ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 315
International Conference on Pedagogy, Communication and Sociology (ICPCS 2019)

Integrating Market Sentiment with Trading Rules - Empirical Study on China Stock Market

Long Qin
School of Economics and Management
Beihang University
Beijing, China

Ruoen Ren
School of Economics and Management
Beihang University
Beijing, China

***Abstract*—The effectiveness of technical analysis in financial market has been discussed by researchers in many years. With the explosion of text data on internet, it is crucial to capture the market sentiment and integrate with market data to optimize the trading strategy. This paper reviewed the literature of technical analysis, text mining application in financial market forecasting and proposed new method to optimize the trading rule. It analyzed the investor comments on top financial websites in China and further generated market sentiment. The empirical study showed that the return of trading rules with market sentiments outperformed the traditional methods.**

Keywords—technical analysis, text mining, market sentiment

I. INTRODUCTION

Technical analysis began in 19th century, which evolved different stages of morphological analysis, indicator analysis, and automated trading rules. Technical analysis predicts the price of financial asset by analyzing historical data and establishing special trading rules. It has been controversial, even although it has been used in financial markets for more than a century. For a long time, the effectiveness of technical analysis was challenged by mainstream financial economists who believed in the weak effectiveness of the market.

The means of technical analysis has been continuously injected with new theories with the development of information technology. Algorithm trading is based on trading rules and analyzing the achievable information to build statistical models. The advantage of algorithmic trading is the ability to respond to market changes in real time by capturing fleeting investment opportunities and delivering trading orders in a very short time. Today, the definition of data itself changes dramatically. In addition to asset price and volume, the text mining technology enables analyst to extract large information in the free text on internet.

The prosperity of trading strategy systems in financial markets over the past decade suggested that appropriate technical analysis methods can make profits. By collecting more valuable information, investors can improve the forecasting accuracy. If the trader can get the data in a timely manner and aggregate the data into useful information, he can respond in advance and increase the probability of getting extra profit.

In this paper, we analyzed the investor sentiment by retrieving the information on internet. And we proposed a new method by considering investor sentiment into trading rules. We also conducted an empirical study on China stock market. The structure of this paper is as follows: section two, literature review of technical analysis and text mining in financial markets; section three, new trading strategy with integrating of investors' sentiment; section four, empirical research on China stock market; section five, conclusion and discussion for future research.

II. LITERATURE REVIEW

Technical analysis has been widely used in different financial markets, such as stocks, options, futures, foreign exchange market, etc. It evolved from simple trading rules to complex algorithm trading supported by advanced statistical models. The early methods failed to adequately deal with trading risks; the statistical tests on the significance of the returns was not performed; the parameter optimization, out-of-sample testing and data snooping of the rules were often overlooked. In contrast, recent research improved the asset price forecasting by considering all available information, such as the transaction costs, transaction risk, out-of-sample evaluation of optimal rules, and statistical significance testing of returns. With the fast growth of internet, information is no longer limited to structured data. It extends to text information distributed on internet. People express their ideas, judgment and insights on internet anytime anywhere. Computer scientist and linguist did great contribution to extract useful information from the text. Behavioral economists analyzed the public sentiment with text mining methods to forecast the financial market.

In this section, firstly, we reviewed the literatures about the effectiveness of technical analysis, which is the foundation of our proposed method; secondly, the history of text mining application in market prediction was reviewed. Last but not the least, we reviewed the text mining applications on China financial markets forecasting.

Corresponding Author: Long Qin, School of Economics and Management, Beihang University, Beijing, China.

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

A. The Effectiveness of Technical Analysis

The effective market theory (EMH) was developed by Eugene Fama [1] in 1970. It faced great challenges by researchers. A few of researchers could get extra benefits through technical analysis methods. Poole (1967) [2] used 10 simple filter rules from 0.1% to 2% for 9 currencies. Its result was that 4 out of 9 currencies generated averagely 25% annual profit. The simple filter rules beat buy and hold strategy. Logue et al. (1977) [3] analyzed CHF/USD using 14 filter rules from 0.7% to 5% with transaction cost into consideration. It turned out that 13 of 14 filter rules outperformed buy and hold strategy. Cornell et al. (1978) [4] analyzed 6 currencies using moving average rules and filter rules from 0.1% to 5%, and its results showed that 4 currencies' annual profit was over 10%. The filer rules generated annual yield from 9% to 32.9%. Dooley et al. (1983) [5] conducted a quantitative study of the profitability of the filter rules. The purpose was to confirm whether the short-term fluctuations in the exchange rate of the foreign exchange market were caused by the extensive use of technical analysis. Sweeney (1986) [6] developed a test to examine the validity of a filter rule, which was mainly based on the risk premium of the constant coefficient. The empirical results showed that even after considering the transaction costs and risk premium adjustments, the filtering rules could still obtain significant exceeding profits.

Besides simple trading rules, the advanced statistical models were introduced to build the trading system. Brock et al. (1992) [7] found that various linear models could not explain the characteristics of the return of technical trading rules. The empirical result indicated that the returns had hidden nonlinear correlations. Gencay (1999) [8] examined the nonlinear prediction ability of technical trading rules. He conducted an empirical analysis of five spot exchange rates from 1973 to 1992 and found that the results of trading based on nonlinear models (nearest neighbors, feedforward neural networks) and moving average rules (1/50 and 1/200) was better than the random walk and GARCH (1,1) model.

Unlike simple technical trading rules that can be accurately defined, most technical graphics are complex nonlinear geometries. Some scholars have done a lot of research work in this field. Curcio et al. (1997) [9] used support and resistance trading rules to analyze 3 currencies. It turned out that 4 out of 30 rule outperformed buy and hold strategy. Chang et al. (1999) [10] introduced head and shoulder pattern and used moving average and momentum strategy. Guillaume (2000) [11] introduced the breakthrough concept into support and resistance pattern. Osler (2000) [12] pointed out that an in-depth study of these complex technical graphs would help people understand the special random phenomena in financial time series.

A large number of studies had solid evidence that technical analysis could generate exceeding profit. While conducting an empirical test, the data may be different and the processing methods are different.

B. Text Mining for Prediction

With the development of internet and social media, finance market prediction evolves into an interdisciplinary area that both behavioral-economics and artificial intelligence are the core value. A lot of researchers try to interpret the sentiment hidden in the online news, financial reports, and free ideas expression on social media. And then they apply market sentiment into financial asset price prediction. According to the behavioral finance, investors may have cognitive biases such as overconfidence, overreaction, and herd behavior [13]. Their behaviors might be shaped by the information they received. Some researchers treat herd behavior as an extreme example of market sentiment [14]. Market sentimental analysis is identifying positive or negative words and processing text with purpose of classifying its emotional stance as positive or negative.

The advanced statistical models combined with text mining techniques are used to forecast financial market movement. Arman et al. [15] predicted intraday directional-movements of a currency-pair in the foreign exchange market based on the text of breaking financial news-headlines. A multi-layer dimension reduction algorithm with semantics and sentiment was proposed. Nan Li et al [16] studied online forums hotspot detection and forecast using sentiment analysis and text mining approaches. A K-means clustering and SVR were combined to develop unsupervised text mining. Al-Hassan et al [17] explored the difficulties and recommendations while using a textual data mining. Samuel W.K. Chan et al. [18] invested more than 2000 financial reports using a text based decision support system that extracts event sequences from shallow text patterns and predicted the likelihood of the occurrence of events using a classifier-based inference engine. Sven S. Groth et al. [19] applied Naïve Bayes, K-Nearest Neighbour, Neural Network, and Support Vector Machine to detect patterns in the text data and measure the financial risk. Michael Hagenau et al [20] predicted stock price based on financial news using context-capturing features. Xiaodong Li et al [21] quantitatively integrated information from both market news and stock prices to improve the accuracy of prediction on stock future price return in an intra-day trading context. Ali Serhan Koyuncugil et al [22] proposed an early warning system (EWS) model based on data mining for financial risk detection. Jasmina Smailovic et al [23] developed an active learning approach and applied into sentiment analysis of tweet streams in the stock market domain. They used SVM classifier to categorize Twitter posts into three sentiment categories of positive, negative and neutral. By analyzing stock market sentiments of a particular company show that changes in positive sentiment probability can be used as indicators of the changes in stock closing prices.

There were also some researchers analyzing China finance market by taken the investors' sentiment into consideration. Baohua Wang et al. [24] combined ARIMA and SVR models to process textual information to aid the financial time series forecasting. The model with market sentiment could generate better results compared to traditional models. Zhang Kaidong et al. [25] used vector auto regression (VAR) and impulse analysis methods to study the stock price effects on the internet public opinion of Enterprise's emergency crisis incident based on microblog. Chen et al. [26] explored how Sina microblog affecting stock market trends by constructing market sentiment. It got the conclusion that previous 1-3 day's market sentiment was a significant predictor for SSE composite index. Xu Xiang et al. [27] studied the correlation between the change of online public sentiment and the price limit of Shanghai securities index. The result showed that there was correlation relationship. Zhu Chang-sheng et al. [28] established SVR regression model to study the stock returns and predicted the future return rate based on network public opinion. They used R software to process the Chinese language. Ding et al. (2017) [29] proposed a Chinese stock market investor sentiment index, study the contagion effect of international crude oil price fluctuations on Chinese stock market investor sentiment. Ni et al. (2015) [30] used panel quantile regression model and find that investor sentiment can greatly lead to the stock mispricing in Chinese stock market. Guo et al. (2017) [31] got the investor sentiment data through semantic analysis of a popular professional social networking site of China, and the data could be used to predict the stock price only when the stock had high investor attention. Zhu et al. (2016) [32] analyzed the mechanism behind the effects of investor sentiment and accounting information on stock price based on the data from China's A-share market. Zhang et al. (2017) advocated the provincial TV audience rating as the novel proxy for the provincial investor sentiment (PIS) and investigated its relation with stock returns.

III. METHODS

As we reviewed in above sections, the effectiveness of technical analysis in financial market was proved by many researchers. And the application of text mining methods integrating with traditional forecasting methods was widely discussed. In this paper, we proposed that a new method of integrating market sentiment with trading rules will generate exceeding returns. For simplicity, we selected common used technical index, i.e., EMA, MACD, RSI, and CC index, as basic trading rules. We will compare the results in empirical study section.

A. Exponential Moving Average

Simple Moving Average (SMA) is an arithmetic moving average of prices. Since the short-term average respond quickly to changes while long-term average reacts slowly. Investors watch for short-term averages to cross above longer-term averages as the signal of price uptrend. On the other hand, if long-term average is above the short-term average, it is the signal of price deprecation. The limitation of SMA is that the same weight for each observation period is applied. The most recent data points should have greater weight and significance. Exponential Moving Average (EMA) weighted moving average reacts more significantly to recent price changes, which overcomes the shortages of SMA.

It expressed as below:

$$EMA_n(P,m) = (1-\alpha) \times EMA_{n-1} + \alpha \times P_n$$

Where P_n is the close price on day n, α is the smooth factor, *m* is the average day. In practice, traders often use 12 and 26 days EMA as the short term EMA; 50 and 200 days' EMA are long term EMA. When the short term curve is above the long term curve, it is a signal of buying asset; and when the long term EMA is higher than the short term curve, it is a signal of selling asset.

B. Moving Average Convergence Divergence (MACD)

EMA has its shortcoming which may often generate pseudo-signal. In order to conquer it, MACD index is created by capturing the characteristics of the aggregation and deviation of short term and long term moving average line. It is defined as:

$$MACD = EMA(close, 12) - EMA(close, 26)$$

And its signal line is defined as:

$$Signal = EMA(MACD, 9)$$

When MACD line crosses above the signal line, it is a call signal. When MACD line crosses below the signal line, it is short signal.

C. Relative Strength Index (RSI)

RSI is used to measure the magnitude of recent price change to evaluate overbought or oversold condition.

It is defined as:

$$RSI(m) = 100 - \frac{U}{U+D} \times 100$$

Where U represents the sum of upward changes, and D represents the sum of downward changes.

The value of RSI is between 0 and 100. When RSI is smaller than 30, it is an oversold status, the price is lower than its instinct value. This is a buy signal. When RSI is greater than 70, it is an overbought status, which means the price is very high. And it is a sell signal.

D. Commodity Channel Index (CCI)

CCI was originally developed to spot long-term trend changes, but has been used by traders on all timeframes. It captures the deviation of the close price from moving average price. It is defined as:

$$CCI(m) = \frac{1}{\alpha} \times \frac{P_n - SMA(P,m)}{\sigma(P)}$$

Where $\alpha = 1.5\%, P = \frac{H+L+C}{3}$, σ is a deviation, and m is the time duration. Usually, CCI is between -100 and 100.

The basic strategy is to track the CCI for the movement above 100, which generates buy signals, and movements below -100, which generates sell signals.

E. Trading Strategy with Market Sentiment

SMA and EMA may generate pseudo-signals, which will lead traders into wrong direction. Especially when major events occur in the financial market, the price trend will change immediately. The forecasting model based on the traditional technical analysis cannot capture the sudden changes and the market sentiment.

At time t, the market sentiment is measured at:

$$S_t = \begin{cases} 1, Positive \\ 0, None \\ -1, Negative \end{cases}$$

Then our trading rule with market sentiment is as below:

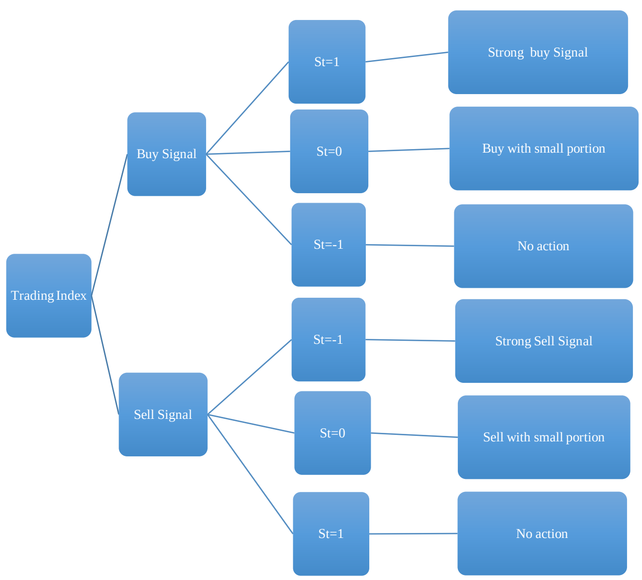

Fig. 1. The New Trading Strategy

When the buy signal appears and the market sentiment is positive, then it is treated as strong buy signal. But if the market sentiment is negative while buy signal appears at the same time, our trading rule will be “no action taken”. When there is neutral sentiment, and the rule will still execute buy strategy with small portion. If the sell signal appears, with the similarity, then the strong sell signal is generated with negative market sentiment. But when market sentiment is optimistic, "no action is taken".

The traditional trading strategy mainly captured the history price and volume data. It could not reflect the market events immediately, and could not reflect people’s feeling and judgment about the whole market. For example, the interest rate changes announced by central bank, economic policy changes, international affairs, etc. will affect the financial market movement immediately. Integrating the market sentiment with traditional trading rules is a great improvement to help traders managing risks and gaining additional profits.

IV. EMIPIRICAL STUDY

We collected the financial comments from three famous online forums in China. They are quite representative of the whole investors’ opinion and also it is enough for the model explanation purpose. The selected websites are very active and popular in China. The process of the empirical study is as below:

Fig. 2. The Workflow

A. Data Source

In this paper, we choose three top active financial websites in China, i.e., Tianya BBS, East Money, Sina Microblog. The comments on these websites are Chinese. So the Chinese text mining techniques are used to process the data.

16990 comments from the above source were collected. In the selected website, the data quality from Sina microblog was the best. The reason was that users on Sina microblog were real name verified. In addition, most of the authors are famous economists or public characters in financial area. We only collect the main article published by the author. The comments from the followers are discarded because it includes some spam information.

TABLE I. DATA SOURCE

Website	URL	Date From	Date To
Tianya BBS	http://bbs.tianya.cn/list-stocks-1.shtml	2018/9/1	2019/3/1
East Money	http://guba.eastmoney.com/	2018/9/1	2019/3/1
Sina Microblog	http://blog.sina.com.cn/lm/search/stocks/	2018/9/1	2019/3/1

In order to avoid the bias on specific companies, we collect comments about the overall market trends. It aligns with our market data source. We use the Shanghai Stock Exchange (SSE) composite index as objective because it is a good presentation of overall market situation. The market data is downloaded from yahoo finance. The open, high, low, close price and trading volume expressed by the overall trading amount is used. The market data timeframe is the same as text data. The future of the composite index can be traded with T+0 rule. This will lead to many trading intraday.

B. Text Processing and Mining

Chinese word segmentation is the basic part of Chinese text processing and the basis for text mining. Words are the smallest unit of independent expression semantics. Chinese is different from English which are composed of words and spaces. In Chinese sentences, there is no clear symbolic division between words. This requires a certain word segmentation method for Chinese word segmentation. Chinese word segmentation is the difficult part in dealing with Chinese text information. Splitting words into correct expression of text semantics is the key of text mining. The Chinese word segmentation method can be divided into the following three categories, i.e., matching based method, statistical based word segmentation method, and semantic based word segmentation method.

After word segmentation, the general text data is converted into structured data such as matrix, and then analyzed using statistical analysis, data mining or machine learning algorithms. Common methods for text mining include word frequency analysis, topic models, text categorization, and association analysis. Classification is a common method used in text mining, mainly including unsupervised classification. Algorithms such as hierarchical clustering, K-Means, String Kernals are often used. K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are widely used to topic model. Nowadays, supervised classification methods have become the mainstream of text classification because they have clear evaluation criteria and have achieved good results in practical applications. Among them, the SVM classifier has good versatility, high classification accuracy, fast classification speed and wide application prospects. In this paper, we used the RTM package which contains various algorithms to process the word segmentation and classification.

C. Sentiment Analysis

There are two commonly used methods to classify the text into positive or negative: one is based on an emotional dictionary; the other is based on machine learning. This paper used mature open source sentiment dictionary, i.e., HowNet Chinese sentiment dictionary [34]. The method based on sentiment lexicon to calculate the emotional score of the text. The dictionary-based approach mainly uses a previously defined sentiment dictionary, by training corpus and emotions to calculate the emotional score of a single sentence in each comment, and finally the emotional score of the aggregated sentence is judged.

In this paper, we use R sentiment package RTextTools developed by Timothy P. Jurka [35]. A sentiment score is generated to each article in the dataset. The R package uses nine algorithms for ensemble classification for positive and negative emotional classification. We used neural networks algorithm in this paper [36]. We calculate average sentiment score as market overall sentiment at the point of the signal appear.

D. Results

While calculating the returns, the transaction cost was not considered in our model. A test of hypothesis was applied to identify if the trading rule generate positive return with statistical significance.

H_0: The return for the trading rule is zero

H_1: The return for the trading rule is positive

Assuming the distribution of the returns follows normal distribution, and the T-test method is applied. If P-value is lower than 0.1, then the H_0 is rejected. It means the trading rule can generate positive return. Table 1 shows the results comparison between trading index and the new method.

TABLE II. TRADING RESULTS ON SSE COMPOSITE INDEX

Index	# of signals	Correct Amount	Correct Rate	Returns
MACD	54	16	29.63%	1.95%*
RSI	93	23	24.73%	-6.82%
CCI	78	19	24.36%	-3.03%
MACD&SA	32	15	46.88%	4.28%*
RSI&SA	42	27	64.29%	-2.52%
CCI&SA	29	16	55.17%	1.91%**

Note: * represents the P=0.1, ** represents the P=0.05

We can see from the results that the new method with the market sentiment consideration generates fewer trading signals, but improves the correct rate significantly. The MACD can generate positive return of 1.95% with statistical significance, where P value equals 0.1; the new method of MACD&SA can improve the correct rate from 29.63% to 46.88%. And the return is improved 219% from 1.95 to 4.28%. However, RSI and CCI could not generate positive returns during this period. Unfortunately, even using new methods on RSI&SA, the return is till negative. But the return was improved as well. The CCI&SA rule has significant improvement compared to standard CCI. The return becomes positive and the P value is 0.05.

V. CONCLUSION

In this study, we reviewed the effectively of technical analysis in financial market. It was a foundation of the paper. And then we extend the definition of data into unstructured text data. Further, market sentiment is extracted from market text data and applied into trading rules. The empirical results showed that there was significant improvement by integrating market sentiment.

However, there are some limitations of this paper which can be further investigated. Only some simple trading rules were used in this paper, more complex trading rules such as graphs, patterns can be evaluated. And for text mining algorithms, we only used neural networks. More advanced algorithms can be used to verify which one can improve the forecasting rate and product higher returns. In addition, the transaction cost was ignored in this paper. By considering the transaction cost, it might cause the reduction of return significantly.

ACKNOWLEDGEMENTS

Thanks are given for the reviewers' detailed review and their very valuable comments. With the reviewer's great insights, the paper's structure and conclusion was further enhanced. Prof. Ren Ruoen provided insights and supervision on the methodology.

REFERENCES

[1] Fama, Eugene (1970). Efficient Capital Markets: A Review of Theory and Empirical Work [J] Journal of Finance. 25 (2): 383–417.

[2] Poole, W. Speculative prices as random walks - an analysis of ten time series of flexible exchange rates [J]. Southern Economic Journal, 1967, 33: 468-478.

[3] Logue, D. E.and Sweeney R. J. White-noise in imperfect markets: the case of the Franc/Dollar exchange rate [J]. Journal of Finance, 1977, 32:761-768.

[4] Cornell, W.B. and Dietrich, J.K. The efficiency of the market for foreign exchange under floating exchange rates [J]. Review of Economics and Statistics 1978, 60:111-120.

[5] Dooley, M.P., Shafer JR. Analysis of short-run exchange rate behavior: march 1973 to september 1975 [A]. Intl'Finance Dissussion Paper 123, FRB, Washington, D.C. 1983.

[6] Sweeny, R. J. Beating the foreign exchange market [J]. Journal of Finance,1986, 41: 163-182.

[7] Brock, W., J. Lakonishock, and B. LeBaron. Simple technical trading rules and the stochastic properties of stock returns [J]. Journal of Finance, 1992, 47: 1731-1764.

[8] Gen çay, R. Linear, Non- linear and essential foreign exchange rate prediction with simple technical trading rules [J]. Journal of International Economics, 1999, 47: 91-107.

[9] Curcio, R., C. Goodhart, D. Guillaume, and R. Payne. Do technical trading rules generate profits? Conclusions from the intra-day foreign exchange market [J]. International Journal of Finance and Economics, 1997, 2: 267-280.

[10] Chang, P. H. K., and C. L. Osler. Methodical madness: technical analysis and the irrationality of exchange-rate forecasts [J]. Economic Journal, 1999, 109: 636-661.

[11] Guillaume, D. M. Intradaily exchange rate movements [M]. Boston, MA: Kluwer Academic Publishers, 2000.

[12] Osler, C. L. Support for resistance: technical analysis and intraday exchange rates [M]. Economic Policy Review, Federal Reserve Bank of New York, 2000, 6: 53-65.

[13] Shiller, Robert J. (2000). Irrational Exuberance. Princeton University Press. pp. 149–153.

[14] Robert Prechter, The Wave Principle of Human Social Behavior, New Classics Library (1999), pp. 152–153.

[15] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, David Chek Ling Ngo:Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment [J]. Expert Systems with Applications, 42 (2015) 306–324.

[16] Nan Li, Desheng Dash Wu:Using text mining and sentiment analysis for online forums hotspot detection and forecast [J]. Decision Support Systems, 48 (2010) 354–368.

[17] Abeer A. Al-Hassan, Faleh Alshameri, Edgar H. Sibley:A research case study: Difficulties and recommendations when using a textual data mining tool [J]. Information & Management, 50 (2013) 540–552.

[18] Samuel W.K. Chan, James Franklin:A text-based decision support system for financial sequence prediction [J]. Decision Support Systems, 52 (2011) 189–198.

[19] Sven S. Groth, Jan Muntermann:An intraday market risk management approach based on textual analysis [J]. Decision Support Systems, 50 (2011) 680–691.

[20] Michael Hagenau, Michael Liebmann, Dirk Neumann:Automated news reading: Stock price prediction based on financial news using context-capturing features [J]. Decision Support Systems, 55 (2013) 685–697.

[21] Xiaodong Li, Xiaodi Huang, Xiaotie Deng, Shanfeng Zhu:Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information [J]. Neurocomputing, 142 (2014) 228–238.

[22] Ali Serhan Koyuncugil, Nermin Ozgulbas:Financial early warning system model anddata mining application for risk detection [J]. Expert Systems with Applications, 39 (2012) 6238–6253.

[23] Jasmina Smailovic´, Miha Grcˇ, Nada Lavracˇ, Martin Zˇnidaršic: Stream-based active learning for sentiment analysis in the financial domain [J]. Information Sciences, 285 (2014) 181–203.

[24] Baohua Wang, Hejiao Huang, Xiaolong Wang: A novel text mining approach to financial time series forecasting [J]: Neurocomputing, 83 (2012) 136–145. (in Chinese)

[25] Chen D, Qi J: Research on the stock price shock effects of the internet public opinion of Enterprise's Emergency Crisis Incident Based on Microblog [J]. Journal of Intelligence, 2015, 34 (03): 132-137+149. (in Chinese)

[26] Chen Yunsong, Yan Fei: Does online sentiment predict stock market indices? The ARDL bounds tests based on Sina-Microblog data [J]. Chinese Journal of Society 2017, 37 (02): 51-73. (in Chinese)

[27] Xu Xiang, Jin Jing: A Correlation Analysis between online public opinion and Shanghai securities composite price limit: A Text Mining Based on LDA Theme Model [J]. Journal of Hangzhou Dianzi University (Social Sciences) 2018, 14 (03): 18-24+31 (in Chinese)

[28] Zhu Chang-Sheng, Sun Xin, Feng Wen-fang: Study on the impact of network public opinion on the stock market based on R-Language [J] Journal of Lanzhou Univeristy of Technology, 2018, 44 (04): 103-108. (in Chinese)

[29] Ding Z, Liu Z, Zhang Y, et al. The contagion effect of international crude oil price fluctuations on Chinese stock market investor sentiment [J]. Applied Energy, 2017, 187: 27-36 (in Chinese)

[30] Ni Z X, Wang D Z, Xue W J. Investor sentiment and its nonlinear effect on stock returns—New evidence from the Chinese stock market based on panel quantile regression model [J]. Economic Modelling, 2015, 50: 266-274. (in Chinese)

[31] Guo K, Sun Y, Qian X. Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market [J]. Physica A: Statistical Mechanics and its Applications, 2017, 469: 390-396. (in Chinese)

[32] Zhu B, Niu F, Chan K, et al. Investor sentiment, accounting information and stock price: Evidence from China [J]. Pacific-Basin Finance Journal, 2016, 38: 125-134. (in Chinese)

[33] Zhang Y, Zhang Y, Shen D, et al. Investor sentiment and stock returns: Evidence from provincial TV audience rating in China [J]. Physica A: Statistical Mechanics and its Applications, 2017, 466 (Complete): 288-294. (in Chinese)

[34] http://www.keenage.com/

[35] https://cran.r-project.org/web/packages/RTextTools/index.html

[36] Tomer Geva, Jacob Zahavi: Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news [J]. Decision Support Systems 57 (2014) 212–223.