

Research Article

Investor Sentiment Combined with Multisource Information to Predict Stock Prices: An Analysis of China's A-Share Market

Xin Huang ¹ and Huilin Song ²

¹*School of Software, Jiangxi Normal University, Nanchang 330031, China*

²*School of International Economics and Trade, Jiangxi University of Finance and Economics, Nanchang 330013, China*

Correspondence should be addressed to Huilin Song; songhuilin@jxufe.edu.cn

Received 13 August 2021; Revised 20 October 2021; Accepted 22 November 2021; Published 7 December 2021

Academic Editor: Yi-Zhang Jiang

Copyright © 2021 Xin Huang and Huilin Song. 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.

Investor sentiment has been widely used in the research of the stock market, and how to accurately measure investor sentiment is still being explored. With the rise of social media, investor sentiment is no longer only influenced by macroeconomic data and news media, but also guided by We-Media and fragmented information. We take the data of China A-shares from January 2020 to December 2020 as the research object and propose a stock price prediction method that combines investor sentiment with multisource information. Firstly, the sentiment of macroeconomic data, brokerage research reports, news, and We-Media is calculated, respectively, and then the investor sentiment vector combining multisource information is obtained by the multilayer perceptron. Finally, the LSTM model is used to represent the stock time series characteristics. The results show that (1) the proposed algorithm is superior to the benchmark algorithm in terms of accuracy and F1-score, (2) investor sentiment vector can effectively measure the investment sentiment of stocks, and (3) compared with vector concatenation, multilayer perceptron can better represent investor sentiment.

1. Introduction

Behavioral finance, which is derived from finance, psychology, communication, and behavioral science, believes that the stock price is not only determined by the intrinsic value of an enterprise but is largely influenced by the psychology and behavior of investors [1]. The idea in behavioral finance is that investors in markets are not completely rational people. In the process of investment decision-making, investors often cannot make a correct and reasonable judgment due to factors such as emotional preference and cognitive bias. In other words, investor sentiment reflects investor behavior and affects the final investment decision to some extent. Researchers try to explain market behavior from the perspective of investors. To verify the effectiveness of investor sentiment, Akerlof and Shiller [2] found a close relationship between investor sentiment and stock price by studying the volatility of investor sentiment and stock price. You and Wu [3] used the “spiral of silence” theory in media

effect research of communication to study the impact of sentiment index on stock asset pricing from the perspective of media. Sentiment, as a factor affecting investors' psychological activities and then their behavior, has gradually become an important research issue in the task of stock price prediction.

Investor sentiment plays an important role in stock price forecasting. Song et al. [4] proposed a method for predicting stock excess returns that integrates research reports and investor sentiment, which can be verified in the Chinese A-share market to effectively improve the accuracy of the forecast. Li et al. [5] also conducted a similar study, and the difference is that the research object is the Hong Kong stock market. Polk and Sapienza [6] showed in their research that investor sentiment is similar to mispricing behavior in the stock market. Other views believe that investor sentiment is formed by the wrong estimation of asset value, which to a certain extent indicates the speculative propensity of investors [7]. Although the definition of investor sentiment

has not yet reached a unified concept, it can be seen from different definitions that investor sentiment is an expectation of future stock returns, and due to investors' irrational behavior and reasons that are not completely based on fundamental analysis, investors will have certain deviations in their expectations [8].

In the current study, measures of investor sentiment can be divided into three categories. The first is the direct measurement method, which uses indicators obtained from market surveys to directly replace investor sentiment. The second is the indirect measurement method, which uses single economic variables and combination variables as proxy variables to measure investor sentiment. The third type uses the machine learning method to extract online text information in social media and further construct an investor sentiment index to measure investor sentiment. The information explosion and fragmented nature of the age of big data make it inadequate to use any of these measures alone to measure investor sentiment. In our opinion, the measurement of investor sentiment should take into account four factors simultaneously: macroeconomic conditions, brokerage research reports, news, and We-Media information. Based on this, we put forward a kind of multisource information fusion method to predict the price of the shares of investor's emotion; first, the sentiment of macroeconomic data, securities research reports, news, and the media is calculated, fusion of multisource information is obtained by concatenation operation ISV (Investor Sentiment Vector, ISV), and finally LSTM model is used to represent the stock time series characteristics. The contribution of this paper is as follows:

- (1) An investor sentiment measurement method integrating multisource information is proposed
- (2) The positive role of investor sentiment in the stock prediction task is verified
- (3) A stock price prediction framework is proposed based on deep learning

The rest of this paper is organized as follows. Section 2 reviews investor sentiment measurement and its relationship with stock prices. Section 3 introduces our proposed method. Section 4 presents the experimental and details. Section 5 presents the experimental results and discussion. Section 6 gives our conclusions and directions for future work.

2. Related Works

With the continuous development of the Internet, the emergence of social media provides a new platform for users to search for information, express their feelings, and exchange opinions. Using social media indices as a proxy for investor sentiment has also become a convenient way to capture investor sentiment in the market. According to the Google search index, Da et al. [9] constructed the investor sentiment index through Google search keywords and found that the index could predict the short-term return and volatility of stocks. Meng et al. [10] use the Baidu search index to measure investor sentiment and find that investor

sentiment has a linkage mechanism with the stock market. Although quantitative indicators are feasible in reflecting investors' attention to the stock market, they are difficult to measure more in-depth investor sentiment information [11].

With the rise of big data, text mining, machine learning, and sentiment analysis technologies, researchers can more quickly and accurately extract valuable information from texts for the construction of investor sentiment [12]. Oliveira et al.'s [13] research shows that investor sentiment extracted from social media platforms has a certain impact on stock prices, and social media also provides a large number of data sources for the construction of investor sentiment. Bollen and Mao [14] analyzed and compared the predictive ability of traditional investor sentiment metrics and social media and found that sentiment indicators extracted from social media have a better predictive effect. Sentiment indicators obtained from text analysis of social media content have been widely used in stock market prediction, but there is no consistency in research conclusions [15]. Ma and Zhang [16] believed that the inconsistent conclusions were caused by the difference in sample data selection and the accuracy of investor sentiment measurement. At present, the research is no longer limited to judging whether investor sentiment can predict the stock market. How to extract valuable information from a large amount of data and apply it to the construction of investor index has become the focus of the research.

Pröllochs et al. [17] analyzed the information in financial news media and found that the sentiment of negative sentences in financial news is correlated with stock prices. In terms of information usefulness, Sprenger et al. [18] point out that many professional and amateur investors and analysts use Twitter to post news comments and opinions, usually more frequently than professional news media. In terms of the speed of information transmission, Sul et al. [19] believe that investors' emotions transmitted through social media are more likely to affect stock prices quickly, while investors' emotions that spread more slowly take longer to affect stock prices and are more likely to predict prices in the next few days. In addition to Twitter, StockTwits [20] and Yahoo Finance [21, 22] are also used to mine investor sentiment.

3. Method

We propose a stock price prediction method, including an investor sentiment module and a stock prediction module. The investor sentiment module separately calculates the four dimensions of macroeconomic status, broker report sentiment, news sentiment, and self-media sentiment through different methods and then obtains the ISV by MLP (multilayer perceptron) [23]. The stock prediction module consists of an LSTM [24], where the first input of LSTM is investor sentiment, and the subsequent input is the stock price. The method flow is shown in Figure 1.

3.1. Investor Sentiment Vector

3.1.1. Macroeconomic Status. MS (macroeconomic status) includes market status and economic status. For the market status, we select five indicators to measure the transaction

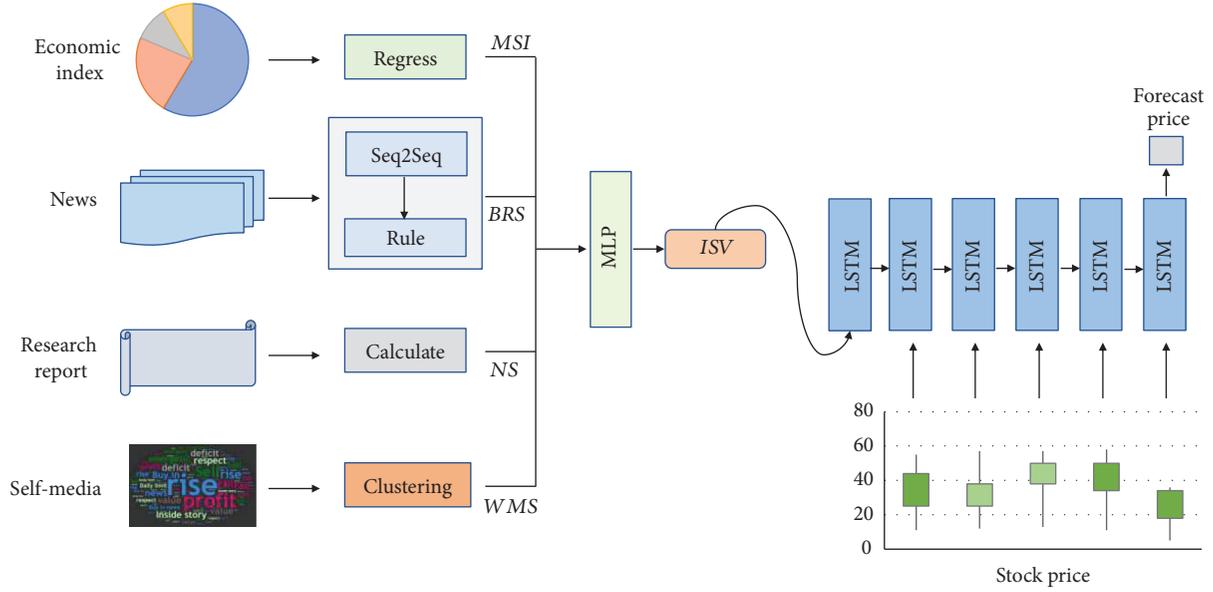


FIGURE 1: A stock price forecasting framework incorporating investor sentiment based on multiple-source information.

volume (VOLUME), the number of new investor accounts (NEWIN), the consumer confidence index (CCI), the closed-end fund discount rate (FUND), and the market turnover rate (HS_TVR). The economic status is measured by four indicators: the resident consumption index (CPI), the amount of new credit (IC), the economic growth rate (GDP), and the money supply (M2). The macroeconomic status measurement can be divided into two steps.

The first step is to calculate the preliminary sentiment index: $Sentiment_1$. Specifically, first standardize the market state indicators, then perform principal component factor analysis on the indicators, and select the three principal components with the highest variance explanation as weights; finally, the factor load, that is, the coefficient of $Sentiment_1$, is obtained after weighted average.

$$Sentiment^1 = a * FUND + b * NEWIN + c * VOLUME + d * HS_{TVR} + e * CCI. \quad (1)$$

The second step is to control the influence of economic state indicators and perform regression analysis on $Sentiment^1_m$, as shown in formula (2). The residual value ϵ_m is the measurement index of macrosentiment (MS).

$$Sentiment^1_m = \alpha_0 + \alpha_1 CPI + \alpha_2 IC + \alpha_3 GDP + \alpha_4 M_2 + \epsilon_m. \quad (2)$$

Among them, m is the month, α_0 is a constant, and $\alpha_1 \sim \alpha_4$ are the regression coefficients to be estimated.

the attention index $RRA_{k,t}$ is constructed by the ratio of the absolute number of reports of stock k on that day to the total number of all reports of stock A-share market on that day. The calculation method is shown as follows:

$$RRA_{k,t} = \frac{N_{k,t}}{N_{A,t}}, \quad (4)$$

where $N_{k,t}$ is the total number of reports of stock k on day t and $N_{A,t}$ is the total number of reports of all stocks in the A-share market on day t .

3.1.2. Brokerage Report Sentiment. Referring to the method proposed in the literature [4], we first split the brokerage research report into attention $RRA_{k,t}$ and rating sentiment $RRRS_{k,t}$ and then take the product of the two as the BRS (brokerage report sentiment), as shown the following equation:

$$BRS_{k,t} = RRA_{k,t} * RRRS_{k,t}. \quad (3)$$

Stock k report rating sentiment ($RRRS_{k,t}$) comprehensively considers the two factors of base rating and rating change. Specifically, the assignment of base rating and rating change is shown in Table 1, and the calculation method is as follows:

$$RRRS_{k,t} = \frac{\sum_1^{N_{k,t}} R_{k,t} \cdot C_{k,t}}{N_{k,t}}, \quad (5)$$

The difference is that attention in this paper is calculated daily (in literature [4], it is calculated monthly). Specifically,

where $R_{k,t}$ is the base rating and $C_{k,t}$ is the rating change. When there are multiple ratings for an individual stock in a single day, the research report ratings are averaged.

TABLE 1: Discrete assignment of base ratings and rating changes.

Variable	Assignment	Variable	Assignment	
Base rating	Strong buy	Rating Change	Up	
	Buy			1.2
	Long-term buy			1.7
	Overweight		1.6	First
	Buy with caution		1.5	
	Highly recommended		1.4	Maintenance
	Recommend		1.3	
	Cautious overweight		1.2	
	Hold		1.1	Down
	Neutral		1.0	
	Recommended with caution		0.9	
	Wait and see		0.8	
	Underweight		0.7	
Sell	0.6	0.9		

3.1.3. *News Sentiment.* News on the Web is a long text, and headlines alone cannot accurately and completely express the text. Therefore, we first generate a summary of the news, obtain the accurate intention of the text, and then calculate the NS (news sentiment).

(1) *News Summary Generation.* We use the architecture of Seq2Seq [25] to generate the summary, where the encoder takes a sequence as input, encodes the information in the sequence as a semantic vector, and then outputs the summary text through the decoder. The model is shown in Figure 2.

The encoder is bidirectional long short-term memory (Bi-LSTM) network. The input news is represented as $x = \{x_1, x_2, x_3, \dots, x_N\}$, and we encode x into hidden state vectors with

$$h = \text{BiLSTM}(x). \quad (6)$$

$h = \{h_1, h_2, h_3, \dots, h_N\}$. Specially, h_N is the result of the merger of two-way last hidden states. The decoder part uses LSTM, whose initial state s_0 is the output h_N of the encoder. On the step t , the decoder receives the previous decoder state s_{t-1} and the previously generated token y_{t-1} , and the decoder current state s_t is calculated as follows:

$$s_t = \text{LSTM}(s_{t-1}, y_{t-1}). \quad (7)$$

This method only uses s_t to connect the encoder and decoder, so the encoder needs to compress the entire sequence information into a fixed-length vector, which is limited. As the length of the input sequence increases, the information entered first is diluted by the information entered later. For better decoding, we use the attention mechanism [26] to instruct the decoder to generate the next word through the probability distribution on the source word. The attention distribution a^t can be calculated by s_t and h_i :

$$\begin{aligned} e_i^t &= v^T \tanh(W_h h_i + W_s s_t), \\ a^t &= \text{soft max}(e^t), \end{aligned} \quad (8)$$

where v, W_h, W_s are learnable parameters and a^t computes the context vector h_t^* :

$$h_t^* = \sum_i a_i^t h_i. \quad (9)$$

h_t^* contains decoding information, and we finally get the probability distribution P_{vocab} of the output words through h_t^* :

$$P_{\text{vocab}}(y_t | x, y_{<t}) = \text{soft max}(V' \tanh(V[s_t; h_t^*])), \quad (10)$$

where V' and V are learnable parameters.

(2) *News Sentiment Computing Based on Rules.* Referring to the study of Qi [27], we construct relevant semantic rules to dig out the real emotions of semantic words in different contexts. Specifically, according to the number of semantic words, the text is divided into multiple clusters, and the emotional value of the text is the sum of the semantic values of the clusters. The calculation formula is as follows:

$$U_a = \prod L_s \times S, \quad (11)$$

$$U_b = \left(\prod L_n \times (0.1 \times C_n - 1)^{c_n} + \prod L_s \right) \times S, \quad (12)$$

where U_a is the sentiment value of the cluster without negative words; U_b is the sentiment value of the cluster with negative words; S is the semantic value of the emotional word; L_s is the degree value of the degree adverb that modifies semantic words; L_n is the degree value of the degree adverb that modifies the negative word; C_n is the number of negative words. Then, the average of all the news semantic of the day was calculated, namely, news semantic (NS).

$$\text{NS} = \sum_1^n U_a + \sum_1^n U_b. \quad (13)$$

3.1.4. *We-Media Sentiment.* The data we chose to calculate WMS (We-Media sentiment) came from the stock BBS (Bulletin Board System). Retail investors communicate in

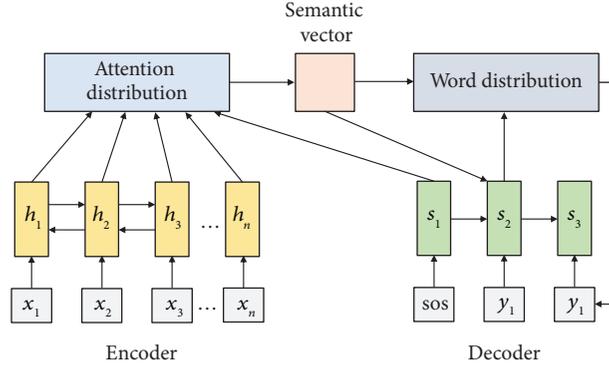


FIGURE 2: Abstract generation of a news text based on Seq2Seq.

the form of posts and replies, and the information they publish is usually short. After analysis, we believe that a post, including the post information and the reply information, represents the investment sentiment, so we combine the short text of the post and reply into a long text and calculate the sentiment of the long text. Refer to formulas (11)–(13) for WMS calculation.

3.1.5. Multisource Information Fusion. We use a multilayer perceptron [23] to fuse the output of four different emotion vectors and then use the aggregate vector as the first input of the stock prediction LSTM. ISV calculation is as follows:

$$ISV = v^T \sigma(W \cdot [MS, BRS, NS, WMS] + b). \quad (14)$$

3.2. Stock Price Prediction Model Based on LSTM. LSTM takes the output of the encoder as the input of $t = -1$ time step to guide the prediction of subsequent stock price, and the output m_{t-1} of LSTM at $t - 1$ time step is the input of t time step. In the training phase, for ISV and stock price $S = (S_0, \dots, S_N)$, the probability formula for predicting the stock price of the next trading day is as follows:

$$\begin{aligned} x_{-1} &= ISV, \\ x_t &= S_t, \quad t \in \{0, \dots, N - 1\}, \\ p_{t+1} &= LSTM(x_t), \quad t \in \{0, \dots, N - 1\}, \end{aligned} \quad (15)$$

where x_{-1} and x_t , respectively, represent the inputs of LSTM in time steps -1 and t , S_0 represents the closing price of stocks at the beginning trading day, and S_N represents the closing price of stocks in the end trading day. The loss function formula of the whole model is as follows:

$$L(I, S) = - \sum_{t=1}^N \log p_t(s_t), \quad (16)$$

where I represents ISV.

4. Experiment

4.1. Data and Preprocessing. The experimental data were selected from January 1, 2020, to December 31, 2020, excluding new shares and long-term suspended stocks. All the web texts are captured by scrapy crawler framework and preprocessed by word segmentation and removal of stop words.

- (1) Macroeconomic data: considering the lag of macroeconomic data, the data is selected from September 2019 to September 2020, and the data source is the WIND database (<https://www.wind.com.cn/NewSite/edb.html>).
- (2) Brokerage research reports: a total of 32724 reports on 2365 A-share companies released by 63 securities institutions are included. Data on the number of published reports, date of release, title, basic rating, and change of rating were obtained from East-money (<https://www.eastmoney.com/>). The brokerage rating confidence data are shown in Table 2.
- (3) News: we selected the news on the four authoritative websites of China Securities Network (<http://www.cs.com.cn/>), Sina Finance (<https://finance.sina.com.cn/>), Netease Finance (<https://money.163.com/>), and Securities Times (<http://www.stcn.com/>) as news data sources and captured the content including news headlines, release time, and news content. After sorting, a total of 96,532 news articles were obtained.
- (4) We-Media: the We-Media data came from Guba (<https://guba.eastmoney.com/>) and Xueqiu (<https://xueqiu.com/>), two BBS where Chinese retail investors discuss stocks. The number of “We-Media” texts after splicing is 183,938.

4.2. Baseline. In this paper, stock returns are used as the research object to predict whether the returns obtained by individual stocks in a certain period in the future are positive, negative, or flat. To ensure robustness, all data are

TABLE 2: Descriptive statistical results of brokerage rating sentiment.

Sentiment	Optimistic (RRRS > 1)	Neutral (RRRS = 1)	Pessimistic (RRRS < 1)
Number	31725	107	892
Proportion (%)	96.948	0.327	2.725

standardized according to the returns of the market in this period. To measure the advantages and disadvantages of the model from different perspectives, this paper selected SVM, LSTM model, RrmsNet [4], and SenticNet [5] as the benchmark methods for comparison with our work.

4.3. *Metrics.* In the experiment, accuracy and F1-score are adopted to evaluate the performances of each method. Let n denote the total number of samples and N_c denote the number of samples whose true label is c . These metrics are defined as follows:

$$\text{accuracy} = \sum_{c \in y} \frac{n_{cc}}{N}, P_c = \frac{n_{cc}}{n_{cc} + \sum_{p \neq c} n_{pc}}, R_c = \frac{n_{cc}}{n_{cc} + \sum_{p \neq c} n_{cp}}, F_1 = \sum_{c \in y} \frac{N_c}{N} \frac{2P_c R_c}{P_c + R_c}, \quad (17)$$

where $y = \{[1, 0, 0], [0, 1, 0], [0, 0, 1]\}$, N_c is the total number of samples in class c , N is the total number of samples, and n_{ij} is the number of samples whose true label is i and the predicted label is j . P_c and R_c are precision and recall.

5. Results

5.1. *Main Result.* Due to the timeliness of information dissemination, we choose the 5th, 15th, and 30th as the window period for experimental observations. Table 3 shows the detailed results of the comparative experiment. In general, the method we proposed has achieved the best results in both accuracy and F1-score, which shows that investor sentiment vectors combined with multisource information can effectively improve the performance of stock price prediction.

Figure 3 shows the accuracy and F1-score for different time windows. It can be seen from the figure that over time, the accuracy and F1-score of all methods have declined. Taking our method as an example, the accuracy rates on the 5th, 15th, and 30th days are 0.749, 0.693, and 0.668, respectively, and the F1-score are 0.723, 0.699, and 0.641, respectively. There are two reasons for this. First, all methods, no matter whether third-party information is included, are based on historical stock prices to predict future stock prices. As a result, the greater the time, the greater the uncertainty of the prediction. Second, investor sentiment based on the comprehensive calculation of different information is essentially an expression of information dissemination, so its influence on stock price prediction will weaken over time. This is consistent with the theory of information communication, that is, the longer the time, the weaker the influence of information.

5.2. *Ablation Experiments.* To better observe the influence of MSI, BRS, NS, and WMS on the performance of stock price prediction, an ablation experiment was carried out in this paper. The main idea is to remove one of the above indicators, respectively, to obtain four models Without_MSI, Without_BRS, Without_NS, and Without_WMS. Then, the accurate value and F1-score are compared with the Our_full

model. The larger the difference, the greater the influence and contribution. The ablation results are shown in Table 4. In general, excluding any index, the accuracy and F1-score are lower than the Our_full model, which indicates that the four indicators measuring investor sentiment have a positive impact on the stock price prediction. Among them, the Without_BRS model excluding the BRS indicator has the largest gap compared with the Our_full model, which shows that, among the four indicators, BRS has the greatest impact on the stock price prediction. There are two reasons. First, BRS, as a professional brokerage research report, is more easily recognized by shareholders. Second, compared with other sentiment indicators, brokerage research reports will directly give buy or sell recommendations, more direct.

Next, we remove the two indicators for further testing the performance of the model. Specifically, one is to get the Without_MSI_BRS model by removing MS and BRS simultaneously and the other is to get the Without_NS_WMS model by removing NS and WMS simultaneously. The reason for this is that MS and BRS information is more formal and comes from official or institutional sources, while NS and WMS information comes from news and comments on the Internet, which is more casual and free. The results show that the gap between the Without_MSI_BRS and Our_full model is larger, which shows that although the amount of news and commentary information from the Internet is greater, the price of stocks is more affected by official economic indexes and brokerages.

Finally, we compare the influence of different fusion methods of four indicators on the stock price prediction. Concatenation model splices four indicators into one-dimensional vectors as LSTM inputs, and the results show that its performance is inferior to Our_full model, indicating that the fusion method proposed in this paper is more suitable for stock price prediction tasks.

5.3. *Long-Term Stock Price Impact Analysis.* To examine whether there is a long-term effect on the impact of information on stock prices, we choose to conduct experiments on the 45th, 60th, and 90th window periods. The

TABLE 3: Stock returns forecast results.

Method	5 days		15 days		30 days	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
SVM	0.672	0.633	0.623	0.607	0.583	0.567
LSTM	0.710	0.684	0.634	0.618	0.599	0.578
RrmsNet [4]	0.727	0.718	0.656	0.643	0.618	0.604
SenticNet [5]	0.719	0.704	0.666	0.640	0.611	0.597
Ours	0.749	0.732	0.693	0.699	0.668	0.641

Bold values represent the best performance.

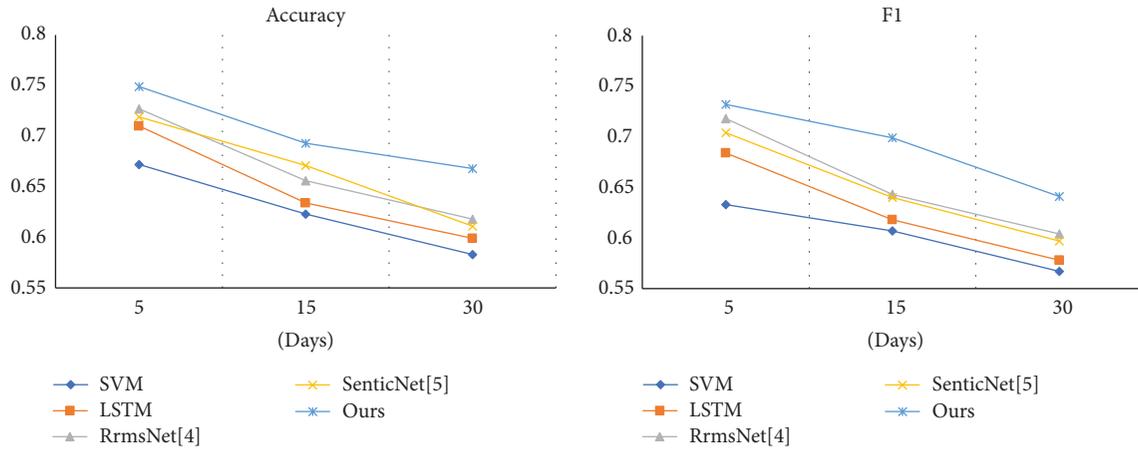


FIGURE 3: Comparison of results in different time windows.

TABLE 4: Results of ablation experiments.

Method	5 days		15 days		30 days	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Without_MSI	0.732	0.719	0.681	0.671	0.639	0.622
Without_BRS	0.688	0.713	0.646	0.670	0.610	0.591
Without_NS	0.721	0.719	0.685	0.665	0.638	0.624
Without_WMS	0.730	0.725	0.680	0.667	0.638	0.617
Without_MSI_BRS	0.701	0.698	0.640	0.642	0.630	0.609
Without_NS_WMS	0.732	0.723	0.677	0.668	0.633	0.621
Concatenation	0.736	0.725	0.687	0.672	0.642	0.626
Our_full	0.749	0.732	0.693	0.699	0.668	0.641

Bold values represent the best performance.

results are shown in Table 5. It can be seen from the table that the accuracy and F1-score of all methods are between 0.50 and 0.60, and it does not clearly show which method has better performance. Through the analysis of the case, it is found that, in the 45th, 60th, and 90th day time window after the information is released, the stock price performance of the predicted stocks is unstable and even presents a certain degree of randomness.

First, the basic methods of stock price prediction in other models except SVM are all based on the LSTM model, and the LSTM model itself has the problem that the input sequence is too long and the gradient disappears. Further, the investor sentiment vector aggregated by our proposed method is used as the initial input of the LSTM model ($t = 1$). As the observation window expands to 45 days or even 90 days, the influence of investor sentiment vectors on subsequent time steps gradually weakens.

Second, in the self-media era, the update cycle of market-related information is relatively short. Among the four information sources we selected, the MSI cycle is updated monthly, the BRS update cycle is about 20 days, the NS update cycle is about 2 weeks, and the WMS is updated daily, as shown in Table 6. In other words, the longest period of all information is 30 days, which means that new news will overwrite old news and affect investors' decision-making.

Finally, China's A-share market is a semiclosed and immature market. Investors' decision-making is often affected by the latest information, leading to frequent transactions and short holding periods. According to statistics, the average holding period of individual investor accounts in the A-share market is less than 20 trading days; even for investment institutions, the average holding period is about 30–40 trading days. The characteristics of the market determine the direction of the market.

TABLE 5: Stock return forecast results in the long term.

Method	45 days		60 days		90 days	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
SVM	0.587	0.533	0.567	0.564	0.544	0.512
LSTM	0.516	0.541	0.526	0.562	0.507	0.531
RrmsNet [4]	0.542	0.550	0.533	0.544	0.512	0.507
SenticNet [5]	0.568	0.549	0.589	0.551	0.527	0.514
Ours	0.548	0.539	0.542	0.577	0.525	0.530

Bold values represent the best performance.

TABLE 6: Stock return forecast results in the long term.

	MSI	BRS	NS	WMS
Update cycle	A month	Three weeks	Twelve days	One day

In summary, the investor sentiment vector calculated from the day of information collection has a limited impact on the stock price 45 days or even 90 days later, which is also the reason for the poor performance of the model.

6. Conclusion

The relationship between investor sentiment and the stock price has always been a hot research topic. In the era of big data, the channels for investors to obtain information have changed from research reports and news dominated by securities brokers to We-Media information. Multiple sources of information have brought new changes to measures of investor sentiment. Based on multisource information fusion, this paper proposes a new measurement method of investor sentiment and incorporates the new investor sentiment into the framework of stock price prediction. In the experiment with the data of China A-shares from January 2020 to December 2020, the results show that (1) investor sentiment is an important factor affecting stock price fluctuations, (2) among the different indicators of investor sentiment, brokerage report sentiment has the greatest impact on stock prices, and (3) multilayer perceptrons can better integrate emotional indicators.

Data Availability

All data in this paper come from public information on the Internet.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Natural Science Foundation of Jiangxi Province (grant no. 20212BAB202016) and the Science and Technology Research Project of Jiangxi Provincial Department of Education (grant no. GJJ200318).

References

- [1] D. B. Dai, Y. S. Lan, T. J. Fan, and M. Zhao, "Stock Forecast with Investors Sentiment by Text Mining and Machine Learning," *China Soft Science*, vol. 4, pp. 166–175, 2019.
- [2] G. A. Akerlof and R. J. Shiller, *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*, Princeton university press, Princeton, NJ, USA, 2010.
- [3] J. You and J. Wu, "Spiral of Silence: Media Sentiment and the Asset Mispricing," *Economic Research Journal*, vol. 47, 2020.
- [4] H. Song, D. Peng, and X. Huang, "Incorporating Research Reports and Market Sentiment for Stock Excess Return Prediction: A Case of Mainland China," *Scientific Programming*, vol. 2020, Article ID 8894757, 7 pages, 2020.
- [5] 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.
- [6] C. Polk and P. Sapienza, "The stock market and corporate investment: a test of catering theory," *Review of Financial Studies*, vol. 22, no. 1, pp. 187–217, 2008.
- [7] B. Debata, S. R. Dash, and J. Mahakud, "Monetary Policy and Liquidity: Does Investor Sentiment matter?" *IIMB Management Review*, vol. 33, 2021.
- [8] K. L. Fisher and M. Statman, "Investor sentiment and stock returns," *Financial Analysts Journal*, vol. 56, no. 2, pp. 16–23, 2000.
- [9] Z. Da, J. Engelberg, and P. Gao, "The sum of all FEARS investor sentiment and asset prices," *Review of Financial Studies*, vol. 28, no. 1, pp. 1–32, 2015.
- [10] X. J. Meng, X. L. Meng, and Y. Y. Hu, "Research on investor sentiment index based on text mining and Baidu index," *Macroeconomics*, vol. 38, no. 1, pp. 144–153, 2016.
- [11] H. Y. Yin and X. Y. Wu, "Predictive effect of high-frequency investor sentiment on the intraday stocks return," *China Industrial Economics*, vol. 8, no. 8, pp. 80–98, 2019.
- [12] H. Song, D. Peng, X. Huang, and J. Feng, "Research on weibo hotspot finding based on self-adaptive incremental clustering," *Journal of Shanghai Jiaotong University*, vol. 24, no. 3, pp. 364–371, 2019.

- [13] N. Oliveira, P. Cortez, and N. Areal, "The impact of micro-blogging data for stock market prediction: u," *Expert Systems with Applications*, vol. 73, pp. 125–144, 2017.
- [14] C. Zhao, Y. W. Ye, and M. H. Yao, "Stock volatility forecast based on financial text emotion," *Computer Science*, vol. 47, no. 5, pp. 87–91, 2020.
- [15] Z. G. Yi and N. Mao, "Research on the measurement of investor sentiment in Chinese stock market: the CICSI's construction," *Journal of Financial Research*, vol. 11, pp. 174–184, 2009.
- [16] R. W. Ma and N. Zhang, "The construction of investor sentiment index for China's stock market: based on the panel data of shanghai A share companies," *Journal of Central University of Finance & Economics*, vol. 7, pp. 42–49, 2015.
- [17] N. Pröllochs, S. Feuerriegel, and D. Neumann, "Negation scope detection in sentiment analysis: decision support for news-driven trading," *Decision Support Systems*, vol. 88, pp. 67–75, 2016.
- [18] T. O. Sprenger, A. Tumasjan, P. G. Sandner, and I. M. Welp, "Tweets and trades: the information content of stock microblogs," *European Financial Management*, vol. 20, no. 5, pp. 926–957, 2014.
- [19] H. K. Sul, A. R. Dennis, and L. I. Yuan, "Trading on twitter: using social media sentiment to predict stock returns," *Decision Sciences*, vol. 48, no. 3, pp. 454–488, 2017.
- [20] M. S. Checkley, D. A. Higón, and H. Alles, "The hasty wisdom of the mob: how market sentiment predicts stock market behavior," *Expert Systems with Applications*, vol. 77, pp. 256–263, 2017.
- [21] 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.
- [22] 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.
- [23] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and translate," 2014, <https://arxiv.org/abs/1409.0473>.
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.
- [26] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," 2015, <https://arxiv.org/abs/1509.00685>.
- [27] T. F. Qi and H. X. Jiang, "Exploring stock price trend using Seq2Seq based automatic text summarization sentiment mining," *Management Review*, vol. 33, no. 5, p. 257, 2021.