SEBERTNets model. In these cases, we can find that BERT layer focus on the entity of the text well. Several entities are not belonging to the corresponding event type, and BERT layer can be well observed. From this figure, we observe that the BERT layer can capture key entities relevant to current event type.

References

1. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://www.aclweb.org/anthology/N19-1423

2. Hong, Y., Zhang, J., Ma, B., Yao, J., Zhou, G., Zhu, Q.: Using cross-entity inference to improve event extraction. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. pp. 1127–1136. Association for Computational Linguistics (2011)
3. Keskar, N.S., Socher, R.: Improving generalization performance by switching from adam to sgd. arXiv preprint arXiv:1712.07628 (2017)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Liu, S., Chen, Y., Liu, K., Zhao, J., et al.: Exploiting argument information to improve event detection via supervised attention mechanisms (2017)
6. Robbins, H., Monro, S.: A stochastic approximation method. The annals of mathematical statistics pp. 400–407 (1951)