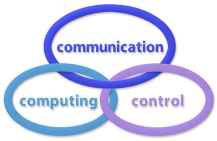
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 6, Month: December, Year: 2026
Article Number: 7226, https://doi.org/10.15837/ijccc.2026.6.7226



CCC Publications



CEPEX: Enhancing Corporate Event Prediction with Stock Price Dynamics and GRU-Attention Networks

X. Liu, Z. Peng

Xiping Liu*
School of Computing and Artificial Intelligence
Jiangxi University of Finance and Economics, China
*Corresponding author: liuxiping@jxufe.edu.cn

Zichun Peng
Anfu Secondary Vocational School
Yuyuan North Road, Anfu Industrial Park, Ji'an, Jiangxi, China
1806708740@qq.com

Abstract

Accurate corporate event prediction (CEP) presents a significant challenge for decision-making in financial markets. In this work, we reframe the CEP task as a problem of forecasting the optimal next step in a corporate life-cycle sequence, and introduce CEPEX, a novel predictive framework that enhances a Gated Recurrent Unit (GRU) attention-based model by uniquely incorporating quantified stock price dynamics alongside historical event sequences. Evaluated on the comprehensive Chinese Corporate Event Prediction (CCEP) dataset, CEPEX achieves a state-of-the-art accuracy of 84.28%. A key contribution of our work is the demonstration that stock market data provide notably stronger predictive power for future corporate actions than financial news. This framework improves decision-support systems in financial informatics and offers a new approach for predictive control in dynamic market environments.

Keywords: Corporate Event Prediction, GRU model, attention mechanism, stock price dynamics, financial informatics.

1 Introduction

Corporations operating within financial markets continuously undertake actions and reach milestones, broadly termed *corporate events* (e.g., dividend declarations, Initial Public Offerings (IPOs), mergers, share buybacks) [27]. When these events happen and in what order matters greatly to investors, analysts, and regulators. This information helps them anticipate a company's next strategic moves and manage market risks. Although markets are naturally unpredictable, many corporate events follow set business steps or regulatory timelines—like the stages of an IPO or regular dividend payments. Because of this built-in structure, such events can be predicted to some extent. This research addresses *Corporate Event Prediction* (CEP), which involves forecasting the next likely corporate event based on historical sequences.

A key difficulty in CEP is selecting the most informative external knowledge to enhance sequential modeling. Existing approaches have focused on modeling sequential dependencies between event types [8, 17] or incorporating external textual data like financial news titles [19]. However, these methods ignore a critical, timely clue: stock price dynamics. Often, a company's decisions and what the market expects can be seen in how its stock price changes before any official announcement is made. We suggest that using numerical data which reflects market sentiment and shifts in company value will lead to more accurate predictions compared to relying on news stories or the event sequence alone.

To address this gap, we propose *CEPEX* (Corporate Event Prediction driven by External Knowledge), a novel predictive framework. CEPEX enhances a Gated Recurrent Unit (GRU) attention-based model by uniquely incorporating quantified stock price dynamics alongside historical event sequences. By considering the CEP task as a problem of forecasting the optimal next step in a corporate lifecycle sequence, our work provides a powerful tool for decision-support systems in financial informatics. Moreover, the ability to accurately forecast future corporate actions is crucial for developing advanced predictive control strategies in dynamic market environments.

The fundamental challenge of CEP is illustrated in Figure 1, which shows that based on a sequence of observed historical events, the model must predict the most probable subsequent corporate action.

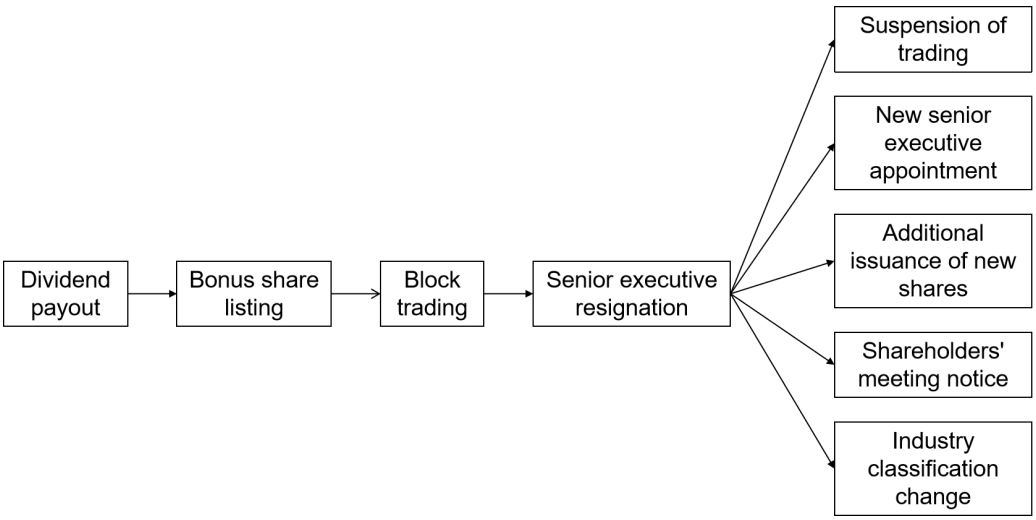


Figure 1: An example of corporate event prediction task

The contributions of this paper are summarized as follows:

1. We propose CEPEX, a novel predictive framework that integrates stock price dynamics with event sequences, enhancing corporate event prediction accuracy by leveraging this highly informative external market knowledge.

2. We demonstrate the utility and effectiveness of our approach, showing that stock market data provide a notable increase in predictive power for future corporate actions compared to models relying solely on historical event sequences.

3. We develop the CCEP dataset, which includes over 1 million events from 2,724 Chinese listed companies paired with processed stock price states, providing a specialized benchmark for CEP research.

4. Through extensive experiments, we validate that CEPEX achieves a state-of-the-art accuracy of 84.28% on the CCEP dataset, offering a promising direction for financial predictive modeling.

This paper is structured as follows. Section 2 presents the related work. Section 3 introduces the proposed CEPEX model. Section 4 details the construction of the CCEP dataset. Section 5 presents the experimental results. Finally, the paper is summarized in Section 6.

Related Work

The CEPEX framework is related to sequence modeling, advanced computational intelligence, and financial informatics. To position this work accurately, this section reviews related methodologies, focusing on the evolution of event prediction tasks, the specialized challenges of corporate financial forecasting, financial time-series prediction, and data-driven predictive control.

1.1 Sequential Event Prediction (SEP) Methodologies

The foundational task underlying CEP is Sequential Event Prediction (SEP), which aims to forecast the next event in a sequence based on general knowledge of typical progressions [23]. SEP methodologies have evolved from early statistical approaches based on event pair co-occurrence [5] to modern deep learning models. Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory networks (LSTMs) and Gated Recurrent Units (GRUs), became standard for modeling event chains, demonstrating an ability to capture non-linear, temporal dependencies. More recently, Graph Neural Networks (GNNs), such as NEEG [8] and Directed Graph Convolutional Networks (DGCN) [16], have proven highly effective for capturing intricate temporal and causal relationships [8, 16]. These models underscore the necessity of capturing both sequential order and relational structure. CEPEX builds upon these advancements by adopting the GRU architecture, extending it to incorporate domain-specific external knowledge.

1.2 Corporate Event Prediction (CEP) and the Role of External Knowledge

The Corporate Event Prediction (CEP) task involves forecasting corporate actions (e.g., mergers, regulatory filings) and presents unique challenges due to event sparsity and the specialized domain knowledge required. Prior dedicated research on CEP has largely focused on unstructured, textual external knowledge. For example, methods using Gated Graph Neural Networks (GGNN) have incorporated financial news titles to enhance event representations [6, 19]. However, textual data suffers from sentiment lag and noise. The core gap addressed by CEPEX is the lack of integration of market signals. We hypothesize that structured quantitative data reflecting market sentiment—specifically, stock price dynamics—offers a demonstrably stronger predictive signal. This is supported by our empirical findings, where the CEPEX framework (84.28% accuracy) notably outperforms the GGNN+News baseline (80.96%), validating the focus on market movements as a primary external knowledge source.

1.3 Financial Time-Series Prediction

Predicting events in financial markets requires techniques capable of handling chaotic and nonlinear time-series data [9, 15]. Deep learning architectures, particularly LSTMs and GRUs, are dominant in this field. The CEPEX methodology relies on preprocessing continuous stock price data into eight discrete market states. This **data granulation** is a validated strategy for managing time-series instability. The conceptual method of **trend division** is supported by research for financial time series data mining [20]. Further work demonstrated that transforming time series into granular time series improved prediction accuracy and scalability [10], directly validating the CEPEX design choice. The subsequent use of the GRU network, coupled with an attention mechanism, aligns with state-of-the-art deep learning practices. Our results show that preserving the distinct representation of financial market dynamics and event semantics through vector concatenation (84.28% accuracy) is superior to complex nonlinear fusion (83.40% accuracy).

1.4 Data-Driven Predictive Control

CEPEX provides a robust foundation for **Advanced Decision Support Systems (DSS)** in financial informatics [18]. Research suggests that AI-driven predictive analytics can facilitate the efficiency of financial services and proactive risk management [13]. The prediction output ($\hat{y} = e_{n+1}$) generated by CEPEX serves as a primary input for informing subsequent control actions within a

dynamic financial environment. This positioning aligns with the principles of **Data-enabled Predictive Control (DeePC)**, which utilizes data to learn control policies without requiring an explicit parametric system model [4]. While DeePC typically focuses on mapping input-output data directly to control actions, CEPEX fulfills the critical prerequisite of system behavior forecasting. By processing historical event chains and processed stock dynamics (input-output data), the GRU-Attention network implicitly models latent market transitions. Consequently, this predictive signal serves as a reliable front-end mechanism necessary for developing sophisticated data-driven control strategies, such as Model Predictive Control (MPC), in specialized financial contexts.

2 Methodology

The task of Corporate Event Prediction (CEP) can be formalized as follows: Given a historical corporate event chain $E = \{e_1, e_2, \cdots, e_n\}$, where e_i represents the i-th historical corporate event and n represents the total number of historical events, the CEP task is to predict the next most likely corporate event e_{n+1}. In this task and subsequent content, the sequence of occurred corporate events is called the *historical event chain*, each event in the chain is called a *historical event*, and potential corporate events being considered for the next step (typically from a predefined set) are called *candidate events*.

2.1 Model Overview

This paper introduces CEPEX, a novel model for CEP that leverages external knowledge. The overall architecture of CEPEX is illustrated in Figure 2.

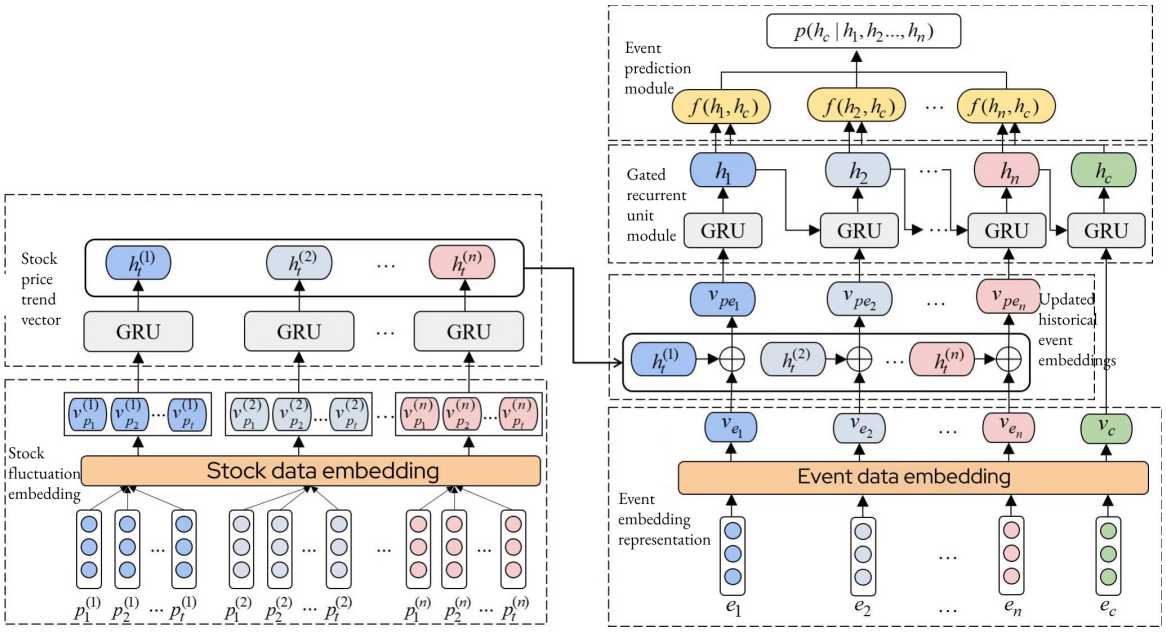


Figure 2: The framework of the CEPEX model. The architecture features dual input streams: (1) **Event Data** (e_i) and (2) **Stock Fluctuation Data** ($p_t^{(i)}$). The left side depicts the Stock Trend Embedding pipeline: discrete stock states are first embedded (Skip-gram), and then processed by a dedicated Gated Recurrent Unit (GRU) to extract a temporal Stock Price Trend Vector ($h_t^{(i)}$). The event embedding (v_{e_i}) is subsequently concatenated ($\oplus$) with the corresponding trend vector ($h_t^{(i)}$) to create an Enhanced Historical Event Embedding (v_{pe_i}). The right side shows the main sequential modeling: the sequence of enhanced embeddings ($v_{pe_1}, \ldots, v_{pe_n}$) is fed into a second GRU network to capture contextual dependencies ($h_1, \ldots, h_n$). Finally, the Event Prediction Module uses an attention mechanism to weigh the contextual states against the encoded candidate event (h_c), selecting the candidate with the highest score as the prediction $p(e_c|h_1, \ldots, h_n)$.

The CEPEX model operates through a series of steps. Initially, we employ the Skip-gram model to generate vector embeddings for both stock fluctuation sequences and events, effectively mapping them into a shared, low-dimensional semantic space. Concurrently, a Gated Recurrent Unit (GRU) network is utilized to extract temporal trend features from the stock fluctuation sequences. Subsequently, we integrate historical event embeddings with their corresponding stock trend features. This combined representation is then fed into another GRU network to model the temporal dependencies between historical events and candidate events, incorporating the stock trend information. Finally, an attention mechanism is applied to weigh the importance of each historical event when calculating similarity scores with candidate events. The candidate event exhibiting the highest similarity score is then selected as the predicted event. Note that we use Skip-gram and GRU instead of other modern networks like Transformer mainly for their efficiency and suitability for learning from discrete, categorical stock state sequences.

The CEPEX model is structured into three primary modules:

(1) **Embedding Module**: This module is responsible for generating various vector representations.

(2) **Gated Recurrent Unit (GRU) Module**: This module focuses on capturing temporal dependencies.

(3) **Event Prediction Module**: This module performs the final prediction.

These modules will be discussed in detail in the following sections.

2.2 The Embedding Module

The embedding module is responsible for generating embeddings for stock prices and events, and combining these embeddings together.

2.2.1 Stock Trend Embedding

This section outlines the methodology for embedding stock price trends into numerical representations. Each historical event in our framework is paired with a stock fluctuation sequences of length t. A typical event chain containing n consecutive events thus comprises n independent fluctuation sequences, each of length t, which serve as input to the embedding module illustrated in Figure 2 (on the left).

The implementation follows a three-stage process:

(1) **Corpus Construction**: We first establish a domain-specific corpus using historical stock fluctuations data. Unlike generic text corpora, this financial corpus is constructed by concatenating normalized fluctuation sequences from multiple firms. This approach preserves the unique characteristics of stock market movements while enabling effective feature learning.

(2) **Initial Embedding Generation**: The Skip-gram architecture is employed to convert discrete stock values into continuous vector representations. For the i-th historical event, this yields a sequence of embeddings $v_{p_1}^{(i)}, v_{p_2}^{(i)}, \cdots, v_{p_t}^{(i)}$ corresponding to the stock values $p_1^{(i)}, p_2^{(i)}, \cdots, p_t^{(i)}$.

(3) **Temporal Pattern Modeling**: A Gated Recurrent Unit (GRU) network processes these embeddings to capture temporal dependencies. The GRU uses a specialized structure with two control gates to adaptively manage the flow of information over time. This allows the model to selectively remember significant long-term stock market trends and filter out non-predictive, short-term volatility when predicting corporate events.

2.2.2 Event Embedding

This subsection outlines the methodology for generating numerical representations of historical and candidate events. Given the specialized nature of our financial event dataset, we constructed a corpus using 1,040,103 event chains (details will be given in Section 3). These chains were consolidated into a single document to train a Skip-gram model, which learns vector representations by predicting contextual events.

Through this training process, we obtained initial d-dimensional embeddings for each historical event: $v_{e_1}, v_{e_2}, \cdots, v_{e_n}$, where $v_{e_i} \in R^d$. Candidate events, being selected from the same event types as

the historical corpus, were similarly embedded as $v_{c_1}, v_{c_2}, \cdots, v_{c_k}$. For events not present in the original dataset (i.e., newly emerging event types), we adopted a zero-vector encoding strategy. This approach ensures dimensional consistency while providing a neutral baseline representation for unseen events during model inference. The complete event embedding pipeline maintains compatibility between historical patterns and candidate predictions through shared semantic spaces.

2.2.3 Enhanced Historical Event Embedding

To incorporate market dynamics into event representations, we introduce two methodologies for updating historical event embeddings using their initial vector v_{e_i} and corresponding stock price trend vector $h_t^{(i)}$. The updated embedding v_{pe_i} is computed as $v_{pe_i} = f(h_t^{(i)}, v_{e_i})$ through the following techniques:

(1) **Nonlinear Transformation Method**: This approach first performs linear combination of event and price vectors, followed by a hyperbolic tangent activation to introduce nonlinearity:

$$v_{pe_i} = \tanh(W_p h_t^{(i)} + W_e v_{e_i} + b)$$

The learned weight matrices W_p and W_e control contributions from price trends and event semantics respectively, while b represents the bias term.

(2) **Vector Concatenation Method**: Alternatively, we vertically concatenate the price trend vector with the event vector:

$$v_{pe_i} = h_t^{(i)} \oplus v_{e_i}$$

This operation preserves original feature dimensions by creating a combined representation $v_{pe_i} \in R^{2d}$, where $\oplus$ denotes vector concatenation. The doubled dimensionality maintains separate information channels for market movements and event characteristics.

Both methods enhance event representations by integrating contextual market data, with the nonlinear approach emphasizing feature interaction and the concatenation method preserving complete information from both modalities.

2.3 Gated Recurrent Unit (GRU) Module

This section introduces a GRU-based architecture designed to model temporal dependencies between financial events. The module's dual-component framework is visualized in Figure 3, with technical implementation details provided below.

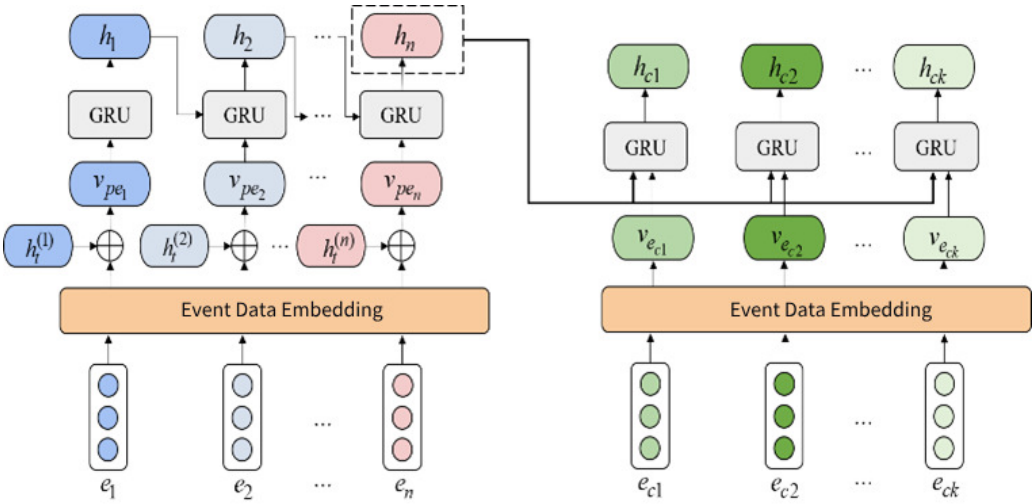


Figure 3: GRU Module Architecture. The left part is the historical event processing pipeline. The right part is the candidate event encoding mechanism.

The architecture comprises two specialized components:

(1) **Contextual Historical Event Modeling**. As illustrated in Figure 3, we first generate price-aware event representations $v_{pe_1}, v_{pe_2}, \cdots, v_{pe_n}$ using the methodology described in Subsection 2.2.3. These enhanced vectors are sequentially fed into a GRU network to capture temporal patterns:

$$h_i = \text{GRU}(v_{pe_i}, h_{i-1})$$

where $h_i \in R^d$ represents the contextual embedding at time step i. The initial hidden state h_0 and GRU parameters are initialized randomly to start learning from scratch.

(2) **Candidate Event Encoding Mechanism**. For candidate events, shown in Figure 3, we first obtain their initial embeddings $v_{c_1}, v_{c_2}, \cdots, v_{c_k}$ following Subsection 2.2.2. Each candidate vector is then combined with the final historical context h_n and processed through the GRU:

$$h_{c_j} = \text{GRU}(v_{c_j}, h_n)$$

This produces candidate-specific embeddings $h_{c_j} \in R^d$ that incorporate both event semantics and historical market context.

This dual-stream architecture enables the model to: (1) learn evolving patterns from historical event sequences, and (2) project candidate events into the same latent space while preserving contextual dependencies.

2.4 Event Prediction Module

This module implements a dual-stage scoring system to predict the most likely subsequent event. Building upon the GRU-derived embeddings from Subsection 3.2, we first compute baseline scores using four correlation metrics, then enhance these scores with an attention mechanism to capture event saliency.

2.4.1 Baseline Event Scoring

Given historical event embeddings $h_1, h_2, \ldots, h_n$ and candidate embeddings $h_{c1}, h_{c2}, \ldots, h_{ck}$, we calculate initial scores for each candidate using:

$$\text{score}_{c_j} = \frac{1}{n} \sum_{i=1}^{n} s(h_i, h_{c_j}) \tag{1}$$

where $s(\cdot, \cdot)$ quantifies the relationship between historical and candidate events. Four correlation functions are evaluated:

- *Cosine Similarity*: Measures vector orientation alignment

$$\cos(h_i, h_{c_j}) = \frac{h_i \cdot h_{c_j}}{\|h_i\| \|h_{c_j}\|}$$

- *Normalized Euclidean Distance*: Captures magnitude-insensitive dissimilarity

$$\text{euclidean}(h_i, h_{c_j}) = -\left\| \frac{h_i}{\|h_i\|} - \frac{h_{c_j}}{\|h_{c_j}\|} \right\|$$

- *Dot-Product Similarity*: Reflects raw feature correlation

$$\text{dot}(h_i, h_{c_j}) = h_i \cdot h_{c_j}$$

- *Neural Similarity*: Learns non-linear relationships via sigmoid activation

$$\text{linear_sigmoid}(h_i, h_{c_j}) = \sigma(W_e h_i + W_c h_{c_j} + b)$$

2.4.2 Attention-Weighted Scoring

To address the limitation of equal weighting in Equation 1, we introduce an attention mechanism that dynamically assigns importance to historical events based on their relevance to each candidate:

- *Attention Weight Calculation*:

$$u_{ij} = \tanh(W_e h_i + W_c h_{c_j} + b)$$

$$\alpha_{ij} = \frac{\exp(u_{ij})}{\sum_{i=1}^{n} \exp(u_{ij})}$$

- *Weighted Score Aggregation*:

$$\text{atten_s}_{c_j} = \sum_{i=1}^{n} \alpha_{ij} s(h_i, h_{c_j})$$

2.4.3 Event Selection

The final prediction is obtained by selecting the candidate with the highest attention-weighted score:

$$\hat{y} = \underset{j \in \{1,\ldots,k\}}{arg\,max} \text{atten_s}_{c_j}$$

This architecture combines: (1) multi-metric event correlation assessment, and (2) context-aware attention weighting to improve prediction accuracy by emphasizing salient historical events during candidate evaluation. The neural similarity component introduces learnable parameters W_e, W_c, b to capture complex relationships beyond manual feature engineering.

3 The CCEP Dataset

This section describes the construction and characteristics of the Chinese Corporate Event Prediction (CCEP) dataset, detailing the preprocessing of stock price data, event chain generation, and alignment with market states to support the CEPEX model evaluation.

3.1 Stock Data Processing

Stock prices exhibit substantial variability across companies and temporal periods. During pre-listing or delisting phases, price data may be entirely unavailable. Such irregularities complicate pattern recognition in financial time series. To address these challenges, we first preprocess the stock price data by categorizing each trading day into one of eight distinct states, reflecting both price fluctuations and trading phase characteristics. The classification criteria are defined as follows:

- *IPP (In Pre-listing Period)*: Designates stocks with no available price data prior to initial public offering (IPO), though event records may exist during this phase.
- *IDP (In Delisting Period)*: Applied to stocks undergoing delisting procedures, where official price quotations cease to exist.
- *ABI/ABD (Abnormal Increase/Decrease)*: Defined by exchange regulations governing excessive price fluctuations. For non-ST stocks, abnormal status is triggered when cumulative price deviation over three consecutive days reaches ±20%. ST stocks follow exchange-specific thresholds:

$$\text{Fluctuation Deviation} = \text{Stock Return} - \text{Index Return} \tag{2}$$

$$\text{Stock Return} = \frac{\text{Current Close}}{\text{Previous Close}} - 1 \tag{3}$$

$$\text{Index Return} = \frac{\text{Current Index}}{\text{Previous Index}} - 1 \tag{4}$$

Days exceeding upper/lower thresholds are classified as ABI/ABD respectively.

- *STA (Stable)*: Assigned to trading days with price movements within $[-1\%, 1\%]$, indicating minimal volatility.
- *NMI/NMD (Normal Increase/Decrease)*: Used for days with moderate price changes (exceeding stability thresholds but below abnormality criteria).
- *SUS (Suspended)*: Indicates trading halts due to corporate actions (e.g., asset restructuring) or regulatory investigations, during which price data is unavailable.

3.2 CEP Sample Construction

We construct our dataset through systematic event chain generation and negative sampling. The process consists of two phases:

Phase 1: Event Chain Formation For each company, we chronologically sort its events and create sequential chains. Each chain contains the first 8 consecutive events from a starting point, with the subsequent 9th event serving as the prediction target. This sliding window approach generates multiple $\langle context, target \rangle$ pairs per company.

Phase 2: Negative Sampling Following [5], we create challenging prediction tasks by pairing each true event chain with four distractors. These negative candidates were randomly selected from the company's event pool after excluding events appearing in either the context window or the true subsequent event. This generates balanced prediction tasks where models must distinguish real event progressions from plausible alternatives.

3.3 Event-Stock State Alignment

This subsection details the methodology for aligning corporate events with corresponding stock market states. We collected daily trading data for 2,738 Chinese listed companies from December 2016 to December 2021 through Tushare's Python API [28]. The dataset contains essential trading indicators including stock codes, transaction dates, open/close prices, price changes, and trading volumes.

Our alignment process comprises two phases:

Phase 1: Stock Status Classification Following the criteria established in Section 3.1, we categorized stock market performance into eight distinct states: In Pre-IPO Period (IPP), In Delisting Process (IDP), Abnormal Increase (ABI), Abnormal Decrease (ABD), Stable (STA), Normal Increase (NMI), Normal Decrease (NMD), and Suspended (SUS).

Phase 2: Temporal Sequence Construction For each historical event chain $E = \{e_1, e_2, \cdots, e_n\}$, we created corresponding stock status sequences through three operational steps:

1. *Date Synchronization*: Align each event's publication date with its corresponding trading day in the stock database.
2. *Status Mapping*: Convert raw stock metrics to pre-defined status labels using the classification rules in Section 3.1.
3. *Window Selection*: Extract t consecutive trading days' statuses backward from each event date.

This procedure generates a stock status sequence $p_1^i, p_2^i, \cdots, p_t^i$ for every historical event e_i, where superscript i denotes association with the i-th event in chain E. The parameter t controls the historical context length considered for stock market influences, with experimental validation detailed in Section 4.

3.4 Dataset Partitioning

We constructed the Chinese Corporate Event Prediction (CCEP) dataset through rigorous temporal partitioning. The dataset contains 1.35 million corporate events from 3,543 listed companies, chronologically divided into training (2017 – 2020), validation (2021 Q1 & Q2), and test (2021 Q3 & Q4) sets to maintain temporal validity.

The dataset originates from regulatory filings of Chinese A-share listed companies, containing 360 distinct event types. We implemented three key filtering steps to ensure data quality:

1. **Company Consistency Filtering**: Excluded firms without continuous event records across all three subsets to ensure temporal continuity.
2. **Market Presence Validation**: Removed companies listed after January 2021 or delisted before 2017 to maintain consistent market participation.
3. **Event Density Standardization**: Adopting criteria from prior studies [8, 11], we enforced minimum event thresholds per company (90 events in training, 9 events in validation/test sets) to ensure sufficient data density.

These procedures yielded the final dataset distribution shown in Table 1, retaining 2,738 firms with stable event histories.

Table 1: Distribution of processed financial market event datasets

	Training Set	**Validation Set**	**Test Set**
Time Span	2017 – 2020	2021 Q1&Q2	2021 Q3&Q4
Number of Events	1,052,514	128,637	84,728
Number of Companies	2,724	2,724	2,724

The curated dataset maintains a 12.4:1.5:1 event ratio across training/validation/test subsets. Each event record includes three core elements: stock identifier, event timestamp, and categorical event type. A total of 360 distinct event types were identified, with the top 10 most frequent categories shown in Table 2, representing 56.6% of total occurrences. Trading suspensions dominate (21.14%), followed by restructuring-related halts (7.16%) and mandatory disclosure suspensions (6.50%).

Table 2: Top 10 event types by frequency

Event Type	**Count**	**Share(%)**
Trading suspension	267843	21.16
Asset restructuring suspension	90653	7.16
Major disclosure suspension	82158	6.49
Abnormal trading activity	44083	3.48
Shareholder meeting notice	42737	3.38
Shareholder meeting convening	43255	3.42
Shareholder registration start	41034	3.24
Block trade execution	40257	3.18
Shareholder record date	37575	2.97
Shareholder registration end	37530	2.96

4 Experiments

This section presents the experimental evaluation of the CEPEX model, comparing its performance against baseline methods on the CCEP dataset. We analyze key parameters, conduct ablation studies, and report results to validate the effectiveness of integrating stock price dynamics into corporate event prediction.

4.1 Baselines and Metrics

To evaluate the performance of CEPEX comprehensively, we compared it against six baseline approaches, which represent foundational and traditional baselines for sequence and graph-based embedding methods in event prediction. We also employed an enhanced multi-source variant.

- **PPMI**: Computes positive point-wise mutual information between historical and candidate events to measure their correlation strength.
- **DeepWalk**: Generates event representations through random walk-based graph embeddings, using Euclidean distance for candidate ranking.
- **Word2Vec**: Trains Skip-gram embeddings on our event chain corpus, evaluating candidates via cosine similarity.
- **GRU**: Models temporal dependencies in event sequences using gated recurrent units.
- **GGNN**: Constructs event graphs from co-occurrence patterns and applies graph neural networks with gated message passing.
- **GGNN+News**: Extends GGNN by incorporating financial news text features during event representation learning.
- **CEPEX+News**: Our ablation variant integrating both stock prices and news data with event sequences.

Following standard evaluation protocols for sequential event prediction [8], we adopt a multiple-choice cloze-style evaluation. For each historical event chain $\{e_1, ..., e_n\}$, the model receives k candidate events $\{e_{c1}, ..., e_{ck}\}$ containing one ground truth successor and $k-1$ randomly sampled distractors. Prediction accuracy is calculated as:

$$\text{Accuracy} = \frac{\#\text{ Correctly Predicted Chains}}{\#\text{ Total Chains}} \tag{5}$$

This metric directly reflects the model's ability to identify genuine temporal progressions from spurious associations, aligning with real-world decision-making scenarios where investors must distinguish plausible future events from irrelevant possibilities.

4.2 Experimental Setup

We employ margin loss as the objective function to optimize model parameters. The loss function is defined as:

$$L(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{k} \max\left(0, m - s_{iy} + s_{ij}\right) + \frac{\lambda}{2}\|\theta\|^2 \tag{6}$$

where n denotes the number of historical events, k the candidate event count, m a tunable margin threshold, s_{iy} the similarity score between the i-th historical event and the ground truth event y, and s_{ij} the score between the i-th historical event and j-th candidate event. The L2 regularization term with coefficient λ helps prevent overfitting.

The model was configured with key hyperparameters including a batch size of 1,000 for efficient mini-batch training, a learning rate of 0.0001 optimized via RMSprop, and uniform vector dimensions of 128 for both event embeddings and GRU hidden states to ensure architectural consistency. We implemented L2 regularization with $\lambda = 1 \times 10^{-8}$ to mitigate overfitting while maintaining model capacity, and set the margin value m to 0.08 in the loss function to enforce discriminative event scoring. Training stability was further enhanced through Xavier initialization and early stopping based on validation loss plateaus.

To enhance generalization, we implemented early stopping based on validation loss plateauing for 5 consecutive epochs. All model components were initialized using Xavier initialization to ensure stable gradient flow.

Table 3: Model performance comparison (Accuracy %)

Model	Accuracy
PPMI	59.78
DeepWalk	72.33
Word2Vec	73.18
GGNN	80.63
GGNN+news	80.96
GRU	81.78
CEPEX+news	82.72
CEPEX (Ours)	**84.28**

4.3 Experimental Results

4.3.1 Main Results

Table 3 presents the prediction accuracy of different models, revealing several key observations.

First, the PPMI baseline achieves the lowest accuracy (59.78%) due to its reliance on simple co-occurrence statistics. This approach fails to capture semantic relationships between events and cannot handle unseen event types in the test set. Second, neural models (DeepWalk, Word2Vec, GGNN, GRU) significantly outperform PPMI (72.33%-81.78% vs 59.78%), demonstrating the importance of learning event interactions. The GRU model's superior performance (81.78%) over graph-based GGNN (80.63%) suggests temporal dependencies are more critical than spatial relationships for this task. Third, our CEPEX model achieves state-of-the-art accuracy (84.28%) by incorporating stock trends. Notably:

- CEPEX outperforms GGNN+news (80.96%) by 3.32%, indicating stock data provides more discriminative signals than news.
- The CEPEX+news variant (82.72%) underperforms CEPEX, suggesting news integration may introduce noise that dilutes stock trend signals.

The results confirm that stock price trends provide complementary information to event sequences, enabling more accurate predictions than pure event-based or news-enhanced approaches. The performance gap between CEPEX and CEPEX+news (1.56%) highlights the importance of focused integration of market dynamics over multi-source fusion.

4.3.2 Discussion of Main Findings and Error Analysis

The central finding—that stock price dynamics provide a significantly stronger predictive signal for corporate events than news text—is rooted in the nature of financial information flow.

Superiority of Stock Trends over News: Stock market movements are often forward-looking indicators of corporate actions. Unlike news articles, which are unstructured, prone to sentiment lag, and subject to editorial bias, stock prices serve as a direct, quantified reflection of collective market anticipation. Corporate decisions and market expectations frequently manifest as sustained price trends (captured by our discrete states) *before* official announcements or public news breaks. The 3.32% relative improvement of CEPEX over the GGNN+news baseline demonstrates that this market-driven signal is more timely and less noisy for forecasting the subsequent corporate event. The underperformance of the CEPEX+news variant (82.72%) compared to pure CEPEX (84.28%) further supports this; incorporating news text likely introduces irrelevant or lagging information that dilutes the highly discriminative signal from the structured market trends.

Analysis of Prediction Errors: While CEPEX achieves high accuracy, mispredictions generally occur in two distinct scenarios:

1. **Sudden or Regulatory Mandates**: The model excels at predicting process-driven events (e.g., dividend announcement following a specific historical pattern). However, errors often arise with sudden, exogenous events, such as unexpected regulatory investigations or unforeseen trading halts mandated by the exchange, which may not be preceded by a clear, anticipatory stock trend.

2. **Ambiguous Market Signals**: Mispredictions can occur when the stock price trend reflects general sector movement or broad macroeconomic volatility rather than an imminent company-specific action. For instance, a period of Normal Increase (NMI) might correctly anticipate a *Shareholder Meeting Notice* (a routine process event) in a predictable case, but fail when the NMI is merely due to an external market surge, leading the model to incorrectly predict a process event that is not due. Conversely, genuine prediction targets are missed when the pre-event stock fluctuation is **Stable (STA)**, providing no discernable signal to the GRU, and the subsequent event lacks a strong sequential link to the context chain.

These patterns highlight that although CEPEX effectively harnesses market dynamics, event prediction remains challenging for outcomes driven by sudden, non-linear shifts or external regulatory pressures that are not implicitly discounted in the stock price trend sequence.

4.3.3 Analysis of Parameters

This subsection mainly conducts comparisons from three aspects: the combination method of stock price fluctuations and historical events, the time window considered for stock price fluctuations, and the similarity measure, to study the impact of different parameter settings on the prediction results.

Combination methods. This study evaluates two methods of combining stock price volatility and historical events: nonlinear combination and vector concatenation. The results are shown in Table 4. We can see that, the concatenation method performs better than the non-linear combination.

Table 4 presents the results of a systematic evaluation comparing two distinct feature fusion paradigms in event prediction tasks. The nonlinear interaction model, which employs cross-modal attention mechanisms to capture complex relationships, achieves 83.40% accuracy. In contrast, the tensor space concatenation approach demonstrates superior performance with 84.28% accuracy through direct integration of orthogonal feature dimensions. The observed 0.88% performance differential validates the hypothesis that preserving dimensional orthogonality in feature representation significantly enhances model generalization capacity.

Table 4: Comparative Analysis of Feature Fusion Strategies

Fusion Methodology	**Accuracy (%)**
Nonlinear Interaction Model	83.40
Tensor Space Concatenation	84.28

Time window of stock price fluctuations. This study investigates the relationship between temporal granularity and model performance through a controlled experiment. Employing time windows of 7, 10, 15, and 21 consecutive trading days, we quantify how varying historical context lengths affect event prediction accuracy. As demonstrated in Table 5, the proposed framework achieves consistent predictive performance across all configurations, with minimal variation observed between optimal (84.28% at 7-day window) and longest-tested (84.04% at 21-day window) durations. Notably, the 7-day window configuration yields the highest observed accuracy, demonstrating its effectiveness in capturing critical market dynamics without overfitting to short-term noise.

Similarity measure. Table 6 presents the empirical evaluation of four similarity quantification strategies in event prediction tasks. The Dot Product method emerges as the optimal choice with 84.28% accuracy, outperforming the Linear-Sigmoid Fusion approach by 1% absolute gain. This performance differential aligns with theoretical expectations: dot product computation inherently captures stronger directional alignment between vectors, while linear-sigmoid fusion introduces potential saturation effects in high-dimensional spaces. Notably, all metrics maintain >83% accuracy, indicating their fundamental capacity for pattern recognition in financial time series analysis.

Table 5: The impact of time window duration on prediction accuracy

Time Window (Days)	Accuracy (%)
7	84.28
10	84.16
15	83.88
21	84.04

Table 6: The impact of similarity metrics on prediction accuracy

Similarity Metric	Accuracy (%)
Linear+Sigmoid Similarity	83.27
Euclidean Distance Similarity	83.33
Cosine Similarity	83.47
Dot Product Similarity	**84.28**

4.3.4 Ablation Studies

Table 7 quantifies the contribution of core components through systematic ablation analysis. The full model achieves baseline accuracy of 84.28%, while component removal induces differential performance degradation:

- Exclusion of price fluctuation features (-price) reduces accuracy by 1.5%, demonstrating that raw price data constitutes a critical information source for capturing market microstructure dynamics
- Removal of attention mechanism (-attention) causes significant performance deterioration of 2.25%, validating its essential role in establishing dynamic event importance weighting across historical sequences

Table 7: Impact of Component Removal on Prediction Accuracy

Model Configuration	Accuracy (%)
CEPEX	84.28
-price	82.78
-attention	82.03

5 Conclusion

This paper introduced CEPEX, a novel and highly accurate framework for Corporate Event Prediction (CEP) that strategically integrates market dynamics into sequential modeling. By addressing the critical research gap of underexplored stock price signals, CEPEX enhances predictive capacity in dynamic financial environments.

5.1 Summary of Contributions and Implications

Our primary contributions are threefold:

1. We introduced a novel integration of granular stock price dynamics with event sequences using a GRU-Attention architecture, establishing the CEPEX framework for the CEP task.
2. We developed and temporally partitioned the CCEP dataset, a specialized benchmark resource comprising over 1.35 million events across 3,543 firms within the Chinese A-share market. This

dataset is designed to drive future research in market-aware CEP under specific regulatory price-limit regimes.

3. We conducted extensive experiments to validate the effectiveness of the proposed method. Experimental results showed that CEPEX achieved a highly competitive accuracy of 84.28%, demonstrating that quantified stock price dynamics can provide a more discriminative predictive signal than conventional news text.

The high-precision predictions generated by CEPEX provide a robust foundation for Advanced Decision Support Systems (DSS) in financial informatics [18]. By forecasting the immediate future corporate state, CEPEX acts as the predictive horizon required to inform control actions within complex financial systems. This aligns the model with the principles of Data-enabled Predictive Control (DeePC) [4], where the prediction serves as the data-driven "front-end" that identifies system transitions. Thus, CEPEX serves as a valuable predictive module for developing sophisticated data-driven control strategies in financial markets.

5.2 Limitations and Future Work

While the CEPEX framework demonstrates promising results, it is subject to certain limitations that define important directions for future research:

1. A primary limitation concerns the generalizability of the model across different regulatory environments. The current validation is confined to the Chinese A-share market, where the quantified price states (ABI/ABD) are specifically calibrated to local price-limit regulations (e.g., the $\pm 10\%$ and $\pm 20\%$ thresholds). Since international markets, such as the S&P 500 or NASDAQ, operate under different volatility management regimes and disclosure timelines, the cross-market stability of these feature definitions remains to be fully established.

2. While we demonstrated the notable effectiveness of stock price dynamics compared to news headlines, future research should investigate integrating a broader range of multi-modal external sources. Incorporating comprehensive analyst reports or full-text regulatory filings may further enhance the model's ability to account for sudden, unpredictable events that are not yet reflected in price movement.

Funding

This research was partially supported by the National Natural Science Foundation of China (62462034, 62272205, and 62272206), the Natural Science Foundation of Jiangxi Province (20232ACB202008 and 20242BAB25119).

Author contributions

Xiping Liu devised the project and the main conceptual ideas, and took the lead in writing the manuscript. Zichun Peng worked out almost all of the technical details, designed and performed the experiments.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Bai, L.; Guan, S.; Guo, J.; Li, Z.; Jin, X.; Cheng, X. (2021). Integrating deep event-level and script-level information for script event prediction, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online and Punta Cana, Dominican Republic, pages 9869–9878, 2021. DOI: 10.48550/arXiv.2110.15706

[2] Benzin, J.-V.; Rinderle-Ma, S. (2025). A survey on event prediction methods from a systems perspective: Bringing together disparate research areas, *ACM Computing Surveys*, 57(12), July 2025. DOI: 10.1145/3743672

[3] Chambers, N.; Jurafsky, D. (2008). Unsupervised learning of narrative event chains, *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Columbus, Ohio, USA, pages 789–797, 2008.

[4] Coulson, J.; Lygeros, J.; Dörfler, F. (2019). Data-enabled predictive control: In the shallows of the DeePC, *Proceedings of the 2019 18th European Control Conference (ECC)*, Naples, Italy, pages 307–312, 2019. DOI: 10.23919/ECC.2019.8795639

[5] Granroth-Wilding, M.; Clark, S. (2016). What happens next? event prediction using a compositional neural network model, *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, Phoenix, Arizona, USA, pages 2727–2733, 2016. DOI: 10.1609/aaai.v30i1.10344

[6] Harl, M.; Weinzierl, S.; Stierle, M.; Matzner, M. (2020). Explainable predictive business process monitoring using gated graph neural networks, *Journal of Decision Systems*, 29(sup1):312–327, 2020. DOI: 10.1080/12460125.2020.1780780

[7] Li, Y.; Zemel, R.; Brockschmidt, M.; Tarlow, D. (2016). Gated graph sequence neural networks, *Proceedings of the 4th International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, April 2016. DOI: 10.48550/arXiv.1511.05493

[8] Li, Z.; Ding, X.; Liu, T. (2018). Constructing narrative event evolutionary graph for script event prediction, *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, Sweden, pages 4201–4207, 2018. DOI: 10.5555/3304222.3304354

[9] Li, W.; Zhang, L.; Zhang, X. (2025). Predicting Multi-Indicator Stock Time Series using Convolutional Neural Networks based on Feature Engineering, *International Journal of Computers Communications & Control*, 20(5):1–12, 2025. DOI: 10.15837/ijccc.2025.5.6774

[10] Lu, W.; Yang, J.; Liu, X. (2014). Numerical Prediction of Time Series Based on FCMs with Information Granules, *International Journal of Computers Communications & Control*, 9(3):210, 2014. DOI:10.15837/IJCCC.2014.3.210

[11] Lv, S.; Qian, W.; Huang, L.; Han, J.; Hu, S. (2019). Sam-net: Integrating event-level and chain-level attentions to predict what happens next, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 33(1):6802–6809, 2019. DOI: 10.1609/aaai.v33i01.33016802

[12] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. (2013). Efficient estimation of word representations in vector space, *Proceedings of the International Conference on Learning Representations (ICLR)*, Scottsdale, AZ, pages 1–12, 2013. DOI: 10.48550/arXiv.1301.3781

[13] Odumuwagun, O. (2023). Time Series-Based Quantitative Risk Models: Enhancing Accuracy in Forecasting and Risk Assessment (2023). International Journal of Computer Applications Technology and Research, 12(11): 29–41, 2023. DOI: 10.7753/IJCATR1211.1006

[14] Perozzi, B.; Al-Rfou, R.; Skiena, S. (2014). Deepwalk: Online learning of social representations, *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, New York, NY, USA, pages 701–710, 2014. DOI: 10.1145/2623330.2623732

[15] Srivastava, A.; Assaf, R.; Pandey, D.K.; Kumar, R. (2024). Stock price crash risk research: current trends and future directions, *Journal of Risk Finance*, 25(5):723–750, 2024. DOI: 10.1108/JRF-12-2023-0331

[16] Wang, L.; Yue, J.; Guo, S.; Sheng, J.; Mao, Q.; Chen, Z.; Zhong, S.; Li, C. (2021). Multi-level connection enhanced representation learning for script event prediction, *Proceedings of the Web Conference (WWW)*, New York, NY, USA, pages 3524–3533, 2021. DOI: 10.1145/3442381.3449894

[17] Wang, Z.; Zhang, Y.; Chang, C.-Y. (2017). Integrating order information and event relation for script event prediction, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark, pages 57–67, 2017. DOI: 10.18653/v1/D17-1006

[18] Wang, Q. (2018). An Overview of Functional Components of Artificial Intelligence Financial Decision Support System, *Open Journal of Social Sciences*, 6(8):115–126, 2018. DOI: 10.4236/jss.2018.68009.

[19] Yang, Y.; Wei, Z.; Chen, Q.; Wu, L. (2019). Using external knowledge for financial event prediction based on graph neural networks, *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, New York, NY, USA, pages 2161–2164, 2019. DOI: 10.1145/3357384.335815

[20] Yang, H.; Gao, X.; Han, L.; Cui, W. (2022). A financial time series data mining method with different time granularity based on trend Division, *International Journal of Computers Communications & Control*, 17(6): 4952, 2022. DOI:10.15837/ijccc.2022.6.4952

[21] Mustafa, H.R.; Imran, R.; Mahmood, A. (2025). Text classification using graph convolutional networks: A comprehensive survey, *ACM Computing Surveys*, March 2025. DOI: 10.1145/3714456

[22] Zhan, C.; Xiang, W.; Chao, L.; Wang, B. (2024). What would happen next? predicting consequences from an event causality graph, *Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Miami, Florida, USA, pages 820–832, 2024. DOI: 10.18653/v1/2024.findings-emnlp.45

[23] Zhao, L. (2021). Event prediction in the big data era: A systematic survey, *ACM Computing Surveys*, 54(5):1–37, 2021. DOI: 10.1145/3450287

[24] Zhou, B.; Chen, Y.; Liu, K.; Zhao, J. (2022). Script event prediction via multilingual event graph networks, *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(2):1–17, 2022. DOI: 10.1145/3557893

[25] Zhou, P.; Wu, B.; Wang, C.; He, L. (2025). An improved hierarchical neural network model with local and global feature matching for script event prediction, *Expert Systems with Applications*, 259: 125325, 2025. DOI: 10.1016/j.eswa.2024.125325

[26] Zhu, F.; Gao, J.; Yu, C.; Wang, W.; Xu, C.; Mu, X.; Yang, M.; Xu, R. (2023). A generative approach for script event prediction via contrastive fine-tuning, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 37(11), 14056-14064, 2023. DOI: 10.1609/aaai.v37i11.26645

[27] [Online]. Available: https://www.socialtables.com/blog/event-planning/types-of-corporate-events/, Accessed on 20 December 2024.

[28] [Online]. Available: https://tushare.pro/, Accessed on 20 October 2024.



Copyright ©2026 by the authors. Licensee Agora University, Oradea, Romania.
This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.
Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

Cite this paper as:



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Liu, X.; Peng, Z. (2025). CEPEX: Enhancing Corporate Event Prediction with Stock Price Dynamics and Attention Mechanisms, *International Journal of Computers Communications & Control*, 21(6), 7226, 2026.
https://doi.org/10.15837/ijccc.2026.6.7226