

Retraction

Retracted: The Influence and Prediction of Industry Asset Price Fluctuation Based on the LSTM Model and Investor Sentiment

Mathematical Problems in Engineering

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

Copyright © 2023 Mathematical Problems in Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

 W. Hu, H. Liu, X. Ma, and X. Bai, "The Influence and Prediction of Industry Asset Price Fluctuation Based on the LSTM Model and Investor Sentiment," *Mathematical Problems in Engineering*, vol. 2022, Article ID 1113023, 8 pages, 2022.



Research Article

The Influence and Prediction of Industry Asset Price Fluctuation Based on The LSTM Model and Investor Sentiment

Wenxiu Hu,¹ Huan Liu ^(b),¹ Xiaoqiang Ma,² and Xiong Bai²

¹School of Economic and Management, Xi'an University of Technology, Xi'an 710054, China ²School of Economics & Management Northwest University, Xi'an 710127, China

Correspondence should be addressed to Huan Liu; 1170511006@stu.xaut.edu.cn

Received 15 February 2022; Revised 28 February 2022; Accepted 3 March 2022; Published 21 March 2022

Academic Editor: Hangjun Che

Copyright © 2022 Wenxiu Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a real-world environment, not only can different levels of market expectations be triggered by factors such as macroeconomic policies, market operating trends, and current company developments have an impact on sector assets, but sector asset rises and falls are also influenced by a factor that cannot be ignored: market sentiment. Therefore, this paper uses LSTM to construct a forecasting model for industrial assets based on investor sentiment and public historical trading data of industry asset markets to determine future trends and obtains two conclusions: first, forecasting models incorporating investor sentiment have better forecasting effects than those without the incorporation of sentiment characteristics, indicating that the factor of investor sentiment should not be ignored when studying the problem of industry asset forecasting; secondly, investor sentiment quantified by different methods.

1. Introduction

As people's living standards rise, consumption is becoming a decreasing proportion of disposable income, replaced by investments, savings, and financial management [1]. Nowadays, investment and financial management are gradually coming into the public eye, and more and more people are concerned about and studying how to invest more effectively in order to allocate their disposable income rationally and maximize the return on their existing funds. The financial markets are developing at an unstoppable pace, and in the process of adapting to economic development, many investment options have emerged, most of which are based on the bond market, futures market, and the sector asset market [2]. Amongst these, the sector asset market has low investment thresholds and high liquidity, i.e., it can be quickly realized when investors need liquidity, making it the most common choice for retail investors to invest in the sector asset market to achieve a reasonable distribution of income. However, changes in the sector asset market are unpredictable and various factors, both imagined and unanticipated, may have an impact on sector assets to a greater or

lesser extent, and the market may not respond to various impacts to the same extent. The impact on asset volatility in the industry is explained in [3].

In the field of industry asset research, scholars hope to discover the overall operation and trends of industry assets through effective technical means, but after years of practice and research analysis, it is found that such ideas are difficult to realize in real life [4]. The efficient market hypothesis is considered a good theory to explain the changes in the development of the industry asset market, which believes that all effective information in the industry asset market is reflected in the historical prices of industry assets and that the future prices of industry assets are mainly affected by future information [5]. However, there are many factors that cause changes in future information that make forecasting future information very difficult, if not impossible, and therefore it is impossible to achieve forecasts of future industry assets using technical analysis under the efficient market hypothesis [6]. With the development of research and the emergence of other innovative theories, researchers have discovered that the role of investors' psychology, i.e., irrational factors, can have an impact on their financial behavior, and this is what behavioral finance studies. With the current poor transmission mechanism in the financial markets and the time lag in information, investors have a certain speculative mentality when investing in sector assets, usually wanting to get relatively high returns for less cost, which confirms the adage "speculation is speculation." In addition, sector assets influence investors' decisions, which in turn influence sector assets, forming a two-way cycle, similar to the bank's "run effect," but investors' decisions are often made with limited rationality mixed with some subjective judgment [7]. Family background, education, social background, etc. as the concept of behavioral finance has entered the public consciousness, some traditional theories are no longer applicable to the current financial markets. For example, while traditional asset pricing is theoretically defined based on the impact of macroeconomic policies, mesoindustry developments, and microfirm operating conditions on industry assets, there is usually a gap between asset prices in real markets and the theoretically expected prices, and behavioral finance has a unique understanding of the existence of this gap, which it sees as a result of investors' emotional decisions. The existence of such a gap is uniquely understood by behavioral finance as a result of the "rational constraints" of investors' emotions, which can create a systematic bias in the market as a whole and can also influence the next step in investor behavior [8].

In this paper, we consider the textual comments that reflect investors' sentiment positively and the proxy indicators that reflect investors' sentiment laterally, and after obtaining the textual data of investors' sentiment, we carry out sentiment identification through sentiment analysis to obtain the negative and positive sentiment classification, based on which we obtain the textual investors' sentiment index that represents investors' sentiment, and together with some of the proxy indicators, we achieve the construction of a comprehensive index of investors' sentiment through factor analysis. Finally, the index is used as the input variable of the LSTM model to build a prediction model for industry asset trends [9]. The validity of the model is tested through comparative analysis of different models and different industry backgrounds. This paper integrates the theories in the field of investor sentiment research with those in the field of industry asset forecasting research and investigates the degree of influence of investor sentiment on different styles of industries classified by CITIC style series subindices, which has certain theoretical significance for enriching industry asset forecasting models [10].

2. Related Work

The low threshold of the sector asset market and the readiness of investors to realize funds when they need them have made the equity market one of the more active markets in the financial investment sector [11]. These characteristics make the stock market a more active market in the financial investment sector. In order to invest wisely and profitably, investors focus their attention on fore-casting the trend of sector assets. The common methods

used to forecast sector assets are fundamental analysis and technical analysis. Among them, the fundamental analysis method is highly subjective and mainly involves some financial researchers analyzing the future ups and downs of industry assets based on public information in the market (e.g. national policies, industry developments, company financial reports, company announcements, etc.) and combining it with their own experience, which is a test of the researcher's professionalism and experience. For example, [12] proposed fundamental analysis methods such as the Delphi method, the principal probability method, and the cross-probability method to qualitatively forecast industry assets. Technical analysis methods are further divided into those based on statistical views and those based on data mining algorithms. Yang et al. [13] used GA-Elman neural networks to construct industry asset forecasting models, which not only achieve better forecasting results, but can also quickly calculate a large amount of data and save running time. An et al. [14] processed the mined news and financial text data and input them into a machine learning model to predict industry assets and analyzed various evaluation indicators to show that the prediction effect of the machine learning model based on text sentiment was significant.

Industry asset forecasting research has been the focus of financial researchers, and scholars have proposed many models for forecasting industry assets based on empirical and optimization studies, such as autoregressive moving average models, GRU models, and artificial neural networktype models. In order to optimize the models to achieve better forecasting results, scholars have conducted various studies: Ma et al. [15] compared BP neural networks, grey GM (1, 1) and their hybrid models, and concluded that hybrid models have better forecasting results; Thakkar and Chaudhari [16] mixed SVM models with GARCH models in order to improve the forecasting results of SVM models, and incorporated sentiment factors. In order to improve the forecasting effect of the SVM model, they mixed it with the GARCH model and incorporated the sentiment factor and investor attention into the forecasting model, and obtained a better forecasting effect. The model with the sentiment incorporated had a better effect than the model without the sentiment incorporated; Nasekin and Chen [17] chose the LSTM model to analyze the industry asset forecasting problem, and the results showed that the LSTM model could not only forecast industry assets better, but also run quickly when the data volume was large, and investors could draw on the forecasting results of the model. The results show that the LSTM model is not only good at predicting industry assets but also fast when the volume of data is large [18].

3. LSTM Model Based on Principal Component Analysis

3.1. Obtaining Training Data. The nine basic data items of the selected sector assets were obtained through the sector asset data stations of the major financial websites and the TuShare financial data interface package for Python: opening price, closing price, high price, low price, previous closing price, up/down amount, up/down range, volume, and turnover amount. These are shown in Table 1. The KDJ and MACD indicators calculated for the underlying data are used together as training data for the model.

The KDJ indicator is a sensitive and fast technical analysis indicator that uses the real volatility of the price fluctuations of sector assets to reflect the strength of the price change trend and can signal a buy or sell before the sector assets have risen or fallen. The highest and lowest prices that have occurred in a period, the last closing price of the period, and the proportional relationship between these three are used to calculate the unripe stochastic value of RSV on the last day of the period, and then the *K*, *D*, and *J* values are calculated based on the sliding average method [19].

The K value is the n-day moving average of the RSV and the K line, which is also known as the fast line, changes at a moderate rate among the 3 curves; the D value is the n-day moving average of the K value and the D line changes at the slowest rate among the 3 lines and is known as the slow line; the J value changes the fastest and is known as the ultra-fast or confirmation line as an aid to observing the buying and selling signals from the K and D lines. The three lines on the same coordinate make up the KDJ indicator, which reflects the trend of price fluctuations.

$$\begin{cases}
\text{RSV} = (C_n - L_n)/(H_n - L_n) \times 100, \\
K = \frac{2}{3}K_p + \frac{1}{3}\text{RSV}, \\
D = \frac{2}{3}D_p + \frac{1}{3}K, \\
J = 3 \times K - 2 \times D,
\end{cases}$$
(1)

where C_n is the closing price on day n, L_n is the lowest price on day n, H_n is the highest price on day n, and K_{p_n} and D_p is the previous day's K and D values, both replaced by 50 if not available.

MACD is also known as the moving average of divergence. The convergence and divergence of the fast and slow averages represent changes in market trends and are a common technical indicator for sector assets. The fast and slow-moving averages, EMA, are generally chosen as the 12day and 26-day moving averages, and their divergence, DIF, and the divergence's 9-day moving average, DEA, are calculated to give the MACD.

$$\begin{cases}
EMA_{(n)} = \frac{n-1}{n+1} \times PEMA_{(n)} + \frac{2}{n+1}, \\
DIF = EMA_{(12)} - EMA_{(26)}, \\
DEA = \frac{n-1}{n+1} \times PDEA_n + DIF, \\
MACD = (DIF - DEA) \times 2,
\end{cases}$$
(2)

where n is the number of days of moving average, C is the closing price of the day, and PEMA and PDEA are the EMA and DEA of the previous day.

3.2. Dimensionality Reduction of Data Using Principal Component Analysis. The principal component analysis is a method of transforming multiple interrelated raw data into a small number of linear combinations of two uncorrelated variables without changing the structure of the sample data by rotating the spatial coordinates. This reduces dimensionality and simplifies complex multidimensional problems by replacing more variables with fewer variables while maximizing the information in the original data.

To extract the principal components, the original data is first standardized, i.e., the mean of the corresponding variable is subtracted and then divided by the variance to eliminate the effect of differences in magnitudes.

$$Y_{ij}^* = \frac{x_{ij} - \overline{x}_j}{S_j}, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n.$$
(3)

The correlation coefficient matrix *R* is then calculated and the eigenvalues (i = 1, 2,...,n) are obtained by solving the characteristic equation $\lambda E \cdot R = 0$.

The eigenvalue is the variance of each principal component. It is used to describe the amount of information contained in the direction of the corresponding eigenvector, i.e., the magnitude of the eigenvalue directly reflects the influence of each principal component. The value of an eigenvalue divided by the sum of all eigenvalues gives the variance contribution of the eigenvector. $\lambda_i / \sum_{k=1}^n \lambda_k$ is the contribution of the i-th principal component. $\sum_{k=1}^{l} \lambda_k / \sum_{k=1}^{n} \lambda_k$ is the cumulative contribution of the first *i* principal components. According to the rules for the selection of the number of principal components, the selected principal components must all have eigenvalues greater than 1 and a cumulative contribution of at least a high percentage (usually greater than 85%). It is guaranteed that the selected principal components contain most of the information of the original data.

Finally, the principal component loadings are calculated and the principal component scores are obtained as new training data.

$$T_{ij} = p(z_i, x_j) = \sqrt{\lambda_i a_{ij}},$$

 $i = 1, 2, \cdots, m; j = 1, 2, \cdots, n.$
(4)

3.3. Prediction Using LSTM Models. The full form of LSTM is long term short term memory artificial neural network, a temporal recurrent neural network suitable for processing and predicting important events with relatively long intervals and delays in a time series [20]. It is a variant of the recurrent neural network, and the LSTM has an additional cellular structure in the algorithm that determines whether the information is useful or not than the recurrent neural network, as shown in Figure 1.

TABLE 1: Industry asset ba	se data.
----------------------------	----------

	Ts code	Trade date	Open	High	Low	Close	Pre close	Change
0	000001.SZ	20190314	12.33	12.62	12.55	12.64	10.20.05	0.184
1	000001.SZ	20190313	12.34	12.55	12.12	12.37	0.01	0.0826
2	000001.SZ	20190312	12.49	12.64	12.24	12.36	12.32	0.04
3999	000001.SZ	20020121	10.19	9.57	9.6	10.18	-0.59	-5.78



FIGURE 1: LSTM cell structure.

The LSTM has 3 gates in a cell: the forgetting gate, the input gate, and the output gate. Once a piece of data enters the LSTM's network, it is judged to be useful according to the rules, and those that match the algorithm's rules are left behind, while those that do not are forgotten through the forgetting gate. Only information that meets the algorithm's certification is left behind, while information that does not meet the rules is forgotten through the forgetting gate.

$$f_t = \sigma \Big(W_f \cdot \big[h_{t-1}, x_t \big] + b_f \Big). \tag{5}$$

The input gate then updates the cell state, first determining the value to be updated through the sigmoid layer, and the vector of candidate values created by the tanh layer, which are multiplied together to obtain the new candidate values.

$$\begin{cases} i_{t} = \sigma(W_{i} \cdot [h_{t-1}, x_{t}] + b_{i}), \\ C_{t} = \tanh(W_{C} \cdot [h_{t-1}, x_{t}] + b_{C}). \end{cases}$$
(6)

The old cell state is then multiplied by the discard information defined by the oblivion gate and the new candidate value is added to obtain the updated cell.

$$C_t = f_t \times C_{t-1} + i_t \times C_t. \tag{7}$$

Finally, based on the current cell state, the output component is determined by the sigmoid layer, which is multiplied by the tanh-processed cell state to obtain the value of the determined output

$$\begin{cases} o_t = \sigma \left(W_o[h_{t-1}, x_t] + b_o \right), \\ h_t = o_t \times \tanh(C_t). \end{cases}$$
(8)

In LSTM models, the model can choose what to keep and what to forget so that the model can analyze the data that is most relevant to the task. LSTM models can also learn a more abstract representation of the data so that the model learns more features of the data. These features allow LSTM models to be more effective in analyzing industry asset trends when applied to industry assets [21].

4. Analysis of Results

Based on the previous analysis, we use the CITIC style index and its constituent stocks as the research object when conducting the construction of the prediction models, and this section mainly presents the results based on the financial subindices. This section constructs LSTM models based on the four forecasting scenarios mentioned above, namely, the LSTM model based on textual investor sentiment index and historical data, the LSTM model based on proxy sentiment index and historical data, the LSTM model based on composite investor sentiment index and historical data, and the LSTM model based on K-line data. Figure 2 shows the prediction results of scenario 3 on the training set compared with the real results. By tuning and optimizing the model, it can be seen that the predicted and actual values basically have the same trend, although there are certain deviations from each other, but they can capture the up and down trends well,



FIGURE 2: Predicted and actual values for the training set of the integrated sentiment model.



FIGURE 3: Predicted and actual values for the test set of the integrated sentiment model.

indicating that the prediction results are more satisfactory and provide a strong reference for investors' buying and selling behavior [22]. Figure 3 shows the comparison between the predicted and actual values of scenario 3 on the test set.

Figures 4–6 show the predicted versus true results for scenarios 1, 2, and 4 on the training and test sets, respectively.

Table 2 shows the performance of each evaluation indicator for each scenario based on the LSTM model. Comparing the scenarios, it can be seen that the LSTM industry asset forecasting model incorporating a composite index of investor sentiment has the best forecasting performance, the LSTM industry asset forecasting model based on a sentiment proxy index has the second-best forecasting



FIGURE 4: Predicted and actual values for the training set of the text sentiment prediction model.



FIGURE 5: Predicted and actual values for the test set of the text sentiment prediction model.



FIGURE 6: Predicted and actual values for the training set of the proxy sentiment prediction model can be seen from Figures 4–6. Scenario 3, which incorporates a composite index of investor sentiment and historical data on industry assets, has the best predictions, with the smallest difference between the true and predicted values on both the test and training sets.

TABLE 2: LSTM model evaluation metrics based on each scenario.

Dradiction achoma	Scheme I		Scheme 2		Scheme 3		Scheme 4	
Prediction scheme	Training set	Test set						
MAE	0.045	0.057	0.041	0.051	0.029	0.039	0.095	0.112
MSE	0.0034	0.0041	0.0026	0.0032	0.00012	0.0001	0.015	0.0121
RMSE	0.0591	0.0648	0.053	0.0558	0.117	0.0142	0.115	0.1104
MAPE	0.52	0.42	0.41	0.48	0.26	0.39	0.4	0.5
Accuracy	0.72	0.73	0.82	0.75	0.89	0.81	0.61	0.59

performance, and the K-line forecasting model based on historical industry asset data only has the worst forecasting performance.

5. Conclusions

In this paper, when using technical methods to build LSTM stock forecasting models, traditional methods often result in poor generalization and poor forecasting due to a large number of input data variables selected, overlapping data information, and the large impact of outliers on training. To address such problems, we propose to use principal component analysis to reduce the dimensionality of the underlying data, then combine the KDJ and MACD as input data together with stock-related technical indicators and adjust the model according to the characteristics of the stock before making predictions. The experimental results show that the PCA-S-LSTM model can reduce the average error of prediction, and predict the closing price of Ping An Bank more accurately.

Data Availability

The experimental 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 regarding this work.

References

- H. Wang, L. Xue, W. Du et al., "The effect of online investor sentiment on stock movements: An LSTM approach," in *Proceedings of the International Conference on Intelligence Science*, pp. 1–14, Springer, Cham, 2021, February.
- [2] Z. Huang and Z. Li, "What reflects investor sentiment? Empirical evidence from China," *Data Science in Finance and Economics*, vol. 1, no. 3, pp. 235–252, 2021.
- [3] M. Baker and J. Wurgler, "Investor sentiment in the stock market," *The Journal of Economic Perspectives*, vol. 21, no. 2, pp. 129–152, 2007.
- [4] M. Li, W. Li, F. Wang, X. Jia, and G. Rui, "Applying BERT to analyze investor sentiment in stock market," *Neural Computing & Applications*, vol. 33, no. 10, pp. 4663–4676, 2021.
- [5] G. Wang, G. Yu, and X. Shen, "The effect of online environmental news on green industry stocks: The mediating role

of investor sentiment," *Physica A: Statistical Mechanics and Its Applications*, vol. 573, Article ID 125979, 2021.

- [6] Y. Li, S. Jiang, X. Li, and S. Wang, "The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach," *Energy Economics*, vol. 95, Article ID 105140, 2021.
- [7] 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.
- [8] Y. Ren, F. Liao, and Y. Gong, "Impact of news on the trend of stock price change: An analysis based on the deep bidirectiona LSTM model," *Procedia Computer Science*, vol. 174, pp. 128–140, 2020.
- [9] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, Article ID 106181, 2020.
- [10] L. Malandri, F. Z. Xing, C. Orsenigo, C. Vercellis, and E. Cambria, "Public mood-driven asset allocation: The importance of financial sentiment in portfolio management," *Cognitive Computation*, vol. 10, no. 6, pp. 1167–1176, 2018.
- [11] W. Gu, L. Zhang, H. Xi, and S. Zheng, "Stock prediction based on news text analysis," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 25, no. 5, pp. 581–591, 2021.
- [12] L. Yang, T. L. J. Ng, B. Smyth, and R. Dong, "Html: Hierarchical transformer-based multi-task learning for volatility prediction," in *Proceedings of the Web Conference 2020*, pp. 441–451, Taipei, Taiwan, 2020, April.
- [13] V. D. Ta, C. M. Liu, and D. A. Tadesse, "Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading," *Applied Sciences*, vol. 10, no. 2, p. 437, 2020.
- [14] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.
- [15] H. Ma, J. Ma, H. Wang, P. Li, and W. Du, "A comprehensive review of investor sentiment analysis in stock price forecasting," in *Proceedings of the 2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)*, pp. 264–268, IEEE, Xi'an, China, 2021, October.
- [16] A. Thakkar and K. Chaudhari, "Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions," *Information Fusion*, vol. 65, pp. 95–107, 2021.
- [17] S. Nasekin and C. Y. H. Chen, "Deep learning-based cryptocurrency sentiment construction," *Digital Finance*, vol. 2, no. 1, pp. 39–67, 2020.
- [18] C. H. Wu, C. C. Lu, Y. F. Ma, and R. S. Lu, "A new forecasting framework for bitcoin price with LSTM," in *Proceedings of the* 2018 IEEE International Conference on Data Mining

Workshops (ICDMW), pp. 168–175, IEEE, Singapore, 2018, November.

- [19] 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 ID 102212, 2020.
- [20] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
- [21] H. Liu, "A research on stock forecasting based on principal component LSTM model," in *Proceedings of the 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 684–688, Dalian, China, 2021.
- [22] H. Vachhani, M. S. Obiadat, A. Thakkar et al., "Machine learning based stock market analysis: A short survey," in Proceedings of the International Conference on Innovative Data Communication Technologies and Application, pp. 12– 26, Springer, Cham, 2019, October.