Extracting Financial Events from Raw Texts via Matrix Chunking

Yusheng Huang[1], Ning Hu[2], Kunping Li[2], Nan Wang[2], Zhouhan Lin[1]*
[1]LUMIA Lab, Shanghai Jiao Tong University
[2]Ping An Technology
huangyusheng@sjtu.sjtu.edu.cn, lin.zhouhan@gmail.com

Abstract

Event Extraction (EE) is widely used in the Chinese financial field to provide valuable structured information. However, there are two key challenges for Chinese financial EE in application scenarios. First, events need to be extracted from raw texts, which sets it apart from previous works like the Automatic Content Extraction (ACE) EE task, where EE is treated as a classification problem given the entity spans. Second, recognizing financial entities can be laborious, as they may involve multiple elements. In this paper, we introduce **CFTE**, a novel task for Chinese Financial Text-to-Event extraction, which directly extracts financial events from raw texts. We further present **FINEED**, a Chinese FINancial Event Extraction Dataset, and an efficient MAtrix-ChunKing method called **MACK**, designed for the extraction of financial events from raw texts. Specifically, FINEED is manually annotated with rich linguistic features. We further propose a novel two-dimensional annotation method for FINEED, which can visualize the interactions among text components. Our MACK method is fault-tolerant by preserving the tag frequency distribution when identifying financial entities. We conduct extensive experiments and the results verify the effectiveness of our MACK method.

Keywords: event extraction, text-to-event extraction, matrix chunking

1. Introduction

Event Extraction (EE), which aims to detect the event trigger and arguments, plays a vital role in various domains, such as finance and e-commerce, since it can produce valuable structured information for downstream tasks (Xiang and Wang, 2019), including knowledge base construction (Li et al., 2018) and text summarization (Wang et al., 2016). For example, as illustrated in Figure 1, an EE system is supposed to identify an *Equity Pledge* event.

However, there are two significant challenges for Chinese financial EE in real-world scenarios that need to be addressed. First, financial events need to be extracted from raw texts. The EE task, as demonstrated in the ACE (Grishman et al., 2005) evaluation program, is typically formulated as a classification problem, where the given entity spans are utilized. But in production, entity spans are not available, and we need to segment the Chinese document since there is no delimiter between words (Chen and Ji, 2009), recognize the related entities, and then extract events. The error accumulation should be minimized whenever possible. Second, recognizing financial entities can be laborious work. Financial entities encompass diverse brands, company titles, domestic and foreign names, as well as job positions, resulting in a comprehensive set of elements. For instance, as depicted in Figure 1, the event argument for Stock Company is composed of Chinese characters, English letters, parentheses, numbers, and punctuation. Besides, the term "controlling shareholder" in the Company Law of the People's Republic of China [1] is a single word, rather than the two independent words "controlling" and "shareholder". "Perest" corresponds to an unconventional transliterated Chinese name. These present a domain-specific challenge.

Figure 1: Example of an *Equity Pledge* event triggered by *pledged* with three corresponding event arguments.

In this paper, we make the following threefold contributions. First, we introduce CFTE, a novel task for Chinese financial text-to-event extraction. The goal of CFTE is to directly extract financial events from raw texts. During the inference phase, when provided with a Chinese document that lacks entity spans, CFTE relies on the designed model's capability to effectively segment the document, accurately recognize financial entities within it, and subsequently identify the event triggers and event arguments. Second, we present FINEED, a financial event extraction dataset for the CFTE task. FINEED is a large-scale human-labeled dataset with rich linguistic features. FINEED contains $5,000$ accurately labeled financial event samples, and

*Zhouhan Lin is the corresponding author.

[1] http://www.mofcom.gov.cn/article/b/bf/

LREC-COLING 2024, pages 7035–7044
20-25 May, 2024.

each sample is annotated with entity types, part-of-speech features, and syntactic dependency parsing features. In order to better visualize and exploit the interactions between text components, we propose a novel two-dimensional (2-D) annotation method to present these features. Third, we propose a novel method named MACK that leverages a matrix-chunking approach to jointly segment Chinese texts and recognize financial entities. MACK is able to correct prediction errors through the statistics of the 2-D matrix, which makes MACK fault-tolerant.

In summary, our work presents the following contributions:

- We propose a new task CFTE to extract Chinese financial events from raw texts.
- We present a human-labeled event extraction dataset FINEED for CFTE. A novel two-dimensional annotation method is proposed to present complex interactions between text components.
- We introduce a novel financial text-to-event model MACK for CFTE, which is fault-tolerant for both Chinese word segmentation and character-level tagging.

2. Related Work

Event extraction plays a crucial role in various fields and applications, enabling advanced analysis, search, and decision-making.

2.1. Event Extraction Datasets

The ACE 2005 (Grishman et al., 2005) corpus [2] is a widely used benchmark to extract events from sentences, which advances the development of EE techniques and fosters research in this area. Sun et al. (2022); Douglass et al. (2022); Mutuvi et al. (2020) separately introduce innovative EE datasets for conducting research on pharmacovigilance, international crises, and the emerging threats of infectious diseases. Veyseh et al. (2022); Wang et al. (2021) further drive EE research towards multilingual and multimodal directions. Recently, Han et al. (2022); Li et al. (2021); Yang et al. (2018); Zheng et al. (2019) propose multiple document-level EE datasets to explore challenges beyond the sentence scope. The primary distinction lies in the scattering of entity mentions across multiple sentences. Our FINEED dataset aligns with this particular characteristic as well. However, the aforementioned datasets operate under the assumption of given entity spans. To the best of our knowledge, we are the pioneering researchers in proposing the dataset for extracting events directly from raw texts within the financial domain.

[2]https://www.nist.gov/itl

2.2. Event Extraction Methods

Current EE methods can be categorized into two main groups: statistical methods and deep learning methods. For the statistical methods, Rusu et al. (2014) propose to discover statistical relations to extract events. The prior domain knowledge is introduced to provide the lexical, syntactical, and semantic features (Hearst, 1998; Mejri and Akaichi, 2017). Borsje et al. (2010) propose to utilize pre-defined patterns for EE. For deep learning methods, a multitude of models have been proposed. These models incorporate a variety of neural network blocks, including Convolutional Neural Networks (CNN) (Yubo et al., 2015), Recurrent Neural Networks (RNN) (Nguyen et al., 2016), Graph Neural Networks (GNN) (Liu et al., 2018; Yan et al., 2019; Huang and Jia, 2021), and the attention mechanism (Ren et al., 2022). Transformer-based Pre-trained Language Models (PLM), including BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019), have also become a standard component in these models. However, there have been few studies endeavored to tackle event extraction from raw Chinese financial texts, and we are the first to undertake this task.

2.3. Origin of the Name MACK

In cognitive psychology, chunking is a cognitive process that involves binding small individual pieces of information together to form a cohesive and meaningful whole in memory.

Specifically, according to the chunking hypothesis(Miller, 1956), repeated exposure to a set of stimuli results in the stimuli being organized and represented using larger and larger chunks. Initially, when encountering a stimulus set, each item may be encoded as an individual chunk. However, with repeated exposure, multiple items can be encoded as a single chunk. This hypothesis highlights the concept of our cognitive system actively identifying patterns in stimuli and the environment, and encoding them as increasingly larger chunks of knowledge, thus facilitating learning.

In line with this notion of chunking, our model, MACK (Matrix Chunking), also embraces this idea.

3. Preliminaries

We illustrate the following primary notions: (1) *event mention*: an event mention is a sentence or a document describing a change of states with an event trigger and several event arguments with required roles. (2) *event type*: the type of event mention. (3) *event trigger*: an event trigger is a word that most clearly expresses the event occurrence in its scope, which is usually the main verb. (4) *argument role*: an argument role is a pre-defined field of the event that requires being filled. (5) *event argument*: an event argument is an entity, a value, or an expression that can be filled in a specific set of argument roles. For example, in Figure 1, the event mention contains an *equity pledge* event with the event trigger *pledged* and three argument roles.

4. CFTE Task Formulation

We present CFTE, a novel task for Chinese financial text-to-event extraction, aimed at extracting events from raw texts in the financial domain. This task emulates the extraction of financial events in real-world application scenarios, distinguishing it from the ACE event extraction task, which is formulated as an entity classification problem.

The CFTE task is formulated as follows. Given a document comprised of m Chinese characters and a set of p entities $\mathcal{E} = \{e_i\}_{i=1}^{p}$, where $e_i = \{m_j\}_{j=1}^{q}$ refers to q entity mentions and m_j covers a span of Chinese characters, this task aims to extract events with the pre-defined event types, event triggers, and corresponding event arguments. The event trigger is usually a verb with a part-of-speech tag *Verb*. Event arguments are usually financial entities in $\mathcal{E}$. Both of them contain multiple Chinese characters, which requires aggregating character embeddings.

5. FINEED Dataset Construction

Our FINEED dataset is a meticulously human-annotated collection of Chinese financial EE dataset derived from raw texts. It encompasses essential information, such as Chinese word boundaries, entity types, grammatical features, and event mentions.

The construction of the FINEED dataset involves four steps: financial text collection, Chinese word segmentation, financial entity recognition, and grammatical feature identification. Furthermore, a novel 2-D annotation method is proposed to present these features.

5.1. Financial Text Collection

To ensure the authenticity of our data sources, we gather financial announcements by crawling online websites. Specifically, the following media platforms are included: (1) Sina Finance [3], the largest and most influential financial online media in China, covering over 80% of important industry conferences and events; (2) Xinhua Finance [4], the national financial information platform, which offers information services for financial institutions through branches of Xinhua News Agency and a global network of professional economic analysts.

[3]https://finance.sina.com.cn/

Dataset	Train	Dev	Test	Total
ACE 2005	529	30	40	599
FINEED	**3000**	**1000**	**1000**	**5000**

Table 1: Comparison between the FINEED dataset and ACE 2005 event extraction dataset.

Our focus centers around five event types that significantly impact the valuation of listed companies: equity pledge, equity repurchase, equity underweight, equity overweight, and equity freeze. To ensure data quality, we enforce a requirement that, for each event mention, the number of existing event arguments exceeds at least half of the number of specified argument roles. We discard documents with a character count below 128. For each event type, we randomly select $1,000$ documents for human annotation. Overall, the FINEED dataset contains $3,000$ documents for the training set, $1,000$ documents for the development set, and $1,000$ documents for the testing set. Compared with the ACE 2005 event extraction dataset (Grishman et al., 2005), as shown in Table 1, our FINEED dataset is 8.3 times larger in total. [5]

5.2. Chinese Word Segmentation

To ensure efficient annotation of Chinese word segmentation, we initially employ the widely-used Stanford CoreNLP toolkit to segment the event text. To ensure annotation consistency, annotators initially undergo standardized annotation training and review fundamental financial knowledge. Subsequently, an annotator reviews and validates the segmentation results, with particular attention given to financial entities such as company titles and job positions. Next, another annotator conducts a secondary inspection to ensure precision. When two annotators have differing opinions, a third annotator is introduced into the annotation process, and the majority rule is applied. This means that the annotation result with the most consensus among the annotators is chosen.

[4]https://www.cnfin.com/

[5]FINEED, being manually annotated, is not compared to datasets constructed using distant supervision methods.

Dataset	Entity Type	Total
FEED (Li et al., 2021)	person, organization, date, number	4
FINEED	company title, branch office, national institution, organization name, department, person name, job position, job title, location name, number with units, time point, time interval	12

Table 2: Comparison of entity types between different event extraction datasets.

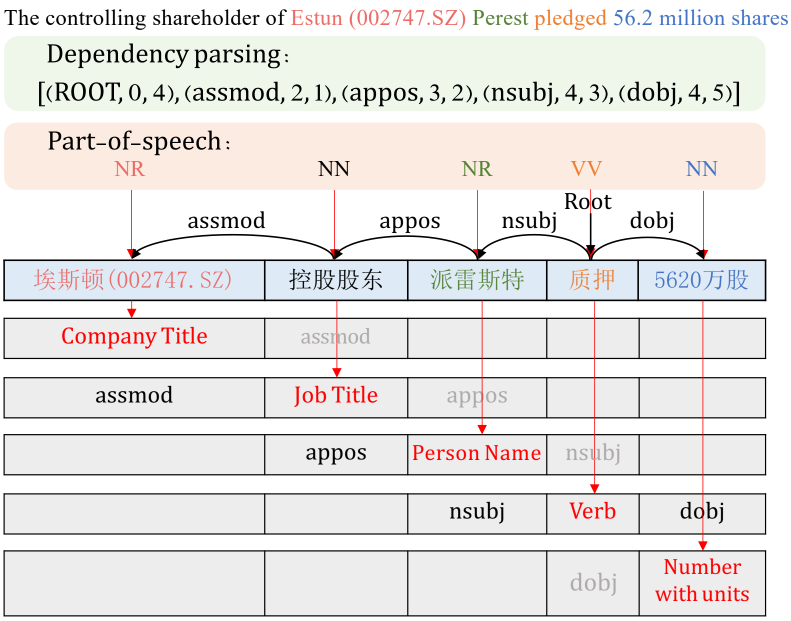

Figure 2: An example of the 2-D matrix. The red tags correspond to each attentive position. The black tags are directly assigned according to dependency parsing tags. The grey tags are determined based on symmetry. The empty places are O tags, indicating an absence of any tag.

5.3. Financial Entity Recognition

Relevant entity types have been found to contribute to the improvement of information extraction models, as evidenced by previous studies (Xu et al., 2021b; Zhou and Lee, 2022). This is because certain event arguments exhibit a significant correlation with the entity type. Therefore, we have summarized a set of entity types in the financial domain as shown in Table 2. Notably, the annotation of entity types remains a rarity across the majority of event extraction datasets (Yang et al., 2018; Zheng et al., 2019), with our investigation yielding the identification of this characteristic within the recent FEED (Li et al., 2021) dataset. Our FINEED dataset, in contrast to FEED, not only encompasses entity types that bear enhanced relevance to the financial domain, but also provides a level of detail that surpasses the former.

We begin by training a commonly used BI-LSTM-CRF sequence tagging model (Huang et al., 2015) to recognize entity mentions based on the designed entity types. To ensure the precision and effectively tackle the challenge of coreference, an annotator conducts an examination, meticulously verifying the recognition of entity mentions and subsequently resolving references pertaining to the same entity. In this way, we provide coreference information for entities.

5.4. Grammatical Feature Identification

Grammatical features, including part-of-speech features and syntactic dependency parsing features, have proven to be advantageous for event extraction by offering valuable auxiliary information (Ritter et al., 2012; Liu et al., 2018; Yan et al., 2019; Shuo et al., 2020; Nguyen et al., 2021). To obtain these tags for our FINEED dataset, we leverage the Stanford CoreNLP toolkit to generate them. Subsequently, annotators undertake the task of carefully examining and rectifying the potential errors, resulting in significant time savings.

5.5. 2-D Annotation Method

We propose an innovative 2-D annotation method to visually represent the aforementioned features. Specifically, as illustrated in Figure 2, for an n-word document segmented from m Chinese characters in §5.2, we annotate a 2-D matrix with the size of $n \times n$. For the i_{th} row in the matrix, the i_{th} entry is the primary attentive position. If the i_{th} word is a financial entity with a corresponding type listed in Table 2, we assign the entity type to the (i, i) entry of the matrix. Otherwise, we assign the part-of-speech tag identified in §5.4 to the (i, i) matrix entry. For instance, the $(4, 4)$ matrix entry in Figure 2 is annotated with the part-of-speech tag *Verb*. For the remaining off-diagonal entries, we assign the dependency parsing tags generated in §5.4 to the corresponding entries in the 2-D matrix based on the coordinates of the head and dependent words. For example, the *nsubj* tag is assigned to the $(4, 3)$ entry on the 2-D matrix. Furthermore, these off-diagonal tags are also assigned to symmetric entries in the 2-D matrix. This allows each row or column to conveniently retrieve the dependency relations between the attentive word and all other words. Otherwise, as shown in the 5_{th} row in Figure 2, this row where the event argument *56.2 million shares* is located will not be able to query the relations with other words.

6. MACK Method

An illustration of MACK is shown in Figure 3.

6.1. Character Tagging

Given an m-character text, m different tag sequences are predicted, forming an $m \times m$ square matrix $M \in \mathbb{R}^{m \times m}$. For each attentive character position $p \in \{1, 2, ..., m\}$, we predict the character tag at position p as well as the tags for all other characters in the same row of the matrix. Following the 2-D annotation method outlined in §5.5, characters within a word are assigned the same tag.

Specifically, considering a sequence s_i centered on the position p, we feed s_i to the Pre-trained Language Model (PLM), e.g., BERT (Devlin et al., 2019), after adding the special tokens [CLS] and [SEP] to generate the contextual representations,

$$\mathcal{B}_{S_i} = \mathrm{PLM}(s_i), \tag{1}$$

where $\mathcal{B}_{S_i} \in \mathbb{R}^{d \times h}$ denotes the hidden states of the last layer. $d = |s_i| + 2$ and h is the embedding dimensionality.

Then, we adopt a Feed-Forward Network (FFN) with LeakyReLU to predict tags T_{s_i} for s_i,

$$p_{s_i} = \sigma(\mathrm{FFN}(\mathcal{B}_{s_i})), \tag{2}$$

$$T_{s_i} = \mathrm{Argmax}(p_{s_i}), \tag{3}$$

where σ denotes the softmax function.

We utilize the cross-entropy loss $\mathcal{L}_{Tag}$ for optimization, which is computed based on the character target tags y_{s_i} as annotated in §5.5,

$$\mathcal{L}_{Tag} = -\log p(y_{s_i}|p_{s_i}). \tag{4}$$

6.2. Chinese Word Segmentation

Given the square matrix $M \in \mathbb{R}^{m \times m}$, we segment the raw text based on the criterion that the number of different adjacent non-O tags between two adjacent rows and columns at the corresponding position in M is more than half. For example, as shown in Figure 3(b), the text is segmented between characters c_3 and c_4 because every two adjacent tags in the 3_{rd} and 4_{th} rows of the matrix, as well as in the 3_{rd} and 4_{th} columns, are different. Similarly, the text is not segmented between characters c_6 and c_7. The reason for simultaneously counting the data of rows and columns in M is to increase fault tolerance based on the symmetry of M.

The criterion is formulated as follows,

$$\mathrm{Seg}_i = \lceil \frac{\alpha_i + \beta_i}{2m - \theta_i(O)} - \frac{1}{2} \rceil, \tag{5}$$

where α_i and β_i are the number of different adjacent tags in the i_{th} and $(i+1)_{th}$ rows and the i_{th} and $(i+1)_{th}$ columns respectively, calculated as,

$$\alpha_i = \sum_{1 \le j \le m} \chi_{(M_{i,j} \neq M_{i+1,j})}, \tag{6}$$

$$\beta_i = \sum_{1 \le j \le m} \chi_{(M_{j,i} \neq M_{j,i+1})}. \tag{7}$$

$\theta_i(O) = \gamma_i + \delta_i$ denotes the number of adjacent tags that are both the O tag, where

$$\gamma_i = \sum_{1 \le j \le m} \chi_{(M_{i,j} = M_{i+1,j} = O)}, \tag{8}$$

$$\delta_i = \sum_{1 \le j \le m} \chi_{(M_{j,i} = M_{j,i+1} = O)}. \tag{9}$$

Notably, χ_p represents the characteristic function, where p denotes the condition, i.e., $\chi_p = 1$ if p is true, else $\chi_p = 0$.

Finally, we segment the text at position i if $\mathrm{Seg}_i = 1$ for $\forall 1 \le i \le m - 1$.

6.3. Matrix Chunking

We first partition the matrix $M \in \mathbb{R}^{m \times m}$ into $M \in \mathbb{R}^{n \times n}$ using the word coordinates, where n is the number of words. Then, we assign the mode tag of each submatrix as the designated tag,

$$T_{i,j} = \mathrm{Mode}(\phi(m_{i,j}) \oplus \phi(m_{j,i})), \tag{10}$$

where $m_{i,j} \in \mathbb{R}^{\mathrm{len}(\mathrm{L}_{i+1} - \mathrm{L}_i) \times \mathrm{len}(\mathrm{L}_{j+1} - \mathrm{L}_j)}$ is the corresponding submatrix of M, and $\mathrm{L}_i = i$ when

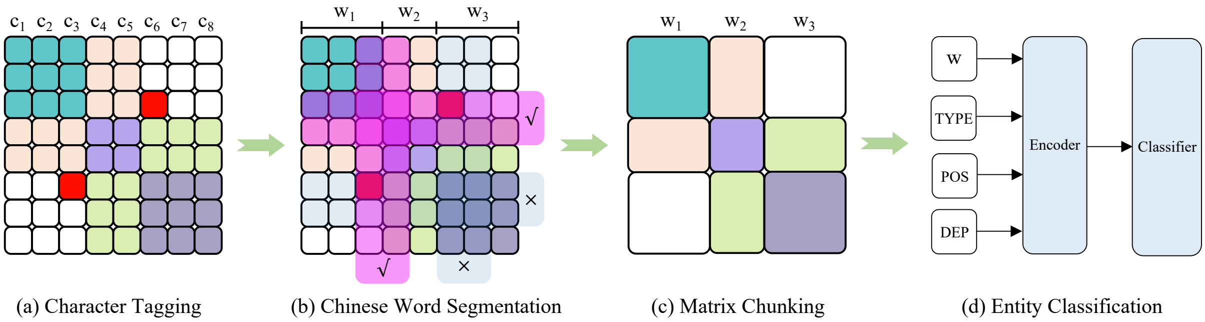

Figure 3: An illustration of the MACK method, including four key tiers. (a) Implement character-level tagging using distinct colors for different tags. Tags in red denotes incorrect predictions. (b) Segment the text based on the statistics of adjacent rows and columns at the corresponding position. (c) Partition the matrix into submatrices using word coordinates, and assign the mode tag of each submatrix as the designated tag. (d) Detect event triggers and classify entities into the event roles by leveraging multiple features.

$\text{Seg}_i = 1$. ϕ denotes the frequency distribution of tags in $m_{i,j}$. $\oplus$ denotes the element-wise addition. Mode denotes returning the mode in statistics.

Comparison with BIEO tagging. The BIEO tagging scheme (Dai et al., 2019) maintains no fault tolerance. But our MACK method is fault-tolerant by preserving the frequency distribution and selecting the mode tag. For example, in Figure 3(c), the red error tags shown in Figure 3(b) have been corrected. Additionally, the process is traceable and visible, allowing us to backtrack to the intermediate results of our MACK method and identify the cause of incorrect predictions. This traceability is crucial in real-world application scenarios.

6.4. Entity Classification

Given an entity mention that spans j characters, we utilize the log-sum-exp pooling to derive the representation $m_i \in \mathbb{R}^h$,

$$m_i = \log \sum_{k=1}^{j} \exp\left(c_k\right), \tag{11}$$

where $c_k \in \mathbb{R}^h$ is the k_{th} character embedding.

To capture the grammatical features between the entity mention and other words, we propose a dependency attention mechanism for information aggregation. Specifically, if the entity mention is the i_{th} word, we extract the corresponding i_{th} row of the matrix $M \in \mathbb{R}^{n\times n}$ that the entity mention is being attended to. Next, we generate the existing t dependency parsing tag embeddings $D_i \in \mathbb{R}^{t\times h}$ and their corresponding word embeddings $W_i \in \mathbb{R}^{t\times h}$ as in equation 11, and concatenate them as $D^{dep} = [D_i; W_i] \in \mathbb{R}^{t\times 2h}$. Then, the dependency attention d_i is calculated as follows,

$$r_{ij} = \mathbf{v}^T \tanh(\mathbf{W}^d \cdot D^{dep} + \mathbf{b}^d), \tag{12}$$

$$\alpha_{ij} = \text{softmax}(r_{ij}) = \frac{\exp\left(r_{ij}\right)}{\sum_{k=1}^{t} \exp\left(r_{ik}\right)}, \tag{13}$$

$$d_i = \sum_{j=1}^{t} \alpha_{ij} * D^{dep}, \tag{14}$$

where $\mathbf{W}^d$, $\mathbf{v}$, and $\mathbf{b}^d$ are learnable parameters.

We feed the concatenation of the entity mention representation m_i, the entity type embedding t_i, and the dependency attention d_i to an FFN to generate the contextualized grammatical entity mention representation $\boldsymbol{h}_i$,

$$\boldsymbol{h}_i = \text{FFN}\left([m_i; t_i; d_i]\right). \tag{15}$$

Finally, the entity representation $\boldsymbol{e_i}$ is computed by aggregating q entity mentions,

$$\boldsymbol{e_i} = \log \sum_{i=1}^{q} \exp\left(\boldsymbol{h}_i\right). \tag{16}$$

The prediction of the event type relies on the embedding of the special token [CLS] from the pre-trained language model. We utilize a linear classifier to determine the event type,

$$p_{type} = \sigma(\mathbf{W} \cdot \boldsymbol{e}_{\text{CLS}} + \mathbf{b}), \tag{17}$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters. The cross-entropy loss is employed for optimization,

$$\mathcal{L}_{Type} = -\log p(y_{type}|p_{type}), \tag{18}$$

where y_{type} is the target event type.

The event trigger is predicted by applying feed-forward networks (FFN) to the word embeddings labeled as *Verb* in the diagonal of the 2-D matrix,

$$p_{trig} = \text{softmax}(\text{FFN}(\boldsymbol{e}_{\text{Verb}})). \tag{19}$$

We leverage the cross-entropy loss for trigger classification as follows,

$$\mathcal{L}_{Trig} = -\log p(y_{trig}|p_{trig}), \tag{20}$$

Model	Precision	Recall	F1-score
Doc2EDAG (Zheng et al., 2019)	75.32	69.80	72.46
GIT (Xu et al., 2021a)	77.34	70.02	73.50
SCDEE (Huang and Jia, 2021)	78.57	72.45	75.39
MACK	**82.15** $_{\uparrow +3.58}$	**80.53** $_{\uparrow +8.08}$	**81.33** $_{\uparrow +5.94}$

Table 3: Main results (%) on the FINEED test set. The best results are in bold. Improvements are compared with the SCDEE baseline.

Model	Precision	Recall	F1-score
MACK	**82.15**	**80.53**	**81.33**
MACK w/o ET & DA	80.03	74.97	77.42
MACK w/o ET	80.78	77.24	78.97
MACK w/o DA	81.66	78.47	80.03

Table 4: Results (%) on the FINEED test set for ablation study. ET denotes the entity type. DA denotes the dependency attention.

where y_{trig} denotes the target event trigger.

The event arguments are predicted by applying an FFN on entity representations e_i, optimized by the binary cross-entropy loss denoted as $\mathcal{L}_{Role}$,

$$p_{role} = \mathrm{sigmoid}(\mathrm{FFN}(\boldsymbol{e}_i)), \tag{21}$$

$$\begin{aligned}\mathcal{L}_{Role} = -\sum_{i=1}^{C} y^i_{role} \log p^i_{role} + \\ (1 - y^i_{role}) \log(1 - p^i_{role}),\end{aligned} \tag{22}$$

where y_{role} is the target argument and C denotes the number of event argument roles.

The final objective function of our MACK method is as follows,

$$\mathcal{L}_{sum} = \mathcal{L}_{Tag} + \mathcal{L}_{Type} + \mathcal{L}_{Trig} + \mathcal{L}_{Role}. \tag{23}$$

7. Experiments

We perform extensive experiments to evaluate the FINEED dataset and the MACK method. Our experimental results confirm the difficulty of financial word segmentation and extracting financial events from raw texts for existing methods.

7.1. Experiment Settings

We utilize the hidden states from the last layer of BERT Chinese pre-trained model with 768 dimensions as the character embeddings. The dropout rate is set to be 0.2 in FFN in both equation 2 and equation 15. The dropout rate of FFN in equation 19 and equation 21 is 0.4. We employ the AdamW (Loshchilov and Hutter, 2019) optimizer for model optimization.

An event argument is considered correct when both the event type and its corresponding argument role are correct. The performance is evaluated by precision, recall, and F1-score.

7.2. Baselines

We compare the proposed MACK method with the following competing models:

- Doc2EDAG (Zheng et al., 2019) generates an entity-based directed acyclic graph to extract events by using sequential path-expanding.
- GIT (Xu et al., 2021a) constructs a heterogeneous graph interaction network to capture the global interactions among entity mentions.
- SCDEE (Huang and Jia, 2021) proposes a document graph construction method to leverage the entity and sentence interactions.

These baseline models primarily concentrate on entity classification without incorporating entity recognition. For fair comparisons, these reimplemented models utilize the entities extracted by the MACK method.

7.3. Quantitative Results

Table 3 presents the experimental results for event extraction on the FINEED test set. MACK achieves impressive precision of 82.15%, recall of 80.53%, and an F1-score of 81.33%, surpassing the performance of the baseline models by a significant margin. These results highlight the effectiveness of our method. Our MACK model efficiently captures text features, facilitates integrated processing of multiple steps, including Chinese word segmentation and entity recognition, and proficiently extracts

Model	**Precision**	**Recall**	**F1-score**
Stanford Corenlp V.3.9.0	86.26	85.34	85.78
Jieba V.0.42.1	87.31	84.75	86.01
Stanford Corenlp V.4.5.4	90.23	88.74	89.47
MACK	**97.24**	**96.55**	**96.89**

Table 5: Results (%) on the FINEED test set for Chinese word segmentation in the financial domain.

The controlling shareholder of Estun (002747.SZ), Perest, pledged 56.2 million shares

Stanford Corenlp Version 3.9.0:	埃斯顿, (, 002747. SZ,), 控股, 股东, 派, 雷斯特, 质押, 5620万, 股
Stanford Corenlp Version 4.5.4:	埃斯顿, (, 002747. SZ,), 控股, 股东, 派雷斯特, 质押, 5620万, 股
Jieba Version 0.42.1:	埃斯, 顿, (, 002747, . , SZ,) ,控股, 股东, 派, 雷斯特, 质押, 5620, 万股
MACK (Target):	埃斯顿(002747. SZ), 控股股东, 派雷斯特, 质押, 5620万股

Figure 4: A case study on Chinese financial word segmentation using different toolkits. Distinct colors are used to highlight different financial entities.

events from complex raw financial texts. Notably, compared to the SCDEE model, the MACK method exhibits an 8.08% increase in recall, indicating that our MACK method can retrieve event arguments more effectively.

7.4. Ablation Study

We conduct an ablation study to evaluate the effectiveness of financial entity types and dependency parsing features. Specifically, we remove the entity type embedding t_i and the dependency attention d_i from equation 15, and the results are presented in Table 4. First, we examine the model performance when both t_i and d_i are removed, referred to as MACK w/o ET & DA. This configuration leads to a decline of 2.12% in precision, 5.56% in recall, and 3.91% in F1-score, highlighting the importance of these features. Next, we investigate the impact of financial entity types by removing t_i, referred to as MACK w/o ET. This modification results in a decrease of 2.36% in F1-score, confirming the significance of entity types. Furthermore, we assess the influence of grammatical features by removing d_i, referred to as MACK w/o DA. The removal of dependency attention leads to a decline of 2.06% in recall, underscoring the usefulness of dependency features.

7.5. Detailed Analysis

We evaluate the performance of Chinese word segmentation, as presented in Table 5. As baselines, we compare our approach with Stanford CoreNLP Version 3.9.0, Version 4.5.4, and Jieba Version 0.42.1, which are popular Chinese word segmentation toolkits. In terms of overall performance on the FINEED test set, MACK achieves 97.24% precision, 96.55% recall, and 96.89% F1-score, surpassing the baselines by a significant margin. These results provide strong evidence for the effectiveness of our MACK method in Chinese financial word segmentation.

7.6. Case Study for Word Segmentation

We present a case study comparing the results of Chinese financial word segmentation from different methods. Figure 4 illustrates the results obtained from our MACK method, Stanford Corenlp Version 3.9.0, Version 4.5.4, and Jieba Version 0.42.1. The latter three toolkits fail to recognize the stock company Estun (002747.SZ). Besides, given their restricted understanding of domain-specific knowledge in the financial field, these baseline methods also struggle to precisely identify the term "controlling shareholder" (in black) in the Company Law. Stanford Corenlp Version 3.9.0 and Jieba Version 0.42.1 also struggle with the English transliteration name *Perest*. Identifying numbers with units, such as 56.2 million shares, presents a challenge. These results validate the domain-specific challenge in the financial field, showcasing the promise of our proposed CFTE task and the MACK method.

8. Conclusion

We introduce CFTE, a novel task for extracting Chinese financial events from raw texts. To support this task, we present FINEED, a large-scale human-labeled event extraction dataset specifically designed for the financial domain. FINEED utilizes a novel two-dimensional annotation method, which effectively presents the complex features of financial texts. To address the challenges of CFTE, we propose the MACK method, which employs a matrix chunking approach for text-to-event extraction. One advantage of the MACK method is its fault-tolerant tagging capability, which maintains robustness in event extraction. Experimental results demonstrate the effectiveness of our MACK method in the context of CFTE.

9. Acknowledgements

This work is supported by the National Key Research and Development Program of China Grant (No. 2022YFB3904201) and the National Natural Science Foundation of China (NSFC) Grant (No. 62106143).

Jethro Borsje, Frederik Hogenboom, and Flavius Frasincar. 2010. Semi-automatic financial events discovery based on lexico-semantic patterns. *International Journal of Web Engineering and Technology*, 6(2):115.

Zheng Chen and Heng Ji. 2009. Language specific issue and feature exploration in chinese event extraction. In *HLT-NAACL (Short Papers)*, pages 209–212. The Association for Computational Linguistics.

Dai Dai, Xinyan Xiao, Yajuan Lyu, Shan Dou, Qiaoqiao She, and Haifeng Wang. 2019. Joint extraction of entities and overlapping relations using position-attentive sequence labeling. In *AAAI*, pages 6300–6308. AAAI Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*, pages 4171–4186. Association for Computational Linguistics.

Rex W. Douglass, Thomas Leo Scherer, J. Andrés Gannon, Erik Gartzke, Jon R. Lindsay, Shannon Carcelli, Jonathan Wilkenfeld, David M. Quinn, Catherine Aiken, Jose Miguel Cabezas Navarro, Neil Lund, Egle Murauskaite, and Diana Partridge. 2022. Introducing the icbe dataset: Very high recall and precision event extraction from narratives about international crises. *CoRR*, abs/2202.07081.

Ralph Grishman, David Westbrook, and Adam Meyers. 2005. Nyu's english ace 2005 system description. *ACE*, 5.

Cuiyun Han, Jinchuan Zhang, Xinyu Li, Guojin Xu, Weihua Peng, and Zengfeng Zeng. 2022. Duee-fin: A large-scale dataset for document-level event extraction. In *NLPCC (1)*, volume 13551 of *Lecture Notes in Computer Science*, pages 172–183. Springer.

Marti A Hearst. 1998. Automated discovery of wordnet relations. *WordNet: an electronic lexical database*, 2.

Yusheng Huang and Weijia Jia. 2021. Exploring sentence community for document-level event extraction. In *EMNLP (Findings)*, pages 340–351. Association for Computational Linguistics.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *CoRR*, abs/1508.01991.

Guozheng Li, Peng Wang, Jiafeng Xie, Ruilong Cui, and Zhenkai Deng. 2021. FEED: A chinese financial event extraction dataset constructed by distant supervision. In *IJCKG*, pages 45–53. ACM.

Zhongyang Li, Xiao Ding, and Ting Liu. 2018. Constructing narrative event evolutionary graph for script event prediction. In *IJCAI*, pages 4201–4207. ijcai.org.

Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. In *EMNLP*, pages 1247–1256. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *ICLR (Poster)*. OpenReview.net.

Mohamed Mejri and Jalel Akaichi. 2017. A survey of textual event extraction from social networks. In *LPKM*.

George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Stephen Mutuvi, Antoine Doucet, Gaël Lejeune, and Moses Odeo. 2020. A dataset for multilingual epidemiological event extraction. In *LREC*, pages 4139–4144. European Language Resources Association.

Minh Van Nguyen, Tuan Ngo Nguyen, Bonan Min, and Thien Huu Nguyen. 2021. Crosslingual transfer learning for relation and event extraction via word category and class alignments. In *EMNLP (1)*, pages 5414–5426. Association for Computational Linguistics.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 300–309.

Yubing Ren, Yanan Cao, Fang Fang, Ping Guo, Zheng Lin, Wei Ma, and Yi Liu. 2022. CLIO: role-interactive multi-event head attention network for document-level event extraction. In *COLING*, pages 2504–2514. International Committee on Computational Linguistics.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *KDD*, pages 1104–1112. ACM.

Delia Rusu, James Hodson, and Anthony Kimball. 2014. Unsupervised techniques for extracting and clustering complex events in news. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 26–34.

Wang Shuo, Rao Yuan, Fan Xiaobing, and Qi Jiangnan. 2020. Joint Event Extraction Model based on Multi-feature Fusion. *Procedia Computer Science*, 174:115–122.

Zhaoyue Sun, Jiazheng Li, Gabriele Pergola, Byron C. Wallace, Bino John, Nigel Greene, Joseph Kim, and Yulan He. 2022. PHEE: A dataset for pharmacovigilance event extraction from text. In *EMNLP*, pages 5571–5587. Association for Computational Linguistics.

Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Nguyen. 2022. MEE: A novel multilingual event extraction dataset. In *EMNLP*, pages 9603–9613. Association for Computational Linguistics.

Chengyu Wang, Rong Zhang, Xiaofeng He, Guomin Zhou, and Aoying Zhou. 2016. Event phase extraction and summarization. In *WISE (1)*, volume 10041 of *Lecture Notes in Computer Science*, pages 473–488.

Shuo Wang, Qiushuo Zheng, Zherong Su, Chongning Na, and Guilin Qi. 2021. MEED: A multimodal event extraction dataset. In *CCKS*, volume 1466 of *Communications in Computer and Information Science*, pages 288–294. Springer.

Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.

Runxin Xu, Tianyu Liu, Lei Li, and Baobao Chang. 2021a. Document-level event extraction via heterogeneous graph-based interaction model with a tracker. In *ACL/IJCNLP (1)*, pages 3533–3546. Association for Computational Linguistics.

Wang Xu, Kehai Chen, and Tiejun Zhao. 2021b. Discriminative reasoning for document-level relation extraction. In *ACL/IJCNLP (Findings)*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 1653–1663. Association for Computational Linguistics.

Haoran Yan, Xiaolong Jin, Xiangbin Meng, Jiafeng Guo, and Xueqi Cheng. 2019. Event detection with multi-order graph convolution and aggregated attention. In *EMNLP/IJCNLP (1)*, pages 5765–5769. Association for Computational Linguistics.

Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfee: A document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55.

Chen Yubo, Xu Liheng, Liu Kang, Zeng Daojian, Zhao Jun, et al. 2015. Event extraction via dynamic multi-pooling convolutional neural networks.

Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. 2019. Doc2edag: An end-to-end document-level framework for chinese financial event extraction. *arXiv preprint arXiv:1904.07535*.

Yang Zhou and Wee Sun Lee. 2022. None class ranking loss for document-level relation extraction. In *IJCAI*, pages 4538–4544. ijcai.org.