Research Article

Incorporating Research Reports and Market Sentiment for Stock Excess Return Prediction: A Case of Mainland China

Huilin Song,[1] Diyun Peng,[2] and Xin Huang [3]

[1]*School of Management, Nanchang University, Nanchang 330031, China*
[2]*School of Economics Management, Nanchang University, Nanchang 330031, China*
[3]*School of Software, Jiangxi Normal University, Nanchang 330031, China*

Correspondence should be addressed to Xin Huang; 1610466@tongji.edu.cn

The prediction of stock excess returns is an important research topic for quantitative trading, and stock price prediction based on machine learning is receiving more and more attention. This article takes the data of Chinese A-shares from July 2014 to September 2017 as the research object, and proposes a method of stock excess return forecasting that combines research reports and investor sentiment. The proposed method measures individual stocks released by analysts, separates the two indicators of research report attention and rating sentiment, calculates investor sentiment based on external market factors, and uses the LSTM model to represent the time series characteristics of stocks. The results show that (1) the accuracy and F1 evaluation indicators are used, and the proposed algorithm is better than the benchmark algorithm. (2) The performance of deep learning LSTM algorithm is better than traditional machine learning algorithm SVM. (3) Investor sentiment as the initial hidden state of the model can improve the accuracy of the algorithm. (4) The attention of the split research report takes the two indicators of investor sentiment and price as the input of the model, which can effectively improve the performance of the model.

1. Introduction

Stock price prediction is a method of predicting stock prices in the future based on stock price information at the past or current moment. Traditional quantitative investment methods are mostly based on experience in forecasting future stock prices. Such methods often have weak antirisk capabilities, poor long-term forecasting capabilities, and slow analysis speed and are not convenient for dissemination and promotion. Randomly appearing statistical and financial-based stock analysis methods belong to the traditional machine learning category. Most of them use autoregressive models, random fluctuation models, and Markov models to make predictions. Compared with empirical methods, this method is faster. It is fast and accurate, but the disadvantage is that it can process less information and cannot fully deal with many factors of external market data that cause stock price fluctuations. Thanks to the massive financial data provided by the continuous development of big data

technology, it is possible for artificial intelligence methods to enter the financial analysis field. Therefore, more and more researchers have begun to use machine learning or deep learning methods to analyze stock prices and make predictions. Related methods in the field of artificial intelligence have demonstrated incomparable excellent performance on large-scale datasets. This has been verified in areas such as images [1, 2] and text [3–5]. It is foreseeable that artificial intelligence-related methods can solve many problems in current stock price prediction models. Because of its policies, internal environment, and investor attributes, stocks have different rules in different markets. The stock market in mainland China belongs to an emerging capital market. The imperfect regulatory policies and the characteristics of most investors are retail investors, and making media reports to a great extent can affect the trend of stock prices. Ding and Sun's [6]research shows that in the Chinese A-share market, the behavior of ordinary investors in buying and selling stocks will be largely affected by research reports issued by

financial institutions. Compared with the news media reports, which focus on the occurrence of events and describe the original events, the research reports are more focused on financial and market-related attributes related to stock prices, with the purpose of predicting stock prices. At the same time, as an information publisher, securities analysts have a more professional industry background and richer information channels than ordinary financial news reporters, so for ordinary investors, direct and professional research reports are important reference objects for investment decisions.

In order to better predict the stock price of the Chinese mainland stock market, we propose a method that combines research reports and market sentiment to predict abnormal stock returns. The proposed method measures individual stocks released by analysts and splits the research report. Attention and rating sentiment, calculate investor sentiment based on external market factors, and use the LSTM model to represent the time series characteristics of the stock. We selected A-share data from July 1, 2014, to September 30, 2017, for experiments and compared different algorithms and different inputs. Based on the experimental results, we found that

(1) The accuracy of the proposed algorithm and the evaluation of F1 are better than the benchmark algorithm

(2) The performance of deep learning LSTM algorithm is better than traditional machine learning algorithm SVM

(3) Investor sentiment as the initial hidden state of the model can improve the accuracy of the algorithm

(4) The attention of the split research report and the two indicators of investor sentiment and price are used as input for the model, which can effectively improve the performance of the model

The rest of this article is organized as follows. Section 2 reviews the literature that separately introduced the impact of machine learning-based stock price predictions and research reports on stock prices. Section 3 introduces our proposed method. Section 4 presents the experimental design and details. Section 5 presents the experimental results and discussion. Section 6 gives our conclusions and directions for future work.

2. Related Work

2.1. Machine Learning-Based Stock Price Prediction. In the traditional machine learning field, Xiang [7]used an improved gradient boosting decision tree (GBDT) to predict stock prices. This model can mine the relevant features of the current stock series, but the GBDT model structure itself is not suitable for solving serial data problems like stocks.

Du et al. [8] used a Bayesian learning (BL) model to predict stock prices in the research. This model is actually similar to the autoregressive integrated moving average (ARIMA) model. Based on statistical knowledge, it learns the characteristics of the stock sequence. However, the BL

model itself is actually not suitable for sequence data. In the field of deep learning, Tsantekidis et al. [9] in their research proposed a stock price prediction model based on a CNN encoder. CNN is a very effective model for image input. In order to adapt it to sequence data, first use the encoder to encode the sequence data, and then use CNN for training. This method is very similar to the signal and system. In filtering theory, the sequence data can be regarded as a time signal, and the CNN can be regarded as a filter for convolution. Bao et al. [10] in combination with the long short-term memory (LSTM) of the autoencoder (AE) constructed a special algorithm based on the recurrent neural network (RNN). The neural unit structure makes it very suitable for processing sequence data such as stocks. This method even adds an autoencoder to encode the stock sequence through training and then uses the LSTM network for training. Based on the basic deep learning models, more studies have considered the basic characteristics of the stock market and incorporated them into the method. Zhang and Tan [11] used historical price data to predict the future return ranking of stocks through a new stock selection model based on deep neural networks. Li et al. [12]have established a system that uses deep learning architecture to improve feature representation and uses extreme learning machines to predict market impact. They concluded that the feature representation of deep learning together with extreme learning machines can provide better accuracy of market impact predictions. Li et al. [13] emotion vectors are obtained through sentiment analysis of news articles, and sentiment vectors are added to the LSTM model to predict stock prices. The experiments on the Hong Kong stock market have shown good performance.

2.2. Stock Price Prediction Based on Research Reports. Lee et al. [14] believe that after being affected by media sentiment, investors will form a subjective and objective comprehensive judgment on future capital flows and investment risks, which is called "investor sentiment." When investor sentiment is extremely optimistic or pessimistic, stock price volatility increases. At the same time, for the purpose of promotion, commission income, contracting customers or business, etc., the research report written by the securities company is not always neutral, and they will convey information with serious selective deviations to the market to meet the needs of investors, and such deviations are often optimistic. Using data from the "Abreast of the Market" column on the Wall Street Journal's website as a sample, Tetlock [15] constructed a media pessimism index and found that abnormally high or abnormally low media pessimism can cause temporary activity in market trading behavior. Hribar and McInnis [16]in their research found that when investor sentiment is high, analysts' optimism tends to be more obvious. The existence of optimism tends to distort stock prices and seriously affects investor decisions. Zhao et al. [17] proposed that in companies with high stock price synchronization, analysts' optimism tends to have a weaker impact on the accuracy of their subsequent earnings forecasts. Xu et al. [18] believe that optimism tends to lead to

high transaction volumes, but it also easily leads to negative news of listed companies not being disclosed in a timely manner, and the risk of future stock price crashes. Lu and Chen [19] found that the impact of extreme optimism and extreme pessimism on the stock price index is asymmetric, and short-term extreme pessimism has a negative relationship with the stock price index.

2.3. Task. The ups and downs of stocks determine the rate of return on stocks, and most stock prices fluctuate with changes in the stock market environment. When the stock market is in a "bull market," most stocks will rise, and when the stock market is in a "bear market," most stocks will follow the trend and fall. Simply predicting the rise or fall of a stock on the next trading day cannot objectively reflect the stock income. Therefore, this article uses stock excess returns as a research object to explore whether the excess returns obtained by individual stocks in a certain time interval in the future are positive, negative, or par. The calculation of excess returns for individual stocks is as follows:

$$AR_{k,t} = R_{k,t} - ER_{k,t}, \tag{1}$$

where $AR_{k,t}$ is the abnormal rate of return on day t of stock k, $R_{k,t}$ is the actual rate of return on day t of stock k, and $ER_{k,t}$ is its expectation yield (or expected normal return). There are many methods to calculate the expected normal rate of return. In order to exclude the part of the return that is related to market returns, this article uses the Malkiel and Fama [20] market model to measure. Therefore, the actual yield of individual stocks can be expressed as

$$R_{k,t} = \alpha_k + \beta_k R_{m,t} + \mu_{k,t}, \tag{2}$$

where $R_{m,t}$ is the market return rate on day t, $\mu_{k,t}$ is a random error term, and estimates of α_k and β_k obtain the values $\hat{\alpha}_k$ and $\hat{\beta}_k$. The models that measure expected normal returns are

$$ER_{k,t} = \hat{\alpha}_k + \hat{\beta}_k R_{m,t}. \tag{3}$$

Finally, calculate the cumulative abnormal return of stock k during the event window:

$$CAR = \sum AR_{k,t}. \tag{4}$$

Due to the high turnover rate of ordinary investors in China's A-share market and the average holding time is about one month, this article selects the excess returns from one trading day to the 5, 15, and 30 trading days as the forecast target.

3. Methods

3.1. Overview. In order to better understand the time series relationship between stock prices, our method uses LSTM networks [21] as the basic unit of the model, and based on this, a research report and market sentiment are quantified into the process of stock prices. Figure 1 shows the structure of our method model.
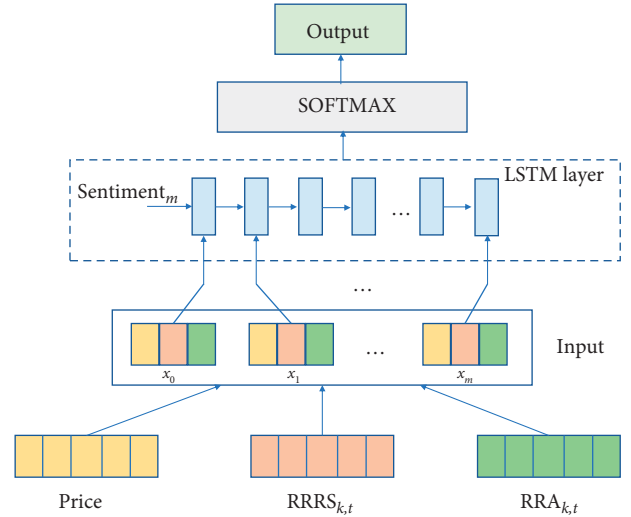

Figure 1: A stock price prediction model incorporating research reports and market sentiment.

First of all, in order to better indicate the status of a stock in the current stock market, we sort the stock price (Price), research report rating sentiment ($RRA_{k,t}$), and research report attention ($RRRS_{k,t}$) which is concatenated to get x_t, and the calculation method of x_t is shown in the following formula:

$$x_t = \left[\text{Price}; RRA_{k,t}; RRRS_{k,t} \right]. \tag{5}$$

Then, x_t is used as the input of LSTM, and the initial hidden state of LSTM is investor sentiment Sentiment_m. The reason for this is that relative to the positive volatility of the stock price, Sentiment_m is stable for a period of time and can be regarded as an indicator of market sentiment in the short term. Finally, the output of the LSTM is calculated by the SOFTMAX function to obtain the final output of the model. The calculation methods of LSTM, Sentiment_m, $RRA_{k,t}$, and $RRRS_{k,t}$ will be described in detail in the following sections.

3.2. Long Short-Term Memory Network. The LSTM proposed by Hochreiter and Schmidhuber [21] in 1997 can effectively deal with the long-term dependencies in the sequence, and its structure is shown in Figure 2.

The core of LSTM lies in its memory unit, and related information is transmitted backward through the memory unit. Theoretically, the memory unit can transfer information during the entire sequence propagation process so that the information at the previous time can be used to predict the output at the later time, so it can solve the short-term memory problem of the traditional recurrent neural network. In addition, during the backward transfer of information in the memory unit, the LSTM adds or deletes information in the memory unit through three gates. These gates can be seen as different neural networks, which can be trained to automatically learn what information to keep or forget. The process of LSTM processing information is as follows. First, the LSTM will use the "forget gate" to determine which information should be removed. The input is
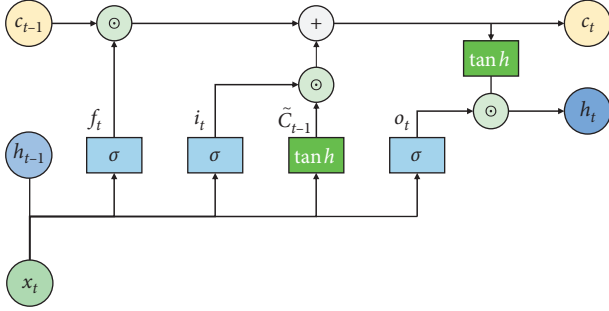
FIGURE 2: Long short-term memory network structure.

mapped between 0 and 1 by the Sigmoid function. The trend to "1" means to retain the information; otherwise, it means to forget information. The "input gate" is used to determine which information needs to be updated. The Sigmoid function is used to determine whether it needs to be retained. Then, the tan h function maps the input value to $[-1, 1]$, thereby generating a new memory unit state and adding it to the original memory unit. Next, to update the value of the memory unit, first multiply the memory unit by the forget gate, discard the information that needs to be forgotten, and then add the input information obtained from the input gate to obtain the new memory unit value. Finally, the "output gate" decides which memory unit information to output, that is, the hidden state. The calculation formula of LSTM is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{6}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right), \tag{7}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right), \tag{8}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{9}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right), \tag{10}$$

$$h_t = o_t \odot \tanh\left(C_t\right), \tag{11}$$

where f_t, i_t, o_t, and C_t, respectively represent the forget gate, input gate, output gate, and memory unit. For the t time step, h_{t-1} represents the hidden state at the previous time step, W represents the weight matrix, σ represents the Sigmoid function, and $\odot$ represents the point multiplication operation.

3.3. Measure of Research Report.

We measure the research report from the two dimensions of attention and rating sentiment. The attention of the research report can measure the popularity of the stock in the entire market by analysts. The rating sentiment indicates an analyst's judgment on the future trend of the target stock.

3.3.1. Attention of Research Report.

Different stocks in the market receive different degrees of attention. We calculate the ratio of the absolute number of research reports on a stock day to the total number of all stock research reports in the A-share market for the current month as a measure of the stock's attention, $\text{RRA}_{k,t}$ and $\text{RRRS}_{k,t}$. The larger the value, the higher the stock analyst's attention. $\text{RRA}_{k,t}$ is calculated as shown in the following formula:

$$\text{RRA}_{k,t} = \frac{N_{k,t}}{N_{A,m}}, \tag{12}$$

where $N_{k,t}$ is the total number of research reports of stock k on the t-th day and $N_{A,m}$ is the total number of research reports of all stocks in the A-share market in that month (m).

3.3.2. Rating Sentiment Measures in Research Reports.

The text of the research report released by the analyst contains the optimistic or pessimistic attitude of the individual company's operating status, future prospects, earnings expectations, investment recommendations, and risk warnings. The report contains two important key pieces of information: first, the current rating, and second, the rating change. Among them, the current rating provides investment advice on buying, selling, or holding of individual stocks; the rating change indicates the current rating and the previous rating change are reported. In the previous literature, when discussing the impact of stock investment ratings on their abnormal returns, the basic ratings and rating changes were always analyzed separately, making it difficult for investors to choose when the two ratings were inconsistent. For example, when an analyst gives a "Hold" rating to individual stocks, and the rating changes to "Down," it is more difficult for investors to determine whether to buy or sell, so this article innovatively proposes a "research report rating sentiment" index, $\text{RRRS}_{k,t}$, taking into account the two major factors of basic rating and rating change; the calculation method of $\text{RRRS}_{k,t}$ is shown in the following formula:

$$\text{RRRS}_{k,t} = \frac{\sum_1^{N_{k,t}} R_{k,t} \cdot C_{k,t}}{N_{k,t}}, \tag{13}$$

where $R_{k,t}$ is the base rating and $C_{k,t}$ is the rating change. When there are multiple rating results for a stock within a day, this article will average the ratings of these research reports to find a comprehensive rating sentiment.

3.4. Measure of Investor Sentiment.

In the actual trading process, investor sentiment will affect investors subjective judgments on future returns. When investor sentiment rises or becomes more pessimistic, it will trigger its "irrational" behavior of information and cause market anomalies. For the calculation method of investor sentiment quantification, this article draws on the ideas of Hai-Yuan [22], selects Shanghai and Shenzhen cities A transaction volume (VOLUME), the number of new investor accounts (NEWIN), and consumer confidence index (CCI). The five indicators of closed-end fund discount rate (FUND) and broad market turnover rate (HS_TVR) were used for principal component factor analysis, and the initial investor sentiment index for each month was calculated using the

respective variance contribution rate as the weight. Then, macroeconomic control was introduced, regression analysis was performed on the variables, and the calculated residual value was used as the investor sentiment index. Finally, the simple investor average of the period i lagging behind the study report date (where $i = 3$) was used to obtain the final investor sentiment index.

First of all, the abovementioned five indicators are standardized, and principal component factor analysis (PCA) is performed on these standardized variables. The three principal components with the highest variance explanations are selected, and the respective feature values are used as weights to obtain the factor load after weighted averaging, and as the principal component coefficients of the preliminary sentiment index, the preliminary sentiment index is shown in the following formula:

$$\begin{aligned} \text{Sentiment} = &-0.173\text{FUND} + 0.242\text{NEWIN} \\ &+ 0.737\text{VOLUME} + 1.047\text{HS\_TVR} + 0.512\text{CCI}. \end{aligned}$$

$$(14)$$

Then, control the impact of macroeconomic variables. Take the abovementioned preliminary sentiment indicators as the explanatory variables, and the consumer consumption index (CPI), the amount of new credit (IC), the rate of economic growth (GDP), and the money supply (M_2) as the explanatory variables (standardize the data in advance to eliminate dimensional impact), regression analysis of Sentiment_m, the residual sequence can be used as an indicator of investor sentiment: CSI (China Sentiment Index).

$$\text{Sentiment}_m = \alpha_0 + \alpha_1\text{CPI} + \alpha_2\text{IC} + \alpha_3\text{GDP} + \alpha_4 M_2 + \varepsilon_m,$$

$$(15)$$

where Sentiment_m is the preliminary investor sentiment index for m months, α_0 is a constant, and α_{1-4} is the regression coefficient to be estimated.

Finally, considering the lag of investor sentiment, the sentiment index of the three months before the month on which the research report was published was selected to calculate the average value as the final investor sentiment index.

4. Experiments

4.1. Data Collection.
This article selects the research report on 2,225 Chinese A-share companies issued by 66 securities institutions between July 1, 2014, and September 30, 2017, as the research object. The data on the number of research reports published, the date of publication, the title, the basic rating, and the rating changes are from the Oriental Fortune website. Economic data such as stock returns, market value of stocks in circulation, and stock turnover rate are taken from the wind database. For the selected time period, the trend of the Chinese A-share market can be roughly divided into two phases: July 2014 to June 2015 is the rising period of the stock market, which belongs to the "bull market," and July 2015 to September 2017 is the decline of the stock market period, belonging to the "bear market." The sample time spans a bull-bear cycle, which can more effectively verify the robustness of the algorithm.

4.2. Data Preprocessing

4.2.1. Data Culling.
This article deletes some anomalous data: first, unrated or ambiguous research reports; second, new stock data, because during the continuous daily limit of new stock sales, stock price fluctuations do not truly reflect market fluctuations, and during this period, few investors can successfully buy new stocks, so the new stocks issued from July to September 2017 are uniformly eliminated; third, the individual stock research report at the time of long-term suspension, because the long-term suspension of stocks cannot be traded, and the price of the stock cannot be compared with the average market price without fluctuation (temporarily suspended stocks are not excluded).

4.2.2. Rating Consolidation.
In all the reports collected, a total of 27 different basic ratings were included, and we obtained a total of 14 different ratings after synonym merger. At the same time, we sorted out four different rating changes. Referring to the research by You et al. [23], we use discrete values to assign ratings and rating changes. This article assigns the "Neutral" rating to 1.0 and increases or decreases by 0.1 according to the intensity change to obtain the basic rating G, $G \in [0.6, 1.9]$. The specific values of the rating are shown in Table 1:

For the four rating changes of "Up," "First," "Maintenance," and "Down," we assign "maintenance" to 1.0 and increase or decrease by 0.1 according to the rating change to get the rating change C, $C \in [0.9, 1.2]$; the specific value of the rating change assignment is shown in Table 2:

4.3. Detail.
All data are divided into training set, validation set, and test set according to the ratio of 80%, 10%, and 10%. We use categorical cross-entropy as a loss function to optimize the target parameters during the model's backpropagation, which is defined as

$$\text{Loss} = -\frac{1}{n}\sum_{t=1}^{n} y_t \log p(\hat{y}_t),$$

$$(16)$$

where y_t is ground-truth in the form of one-hot, and $\hat{y}_t$ is the model's predicted probability that the excess return is a "positive," "negative," and "par" vector. During the model training process, the Adam [24] function was selected for optimization, where the initial learning rate was set to $1e{-}4$ and the minimum batch size was set to 32.

5. Results and Discussion

In order to measure the performance of the model from different angles, this paper chooses the classic classification algorithm SVM and the vanilla LSTM model as the benchmark method to compare with our work. Table 3 shows the results of the experiment. The results show that in the excess return forecast on the 5th, 15th, and 30th, our proposed method achieves the best performance regardless of the accuracy rate or F1 measurement. In the comparison of the 5th, 15th and 30th, the accuracy of the excess return prediction on the 15th by all methods is the highest. This is

TABLE 1: Assignment of research report ratings.

Rating	Value	Rating	Value
Strong buy	1.9	Cautious overweight	1.2
Buy	1.8	Hold	1.1
Long-term buy	1.7	Neutral	1.0
Overweight	1.6	Recommended with caution	0.9
Buy with caution	1.5	Wait and see	0.8
Highly recommended	1.4	Underweight	0.7
Recommend	1.3	Sell	0.6

TABLE 2: Assignment of research report rating changes.

Change	Up	First	Maintenance	Down
Value	1.2	1.1	1.0	0.9

TABLE 3: Comparison of excess return forecast results by different methods.

Methods	5 days		15 days		30 days	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
SVM	0.631	0.642	0.7081	0.6965	0.6660	0.6705
LSTM	0.6598	0.6671	0.7217	0.7201	0.6876	0.6911
OURS	**0.6833**	**0.6792**	**0.7381**	**0.7399**	**0.7087**	**0.6986**

TABLE 4: Comparison of excess return forecast results in different market environments.

Methods	5 days		15 days		30 days	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
Bull market						
SVM	0.6810	0.6931	0.7335	0.7398	0.7161	0.7176
LSTM	0.6938	0.6978	0.7534	0.7583	0.7216	0.7271
OURS	**0.7155**	**0.7203**	**0.7801**	**0.7868**	**0.7337**	**0.7312**
Bear market						
SVM	0.6688	0.6724	0.7186	0.7113	0.6922	0.6950
LSTM	0.6814	0.6957	0.7476	0.7561	0.7284	0.7189
OURS	**0.7076**	**0.7055**	**0.7465**	**0.7493**	**0.7195**	**0.7112**

TABLE 5: Comparison of excess return forecast results with different inputs.

Methods	5 days		15 days		30 days	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
LSTM	0.6598	0.6671	0.7217	0.7201	0.6876	0.6911
LSTM + Sentiment$_m$	0.6601	0.6766	0.7316	0.7287	0.6891	0.6934
LSTM + Sentiment$_m$ + RRA	0.6768	0.6784	0.7301	0.7287	0.6965	0.7011
OURS_full	**0.6833**	**0.6792**	**0.7381**	**0.7399**	**0.7087**	**0.6986**

related to the release cycle of the research report. According to statistics, the average cycle of all stock research reports in the dataset is 18.9 days. When the period exceeds 20 days, there will be multiple reports overlapping, and the latest report will affect stock price fluctuations and thus affect excess returns.

We included a "bull market" and a "bear market" in our selected trading cycle. Different market states present different trading sentiments, so we further trained the "bull market" and "bear market" data separately and tested them on the test set, Table 4 shows the comparison results. Compared with the training of "bull market" and "bear market" data aggregation, the accuracy of all methods after

training separately according to different market conditions has improved. Similarly, our proposed method achieved the best performance in the 5th, 15th, and 30th excess return forecasts, with the highest accuracy rate as the "bull" 15th excess return forecast. The table shows that the overall accuracy of the "bull market" is higher than that of the "bear market," and this conclusion is consistent with the research results of Hai-Yuan [22].

In order to verify the effectiveness of the increased research report metrics and investment sentiment on the model, we delete the corresponding inputs and compare them. Table 5 shows the results. Among them, LSTM + Sentiment$_m$ indicates that the original hidden

LSTM uses Sentiment$_m$ as the initial hidden state of the model and the input of the model is only the stock price. Based on Sentiment$_m$, the input of the model is the concatenation of the stock price and RRA, and OURS_full represents our complete model. The results show that the addition of Sentiment$_m$, RRA, and RRRS can gradually improve the accuracy of the model.

6. Conclusion

Regarding the prediction of excess returns in the mainland Chinese stock market, first of all, this article measures the research report released by the analyst and splits the research report into two indicators: the attention degree of the research report and the rating sentiment; secondly, we quantify the external environment that may affect stock price changes as investor sentiment; then, we split the research report indicators and investor sentiment as the input and initial hidden state of LSTM; finally, in the comparison of experiments, our proposed method achieved the best performance.

Data Availability

All data in this article are from public websites (https://www.eastmoney.com/).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, NV, USA, December 2012.

[2] 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.

[3] S. Kombrink, T. Mikolov, M. Karafiát, and L. Burget, "Recurrent neural network based language modeling in meeting recognition," in *Proceedings of the Twelfth Annual Conference of the International Speech Communication Association*, Florence, Italy, August 2011.

[4] A. Mnih and G. Hinton, "Three new graphical models for statistical language modelling," in *Proceedings of the 24th international conference on Machine learning—ICML '07*, pp. 641–648, Corvallis, OR, USA, June 2007.

[5] S. Huilin, P. Diyun, H. Xin, and F. Jun, "Research on weibo hotspot finding based on self-adaptive incremental clustering," *Journal of Shanghai Jiaotong University (Science)*, vol. 24, no. 3, pp. 364–371, 2019.

[6] L. Ding and H. Sun, "A study of the effect of recommending stocks to China' s stock Market%," *Manage. World*, vol. 000, no. 5, pp. 111–116, 2001.

[7] L. Xiang, *Multi-factor Quantitative Stock Selection Plan Planning Based on XGBoost Algorithm*, Shanghai Normal University, Shanghai, China, 2017.

[8] B. Du, H. Zhu, and J. Zhao, "Optimal execution in high-frequency trading with Bayesian learning," *Physica A: Statistical Mechanics and Its Applications*, vol. 461, pp. 767–777, 2016.

[9] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Forecasting stock prices from the limit order book using convolutional neural networks," in *Proceedings of the 2017 IEEE 19th Conference on Business Informatics (CBI)*, pp. 7–12, Thessaloniki, Greece, July 2017.

[10] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, 2017.

[11] X. Zhang and Y. Tan, "Deep stock ranker: a LSTM neural network model for stock selection," in *Data Mining and Big Data*, pp. 614–623, Springer, Berlin, Germany, 2018.

[12] X. Li, J. Cao, and Z. Pan, "Market impact analysis via deep learned architectures," *Neural Computing and Applications*, vol. 31, no. 10, pp. 5989–6000, 2019.

[13] 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*, Article ID 102212, 2020.

[14] C. M. C. Lee, A. Shleifer, and R. H. Thaler, "Investor sentiment and the closed-end fund puzzle," *The Journal of Finance*, vol. 46, no. 1, pp. 75–109, 1991.

[15] 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.

[16] P. Hribar and J. McInnis, "Investor sentiment and analysts earnings forecast errors," *Management Science*, vol. 58, no. 2, pp. 293–307, 2012.

[17] L. Zhao, Z. Li, and J. Liu, "The managers preferences, the optimization in the evalution of the investment level and the obtainment of the private information," *Manage. World*, vol. 4, pp. 33–47, 2013.

[18] N. Xu, X. Jiang, Z. Yi, and X. Xu, "Conflicts of interest, analyst optimism and stock price crash risk," *Economics Research Journal*, vol. 7, no. 127, p. r140, 2012.

[19] J. Lu and J. Chen, "Asymmetric relationship between extreme investor sent iment and stock index," *Systems Engineering*, vol. 2, pp. 13–22, 2013.

[20] B. G. Malkiel and E. F. Fama, "Efficient capital markets: a review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] Y. I. N. Hai-Yuan, "A study on effect of media reports on investor sentiment:evidence from China's stock market," *Journal of Xiamen University (Arts & Social Sciences)*, vol. 2, p. 11, 2016.

[23] J. You, Y. Qiu, and C. Liu, "'Changed face phenomena of security analysts' forecasting behaviors: a reputation game model and evidences," *Journal of Management Science. China*, vol. 16, no. 6, 2013.

[24] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv Prepr. arXiv1412, 6980, 2014.