Received: 29 February 2020 | Revised: 22 September 2020 | Accepted: 16 October 2020
DOI: 10.1002/cpe.6076

RESEARCH ARTICLE

WILEY

Stock movement prediction with sentiment analysis based on deep learning networks

Yong Shi[1,2,3,4] | Yuanchun Zheng[2,3,5] | Kun Guo[1,2,3] | Xinyue Ren[2,3,6]

[1]School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China
[2]Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing, China
[3]Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China
[4]College of Information Science and Technology, University of Nebraska at Omaha, Omaha, Nebraska, USA
[5]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China
[6]Sino-Danish College, University of Chinese Academy of Sciences, Beijing, China

Correspondence
Kun Guo, School of Economics and Management, University of Chinese Academy of Sciences, No. 80 of Zhongguancun East Street, Haidian District, Beijing 100190,China.
Email: guokun@ucas.ac.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 71501175, 71932008, 91546201; Fundamental Research Funds for the Central Universities, Grant/Award Number: E0E48944X2

Abstract
With the development of Internet and big data, it is more convenient for investors to share opinions or have a discuss with others via the web, which creates massive unstructured data. These data reflect investors' emotions and their investment intentions, and it will further affect the movement of the stock market. Although researchers have been attempted to use sentiment information to predict the market, the sentiment features used are driven by outdated emotion extraction systems. In this article, we proposed a new sentiment analysis system with deep neural networks for stock comments and applied estimated sentiment information to the stock movement forecasting. The empirical results showed that our deep sentiment classification method achieved a 9% improvement over the logistic regression algorithm, and provided an accurate sentiment extractor for the next predicting step. In addition our new hybrid features that mix stock trading data and sentiment information achieved 1.25% improvement among 150 Chinese stocks in the testing dataset. For American stocks, the sentiment information would reduced the predicting results. We found that emotion features extracted from comments are indeed effective for stocks with a higher price to book value and a lower beta risk value in China.

KEYWORDS
deep learning, financial comments, sentiment analysis, stock movement prediction

1 | INTRODUCTION

Sentiment analysis is one of the most important research directions in artificial intelligence and has been applied in many fields, such as stock predicting, political science, and health science.[1,2] Among these applications, stock movement prediction has wide commercial potential and market prospects. However, the fluctuation of the stock market is closely related to the development of the national economy, and the market itself is also a complex nonlinear dynamic system that is susceptible to many factors.[3] Due to the complexity of market rules, the volatility of prices and the diversity of factors also affect the market. Generally, there are approximately two basic methods for stock prediction. The first is a kind of qualitative analysis method in which the effectiveness of forecasts depends largely on the capabilities and experience of experts. The other is the technical analysis method, which contains statistical methods and a data mining algorithm. According to academic studies and practical applications, methods relying solely on expert strategic analysis are insufficient to improve the prediction accuracy. Therefore, researchers have begun to pay more attention to the multisource heterogeneous data and to introduce deep learning methods into this field.

In recent decades, the rapid development of the mobile Internet and online applications has brought the whole world into the big data era. It is more convenient for investors to share their opinions or have a discussion with others by pushing messages. People are also likely to browse the web before making investment decisions. All these activities provide multisource data for scientific research. Related studies on behavioral finance in recent years have also proven that sentiment can interfere with investment decisions, which makes stock forecasting reasonable with sentiment analysis. Therefore, different methods for predicting the stock price and movement have been exploited, and extra information, such as attention and sentiment extracted from news and social media, has also been treated as auxiliary features.

However, compared with foreign mature financial markets, the financial market in China is still in the development stage, and individual investors still play an important role in the stock market. Thus, the study of the intention of individual investors, will greatly improve the overall control of the stock market. In this article, a new sentiment analysis system with a deep neural network was built from scratch. Compared with the logistic regression (LR) method, the new system achieved a 9% improvement in the sentiment classification. The sentiment information estimated by the analysis system was also further applied to the stock movement prediction task and obviously improved compared with the trading data only input methods.

The remainder of the article is organized as follows. Section 2 is a review of sentiment analysis and stock return prediction. Section 3 proposed the new sentiment analysis method based on deep neural networks and explored stock movement prediction with sentiment information. Section 4 gives some experiments and compares the difference between raw trading input data and hybrid sentiment data. Section 5 provides our final conclusion and some discussion on the Chinese stock market.

2 | RELATED WORK

Sentiment analysis has drawn much attention in many fields and become one of the most attractive topics in academia and the stock market. In addition, with the development of research, new models, such as deep learning networks, were introduced to sentiment analysis to predict the movement of the stock market.

2.1 | Stock prediction methods

Stock data are a kind of time series data full of complex factors that makes it difficult to predict. Researchers proposed different predicting methods based on autoregressive integrated moving average (ARIMA). For instance, Ariyo et al. used the ARIMA model to forecast stock prices, and the results showed that this model is suitable for short-term stock price prediction.[4] However, there are still unpredictable emergencies in the stock market that cause aperiodic fluctuations. Therefore, Adebiyi et al. tested the forecasting performance of the ARIMA and artificial neural network (ANN) models, and the results indicated that the ANN model achieved better performance.[5] With an increasing number of machine learning algorithm introduced to predict stock prices and trends, some scholars have made comparisons between different machine learning algorithms. Mittal and Goel[6] combined the machine learning method with sentiment analysis and used LR and support vector machine (SVM) models with public mood which were estimated from Twitter data to predict stock market movements. Ballings et al.[3] compared the performance between ensemble methods and single classifier models and found that random forest had the best result. Weng et al. used three different machine learning methods, including decision tree, neural networks and SVMs to predict AAPL stock with combined disparate online data sources and traditional time-series indicators. The result shows that these disparate data sources can improve the performance of financial expert systems.[7] Some researchers also studied the behavior of the hybrid model.[8-10] Recently, an increasing number of neural network models are introduced into this field. Ince and Trafails[11] compared the performance between support vector regression (SVR) and multilayer perceptron (MLP) and illustrated that even though they require different inputs, the MLP networks are better than the SVR method. Ding et al.[12] extracted events from the news and trained dense event vectors with neural networks. In addition, they obtained nearly 6% improvements in the S&P index prediction with a convolutional neural network (CNN). Akita et al.[13] used long short-term memory (LSTM) for financial time series forecasting, and the results showed that LSTM outperforms SVM and the LSTM unit was able to capture long-term comments. Zhou et al.[14] also adopted a LSTM network in the CSI300 prediction with Baidu search volume. However, neural network methods do not always achieve better performance than traditional machine learning-based methods. Atsalakis and Valavanis[15] summarized the current development of conventional techniques used to predict stock market price and concluded that conventional forecasting techniques are still promising for future research. Chiong et al.[16] also found that their SVM-based method achieved better performance than deep learning-based methods. Different models provide different ideas for stock market forecasting, but based on the volume of daily stock data, machine learning especially SVM is the most appropriate model for stock movement prediction.

In addition to research on the stock market in developed countries, scholars have attempted to predict Chinese stocks with local social medias. Chen et al.[17] proposed a Weibo emotion mining approach to predict stock market price and noted that financial decisions are significantly driven by emotions. Guo et al. collected comments and tweets from the Chinese stock market website and proposed a thermal optimal path to analyze the relationship between investor sentiment and the stock market. They found that predictions may fail when the stock has a low investor attention.[18] This approach also exposed the difference between the Chinese stock market and the American stock market when predicting the movement with sentiment information. Oliveira et al. utilized several machine learning-based methods for stock market prediction with the sentiment and attention extracted from microblogs and survey indices. Their results showed the usefulness of social media sentiment for financial expert systems.[19] Zhou et al.[20] studied over 3 million stock-relevant investors from Weibo, and found that five stock attributes in China can be predicted by emotions. They believed that the Chinese market might be more emotional than its western counterparts.

2.2 | Sentiment extraction methods

In addition to innovation in prediction methods, research on the algorithm input itself is also constantly evolving . With the rise of social networks and stock review websites in recent years, scholars have attempted to utilize sentiment information as an auxiliary input for stock movement prediction. Therefore, different sentiment extraction methods have been proposed, including using existing extraction tools, using search engines to obtain emotions, or building an emotion analysis engine. Bollen et al.[22] analyzed the context of tweets by two mood tracking tools referred to as the OpinionFinder and Google-Profile of Mood States. The first tool measured positive and negative moods while the second tool measured moods with 6 dimensions (calm, alert, sure, vital, kind a happy). Their results of the investigators showed that these public mood extractors can significantly improve the accuracy of the Dow Jones industrial average significantly. Hagenau et al.[23] enhanced an existing text mining methods and used more expressive features for stock price prediction. Wang et al.[27] proposed an novel algorithm with multiple strategies and could be used for numerical optimization for data processing, and Cui et al.[28] improved an optimization algorithm for many-object optimization problems, all these algorithm can be used in the stock prediction task. Li et al.[29] implemented a generic prediction framework and used two sentiment dictionaries to obtain news sentiment and obtained a better prediction than bag-of-word methods. Li et al.[24] proposed a trading signal mining platform that utilized an extreme learning machine to render stock price predictions with market news and historical price information. Based on their platform, accurate and fast prediction results were achieved. Pagolu et al.[21] used two different representation methods, word2vec and N-gram when extracting public sentiment in tweets. Chan and Chong[25] proposed a sentiment analysis engine for stock prediction based on linguistic analyses, they extracted collective expressions both at the word level and at the phrase level. The evaluation results showed that mood created by their engine can improve the predictive results. Akita et al. employed distributed representations of news articles and tried to use the correlation between multiple companies within the same industry to train the model with LSTM. The results showed that textual information is better than the numerical-data-only methods and bag-of-words-based methods.[13] Liu et al.[26] used the ensemble method as a sentiment extractor for microblog data and got a good result. In summary, sentiment information determines the final prediction results directly, and existing sentiment extractors are not guaranteed to achieve the same effect on a new stock dataset, and deep learning has shown great power in the field of nature language processing, so it is necessary to build a sentiment extractor with deep learning from scratch.

Table 1 is summary table about research goal, models, and dataset used in the related works. During the investigation, sentiment analysis with news and social media was found to significantly improve the predictive performance significantly. On the one hand, previous researchers used outdated sentiment extractors or rough analysis systems, which caused errors in sentiment analysis. These errors also led to the failure of stock market forecasting. On the other hand, the different trading environments and national conditions caused a large difference between the stock market in China and developed countries. There are only a few studies on the Chinese stock market based on sentiment analysis. Based on this, a new sentiment analysis system was built from scratch, and the latest deep learning technique was introduced into the system which further improves the performance of the entire stock movement prediction.

3 | METHODOLOGY

For the forecasting work, there are three important parts involved which indicated as Figure 1: the first part is a data collector with a series of web crawlers, and each crawler is used to collect different types of data from the Internet; the second main part is a sentiment analysis engine based on a deep learning network which is mainly used to provide extra sentiment information for individual stock; and the third part is a stock movement prediction system based on stock trading data and sentiment data.

3.1 | Data introduction

The entire system was built from scratch, and all datasets used in the system were also collected from scratch. As the most important foundation, a spider engine was designed to crawl stock comments from the web and four datasets have been proposed.

3.1.1 | Snowball dataset

Established in 2011, Snowball [1]is a representative financial online community in China. It provides comprehensive financial services including the real-time market, news information, investment strategy, and trading services. Users in this platform are mostly individual investors, and comments pushed by them would better reflect their investment intents and emotions. We designed a web crawler to collect webpages from the beginning

[1]http://xueqiu.com

TABLE 1 Summary of related works

Purpose	Model	Dataset	Reference
Model comparing	ARIMA and neural network	stock data (New York Stock Exchange)	5
	Ensemble methods	stock data (European companies)	3
Movement predicting	Neural network	stock data (DJIA values) and Twitter feeds	6
	SVM and neural network	stock data (NASDAQ index)	10
	Deep learning	textual data (S& P 500 and individual stocks)	12
	LSTM	search volume index (CSI300 on Baidu)	14
	SVM	textual data (news disclosures of German)	16
	Sentiment analysis	textual data (Twitter about S&P 500 index)	19
	Sentiment analysis	textual data (tweets in social media)	21
Price predicting	ARIMA	stock data (New York and Nigeria Stock Exchange)	4
	Decision tree, neural network, SVM	stock data (AAPL stock values)	7
	Neural network and decision tree	stock data (electron industry in Taiwan)	8
	Genetic algorithm and SVM	stock data (TCS and Infosys)	9
	PAC, SVR, and neural network	technical indicators and fundamentals	11
	LSTM	textual data (Nikkei newspaper) and stock data (10 companies from Nikkei 225)	13
	Topic sentiment analysis	textual data (social news from Weibo)	17
	Sentiment analysis	textual data (comments of Chinese stocks)	18
	Logistic regression and SVM	textual data (tweets from Weibo) and stock data	20
	Mood tracking tools and neural network	textual data (Twitter feeds)	22
	Feature selection	textual data (financial news)	23
	Sentiment tools	textual data (Financial news of Hong Kong)	24
Sentiment analysis	Sentiment analysis engine	textual data (financial texts)	25
	Random forest	textual data (microblog comments)	26
Sentiment analysis and movement predicting (ours)	Word2vec, deep learning	stock data and textual data (150 Chinese Stocks and 30 American Stocks))	

Abbreviations: ARIMA, autoregressive integrated moving average; DJIA, Dow Jones industrial average; LSTM, long short-term memory; SVM, support vector machine.

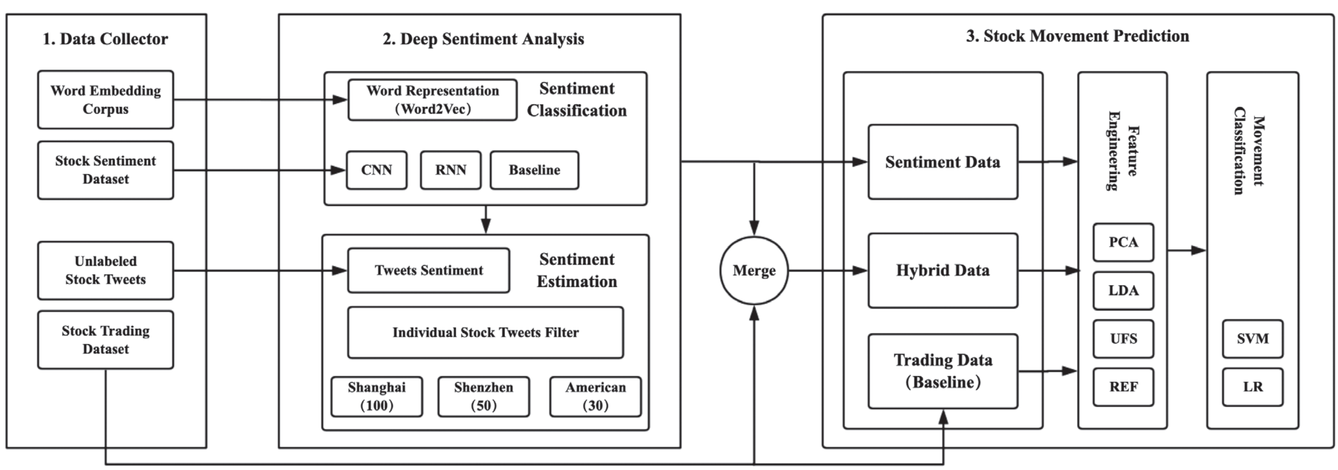

FIGURE 1 Pipeline of stock movement prediction

TABLE 2 Statics of financial comments in fixed format

Type	No.	Sentiment	Example
User	10380876	Positive/negative/neutral	Sanquan traditional sweet dumplings are very popular.
Rule1	5952793	Positive	I just followed **PFYH(SH600000)**, current price is 14.68.
			At 23.91 followed **SQSP(SZ002216)**
Rule2	395337	Positive	Buying **RMW(SH603000)** with 68.67
Rule3	176529	Negative	Selling **ZJYH(SH600036)** with 19.68
Rule4	15792	Neutral	Until 30 June 2015, my most profitable is …
Rule5	1635830	Positive	I just followed the snowball group **BFT-Top5(ZH002308)**
Rule6	134054	Positive/negative	I just changed the position of **ZQJQ(ZH000734)**.
Rule7	1158	Neutral	I just write my investment summary of 2014.
Rule8	19547	Neutral	Change the buying target price of **EDU** to $30.
Rule9	139916	Positive/negative	I just changed the group of **BFT-Top5(ZH002308)**

of 2012 to the middle of 2015 and obtained approximately 2,225,000 documents in total. Finally, data from 2014 to the middle of 2015 were chosen as the Snowball dataset, and each text was estimated to be a sentiment labeled positive/neutral/negative by our sentiment analysis methods. Some comments are pushed automatically when user triggered specific operations. For example, the Snowball platform will automatically push a comments such as “Someone following a stock at price -” when users add a stock to their account. Table 2 listed all types of stock comments.

3.1.2 | Stock sentiment dataset

The sentiment analysis was regarded as a kind of classification problem. To train a deep sentiment analysis engine, an annotated stock sentiment dataset must be created in advance. Based on our previous work, a labeled stock sentiment dataset was built with 9627 positive samples, 3234 neutral samples, and 4592 negative samples randomly chosen from the Snowball dataset. Each comment was marked by three students who majored in economics or finance, ensuring the effectiveness, and professionalism of the result.

3.1.3 | Word embedding corpus

Most comments in Snowball dataset are short texts because the properties of the platform itself. To train word embedding with a domain corpus, guba [2] was chosen as data source for the text corpus. It is a professional stock forum in the Chinese stock market, and it also provides real-time stock comments and experimental articles. This word embedding corpus has approximately 42.2 million sentences and covers approximately 31 thousand words in dictionary. Another open domain pretrained word embedding provided by Tencent[30] was also tested in this article.

3.1.4 | Stock trading dataset

For the basic input of the predicting model, opening access trading information including the open and close price, highest and lowest prices, turnover rate, and trading volumes are downloaded from the web. After sorting the number of comments of all stocks in three different exchanges, the top 100 stocks in the Shanghai exchanges, the top 50 stocks in Shenzhen exchanges and the top 30 stocks in American stocks are chosen as the final stocks, and the remainder are those with only a few comments.

3.2 | Sentiment analysis with deep learning

There are two major differences between deep learning-based sentiment analysis and past approaches: (i) the dense embedding input layer and (ii) multiple hidden layers.[31] Texts and documents need to be converted into dense numerical vectors before being initialized as the input layer of

[2] http://guba.com.cn

deep learning. Compared with the sparse one-hot encoding method, word embedding learned semantic information in vector space, and reduced the computational time which is really important for practical applications. At the same time, more hidden layers enhanced the ability of feature extraction and transformation of deep learning. This benefits from the nonlinear activation function of neuron units. In the neural network, each neuron unit can be regarded as a feature processor. Original text information is transformed layer by layer until a decision is made at the final output layer. In general, higher layers obtain more abstract features than lower layers. For example, a hidden layer close to the input layer may capture the character level or syntactic feature, and a hidden layer close to the output layer may learn semantic and pragmatic level information.

3.2.1 | Word representation

Word representation learning bridges the gap between deep networks and text representation and uses a neural network to calculate the joint probability of the n-gram language model. Assume $S = \{w_1, w_2, \dots w_n\}$ is one text from the stock datasets and language models aim to predict target word with context words. In this case, we use $C(w)$ to represent the context words of w, these words may be all previous $n-1$ words, or adjacent $2k$ words in word2vec where k is the context window size. The goal is to maximize its joint probability to ensure that the sentence is correct. Euq 1[32] shows the object function of the language model. Unlike other fields, word embedding is also a parameter of the neural network language model and needs to be learned by back propagation.

$$P(S) = \sum_{i=1}^{n} p(w_1)p(w_i|w_1, w_2, \dots , w_n) = \sum_{i=1}^{n} p(w_i|C(w_i)). \tag{1}$$

3.2.2 | Sentiment classification

Word2vec is an implementation toolkit of word representation learning and provides efficient text conversion for deep sentiment analysis. Three different sentiment classification models including a CNN, a recurrent neural network (RNN), and LR were chosen as the candidate sentiment analysis models. Among them, the LR method is one of the most commonly used machine learning-based methods, so we choose this model as a baseline model.

A CNN is a simple and efficient model for text classification. This model is derived from computer vision, different image patterns are extracted by convolutional layers and the most important features are filtered by a pooling strategy. In fact, convolution kernels play the same role as image operators, but the kernel is learned during the training process, while the operator is designed manually. When deep convolution is applied to sentiment classification, the biggest difference is that sentence belongs to sequence data, so convolution kernels must be shrunk to one-dimensional filters. Figure 2 indicates the structure of the parallel CNN network that we used. From the diagram, each word in the text was replaced by word embedding (the vector size is 4 in this example). Four different convolution kernels (kernel size is $h = \{2, 2, 3, 3\}$) move from the sentence start to the end, and four different sentence representations are obtained after the first convolutional layer, which has been drawn by single colored rectangles. These results represent different features of the original sentence. Then, the pooling layer filters the most distinctive value and a latent vector (the multiple color rectangle) is composed, this latent vector is a sentiment representation of the raw comments. The second half of the model is a neural network with multiple layers of output.

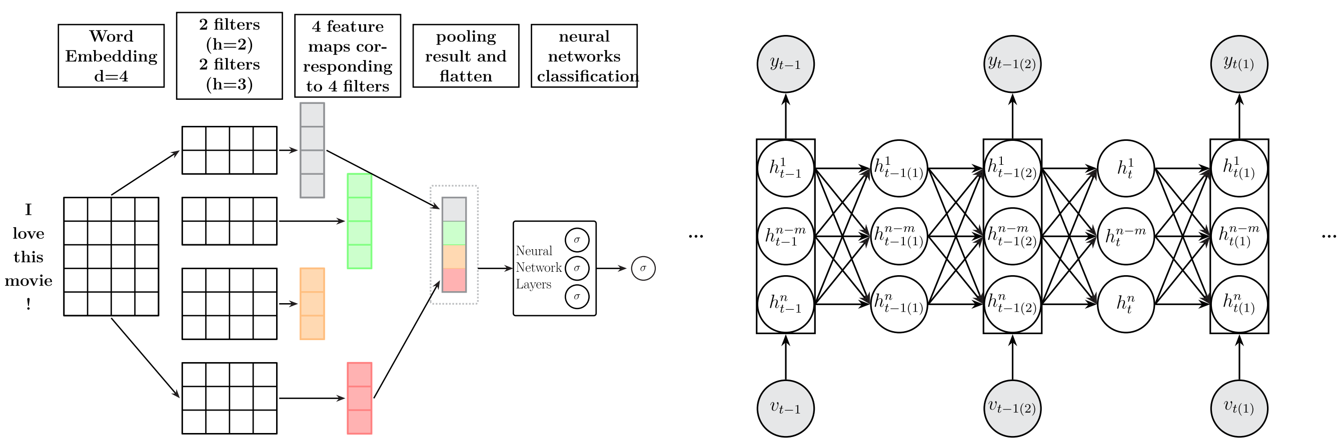

FIGURE 2 Diagram of sentiment classification models (left: text CNN, right: RNN). CNN, convolutional neural network; RNN, recurrent neural network

CNNs can capture different n-gram information and select the most significant features for classification. However, these networks ignore word order which is important for emotional expression. To solve this problem, another neural network called the RNN was introduced. An RNN treats each word as timestamps in the sequence sentence, and uses the same network at each point. As shown in Figure 2,[33] each hidden layer not only receives the input layer, but also receives the output of the hidden layer at last moment. Therefore, the RNN has the ability to remember historic information. Equation (2)[33] is the mathematical expression of a common RNN structure where x_t represents the word vector of w_t (t means it is the tth word in a sentence), **U** is a transformation matrix between input layer to hidden layer. h_{t-1} is the information from previous word, and **W** is another transformation matrix between hidden layer. In fact, h_t explains why the RNN is called a recurrent network.

$$\begin{aligned} h_t &= \sigma_1(\mathbf{W} \cdot h_{t-1} + \mathbf{U} \cdot x_t), \\ y_t &= \sigma_2(\mathbf{V} \cdot h_t). \end{aligned} \tag{2}$$

However, the recurrent structure in the RNN leads to gradient vanishing or exploding problem during backward propagating along time direction. One solution is to replace neurons with gated recurrent unit (GRU). As shown in Equation (3),[33] it introduced two gates named reset gate r_t and update gate z_t. Utilized sigmoid activation function, these two gates control the flow of data through whole network. r_t is used to control how much historic memory needs to be kept, and nonhistoric information will be left when it was reduced to zero. z_t controls how much information needs to be forgotten from the last hidden layers.

$$\begin{aligned} r_t &= \sigma_1(\mathbf{W}_z \cdot [h_{t-1}, x_t]), \\ z_t &= \sigma_2(\mathbf{W}_r \cdot [h_{t-1}, x_t]), \\ \hat{h}_t &= \tanh(\mathbf{W} \cdot [r_t * h_{t-1}, x_t]), \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \hat{h}_t. \end{aligned} \tag{3}$$

3.2.3 | Sentiment estimation

The well trained deep sentiment classification model was used to estimate all other unlabeled stock comments in the Snowball dataset. In addition, people's attention to stocks and the total number of comments in the chosen period (2014 to middle of 2015) are very different according to the popularity of stock trading. The number of comments are sorted for stocks in different exchange and ensured an average of more than 10 reviews per trading day for each stock. Finally, 100 stocks in Shanghai, 50 stocks in Shenzhen and 30 stocks in American were selected.

As indicated in Figure 3, three stocks in the Shanghai, Shenzhen, and American stock exchange are randomly chosen to show their statistical information and relationship between the sentiment indicators and the closing price. Subfigures in the left part describe the comment number distribution in 24 h, the blue points represent all types of comments including user pushed comments and automatically pushed comments indicated in Table 2, and the red points only cover user pushed comments. It can be learned from these three subfigures that investors always become active an hour earlier than the market opening, and they become busy completing their respective trading plans during the opening time. These comments, either manual or automatic messages, provide real and effective information for our prediction research. Another three subfigures in the right part show the relationship between different sentiment indicators and close price. The green line represents the actual closing price on every trading day, and the other three lines represent the total number of comments and the number of positive and negative comments on that day, respectively. Compared with American stocks, the sentiment information of Chinese stocks are more strongly correlated with the closing price. The correlation coefficients between the closing price and daily positive comment numbers are 0.819 (for Shanghai stocks), 0.725 (for Shenzhen stocks), and 0.066 (for American stocks).

3.3 | Stock movement prediction

3.3.1 | Feature engineering

In most machine learning applications, data preprocessing is important because there are some shortcomings in raw data, such as missing values, invalid data, and noise data. Usually, data standardization and normalization, data binarization and type conversion, missing values padding, and denoising are all belong to data preprocessing. Then, feature engineering is an important chain in machine learning pipelines, which aims to improve model performance by transforming features or creating new features. In addition, feature engineering will reduce the input dimension and prevent the overfitting of the training model so that it can greatly accelerate the training speed.

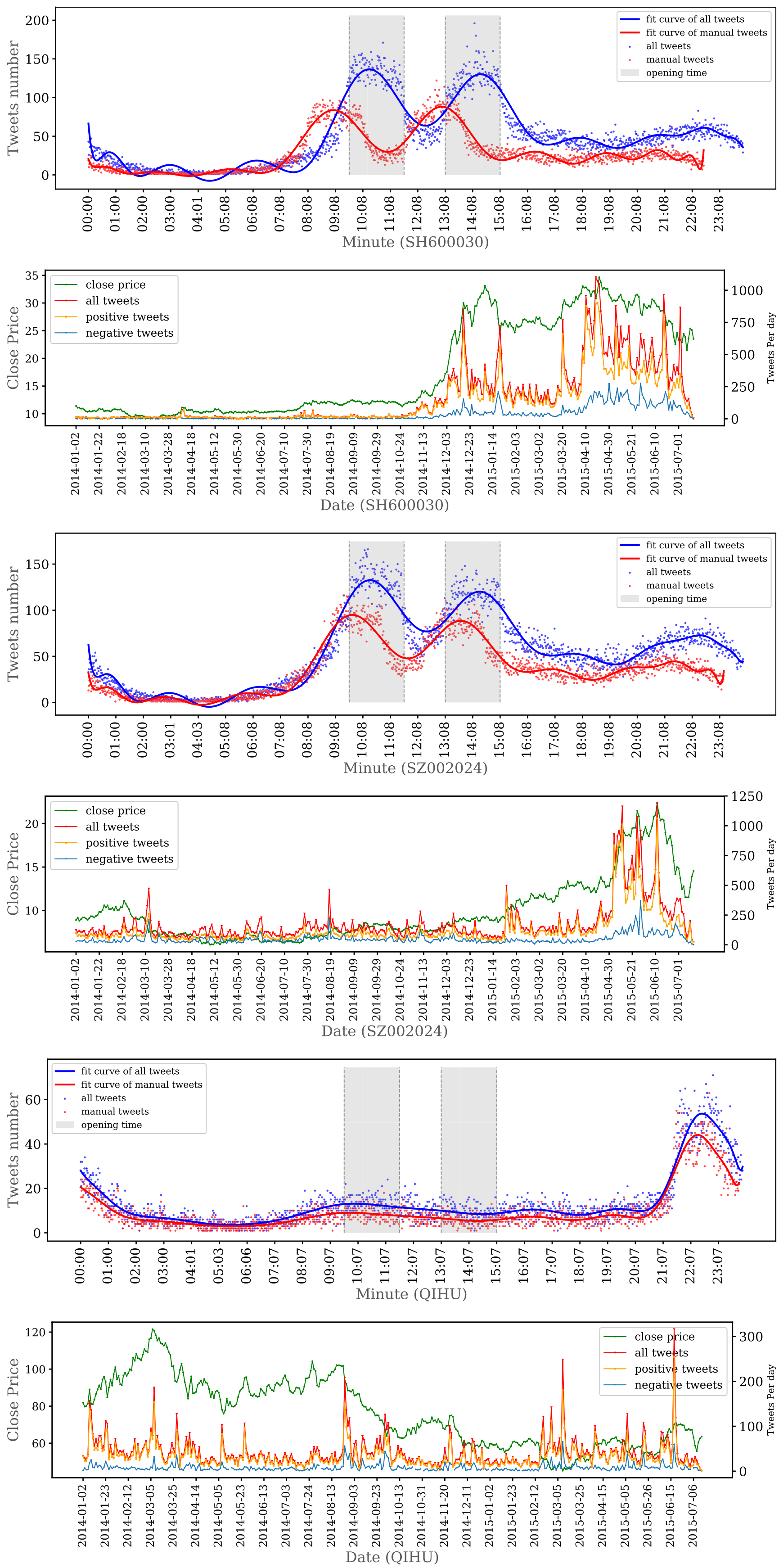

FIGURE 3 Statistics of comments number in minutes-level and its daily comments trend of three representative stocks

Feature selection transforms samples from a high-dimensional space into a low-dimensional space by removing redundant and irrelevant features. The principle of univariate feature selection is to calculate a statistical index of each variable separately, and measure the importance of each feature with this index. Finally, features would be kept if they are important and removed if they are not. For the classification task, the Chi-square test, f-test, and mutual information can be used to obtain the index. The Chi-square test is used to describe the independence of two events or to describe the degree of deviation between actual observations and expectations. The larger the value is, the greater the deviation between the actual observation value and the expected value is, which means the weaker the mutual independence between two events is. The use of o_i represents the value vector of ith feature, l represents sentiment label vector, and Equation (4) shows how to calculate Chi-square. And mutual information method is also be proposed to evaluate the correlation between independent variables and dependent variables. In order to process quantitative data, algorithm named maximum information coefficient method is proposed and its formula is as Equation (5)

$$X_i^2(o_i;l) = \sum_{j=1} \frac{(o_{i,j} - l_j)^2}{l_j}, \tag{4}$$

$$I(o_i;l) = \sum_{x \in o} \sum_{y \in l} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{5}$$

Another feature selection method is recursive feature elimination which belongs to the wrapper family. Features are removed gradually based on the results of a reference model. In each round of the feature deletion process, the evaluation of feature importance is no longer a statistical index, but is based on the results of the model.

Feature selection methods only keep a subset of the original dataset, and do not involve any linear or nonlinear combination operations. Another efficient feature method is feature decomposition, which uses space transformation and feature reorganization techniques to obtain new feature representations in a new space. Principal component analysis (PCA) is an unsupervised feature decomposition method, and its purpose is to find a new orthogonal coordinate system in feature space that maximizes the variance of the dataset. PCA can be regarded as a space transforming process, which is defined in such a way that the first principal component has the largest possible variance, and the next principal has the second largest variance in other orthogonal axes. Therefore, later principal components have smaller variances and represent higher frequency information in raw data which have limit help to the classification task. Linear discriminant analysis (LDA) is a linear dimension reduction algorithm similar to PCA, but it is a supervised process. Generally, PCA only chooses directions in which components change significantly, while LDA mostly considers data categories, making the projected samples as separable as possible. By choosing a projection hyperplane in low k-dimensional space, LDA makes the distance between projection points of same classes as close as possible and the distance between projection of different classes as far as possible.

3.3.2 | Stock movement prediction

SVM[34] is a classical binary classification algorithm, which is more robust than other algorithms because of its segmentation hyperplanes. Therefore, SVM is widely used in many tasks. In our stock movement prediction system, an SVM-based method with hybrid sentiment feature input was proposed as a movement predictor. Each stock has two input features x_1 and x_2, and all stocks can be written as $S = \{(x_1, x_2)\}_{n=1}^{N}$, where $y_n \in \{+1, -1\}$ is the label for up and down stocks, respectively. If the S values are linearly separable, then there exists a hyperplane written as Equation (6) to separate the two types of points.

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0. \tag{6}$$

So for every points, there are $y^{(n)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} + b) > 0$. And distance of each point to the hyperplane is:

$$d^n = \frac{||\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} + b||}{||\mathbf{w}||} = \frac{y^{(n)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} + b)}{||\mathbf{w}||}. \tag{7}$$

We define the shortest distance between all points in S and the hyperplane as margin d:

$$d = \min_n d^{(n)}. \tag{8}$$

In addition, SVM models are more stable if the margin is larger, and the goal of SVM is to find a hyperplane (w, b) that is maximized d, as written as Equation (9):

$$\begin{aligned} &\max_{\mathbf{w},b} \quad d \\ &\text{s.t.} \quad \frac{y^{(n)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} + b)}{||\mathbf{w}||} \geq d, \forall n. \end{aligned} \tag{9}$$

Algorithm 1. Individual Stock Movement Prediction

Input: Stock data and textual data

1: Individual stock trading values (373 trading days for 150 Chinese stocks and 383 trading days for 30 American stocks)
2: Individual stock comments (17,435 labeled texts and 2,225,000 unlabeled texts)

Output: stock movement predicting of individual stocks

3: Training word embedding with textual comments with word2vec tool
4: Training sentiment classifier with CNN and RNN based algorithms with labeled texts and choosing the best model as **C**
5: Estimating the sentiment polarity of each unlabeled texts with **C**
6: Individual Stock Filter: analyze daily sentiment index of each stock
7: Extracting trade values in previous 5 days: $X_{\text{trade}} = \{x^{t-5}, x^{t-4}, x^{t-3}, x^{t-2}, x^{t-1}\}$
8: Extracting sentiment index in previous 5 days: $X_{\text{news}} = \{s^{t-5}, s^{t-4}, s^{t-3}, s^{t-2}, s^{t-1}\}$
9: Get $x^{i}_{\text{sentiment}} = \{\text{total}^{i}, \text{pos}^{i}, \text{neg}^{i}, \text{total2}^{i}, \text{pos2}^{i}, \text{neg2}^{i}\}$
10: $r\_\text{pos} = \text{pos}/\text{total}$; $r\_\text{neg} = \text{neg}/\text{total}$; $r\_\text{pos2} = \text{pos2}/\text{total}$; $r\_\text{neg2} = \text{neg2}/\text{total}$
11: $X_{\text{hybrid}} = \text{Merge}(X_{\text{trade}}; X_{\text{sentiment}})$
12: Feature selection of $X_{\text{hybrid}} = \{x^{t-5}_{\text{hybrid}}, x^{t-4}_{\text{hybrid}}, \cdots, x^{t-1}_{\text{hybrid}}\}$ and get X_{input}
13: Stock movement predicting with SVM and LR based on X_{input}

LR is also introduced to predict the stock movement. Compared with SVM models, the LR model is more suitable for the nonlinear transformation of input features. Function g (written as Equation (10)) is used to predict the posterior probability of labels. In addition, $g(\cdot)$, called the activation function, is an S-shaped curve and reduces the range of the linear function to (0, 1), which can be used to represent probability. The parameters **w** and **b** can be learned by the back propagation algorithm.

$$z = \mathbf{w}^{T} x + \mathbf{x}$$
$$\sigma(z) = \frac{1}{1 + e^{-z}}. \tag{10}$$

In summary, there are three main steps in our work, the first is the textual comments collection and word representation learning process which word2vec tools was used in line 3 in Algorithm 1. The second is how to train a sentiment classifier with word vectors and deep learning models, a well labeled stock sentiment dataset is necessary in this part. The last step is analyzing daily sentiment index with the sentiment classifier and get the most important features from combined data with stock daily trading and sentiment data. All steps and key procedures are described with Algorithm 1.

4 | EXPERIMENT

4.1 | Deep sentiment analysis

Different models with various text representation learning inputs were investigated in our previous sentiment classification work. The results showed that LR obtained the most stable performance, the CNN and GRU had the highest accuracy among all classifiers. Therefore, these three models with word embedding input were chosen as the final classifier for labeling the whole stock comments. For word embedding, a 300-dimensional size is the optimal setting when training word vectors. To obtain more comprehensive results, a three-class classifier is used instead of the original binary classifier. Table 3 shows the classification accuracy of three models with word embedding. It can be learned from the results that deep learning-based sentiment classification achieved 9% improvement over LR, so CNN based and GRU based methods were chosen as the sentiment extraction methods.

4.2 | Stock movement prediction

The research mainly focuses on the influences of the investors' sentiment on the movement of the stock market. At first, baseline evaluations are built with only stock trading features, and enhanced features that covered sentiment information is used as input of the comparative experiment group. All features in the past 5 days are utilized to predict the stock movement in the next day. Using $X = \{x^{t-5}, x^{t-4}, x^{t-3}, x^{t-2}., x^{t-1}\}$ to represent the input of classification models, the ith day trading feature is $x^{i}_{\text{trade}} = \{\text{open}^{i}, \text{close}^{i}, \text{high}^{i}, \text{low}^{i}, \text{volume}^{i}, \text{present}^{i}, \text{chg}^{i}\}$, and the ith day sentiment

TABLE 3 Performance of three sentiment classification models (word embedding input)

Model	Category	Precision	Recall	f1	Support
LR	Pos	0.79	0.83	0.81	909
	Neu	0.58	0.45	0.50	185
	Neg	0.69	0.68	0.68	447
	Micro	0.74	0.74	0.74	1541
CNN	Pos	0.83	0.90	0.86	919
	Neu	0.65	0.49	0.56	176
	Neg	0.79	0.73	0.76	446
	Micro	0.80	0.80	0.80	1541
GRU	Pos	0.86	0.91	0.88	902
	Neu	0.74	0.53	0.62	178
	Neg	0.80	0.79	0.79	461
	Micro	0.83	0.83	0.83	1541

Abbreviations: CNN, convolutional neural network; GRU, gated recurrent unit; CNN, convolutional neural network.

feature is $x^i_{\text{sentiment}} = \{\text{total}^i, \text{pos}^i, \text{neg}^i, \text{neu}^i, \text{total2}^i, \text{pos2}^i, \text{neg2}^i\}$. For trading features, the first discrete difference of the first five elements are also be added, so there are 12 dimensions in the original trading features. For the sentiment feature, we add the first discrete differences for all elements, and add another four ratio features $r\_\text{pos} = \text{pos/total}$, $r\_\text{neg} = \text{neg/total}$, $r\_\text{pos2} = \text{pos2/total}$, and $r\_\text{neg2} = \text{neg2/total}$, so there are 18 dimensions in the original sentiment features.

In the Snowball dataset, the stock code and its name will always be embedded into a hyperlink in each html snippet. Sometimes only one stock exists in a comment, and sometimes several stocks are mentioned in one comment at the same time. Based on this trend, three different types of tweet filters are proposed: "*only*" means that comments include only one stock, "*first*" represents tweets with more than one stock, but the first stock mentioned is the target stock, and "*all*" represents comments with more than one stock, and the target stock is in these stocks no matter its position. Therefore, the relationships among these three patterns are satisfied with $S_{\text{only}} \in S_{\text{first}} \in S_{\text{all}}$.

The Stock dataset and trading dataset span one and a half years and contains 373 trading days in the Chinese stock market and 383 trading days in the American stock market. Considering that the historical window day is 5, and adding a discrete difference feature, there are only 367/377 valid samples in our dataset. The previous 80% samples are used as a training set and another 20% samples left are used as a testing set. Different feature combinations with comment type (only, first and all) and sentiment classification methods (CNN, GRU) are all trained with the SVM and LR methods with four different feature engineering algorithms for each stock. After obtaining prediction results of all stocks, the average accuracy is calculated to show the predictive ability of the model in each exchange.

For each stock, there are six different combinations of enhanced feature inputs. As the result show in Table 4, the "Trade only" column represents our baseline results with only trading features as the input, and the next six columns are different hybrid features. Each row represents the average accuracy with a certain classification model and feature engineering algorithm. The results indicate the following: (i) All hybrid features achieve better performance than the baseline features regardless of the comment filter strategy and the sentiment analysis methods chosen. (ii) PCA feature decomposition achieves the best performance both in three different exchanges. (iii) American stocks achieve a higher accuracy with baseline input features (underline result in trade column only).

The best performance with baseline features is marked with a dashed box, and the best performance with hybrid features is marked with gray color box. The sentiment gain rate is used to measure the improvements that sentiment information brings to the model. $\text{Accu}^{\text{trade}}_{\text{best}}$ indicates the highest accuracy of the trade only features, and $\text{Accu}^{\text{hybrid}}_{\text{best}}$ indicates the highest accuracy of the hybrid features. Therefore, the gain ratio was defined as the first formula in Equation (11), and get result for each exchanges as indicated in American.

$$
\begin{aligned}
r_{\text{gain}} &= \frac{\text{Accu}^{\text{hybrid}}_{\text{best}} - \text{Accu}^{\text{trade}}_{\text{best}}}{\text{Accu}^{\text{trade}}_{\text{best}}}, \\
r_{\text{gain}}(\text{SH}) &= \frac{0.8189 - 0.7232}{0.7232} = 13.23\%, \\
r_{\text{gain}}(\text{SZ}) &= \frac{0.8595 - 0.7512}{0.7512} = 14.42\%, \\
r_{\text{gain}}(\text{USA}) &= \frac{0.9470 - 0.8829}{0.8829} = 10.72\%.
\end{aligned}
\tag{11}
$$

TABLE 4 Average training accuracy of different models with different input features

Stock	Model	Trade only	All cnn	All gru	First cnn	First gru	Only cnn	Only gru
ShangHai (100 stocks)	SVM(UFS)	0.6503	0.7480	0.7495	0.7490	0.7486	0.7497	0.7481
	SVM(REF)	0.6586	0.7600	0.7590	0.7589	0.7617	0.7597	0.7591
	SVM(PCA)	0.7232	0.8189	0.8178	0.8166	0.8181	0.8142	0.8119
	SVM(LDA)	0.6541	0.7985	0.7990	0.7987	0.7986	0.7919	0.7948
	LR(UFS)	0.6260	0.7229	0.7229	0.7227	0.7218	0.7222	0.7211
	LR(REF)	0.6325	0.7347	0.7354	0.7352	0.7340	0.7363	0.7335
	LR(PCA)	0.6002	0.6833	0.6806	0.6836	0.6830	0.6822	0.6841
	LR(LDA)	0.6487	0.7932	0.7938	0.7930	0.7908	0.7860	0.7877
ShenZhen (50 stocks)	SVM(UFS)	0.6592	0.7579	0.7584	0.7612	0.7603	0.7612	0.7620
	SVM(REF)	0.6682	0.7682	0.7688	0.7721	0.7728	0.7765	0.7743
	SVM(PCA)	0.7512	0.8588	0.8510	0.8595	0.8527	0.8571	0.8524
	SVM(LDA)	0.6598	0.8010	0.8015	0.8053	0.7995	0.7973	0.7974
	LR(UFS)	0.6263	0.7272	0.7242	0.7273	0.7267	0.7295	0.7270
	LR(REF)	0.6322	0.7380	0.7367	0.7412	0.7378	0.7429	0.7397
	LR(PCA)	0.6051	0.6894	0.6838	0.6857	0.6879	0.6895	0.6864
	LR(LDA)	0.6524	0.7971	0.7960	0.7999	0.7945	0.7937	0.7926
American (30 stocks)	SVM(UFS)	0.7254	0.7968	0.7952	0.7969	0.7970	0.7981	0.7923
	SVM(REF)	0.7465	0.8105	0.8067	0.8115	0.8086	0.8140	0.8184
	SVM(PCA)	0.8829	0.9470	0.9454	0.9464	0.9462	0.9468	0.9443
	SVM(LDA)	0.6552	0.7983	0.7979	0.8014	0.7965	0.7972	0.7935
	LR(UFS)	0.6322	0.7329	0.7355	0.7348	0.7346	0.7392	0.7385
	LR(REF)	0.6402	0.7468	0.7503	0.7481	0.7510	0.7530	0.7525
	LR(PCA)	0.6073	0.6791	0.6819	0.6809	0.6832	0.6849	0.6870
	LR(LDA)	0.6450	0.7943	0.7920	0.7967	0.7927	0.7926	0.7883

Abbreviations: LDA, linear discriminant analysis; LR, logistic regression; PCA, principal component analysis; SVM, support vector machine; UFS, univariate feature selection.

It can be learned from the gain ratios that extra sentiment information in China stocks especially in Shenzhen and Shanghai exchange perform better than stocks in American.

The best performance of hybrid features has been highlighted by the gray box in Table 4, and the best model settings and feature combinations were selected from the results and be tested with testing data. To forecast the model performance in different future time periods, the model forecasts stock movements in the next 10 days, next 20 days until the next 70 days. The average testing accuracy is listed in Table 5. From the results we learn the following: (i) In all results, with an increase in forecasting time, the accuracy tends to decrease in all three exchanges. However, the stocks in the American exchange almost maintain the same performance during the first four time periods while stocks in China show an downtrends. (ii) In the Chinese stock market, sentiment information is useful for stocks. However, it is expected that the maximum improvement would be achieved by the first time period in the Shanghai exchange, but it fails. A detailed analysis and explanation are introduced in the following section.

4.3 | Stock analysis

4.3.1 | Influence of the reference days

In previous experiments, historical reference days are set to 5 because we assume that 5 days is a common trading cycle. To analyze the influence of reference days on the prediction results, other experiments with days ranges from 1 to 10 are completed with the same settings. As shown in

TABLE 5 Average testing accuracy in different time periods

Stocks	Feature	[0, 10]	[0,20]	[0,30]	[0,40]	[0,50]	[0,60]	[0,70]
ShangHai (100 stocks)	Trade	0.5570	0.5390	0.5270	0.5300	0.5278	0.5195	0.5064
	Sentiment	0.5620	0.5355	0.5253	0.5268	0.5224	0.5167	0.5049
	$\text{Hybrid}_{\text{best}}$	0.5500	0.5515	0.5350	0.5365	0.5336	0.5288	0.5130
ShenZhen (50 stocks)	Trade	0.5560	0.5430	0.5240	0.5260	0.5176	0.5180	0.4943
	Sentiment	0.5560	0.5200	0.5047	0.5185	0.5152	0.5160	0.5046
	Hybrid	0.5580	0.5510	0.5333	0.5305	0.5232	0.5217	0.5034
American (30 stocks)	Trade	0.5433	0.5450	0.5478	0.5408	0.5307	0.5300	0.5286
	Sentiment	0.4733	0.5000	0.5067	0.5042	0.4967	0.5017	0.5005
	$\text{Hybrid}_{\text{best}}$	0.5367	0.5317	0.5311	0.5275	0.5173	0.5106	0.5143

Figure 4, the average trading and testing accuracy of stocks in Shanghai (left subfigure) and Shenzhen (right subfigure) is demonstrated. From the results, it can be learned that the training accuracy increases as the reference days become longer. However, different results are shown in the testing set. More reference days bring more useful features for the prediction model in the training set but these extra features do not provide any benefit in some moments in the testing data. Throughout the model performance, it is reasonable to choose 5 days as the reference days.

4.3.2 | Visualization of the sentiment features

Feature selection algorithms filter a subset of the original features based on their importance. To explore how much sentiment information would be retained after the selection method, the total occurrences of each feature were visualized in Figure 5. In this figure, from top to the bottom, each

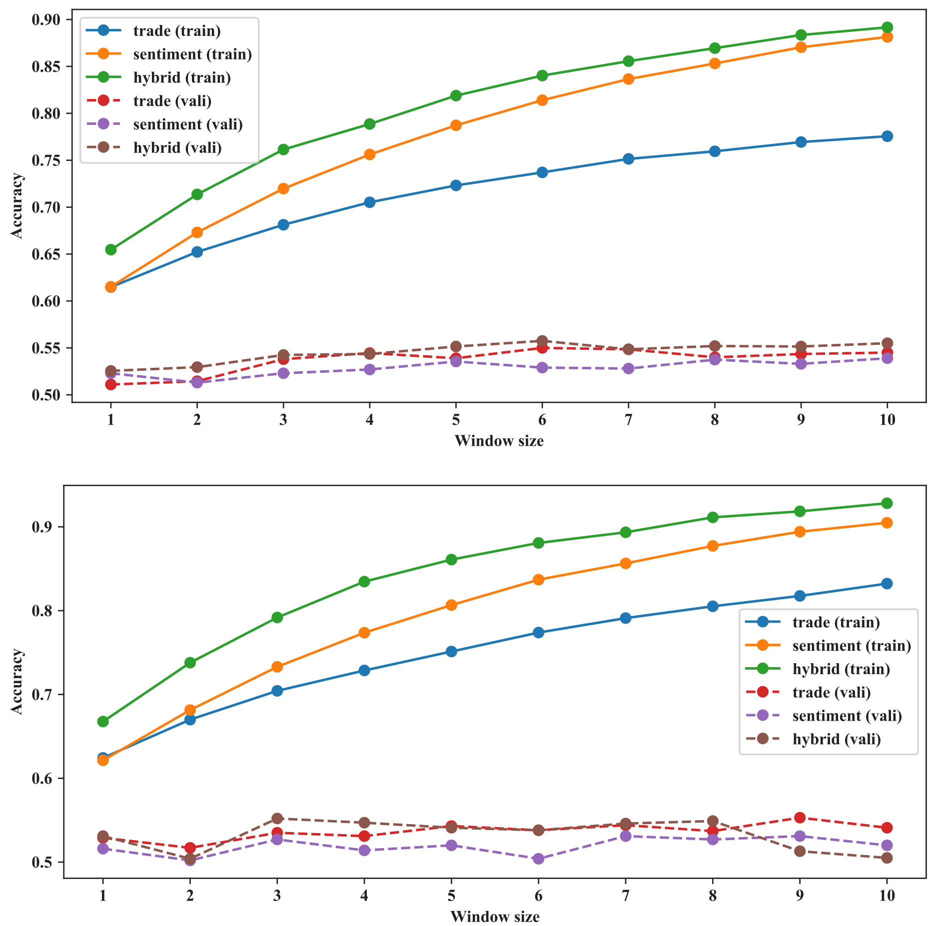

FIGURE 4 Average accuracy with different reference days in Shanghai and Shenzhen stock exchange

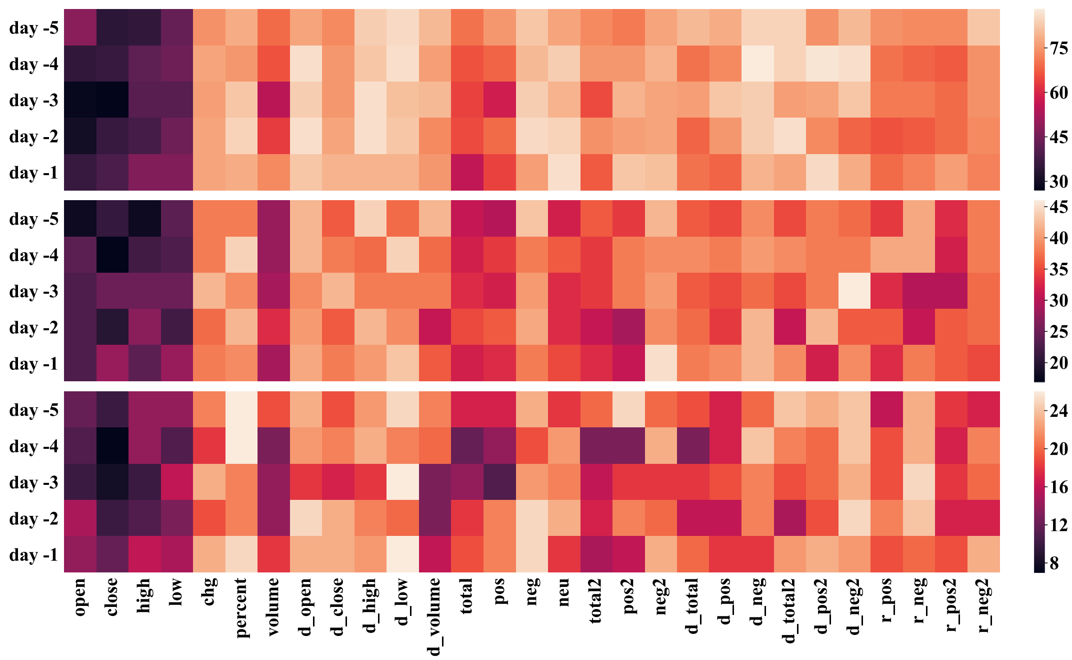

FIGURE 5 Heatmap of occurrences of each feature

subfigure represents stocks in the Shanghai, Shenzhen, and American exchanges. The horizontal axes show features including trade features (left part) and sentiment features (right part), and the vertical axes show reference days from the previous first day to the previous fifth day. Every small square represents accumulated occurrences of the corresponding features, and the brighter positions means that they occur more than others. It can be learned from the heatmap that in the forecasting, the sentiment features play an important role, which is particularly prominent in the Chinese stock market.

4.3.3 | Individual stock analysis

Previous results are all based on the average analysis of all stocks. However, in the real stock market, sentiment promotes prediction of some stocks, especially in the Chinese exchanges and sometimes may fail. In this section, sentiment features were analyzed from the perspective of individual stocks.

The price to book value (P/B ratio) is the ratio of market value of a company's shares over its book value of equity. This ratio is always used to reflect the future development prospects of a company. When the market value is higher than the book value, buying stock is a better opportunity. By contrast, when the quality of the assets is poor, it is better not to buy stock. The left two subfigures in Figure 6 indicate the P/B ratio of stocks, the blue box plot indicates stocks in which the sentiment information has a positive effect on the predictions, while the orange box plot indicates a negative effect. It can be concluded that comments on stocks with high P/B ratios are more closer to the real trend of the market.

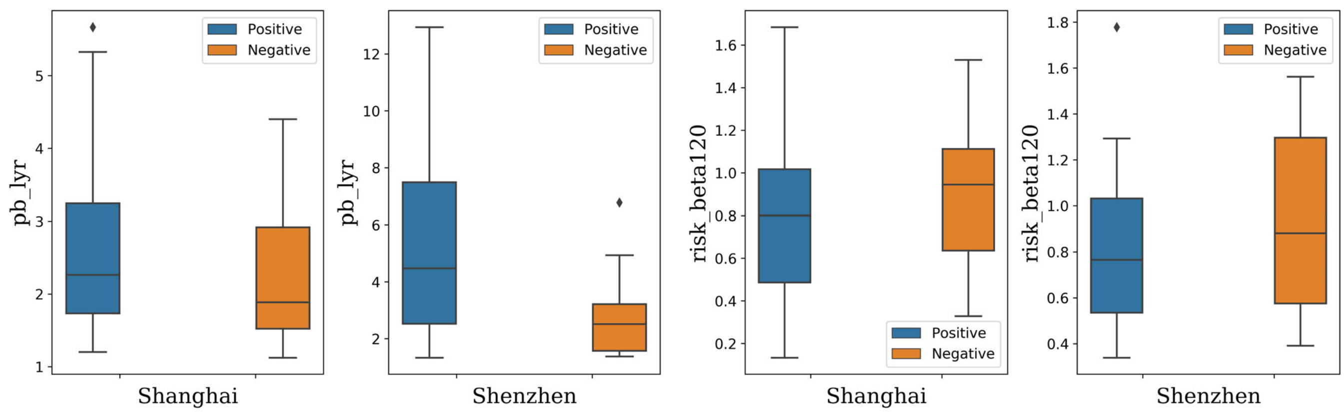

FIGURE 6 Boxplot of price to book ratio and beta value among Chinese stocks

The beta coefficient is a risk index that measures the price movement of individual stocks and is relative to the entire stock market. The larger the absolute value of beta is, the larger the movement of the stock market is. If the coefficient is negative, then it shows that the direction of the stock movement is opposite to that of the market. The right two subfigures in Figure 6 indicates the beta value of the improved and reduced stocks. It is obvious that the beta values of most stocks with improved prediction were smaller than these stocks with reduced performance. We can learn from this comparison that most investors prefer low risk stocks and present negative comments for these high risk stocks.

From the perspective of individual stocks, sentiment information is not valid for all stocks as shown in Figure 6. In addition, in Table 5, we also studied the reason why the average testing accuracy in the range [0,10] did not achieve the expected results. By reviewing the test dataset selection, we can know that the first range [0,10] is exactly the beginning of a bull market, and investors' behavior with respect to their comments are very different from their usual comment behavior. This reason causes the abnormality. Overall, extra sentiment information is more powerful than the trade only features.

5 | CONCLUSION

Sentiment analysis plays an increasingly important role in many fields, especially in financial prediction. With the development of social media and mobile applications, unstructured data, such as comments and news from the Internet provide streaming data for analysis. In this article, we proposed a sentiment information extraction method based on deep learning networks and applied sentiment features to stock movement prediction. Through the experimental analysis, we can draw the following conclusions:

- Sentiment analysis based on deep learning methods is practicable for stock movement prediction, and sentiment features have a positive effect on the prediction. In the sentiment feature extraction process, although deep neural networks, such as CNN and GRU achieved the best polarity classification results, they have almost the same contributions to the final prediction with the LR method. Therefore, machine learning-based feature extraction can achieve similar performance to deep learning-based methods.
- For both trade only and hybrid features, PCA is an effective and powerful feature decomposition method. It achieved the best performance among the feature selection methods. Through feature visualization, we can learn that more sentiment information was filtered after feature engineering.
- The most reasonable historical reference days are five which is exactly a trading cycle in one week. Based on this, the prediction model can only forecast movements in the next few days, and fails in the long term prediction.
- Sentiment analysis is useful for stocks with higher P/B ratios and lower beta risk values. This behavior illustrates that investors are more likely to focus on stocks with a development space and make plans based on the overall sentiment of the market.

In the future research, except for improving prediction accuracy, a detection method for important stock events must be studied before bull and bear markets. At the same time, appropriate investment strategies need to be combined with current prediction research in order to apply to the real stock market.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China No. 91546201, No. 71501175, and No. 71932008 University of Chinese Academy of Sciences and the Open Project of Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences.

ORCID

Yuanchun Zheng https://orcid.org/0000-0001-5757-0598
Kun Guo https://orcid.org/0000-0002-4787-8406

REFERENCES

1. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. *Wiley Interdiscipl Rev Data Mining Knowl Discov*. 2018;8(4):e1253.
2. Kumar A, Jaiswal A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. *Concurr Comput Pract Exper*. 2020;32(1):e5107.
3. Ballings M, Poel D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst Appl*. 2015;42(20):7046-7056.
4. Ariyo AA, Adewumi AO, Ayo CK. Stock price prediction using the ARIMA model. In Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge; 2014:106-112. Accessed February 23, 2015.
5. Adebiyi AA, Adewumi AO, Ayo CK. Stock price prediction using the ARIMA model. In Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. Cambridge; 2014;106-112. Accessed Date: February 23, 2015.
6. Mittal Anshul, Goel Arpit. *Stock Prediction using twitter sentiment Analysis.* Stanford University Working Paper. 2012;15.
7. Weng B, Ahmed MA, Megahed FM. Stock market one-day ahead movement prediction using disparate data sources. *Expert Syst Appl*. 2017;79:153-163.

8. Tsai CF, Wang SP. Stock price forecasting by hybrid machine learning techniques. Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong; 2009. Accessed March 20, 2009.
9. Choudhry R, Garg K. A hybrid machine learning system for stock market forecasting. *World Acad Sci Eng Technol*. 2008;39(3):315-318.
10. Lee M-C. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Syst Appl*. 2009;36(8):10896-10904.
11. Ince H, Trafalis TB. Kernel principal component analysis and support vector machines for stock price prediction. *IIE Trans*. 2007;39(6):629-637.
12. Ding X, Zhang Y, Liu T, Duan J. Deep learning for event-driven stock prediction. Proceedings of the 24th International Joint Conference on Artificial Intelligence, Atlanta; 2015. Accessed August 01, 2015.
13. Akita R, Yoshihara A, Matsubara T, Uehara K. Deep learning for stock prediction using numerical and textual information. Paper presented at: Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS); 2016:1-6. Accessed August 25, 2016.
14. Zhou Y-L, Han R-J, Xu Q, Jiang Q-J, Zhang W-K. Long short-term memory networks for CSI300 volatility prediction with Baidu search volume. *Concurr Comput Pract Exper*. 2019;31(10):e4721.
15. Atsalakis G, Valavanis Kimon P. Surveying stock market forecasting techniques-Part I: conventional methods. *J Comput Optimiz Economs Finan*. 2010;2(1):45-92.
16. Chiong R, Fan Z, Hu Z, Adam MT, Lutz B, Neumann D. A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. Paper presented at: Proceedings of the Genetic and Evolutionary Computation Conference Companion, Kyoto; 2018:278-279. Accessed July 19, 2018.
17. Chen W, Cai Y, Lai K, Xie H. A topic-based sentiment analysis model to predict stock market price movement using Weibo mood. *Web Intelligence*. Amsterdam: IOS Press; 2016:287-300.
18. Guo K, Sun Y, Qian X. Can investor sentiment be used to predict the stock price? dynamic analysis based on China stock market. *Phys A Stat Mech Appl*. 2017;469:390-396.
19. Oliveira N, Cortez P, Areal N. The impact of microblogging data for stock market prediction: using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Syst Appl*. 2017;73:125-144.
20. Zhou Z, Xu K, Zhao J. Tales of emotion and stock in China: volatility, causality and prediction. *World Wide Web*. 2018;21(4):1093-1116.
21. Pagolu Venkata Sasank, Reddy Kamal Nayan, Panda Ganapati, Majhi Babita. Sentiment analysis of twitter data for predicting stock market movements. Paper presented at: Proceeding sof the 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES); 2016:1345-1350. Accessed June 26, 2017.
22. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *J Comput Sci*. 2011;2(1):1-8.
23. Hagenau M, Liebmann M, Neumann D. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decis Support Syst*. 2013;55(3):685-697.
24. Li X, Xie H, Wang R, et al. Empirical analysis: stock market prediction via extreme learning machine. *Neural Comput Appl*. 2016;27(1):67-78.
25. Chan SWK, Chong MWC. Sentiment analysis in financial texts. *Decis Support Syst*. 2017;94:53-64.
26. Liu X-Q, Wu Q-L, Pan W-T. Sentiment classification of micro-blog comments based on randomforest algorithm. *Concurr Comput Pract Exper*. 2019;31(10):e4746.
27. Wang Y, Wang P, Zhang J, et al. A novel bat algorithm with multiple strategies coupling for numerical optimization. *Mathematics*. 2019;7(2):135.
28. Cui Z, Zhang J, Wang Y, et al. A pigeon-inspired optimization algorithm for many-objective optimization problems. *SCIENCE CHINA Inf Sci*. 2019;62:070212.
29. Li X, Xie H, Chen L, Wang J, Deng X. News impact on stock price return via sentiment analysis. *Knowl-Based Syst*. 2014;69:14-23.
30. Song Y, Shi S, Li J, Zhang H. Directional skip-gram: explicitly distinguishing left and right context for word embeddings. Paper presented at: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans; 2018:175-180. Accessed June 06, 2018.
31. Do HH, Prasad PWC, Maag A, Alsadoon A. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Syst Appl*. 2019;118:272-299. https://doi.org/10.1016/j.eswa.2018.10.003.
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. Paper presented at: Proceedings of the International Conference on Learning Representations, Scottsdale; 2013:1-12. Accessed September 07, 2013.
33. Chung J, Gulcehre C, Cho KH, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling; 2014. arXiv:1412.3555.
34. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565-1567.

How to cite this article: Shi Y, Zheng Y, Guo K, Ren X. Stock movement prediction with sentiment analysis based on deep learning networks. *Concurrency Computat Pract Exper*. 2020;e6076. https://doi.org/10.1002/cpe.6076