Related Research on News Sentiment Tendency and Stock Price Fluctuation

Hongye Li

Department of Software Engineering, Southeast University, Nanjing, China

Abstract—In recent years, with further development of the Internet and artificial intelligence, the research on using news sentiments, and then predicting investor behavior has become more feasible. Based on the news information obtained from the Internet, this article analyzes the investor 's emotions about the news, further predicts the behavior of investors, and then the rise and fall of investment asset prices. We use news sentiment tendency and stock price of listed companies to analyze whether the news sentiment really affects the stock price and the degree of its influence.

The research results of the thesis are divided into two parts: (1) A CNN neural network capable of analyzing news that may have an impact on stock prices in the near future is designed and trained to obtain a timely response and form a simulated investment decision for investors to adjust the investment strategy timely; (2) Combined with the economic indicators, this article uses the LSTM neural network to comprehensively investigate whether the emotional tendency of the news has a significant impact on the stock price. This article draws four kinds of images, which can test the accuracy of the model from different angles and guide consumers' investment.

1 Introduction

Predicting the behavior of the stock market has always been very attractive for academic researchers and practitioners in the financial industry.

Although it is a very risky activity, it can bring huge benefits to people, which is really hard to reject; and predicting the stock market based on new information is a more interesting area, The efficient market hypothesis EMH (Efficient Market Hypothesis) proposed by the Finance Professor Eugene Fama (1970) from University of Chicago shows that when a good news or a bad news just came out, the stock price began to change[1]. When it was already known to all people, the stock's Prices have also risen or fallen to a appropriate price level (i.e. at any time prices fully reflect all available information), but scholars are still willing to challenge this theory.

Although scholars have a keen interest in this field, and studying this field can bring huge profits, the analysis of unstructured texts, and then predicting the stock market prices, has earned little effect. In practice, as one of the important resources of market information, the analysis of text content such as news and comments has been widely used by investors, and the analysis methods have also been studied in detail by scholars. Financial news conveys new information to market participants and investors respond to these news in a timely manner to make investment decisions.

In fact, news on the Internet affects the decisions and behaviors of investors, which in turn affects the fluctuation of various asset prices in the capital market. Due to the availability and timeliness of financial news on the Internet, as Transmission of news has become more and more widely, information spreads faster and faster, and the recipients of such information (generally investors) will respond more quickly to this information. They will have different perceptions of the information contained in the news and make different investment behaviors based on individual or collective judgments, which will have an impact on the prices of assets and the stock market. With the advancement of text mining technology, there have been some research results exploring the relationship between news sentiment and stock prices, but not so many.

This article analyzes the relationship between news sentiment of listed companies and stock prices. Through theoretical analysis and practical tests, it is to find out what extent that news sentiment affects stock price fluctuations; and then explores the relationship between the two from a novel perspective so as to gain an appropriate prediction result based on the researched relationship. In order to make the results more accurate, this article selects some economic indicators that can affect stock prices, combines news sentiment tendency and economic indicators, and comprehensively predicts the relationship between news sentiment and stock prices.

The paper selects the news from the financial section of Reuters and the top-ranked news on World News

213171176@seu.edu.cn

section of the Reddit Forum, Yahoo Financial's stock data, and three leading economic indicators provided by the OECD: BCI (Business Confidence Index), Consumer Confidence Index (CCI), Composite Leading Index (CLI) as the research samples.

This article trains a CNN neural network that can analyze investment decisions for news headlines. By obtaining the news headlines and text of the website from a recent period of time (maybe within a day), after preprocessing, the CNN network obtained from the previous training is used. According to the rise and fall of the stock company corresponding to the news on the day, the investment decisions that investors will make are simulated. There are five types of investment decisions: "Strong Sell", "Sell", "Unknown", "Buy", "Strong Buy ". Then based on the number of buying and selling in the investment decision, a stock recommendation decision list and a word cloud map as well as the news links would be formed to facilitate investors to browse news information; investors can combine the previously formed recommendation decision list with the corresponding company news to make their own investment decisions quickly.

This article also analyzes the sentiment tendency in the news and uses the LSTM neural network to comprehensively examine the impact of news sentiment tendency on stock prices combined with the economic indicators previously obtained. The four kinds of images drawn in this article are: (1)stock price trend chart in units of one day, (2)stock price trend chart in units of one month, the above two are convenient for investors to observe the accuracy of the prediction model; (3)buy and sell point charts , test the model's implementation of the "buy low and sell high" principle; (4)Consumer income map, based on the previous calculation of the buy and sell points, it can quantitatively calculate the consumer's income according to the investment decision specified in this article so as to test the effectiveness of the model.

2 Literature Review

In recent years, With the improvement of computing performance and data processing ability of computer, more people choose to use computers to help people make transaction choices. At the same time, along with advances in the fields of machine learning and natural language processing, more people are using unstructured text data as a resource for investment analysis.

In the field of natural language processing, by far, sentiment analysis can be the most influential element of stock market price prediction. In this section, this article reviews the predecessors' discussions on the relationship between text information and investment, consumption choices, and analysis of news sentiment.

2.1 The Analyze of the Correlation between Text Information and Investment

In practice, news about various events and comments on events are one of the main sources of market information. They are also widely used and analyzed by investors. It

is a novel research idea and research direction to carry out information mining on the above text information, analyze the amount of text information or text content, and obtain indicators such as emotional orientation of the text, and then analyze the relationship between these indicators and investment or consumption behavior.

2.1.1 The Correlation between News Amount and Stock Price

In early studies of the relationship between news and stock prices, the text content of news was not or was rarely involved, mainly using news titles and news quantities to represent news, and then supplemented by statistical methods to analyze the impact of news on stocks.

Mitchell and Mulherin (1994) found that through research, the number of Dow Jones daily announcements is directly related to stock trading volume and stock returns[2]. Klibanoff (1998) and others used the headlines of the New York Times to take the closed-end national funds of various countries as a sample to study the relationship between the fluctuation of fund prices and news reports, which found that whether a country has news in the New York Times would affect the pricing of the country's closed-end national funds[3]. Huberman and Regev (2001) used a case study: The New York Times published a news article on a weekend "A biopharmaceutical company named EntreMed develops a drug that can treat cancer", after the press released later, the company's stock price increased from the closing price of $ 12.063 last Friday to an opening price of $ 85 on Monday[4]. The closing price on the day was close to $ 52, an increase of more than 330%. Other pharmaceutical companies also rose affected by this news. This case strongly proves that news affects the price of stocks.

Two Japanese scholars Takeda and Yamazaki (2006) further used the event research method to study the impact of the documentary titled "Project X" on the company's stock price broadcast by the national television station NHK. In the 190 episodes from 2000 to 2005, they selected 69 Japanese listed companies as research samples and classified them based on industry reports and viewing status. The research results show that the documentary has a certain impact on the company's stock prices in all industries, especially when the content of the program involves the company's operating capabilities and development status, its impact on the company's stock price is more obvious[5]. Similarly, Li Qing et al. (2014) studied the relationship between public and media sentiment and the stock market. The study found that media coverage of a company's basic information can increase the amount of information for investors and significantly affect investment behavior. Public sentiment has a significant impact on the emotional fluctuations and investment decisions of individual investors[6].

2.1.2 The Correlation between Online Comment and Investment

The research on the impact of online reviews on individual economic behaviors mainly quantifies the impact of online reviews on consumer purchasing decisions and the impact of online reviews on consumer investment decisions.

Godes and Mayzlin (2004) have found that the increase in the average score of online book reviews is positively related to book sales[7]. Similar studies include Chevalier and Mayzlin (2006) [8] and Qiang Ye (2009) [9]. Chevalier and Mayzlin use Amazon.com and Barnesandnoble.com as examples to test the impact of online reviews from customers on book sales. They found that an increase in consumer evaluation scores for books can increase corresponding sales, and that for most examples, one-star reviews have a greater impact than five-star reviews. Qiang Ye and others studied the relationship between online reviews and hotel room sales, and analyzed data on five indicators: review number, review level, room price, city and star rating. Online booking of rooms plays an important role.

Dellarocas (2005) conducted a more in-depth study. He examined the comprehensive rating of reviews, the number of reviews, and the product sales cycle in which online reviews are located. He concluded that online reviews can be used to predict, but consumers' attitudes to products cannot be evaluated. Early online reviews can effectively predict product sales, while later online review information does not have a significant predictive effect on film review revenue[10].

Mudambi and Schoff (2006) analyzed product reviews on Amazon to analyze what kind of reviews are useful to buyers and pointed out that there is a close relationship between review depth, product type, and consumers. Buying decisions have also had an impact[11]. And Moe and Schweidel (2012) for the first time proved that different consumers have different text reviews, that is, there are significant differences in the responses of different individuals to different reviews[12].

2.1.3 Some Results of Analyzing Stock Prices with Text Mining Technology

In addition to the sentiment analysis of news texts, researchers have also adopted other text mining techniques to extract the text information that affects the stock market from the media, quantify the news text, and apply it to stock analysis.

Feldman and Dagan (1995) combined the text paradigm with KDD to summarize a set of KDT framework structures. Based on this model, 22,173 financial news articles published on Reuters were used as research objects to dig up distribution of the country, subject, person, organization and stock. As one of the earliest text mining models, KDT provides an effective method for mining news text information[13].

Since then, the popularity of text mining technology in research has gradually increased. Tan (1999) proposed another classic text mining model: a two-stage model.

This model divides text mining into two stages: the first stage is Text Refining, which converts unstructured text into structured text; the second stage is Knowledge Distillation, that is, Extract patterns and knowledge from text. Two-stage text mining models have been widely used since they were proposed[14].

Lægreid and Sandal (2006) used financial news on the Norwegian market as a research object, using part-of-speech tagging, name entity identification, and traditional information retrieval as well as information extraction and other methods to summarize the process of news text mining; therefore, entity attributes such as company name and people can be obtained from the news text, and these entities and their relationships can be monitored over time[15]. Schumaker and Chen (2006) vectorized news in different text representations and used support vector machine (SVM) to predict stock price 20 minutes after news release[16].

2.1.4 The Correlation between News Sentiment and Stock Price

As one of the text mining techniques, sentiment analysis has also been adopted by researchers. In some stock market news, this article often observed the subjective views of some news authors or stockholders, such as bullish, bearish, and rebound. These subjective views are also a kind of news sentiment. They are converted into suggestions of investors and accepted by investors.

Antweiler and Frank (2006) collected news related to the stocks of 45 companies on the Dow Jones Industrial Index and the Dow Jones Internet Index published on "Yahoo! Finance" and "Raging Bull"[17], they used Lewis' (1998) [18] Naive Bayes text classification, and divides these news texts into three types of emotional tendencies: bullish, bearish, and neither; so the news is divided into three suggestions for investors buy, sell and hold. They adopted time series analysis and the GARCH model to find out bullish information that helps predict active level of the stock market.

Tetlock (2007) collected historical data from the Wall Street Journal's 1984-1999 "Abreast of the Market" column, and divided the column information into two types of optimism and pessimism; then they studied its relationship with the stock market by using the principal component analysis method to quantify textual information, and find that pessimistic news will bring downward price pressure and subsequent price reversals, besides, pessimistic news has a more negative impact on stock prices than positive news on stock prices influences[19].

Tetlock et al. (2008) based on the Harvard-IV-4 dictionary to quantify news sentiment in Dow Jones News Service and Wall Street Journal newspapers. General Inquirer software was used to calculate the positive and negative words in each news. The ratio is used to measure the emotional tendency of the news and study its relationship with stock returns. Studies have shown that negative news coverage by the company means that the company will have lower profits, indicating a lower return on the stock price[20].

Dougal et al. (2011) used the Wall Street Journal's exogenous scheduling of Wall Street Journal columnists as a sample, and by adding a columnist as a control variable, multiple regressions between financial news reporting and the stock market relationship. The study found that media news reports are closely related to the stock market, and the writing style of columnists has a significant impact on investor behavior[21].

2.1.5 The Correlation between Artificial Neural Network and Stock Price

In recent years, the Artificial Neural Network is one of the hottest areas of computer research. Neural networks have been widely used in stock prices, financial returns, exchange rate risks, and other fields due to their non-linear mapping ability, good self-learning, and adaptive performance.

Since stock data, trading volume, and other data contain a lot of information that affects stock price changes, artificial neural networks can learn the historical data of stocks to find the laws of stock prices. But the actual financial data is affected by many factors, and the time series formed by it is more random, and usually have multiple levels and scales. Therefore, the prediction model of single neural network has limitations and has a certain impact on the prediction accuracy of stock prices.

In neural networks, Recurrent Neural Networks (RNNs) is a powerful model for processing sequential data. It can be used to process and analyze sound, time series data, and written natural language[22]. Some RNNs networks have been designed to make stock market predictions [23] [24].

Long Short-Term Memory (LSTM) is one of the most successful RNNs architectures. LSTM introduces a memory cell, a computing unit that replaces traditional artificial neurons in the hidden layer of the network. With these storage units, the network can effectively associate storage and timely input remote data, so it is suitable for dynamically grasping data structures that change over time with high predictive power[25].

2.2 The Analyze of the Correlation between Economic Standards and Stock Price

In addition to news text, some economic indicators may also affect stock prices.

One way to analyze stock price changes is to use Consumer Confidence and Business Confidence, which have also attracted great interest from researchers.

Şakir Görmüş and Sevcan Güneş (2010) used several econometric techniques to study the impact of CCI on stock prices between 2002: 1 and 2008: 12, and the results of the GARCH-M and OLS models showed that CCI affects stock prices[26].

Khan and Upadhayaya (2019) studied the impact of Business Confidence on investment decisions. They used US business confidence survey data between 1955Q1-2016Q4 and found that Business Confidence has predictive power for investment growth. Compared to traditional forecasting indicators, Business Confidence has a stronger ability to predict investment downturns in the first three quarters and signs of investment growth in the second quarter[27].

Jan Dovolil (2016) analyzed various economic indicators and used them as indicators to predict the development of the S & P 500 stock index. The results show that the leading indicator serves as an indicator of the future development of the S & P 500 stock index[28].

3 Methodology

3.1 Data Collection

3.1.1 News Data and Stock Price

First, according to the research results of Zhang and Skiena (2010), when a stock's market capitalization is large, the daily trading volume and the number of news articles cited by its corresponding company will also increase[29]. This means that analysis of companies with large market capitalization can better find the relationship between investment choices and the number of news; at the same time, it also shows that there are more news reports on actively-traded stocks than relatively inactive stocks, which means that this article can get more news from stocks with higher market capitalization. So this article downloaded the NASDAQ website from the NASDAQ website corresponding to the market capitalization of the top 20% of the company's ticket list which contains the corresponding company's stock ticker, the company's full name, the name of the company's stock exchange and the company's corresponding market cap, and this table is stored for later use.

Then, this article on Reuters, using the tickers list obtained earlier, to crawl the financial news of the company corresponding to the ticker from the time range of September 2011 to August 2019. According to research by Radinsky et al. (2012) and Ding et al. (2014), this article uses the headline of the news to make predictions rather than the entire news article, since they found that using the headline to make predictions can achieve better effect. The news title corresponding to each company is listed behind their corresponding ticker, and the news titles are separated by "|", so that the news title can better correspond to the stock, and vice versa[30] [31].

In terms of forum information, this article uses the Wayback Machine to collect the hottest world news headlines of the Reddit Forum news section from May 2010 to July 2019 to obtain the trend of global news worldwide and to facilitate the analysis of the world, in which way, obtaining changes in sentiment within the news, and the health of the overall market.

In terms of stock prices, this article obtained "Yahoo! Finance" according to the previously obtained ticker list. The stock price information of various companies in

December 2009 and August 2019, including opening prices, closing prices, and so on.

3.1.2 Economic Indicators

For economic indicators, this article selects the Business Confidence Index, Consumer Confidence Index and Composite Leading Index provided by OECD.

The Business Confidence Index provides information on future developments based on opinion surveys on the production, order, and inventory development of manufactured goods in the industrial sector. It can be used to monitor output growth and forecast turning points in economic activity. Index above 100 indicate increased confidence in recent business performance, and index below 100 indicate pessimism about future performance[32].

The Consumer Confidence Index provides an indicator of the future development of household consumption and savings based on responses to households' expected financial status, overall economic conditions, unemployment rate, and ability to save[33]. An indicator above 100 indicates that consumers have greater confidence in the future economic situation, so they are less inclined to save and more likely to spend money on major purchases in the next 12 months. A value below 100 indicates a pessimistic attitude towards future economic development, which may lead to an increase in savings and a decrease in consumption.

The OECD CLI is used to predict cycles in a reference sequence that is representative of economic activity[34]. Fluctuations in economic activity are changes in economic output relative to its long-term potential. The difference between potential and observed output is often referred as the output gap, and fluctuations in the output gap are called as the business cycle. However, the output gap cannot be directly observed; in fact, it is estimated as part of the entire CLI production process.

The above three economic indicators can be used to study the overall market health and are adopted as economic indicators for analyzing the investment behavior of investors and then predicting the trend of stock prices. This article selects the indicators for the time range from January 2014 to June 2019 as a sample.

3.2 Stock Prediction

3.2.1 Convolutional Neural Network for Investment Prediction

This paper uses a convolutional neural network to classify the financial news appearing on the day of the company according to the rise and fall of the stock price on the day to obtain the target of using financial news to predict investment decisions.

Based on the principles proposed by Kim (2014), this paper designs a model of CNN and classifies news into two categories for training according to the price changes of the corresponding company on the day[35].

Before using a convolutional neural network to classify news headlines, this article first preprocesses the obtained news content. After preprocessing the obtained news content, this article reads the news file, counts the word frequency, and generates a word cloud map by word cloud module.

The network constructed in this paper is roughly as follows: in the first layer, words are embedded in low-dimensional vectors; in the next layer, multiple filter sizes are used to perform convolution on embedded word vectors; next, the convolutional layer's result "max-pool" is taken ad a long feature vector, adding dropout regulation, and then the result is classified by softmax layer.

In implementation, this article makes some simplifications to the model proposed in the original paper:

(1) This article will not use pre-trained word vectors for word embeddings. Instead, word embeddings from scratch will be taken into action;

(2) This article will not enforce L2 norm constraints on the weight vectors. According to the results of Zhang (2015) and others related, the constraints had little effect on the end result[36].

After that, this article will continue to divide the result of the softmax layer classification (here represented by digits) to form an investment signal, that is, investment advice for each news title:

$$
signal(digit) = \begin{cases} Strong\ Buy, & digit > 0.5 + 0.4 \\ Buy, & 0.5 + 0.05 < digit \le 0.5 + 0.4 \\ Unknown, & 0.5 - 0.05 < digit \le 0.5 + 0.05 \\ Sell, & 0.5 - 0.4 < digit \le 0.5 - 0.05 \\ Strong\ Sell, & digit \le 0.5 - 0.4 \end{cases} \quad (1)
$$

In the process of selecting a recommendation list, this article based on a company's news headlines within one day, comprehensively derived investment recommendations to recommend investors to "buy", "hold" or "sell". The logic for forming the recommendation list is as follows: we form it based on the maximum number of signals, where the number of Strong Buy / Sell is multiplied by 4, and the number of Unknown is divided by 4.

3.2.2 Long short-term Memory Network for Stock Price Prediction

In this article's stock price prediction method, the LSTM model is also designed to better analyze and predict the stock price trend (including short-term and long-term), investment advice, and calculate the income obtained by investing according to the investment advice to verify that the investment advice in this article is reliable.

Financial Data. As mentioned above, according to the downloaded ticket list, this article obtained the daily stock data of the corresponding company from "Yahoo! Finance," and the basic information of the data shown in Table 1. For the 507 companies on the ticket list, each company has 2424 daily data from 2009.12.31 to 2019.08.19.

TABLE1. THE DESCRIPTION OF THE INDEXES IN DAILY STOCK DATA

Index	Description
Date	Date of data
High	Maximum price for the corresponding date
Low	Lowest price for the corresponding date
Open	Opening price for the corresponding date
Close	Closing price for the corresponding date
Volume	Number of shares that are sold or traded for the corresponding date
Adj Close	Closing price adjusted for splits and dividend distributions for the corresponding date

Among them, this article selects the closing prices of several companies from 2014.01.01 to 2019.06.01. When training the model, this article uses "close" to express a company's close price on a certain date as a training feature.

News Data and Sentiment Data. This article uses the news data obtained from the Reuters Financial News section and the Reddit forum to analyze the sentiment of the news using the VADER (Valence Aware Dictionary for sEntiment Reasoning) algorithm model [37]. After using VADER, this article obtains the information contained in the news title. Scores corresponding to three emotional attributes: positive, negative, and neutral. In addition, this article also uses TextBlob [38] to measure the sentiment of news using subjectivity and polarity. Among them, subjectivity is used to measure the degree of subjectivity or objectivity of the text. The value range is [0, 1], where 0 is completely objective and 1 is completely subjective. Polarity is used to measure the degree of optimism or pessimism in the text. The value range is [-1, +1], -1 means very negative, +1 means very positive. This article takes 14 days and 7 days as the unit and uses the above two modules to obtain the news sentiment index of each company and Reddit forum, including the positive, negative, neutral, subjectivity, and polarity mentioned above. This article uses pos, neg, neu to represent positive, negative, and neutral respectively.

When training the model, this article uses stock_polarity, stock_subjectivity, stock_pos, stock_neg, stock_neu to represent the five sentiment indicators of Reuters financial news, and market_neg, market_neu, market_pos, market_polarity, and market_subjectivity to represent the five sentiment indicators of Reddit forum news.

Economic Indicators Data. This article downloads three indicators from the OECD website, all of which describe the economic phenomenon in one month. This paper summarizes these three data into a table. Some data samples in the table are shown in Table2 below.

When training the model, this article uses CCI, BCI, and CLI to represent three economic indicators.

TABLE2. THE SAMPLE DATA OF THREE ECONOMIC INDICATORS

TIME	CCI	BCI	CLI
2014/1/1	99.31528	100.183	100.5739
2014/2/1	99.3845	100.2308	100.6074
2014/3/1	99.41634	100.3332	100.6685
2014/4/1	99.48846	100.4209	100.7376
2014/5/1	99.47501	100.4788	100.7984

The Model. For each company, this article uses its close, stock_polarity, stock_subjectivity, stock_pos, stock_neg, stock_neu, CCI, BCI, CLI, market_neg, market_neu, market_pos, market_polarity, market_subjectivity as the features of training, verification, and test set samples. The structure of our LSTM model is roughly shown in Figure 1: the training sample data is 915 from 2014.01.01 to 2018.01.17, the verification sample data is 135 from 2018.01.17 to 2018.06.01, and the test sample is the remaining 100 data from 2018.06.01 to 2019.06.01 (if no validation set is used, the test sample is 235 data from 2018.01.17 to 2019.06.01) Using the above data, this article predicts the closing price of the stock price.
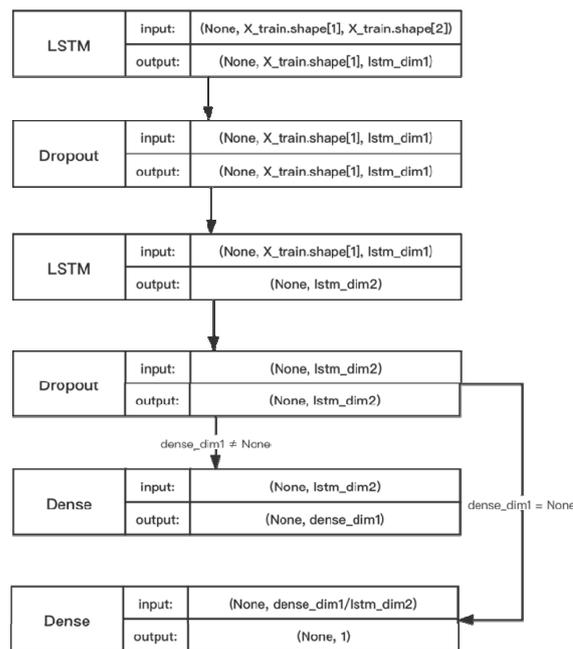
Figure 1. The Structure of LSTM Model

This article draws the following four images (all based on the test set) to better test the model and guide consumers' decisions more clearly and intuitively. The specific steps for drawing an image are as follows:

- Plot the stock price trend in units of one day. This article compares the predicted closing price of each day with the closing price in the test set to obtain the final image.
- Draw a stock price trend chart for a time period. This article uses a period of time (the default is

30 days) to draw an image and compare it with the closing price in the test set.

- Draw buying and selling points. Based on the basic "buy low and sell high" principle of the stock market, this article decides the buying and selling strategies as follows: Given a date, this article first predicts the price that it will be for a period of time (the default is 30 days) in the future, gets a series of prices, and finds out the predicted maximum price and minimum price from it. Then, this paper calculates

$$\begin{cases} buyreturn = \dfrac{maxprice - startprice}{startprice} \\ sellreturn = \dfrac{startprice - minprice}{startprice} \end{cases} \quad (2)$$

Among them, startprice is the stock price on a given date, and maxprice and minprice are the maximum and minimum prices of the price series, respectively.

Subsequently, this article calculates whether the given date can be used as a buy / sell point based on the size relationship between buyreturn and sellreturn and the set threshold (represented by return_threshold).

$$decision(buyreturn, sellreturn, return_{threshold}) = \begin{cases} Buy, & buyreturn \geq sellreturn \\ & or\ buyreturn \geq return_{threshold} \\ Sell, & sellreturn \geq buyreturn \\ & or\ sellreturn \geq return_{threshold} \\ None, & else \end{cases} \quad (3)$$

Among them, this article uses red to indicate the selling point and green to indicate the buying point, which is plotted on the closing price curve.

- Draw the consumer income chart. This article first uses the algorithm for calculating buy and sell points to determine the buy and sell time points within the time range of the test set data. This article sets the number of shares purchased / sold each time as stock_per_trade (fixed value). It uses portfolio to represent the total number of shares at the time of buying / selling, and cash to represent the amount of money at the time of buying / selling.

When buying / selling, this article calculates

$$\begin{cases} portfolio = portfolio \pm stock\_per\_trade \\ cash = cash \mp stock\_per\_trade * price[date] \\ curr\_value = cash + portfolio * price[date] \\ curr\_return = 100 * \dfrac{curr\_value - initial\_cash}{initial\_cash} \end{cases} \quad (4)$$

In this paper, curr_return is used to draw the consumer income chart in order of date to observe the accuracy of the model more intuitively.

3.2.3   Other Features

In other respects, this article also does some work to help consumers better invest.

This article not only provides the regular K-line chart and line chart display, but also provides a fuzzy search for stocks. You can find the corresponding company in

the fuzzy search bar below by the company code; After a stock is added as a favorite, consumers can find the corresponding stock in their favorite list, which is convenient for consumers to browse the stocks they are interested in.

4   Findings and Results

4.1 The Evaluation of CNN Model

4.1.1   Word Cloud Diagram

For the news of different companies, this article extracts the keywords, and uses different colors and text sizes to create a word cloud diagram. The word cloud diagram of AAPL company is shown in Figure 2.

Aiming at the word cloud diagram, consumers can judge the keywords that have been extracted without having to browse the main content of the news, which is convenient for consumers to look at the overall situation and have a overall understanding of the news.

4.1.2   The Accuracy of CNN Model

According to the evaluation, the prediction rate of a news headline, the accuracy of the CNN model is 58%. Among them, some companies' comprehensive investment opinions based on multiple news headlines are shown below. Table 3, Table 4, and Table 5 respectively represent three different investment opinions obtained by three different companies after analysis.

In actual practice, the probability of Strong Buy / Sell is very small, which may be related to the less and more objective information reflected in the news headline.

Figure 2. The Word Cloud Diagram of AAPL

TABLE 3. TSLA BUY/SELL DECISION

Headline	Signal
Tesla Waives $1	Unknown
500 Solar Removal Fee for Rest of September	Unknown
Tesla (TSLA) Stock Sinks As Market Gains: What You Should Know	Buy
The Pros Say No: 7 Large-Cap Stocks to Sell or Avoid	Buy
Safety Group Calls For Tesla Recall After NTSB Report	Unknown
Nio Stock Shows Little Potential for Getting	Unknown

Headline	Signal
out of Reverse	
Tesla and Lyft Have In Common?	Unknown
Chinese Electric Carmaker Nio Announces $200M Debt	Unknown
TSLA's Investment advice:: Buy	

TABLE 4. SBUX BUY/SELL DECISION

Headline	Signal
Coffee Has a Dire Week as Glut Sends Robusta to Nine-Year Low	Unknown
UPDATE 1-Former Starbucks CEO won't seek U.S. presidency in 'broken system'	Unknown
Coffee Snobs Are Shelling Out $3.25 for a Jolt of Instant	Unknown
The Pros Say No: 7 Large-Cap Stocks to Sell or Avoid	Buy
Starbucks to offer mental health benefits	Sell
Starbucks Trims 2020 Earnings Forecast	Unknown
Howard Schultz Won't Be Running For President	Unknown
SBUX's Investment advice:Unknown	

TABLE 5. MMC BUY/SELL DECISION

Headline	Signal
Marsh & McLennan (MMC) Down 1.1% Since Last Earnings Report: Can It Rebound?	Sell
A Holistic Look At Marsh & McLennan Companies	Unknown
What Type Of Shareholder Owns Marsh & McLennan Companies	Unknown
Why You Should Keep Marsh & McLennan (MMC) in Your Portfolio	Buy
Top Analyst Reports for Caterpillar	Sell
Edited Transcript of MMC earnings conference call or presentation 30-Jul-19 12:30pm GMT	Unknown
MMC's Investment advice: Sell	

4.2 The Accuracy of LSTM Model

For different companies, this article has generated four kinds of images, using the model's RMSE as the standard, and obtained relatively reasonable parameters for each company's LSTM model to achieve better results; in four images, the former two images can more intuitively reflect the timely and overall trend of the company's stock price, and the latter two images can reflect the effect of the more comprehensive forecasting decision. Some images are shown as below.

4.2.1 Tuning Parameters

For the architecture given by Figure 1, this article sets a few values for the parameters that appear in them: lstm_dim1, lstm_dim2, dropout, dense_dim1, and lookback values. For different companies, within these value ranges, they all have the lowest RMSE. These

values make the parameters relatively reasonable for the model, which will better reduce the model's errors.

4.2.2 Four Kinds of Prediction Curves

In the stock price trend chart in one day, the forecast curve for the stock price is purple, and the actual stock price trend is red; in the stock price trend chart for a period of time, the actual stock price trend is black, and the stock price trend in the forecast stage is colored. The test sample without TSLA company is used as a sample. The two images are drawn in Figure 3 and Figure 4. Figure 3 shows the stock price trend of TSLA in one day. Figure 4 Shows 30-day TSLA stock price charts.

It can be seen that if TSLA stocks are bought / sold in this period according to the decision-making method described in this article, a very good return can be achieved. Among them, Figure 5 shows the situation of TSLA buy / sell points, and Figure 6 shows the profit that can be obtained by operating the buy / sell points shown in Figure 5.

5 Discussion

5.1 Theoretical Discussion

Based on the above, the contributions of this article in academic theory are:

5.1.1 For the pre-processing of constructing sentiment analysis models, this paper construct sentiment analysis models for specific types of text content

The current application area of sentiment analysis technology is very wide, while it lacks the pertinence of the scene to this technology' application. The sentiment analysis method of news in a specific field is different in this article. Emotional models are often used directly, and the positive and negative emotions given directly by crawling news are not accurate.

For example, when this article analyzes financial news, this article must be particularly sensitive to certain words in financial news. if a piece of news shows that a certain company is involved in "sanctions" of the country, or involved in a "trade war" between countries, then the company's stock price will certainly be greatly affected, and then the sentiment of such news is also very negative. The original sentiment analysis model has difficulty perceiving specific news feelings.
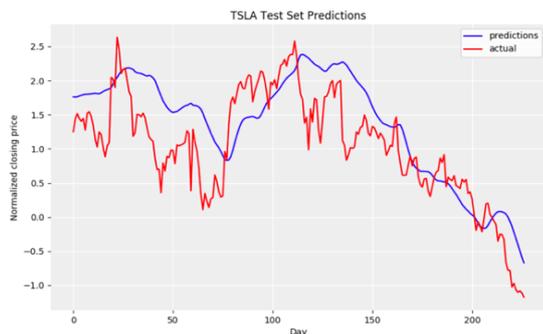
Figure 3. One-day curve for TSLA Test Set
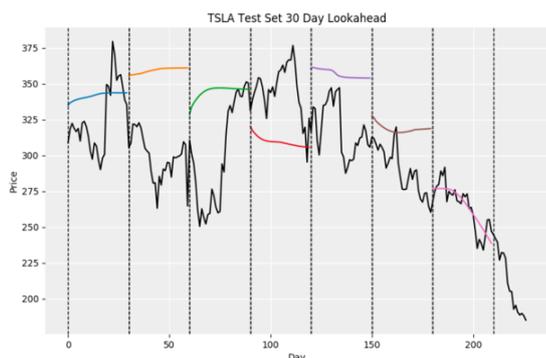*(stock price chart in one day, more timely)*

Figure 4. 30-days curve for TSLA Test Set *(based on stock price chart in units of 30 days, the comprehensive trend over a period of time could be seen)*
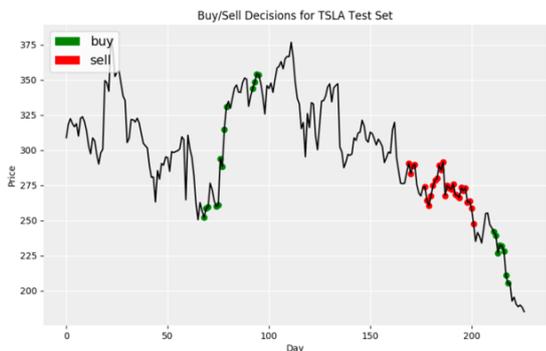

Figure 5. Bull-Sell Points Curve for TSLA Test Set
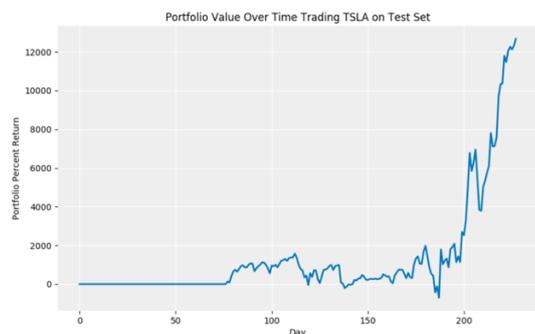*(buy and sell points chart to see if "buy low and sell high")*

Figure 6. Portfolio Value Curve for TSLA Test Set

Therefore, this article needs to increase the sensitivity to certain words when constructing the model,

so that the sentiment analysis model obtained in this article is more accurate and can better predict the rise and fall of stock prices.

5.1.2 Based on short-term situation, this paper provides a way to make investment decisions in a timely manner

In this paper, through the relationship between the stock price fluctuations and news of the day, we trained a CNN network that can comprehensively extract and analyze news in a timely manner. Through a comprehensive analysis of the news that occurred on the day (or before the stock market opened), we can obtain investment decisions with relatively high accuracy, which can free people from the complicated analysis of news content, find the keywords of the news, and obtain relative reasonable decision-making, cooperate with one's own judgment, and make the final investment action.

5.1.3 Based on Long-term situation, this paper can comprehensively cooperate with news sentiment and financial indicators to predict the stock price trend more accurate

By combining financial and economic indicators, world forums, and the sentiment of financial news, this article more accurately predicts the long-term (about 230 days) trend of the stock. Among them, using the LSTM model to analyze stock trends through news sentiment and economic changes over time according to time series can bring some novel ideas to the prediction of stock prices. Combining the two predicted stock price charts, investors can also see whether the model is accurate from a micro and relatively macro perspective.

5.1.4 This paper provides a new method to check whether the model is accurate

This paper analyzes and checked the predicted price and the buying and selling points by indirectly analyzing the predicted prices and buying and selling points, as well as by simulating the method of calculating the returns on the buying and selling dates and drawing the yield curve, which is also a very smart test.

5.2 Recommendation

Based on the above information, the contributions of this article in practical applications are in following:

5.2.1 This paper can provide investment decisions in time to facilitate investors' decision

As ordinary people, there may be work and study in daily, who will not have a timely insight into the stock market as people in the financial industry. The model can provide investment advice in a timely and convenient manner, which can save time reading news

and investment analysis, and facilitate various types of investors to make decisions.

5.2.2 *This paper is able to make investment choices intuitively and observe the effect of investment*

It is sometimes difficult for investors to make investment choices. After making investment choices, they may not have a very intuitive feeling about their investment effects. By observing the buy / sell point chart and the consumer income chart, they can directly find out when to make investments and the income from investing at these times, which greatly gives consumers investment confidence and helps investors make better decisions.

6 Summary

This article tells how to obtain news from listed companies, how to analyze news scientifically and give sentiment analysis models for financial news, how to integrate multiple comprehensive indicators of sentiment and economy to predict stock prices, make purchases and sell decisions and analyze the benefits people can get from making them.

In the future, this article will continue to look for more scientific financial news analysis methods and financial indexes, and train better sentiment analysis models and stock price prediction models to facilitate people to make more reasonable investment decisions. The article also hope to obtain people's real investment ideas and decisions to form investment recommendations for specific investors.

References

1. Eugene F. Fama, 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance* 25(2):383-417.

2. Mark L. Mitchell, J. Harold Mulherin, 1994, The Impact of Public Information on the Stock Market, *The Journal of Finance* 49(3):923-950.

3. Peter Klibanoff, Owen Lamont, Thierry A. Wizman, 1998, Investor Reaction to Salient News in Closed-End Country Funds, *The Journal of Finance* 53(2):673-699.

4. Gur Huberman, Tomer Regev, 2001, Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar, *The Journal of Finance* 56(1):387-396.

5. Takeda F, Yamazaki H, 2006, Stock Price Reactions to Public TV Programs on Listed Japanese Companies, *Economics Bulletin* 13(7):1-7.

6. Qing Li, Tiejun Wang, Ping Li, Ling Liu, Qixu Gong, Yuanzhu Chen, 2014, The Effect of News and Public Mood on Stock Movements, *Information Sciences* 278(9):826-840.

7. David Godes, Dina Mayzlin, 2004, Using Online Conversations to Study Word-of-Mouth Communication, *Marketing Management* 23(4):545-560.

8. Judith A. Chevalier, Dina Mayzlin, 2006, The Effect of Word of Mouth on Sales: Online Book Reviews, *Journal of Market Research* 43(3):345-354.

9. Qiang Ye, Rob Low, Bin Gu, 2009, The Impact of Online User Reviews on Hotel Room Sales, *International Journal of Hospitality Management* 28(1):180-182.

10. Chrysanthos Dellarocas, 2005, Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard, *Information Systems Research* 16(2):209-230.

11. Susan M. Mudambi, David Schoff, 2010, What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com., *MIS Quarterly* 34(1):185-200.

12. Wendy W. Moe, David A. Schweidel, 2011, Online Product Opinions: Incidence, Evaluation, and Evolution, Marketing Science 31(3):372-386

13. Ronen Feldman, Ido Dagan, 1995, Knowledge Discovery in Textual Databases (KDT), Proceedings of the First International Conference on Knowledge Discovery and Data Mining(KDD-95): 112-117.

14. Ah-Hwee Tan, 2000, Text Mining: The State of the Art and the Challenges, Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases:65-70.

15. Tarjei Lægreid, Paul Christian Sandal, 2006, Financial News Mining: : Extracting useful Information from Continuous Streams of Text.

16. Robert Schumaker, Hsinchun Chen, 2006, Textual Analysis of Stock Market Prediction Using Financial News Articles, Americas Conference on Information Systems.

17. Werner Antweiler, Murray Z. Frank, 2006, Do Us Stock Markets Typically Overreact to Corporate News Stories?, *SSRN eLibrary.*

18. Lewis, David D., 1998, Naive (Bayes) at Forty: The Independence Assumption in Infor- mation Retrieval, in Proceedings of the 10th European Conference on Machine Learning (ECML), Springer-Verlag, New York.

19. Paul C. Tetlock, 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance* 62(3):1139-1168

20. Paul C. Tetlock, Maytal Saar-Tsechansky, Sofus A. Macskassy, 2008, More Than Words: Quantifying Language to Measure Firms' Fundamentals, *The Journal of Finance* 63(3):1437-1467.

21. Casey Dougal, Joseph Engelberg, Diego Garcia, Christopher A. Parsons, 2011, Journalists and the Stock Market, *Review of Finance Studies* 25(3):639-679.

22. Zachery C. Lipton, John Berkowitz, 2015, A Critical Review of Recurrent Neural Networks for Sequence Learning, arXiv: 1506.00019 [cs.LG].

23. Donald C. Wunsch, Emad W. Saad, Danil V. Prokhorov, 1998, Comparative Study of St e Study of Stock Trend Prediction using Time Dela ediction using Time Delay, Recurrent and Probabilistic Neural Networks, *IEEE transactions on neural networks* 9(6): 1456-1470.

24. Akhter Mohiuddin Rather, Arun Agarwal, V. N. Sastry, 2015, Recurrent neural network and a hybrid model for prediction of stock returns, Expert Systems with Applications 42(6): 3234-3241.

25. S. Hochreiter and J. Schmidhuber, 1997, Long short-term memory, Neural computation 9(8): 1735–1780.

26. Şakir Görmüş, Sevcan Güneş, 2010, Consumer confidence, stock prices and exchange rates: The case of Turkey, *Applied Econometrics and International Development* 10(2):103-114.

27. Hashmat Khan, Santosh Upadhayaya, 2017, Does Business Confidence Matter for Investment**?**, *Carleton Economic Papers* 17-13, Carleton University, Department of Economics, revised 20 Mar 2019.

28. Jan Dovolil, 2016, The Use of Economic Indicators as a Tool for Predicting S&P 500 Stock Index, Acc Journal 22(2):7-14.

29. Wenbin Zhang, Steven Skiena, 2010, Trading Strategies to Exploit Blog and News Sentiment, *Proceedings of the Fourth International Conference on Weblogs and Social* Media.

30. Kira Radinsky, Sagie Davidovich, Shaul Markovitch, 2012, Learning Causality for news Events Prediction, *Proceedings of the 21st International Conference on World Wide Web*, ACM, pages 909–918.

31. Xiao Ding, Yue Zhang, Ting Liu, Junwen Duan, 2014, Using Structured Events to Predict Stock Price Movement: An Empirical Investigation. *Proceedings of the 2014 Con- ference on Empirical Methods in Natural Lan- guage Processing (EMNLP)*. Association for Computational Linguistics, pages 1415–1425.

32. OECD (2020), Business confidence index (BCI) (indicator). doi: 10.1787/3092dc4f-en (Accessed on 05 February 2020).

33. OECD (2020), Consumer confidence index (CCI) (indicator). doi: 10.1787/46434d78-en (Accessed on 05 February 2020).

34. OECD (2020), Composite leading indicator (CLI) (indicator). doi: 10.1787/4a174487-en (Accessed on 05 February 2020).

35. Yoon Kim, 2014, Convolutional Neural Networks for Sentence Classification, EMNLP, 1746-1751.

36. Ye Zhang, Byron Wallace, 2015, A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification, arXiv:1510.03820 [cs.CL].

37. C. J. Hutto, Eric Gilbert, 2015, VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, Eighth International AAAI Conference on Weblogs and Social Media.

38. S. Loria, 2014, Textblob: simplified text processing, Secondary TextBlob: Simplified Text Processing.