



Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/infus

Full length article

Multi-source aggregated classification for stock price movement prediction

Yu Ma^a, Rui Mao^b, Qika Lin^c, Peng Wu^d, Erik Cambria^{b,*}^a School of Economics and Management, Nanjing University of Science & Technology, 200, Xiaolingwei Road, Nanjing, Jiangsu, 210094, China^b School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798, Singapore^c School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China^d School of Intelligent Manufacturing, Nanjing University of Science & Technology, 200, Xiaolingwei Road, Nanjing, Jiangsu, 210094, China

ARTICLE INFO

Keywords:

Stock prediction
Event-driven investing
Multi-source aggregating
Sentiment analysis

ABSTRACT

Predicting stock price movements is a challenging task. Previous studies mostly used numerical features and news sentiments of target stocks to predict stock price movements. However, their semantics-based sentiment analysis is sub-optimal to represent real market sentiments. Moreover, only considering the information of target companies is insufficient because the stock prices of target companies can be affected by their related companies. Thus, we propose a novel Multi-source Aggregated Classification (MAC) method for stock price movement prediction. MAC incorporates the numerical features and market-driven news sentiments of target stocks, as well as the news sentiments of their related stocks. To better represent real market sentiments from the news, we pre-train an embedding feature generator by fitting the news to real stock price movements. Embeddings given by the pre-trained sentiment classifier can represent the sentiment information in vector space. Moreover, MAC introduces a graph convolutional network to capture the news effects of related companies on the target stock. Finally, MAC can predict stock price movements for the next trading day based on the aforementioned features. Extensive experiments prove that MAC outperforms state-of-the-art baselines in stock price movement prediction, Sharpe Ratio, and backtesting trading incomes.

1. Introduction

Recent developments in artificial intelligence and machine learning have caused significant impacts on the financial industry, specifically on quantitative stock prediction [1–3]. The development of an effective method for stock price movement prediction is critical for investors to grasp investment opportunities and also for security companies to reduce the risk caused by margin trading. The stock market is highly complex because stock prices are affected by many factors [4, 5]. Therefore, aggregating and using adequate information from different sources to improve the prediction performance of stock price movements is a challenging task in both academia and industry.

Extensive research on stock price movement prediction has been performed by using quantitative indicators, including historical transaction data and technical indicators [2,6]. These indicators have been proven effective for predicting stock price movements. Event-driven trading is another widely used investment strategy because stock price movements are influenced with the release, dissemination and absorption of information [7–10]. In this sense, financial news is also a critical information source for investors, which affects their sentiments and investment behaviors, directly impacting stock prices [11,12].

By using Natural Language Processing (NLP) techniques, current studies combined financial news sentiment analysis with quantitative indicators to predict stock price movements [13–15]. The predefined sentiment dictionaries have been widely applied to analyze news sentiments for stock price prediction with their lexical sentiment scores used as sentiment features [14–18]. However, the performance of these dictionary-based methods highly depends on the knowledge quality and coverage of the dictionaries [19–22]. The lexical sentiment knowledge cannot be updated in time according to different market environments, e.g., bull and bear markets. Recently, deep learning-based models have been applied for sentiment analysis of financial news to improve stock price prediction [23–25]. These approaches used labeled data to learn sentiment polarities, i.e., positive, neutral, and negative, to fit stock price movements. However, these methods have two shortcomings: (1) Annotating a huge domain-specific sentiment analysis dataset is expensive and time-consuming; (2) The sentiment polarities based on semantic understanding and human annotation cannot reflect the real market sentiments [26] because these annotators are not investors who really have effects on market prices in real time.

* Corresponding author.

E-mail addresses: mayu@njust.edu.cn (Y. Ma), mao.r@ruimao.tech (R. Mao), qika@sentic.net (Q. Lin), wupeng@njust.edu.cn (P. Wu), cambria@ntu.edu.sg (E. Cambria).

<https://doi.org/10.1016/j.inffus.2022.10.025>

Received 19 April 2022; Received in revised form 7 October 2022; Accepted 26 October 2022

Available online 8 November 2022

1566-2535/© 2022 Elsevier B.V. All rights reserved.

In addition, previous studies in event-driven stock price prediction mainly focused on a one-to-one problem, assuming the stock prices of a company are only affected by the news about the company itself, rarely considering the effects that the news about related companies could have [14–16]. However, with the rapid development of the market economy, listed companies in stock markets are interconnected by various relationships, such as the supply chain [27], the shareholding [28] and the industry competition relationships [29]. The crisis spillover effect indicates that the negative effects of a crisis in a company do not only affect the company itself but also spread to other related companies [30]. Therefore, it is essential to incorporate news about related companies into the stock price movement prediction of a target company.

In this paper, we propose a novel Multi-source Aggregated Classification (MAC) model for stock price movement prediction. Conceptually, the decision-making framework in stock market investment involves two main tasks, namely portfolio management and stock trading [31]. Portfolio management focuses on managing a batch of selective stocks in a long-term investment, while stock trading studies short-term trading strategies for individual stocks. Our research focuses on stock trading. The proposed MAC model can assist investors in achieving better decision-making by predicting the price movement of a specific stock on the next trading day. If the model predicts an ‘up’ signal for the next trading day, an investor may buy or hold the stock today. On the contrary, the investor may not buy or hold the stock if the model predicts a ‘down’ signal.

The MAC model integrates three feature sources, i.e., transaction data and technical indicators of target stocks, news about the target stocks, and news about their related stocks. We employ 31 robust quantitative indicators to represent the morphology of stock price movements. To represent news information, we pre-train a Chinese RoBERTa-based [32,33] embedding generator by fitting the news of a company to its real stock price movements on the next trading day, i.e., up and down. Thus, the news embeddings can reflect the real market sentiment, i.e., up is positive and down is negative. Then, we use the Graph Convolutional Networks (GCN) [34] to model the connections between a target company and its related companies. Lastly, the multi-source features are concatenated and fed into the Bidirectional Long Short-Term Memory (BiLSTM) networks [35] to predict stock price movements on the next trading day.

We demonstrate the validity of our MAC method with six target stocks corresponding to six different industries in the Chinese stock market. The results indicate that, in the evaluation of the stock price movement classification task, our MAC outperforms the state-of-the-art baseline, i.e., eLSTM [17] by 2.38% in accuracy (ACC) and 4.62% in Matthews Correlation Coefficient (MCC) on average. In terms of the financial evaluation, MAC exceeds eLSTM by 0.68 in Sharpe Ratio [36] on average, achieving large margins in the backtesting trading incomes, based on the recommended trading strategies. We also conduct systematic ablation studies to demonstrate the effectiveness of each feature source and of different pre-training methods, i.e., market-driven sentiment analysis and semantics-based sentiment analysis.

The contributions of this work can be summarized as follows:

- We propose a novel model for stock price movement prediction that incorporates features from different sources. To the best of our knowledge, this is the first method that incorporates news about related companies in this task.
- We propose an effective pre-training task for market-driven sentiment analysis to learn the representations of news in vector space.
- Our experimental results demonstrate the superiority of our model compared with state-of-the-art baselines. We also prove that the market-driven sentiment analysis pre-training is more effective than the semantics-based sentiment analysis pre-training.

The remainder of this paper is organized as follows: Section 2 reviews the related work; Section 3 explains the proposed MAC method; Section 4 shows our experimental setups; Section 5 presents the results, the ablation studies, the hyperparameter analysis and the breakdown analysis; Finally, Section 6 contains the conclusions of the present work and the futures perspectives.

2. Related work

2.1. Information and stock price movements

The stock price movements are the results of the release, dissemination and absorption of multi-source information [7,37]. Material news events are essential sources of stock return volatility [9]. Jeon et al. [9] found that stock return volatility is significantly related to news quantity and content. Shi and Ho [38] indicated that negative news events would lead to abnormal volatility of stock prices in the short term. Strauss et al. [39] argued that excessive media attention and negative news would have negative effects on stock returns. In conclusion, news events are important driving forces of stock price movements.

The development of the market economy has caused the interconnection of listed companies through different relationships, such as the supply chain [27], the shareholding chain [28] and the industry competition [29]. In fact, the stock price movements of a company are influenced not only by the news about the company itself but also by the news about its related companies. Relevant studies on marketing and crisis management indicate that there is a spillover effect between related companies meaning that information about Company A will affect public cognition and judgment of Company B [40]. The magnitude of the spillover depends on the strength of their correlation [40]. Jacobs and Singhal [41] showed that negative public opinions of a listed company have a vertical spillover effect in the supply chain, which can lead to a decline in the stock price of its suppliers and customers. Moreover, when a parent brand suffers an adverse event, the sub-brands become vulnerable [42], and vice versa [40]. Moreover, several studies found that a brand crisis of a listed company may also spill over to its competitors [29,41,43], subsequently impacting the whole industry. Therefore, financial news about companies in the relationships of the supply chain, shareholding and competition is an important feature for predicting the stock prices of a target company.

The related studies show that it is critical to consider news-driven stock price prediction, where the news should be about both target companies and their related companies.

2.2. Stock price movement prediction

The prediction of stock price movements is a popular research area in academia and industry [1,2]. Quantitative indicators, including daily prices and technical indicators, such as the Moving Average (MA), the Williams Indicator (WR) and the Relative Strength Index (RSI), have been widely used for stock price prediction [1,6,28] as they can represent the morphology and trading behaviors of stocks in time series.

With the development of NLP, many studies combined financial news with quantitative indicators to improve the performance of stock prediction based on the behavioral finance theory [13,44]. Several studies applied sentiment lexicons to extract financial news features for stock price prediction [14–17,45]. Chen et al. [15] input the quantitative indicators and the news features extracted by a sentiment dictionary into their RNN-boost model improving the stock prediction performance in the Chinese stock market. Chen et al. [16] also improved the prediction performance of stock prices by incorporating daily price and news sentiment features into the LSTM [46] model. Li et al. [14] combined the quantitative indicators and news sentiments extracted from sentiment dictionaries into the LSTM-based model for stock price prediction in the Hong Kong market.

Table 1
Descriptions of symbols.

Notation	Description
CP_t	The closing price of a target stock on day t .
T	The number of trading days before the stock price movement prediction day.
K	The top- K related stocks to a target stock.
$\mathbf{q}_{tar,t}$	$\mathbf{q}_{tar,t} \in \mathbb{R}^{1 \times 31}$, the quantitative indicators of a target stock on day t .
$Q_{tar,t}$	$Q_{tar,t} \in \mathbb{R}^{T \times 31}$, the quantitative indicators of a target stock within a time window T .
\mathbf{e}	The market-driven sentiment embedding for a news title.
\mathbf{n}	A daily news feature.
$MSC(\cdot)$	The news embedding generator based on the pre-trained market-driven sentiment classifier.
$\mathbf{n}_{tar,t}$	$\mathbf{n}_{tar,t} \in \mathbb{R}^{1 \times 768}$, the news features of a target stock on day t .
N_{tar}	$N_{tar} \in \mathbb{R}^{T \times 768}$, the news features of a target stock within a time window T .
$G = (V, E)$	G is a stock relation graph. V is the set of nodes (stocks). E is the set of edges (relations).
$N_{rel,t}$	$N_{rel,t} \in \mathbb{R}^{K \times 768}$, the news features of K related stocks on day t .
N_{rel}	$N_{rel} \in \mathbb{R}^{(T \times K) \times 768}$, the news features of K related stocks within a time window T .
A	$A \in \mathbb{R}^{K \times K}$, an adjacency matrix.
M	$M \in \mathbb{R}^{(T \times K) \times (T \times K)}$, an adjacency matrix with T time steps.
H	$H \in \mathbb{R}^{(T \times K) \times 200}$, the hidden states of GCN, representing the features of related stocks.
R_{tar}	$R_{tar} \in \mathbb{R}^{T \times 200}$, the news features of related stocks to a target stock within a time window T .

They reported that the prediction models using news sentiments extracted from finance domain-specific sentiment dictionaries could outperform the models that only use either technical indicators or news sentiments given by general sentiment dictionaries. Besides, machine learning-based methods have been used to extract financial news features for stock price prediction [1,47]. Researchers commonly used manually labeled positive and negative sentences to train a sentiment classifier for financial news [23–25]. For example, De Oliveira Carosia et al. [23] trained an ANN-based classifier by financial news in Brazilian Portuguese to extract news features. Their news had been manually labeled by experts according to the semantic sentiment. Sousa et al. [24] manually annotated 582 stock news articles as positive, neutral or negative to fine-tune a BERT [48] model for sentiment analysis of financial news.

However, the mentioned publications presented some shortcomings. On one hand, the dictionary-based methods commonly obtained sentiment features by using either the positive and negative word frequencies or the lexical sentiment scores. However, these methods highly depend on the quality and coverage of dictionaries, and they do not take into account the dependency information in the overall context [21]. Besides, the development of dictionaries relies on the great efforts of annotators, which can hardly be updated in time according to different stock market environments, e.g., bull and bear markets. On the other hand, the machine learning-based methods used human annotated sentiment analysis datasets to conduct supervised learning to obtain sentiment features. However, the datasets were annotated by the semantic interpretation of texts. The annotators are not real investors whose investments can impact the stock price movements in real time. Thus, there is a gap between the semantics-based sentiment features and the real market sentiment features. Lastly, the dictionary-based and machine learning-based methods did not take the news about related companies into account, although it has been reported that the related company news indeed impacts the stock prices of a target company according to the above empirical and theoretical studies.

3. Methodology

In this work, we propose a novel model that incorporates information from different sources, i.e., quantitative indicators of target stocks, news about target stocks and news about their related stocks. Following previous studies [14,17], the task of our study is to predict stock price movement directions, i.e., up or down, of a target stock on the next trading day. The label y_{t+1} is 1, if $CP_{t+1} - CP_t \geq 0$. Otherwise, the label y_{t+1} is 0. CP_{t+1} is the closing price of a target stock on the trading day $t + 1$, and CP_t is the closing price on day t . We do not introduce an extra label for $CP_{t+1} - CP_t = 0$ because it is uncommon that the closing prices are exactly the same on two consecutive trading days.

As shown in Fig. 1, the proposed MAC model consists of four technical parts. Firstly (Section 3.1), we extract the quantitative indicators¹ of the target stocks and the subseries of the quantitative indicators within a period of time T . The target stocks are the stocks for which we intend to predict the stock price movements on the next trading day. T denotes the number of trading days before the stock price movement prediction day. Q_{tar} represents the quantitative indicators of a target stock (tar) in T trading days. Secondly (Section 3.2), the news features of target stocks in T trading days (N_{tar}) are given by an embedding generator, which is a pre-trained market-driven sentiment classifier based on Chinese-RoBERTa [33]. We use this classifier to obtain the embedding feature of each news title. The news embeddings of a target stock in a day are averaged. Subsequently, we obtain the news representations of the target stock in T trading days. Thirdly (Section 3.3), we identify the important related stocks of the target stocks according to the cosine similarity of stock nodes. The news features of the important related stocks are also generated from the pre-trained classifier and averaged over a day, yielding the news representations of the important related stocks (N_{rel}) in T trading days (N_{rel}). We use GCN to capture the relationships of stocks and fuse the news representations (N_{rel}) of the relevant stocks, yielding the related stock news representations of the target stocks (R_{tar}). Lastly (Section 3.4), the multi-source information (Q_{tar} , N_{tar} and R_{tar}) is concatenated and fed into a BiLSTM-based classifier to predict the stock price movements on the next trading day (y_{t+1}). The technical components are detailed in the following subsections. Table 1 summarizes the symbols used in this section.

3.1. Extraction of quantitative indicators for a target stock

Quantitative indicators are widely used as numerical features for stock price prediction. Quantitative indicators are composed of historical transaction data and technical indicators. Following previous studies [14,17,45], 31 popular quantitative indicators are selected in this study. The details of our employed quantitative indicators are shown in Table 2. The technical indicators are given by the equations presented in Appendix. The historical quantitative data of target stocks were collected using the RESSET platform.²

¹ Quantitative indicators are the collection of transaction data and technical indicators over several trading days. Technical indicators refer to mathematical operation outputs given historical transaction data, e.g., price, trading volume or time span. Different mathematical operations can yield different technical indicators to represent the morphology and investment behaviors of stock prices.

² RESSET (<http://www.resset.com/>) is a popular provider for financial information services in China.

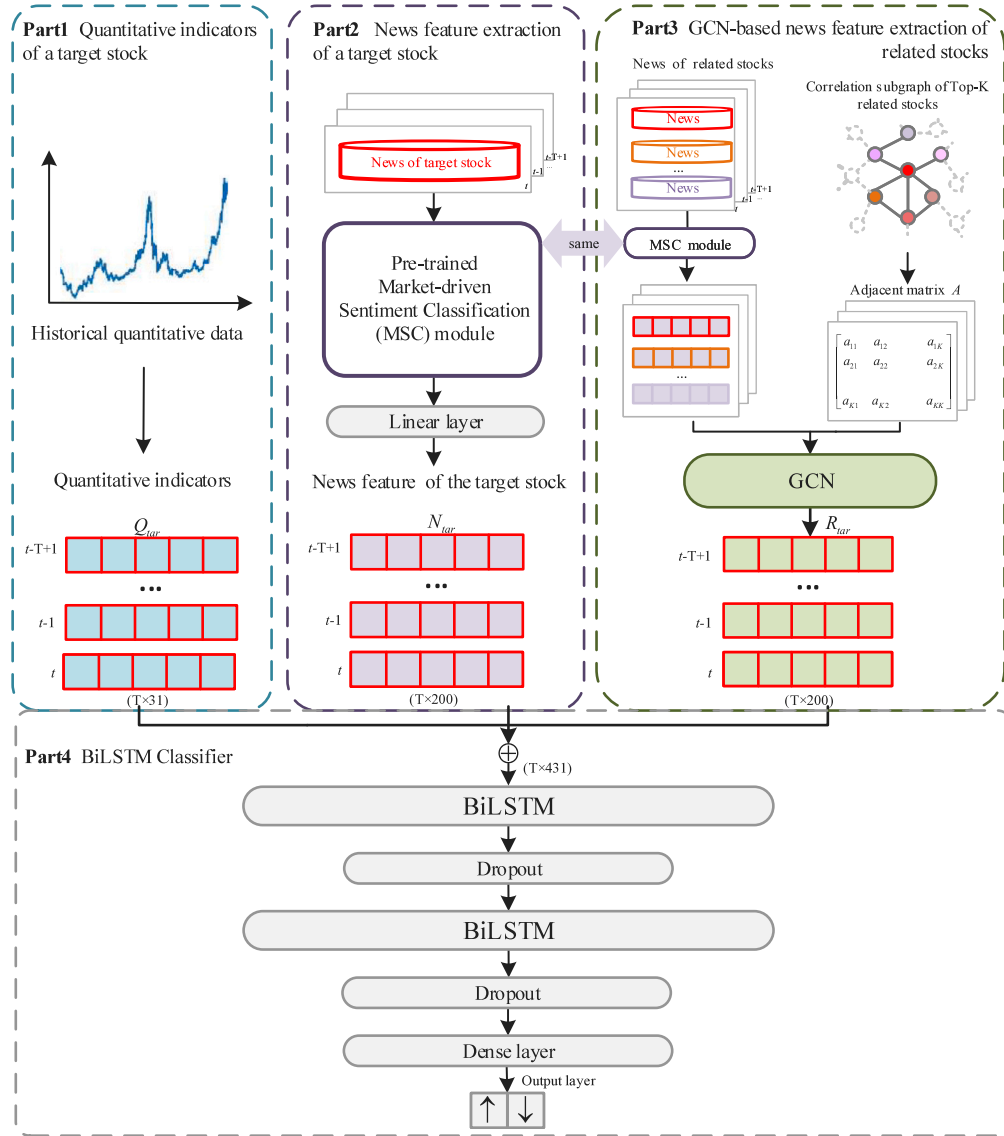


Fig. 1. Overview of the MAC framework.

Due to different value ranges, each quantitative indicator is normalized using Min–Max normalization [45]. Then, the quantitative indicators of a target stock (tar) in the trading day t are denoted as

$$\mathbf{q}_{tar,t} = [OP_{tar,t}, CP_{tar,t}, \dots, LowerBand_{tar,t}], \quad (1)$$

where $\mathbf{q}_{tar,t} \in \mathbb{R}^{1 \times 31}$ is a 31-dimensional vector.

We gather the quantitative indicators of a target stock over n trading days, where n denotes the time span of our training and testing data. For each training step that learns the stock price movement (y_{t+1}) at day $t + 1$, we use the quantitative indicators from T ($T \in n$) trading days before day $t + 1$ to capture the time series information.

We do not consider all the historical data from day 1 because the ancient data could bring noise to the prediction of stock price movements on day $t + 1$. We examine T with different values in Section 5.4. Then, the quantitative indicators within the window T are given by

$$Q_{tar} = [\mathbf{q}_{tar,t-T+1}, \dots, \mathbf{q}_{tar,t-1}, \mathbf{q}_{tar,t}], \quad (2)$$

where $Q_{tar} \in \mathbb{R}^{T \times 31}$ is the quantitative indicator feature matrix of a target stock.

3.2. Extraction of the news features for a target stock

In this section, we first introduce the pre-trained Market-driven Sentiment Classifier (MSC) that is used for generating news embeddings (Section 3.2.1). Then, we present the process of preparing the news features for a target stock based on MSC (Section 3.2.2).

3.2.1. Pre-trained market-driven sentiment classifier

In this section, we aim to pre-train a sentiment classifier driven by market sentiments to generate effective embedding features for news. The motivation is that there is a gap between semantic sentiment and market sentiment because the semantics of financial news cannot directly reflect investment behaviors. For example, the decrease in debt ratio and investment may have different effects on a company's

Table 2
The employed quantitative indicators of target stocks.

Feature name	Feature name
Opening/Closing Price (OP/CP)	Moving Average (MA10/MA20/MA30)
Highest/Lowest Price (HP/LP)	Weight Moving Average (WMA)
Volume of Trade (Vol)	Exponential Moving Average(EMA)
Turnover Rate (TurR)	Relative Strength Index (RSI6/RSI12/RSI24)
Daily Amplitude (DA)	Money Flow Index (MFI)
Daily Return (DR)	Stochastic Indicator KDJ (Slowk/Slowd)
Price-to-Earnings Ratio (P/E)	True Range (TR)
Price-to-Book Ratio (P/B)	Average True Range (ATR)
Price-to-Sales Ratio (P/S)	Williams Indicator (WR)
Moving Average Convergence and Divergence (DIFF/DEA/MACD)	Bollinger Bands (UpperBand/MiddleBand/LowerBand)

stock,³ although both statements semantically mean that the value of a financial indicator diminishes.

Thus, we hypothesize that the stock price movements can represent the real sentiment of investors towards a piece of news. We use news titles as input because a title can summarize the main idea of a piece of news. It is in line with related works which argue that news titles can extract the information more precisely than the news content in stock prediction [49,50]. The pre-training target is to predict the stock price movements of a company on the next trading day, given a news title of the company on the current day. We first collect 24,000 Chinese finance news about 32 stocks from 02/01/2018 to 12/10/2020.

The data were collected from Hundsun Electronics.⁴ According to the taxonomy of Hundsun Electronics, the collected news can be categorized into three classes, namely company information, business information and financial information.⁵ To avoid information leakage, the pre-training dataset does not contain either the target companies or their related companies that are examined in our main task. Then, according to the actual stock price movements on the next day of the news release, the news titles are labeled as positive (up and unchanged) or negative (down). We use Chinese-RoBERTa [33] as the encoder because it has been proved that RoBERTa is an effective pre-trained language model in many NLP tasks [51–56]. A news title is padded with CLS and SEP⁶ on both sides. Then, the padded sequence is fed into the RoBERTa encoder ($enc(\cdot)$), yielding the hidden states (\mathbf{S}) by

$$\mathbf{S} = enc(CLS, w_1, w_2, \dots, w_l, SEP), \quad (3)$$

where w denotes a token in a news title. l denotes the length of the input news title. $\mathbf{S} \in \mathbb{R}^{l \times 768}$. We use the embedding (\mathbf{e}) that corresponds to the CLS position of \mathbf{S} as the embedding of the news title

$$\mathbf{e} = \mathbf{S}_{|CLS|}, \quad (4)$$

where $\mathbf{e} \in \mathbb{R}^{1 \times 768}$. Next, \mathbf{e} is fed into a feedforward neural network (FNN) and a softmax to predict the market-driven sentiment labels (\hat{y}^{pre}) during pre-training

$$\hat{y}^{pre} = softmax(FNN(\mathbf{e})). \quad (5)$$

³ Investors may be positive about the decrease of the debt ratio because a lower debt ratio means lower corporate operational risk. In contrast, investors may be negative about the decrease in investment, e.g., the investment in marketing, research, and development, because lower investment means that the company's development has encountered obstacles.

⁴ Hundsun Electronics (<https://www.hundsun.com>) is a leading financial technology company in China to provide one-stop financial technology services.

⁵ Company information is about the change of business and managers, qualifications and honors and external credit ratings. Business information is about business risks, products, projects, cooperation, contracts, operations, and violations of laws and regulations. Financial information is about equity assets, enterprise finance, financing transactions, and debts.

⁶ CLS and SEP are predefined special tokens in RoBERTa.

Cross Entropy loss is employed for pre-training, which is given by

$$\mathcal{L}^{pre} = CrossEntropy(\hat{y}^{pre}, y^{pre}), \quad (6)$$

where y^{pre} denotes the ground truth pre-training label, i.e., positive or negative.

Thus, we can use the pre-trained sentiment classifier to generate market-driven sentiment embeddings (\mathbf{e}) for a given input news title. In our downstream task applications, there is likely more than one news for a stock in a day. We average the news embeddings, yielding a daily news feature representing all news titles in a day

$$\mathbf{n} = \frac{1}{m} \sum_i^m \mathbf{e}_i, \quad (7)$$

where m is the number of news titles of the stock in a day. The workflow of generating news embeddings can be viewed in Fig. 2.

For simplification, we combine Eqs. (3), (4) and (7). Given a set of news titles of the stock on a trading day, we represent the daily news feature by

$$\mathbf{n} = MSC(news_1, news_2, \dots, news_m), \quad (8)$$

where $MSC(\cdot)$ denotes the news embedding generator based on the pre-trained market-driven sentiment classifier.

3.2.2. The news features of a target stock

We employ the pre-trained MSC module ($MSC(\cdot)$ in Eq. (8)) to extract news features ($\mathbf{n}_{tar,t} \in \mathbb{R}^{1 \times 768}$) for a target stock in day t . If there is no financial news about the target stock on a trading day, the news feature is a 0-vector. Then, we prepare the news features of each target stock over n trading days for training and testing. Similarly, for each training step, we look up the news features from T trading days before the day (day $t+1$) of stock price movement prediction. MAC employs a linear layer ($Linear(\cdot)$) on the MSC output to reduce the dimensionality of news features for the target stock. Thus, the news features of a target company are given by

$$N_{tar} = Linear([\mathbf{n}_{tar,t-T+1}, \dots, \mathbf{n}_{tar,t-1}, \mathbf{n}_{tar,t}]), \quad (9)$$

where $N_{tar} \in \mathbb{R}^{T \times 200}$ is the news feature matrix of a target stock. $Linear(\cdot)$ is updated during the MAC training.

3.3. Extraction of news features for related stocks

The extraction of news features for related stocks is divided into three parts: firstly, the discovery of related stocks (Section 3.3.1), secondly, the news features of related stocks (Section 3.3.2), thirdly, the GCN-based information fusion of related stocks (Section 3.3.3).

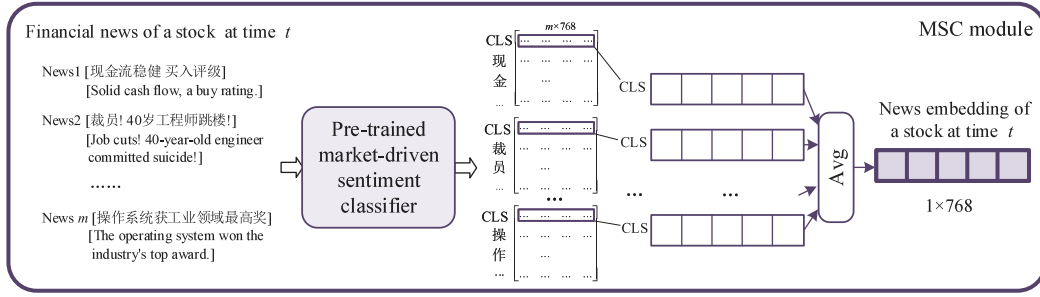


Fig. 2. The diagram of the MSC module for extracting daily financial news features.

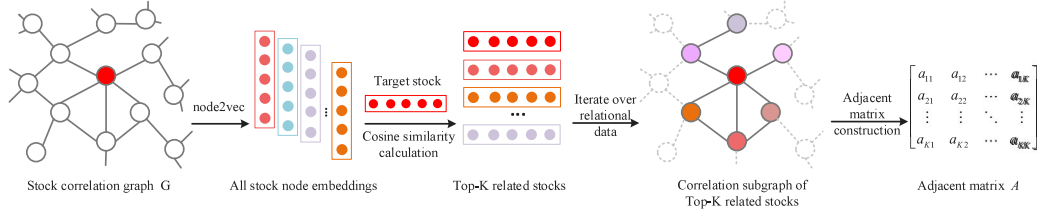


Fig. 3. Illustration of related stock discovery and adjacent matrix construction.

3.3.1. Discovery of related stocks

The discovery of related stocks refers to the identification of the most related stocks to a target stock. We select several of the most related stocks in order to avoid introducing extra noise from less relevant news and also to reduce the computation cost. The stock relationships are collected from Qichacha.⁷

In total, we collected the relationship data for 3495 stocks in the Chinese stock market. For each stock, we collected its relation data, including the supply chain (suppliers and customers), the shareholding chain (holding and being held) and the industry competitor relationships. Next, we identify the most related stocks to a target stock. Fig. 3 illustrates the process of the relevant stock discovery and the adjacent matrix construction for the most related stocks. First, we use the collected data about relationships to construct a stock relation graph. The graph ($G = (V, E)$) includes 3495 nodes (V) and 22,886 edges (E). Each node denotes a stock, and each edge denotes the connection between two stocks. Then, we use a node2vec [57] algorithm to embed the graph, yielding vector representations for all nodes. Next, given a target stock and its node2vec embedding, we use cosine similarity to identify the top- K related stocks to the target stock. The most related one among the top- K is the target stock itself because its cosine similarity is the highest. Finally, we construct an adjacency matrix $A \in \mathbb{R}^{K \times K}$ which represents the relationships among the K most related stocks to each other. K is a hyperparameter. We examine K with different values in Section 5.4.

The adjacency matrix A is defined as Eq. (10). $a_{i,j}$ is the element of Matrix A . If there is a connection between stocks s_i and s_j , the conjunction $a_{i,j}$ is 1, otherwise is 0. The row and column are indexed with the relevant stocks that are sorted in descending order by the values of the cosine similarities.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,K} \\ a_{2,1} & a_{2,2} & \dots & a_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K,1} & a_{K,2} & \dots & a_{K,K} \end{bmatrix}. \quad (10)$$

⁷ Qichacha (<https://www.qcc.com/>) is a leading information service in China, which focuses on enterprise credit and relationship information of companies.

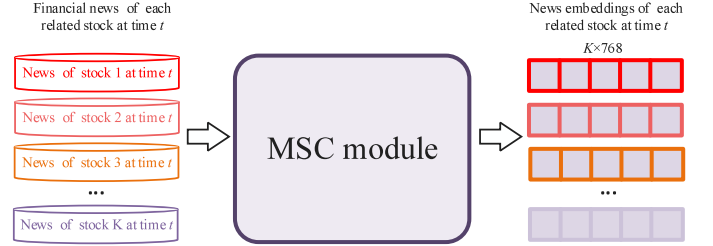


Fig. 4. Illustration of the construction of the feature matrix for related stocks.

3.3.2. The news features of related stocks

In this step, we construct the feature matrix by taking financial news of related stocks into account. Fig. 4 shows the construction of the feature matrix for related stocks. After collecting the financial news titles about each stock among Top- K related stocks on the trading day t , the MSC module (see Section 3.2.1) is also applied to extract news features for each related stock. Given Eq. (8), we extract the news features ($\mathbf{n}_{i,t}$) for the i th important related stock on the trading day t . Then, we obtain the feature matrix $N_{rel,t} \in \mathbb{R}^{K \times 768}$ that represents the news features of the K important related stocks on the trading day t by

$$N_{rel,t} = [\mathbf{n}_{1,t}, \mathbf{n}_{2,t}, \dots, \mathbf{n}_{i,t}, \dots, \mathbf{n}_{K,t}]. \quad (11)$$

In each training step, we also look up the news features (N_{rel}) of the important related stocks in the time window T

$$N_{rel} = [N_{rel,t-T+1}, \dots, N_{rel,t-1}, N_{rel,t}], \quad (12)$$

where $N_{rel} \in \mathbb{R}^{(T \times K) \times 768}$. The corresponding adjacency matrix (M) with T time steps is defined as

$$det(M) = \underbrace{[A, A, \dots, A]}_T, \quad (13)$$

where $M \in \mathbb{R}^{(T \times K) \times (T \times K)}$. $A \in \mathbb{R}^{K \times K}$ is an adjacency matrix (see Eq. (10)) on the diagonal of M . The rest of the positions in M are zeros.

3.3.3. GCN-based information fusion of related stocks

Since our proposed MAC model intends to incorporate news about related stocks and we represent the relationships of a target stock

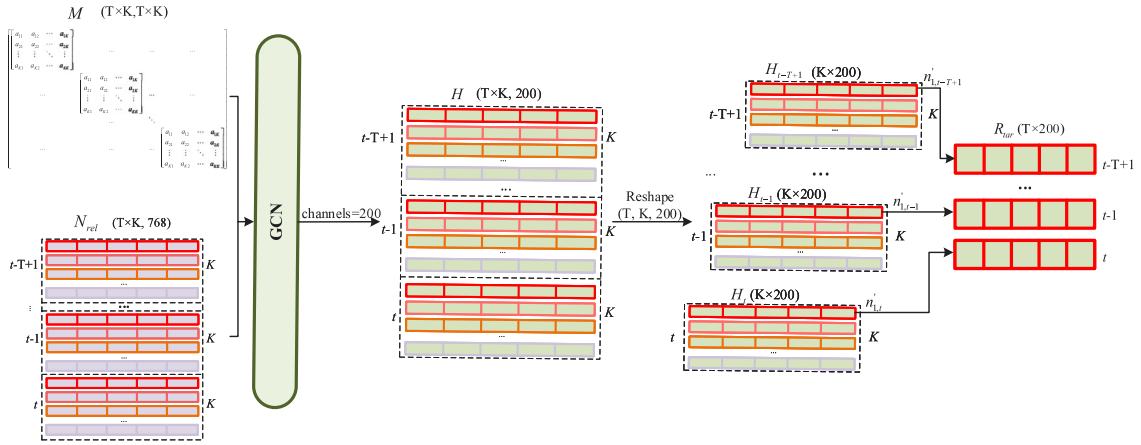


Fig. 5. Information fusion of related stocks based on GCN.

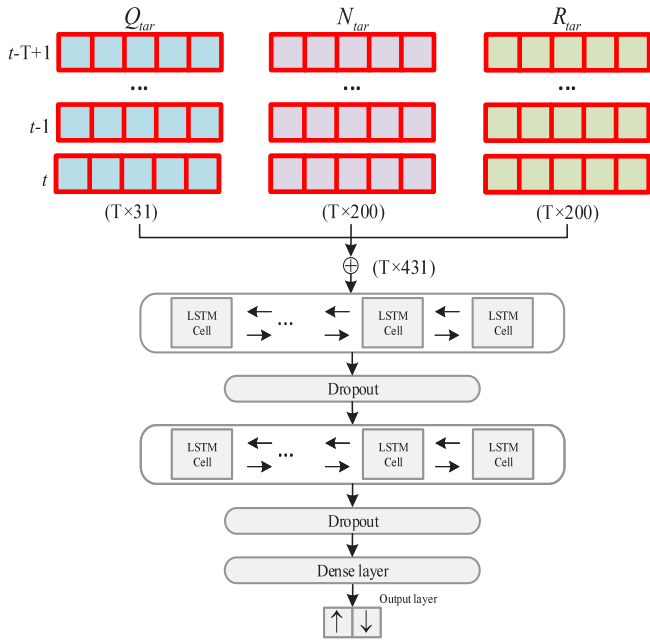


Fig. 6. The prediction of stock price movements based on BiLSTM.

and its related stocks in a graph, we use the GCN model from the work of Kipf and Welling [34] to learn the graph, thereby fusing the information of related stocks. GCN is selected as it can effectively capture the structural information of nodes and edges [58,59]. The input of the GCN model contains: (1) a feature matrix including the features of each node and (2) an adjacency matrix showing the relations between these nodes. Through the convolutional operation, the feature of each node is updated by using the given adjacency matrix to fuse the feature representations of its neighbors. Fig. 5 shows the procedure of the information fusion of related stocks based on the GCN model. We first build the news feature matrix N_{rel} of related stocks by a rolling window with the time step T as input of the GCN (see Eq. (12)).

We also build the corresponding adjacency matrix (M) with T time steps (see Eq. (13)). Then, the news feature matrix (N_{rel}) and the adjacency matrix (M) are fed into a GCN layer for the convolutional operation, which is presented as:

$$H = ReLU(\tilde{D}^{-1/2} M \tilde{D}^{-1/2} N_{rel} W), \quad (14)$$

where $\tilde{D}^{-1/2} M \tilde{D}^{-1/2} \in \mathbb{R}^{(T \times K) \times (T \times K)}$ is a Laplacian matrix. $\tilde{D} \in \mathbb{R}^{(T \times K) \times (T \times K)}$ is a diagonal degree matrix, where $\tilde{D} = \sum_j M_{i,j}$. $M \in$

$\mathbb{R}^{(T \times K) \times (T \times K)}$ is the adjacency matrix. $N_{rel} \in \mathbb{R}^{(T \times K) \times 768}$ is the news feature matrix of the related stocks. $W \in \mathbb{R}^{768 \times 200}$ is the trainable weight matrix of the GCN layer. $ReLU(\cdot)$ is the activation function. $H \in \mathbb{R}^{(T \times K) \times 200}$ is the output of the GCN layer.

The feature vector of each node is updated by fusing the information of its neighbor nodes in GCN. In this way, the updated representation of each stock node contains the information of its related stocks on the same trading day. Next, we reshape the matrix H from 2-dimension ($\mathbb{R}^{(T \times K) \times 200}$) to 3-dimension ($\mathbb{R}^{T \times K \times 200}$). It is represented as

$$H = [H_{t-T+1}, \dots, H_{t-1}, H_t], \quad (15)$$

where $H_t \in \mathbb{R}^{K \times 200}$ is the matrix that includes K related stock feature vectors on the trading day t .

H_t can be represented as

$$H_t = [\mathbf{n}'_{1,t}, \mathbf{n}'_{2,t}, \dots, \mathbf{n}'_{i,t}, \dots, \mathbf{n}'_{K,t}], \quad (16)$$

where $\mathbf{n}'_{1,t} \in \mathbb{R}^{1 \times 200}$ is the updated feature vector of the Top-1 related stock (the target stock as explained before) by fusing the information of its related stocks from Top-2 to Top- K on the trading day t .

Lastly, we extract the feature vector \mathbf{n}'_i on each trading day in the window size T to build the matrix $R_{tar} \in \mathbb{R}^{T \times 200}$, which represents the sequential news features of related stocks of a target stock in the past T trading days. R_{tar} is defined as

$$R_{tar} = [\mathbf{n}'_{1,t-T+1}, \dots, \mathbf{n}'_{1,t-1}, \mathbf{n}'_{1,t}]. \quad (17)$$

3.4. Prediction of stock price movements

BiLSTM is an effective sequential learning model, which achieves great performance in processing sequential data [60]. In this part, we leverage BiLSTM to capture temporal financial information to predict stock price movements. We also tested other popular NLP encoders, i.e., LSTM [46], GRU [61], Bi-GRU and Transformer [62], however, BiLSTM is the most robust model to perform this task. The process of BiLSTM-based stock price movement prediction is shown in Fig. 6. Firstly, given the rolling window size of T , the quantitative indicators of a target stock Q_{tar} , the news features of the target stock N_{tar} and the news features of its related stocks R_{tar} are concatenated, which is denoted as X . The input of BiLSTM is

$$X = [Q_{tar} \oplus N_{tar} \oplus R_{tar}], \quad (18)$$

where $X \in \mathbb{R}^{T \times 431}$. \oplus denotes concatenation. $\mathbf{x}_{t-T+1}, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t$ are the concatenated daily features in X , where $\mathbf{x}_t \in \mathbb{R}^{1 \times 431}$.

Then, X is fed into two BiLSTM layers (the number of BiLSTM layers is verified in Section 5.4) to discover temporal patterns for stock price movement prediction. We employ dropout ($D(\cdot)$) after each BiLSTM

$$h_{t+1} = D(BiLSTM_2(D(BiLSTM_1(X)))) \quad (19)$$

Table 3
The basic information of each target stock.

Code	Name	Industry
000063	ZTE Corporation	Manufacturing of communications
000651	Gree Electric Appliances. Inc of Zhuhai	Manufacturing of electrical machinery
601800	China Communications Construction Ltd.	Civil engineering construction
000876	New Hope Liuhe Ltd.	Farm and sideline food processing
600104	Saic Motor Ltd.	Automobile industry
601933	Yonghui Superstores Ltd.	Wholesales and retail trade

Table 4
Dataset statistics. The # tar denotes the number of target stock news, the # rel denotes the number of related stock news, the # days denotes the number of trading days, the # up denotes the number of up days and the # down denotes the number of down days.

Stock	Dataset	Start	End	# tar	# rel	# days	# up	# down
000063	Train	02/01/2018	21/10/2020	1992	3924	640	325	315
	Val.	22/10/2020	18/02/2021	540	2157	80	37	43
	Test	19/02/2021	18/06/2021	1011	2732	80	37	43
	All	02/01/2018	18/06/2021	3543	8813	800	399	401
000651	Train	02/01/2018	14/10/2020	1422	2054	668	329	339
	Val.	15/10/2020	09/02/2021	404	1547	83	46	37
	Test	10/02/2021	18/06/2021	784	2153	83	40	43
	All	02/01/2018	18/06/2021	2610	5754	834	415	419
601800	Train	02/01/2018	12/10/2020	209	3850	672	315	357
	Val.	13/10/2020	08/02/2021	92	1424	84	32	52
	Test	09/02/2121	18/06/2021	67	2162	84	40	44
	All	02/01/2018	18/06/2021	368	7436	840	387	453
000876	Train	02/01/2018	12/10/2020	623	2530	672	349	323
	Val.	13/10/2020	08/02/2021	429	1561	84	40	44
	Test	09/02/2121	18/06/2021	569	2259	84	31	53
	All	02/01/2018	18/06/2021	1621	6350	840	420	420
600104	Train	02/01/2018	12/10/2020	863	4474	672	331	341
	Val.	13/10/2020	08/02/2021	464	2315	84	40	44
	Test	09/02/2121	18/06/2021	695	3578	84	40	44
	All	02/01/2018	18/06/2021	2022	10367	840	411	429
601933	Train	02/01/2018	13/10/2020	399	1927	672	329	343
	Val.	14/10/2020	09/02/2021	203	1221	84	35	49
	Test	10/02/2021	18/06/2021	274	2061	83	34	49
	All	02/01/2018	18/06/2021	876	5209	839	398	441
Overall		02/01/2018	18/06/2021	11040	43929	4993	2430	2563

where $h_{t+1} \in \mathbb{R}^{1 \times 256}$ is the concatenation of the last hidden states of the forward and backward outputs of $BiLSTM_2(\cdot)$ with dropout. It is used for predicting the stock price movement on day $t + 1$.

Finally, h_{t+1} is fed into a dense layer and softmax to predict the probability of a movement label (\hat{y}_{t+1})

$$\hat{y}_{t+1} = \text{softmax}(\text{Dense}(h_{t+1})). \quad (20)$$

We employ Cross Entropy loss to update the neural network

$$\mathcal{L} = \text{CrossEntropy}(\hat{y}_{t+1}, y_{t+1}), \quad (21)$$

where y_{t+1} denotes the ground truth label, i.e., up or down.

4. Experiments

4.1. Datasets

The stocks used in our experiments are from the Chinese stock market. We select six representative target stocks (no intersection with the pre-training data) from different industries. These stocks have higher market capital and more financial news in their respective industries. The basic information of each target stock is shown in Table 3. We collected 54,969 news titles from January 2, 2018 to June 18, 2021, 11,040 of which were about the six target stocks, and the rest (43, 929) were about the respectively related stocks. The statistics of training (80%), validation (10%) and testing (10%) sets are summarized in Table 4. Following previous research [14,63], to maximally utilize the available data, we conduct a ‘walk forward testing’ method [63] to train and validate the model.

4.2. Baselines

To evaluate the performance of the proposed approach, we benchmark it against the baselines, including eLSTM, LSTM-news, GCN-S and RoBERTa. The baselines are described as follows:

- eLSTM [17] is a tensor-based event-driven LSTM model that effectively utilizes the news signals for stock price movement prediction. The authors developed a dictionary to generate sentiment features. The original settings of eLSTM and their developed dictionary are adopted in our benchmarking.
- LSTM-news [14] is an LSTM-based model for stock price prediction that incorporates technical indicators and news sentiments of target stocks. The news sentiment features of the original work are given by an English financial sentiment dictionary. However, we cannot find the paired dictionary in Chinese. Thus, we employ our pre-trained MSC to generate news features. The quantitative indicators and model structure are implemented according to the original settings.
- GCN-S [28] is a GCN-based model for stock price movement prediction by incorporating a shareholding graph. The employed features are OP, CP, HP, LP and Vol (see Table 2). The original settings of GCN-S are adopted.
- RoBERTa [33] is one of the most popular and competitive methods for text analysis. The pre-trained RoBERTa is adopted to extract news features. In the experiments, the news features of target stocks on a trading day t are directly used to predict stock price movements on the trading day $t + 1$.

Table 5
The results of different prediction models.

Methods	Metr.	000063	000651	601800	000876	600104	601933	Avg.
RoBERTa	ACC	0.6017	0.5941	0.5688	0.5652	0.5909	0.5962	0.5862
	MCC	0.1949	0.1855	0.0899	0.0685	0.1750	0.1744	0.1480
GCN-S	ACC	0.5676	0.5613	0.5962	0.5962	0.5897	0.5897	0.5835
	MCC	0.1471	0.1272	0.2026	0.1348	0.1759	0.1136	0.1502
LSTM-news	ACC	0.6096	0.6078	0.5714	0.5897	0.5909	0.5974	0.5945
	MCC	0.2174	0.2717	0.1144	0.1524	0.1840	0.1782	0.1864
eLSTM	ACC	0.6795	0.6383	0.6646	0.6218	0.6667	0.6220	0.6488
	MCC	0.3548	0.3160	0.3147	0.2347	0.3326	0.2414	0.2990
Ours	ACC	0.7000	0.6828	0.6859	0.6346	0.6923	0.6402	0.6726
	MCC	0.4017	0.3697	0.3593	0.2775	0.3821	0.2808	0.3452

We do not benchmark with FinBERT (a pre-trained language model in the financial domain) [64] because it does not support Chinese text.

4.3. Setups

The basic framework of our model is performed by using Keras and TensorFlow. The batch size is 32 and the maximum number of training epochs is 200. The early-stop mechanism is employed to automatically stop the training process and reduce the overfitting risk.

We adopt the Adam optimizer with an initial learning rate of 0.001 to train the model parameters. The dimension of BiLSTM hidden states is 128. The activity regularization L2 (0.001) is applied to the weight matrix of kernel. In particular, several experiments are carried out to explore the optimal time step T within $\{1, 2, \dots, 14\}$, the number of related stocks K within $\{1, 2, \dots, 15\}$, and the number of BiLSTM layers within $\{1, 2, 3, 4\}$, based on the validation sets. These hyperparameters are verified in Section 5.4. We report the benchmarking results on the testing sets based on the optimal time step ($T = 6$), the number of related stocks ($K = 10$) and two BiLSTM layers.

4.4. Evaluation metrics

The evaluation of a data science model is difficult when applied to the financial domain [44]. Thus, we evaluate the proposed model in terms of classification performance and financial evaluation. To evaluate classification performance, following previous studies [17,49,50], accuracy (ACC) and Matthews Correlation Coefficient (MCC) are adopted as the evaluation metrics. MCC can avoid data imbalance biases. The two metrics are defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (22)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (23)$$

where TP is true positive, FP is false positive, TN is true negative and FN is false negative.

For financial evaluation, we simulate stock trading according to the prediction results of stock price movements. We focus on trading returns and risks. Total money (TMoney) [65] and Sharpe Ratio [36] are adopted to evaluate the simulated trading performance. Sharpe Ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk. The annual risk-free rate is 3.0%. The two metrics are defined by:

$$TMoney = Available\ Capital + Closing\ Price \times Number\ of\ Stocks, \quad (24)$$

$$SharpeRatio = \frac{Rate\ of\ Return - Risk\ Free\ Rate}{Standard\ Deviation\ of\ Return}. \quad (25)$$

5. Results

5.1. Results of stock price movement prediction

In this section, we compare the classification performance of our MAC model against the state-of-the-art baselines. The evaluation results of different models for six target stocks are shown in Table 5. It can be observed that our proposed method achieves the best performance for the six stocks in terms of both the ACC and MCC metrics, yielding at least 2.38% gains in ACC and 4.62% gains in MAC on average. Compared with eLSTM (the strongest baseline), we find that the GCN module can better capture the structural information of the stock correlation graph and aggregate the effects of related stocks.

The pre-training of market-driven sentiment classification also performs better. Besides, compared with LSTM-news which only uses the quantitative indicators and news features of the target stocks, the improvements of our method are presented with 7.81% and 15.88% gains in ACC and MCC, respectively, which demonstrates the effectiveness of using the news about related stocks. Next, compared with GCN-S, which uses a single shareholding graph and price features, the results demonstrate that our method can capture more effective stock relations and useful news features. The result of RoBERTa shows that it is a useful module for learning news features. However, the gap between RoBERTa and our MAC proves the advantage and necessity of both learning the time series information in the financial domain and integrating features from multiple sources.

In summary, the results indicate that the proposed method achieves significant improvements over the baselines by considering the sequential dependencies of financial data, market-driven sentiment pre-training and incorporating information from multiple sources, such as news about related stocks.

5.2. Ablation studies

We conduct several ablation experiments to analyze the effectiveness of each feature of MAC, including (A) the quantitative indicators of the target stocks (Q_{tar}), (B) the news features of the target stocks (N_{tar}) and (C) the news features of the related stocks (R_{tar}). We use A+B+C to represent the proposed full MAC model. A, B, C, A+B and A+C are considered the baselines of the ablation study. The news features in B and C are given by our pre-trained MSC module. To demonstrate the utility of MSC and pre-training, we examine the other two feature generators as baselines. A+B₀+C₀ represents the concatenation of all the features, where the news features are extracted from the original Chinese-RoBERTa without domain-specific pre-training. A+B₁+C₁ represents the concatenation of all the features, where the news features are given by a pre-trained semantics-based sentiment classifier. The classifier is also powered by Chinese-RoBERTa. However, we use a typical Chinese financial news sentiment analysis dataset ('THUC-News' [66,67]) whose labels are given by the semantic interpretations of annotators to pre-train the classifier.

Table 6

Results of the ablation analysis. ‘A’ denotes the quantitative indicators of the target stocks (Q_{tar}); ‘B’ denotes the news features of the target stocks (N_{tar}); ‘C’ denotes the news features of the related stocks (R_{tar}); ‘A+B₀+C₀’ denotes the concatenation of all the features, in which the news features are extracted from the original Chinese-RoBERTa that is not pre-trained by a domain-specific corpus; ‘A+B₁+C₁’ denotes the concatenation of all the features which are extracted from Chinese-RoBERTa that is pre-trained with a semantics-driven financial news sentiment analysis dataset and, lastly, ‘A+B+C’ denotes our proposed full MAC model.

Features	Metri.	000063	000651	601800	000876	600104	601933	Avg.
A(Q_{tar})	ACC	0.5779	0.5652	0.5803	0.5803	0.5741	0.5741	0.5753
	MCC	0.1531	0.1274	0.1274	0.1274	0.1408	0.0864	0.1191
B(N_{tar})	ACC	0.5890	0.6144	0.5741	0.5897	0.5769	0.5962	0.5901
	MCC	0.1701	0.2682	0.0903	0.1065	0.1505	0.1744	0.1600
C(R_{tar})	ACC	0.5867	0.6138	0.5786	0.5833	0.5833	0.5976	0.5906
	MCC	0.1656	0.2294	0.1218	0.1161	0.1560	0.1250	0.1523
A+B	ACC	0.6313	0.6340	0.5926	0.6169	0.6234	0.6173	0.6193
	MCC	0.2560	0.3090	0.1660	0.1831	0.2405	0.2219	0.2294
A+C	ACC	0.6233	0.6340	0.5926	0.6169	0.6115	0.6013	0.6133
	MCC	0.2438	0.3090	0.1441	0.1831	0.2188	0.1593	0.2097
A+B ₀ +C ₀	ACC	0.6800	0.6621	0.6539	0.6218	0.6795	0.6220	0.6532
	MCC	0.3671	0.3281	0.2986	0.2710	0.3583	0.2252	0.3081
A+B ₁ +C ₁	ACC	0.6953	0.6690	0.6654	0.6218	0.6859	0.6282	0.6609
	MCC	0.3872	0.3432	0.3165	0.2710	0.3706	0.2202	0.3181
A+B+C	ACC	0.7000	0.6828	0.6859	0.6346	0.6923	0.6402	0.6726
	MCC	0.4017	0.3697	0.3593	0.2775	0.3821	0.2808	0.3452

The results of the ablation study are shown in Table 6. When we use the features individually (A vs. B vs. C) for stock price movement prediction, the news features (B and C) are more effective than the quantitative indicators (A). When we integrate two feature sources, the performance is improved with A+B outperforming A+C. This means that using the quantitative indicators and the news features of target stocks is more effective than using the quantitative indicators of target stocks and the news features of their related stocks. When we combine all the features in our MAC model, A+B+C obtains the highest ACC and MCC values proving the effectiveness of our model. On the other hand, the use of the original Chinese-RoBERTa (A+B₀+C₀) achieves the worst average results in comparison with A+B₀+C₀, A+B₁+C₁ and A+B+C.

Pre-training a Chinese-RoBERTa with a semantics-driven sentiment analysis dataset (A+B₁+C₁) can yield better results (+0.77% in ACC; +1.00% in MCC) than the original Chinese-RoBERTa (A+B₀+C₀). However, its performance is still weaker (−1.94% in ACC; −3.71% in MCC) than our proposed market-driven sentiment classification pre-training (A+B+C). This marked difference proves the effectiveness of our proposed market-driven sentiment classification pre-training due to more direct corresponding to actual stock market reactions towards news than the semantics-based sentiment polarity analysis.

5.3. Financial evaluation

In terms of financial evaluation, we performed simulations of stock trading (backtesting) according to the prediction results of different models. The backtesting data is based on the combination of the validation and testing sets of each target stock. The Buy & Hold strategy is adopted as an additional financial benchmark in this evaluation task. We take 10,000 RMB as the initial investment capital. And we assume that the transaction cost is zero during the backtesting trading. If the predicted label is 1 (up) on the next trading day ($t + 1$), we buy the target stock at the closing price of the current trading day (t), using all of the currently available capital. If the predicted label is 0 (down), we sell the stock at the closing price of the current trading day (t).

If the predicted label for the next trading day ($t + 1$) is the same as yesterday’s predicted label for the current trading day (t), we do not buy or sell. For the Buy & Hold strategy, we buy the stock at the beginning with all of the initial capital and hold it without any further operations [68,69]. Thus, the capital of Buy& Hold also reflects the price movements of the stock in real time. The backtesting trading

Table 7

The Sharpe Ratio of different methods.

Model	000063	000651	601800	000876	600104	601933
B&H	0.0065	−1.1768	−0.8146	−2.0120	−0.0249	−2.5294
GCN-S	1.0378	−0.4178	0.8530	−0.0029	1.5067	−1.7214
RoBERTa	1.2566	−0.0382	0.8782	0.1258	1.2418	−1.7184
LSTM-news	1.3609	0.1043	0.9847	0.1969	1.2621	−1.9863
eLSTM	2.3786	0.2975	1.6989	0.5676	1.8346	−0.9372
Ours	3.0053	1.0384	2.1627	0.9668	2.8595	−0.1238

results of all the target stocks are shown in Fig. 7. As seen, the proposed MAC model can gain more profits for all the target stocks than the baseline methods. The proposed method achieves positive returns on five target stocks, except for Stock 601933. However, MAC achieves lower losses than the baselines for Stock 601933, yielding more returns (745.96 RMB) than eLSTM. Therefore, the results of backtesting trading show that the proposed method can not only obtain a higher ACC (shown in Section 5.1), i.e., more accurate classification results, but also can capture more effective transaction signals to subsequently achieve higher profits. The Sharpe Ratio of different methods is shown in Table 7. A higher Sharpe Ratio means a higher return relative to the amount of risk taken. As seen in Table 7, the proposed MAC method achieves a higher Sharpe Ratio for all six target stocks.

The negative Sharpe Ratio of stock 601933 means the return does not exceed the risk-free rate (3.0%). To a large extent, this is due to the dramatic decline in its market performance, as shown by the capital line of the Buy & Hold strategy in Fig. 7. Nevertheless, the proposed method still achieves relatively better performance than the baseline methods. These results further indicate the validity and superiority of MAC.

5.4. Hyperparameter analysis

Financial data are typical time-series data. To capture financial temporal information, we perform several experiments in order to explore the optimal window size T within $\{1, 2, \dots, 14\}$. Besides, we also test the number of related stocks K within $\{1, 2, \dots, 15\}$ to capture enough useful information from related stocks, as well as to avoid introducing noise that can mislead target stock predictions. Last, we report the MAC model performance with different numbers of BiLSTM layers, i.e., 1, 2, 3 and 4. These experiments are conducted based on the validation sets of all the target stocks. The ACCs corresponding to different window

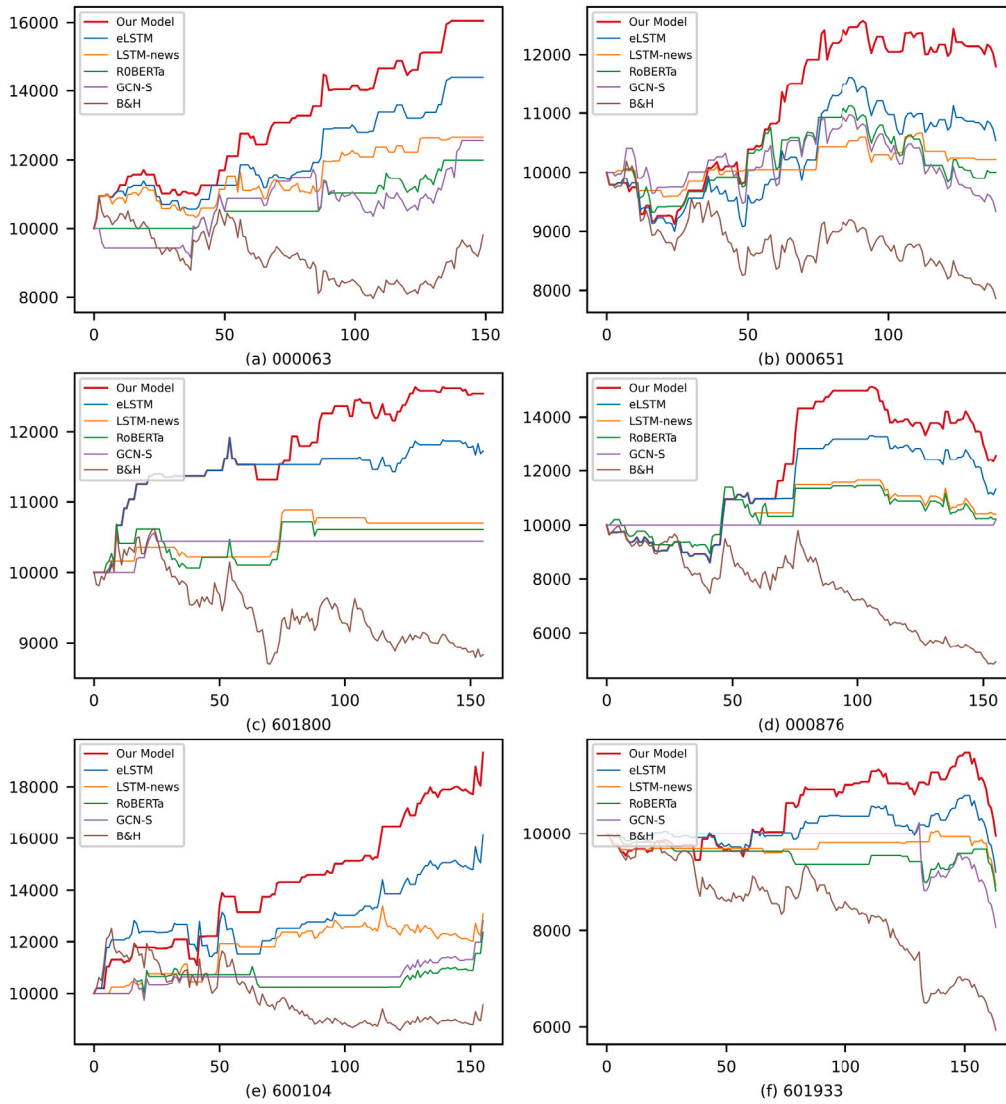


Fig. 7. TMoney of different models. ‘B&H’ denotes a Buy & Hold trading strategy.

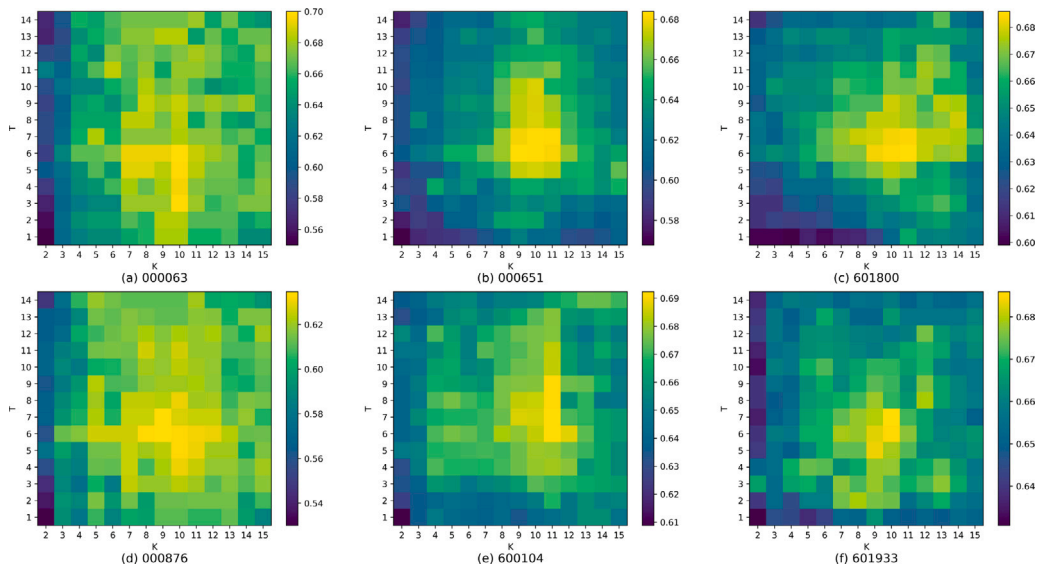


Fig. 8. The prediction ACCs by different window sizes (T) and the different number of related stocks (K). The results are based on the validation sets of the six target stocks.

Table 8

Results of the breakdown analysis. ‘A+B+C’ denotes our proposed full MAC model, using the quantitative indicators of a target stock (A), all types of news about the target stock (B) and it related stocks (C). ‘A+B_C+C_C’, ‘A+B_B+C_B’, and ‘A+B_P+C_P’ use news about company information (C), business information (B) and financial information (P), respectively. ‘A+B_P+C_P’ and ‘A+B_N+C_N’ use positive news (P) and negative news (N), respectively.

Features	Metr.	000063	000651	601800	000876	600104	601933	Avg.
A+B _C +C _C	ACC	0.6223	0.6306	0.6064	0.6017	0.6239	0.6098	0.6158
	MCC	0.2410	0.2617	0.1684	0.1949	0.2468	0.1963	0.2182
A+B _B +C _B	ACC	0.6310	0.6343	0.6104	0.6096	0.6330	0.6037	0.6203
	MCC	0.2577	0.2675	0.1786	0.2152	0.2590	0.2155	0.2323
A+B _P +C _P	ACC	0.6388	0.6370	0.6154	0.6104	0.6463	0.6159	0.6273
	MCC	0.2717	0.2706	0.1911	0.2168	0.2916	0.1962	0.2397
A+B _P +C _P	ACC	0.6407	0.6403	0.6096	0.6090	0.6296	0.6154	0.6241
	MCC	0.2788	0.2944	0.1782	0.1904	0.2634	0.2508	0.2427
A+B _N +C _N	ACC	0.6543	0.6414	0.6194	0.6098	0.6667	0.6201	0.6353
	MCC	0.3080	0.2850	0.2144	0.1868	0.3309	0.2399	0.2608
A+B+C	ACC	0.7000	0.6828	0.6859	0.6346	0.6923	0.6402	0.6726
	MCC	0.4017	0.3697	0.3593	0.2775	0.3821	0.2808	0.3452

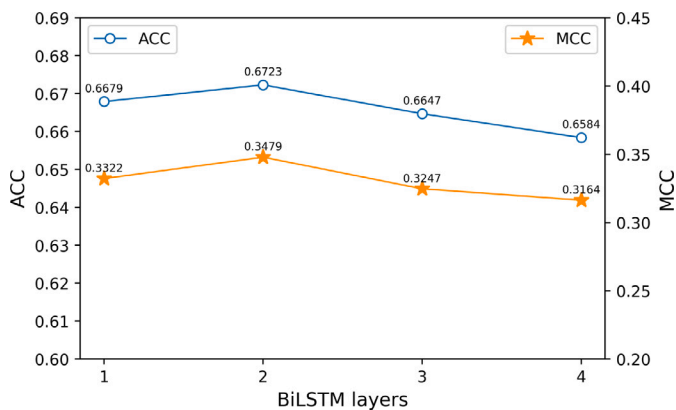


Fig. 9. The prediction ACC and MCC by different numbers of BiLSTM layers. The results are averaged over the six target stocks on their validation sets.

sizes (T) and related stocks (K) are shown in Fig. 8, which shows that most stocks achieve higher ACC by $T = 6$ and $K = 10$ (the lighter areas). Larger T and K do not bring extra gains to the model. Fig. 9 shows that using two BiLSTM layers achieves the optimal performance (the highest averaged ACC and MCC over the six stocks). Thus, we employ the above hyperparameters in our proposed MAC model.

5.5. Breakdown analysis

As mentioned in Section 3.2.1, the news about listed companies can be categorized as company information, business information and financial information. On the other hand, news can also be categorized by sentiment polarities, e.g., positive news regarding stock rises and negative news regarding stock falls.

Here, we conduct the breakdown analysis to explore what types of news have higher utilities in predicting stock price movements based on the above categories. To conduct the experiments, we consider the news features from each class together with the quantitative indicators. The results of the breakdown analysis are shown in Table 8. Compared with the model integrating news about company information (A+B_C+C_C) and business information (A+B_B+C_B), the model integrating news about financial information (A+B_P+C_P) achieves better performance in ACC and MCC. It shows that financial information-related news has a higher utility in stock price movement prediction. The negative news-based model (A+B_N+C_N) outperforms the positive news-based model (A+B_P+C_P) by 1.12% in ACC and 1.81% in MCC on average. It indicates that negative news regarding stock falls has a higher utility in stock movement prediction than positive news regarding stock rises.

When all the news types are considered in our MAC model (A+B+C), the model achieves the highest ACC and MCC. The results demonstrate that the employed news types can complement each other, presenting the highest utility in the stock price movement prediction task.

6. Conclusion

In order to capture the news effect of related stocks, we propose a model by combining GCN and BiLSTM to predict stock price movements in the Chinese stock market. Our model integrates multi-source information, e.g., quantitative indicators, news about a target company, and news about the related companies of the target company. To achieve effective news embeddings, we pre-train a market-driven sentiment classifier, which reflects the sentiment of the real market and investors to a piece of news.

We examine our method on six target stocks that are from different industries. The experimental results show that our method surpasses the strongest baselines by at least 2.38% and 4.62% in terms of the averaged ACC and MCC, respectively. Regarding the financial evaluation, our model also performs better, showing more profits and a higher Sharpe Ratio than the state-of-art baselines. These results demonstrate that the prediction performance can be improved significantly by aggregating news features of related stocks. Moreover, the pre-trained market-driven sentiment classifier is more effective for news feature embeddings than the semantics-based sentiment polarity analysis due to more direct corresponding to actual stock market reactions towards news. In summary, the proposed method can achieve better performance for stock price movement prediction and provides better decision supports for investors to optimize their investment strategies and reduce risks caused by margin trading.

Our model presents some limitations that should be addressed in future work. First, the model does not detect if a piece of news is real. A piece of news about a target stock may be fake information published by competitors for market competition. A fake news detector [68] should be introduced and the impacts of real and fake news on stock markets should be investigated in the future. Second, we find that figurative languages, e.g., metaphors, widely appear in financial news, which creates challenges for computational language understanding. To tackle this problem, metaphor processing technologies [69,70] could be introduced in order to paraphrase the metaphorical expressions into machine-friendly literal ones, thereby improving the prediction performance of stock price movements.

CRediT authorship contribution statement

Yu Ma: Conceptualization, Methodology, Writing – original draft. **Rui Mao:** Conceptualization, Methodology, Writing – review & editing. **Qika Lin:** Methodology, Software, Visualization. **Peng Wu:** Writing – review & editing, Funding acquisition. **Erik Cambria:** Writing – review & editing, Project administration, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project #A18A2b0046). The paper is also supported by the National Natural Science Foundation of China (project numbers are 72274096 and 71774084), Program for Jiangsu Excellent Scientific and Technological Innovation Team, China (project number is [2020]10).

Appendix

The Opening/Closing Price (OP/CP), Highest/Lowest Price (HP/LP), Volume of Trade (Vol), Turnover Rate (TurR), Daily Amplitude (DA), Daily Return (DR), Price-to-Earnings Ratio (P/E), Price-to-Book Ratio (P/B) and Price-to-Sales Ratio (P/S) in the Table 2 can be directly obtained from RESSET platform. The rest of the features are calculated as follows:

Moving Average (MA):

$$MA_t(CP, n) = \frac{1}{n} \sum_{k=t-n+1}^t CP_k, \tag{A.1}$$

where n is the length of the time period.

Exponential Moving Average (EMA):

$$EMA_t(CP, n) = \alpha \times CP_t + (1 - \alpha) \times EMA_{t-1}(n - 1), \tag{A.2}$$

where $\alpha = \frac{2}{n+1}$ is the smoothing factor.

Weight Average (WMA):

$$WMA_t(CP, n) = 100 \times \frac{\sum_{k=t-n+1}^t [(n-t+k) \times CP_k]}{\sum_{m=1}^n m}. \tag{A.3}$$

Moving Average Convergence and Divergence (MACD):

MACD includes DIF, DEA, and MACD histogram:

$$\begin{aligned} DIF_t &= EMA_t(CP, 12) - EMA_t(CP, 26) \\ DEA_t &= EMA_t(DIF_t, 9) \\ MACD \text{ histogram} &= 2 \times (DIF_t - DEA_t), \end{aligned} \tag{A.4}$$

Relative Strength Index (RSI):

$$RSI_t(n) = 100 - \frac{100}{1 + \frac{Average \ Gain}{Average \ Loss}}, \tag{A.5}$$

Money Flow Index (MFI):

$$MFI_t = 100 - \frac{100}{1 + \frac{Positive \ Money \ Flow}{Negative \ Money \ Flow}}, \tag{A.6}$$

where $Money \ Flow = Typical \ price \times Volume$. $Typical \ price = \frac{HP+LP+CP}{3}$.

Stochastic Indicator KDJ:

$$\begin{aligned} RSV_t(n) &= 100 \times \frac{CP_t - Min\{L\}}{Max\{H\} - Min\{L\}} \\ Slowk_t &= \frac{2}{3} \times Slowk_{t-1} + \frac{1}{3} \times RSV_t \\ Slowd_t &= \frac{2}{3} \times Slowd_{t-1} + \frac{1}{3} \times Slowk_t, \end{aligned} \tag{A.7}$$

where H and L are sets of highest prices and lowest prices in recent n days, respectively.

True Range (TR):

$$TR_t = Max\{HP_t - LP_t; HP_t - CP_{t-1}; LP_t - CP_{t-1}\}, \tag{A.8}$$

Average True Range (ATR):

$$ART_t = \frac{ART_{t-1} \times 13 + TR_t}{14}, \tag{A.9}$$

Williams Indicator (WR):

$$WR_t(n) = \frac{Max\{H\} - CP_t}{Max\{H\} - Min\{L\}}, \tag{A.10}$$

where H and L are sets of highest prices and lowest prices in recent n days, respectively. *Bollinger Bands:*

Bollinger Bands include UpperBand, MiddleBand, and LowerBand:

$$\begin{aligned} UpperBand_t &= MA_t(CP, 20) + 2 \times std(CP, 20) \\ MiddleBand_t &= 20 \times MA_t(CP, 20) \\ LowerBand_t &= MA_t(CP, 20) - 2 \times std(CP, 20), \end{aligned} \tag{A.11}$$

where $std(CP, 20)$ is standard deviation of CP in recent 20 days.

References

- [1] W. Jiang, Applications of deep learning in stock market prediction: Recent progress, *Expert Syst. Appl.* 184 (2021) 115537, <http://dx.doi.org/10.1016/j.eswa.2021.115537>.
- [2] O. Bustos, A. Pomares-Quimbaya, Stock market movement forecast: A systematic review, *Expert Syst. Appl.* 156 (2020) 113464, <http://dx.doi.org/10.1016/j.eswa.2020.113464>.
- [3] S. Merello, A. Picasso, L. Oneto, E. Cambria, Ensemble application of transfer learning and sample weighting for stock market prediction, in: *IJCNN*, 2019.
- [4] Q. Li, Y. Chen, L.L. Jiang, P. Li, H. Chen, A tensor-based information framework for predicting the stock market, *ACM Trans. Inf. Syst.* 34 (2) (2016) <http://dx.doi.org/10.1145/2838731>.
- [5] F. Xing, L. Malandri, Y. Zhang, E. Cambria, Financial sentiment analysis: An investigation into common mistakes and silver bullets, in: *COLING*, 2020, pp. 978–987.
- [6] I.K. Nti, A.F. Adekoya, B.A. Weyori, A systematic review of fundamental and technical analysis of stock market predictions, *Artif. Intell. Rev.* 53 (4) (2020) 3007–3057, <http://dx.doi.org/10.1007/s10462-019-09754-z>.
- [7] Q. Li, Y. Chen, J. Wang, Y. Chen, H. Chen, Web media and stock markets : A survey and future directions from a big data perspective, *IEEE Trans. Knowl. Data Eng.* 30 (2) (2018) 381–399, <http://dx.doi.org/10.1109/TKDE.2017.2763144>.
- [8] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: Quantifying language to measure firms' fundamentals, *J. Financ.* 63 (3) (2008) 1437–1467, <http://dx.doi.org/10.1111/j.1540-6261.2008.01362.x>.
- [9] Y. Jeon, T.H. McCurdy, X. Zhao, News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies, *J. Financ. Econ.* (2021) <http://dx.doi.org/10.1016/j.jfineco.2021.08.002>.
- [10] Y. Shi, W. Li, L. Zhu, K. Guo, E. Cambria, Stock trading rule discovery with double deep Q-network, *Appl. Soft Comput.* 107 (107320) (2021).
- [11] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, Y. Chen, The effect of news and public mood on stock movements, *Inform. Sci.* 278 (2014) 826–840, <http://dx.doi.org/10.1016/j.ins.2014.03.096>.
- [12] L. Malandri, F. Xing, C. Orsenigo, C. Vercellis, E. Cambria, Public mood-driven asset allocation: the importance of financial sentiment in portfolio management, *Cogn. Comput.* 10 (6) (2018) 1167–1176.
- [13] A. Thakkar, K. Chaudhari, Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions, *Inf. Fusion* 65 (2021) 95–107, <http://dx.doi.org/10.1016/j.inffus.2020.08.019>.
- [14] X. Li, P. Wu, W. Wang, Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, *Inf. Process. Manage.* 57 (5) (2020) 102212, <http://dx.doi.org/10.1016/j.ipm.2020.102212>.
- [15] W. Chen, C.K. Yeo, C.T. Lau, B.S. Lee, Leveraging social media news to predict stock index movement using RNN-boost, *Data Knowl. Eng.* 118 (2018) 14–24, <http://dx.doi.org/10.1016/j.datak.2018.08.003>.
- [16] M.Y. Chen, C.H. Liao, R.P. Hsieh, Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach, *Comput. Hum. Behav.* 101 (2019) 402–408, <http://dx.doi.org/10.1016/j.chb.2019.03.021>.
- [17] Q. Li, J. Tan, J. Wang, H. Chen, A multimodal event-driven LSTM model for stock prediction using online news, *IEEE Trans. Knowl. Data Eng.* 33 (10) (2021) 3323–3337, <http://dx.doi.org/10.1109/TKDE.2020.2968894>.
- [18] S. Merello, A. Picasso, Y. Ma, L. Oneto, E. Cambria, Investigating timing and impact of news on the stock market, in: *Proceedings of ICDM Workshops*, 2018, pp. 1348–1354.
- [19] F. Xing, F. Pallucchini, E. Cambria, Cognitive-inspired domain adaptation of sentiment lexicons, *Information Processing and Management* 56 (3) (2019) 554–564.

- [20] W. Li, L. Zhu, Y. Shi, K. Guo, E. Cambria, User reviews: sentiment analysis using lexicon integrated two-channel CNN-LSTM family models, *Applied Soft Computing* 94 (106435) (2020).
- [21] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, D.C.L. Ngo, Text mining for market prediction: A systematic review, *Expert Syst. Appl.* 41 (16) (2014) 7653–7670, <http://dx.doi.org/10.1016/j.eswa.2014.06.009>.
- [22] E. Cambria, Q. Liu, S. Decherchi, F. Xing, K. Kwok, SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis, in: *Proceedings of the 13th Language Resources and Evaluation Conference, European Language Resources Association, 2022*, pp. 3829–3839.
- [23] A.E. de Oliveira Carosia, G.P. Coelho, A.E.A. da Silva, Investment strategies applied to the Brazilian stock market: A methodology based on sentiment analysis with deep learning, *Expert Syst. Appl.* 184 (2021) 115470, <http://dx.doi.org/10.1016/j.eswa.2021.115470>.
- [24] M.G. Sousa, K. Sakiyama, L.d.S. Rodrigues, P.H. Moraes, E.R. Fernandes, E.T. Matsubara, BERT for stock market sentiment analysis, in: *2019 IEEE 31st International Conference on Tools with Artificial Intelligence, ICTAI, 2019*, pp. 1597–1601, <http://dx.doi.org/10.1109/ICTAI.2019.00231>.
- [25] M. Van de Kauter, D. Breesch, V. Hoste, Fine-grained analysis of explicit and implicit sentiment in financial news articles, *Expert Syst. Appl.* 42 (11) (2015) 4999–5010, <http://dx.doi.org/10.1016/j.eswa.2015.02.007>.
- [26] N. Pröllochs, S. Feuerriegel, D. Neumann, Statistical inferences for polarity identification in natural language, *PLoS One* 13 (12) (2018) 1–21, <http://dx.doi.org/10.1371/journal.pone.0209323>.
- [27] S. Pandit, C.E. Wasley, T. Zach, Information externalities along the supply chain: The economic determinants of suppliers' stock price reaction to their customers' earnings announcements, *Contemp. Account. Res.* 28 (4) (2011) 1304–1343, <http://dx.doi.org/10.1111/j.1911-3846.2011.01092.x>.
- [28] Y. Chen, Z. Wei, X. Huang, Incorporating corporation relationship via graph convolutional neural networks for stock price prediction, in: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18, Association for Computing Machinery, 2018*, pp. 1655–1658, <http://dx.doi.org/10.1145/3269206.3269269>.
- [29] S. He, H. Rui, A.B. Whinston, Social media strategies in product-harm crises, *Inf. Syst. Res.* 29 (2) (2018) 362–380, <http://dx.doi.org/10.1287/isre.2017.0707>.
- [30] T. Yu, M. Sengul, R.H. Lester, Misery Loves company: The spread of negative impacts resulting from an organizational crisis, *Acad. Manage. Rev.* 33 (2) (2008) 452–472, <http://dx.doi.org/10.5465/amr.2008.31193499>.
- [31] S.B. Smart, C.J. Zutter, *Fundamentals of Investing, fourteenth ed*, Pearson, London, United Kingdom, 2019.
- [32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach, 2019, <http://dx.doi.org/10.48550/ARXIV.1907.11692>.
- [33] Z. Zhao, H. Chen, J. Zhang, X. Zhao, T. Liu, W. Lu, X. Chen, H. Deng, Q. Ju, X. Du, UER: An open-source toolkit for pre-training models, 2019, <http://dx.doi.org/10.48550/ARXIV.1909.05658>.
- [34] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 2017*, pp. 24–26, <http://dx.doi.org/10.48550/ARXIV.1609.02907>.
- [35] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Netw.* 18 (5–6) (2005) 602–610, <http://dx.doi.org/10.1016/j.neunet.2005.06.042>.
- [36] S.A. Sharpe, Financial market imperfections, firm leverage, and the cyclicity of employment, *Am. Econ. Rev.* 84 (4) (1994) 1060–1074.
- [37] E.F. Fama, The behavior of stock-market prices, *J. Bus.* 38 (1) (1965) 34–105.
- [38] Y. Shi, K.Y. Ho, News sentiment and states of stock return volatility: Evidence from long memory and discrete choice models, *Financ. Res. Lett.* 38 (2021) 101446, <http://dx.doi.org/10.1016/j.frl.2020.101446>.
- [39] N. Strauss, R. Vliegthart, P. Verhoeven, Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands, *Publ. Relat. Rev.* 42 (4) (2016) 548–555, <http://dx.doi.org/10.1016/j.pubrev.2016.03.010>.
- [40] J. Lei, N. Dawar, J. Lemmink, Negative spillover in brand portfolios: Exploring the antecedents of asymmetric effects, *J. Mark.* 72 (3) (2008) 111–123, <http://dx.doi.org/10.1509/JMKG.72.3.111>.
- [41] B.W. Jacobs, V.R. Singhal, Shareholder value effects of the Volkswagen emissions scandal on the automotive ecosystem, *Prod. Oper. Manag.* 29 (10) (2020) 2230–2251, <http://dx.doi.org/10.1111/poms.13228>.
- [42] D.Y. Joe, F.D. Oh, Spillover effects within business groups: The case of Korean chaebols, *Manage. Sci.* 64 (3) (2018) 1396–1412, <http://dx.doi.org/10.1287/mnsc.2016.2596>.
- [43] P. Zou, G. Li, How emerging market investors' value competitors' customer equity: Brand crisis spillover in China, *J. Bus. Res.* 69 (9) (2016) 3765–3771, <http://dx.doi.org/10.1016/j.jbusres.2015.12.068>.
- [44] F.Z. Xing, E. Cambria, R.E. Welsch, Natural language based financial forecasting: a survey, *Artif. Intell. Rev.* 50 (1) (2018) 49–73, <http://dx.doi.org/10.1007/s10462-017-9588-9>.
- [45] A. Picasso, S. Merello, Y. Ma, L. Oneto, E. Cambria, Technical analysis and sentiment embeddings for market trend prediction, *Expert Syst. Appl.* 135 (2019) 60–70, <http://dx.doi.org/10.1016/j.eswa.2019.06.014>.
- [46] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [47] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, Deep learning based text classification: A comprehensive review, *ACM Comput. Surv.* 54 (3) (2021) 62, <http://dx.doi.org/10.1145/3439726>.
- [48] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2018, <http://dx.doi.org/10.48550/ARXIV.1810.04805>.
- [49] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, P.S. Yu, Improving stock market prediction via heterogeneous inf. fusion, *Knowl.-Based Syst.* 143 (2018) 236–247, <http://dx.doi.org/10.1016/j.knsys.2017.12.025>.
- [50] H. Huang, X. Liu, Y. Zhang, C. Feng, News-driven stock prediction via noisy equity state representation, *Neurocomputing* 470 (2022) 66–75, <http://dx.doi.org/10.1016/j.neucom.2021.10.092>.
- [51] A.F. Adoma, N.-M. Henry, W. Chen, Comparative analyses of Bert, Roberta, Distilbert, and Xlnet for text-based emotion recognition, in: *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP, 2020*, pp. 117–121, <http://dx.doi.org/10.1109/ICCWAMTIP51612.2020.9317379>.
- [52] Q. Liu, X. Geng, H. Huang, T. Qin, J. Lu, D. Jiang, MGRC: An end-to-end multigranularity reading comprehension model for question answering, *IEEE Trans. Neural Netw. Learn. Syst.* (2021) 1–12, <http://dx.doi.org/10.1109/TNNLS.2021.3107029>.
- [53] R. Mao, X. Li, Bridging towers of multitask learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification, in: *Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021*, pp. 13534–13542.
- [54] M. Ge, R. Mao, E. Cambria, Explainable metaphor identification inspired by conceptual metaphor theory, in: *Proceedings of the 36th AAAI Conference on Artificial Intelligence, 2022*, pp. 10681–10689.
- [55] K. He, L. Yao, J. Zhang, Y. Li, C. Li, Construction of genealogical knowledge graphs from obituaries: Multitask neural network extraction system, *J. Med. Internet Res.* 23 (8) (2021) e25670, <http://dx.doi.org/10.2196/25670>.
- [56] R. Mao, Q. Liu, K. He, W. Li, E. Cambria, The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection, *IEEE Trans. Affect. Comput.* (2022) 1–11, <http://dx.doi.org/10.1109/TAFFC.2022.3204972>.
- [57] A. Grover, J. Leskovec, Node2vec: Scalable feature learning for networks, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, USA, 2016*, pp. 855–864, <http://dx.doi.org/10.1145/2939672.2939754>.
- [58] Q. Lin, Y. Zhu, H. Lu, K. Shi, Z. Niu, Improving university faculty evaluations via multi-view knowledge graph, *Futur. Gener. Comput. Syst.* 117 (2021) 181–192, <http://dx.doi.org/10.1016/j.future.2020.11.021>.
- [59] Q. Lin, J. Liu, L. Zhang, Y. Pan, X. Hu, F. Xu, H. Zeng, Contrastive graph representations for logical formulas embedding, *IEEE Trans. Knowl. Data Eng.* (2021) <http://dx.doi.org/10.1109/TKDE.2021.3139333>, 1–1.
- [60] R. Mao, C. Lin, F. Guerin, End-to-end sequential metaphor identification inspired by linguistic theories, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019*, pp. 3888–3898, <http://dx.doi.org/10.18653/v1/P19-1378>.
- [61] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014*, pp. 1724–1734.
- [62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [63] A.P. Ratto, S. Merello, L. Oneto, Y. Ma, L. Malandri, E. Cambria, Ensemble of technical analysis and machine learning for market trend prediction, in: *2018 IEEE Symposium Series on Computational Intelligence, SSCI, 2018*, pp. 2090–2096, <http://dx.doi.org/10.1109/SSCI.2018.8628795>.
- [64] Y. Yang, M.C.S. Uy, A. Huang, FinBERT: A pretrained language model for financial communications, 2020, arXiv preprint [arXiv:2006.08097](https://arxiv.org/abs/2006.08097).
- [65] W. Chen, M. Jiang, W.-G. Zhang, Z. Chen, A novel graph convolutional feature based convolutional neural network for stock trend prediction, *Inform. Sci.* 556 (2021) 67–94, <http://dx.doi.org/10.1016/j.ins.2020.12.068>.
- [66] J. Li, M. Sun, Scalable term selection for text categorization, in: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007*, pp. 774–782.
- [67] J. Li, M. Sun, X. Zhang, A comparison and semi-quantitative analysis of words and character-bigrams as features in chinese text categorization, in: *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, 2006*, pp. 545–552.
- [68] S. Han, F. Ciravegna, Rumour detection on social media for crisis management, in: *ISCRAM, 2019*.
- [69] R. Mao, C. Lin, F. Guerin, Word embedding and WordNet based metaphor identification and interpretation, in: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018*, pp. 1222–1231.
- [70] R. Mao, X. Li, M. Ge, E. Cambria, MetaPro: A computational metaphor processing model for text pre-processing, *Inf. Fusion* 86–87 (2022) 30–43, <http://dx.doi.org/10.1016/j.inffus.2022.06.002>.