CFinDEE: A Chinese Fine-Grained Financial Dataset for Document-Level Event Extraction

Tian Zhang
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
zhangtian@wust.edu.cn

Maofu Liu*
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
liumaofu@wust.edu.cn

Bingying Zhou
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
budding@wust.edu.cn

ABSTRACT

Document-level event extraction faces numerous challenges in accurately modeling real-world financial scenarios, particularly due to the inadequacies in existing datasets regarding data scale and fine-grained annotations. The development of datasets is a crucial factor in driving research progress; therefore, we present a high-quality Chinese document-level event extraction dataset, CFinDEE. This dataset, grounded in real-world financial news, defines 22 event types and 116 argument roles, annotating 26,483 events and 107,096 event arguments. CFinDEE aims to address these shortcomings by providing more comprehensive annotations and data augmentation, offering richer resources for document-level event extraction in the financial domain. CFinDEE extends data both horizontally and vertically, where horizontal expansion enriches the types of financial events, enhancing the diversity of the dataset; vertical expansion, by increasing the scale of the data, effectively boosts the practical value of the dataset. Experiments conducted on multiple advanced models have validated the high applicability and effectiveness of the CFinDEE dataset for document-level event extraction tasks in the financial field.

CCS CONCEPTS

• **Computing methodologies** → *Information extraction.*

KEYWORDS

Datasets, document-level event extraction, financial event, multi-event extraction

ACM Reference Format:
Tian Zhang, Maofu Liu, and Bingying Zhou. 2024. CFinDEE: A Chinese Fine-Grained Financial Dataset for Document-Level Event Extraction. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3589335.3651921

*Corresponding author.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3651921

1 INTRODUCTION

In today's rapidly advancing information technology landscape, event extraction (EE) has emerged as a critical task in the field of Natural Language Processing (NLP), attracting widespread attention. This task aims to extract structured event information from unstructured text, including event types and argument roles, to mine valuable information from vast amounts of textual data effectively. With the rise of financial technology, the importance of financial event extraction in the field of NLP has become increasingly prominent [1]. The sharp increase in financial data has made it an urgent need for the financial industry to extract key events from a large volume of text efficiently using natural language processing techniques. For financial practitioners, event extraction enables the rapid and accurate acquisition of critical information such as market dynamics, corporate actions, and regulatory changes. For investors, it provides a more comprehensive market analysis and risk assessment, assisting them in making more informed investment decisions. For regulatory bodies, it allows timely monitoring of market anomalies and risk events, helping maintain the financial markets' stability and security. Therefore, research on financial event extraction deepens the understanding of the operational patterns of financial markets and corporate behavior and promotes market transparency and efficiency, paving new pathways for academic research.

Current datasets in the financial domain focus on annotating sentences such as financial headlines, yet the financial events contained within a single sentence are often incomplete [9]. As illustrated in Figure 1, the arguments for financial profit events annotated in the DuEE-Fin dataset [5] are dispersed across multiple sentences. To more comprehensively understand financial events, it is necessary to conduct analyses that span multiple sentences. For example, a company's financial report may contain various information, including performance, financial indicators, and business outlook. Only by capturing the information from the entire document can the completeness of financial event extraction be effectively achieved. Therefore, document-level event extraction (DEE) has become an important direction of development in the financial domain.

Document	Original Annotation Event	Supplementary Annotation Event
• **ChFinAnn** … **2015年7月6日**…金世旗控股分别增持**21582594股**…公司股价从6月12日收盘价17.3元跌至**7月6日**最低点的**8.27元**，跌幅达**52.20%**。 … On **July 6th, 2015**, **Jin Shi Qi Holdings** increased its holdings by **21,582,594 shares**…The company's stock price fell from its closing price of CNY 17.3 on June 12th to a low of **8.27 yuan** on **July 6th**, representing a decrease of **52.20%**.	Event Type 1:Equity Overweight EquityHolder: 金世旗控股(Jin Shi Qi Holdings) TradedShares: 21582594股(21582594 shares) StartDate: 2015年7月6日(July 6th, 2015) EndDate: 2015年7月6日(July 6th, 2015)	Event Type 2:Stock Price Decline 股票简称(Stock Abbreviation)： 金世旗控股(Jin Shi Qi Holdings) 股价(Stock Price)： 8.27元(8.27 yuan) 下跌幅度(Decrease percentage)： 52.20% 事件时间(Event time)： 7月6日(July 6th)
• **DueeFin:** 豆神教育（300010）发布**2020年**三季度报告，…归属于上市公司股东的净利润亏损**93,499,178.39**元。其中第三季度盈利**1,216,185.18**元，比上年同期下滑**94.03%**。 **DouShen Education** (stock code: 300010) released its **third-quarter report for the year 2020**…The net profit attributable to the shareholders of the listed company incurred a **loss** of **93,499,178.39 yuan**. Specifically, the company recorded a **profit** of **1,216,185.18 yuan** in **the third quarter**, which was **a decrease of 94.03%** compared to the same period in the previous year.	Event Type 1:Financial Loss Trigger：亏损(loss) 公司名称(Company Name)： 豆神教育(DouShen Education) 财报周期(Financial Reporting Cycle)： 2020年三季度(third-quarter report for the year 2020) 净亏损(Net loss)： 93,499,178.39元(93,499,178.39 yuan)	Event Type 2:Financial Profit Trigger：盈利(profit) 公司名称(Company Name)： 豆神教育(DouShen Education) 财报周期(Financial Reporting Cycle)： 第三季度(the third quarter) 盈利金额： 1,216,185.18元(1,216,185.18 yuan) 盈利变化： 下滑94.03%(a decrease of 94.03%)

Figure 1: Data Examples for ChFinAnn and DuEE-Fin

While some DEE datasets exist in the financial domain, they are limited in data scale and fine-grained categorization. For example, the ChFinAnn dataset [23] includes 32,040 financial announcements, but the financial announcements only focus on company-specific information, such as financials and transactions. In comparison, financial news reports, with their content diversity and comprehensiveness, have a clear advantage in showcasing the actual complexity of financial markets. Moreover, this dataset only annotates five types of events, mainly focusing on equity changes, which limits its ability to capture the full scope of the financial market. As shown in Figure 1, ChFinAnn only marks the equity overweight event and neglects the stock price decline event. Furthermore, although the DuEE-Fin dataset expands the event types to 13 and annotates 8,168 news articles, it fails to cover other critical events in the financial sector comprehensively. For instance, as depicted in Figure 1, DuEE-Fin only marks the financial loss event while ignoring the financial profit event, failing to reflect the completeness of financial report information. Additionally, the smaller scale of DuEE-Fin limits its effectiveness in comprehensively capturing financial information. Consequently, there is an urgent need to construct a larger-scale, finer-grained, and more comprehensive DEE dataset to facilitate the development of research and practice in Chinese document-level event extraction within the financial domain.

In the paper, we present CFinDEE, a large-scale, human-annotated Chinese financial DEE dataset.CFinDEE aims to address the shortcomings of existing datasets and meet the need for more profound and more comprehensive financial event information, providing a more comprehensive and challenging research resource for the financial DEE task. The CFinDEE dataset features the following characteristics: (1) Large-scale manual annotation: CFinDEE includes 16,372 news articles, with 26,483 events and 107,096 event arguments annotated, significantly exceeding the scale of existing datasets. This ensures the richness of the dataset, bringing it closer to the real-world financial document environment. (2) Fine granularity: CFinDEE defines 22 types of financial events and 116 argument roles, and introduces the concept of opposing events. By expanding the types of events and argument roles, CFinDEE provides a solid foundation for the differentiation and precise extraction of various financial events, thereby enhancing the effectiveness of financial event extraction. (3) High challenge: CFinDEE faces challenges in argument dispersion and the extraction of multiple events, where 39.9% of the documents contain multiple events, reflecting the complexity of financial documents in real scenarios, providing researchers with more challenging task scenarios.

The contributions of this paper are mainly reflected in three aspects:

- This paper constructs a large-scale dataset for DEE tasks in the financial domain, CFinDEE. This dataset covers various financial event types and provides a realistic and challenging data resource for research in the financial domain.
- To comprehensively evaluate CFinDEE, this paper conducts experiments on nine advanced DEE models. The results demonstrate the dataset's performance and robustness across different models, proving its widespread applicability and effectiveness in financial event extraction.
- By conducting comparative experiments with multiple datasets, not only is CFinDEE's challenge in multi-event extraction tasks validated, but also its unique advantages in terms of data scale and fine-grained classification are demonstrated. Furthermore, the paper analyzes the extraction effects of various event types within CFinDEE and the reasons for performance differences.

2 RELATED WORK

2.1 Task Introduction

The event extraction task comprises two core components: event detection and event argument extraction. Initially, event detection involves identifying and classifying trigger words in the text, a process divided into two sub-tasks: trigger word detection and event type classification [11]. Trigger word detection aims to identify the specific words or phrases that initiate an event, while event type classification categorizes these trigger words into predefined event categories. Subsequently, the text is analyzed through event argument extraction to identify arguments related to a specific event and annotate their argument roles [22]. This stage includes two sub-tasks: argument identification and argument role classification. The former is responsible for extracting entities related to the event from the text, and the latter categorizes these entities into corresponding argument roles. The event extraction framework is designed to precisely extract and understand significant event information from unstructured text, meeting users' cognitive needs for critical financial events.

2.2 Dataset

The development of datasets is crucial for advancing research progress. Datasets for event extraction are divided into two main categories: sentence-level and document-level.

In the realm of sentence-level event extraction(SEE), in 2005, the Linguistic Data Consortium released the ACE-2005 dataset [2], which includes 8 event types and 33 subtypes involving 36 argument types. The ACE-2005 dataset compiles 599 documents from six media types, containing 6000 events, and covers three languages: Chinese, English, and Arabic. In 2020, Baidu Inc. released the DUEE dataset [10] for the event extraction competition of the Language and Intelligent Technology Contest. As the largest publicly available sentence-level Chinese event extraction dataset, it encompasses 65 event types, 19,640 events, and approximately 17,000 data entries.

In 2021, Zhou et al. [24] constructed a few-shot Chinese event extraction dataset named FewFC, focusing on the financial domain. This dataset is primarily sourced from internet news reports and announcements published by listed companies, containing 10 financial event types and 19 argument types, covering 8982 sentences. Although these datasets provide rich resources for financial event extraction, the limited event information in sentences cannot fully capture the contextual environment of events, highlighting the urgent need for document-level datasets.

Recently, a series of financial domain DEE datasets have been successfully introduced. In 2018, Yang et al. [19] constructed the first document-level Chinese event extraction dataset in the financial field, DCFEE, using a distant supervision approach. This dataset encompasses 2,976 financial announcements, containing only four event types: equity freeze, equity pledge, equity repurchase, and equity overweight. In 2019, Zheng et al. released the ChFinAnn dataset, employing a similar distant supervision technique, focusing on Chinese financial announcements from 2008 to 2018. It annotated 32,040 documents, encompassing five types of equity-related events: equity freeze, equity repurchase, equity underweight, equity overweight, and equity pledge, along with 35 types of argument roles. In 2021, Li et al. [7] utilized a distant supervision algorithm to build a Chinese financial event extraction dataset named FEED, drawing from 31,748 company announcements on Chinese financial portal websites from 2008 to 2018. FEED covers the same event types and argument roles as the ChFinAnn dataset, totaling 46,960 event instances. In 2022, Ren et al. [14] introduced a fine-grained event extraction dataset named IREE from an investment perspective, comprising five major news categories and 59 types of risk events. The same year, Han et al. released the DuEE-Fin dataset, covering financial announcements, judicial documents, and news articles, without releasing a test set. DuEE-Fin annotated 8,168 documents, including 13 event types and 92 argument roles. Despite these datasets contributing to the development of event extraction at the financial document level, they still have limitations regarding data scale, quality, and diversity of event types. For instance, datasets constructed using distant supervision methods may contain more noise and are primarily based on financial announcements, limiting content diversity. Moreover, existing datasets have not fully covered key event types within the financial domain. These limitations reveal the challenges and opportunities in developing more comprehensive and high-quality datasets in the financial field for the future.

3 DATASET CONSTRUCTION

This paper constructs a large-scale financial DEE dataset to provide a comprehensive and challenging data resource for event extraction from financial documents. This section details the dataset construction process, divided into three main steps: event schema construction, candidate data collection, and data annotation. The overall construction process of the CFinDEE dataset is illustrated in Figure 2.

3.1 Event Schema Construction

When constructing a dataset, clearly defining the event pattern is first necessary. This process requires not only the identification of key elements of financial events, such as companies, stocks, timing, and amounts, but also the clarification of the relationships among these elements to ensure the dataset's structural and semantic consistency. The precise definition of event patterns provides a framework for event information extraction and understanding, enabling the dataset to capture event details more, thereby supporting complex analyses and predictions about the financial market. Thus, defining event patterns is crucial for the dataset's quality and the analysis's effectiveness. However, existing datasets predominantly focus on stock transaction events, failing to cover all financial events comprehensively. Therefore, to enhance the practicality of the dataset and fully reflect the complexity and dynamism of the financial sector, we construct financial event patterns based on the actual business needs of enterprises.

Based on the needs of financial enterprises and following confirmation with financial experts, we have categorized financial events into six major categories: equity changes, company changes, performance changes, stock market fluctuations, executive changes, business changes, and regulatory penalties. Subsequently, we refined each category to ensure that every event type defined in the model is clear and non-overlapping. In defining the types of events, we also introduced the concept of opposing events, aimed at capturing the dynamic changes and contrary trends in the financial domain, such as financial loss versus profit, stock price increase versus decrease, company listing versus delisting, and market open versus close events, among others. Ultimately, we defined 22 event types and 116 argument roles, constructing a comprehensive and detailed event model. This model encompasses the influential event types in the financial sector. Table 1 displays some defined event types and argument roles. The complete event schema is included in Table 7 in the appendix.

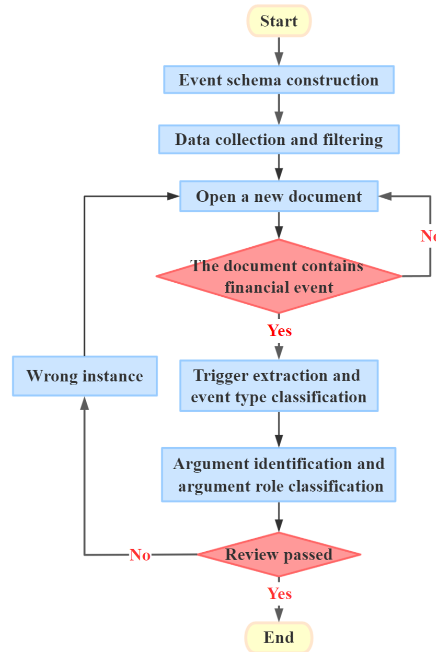

Figure 2: Overall Construction Process of CFinDEE

Table 1: Four Examples of Event Schema in CFinDEE

Event type	Argument roles
Equity Underweight	Stock Ticker, Number of Traded Shares, Price Per Share, Transaction Amount, Transaction Completion Time, Reducing Shareholder, Proportion of Holdings, Proportion of Total Share Capital
Equity Freeze	Equity Holder, Legal Institution, Froze Shares, Amount Frozen, Start Date, Unfreeze Date, Proportion of Holdings, Proportion of Total Share Capital
Financial Profit	Company Name, Financial Reporting Cycle, Earnings Amount, Change in Earnings
Stock Price Increase	Stock Ticker, Growth Rate, Stock Price, Event time

Event Type	企业收购(Corporate Acquisition)	质押(Pledge)	股权冻结(Equity Freeze)
Event Argument Role	• 收购方(Acquirer) • 收购完成时间(Acquisition Completion Date) • 交易金额(Transaction Amount) • 被收购方(Acquiree) • 收购标的(Acquisition Target)	• 质押方(Pledger) • 质押物所属公司(Company of Pledged Asset) • 质押物占持股比(Percentage of Shares Pledged) • 质押物(Pledged Asset) • 质权方(Pledgee)	• 股权持有者(Equity Holder) • 冻结股份(Froze Shares) • 法律机构(Legal Institution)

Figure 3: An Example from CFinDEE

3.2 Candidate Data Collection

News serves as the primary channel for understanding financial events, and its extensive and in-depth content more accurately reflects the realities of financial markets. Consequently, this paper selects news documents as the source for data annotation. Based on their comprehensive coverage of financial events and unique reporting styles, we have carefully selected seven representative financial news websites, including Sina Finance[1], Flush Finance[2], China Economic Net[3], Zhongtong Finance Net[4], China Securities Net[5], East Money[6], and Jiemian News[7]. These websites provide a wide range of content, from instant news to in-depth analysis, covering domestic and international financial news, macroeconomic policies, and individual stock information, ensuring comprehensive coverage of different document formats and content styles. Moreover, the high reliability and frequent updates of these sites ensure the timeliness and quality of the data, facilitating the capture of current market trends.

To ensure the quality of the dataset, this study adopted a rigorous data selection process. We crawled more than 40,000 pieces of data from target websites during the period from 2020 to 2023. After several key filtering steps, we finally selected 21,570 pieces of data that are highly relevant to financial events: (1) Deduplication: In the crawled data, there were instances where the same news event was reported multiple times across different websites, necessitating the removal of duplicate articles. (2) Format verification: Articles that do not meet the standard format, such as those in PDF format or the form of tables, were deleted. (3) Relevance Filtering: Utilizing natural language processing technology for text analysis, news unrelated to financial events, such as non-financial advertisements and personal opinion articles, was filtered out. (4) Authenticity Screening: Based on announcements from exchanges, we excluded data from companies with financial fraud issues to avoid the negative impact of false data on model training. (5) Manual review: The data's relevance, accuracy, and authenticity were further verified through manual review.

3.3 Data Annotation

We adopted the Doccano[8] annotation tool for data annotation. Before the annotation process, we invited four experts in the finance sector and eight experienced annotators for a discussion. Combining insights and suggestions from the experts, we established rigorous annotation guidelines. To ensure the quality of annotation, a mechanism of expert review and cross-validation was implemented to minimize annotation errors as much as possible. As shown in Figure 3, we present a specific annotation example from the CFinDEE dataset, where the annotation process is divided into two main stages: Event detection and event argument extraction.

3.3.1 Event Detection. In the event detection stage, the task of annotators is to categorize documents according to predefined event types, constituting a multi-label classification task. Specifically, annotators are required for each financial news document to determine whether it pertains to one or multiple predefined event categories. If it does, the corresponding event label should be assigned, along with the annotation of the trigger word for the event. Conversely, if the document does not involve predefined events, it should be classified under the "other" category. To enhance accuracy, each news document is independently annotated by two annotators. If there is a discrepancy in the classification results for the same document between the two annotators, the document is submitted to an expert for final adjudication.

As shown in Figure 3, the document contains three financial events: corporate acquisition, pledge, and equity freeze. Among these, the trigger word for the corporate acquisition event is annotated as "purchase," the trigger word for the pledge event as "pledged," and the trigger word for the equity freeze event as " frozen."

3.3.2 Event Argument Extraction. In the event argument extraction stage, annotators label the arguments of each event according to the argument roles defined in the event schema, which is a sequence labeling task. Initially, documents are classified according to event

[1]https://finance.sina.com.cn/
[2]https://stock.10jqka.com.cn/gegugg$_list/index_1.shtml$
[3]http://finance.ce.cn/stock/gsgdbd/index.shtml
[4]https://www.zhitongcaijing.com/
[5]https://www.cnstock.com/
[6]https://www.eastmoney.com/
[7]https://www.jiemian.com/lists/800.html
[8]http://doccano.herokuapp.com/

types, and then, for each event type, the corresponding argument roles are annotated. During the annotation process, annotators are required to label all entities corresponding to each argument role. Entities mentioned repeatedly are annotated only once. It is permitted for the same entity to play different argument roles in different events. A single annotator will independently complete each article, and each annotator is responsible for three different types of events. Each event type is assigned to two annotators, and upon completion, a cross-review is conducted to ensure the consistency and accuracy of the annotation results. If there is a disagreement in the annotation results, the relevant data will be submitted to an expert for final adjudication.

As illustrated in Figure 3, for corporate acquisition events, the annotator labeled "Asia-Pacific Industry" as the acquirer, "March 2020" as the acquisition completion date, "291 million yuan" as the transaction amount, "Lingang Yarno Chemical" as the acquiree, and "equity" as the Acquisition Target. Similarly, in equity pledge events, the annotator needs to label roles such as the pledger, the company of pledged asset, the percentage of shares pledged, the pledged asset, and the pledgee. In the case of equity freeze events, roles such as the equity holder, the frozen shares, and the legal institution must be annotated. Notably, the same entity, such as the "Asi-Pacific Industry," can play different roles in different events; for example, it acts as the acquirer in corporate acquisition events and as the pledger in equity pledge events.

4 DATA ANALYSIS

4.1 Overall Statics

The paper constructs the CFinDEE dataset, which contains 16,372 valid documents, annotating 26,483 financial events and 107,096 arguments. Statistics show that each document contains an average of 408 characters, with the most extended document having 3,558 characters. On average, each document records about 1.6 events and 6.6 arguments, with a single document containing up to 16 events and 56 arguments. Among them, approximately 39.9% of the documents contain multiple events, highlighting the interconnectivity and complexity of events in the financial domain.

CFinDEE encompasses 22 financial event types and 116 argument roles. Figure 4 illustrates the distribution of event types in the dataset, reflecting the diversity and imbalance of financial events in the real world. Among them, equity repurchase, stock price increase, and equity underweight have a higher proportion in the data, while equity freeze, company delisting, and corporate bankruptcy events are relatively rare. Each event type in the dataset includes more than 300 instances, with 50% of the event types having over 1000 instances and 72.7% exceeding 800 instances. Moreover, there is significant co-occurrence between event types such as equity freeze and equity pledge, corporate bankruptcy and company delisting, stock price increase, and financial profit, which is crucial for a deep understanding of the interconnectivity of financial events. Each event type, on average, includes 5 argument roles, with the equity overweight event having the most argument roles, totaling 8, and the company delisting event having the fewest, with only 3. Due to the variation in the number of argument roles among different event types, this presents a requirement for the model to adapt to a complex and variable data structure.

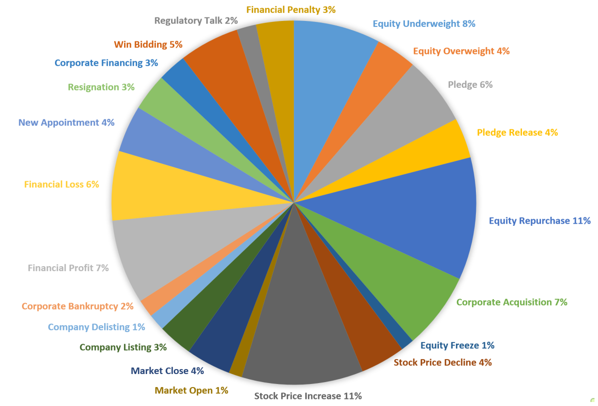

Figure 4: Distribution of All Event Types in CFinDEE

4.2 Comparison with Existing Datasets

As shown in Table 2, this study has conducted a comparative analysis of the CFinDEE dataset with current mainstream event extraction datasets. This comparison spans sentence-level datasets (such as ACE2005 [2] and DuEE [10]) to document-level datasets (such as MUC-4 [4], RAMS [3], DCFEE [19], ChFinAnn [23], FEED [7] and DuEE-Fin [5]). A multi-dimensional set of metrics was employed for evaluation, including dataset quality (Ann., trigger), scale (Instances, Events, Args), granularity (ETs, Roles), complexity (MER), and domain adaptability. The comprehensive analysis indicates that the CFinDEE dataset exhibits significant advantages in key performance metrics. Among these, the CFinDEE's MER is the highest among all datasets, highlighting the complexity of the events covered by the dataset. In datasets with manual annotations, CFinDEE leads significantly in the number of documents, events, and arguments, showcasing the dataset's extensive coverage and the detail of its content. Among datasets in the financial domain, CFinDEE has the most financial event types and argument roles, underscoring the dataset's capability to adapt to diverse financial scenarios. Additionally, triggers were annotated to support related sub-tasks, such as event detection.

5 EXPERIMENT

5.1 Experimental Setup

In this study, the dataset was divided into training, validation, and test sets in an 8:1:1 ratio. As shown in Table 3, we detailed the number of documents, events, arguments, and the MER in each set. Furthermore, all parameter settings used in the experiments followed the standards proposed in the original paper, with the training cycles of all models fixed at 100 epochs and all pre-trained language models adopting the base version to ensure consistency and comparability of results.

5.2 Evaluation Metrics

To ensure fairness and consistency in the evaluation, this study adopts the same evaluation criteria as Doc2EDAG, assessing the model by comparing the predicted event tables with the ground-truth records. Specifically, we select the ground-truth record for

Table 2: Statistics of EE Datasets(Ann.: Whether the dataset is manually annotated, Instances: The number of sentences in SEE and the number of documents in DEE, ETs: The number of event types, Events: The number of events, Roles: The number of event argument types, Args: The number of arguments, MER: Multi-event rate)

Dataset	Ann.	trigger	Type	Instances	ETs	Events	Roles	Args	MER	Domain
ACE2005	Y	Y	SEE	7914	33	3333	35	6198	9.2%	General
DuEE	Y	Y	SEE	16,956	65	19,640	121	41,520	14.1%	General
MUC-4	Y	N	DEE	1,700	4	1514	5	2641	18.9%	General
RAMS	Y	Y	DEE	9,124	139	8,823	65	21,237	0.0%	General
DCFEE	N	N	DEE	7,144	4	9,493	14	34284	24.7%	Finance
ChFinAnn	N	N	DEE	32,040	5	47,824	35	289,871	29.0%	Finance
FEED	N	N	DEE	31,748	5	46,960	35	338066	28.4%	Finance
DuEE-Fin	Y	Y	DEE	8,186	13	11,031	92	56,806	29.2%	Finance
CFinDEE	Y	Y	DEE	16,372	22	26,483	116	107,096	39.9%	Finance

Table 3: Data Partitioning

	Train	Dev	Test	Total
Documents	13057	1334	1981	16372
Events	21220	2111	3152	26483
Arguments	85864	8437	12795	107096
MER	40.0%	36.5%	41.6%	39.9%

each predicted event that matches the event type and has the highest argument-matching degree. We then count the number of all matched arguments to calculate the model's precision, recall, and F1 score. Considering that event types usually involve multiple roles, we compute micro-averaged role-level scores as the final metric for evaluating the performance of the DEE model.

5.3 Baseline

To comprehensively evaluate our dataset and demonstrate its potential in financial DEE tasks, we selected the following models as baselines:

- DCFEE [19] employs an argument-completion strategy and critical event detection techniques to generate document-level event records. This model includes two variants: DCFEE-O, which generates an event record from a single key event sentence, and DCFEE-M, which extracts multiple possible argument combinations within the closest distance to the vital event sentences.
- Doc2EDAG [23] is an end-to-end model that achieves document-level event extraction by transforming documents into entity-based directed acyclic graphs and filling in event tables directly using entity-based path extensions.
- Greedy-Dec [23], a variant of Doc2EDAG, adopts a greedy decoding strategy to fill only one event table entry.
- GIT [18] implements document-level event extraction by constructing a heterogeneous graph interaction network and introducing a tracker module. Using a heterogeneous graph, this model simulates interactions between sentences and entity mentions and continuously tracks the extracted event records with the tracker module to consider global dependencies.
- DE-PPN [21] acquires document-aware representations through a document-level encoder and then employs a multi-granularity non-autoregressive decoder to extract all events in parallel.
- PTPCG [25] is a lightweight DEE model that introduces an event argument combination strategy and combines it with a non-autoregressive decoding algorithm based on automatically selected pseudo triggers to build a pruned complete graph.
- IPGPF [6] eliminates the dependence on the generation order of argument roles by parallel generating event arguments and iteratively generating event records while adopting a pre-filling strategy to mitigate training deficiencies and zero precision issues in a parallel generation.
- ProCNet [17] uses event proxy nodes to establish connections between entities and context for efficiently capturing global information. It then optimizes global training by minimizing the Hausdorff distance, effectively capturing interactions between events.

5.4 Experimental Summary and Analysis

5.4.1 Overall Analysis. Table 4 presents the experimental results on the CFinDEE dataset across various baseline models, demonstrating satisfactory overall performance, which validates our dataset's high quality. Among all baseline models, ProCNet exhibits the best performance across all evaluation metrics, attributable to its use of event proxy nodes and a strategy to minimize the Hausdorff distance. These strategies collectively facilitate global learning of events. Following ProCNet, the GIT model also performs well on most evaluation metrics, thanks primarily to its precise modeling of global interactions and interdependencies. However, DCFEE-M shows the weakest performance, limited by its insensitivity to the context of event arguments, resulting in ineffective handling of scattered event arguments within the dataset. In contrast, DCFEE-O performs slightly better, suggesting that predicting multiple events from key event sentences alone is ineffective. Doc2EDAG shows nearly a 9% improvement over both DCFEE-O and Greedy-Dec because Greedy-Dec models only entity-level representations, whereas Doc2EDAG utilizes global information. DCFEE focuses on

Table 4: Overall Precision (P.), Recall (R.), and F1-Score (F1) of Baselines

Model	P.	R.	F1	F1(S.)	F1(M.)
DCFEE-O	63.0	57.4	60.0	67.8	55.3
DCFEE-M	48.8	59.8	53.7	60.2	50.1
Greedy-Dec	69.1	51.9	59.3	72.2	49.7
Doc2EDAG	76.7	62.9	69.1	74.5	65.6
DE-PPN	65.9	46.4	54.5	56.7	53.2
PTPCG	74.3	64.7	69.2	79.0	62.4
GIT	77.2	65.6	70.9	76.5	66.4
IPGPF	67.4	54.1	60.0	65.6	56.6
ProCNet	83.3	76.2	79.6	86.0	75.6

Table 5: F1 Score of Single-Event and Multi-Event on DuEE-Fin and CFinDEE Dataset

Model	DuEE-fin		CFinDEE	
	S.	M.	S.	M.
GIT	73.7	63.8	76.5	66.4
ProCNet	80.0	72.1	86.0	75.6

Table 6: F1 Score of Four Datasets

Model	ChFinAnn	DuEE-Fin	CFinDEE*	CFinDEE
PTPCG	79.4	66.0	62.6	69.2
ProCNet	83.0	75.6	73.5	79.6

extracting events within a single sentence, whereas Doc2EDAG can handle complex events across sentences within a document. The DE-PPN model encodes sentences and entities using Transformer, but compared to Doc2EDAG, which uses a path extension decoding strategy, its F1 score decreases by 5%. IPGPF and DE-PPN employ parallel generation strategies, yet the former's F1 score is significantly higher than the latter's by 5.6%, indicating that IPGPF's pre-filling strategy and iterative parallel generation method are superior to traditional parallel processing approaches.

Although the experimental results indicate that the model has achieved certain success on the CFinDEE dataset, the characteristics such as dispersed arguments and multi-event extraction within the dataset still pose significant challenges to the existing event extraction models. The F1 score of the current best model is only 79.6%, which suggests that DEE remains a challenging task, requiring future technological advancements to enhance performance further.

5.4.2 Single-Event vs. Multi-Event. To deeply evaluate the model's ability to process events of varying complexities, we divided the test set into a single-event subset (S.) and a multi-event subset (M.). Each document involves only one event in the single-event subset, while each contains multiple events in the multi-event subset. As shown in Table 4, all models exhibit a significant decrease in F1 scores in multi-event scenarios compared to single-event scenarios, confirming the increased difficulty in event extraction from multi-event documents. Notably, the ProCNet model performs exceptionally well in single-event and multi-event scenarios, demonstrating its effectiveness and robustness in handling both events. The PTPCG model performs best in single-event scenarios, highlighting the training and inference efficiency of its pruned complete graph structure. The GIT model performs best in multi-event scenarios, indicating its significant advantage in processing complex events. Table 5 compares the extraction performance of single and multiple events in the two datasets, DuEE-Fin and CFinDEE, under two optimal models. As illustrated, the score difference between single and multiple event extractions in CFinDEE surpasses that in DuEE-Fin due to the higher proportion of multi-events in CFinDEE by more than 10% compared to DuEE-Fin, further validating our dataset's increased challenge in multi-event extraction.

5.4.3 Impact of Data Scale and Granularity. To evaluate the impact of data scale and fine-grained classification on model performance, Table 6 displays the F1 scores for two advanced models on four financial datasets, among which CFinDEE* represents a subset selected from the CFinDEE dataset to match the data scale of the DuEE-fin dataset. The experimental results indicate that the ChFinAnn dataset performs the best among all datasets, which can be attributed to its most enormous data scale and the fewest number of event categories. This highlights the importance of data scale and fine-grained classification in event extraction tasks. When the data scale is the same, the score of CFinDEE* is lower than that of DuEE-Fin, which verifies that the fine-grained annotation in CFinDEE increases the difficulty, thereby reducing the model's effectiveness. When the number of event categories remains unchanged, the score of CFinDEE significantly surpasses that of CFinDEE*, further emphasizing the importance of data scale in enhancing model performance.

5.4.4 Per-Event-Type Results. We evaluated the extraction performance of 22 financial event types in the CFinDEE dataset using the optimal model ProCNet. As shown in Figure 5, the blue lines represent event categories shared with the DuEE-Fin dataset, while the red lines indicate newly added event types. The results show that the newly added event types generally exhibit better performance, suggesting that introducing more granular event classifications makes the relationships between events clearer, thereby effectively enhancing the model's understanding of complex event relationships and improving overall performance. However, ProCNet's performance is unsatisfactory in dealing with new appointments, corporate bankruptcy, and stock price decline. This performance gap may be attributed to the following factors: (1) The long-tail phenomenon: Some event types have relatively less data, especially for events like bankruptcies that are uncommon in real life, failing to provide the model with ample learning resources. (2) Cross-sentence arguments: Arguments are more dispersed due to the longer document length. ProCNet has limitations in capturing long-term dependencies between sentences and cannot directly model the connections between them. (3) Complexity of multiple events: The high frequency of multiple events occurring within documents increases the complexity of model processing. These challenges highlight the complexity of event extraction tasks in the financial domain and point out directions for future research and model optimization. In the future, we need more advanced models and strategies to overcome these difficulties and improve the model's overall performance in handling various financial events.

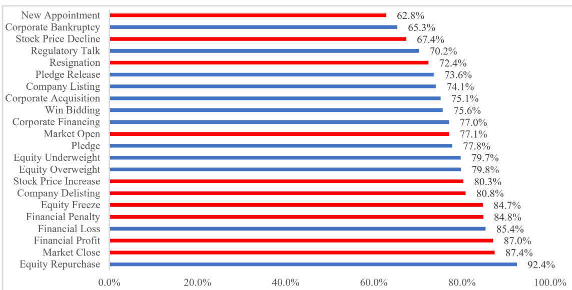

Figure 5: F1-Score of 22 Event Types on ProCNet

5.4.5 Limitations Analysis. The CFinDEE dataset constructed in this study, while showing strengths in data scale, granularity, and multi-event rate, still has certain limitations. Firstly, the imbalanced distribution of the dataset may lead to model overfitting on common events while neglecting rare events, thereby affecting the overall performance and generalization ability of the model. For event types with fewer samples, few-shot learning methods can be adopted, with current solutions including data augmentation [13], prompt learning [8], transfer learning [12], meta-learning [20], generative adversarial networks [16], and sample weighting [15], among others. Secondly, the current dataset lacks annotations for event relationships, limiting the in-depth understanding of the interactions between events in financial news. In future work, we plan to explore the complex relationships between events and enhance the dataset's completeness by adding annotations for event relationships.

6 CONCLUSION AND FUTURE WORK

In this paper, we construct CFinDEE, a high-quality Chinese DEE dataset based on real-world financial news, containing 26,483 events and 107,096 theses. This dataset enhances the utility and complexity with its larger data scale, richer event types, and higher multi-event rate, addressing the deficiencies of existing financial DEE datasets and providing a solid data foundation for research and development in financial document-level event extraction tasks. It is expected to promote the development of this field. In the future, we plan to explore the complex event relations within CFinDEE to offer deeper insights into financial market analysis and risk prediction, thereby fostering innovation and progress in the financial sector.

7 ACKNOWLEDGMENTS

The work presented in this paper is supported by the National Key Research and Development Program of China under Grant No.2020AAA0108501.

REFERENCES

[1] Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. NLP in FinTech applications: past, present and future. *arXiv preprint arXiv:2005.01320* (2020).

[2] George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation.. In *Lrec*, Vol. 2. Lisbon, 837–840.

[3] Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. 2020. Multi-Sentence Argument Linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 8057–8077. https://doi.org/10.18653/v1/2020.acl-main.718

[4] Ralph Grishman and Beth M Sundheim. 1996. Message understanding conference-6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics.*

[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 *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer, 172–183.

[6] Guanhua Huang, Runxin Xu, Ying Zeng, Jiaze Chen, Zhouwang Yang, and E Weinan. 2023. An Iteratively Parallel Generation Method with the Pre-Filling Strategy for Document-level Event Extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 10834–10852.

[7] Guozheng Li, Peng Wang, Jiafeng Xie, Ruilong Cui, and Zhenkai Deng. 2021. Feed: A chinese financial event extraction dataset constructed by distant supervision. In *Proceedings of the 10th International Joint Conference on Knowledge Graphs*. 45–53.

[8] Haochen Li, Tong Mo, Hongcheng Fan, Jingkun Wang, Jiaxi Wang, Fuhao Zhang, and Weiping Li. 2022. KiPT: knowledge-injected prompt tuning for event detection. In *Proceedings of the 29th International Conference on Computational Linguistics*. 1943–1952.

[9] Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, et al. 2022. A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems* (2022).

[10] Xinyu Li, Fayuan Li, Lu Pan, Yuguang Chen, Weihua Peng, Quan Wang, Yajuan Lyu, and Yong Zhu. 2020. DuEE: a large-scale dataset for Chinese event extraction in real-world scenarios. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II 9*. Springer, 534–545.

[11] Jian Liu, Yubo Chen, Kang Liu, Yantao Jia, and Zhicheng Sheng. 2020. How does context matter? On the robustness of event detection with context-selective mask generalization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. 2523–2532.

[12] Qing Lyu, Hongming Zhang, Elior Sulem, and Dan Roth. 2021. Zero-shot event extraction via transfer learning: Challenges and insights. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 322–332.

[13] Manfu Ma, Xiaoxue Li, Yong Li, Xinyu Zhao, Xia Wang, and Hai Jia. 2022. Small sample medical event extraction based on data augmentation. In *International Conference on Biomedical and Intelligent Systems (IC-BIS 2022)*, Vol. 12458. SPIE, 823–833.

[14] Junxiang Ren, Sibo Wang, Ruilin Song, Yuejiao Wu, Yizhou Gao, Borong An, Zhen Cheng, and Guoqiang Xu. 2022. IREE: A Fine-Grained Dataset for Chinese Event Extraction in Investment Research. In *China Conference on Knowledge Graph and Semantic Computing*. Springer, 205–210.

[15] Jun Shu, Xiang Yuan, Deyu Meng, and Zongben Xu. 2023. Cmw-net: Learning a class-aware sample weighting mapping for robust deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[16] Rui Wang, Deyu Zhou, and Yulan He. 2019. Open event extraction from online text using a generative adversarial network. *arXiv preprint arXiv:1908.09246* (2019).

[17] Xinyu Wang, Lin Gui, and Yulan He. 2023. Document-Level Multi-Event Extraction with Event Proxy Nodes and Hausdorff Distance Minimization. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 10118–10133. https://doi.org/10.18653/v1/2023.acl-long.563

[18] Runxin Xu, Tianyu Liu, Lei Li, and Baobao Chang. 2021. Document-level Event Extraction via Heterogeneous Graph-based Interaction Model with a Tracker. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 3533–3546.

[19] 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*. 50–55.

[20] Hang Yang, Yubo Chen, Kang Liu, and Jun Zhao. 2021. Meta Learning for Event Argument Extraction via Domain-Specific Information Enhanced. In *Knowledge Graph and Semantic Computing: Knowledge Graph and Cognitive Intelligence: 5th China Conference, CCKS 2020, Nanchang, China, November 12–15, 2020, Revised*

Selected Papers. Springer, 160–172.

[21] Hang Yang, Dianbo Sui, Yubo Chen, Kang Liu, Jun Zhao, and Taifeng Wang. 2021. Document-level event extraction via parallel prediction networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 6298–6308.

[22] Yunyan Zhang, Guangluan Xu, Yang Wang, Daoyu Lin, Feng Li, Chenglong Wu, Jingyuan Zhang, and Tinglei Huang. 2020. A question answering-based framework for one-step event argument extraction. *Ieee access* 8 (2020), 65420–65431.

[23] Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. 2019. Doc2EDAG: An End-to-End Document-level Framework for Chinese Financial Event Extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 337–346.

[24] Yang Zhou, Yubo Chen, Jun Zhao, Yin Wu, Jiexin Xu, and Jinlong Li. 2021. What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 14638–14646.

[25] Tong Zhu, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai, Nicholas Yuan, and Min Zhang. 2022. Efficient Document-level Event Extraction via Pseudo-Trigger-aware Pruned Complete Graph. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, Lud De Raedt (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4552–4558. https://doi.org/10.24963/ijcai.2022/632 Main Track.

APPENDIX

Table 7: Event Schema in CFinDEE

Major Event Categories	Event Type	Argument Roles
Equity Change	Equity Underweight	Stock Ticker, Number of Traded Shares, Price Per Share, Transaction Amount, Transaction Completion Time, Reducing Shareholder, Proportion of Holdings, Proportion of Total Share Capital.
	Equity Overweight	Stock Ticker, Number of Traded Shares, Price Per Share, Transaction Amount, Transaction Completion Time, Reducing Shareholder, Proportion of Holdings, Proportion of Total Share Capital, Buyer.
	Pledge	Pledger, Number of Pledged Shares, Percentage of Shares Pledged, Percentage of Pledged Shares in Total Capital, Company of Pledged Asset, Event Time, Pledged Asset, Pledgee.
	Pledge Release	Pledger, Number of Pledged Shares, Percentage of Shares Pledged, Percentage of Pledged Shares in Total Capital, Company of Pledged Asset, Event Time, Pledged Asset, Pledgee.
	Equity Repurchase	Repurchasing Party, Percentage of Total Company Capital, Number of Shares Repurchased, Price Per Share, Completion Time of Repurchase, Transaction Amount.
	Corporate Acquisition	Acquirer, Acquisition Completion Date, Acquisition Target, Transaction Amount, Acquiree.
	Equity Freeze	Equity Holder, Legal Institution, Froze Shares, Amount Frozen, Start Date, Unfreeze Date, Proportion of Holdings, Proportion of Total Share Capital
Company Change	Company Listing	Listed Company, Security Code, Event Time, Issuing Price, Fundraising Amount, Market Value.
	Company Delisting	Delisted Company, Security Code, Event Time.
	Corporate Bankruptcy	Bankrupt Company, Scale of Debt, Bankruptcy Time, Creditors.
Performance Change	Financial Profit	Company Name, Financial Reporting Cycle, Earnings Amount, Change in Earnings
	Financial Loss	Company Name, Financial Reporting Period, Change in Losses, Net Loss.
Stock Market Fluctuation	Stock Price Increase	Stock Ticker, Growth Rate, Stock Price, Event time
	Stock Price Decline	Stock Ticker, Decline Percentage, Stock Price, Event Time.
	Market Open	Stock Ticker, Closing Price, Event Time, Stock Price Change.
	Market Close	Stock Ticker, Opening Price, Event Time, Stock Price Change.
Executive Change	New Appointment	Executive's Name, Company Name after Change, Executive Position, Position after Change, Title of Position, Event Time
	Resignation	Executive's Name, Executive Position, Company of Employment, Event Time
Business Change	Corporate Financing	Investor, Investee, Financing Amount, Investment Round, Event Time, Lead Investor.
	Win Bidding	Winning Company, Bid Amount, Tendering Party, Bid Date, Subject of the Bid.
Regulatory Penalty	Regulatory Talk	Company Name, Time of Summons, Summoning Authority.
	Financial Penalty	Entity Fined, Punishing Authority, Fine Amount, Event Time.