Research Article

Modeling and Prediction of Stock Price with Convolutional Neural Network Based on Blockchain Interactive Information

Wei Zhang [ID],[1] Ke-xin Tao,[1] Jun-feng Li [ID],[2,3] Yan-chun Zhu,[4] and Jing Li[4]

[1]*School of Information, Central University of Finance and Economics, Beijing 100081, China*
[2]*School of Finance, Central University of Finance and Economics, Beijing 100081, China*
[3]*Private Investment Fund Research Center, Central University of Finance and Economics, Beijing 100081, China*
[4]*Business School, Beijing Normal University, Beijing 100875, China*

Correspondence should be addressed to Jun-feng Li; lijf@cufe.edu.cn

The interactive information in blockchain architecture establishes an effective communication channel between users and enterprises, enabling them to communicate in a comprehensive and effective manner. Therefore, taking blockchain interactive information as the research object, this paper explores how the intervention of official information on investors affects the stock price movement and then makes predictions on stock prices according to the emotional tendency of interactive information. With the contextual information fusion, a sentiment computing model based on a convolutional neural network is established to extract and quantify the emotional features of blockchain interactive information. Combined with investors' emotional features, the stock price prediction model based on long short-term memory is proposed. The experiment results show that the accuracy of the model has been improved by incorporating the intervened emotional features, thereby proving that information clarification can have a positive effect on the stock price.

1. Introduction

Stock price modeling and prediction have been challenging objectives for researchers and speculators because of the noisy and nonstationary characteristics of samples [1]. Since Fama proposed the efficient market hypothesis in 1965, it has been generally accepted in the traditional financial field [2]. This hypothesis assumes that the stock market will be affected by market information and that stock prices will reflect the available information about the asset. When relevant information changes, the stock price will also change accordingly. Given that information related to the stock market is one of the most important factors that can cause stock market volatility, it should thus be included in the factors affecting stock market volatility [3].

With the advancement of blockchain technology [4–8], relevant information about the stock market has been widely stored and disseminated in blockchain architecture [9]. Therefore, numerous studies have been conducted on the influence of online media on the stock price movement based on the public opinion generated by online media, such as message boards and social media [10, 11]. Nguyen et al. used historical stock prices and Yahoo forum texts to predict stock prices [12]. Upon comparing the prediction results of models that only used stock historical trading data with those that added emotional features, they found that the results of models that added sentiment analysis were better than those of models that only used historical data. Li et al. used an econometric regression model to analyze the impact of social media on stock prices [13]. After analyzing tweets related to S&P 100 companies, they found that social media sentiment had an impact on stock returns. With the development of computer technology, deep learning models have achieved good results in analysis and prediction. Li et al. also proposed a tensor-based prediction model [14], featuring high-dimensional market information and its internal connections to study the stock trends under the influence of the media.

The above studies have all proven that investor sentiment has been widely used in stock price analysis. However, most of the information sources for this type of research focus on

news, message boards, or social media, and the dissemination and changes of information are not considered. In addition, only a few studies on the interaction and influence of multiple information sources have been conducted thus far.

With the development of Internet technology, new online media channels represented by interactive media are gradually emerging. Digital interactive media rely on certain official channels to build question and answer (Q&A) platforms for both investors and listed companies. The Q&A texts on such platforms include representative consumer sentiments and official real news, which inject information content into the stock market environment based on the official interactions with investors. In order to investigate the dissemination and interaction of information, digital interactive media with both "investor questions" and "official answers" as research objects are worthy of further study.

However, the impact analysis of the stock market based on digital interactive media faces two important challenges. First, the interactive nature of text information makes it difficult to analyze such information. Second, the transaction data of listed companies is continuous at the trading day level, while the disclosure time of Q&A information in digital interactive media is intermittent. Moreover, the times of the two data dimensions are heterogeneous.

Based on the above analysis, this paper designs a stock price prediction model for digital interactive media sentiment analysis in order to solve the challenges mentioned above. First, a method for extracting and quantifying the emotional features of digital interactive media information based on convolutional neural networks (CNN) is established. In this paper, the text to be classified and its contextual information are integrated, and the model is trained using artificially labeled datasets, effectively improving the accuracy of the interactive text analysis model. Second, a stock prediction model based on the long short-term memory (LSTM) method is established. This incorporates investor sentiment characteristics guided by official information into the proposed model and explores the depth and breadth of the influence of emotional factors. Experiment results reveal that the accuracy of the model incorporating the emotional characteristics of the intervention has been improved, thus proving that market information requires effective intervention and guidance.

The rest of this article is organized as follows. The second part introduces related works on the impact of news on stock price fluctuations. The third part constructs the emotional calculation model and stock price prediction model. The fourth part describes the algorithm. The fifth part analyzes the details of the experiment in detail. The sixth part presents the experimental results, and the seventh part summarizes the article.

2. Related Work

In the research on the impact of news on stock price fluctuations, the first widely used index to measure the influence of news is the number of stock-related news. Chan studied the relationship between the number of news and excess stock returns [15]. He extracted the quantitative characteristics of stock-related news articles and used these as an explanatory variable to regress excess stock returns. His results indicated that the investors' reaction to negative news was slower than their reaction to positive news; however, the impact of news on investors was related to its actual content. Thus, the use of news volume to summarize the impact of news on investors may have obvious limitations. Tetlock used the textual information of financial news to make stock price forecasts and proved the effectiveness of news for stock price forecasts [16]. In that research, the author used text mining methods, extracted features in the text that may affect market changes, and improved the feature extraction methods. The results indicated that the stock price prediction method that took financial news into account was more effective than other methods.

With the continuous expansion of the Internet industry and the advancement of web2.0 to web3.0 technologies, investors' sentiments toward the stock market can be shared and spread in an interactive manner through Internet platforms. Sentiments and events are integrated with a tensor for stock prediction [14, 17]. Das and Chen first proposed a method to extract investor sentiment from Yahoo message boards [18]. Their algorithm incorporates a variety of classification algorithms, accurately analyzes investor sentiment in message boards, and empirically proves that posts in the technology sector are closely related to stock trading volume and fluctuations.

In analyzing the impact of investor sentiment on securities prices, the common analysis models used are based on statistical models, econometric models, and machine learning models [19–21]. Among them, econometric models include linear regression models, logistic regression models, and autoregressive integrated moving average (ARIMA) models. The econometric model focuses on analyzing the causal relationship between stock prices and information. Antweiler and Frank studied the impact of more than 1.5 million messages posted on Yahoo [22]. Using a linear regression model to analyze the financial announcements of stock returns in 2000, they found that stock information helped predict market fluctuations. The statistical model uses univariate statistical models or bivariate statistical models to test the relationship between information sources and stock changes under different hypothesis tests. Based on a sample from January 1, 2009, to October 31, 2014, Li et al. developed an LSTM model that is based on investor sentiment extracted from internet stock message boards and market data to conduct out-of-sample forecasts for the open and closing prices of the CSI 300 index in the Chinese stock market. Their results showed that daily investor sentiment can adequately predict the subsequent trading day's market open prices, while the predictive information for the daily closing price was weak [23].

With the development of computer technology, machine learning models are increasingly used to study the relationship between stocks and public opinion. When using these models for information representation, it is possible to extract and merge multiple information sources and then apply them to models, such as neural networks, support vector machines, and Bayes classifiers. Chun et al. presented the

conceptual framework of an emotion-based stock prediction system (ESPS) focused on considering the multidimensional emotions of individual investors. To implement and evaluate the proposed ESPS, emotion indicators (EIs) are generated using emotion term frequency–inverse emotion document frequency. The stock price is predicted using a deep neural network (DNN) [24]. To compare the performance of the ESPS, sentiment analysis and a naïve method are employed. The experiment results showed that the accuracy of prediction using EIs was better than the accuracy of prediction using other methods. Jin et al. conduct a comparative study about the predictive performance of an artificial neural network, support vector regression (SVR), and autoregressive integrated moving average and select SVR to study the asymmetry effect of investor sentiment on different industry index predictions. The results show that the industries affected by investor sentiment are composed of young companies with high growth and high operative pressure [25].

When analyzing the impact of investor sentiment on the stock market in the above-mentioned research, the news sources selected can often only represent one party (the investor or the government) and cannot reflect the information interaction between both parties [26, 27]. In terms of methods applied, machine learning-based models are more commonly used in the literature [28–30]. Although machine learning model-based methods can fuse multiple information sources, the research on the mechanism of mutual influence between information is not sufficient. Therefore, taking digital interactive media as the research object, this paper uses machine learning models (e.g., neural networks) in order to study the interaction between market information and clarify the impact of information on the stock market.

3. Stock Price Prediction Model Based on Sentiment Analysis

Our work, as illustrated in Figure 1, includes media data acquisition (building text corpus), data preprocessing, sentiment analysis, trade data collection, and stock price prediction.

3.1. Media Data Acquisition and Representation. Given the limited amount of news text data obtained, because the content is concentrated in the financial field, the financial news part of CA8, a pretrained Chinese word vector dataset, is selected in the model. Among them, the data of pretrained word vector is trained from 6.2G financial news collected by the CA8 builder. The extracted context features are "Word + Character" mode, and the training method is skip-gram with negative sampling. The skip-gram model consists of three parts: input layer, hidden layer, and output layer. The procedure for each word is presented below.

Step 1. A vocabulary based on the training document is built, after which the words $\omega(t)$ are one-hot coded. The dimension of the transformed vector is denoted as $[1, |v|]$, where v represents the total number of words known to the system.

Step 2. The word $\omega(t)$ is passed from $|v|$ to the hidden layer.

Step 3. The hidden layer calculates the dot product between weight vector $\mathscr{W}[|v|, N]$ and $\omega(t)$, thus obtaining $H[1, k]$, where N represents the number of neurons in the hidden layer.

Step 4. Pass the output vector $H[1, k]$ of the hidden layer to the output layer. The output layer calculates the dot product operation between $H[1, N]$ and weight matrix $\mathscr{W}'[N, |v|]$, thus obtaining the output vector $U[|v|]$.

Step 5. The SoftMax regression classifier is used to calculate the probability of the output vector. The calculation method uses the equation

$$p\left(\omega_{c,j} = \omega_{O,c} \mid \omega_I\right) = \frac{\exp\left(u_{c,j}\right)}{\sum_{I-1}^{V}\exp\left(u_j\right)}, \tag{1}$$

where $\omega_{c,j}$ refers to the word predicted at the c position, $\omega_{O,c}$ refers to the word that actually appears at position c, ω_I refers to the word currently entered, and $U_{c,j}$ refers to the word of vector U at the prediction position c.

3.2. Sentiment Classification. The CNN model consists of an embedded layer, a convolutional layer, a pooling layer, and a full connection layer. In the embedding layer, we input a fixed length $n \times k$ matrix, and the sequence can be in the form of a pretrained word vector and a nonstatic word vector or multichannel. Here, n represents the sequence length, and k represents the word vector dimension corresponding to each word.

At the convolutional layer, a convolution kernel w window $x_{I:I+h-1}$ is used on the input sequence of $n \times k$ for the convolution operation, resulting in feature $c_i = f(w \times x_{i:i+h-1} + b)$. Only one-dimension convolution is needed, because the number of channels of text data is generally one, of which h represents the number of words in the window, w represents the weight matrix of $h \times k$ dimension, b represents the bias parameter, and $x_{i:i+h-1}$ represents the $h \times K$-sized window formed from the ith row to the $i + h - 1$ row of the input matrix.

In the pooling layer, the output of the convolutional layer is maximized by taking the maximum value of the feature in the neighborhood. Then, the dimension of the input data is reduced, and the output of the fixed length is obtained.

In the full connection layer, the features extracted by the convolution and pooling layer are inputted into the classifier for classification.

3.3. Stock Price Prediction. The index obtained from the quantitative text is combined with the fundamental index to predict the stock price based on the long short-term memory (LSTM) model. Figure 2 shows the procedure of the LSTM model.
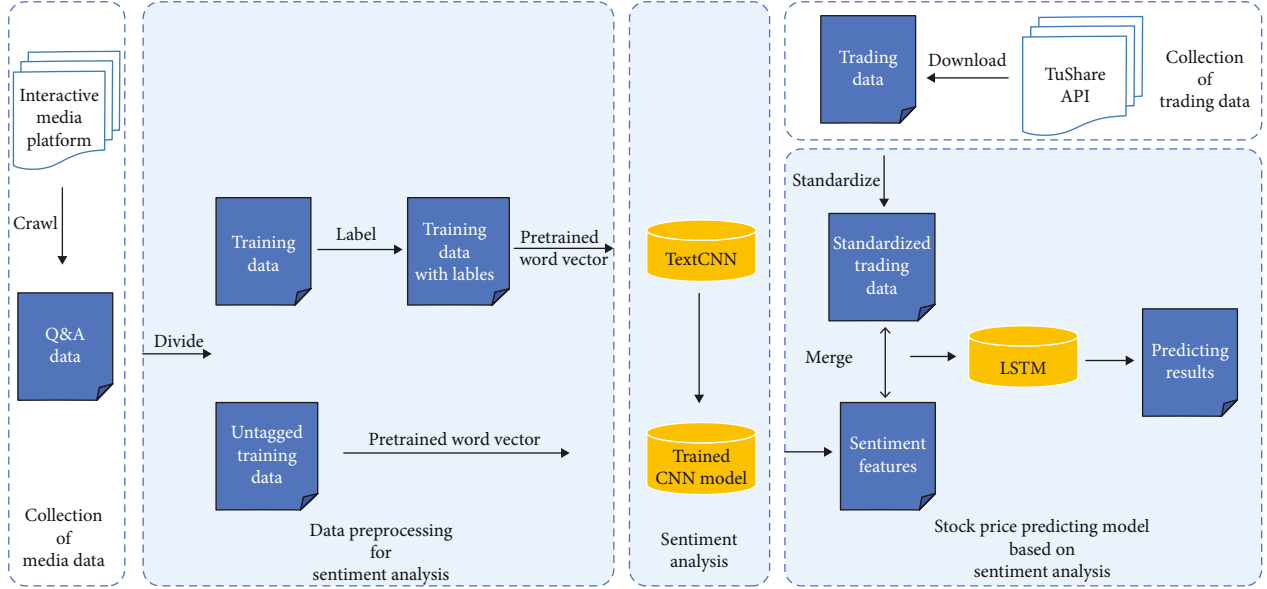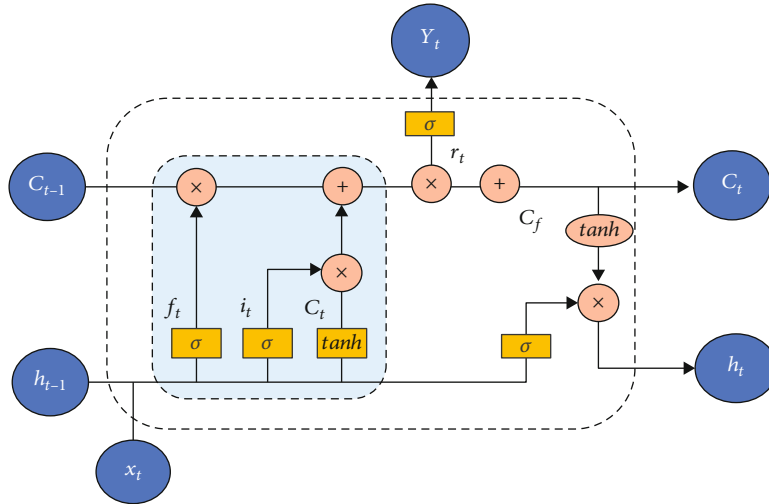
FIGURE 1: Research framework.

FIGURE 2: The flowchart of LSTM.

First, train z, z^i, and z^0 with the current input x^t and the output h^{t-1} of the previous state:

$$z^f = \sigma \left(w^f \times \frac{x^t}{h^{t-1}} \right),$$

$$z^i = \sigma \left(w^i \times \frac{x^t}{h^{t-1}} \right),$$

$$z^0 = \sigma \left(w^0 \times \frac{x^t}{h^{t-1}} \right), \tag{2}$$

$$z = \tanh \left(w \times \frac{x^t}{h^{t-1}} \right).$$

Second, we consider the components of the LSTM model.

In the internal forgetting stage of the model, the trained z^f is used as forgetting gating to determine which parts of c^{t-1} are remembered and forgotten in the previous state.

In the equation, c^{t-1} represents the state of the previous node. Its specific calculation method shall be given in the next stage of the model.

In the stage of selecting memory inside the model, the model selects memory input data x^t under the control of gated z^i. Then, the results of the forgetting stage and the selecting memory stage are added, and the state c^t can be obtained as follows:

$$c^t = z^f \odot c^{t-1} + z^i \odot z^*. \tag{3}$$

In the output phase inside the model, the model undergoes the scaling of c^0 to obtain h^t under the control of z^0, after which we determine the output y^t of the current state as follows:

$$h^t = z^0 \odot \tanh\left(c^t\right),$$
$$y^t = \sigma\left(w' \times h^t\right). \tag{4}$$

4. The Proposed Algorithm

First, this article cleans and quantifies the raw data collected. Then, the CNN is used to predict the emotional tendency of Q&A text, and the overall emotional value of the stock on that day is calculated according to the emotional tendency of all Q&A on that day. Finally, the emotional value and fundamental data are inputted, and the LSTM is used for training. The proposed algorithm is given as follows.

Step 1. Get the Q&A text and divide the words. By writing a web crawler, we are able to collect the raw Q&A data. Then, we merge the main Chinese stopword list and delete the stopwords in the text. Finally, the text data are segmented.

Step 2. Media representation. Here, we introduce the open source pretrained word vector CA8 dataset, transfer the pretrained word vector to the cleaned text data, and quantify the text into a word vector.

Step 3. Train the neural network model. We input the quantized annotated dataset into the CNN and then train the neural network. The validation set is used to check the accuracy of the model obtained by training. If the accuracy fails to pass the test, the parameters are adjusted to retrain the model.

Step 4. Calculate the emotional tendency of the day. The neural network model that has been tested for accuracy is used to classify the text emotions towards the training dataset. Based on the emotional orientation of each text of the day, we calculate the overall emotional orientation of the stock each day. Daily positive or negative tendencies are calculated by dividing the number of positive or negative texts in that part of the day by the total using the following equations:

$$\text{pos}_i = \frac{N_{\text{pos}}}{N_i},$$
$$\text{neg}_i = \frac{N_{\text{neg}}}{N_i}, \tag{5}$$

where pos_i and neg_i, respectively, represent the positive tendency and negative tendency of stock i, N_{pos} and N_{neg}, respectively, represent the positive and negative text quantity of stock i's questions or answers, and N_i represents the total text quantity of stock i's questions or answers.

Step 5. Analyze the text classification results. The emotional tendency variable of the day is integrated with the fundamental indicators, and the stock price is predicted by inputting the LSTM model. By comparing the prediction results of fused and unfused affective indicators, the accuracy of the fused affective indicators can be evaluated.

Table 1: Data description.

Web media	Panorama Interactive	Shanghai E Interactive	SSE Interactive
Coverage	Shanghai +Shenzhen	Shanghai	Shenzhen
Number of shares	3565	1468	1507
Total data amount	1.21 GB	184 MB	51.5 GB

5. Experiment Test

This study took A-share stocks from 2012 to early 2020 as the research object. The dataset consisted of two subdatasets: digital interactive media data and listed company transaction data. As shown in Table 1, the digital interactive media data included query data, time, user name, company name, and company code from each site. Digital interactive media data were collected from Panorama Interactive, Shanghai E Interactive, and SSE Interactive. Founded in 1999, Panorama Interactive is an interactive website for securities investors and provides an effective communication platform between A-share listed enterprises and investors. Shanghai E Interactive and SSE Interactive are digital interactive media platforms officially established by the Shanghai Stock Exchange and Shenzhen Stock Exchange, respectively.

Part of the raw data is shown in Table 2. As can be seen, since February 2020, COVID-19 has caused enterprises to operate under extraordinary situations, thus raising many questions. This situation means that investor sentiments have been significantly affected in the face of such *force majeure*. Moreover, to a certain extent, a company's official introduction regarding the start of the situation can ease investor apprehension.

Next, we filtered and cleaned the text information obtained by crawling according to the following rules:

Rule 1: delete duplicates in the Q&A for each stock.

Rule 2: delete the companies whose trading activities are suspended for more than 10 trading days along with their Q&A data within this period.

Rule 3: delete the companies exiting the market or suspend the listing during the period along with their Q&A data.

Rule 4: only the corresponding stock code, stock name, question text, question time, answer text, and answer time are retained for each Q&A text dataset. The other irrelevant attributes must be deleted.

The trading data of listed companies mainly include stock symbol, trading date, closing price, opening price, high price, low price, and turnover rate.

Stock trading data were collected from the TuShare Financial Big Data Center Interface. The data previews, including market value, turnover, current ratio, closing price, day high, day low, market return, and stock return, are shown in Table 2.

The trading data also excluded companies that were withdrawn from the market, suspended permanently, or suspended for more than 10 days.

After the data cleaning, the digital interactive media data from 2012 to 2015 were divided into training sets. Compared with the method of determining the sentiment of text by

TABLE 2: Trading data (part).

	ts_code	Trade_d	Open	High	Low	Close	Pre_c	Change	pct_chg	Vol	Amount
0	600000SH	20000104	24.98	25.78	24.75	25.57	24.75	0.82	3.31	44961	113946.78
1	600000SH	20000105	25.57	25.98	25.15	25.28	25.57	-0.29	-1.13	52528	134465.43
2	600000SH	20000106	25.18	26.30	25.05	25.99	25.28	0.71	2.81	62297	160059.80
3	600000SH	20000107	26.30	27.50	26.12	26.90	25.99	0.91	3.50	213553	575751.08
4	600000SH	20000110	27.00	27.85	26.71	27.25	26.90	0.35	1.30	165397	450463.52

TABLE 3: Data attributes of the training set.

	Attribute	Count	Ratio
Number of question words	Average	69	—
Questioner sentiment	Positive	9254	73.4%
	Negative	3352	26.6%
Number of answer words	Average	97	—
Respondent sentiment	Positive	7368	58.44%
	Negative	5238	41.56%
Difference between Q&A	Same sentiments	4792	38.01%
	Different sentiments	7814	61.99%
Dataset partitioning	Training set	10131	80%
	Validate set	2532	20%
Number of listed companies	Total	3565	100%
Q&A data quantity	Total	12606	100%

using a sentiment dictionary, the emotional tendency of using manual marking text can ensure the accuracy of the classification of the training set. Therefore, this study used the three-person voting method in the training set to mark the emotional tendency of each Q&A set.

After the dataset was marked, the training set and test set were further divided, and the test set was reserved for model testing. Some attributes of the tag datasets are shown in Table 3.

After the text dataset was cleaned and partitioned, word segmentation was performed.

First, because the text data came from web pages, we removed the tag languages, special symbols, and spaces from all the texts.

Second, given that Chinese text contained some stopwords that were not helpful for text analysis, these were removed in the text during the preprocessing phase.

The three commonly used stopword lists were integrated: the stopword list of Hit University, the Stop Word Library of Machine Intelligence Laboratory of Sichuan University, and the Baidu stopword list. Then, we use the Jieba Library to break up the text. Once the preprocessing was completed, the list for storing text data was previewed.

As the input sequence length of the TextCNN model was fixed, the quantized text sequence should be truncated and supplemented. According to the data exploration in the previous part, the average number of words in each query data was 69, while the average number of words in the combination of Q&As was 166. Therefore, the maximum sequence lengths in the model of Q&A and Q&A statements were set

TABLE 4: Data division.

	Train_ X	Train_ Y	Test_ X	Test_ Y
Ratio	70%	70%	30%	30%
Dimensions without sentiment	7	1	7	1
Dimensions with sentiment	11	1	11	1

to 69 and 166, respectively. For sequences longer than the mean, we truncated them to the mean length. For the sequences shorter than the mean value, all the vacant parts were filled with zeros.

After the regular length sequence was inputted into the embedded layer of the model, the layer outputted the word vector matrix of the sample length multiplied by the word vector dimension. Then, we transformed the good word vector matrix into the convolution layer. As the Chinese text is dominated by two- to four-character words, the sizes of the convolution kernel were set as 2, 4, and 5 to carry out the convolution operation on the word vector matrix. After setting the word vector dimension in the convolution kernel size, the data went through the maximum pooling and full connection layer to complete the classification task.

After completing the training of the CNN model, the text data from January 1, 2016, to February 2020 were then classified using the trained model, and the emotional tendency value of each stock on that day was also calculated.
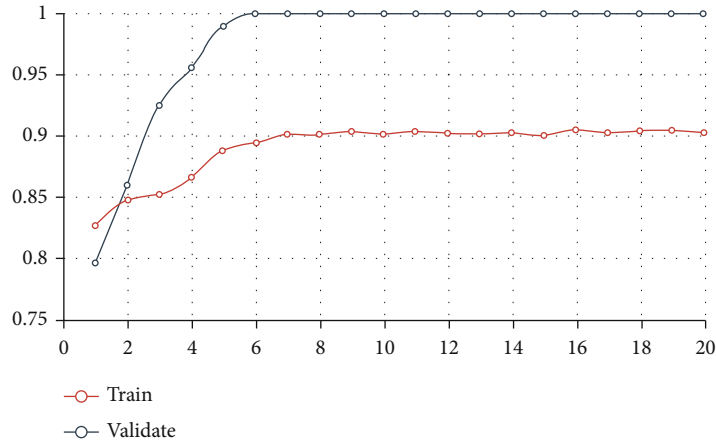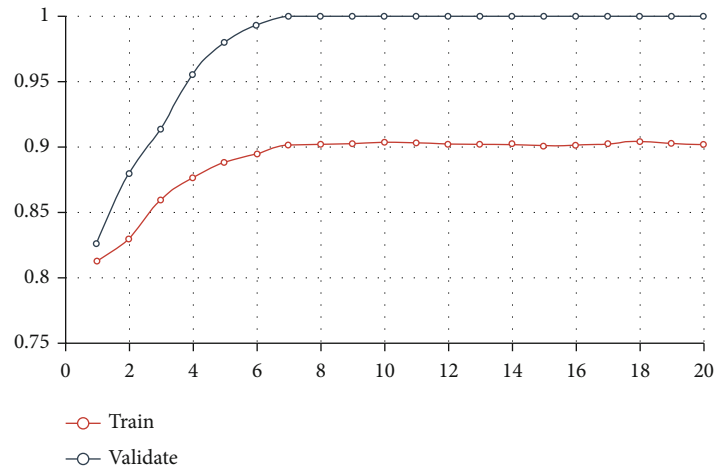
Figure 3: Accuracy of CNN for questions.

Figure 4: Accuracy of CNN for Q&As.

The final step was to process the stock trading data and interactive information data before using the LSTM model to predict the stock price. For stock trading data, it must be standardized. For the text data, because there may be no Q&A on a certain day, the vacant value of the sentiment indicator data should be filled to zero.

Once all the data processing was completed, the training set and test set were divided into the data. We forecast the closing price of the day with and without the inclusion of sentiment indicators (shown in Table 4).

6. Empirical Analysis

Results verified that the accuracy of the model tends to be stable during the fifth to seventh training rounds. As shown in Figures 3 and 4, the accuracy rates of the training set and the test set of the question-asking data are 97% and 90%, respectively, whereas those of the training set and the test set of the question-answering data are 97% and 89%, respectively. The error is within the acceptable range.

As there are many listed companies involved in the study, this paper analyzes the predicted results for PingAn Bank in

Table 5: Model evaluation.

Evaluations	Without sentiment	With sentiment
MSE	0.23003	0.12251
MRSE	0.47961	0.35003
MAE	0.36739	0.26076
R square	0.96752	0.98270

order to verify the results and conclusions of the model more intuitively, as shown in Table 5.

First of all, the determination coefficients of the two predictions are both higher than 95%, indicating that the selected stock trading data have a higher degree of influence on the prediction results. Moreover, the joint influence degree with the addition of emotional indicators is also higher. This indicates that the model has a good effect on stock price prediction whether or not affective indicators are added.

Second, compared with other evaluation indexes, the mean square error, root mean square error, and average absolute coefficient predicted by adding sentiment are smaller. Without the addition of affective indicator prediction, the mean square
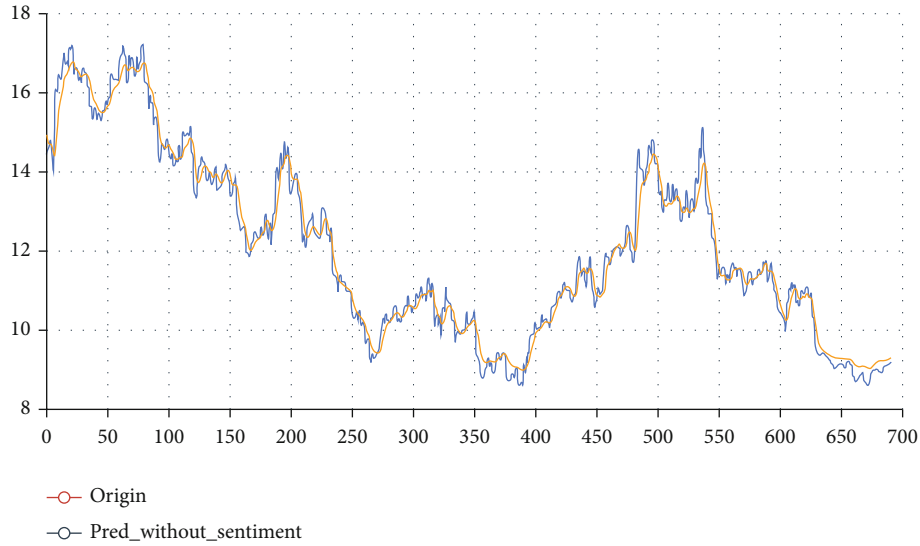
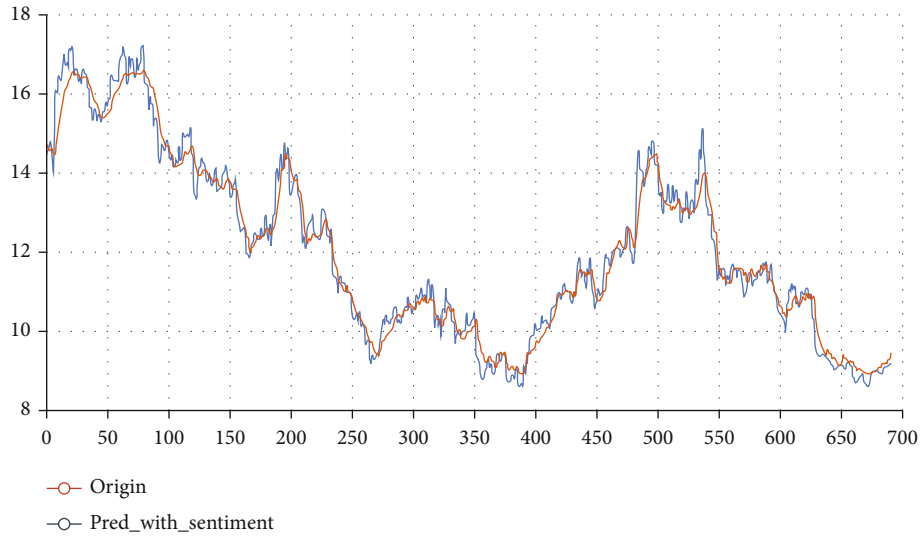Figure 5: Predictions without sentiment.

Figure 6: Predictions with sentiment.

error of the prediction model is 0.23, whereas the root mean square error reaches 0.48. These values indicate that without the addition of affective indicators, the error of the model is large, and the accuracy of the model is not high enough.

The mean square error and root mean square error of the model are 0.12 and 0.35, respectively. These two values are far lower than those of the model without emotion, thereby indicating that the accuracy of model prediction is improved by adding a sentiment index. Similarly, the average absolute error of the model using stock trading data for direct prediction is 0.37, whereas that of the model with sentiment indicators is 0.26. This also indicates that the addition of affective indicators improves the accuracy of model prediction.

Therefore, we have proven that the addition of affective indicators resulted in the significant improvement of the model compared with the condition wherein such indicators are not included.

In addition to the assessment indicators of the model, Figures 5 and 6 also reveal that the results predicted by adding affective indicators are better than those predicted by not adding such indicators.

7. Conclusion

Using several digital interactive media platforms, this research analyzes investor sentiments under the influence of official news and compares the prediction of short-term stock trends with and without sentiment analysis. In order to solve the above-mentioned issues, this research is carried out from two aspects.

First, we construct a text classification index for investor sentiment according to the characteristics of Q&A text. In this study, we used the three-person voting method to manually label the emotional orientation of interactive media text

data from 2010 to 2015. Then, we used this as the training set in training the CNN model. The trained model was used to classify the text data from 2016 to the first quarter of 2020 in order to extract the investor's emotional tendency.

Second, this article verifies the influence of investor sentiment on the accuracy of short-term stock price forecasts. The comparison can prove that investor sentiment and investor sentiment under official guidance can improve the accuracy of such price forecasts. Finally, the good index results prove that the experimental method is accurate and effective.

The findings of this research can guide market participants in crafting their decision-making plans. Specifically, this work provided important theoretical references and practical guidance for safeguarding investors' rights and interests, regulating the behavior of listed companies, and optimizing the stability of the securities market. For market participants, real-time market information must be obtained in a timely manner to avoid making irrational decisions due to information bias. In addition, listed companies must establish a complete rumor rejection mechanism to ensure that the information in the market is true and to maintain the stability of the market.

In the future, this research is aimed at expanding the topic from several aspects. First, we aim to increase the completeness of the data and add platforms that have emerged in recent years, such as the question of the secretary of the East Fortune. Second, in terms of text sentiment analysis, unbalanced classification analysis must be performed on text data, and the extraction of unbalanced text must be integrated to optimize the accuracy of the deep learning model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artificial Intelligence Review*, vol. 53, no. 7, pp. 3007–3057, 2019.

[2] E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, vol. 38, no. 1, pp. 34–105, 1965.

[3] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: a review and taxonomy of prediction techniques," *International Journal of Financial Studies*, vol. 7, no. 2, p. 26, 2019.

[4] H. Liang, J. Zou, K. Zuo, and K. Muhammad Junaid, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, article 106708, 2020.

[5] H. Zheng, W. Guo, and N. Xiong, "A kernel-based compressive sensing approach for mobile data gathering in wireless sensor network systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2315–2327, 2017.

[6] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, "Multimodal representation learning for recommendation in internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.

[7] H. Liang, D. Zou, Z. Li, K. Muhammad Junaid, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.

[8] Z. Liu, B. Hu, B. Huang, L. Lang, H. Guo, and Y. Zhao, "Decision optimization of low-carbon dual-channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, Article ID 2145951, 14 pages, 2020.

[9] H. Al-Shaibani, N. Lasla, and M. Abdallah, "Consortium blockchain-based decentralized stock exchange platform," *IEEE Access*, vol. 8, pp. 123711–123725, 2020.

[10] R. Nair and A. Bhagat, "An application of blockchain in stock market," in *Transforming Businesses with Bitcoin Mining and Blockchain Applications*, pp. 103–118, IGI Global Publisher, 2019.

[11] J. C. Yen and T. Wang, "Stock price relevance of voluntary disclosures about blockchain technology and cryptocurrencies," 2019, https://ssrn.com/abstract=3488483.

[12] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[13] T. Li, J. van Dalen, and P. J. van Rees, "More than just noise? Examining the information content of stock microblogs on financial markets," *Journal of Information Technology*, vol. 33, no. 1, pp. 50–69, 2018.

[14] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen, "A tensor-based information framework for predicting the stock market," *ACM Transactions on Information Systems*, vol. 34, no. 2, pp. 1–30, 2016.

[15] W. S. Chan, "Stock price reaction to news and no-news: drift and reversal after headlines," *Journal of Financial Economics*, vol. 70, no. 2, pp. 223–260, 2003.

[16] P. C. Tetlock, "Giving content to investor sentiment: the role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[17] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 3, pp. 9713–9729, 2020.

[18] S. R. Das and M. Y. Chen, "Yahoo! For Amazon: sentiment extraction from small talk on the web," *Management Science*, vol. 53, no. 9, pp. 1375–1388, 2007.

[19] X. Zhang, J. Shi, D. Wang, and B. Fang, "Exploiting investors social network for stock prediction in China's market," *Journal of Computational Science*, vol. 28, pp. 294–303, 2018.

[20] A. Derakhshan and H. Beigy, "Sentiment analysis on stock social media for stock price movement prediction," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 569–578, 2019.

[21] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, article 102212, 2020.

[22] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.

[23] Y. Li, H. Bu, J. Li, and J. Wu, "The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1541–1562, 2020.

[24] J. Chun, J. Ahn, Y. Kim, and S. Lee, "Using deep learning to develop a stock price prediction model based on individual investor emotions," *Journal of Behavioral Finance*, pp. 1–10, 2020.

[25] Z. Jin, K. Guo, Y. Sun, L. Lai, and Z. Liao, "The industrial asymmetry of the stock price prediction with investor sentiment: based on the comparison of predictive effects with SVR," *Journal of Forecasting*, vol. 39, no. 7, pp. 1166–1178, 2020.

[26] O. C. Sert, S. D. Şahin, T. Özyer, and R. Alhajj, "Analysis and prediction in sparse and high dimensional text data: the case of Dow Jones stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 545, p. 123752, 2020.

[27] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep learning for stock market prediction," *Entropy*, vol. 22, no. 8, p. 840, 2020.

[28] Y. Zhang, R. Zhu, Z. Chen, J. Gao, and D. Xia, "Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data," *European Journal of Operational Research*, 2020.

[29] H. Liang, A. Xian, M. Mao, P. Ni, and H. Wu, "A research on remote fracturing monitoring and decision-making method supporting smart city," *Sustainable Cities and Society*, vol. 62, p. 102414, 2020.

[30] Y. Zhou, D. Zhang, and N. Xiong, "Post-cloud computing paradigms: a survey and comparison," *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 714–732, 2017.