

Research on the Financial Event Extraction Method Based on Fin-BERT

Jing He

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: hj15129793315@163.com

Yongyong Sun

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: sunyongyong@xatu.edu.com

Abstract—Event extraction is based on the event in the text as the subject information, based on the predefined event type and template, the structured event information is extracted, the existing event extraction model is mainly in the general domain, ignoring the prior knowledge in the domain and the dependency information between entities, and the existing methods do not address the problem of event theory dispersion and multiple events. In response to the above issues, this paper proposes a model based on Fin-Bert (Financial Bidirectional Encoder Representation from Transformers) and RATT (Relation-Augmented Attention Transformer). At the same time, this paper will make use of the structured self-attention mechanism to extract the dependencies between entities, use RAAT to fuse the dependency information between entities into sentence coding, and finally use the binary classification method to identify type of event and generate event records. Compared with the baseline method, the F1 value of the event extraction task on the ChFinAnn and Duce-fin datasets was improved by 2.5% and 2.8%, respectively.

Keywords-Event Extraction; Relationship Information; Information Extraction; Natural Language Processing

I. INTRODUCTION

Event extraction takes the event information in the text as the subject information and extracts some specific information according to the predefined event types and templates to obtain structured data. Event extraction, deep learning, natural language processing, knowledge graph and other disciplines are interrelated, and the extracted information can be applied to the fields of public opinion control, the construction of knowledge graphs [1] intelligence collection [2] and information retrieval [3].

In the early days, event extraction was mainly done at the sentence level, which did not take into account the semantic information of the entire context. In order to solve the problem of sentence-level event extraction, we propose document-level event extraction. Nowadays, document-level event extraction is the mainstream method. However, there are two major problems: one is that document-level event contention is dispersed in each sentence, and the other is that there may be multi-event at a documentation.

The extraction of financial events can quickly extract structured financial events from many financial announcements to help investors understand the financial market and provide the most direct reference basis for investment transaction decisions. Nowadays, many document-level event extraction models are in general domain, and the texts in the financial field have their own unique professional characterization, and the existing models have poor extraction effects on financial texts. At the same time, these methods ignore the relationship information between entities, and cannot solve the question of dispersed event parameters and many events. In order to solve the shortcomings in the existing models, this paper studies the event extraction model in the field of Chinese finance, and proposes an event extraction method for Chinese financial corpus, which makes full use of the learning and expression ability of Fin-BERT (Financial Bidirectional Encoder Representation from Transformers), integrates prior knowledge in the financial field, and aims at the problem of scattered event parameters In this paper, a Fin-

BERT-based encoder and an event extraction method fusing entity relationship information Fin-RAAT are proposed, which use the dependency information and Financial information between entities to better extract financial documents with multiple events.

II. RELATED WORK

Early event extraction mainly focused on sentence granularity, extracting event information from the individual sentences of the text. Most of ways are difficult to extract to text scattered in the whole document event information, sentence level event extraction is difficult to extract to the event information in the whole document, so the document level event extraction by the attention of many scholars. Research methods also range from the beginning of pattern matching, machine learning to deep learning.

A. Event extraction method for pattern identification

Event extraction ways of pattern identification is mainly divided into two steps: pattern acquisition and pattern matching. Pattern acquisition is mainly to establish context constraints for the target text and then construct event templates, and pattern matching is based on the constructed event templates to match event information from the text, and the structure of matching depends on the construction of patterns.

Pattern matching event extraction methods based on the smaller datasets size requirements, high accuracy, good interpretability, can be well applied in the corresponding specific domains, but there are also obvious disadvantages: the first model construction relies on the domain expert's understanding of the text, and the effect of the extraction depends on the validity and completeness of the model. Second the model is very sensitive to domain knowledge and difficult to be quickly migrated to other domains. Thirdly, the extraction model is summarized from existing data sets by manual or weakly supervised algorithms, and the extraction model cannot deal with expressions that have not appeared in the data sets or are not representative. Meanwhile, the pattern matching method mainly focuses on the

granularity of sentences, and it is difficult to cope with the situation where the event information is dispersed in different sentences by only relying on the pattern matching method.

B. Machine learning-based event extraction method

Based on the idea of machine learning event extraction method is the event extraction into classification of text fragments, first extract features in the construction model for event extraction, machine learning event extraction method to reduce the dependence on domain knowledge, but also reduce the construction mode of manpower cost, extraction effect, but it also has many shortcomings, first: machine learning method too dependent on the characteristics of manual construction, increase the workload of event extraction, it is also difficult to learn semantic information. Second: great deal high-quality annotated corpus, especially the trigger words in the event detection task. Third: machine learning methods are also stuck on the sentence granularity, and it is difficult to extract the event information scattered throughout the document.

C. Event extraction method based on deep learning

Deep learning based on event extraction methods are mainly used for event extraction by constructing multi-level neural network models. Neural networks can automatically obtain semantic features from the training corpus and has become the current mainstream approach. One class is to use recurrent neural network RNN, long short-term memory neural networks LSTM and Transformer [4] as the main models, which are used to get semantic representations of words and sentences at different granularity, but none of these methods can solve the problems of argument dispersion and multiple events well.

Du [5] et al. shown that document-level event extraction, while focusing on information of whole document, also focuses on the semantics at the sentence level. The thought is to transform event extraction assignment to a sequence notes assignment. Yang [6] et al. proposed the DCFEE method based on this idea, which starts from the

With the rapid development of internet technology, the capability of computational processing of large volumes of text has also been increasing rapidly. This has made the contextual environment of text more complex, and the understanding of contextual semantics by computers directly affects the final outcomes. Consequently, pre-trained models based on language models are being increasingly utilized, among which BERT is one of the most widely used training models. This model was proposed by Google in 2018 and has achieved state-of-the-art results in various tasks within the field of Natural Language Processing (NLP). The BERT model is composed of multiple stacked bidirectional Transformer encoders, which capture the contextual semantic features of text from two directions, thereby better representing natural language text. In practical applications, there is no need to make extensive changes to the structure of specific tasks; instead, by introducing a new output layer and optimizing BERT's performance, advanced models can be established for different tasks. This approach is also referred to as pre-training and fine-tuning.

The Fin-Bert encoder used in our paper is a pre-training model to the Chinese financial domain based on the BERT [10] architecture open-sourced by Entropy Simple Technology AI Lab. The native Chinese BERT input is sliced and diced at the granularity of a word, which doesn't take into account the relationship between co-occurring words or phrases in the domain, and thus fails to learn the implicit prior knowledge in the domain and reduces the model's learning effect. Fin-Bert is trained using full word masks, and the training data are mainly from the financial domain, including financial and financial news, announcements of listed companies and financial encyclopedia entries. The Fin-Bert model takes into account the relationship between Chinese characters and words, which is in line with the linguistic habits of the Chinese language, and permits the model to study the a priori knowledge within financial domain implied between phrases, which makes its extraction in the financial domain better than that of other models. The training model of Fin-Bert is shown in Figure 2.

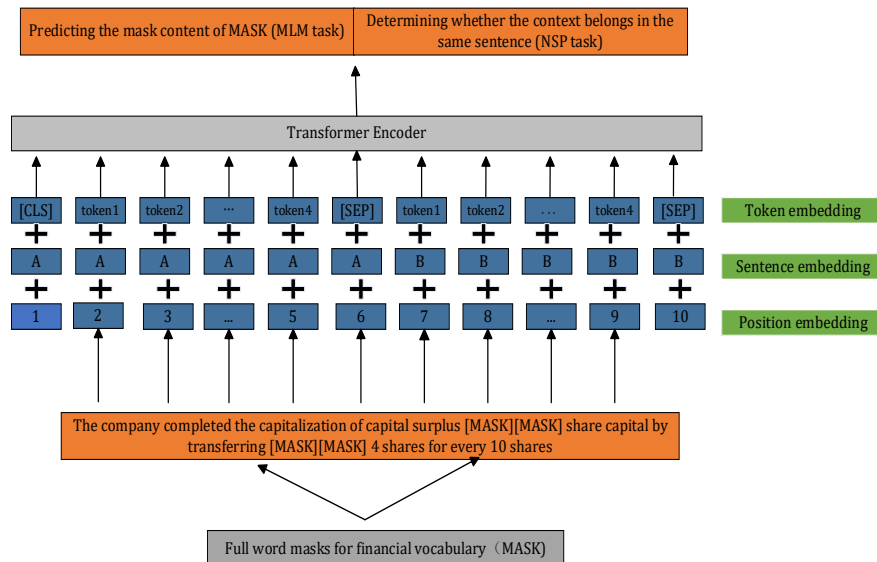


Figure 2. The trained model of Fin-Bert

E. Document relationship extraction

This paper defines the structure between two kinds of entities:

- 1) Co-occurrence structure: Entities appear in the same sentence.
- 2) Two entity references refer to the same entity.

Both structural associations can be present between entities individually or simultaneously. Therefore, this paper has four relationships.

The entities can have both structural associations, either individually or simultaneously. Thus, there are 4 total relational dependencies between entities. The document relationship extraction part utilizes document text D and entity $\{e_1, e_2, \dots, e_j\}$ as data inputs, and outputs the relationship of entities $\{[e_1^h, e_1^t, r_1], [e_2^h, e_2^t, r_2], \dots, [e_k^h, e_k^t, r_k]\}$ pair as the k th ternary, with entities, tail entities and relations represented within the ternary. In order to predict the dependencies between event parameters, structured self-attention network is used for relationship prediction between entities, in this paper, normal cross entropy is used to predict the labels of each entity pair only, and the type of relationship is hypothesized through the following function equation (3):

$$\hat{y}_{i,j} = \arg \max(e_i^T W_r e_j) \quad (3)$$

Where e_i refers to the entity embedding from the encoder model document relational extraction, and W_r denotes biaffine matrix trained by DRE task. The loss function is shown below:

$$L_{dre} = - \sum_{y_{i,j} \in Y} \log P(y_{i,j} | D) \quad (4)$$

Where y_{ij} represents the correct relation label between entities i and j , and D represents all relation pairs of an entity.

F. Entity and sentence encoding

Now embedding the entity mentions and sentences from the Entity Extraction and Representation component and the list of predicted ternary relationships from the Document Relationship Extraction component, the section then encodes the above data and outputs an embedding that is effectively integrated with the relationship information. In this part, this paper will transform the triples into computable matrices

and use the RAAT structure to encode the entity and sentence correlations efficiently, RAAT has a unique attention computation module, which has two parts: self-attention and relationship-enhanced attention computation, and through the alteration module, the dependencies between the entities are integrated efficiently into the document encoding. The RAAT structure is shown in Figure 3.

G. Event record generation

With the output of the previous section, entities and sentence embedding, this event record generation module usually contains two parts: event type classification and event record decoding, given a sentence embedding, for each event type several binary predictions are used to determine whether the corresponding event is recognized or not, and if any of the classifiers recognizes the event kind, then event record decoder is activated to iteratively generating each parameter of the corresponding event type. The loss function is shown below:

$$L_{pred} = - \sum_i \log(P(y_i | S)) \quad (5)$$

The y_i represents the label of the i th event type, and $y_i = 1$ if the event type is the event record of i then $y_i = 1$ else $y_i = 0$, S represents input embedding of sentences.

An EDAG is a series of iterations whose length is equal to the number of roles of a particular type, where the goal every iteration is to forecast the affair parameters of a particular affair character, and the input to each iteration contains embedding of entities and sentences, and the input predicted parameters become part of the input to the next iteration. However, unlike EDAG, this paper employs RAAT instead of the original transformer part in the iteration process, EDAG employs a memory structure to record the extracted parameters and adds a role-type representation to predict the current iteration parameters, however, it is difficult for this method to capture the dependencies between memorized entities, candidate arguments, and sentences, and employs

RAAT instead of the original transformer. The RAAT structure can connect entities and candidate parameters in memory by means of relational triples extracted from the relational extraction section, and a structure representing dependencies can be constructed. Before predicting the event parameters for the current iteration, the matrix T is constructed so that the dependencies can be integrated into the attention computation, and after extracting the parameters and adding them to memory, the new matrix T is generated to accommodate the forecast for next iteration. The RAAT here has the same structure as the RAAT at sentence entity encoding, but the parameters are independent and the event record decoder loss function is defined:

$$L_a = - \sum_{v \in V_D} \sum_e \log(P(y_e | (v, s))) \quad (6)$$

The V_D , v represents the set of nodes and D represents the event parameters of the extracted event records so far s denotes the embedding of the sentence and candidate event parameters and y_e denotes the label of argument candidate e in current step. $y_e = 1$ means that e is the basic event parameter corresponding to the current step event role, else $y_e = 0$.

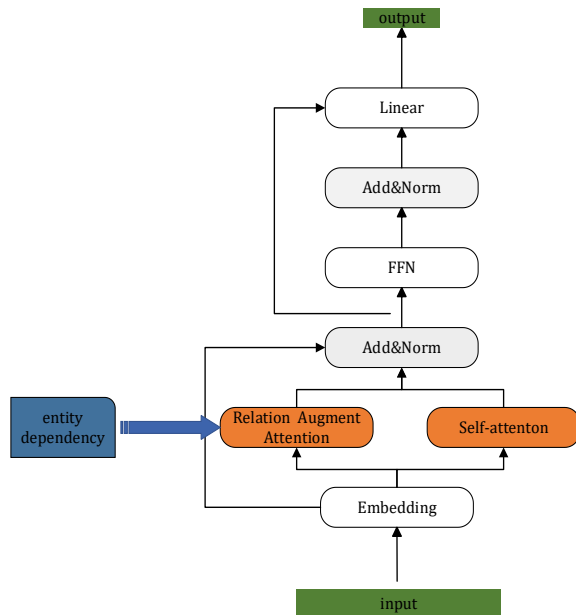


Figure 3. RAAT structure

III. EXPERIMENTS

A. Experimental datasets

This paper use ChFinAnn [7] and Duce-fin [11] as the datasets. The basic information of the datasets is displayed in Table 1.

The ChFinAnn, which contains 32,040 documents, including 9,306 multi-event documents, accounting for about 29% of the total. 98% of the event thesauri are dispersed in different sentences, with the average of 20 sentences contained in each text, the longest text contains about 6,200 Chinese characters. The datasets contain five event types.

The Duce-fin datasets are published by Baidu, Inc. and the results are obtained using an online review in the experiment. The datasets event types are derived from common financial events, 13 in total, which are company listing, equity reduction, shareholder increase, corporate buyback, corporate financing, share repurchase, share pledge, release of pledge, corporate bankruptcy, loss, being interviewed, winning a bid, and executive change. The datasets contain a total of 92 event-theoretic meta-role types.

TABLE I. DATASETS INFORMATION

	Training data	Validation data	Test data	Event types
ChFinAnn	25632	3204	3204	5
Duce-fin	7015	1171	About 3500	13

B. Experimental design

The experimental environment is displayed in Table 2.

TABLE II. EXPERIMENTAL ENVIROMENT

Hardware Information	Configure
Operating system	Windows 10
Internal storage	32GB
CPU	AMD R9-5900HX
Video card	RTX-3080-LAPTOP
Memory	16GB
Exploitation environment	Python 3.7

The following experimental setup was used for this experiment: the same vocabulary as BERT was used for the participle table, the Fin-BERT encoder was used with the base size of the BERT architecture ($dh = 768$, $dl = 12$), a stochastic initialization was used for the parameters to be trained in the model.

C. Experimental results

In this experiment, several baseline algorithms will be selected to compare with this paper's model as a way to display the effectiveness of our paper's model, using the Precision(P), Recall(R), and F1 value(F1) are used as evaluation indexes.

Table 3 demonstrates experimental dismissal of each baseline algorithm of Fin-RAAT and Doc2EDAG, GIT, and PTPCG on the ChFinAnn datasets, and the results prove that the P, R, and F1 value of this paper's model have been improved in comparison with each baseline algorithm.

TABLE III. EXPERIMENTAL EXTRACTION OF CHFANANN DATASETS

	P	R	F1
Doc2EDAG	80.3	70.5	77.5
GIT	82.3	78.4	80.3
PTPCG	88.2	69.1	79.4
Fin-RAAT	84.0	79.9	81.9

Table 4 shows the event extraction results of various baseline algorithms with Fin-RAAT on the Ducee-fin datasets.

TABLE IV. DUEE-FIN DATASETS

	P	Dev		Online test		
		R	F1	P	R	F1
Doc2EDAG	73.7	59.8	66.0	67.1	51.3	58.1
GIT	75.4	61.4	67.7	70.3	46.0	55.6
PTPCG	71.0	61.7	66.0	66.7	54.6	60.0
Fin-RAAT	76.1	71.3	73.0	70.3	56.1	62.8

Analyzing the individual metrics, the F1 value of this paper's model is improved by 6.7%, which is much better than other models. Meanwhile, for

the online test evaluation, this paper's model shows a significant increase of 2.8% in F1 score over the baseline. This result display that this paper's model outperforms existing methods.

Table 5 represents the comparison of the extraction results of the baseline model and the model of this paper on the ChFinAnn datasets in terms of single and multiple events, Single refers to a document in the datasets that contains only 1 event, Multi refers to a document that contains more than one event, and All refers to all the data in the datasets. The results show that Fin-RAAT model is better than other models in extracting multiple events.

TABLE V. COMPARISON OF EVENT EXTRACTION EXPERIMENT RESULTS IN CHFANANN DATASETS

	F1		
	Single	Multi	All
Doc2EDAG	81.0	67.4	77.5
GIT	87.6	72.3	80.3
PTPCG	88.2	69.1	79.4
Fin-RAAT	87.9	75.3	81.9

Figure 4 compares the F1 values of the event extraction results of this paper's method with the baseline model on the ChFinAnn datasets for five types of events, and it can be intuitively seen that for most of the event type. The financial event extraction model proposed in this paper has better extraction results on financial texts.

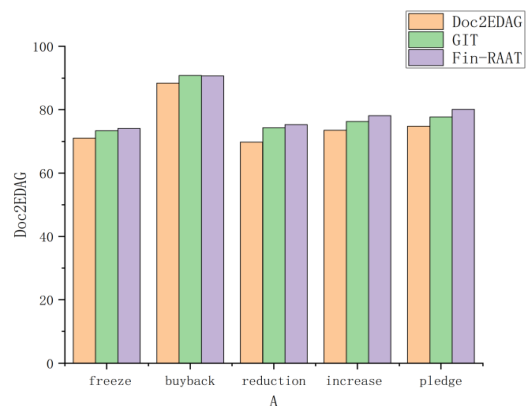


Figure 4. Types of Events Extracted Resulting F1 Values

IV. CONCLUSIONS

In this paper, we propose a financial event extraction model Fin-RAAT based on Fin-BERT as an encoder and fusion of dependencies between entities, which firstly inputs the input documents into the Fin-BERT layer in terms of words for encoding and entity recognition. After obtaining the entity mentions, the structural relationship dependencies between entities are predicted, and the relationship dependencies are fused into the document encoding through RAAT, followed by event type detection through binary classification, and after obtaining the event kinds, the event records are generated basis pre-defined event templates. The experimental results demonstrate that the financial event extraction model proposed in this paper can well solve the problems of event thesis element dispersion and multiple events in document event extraction. The method achieves good results on two benchmark datasets, but there are also areas for improvement in changing the model, the model only considers the structural dependence between entities, and does not consider the causal relationship between entities, and the model is larger, the training cost is high, in the future, we can make the most of this relationship information between entities to improve results of financial event extraction.

REFERENCES

- [1] BOSSELUT A, LE BRAS R, CHOI Y. Dynamic neuro symbolic knowledge graph construction for zero-shot commonsense question answering. Proceedings of the AAAI conference on Artificial Intelligence. 2021, 35(6): 4923-4931.
- [2] LI M, ZHU Y, WANG R. An Empirical Study on Utilizing Neural Network for Event Information Retrieval. International Conference on Computer Science and Communication Technology, 2020, 1621(1): 51-56.
- [3] LI M, ZAREIAN A, LIN Y, et al. GAIA: A fine-grained multimedia knowledge extraction system. Proceeding of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020: 77-86. K Elissa, "Title of paper if known," unpublished.
- [4] WADDEN D, WENBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations. Proceeding of 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5784-5789.
- [5] DU X, CARDIE C. Document-level event role filler extraction using multi-granularity contextualized encoding. Proceedings of Association for Computational Linguistics (ACL). 2020: 634-644.
- [6] YANG H, CHEN Y, LIU K, et al. DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data. Proceeding of Association for Computational Linguistics (ACL). Melbourne, Australia, 2020: 50-55.
- [7] ZHENG S, CAO W, XU W, et al. Doc2EDAG: An end-to-end document-level framework for Chinese financial event extraction. Proceeding of 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019: 337-346.
- [8] XU R, LIU T, LI L, et al. Document-level event extraction via heterogeneous graph-based interaction model with a tracker. Proceeding of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (ACL-IJCNLP). 2021: 3533-3546.
- [9] ZHU T, QU X, CHEN W L, et al. Efficient Document-level Event Extraction via Pseudo-Trigger-aware Pruned Complete Graph. Proceedings of International Joint Conference on Artificial Intelligence. 2022: 4552-4558.
- [10] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceeding of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). 2019: 4171-4186.
- [11] LI, X. et al. Duee-fin: A document-level event extraction dataset in the financial domain released by baidu. (2021) [2023-04-06]. <https://aistudio.baidu.com/competition/detail/46>.