IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 18, 2019, accepted June 8, 2019, date of publication June 19, 2019, date of current version July 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923637

A Sentiment-Aware Trading Volume Prediction Model for P2P Market Using LSTM

XIANGLING FU[1,3], SHUAI ZHANG[1,3], JINPENG CHEN[1], TIANXIONG OUYANG[1,3], AND JI WU[2], (Senior Member, IEEE)

[1]School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[3]Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Jinpeng Chen (chenjinpeng@nlsde.buaa.edu.cn)

This work was supported in part by the National Key Research and Development Program under Grant 2017YFB0803300, in part by the National Natural Science Foundation of China under Grant 61702043, and in part by the Open Project Fund of Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Ministry of Education.

ABSTRACT With the development of Internet lending, the use of peer-to-peer (P2P) as a new financial credit model has been increasing in China. However, this rapid development has led to a major potential risk. A few P2P enterprises operate well in the beginning but close within a short period because of the suspension of business, fraud, illegal fundraising, and blind expansion. Effective supervision of the P2P industry is an urgent problem. Trading volume reflects the operation stability of P2P platforms. Hence, predicting the volume of the P2P market is an important research topic. This paper first analyzes the trading data of a P2P platform. It is found that the sentiment of investor comments is related to the trading volume of the P2P platform. Then, we use the TextCNN model to classify the sentiment of investor comments and obtain the time series of changes in sentiment. It is verified that the time series of change in sentiment and the P2P volume index has statistical causality and a strong correlation. This paper proposes a model that uses the trend of change in investor sentiment to predict P2P trading volume. This model uses the historical time series change in investor sentiment, the P2P volume index, and WeekDay characteristics to predict future P2P trading volume. The experimental results show that the proposed model is better than a few existing baseline methods. Compared with baseline regression, the Pearson coefficient of the predicted and actual values of the proposed model is increased by 13.26%, the mean squared error is decreased by 27.62%, and the R-squared value is increased by 28.48%.

INDEX TERMS Peer-to-peer lending, sentiment tracking, trading prediction, long short-term memory (LSTM).

I. INTRODUCTION

Peer-to-peer (P2P) lending refers to the lending behavior between lenders and investors on online platforms, rather than in conventional banking and other financial institutions [1]. P2P is a new financial credit model that features direct, small, and short-term transactions between borrowers and lenders without expert participation. By 2017, China's domestic P2P trading volume reached 280,849 billion yuan, which was an improvement of 35.9% compared to 2016; the accumulated turnover of the online banking industry exceeded 6 trillion yuan.[1] However, the Chinese P2P market is complicated because of the low barriers to entry. By the end of 2017, 645 P2P enterprises closed because of suspension of business, fraud, and other problems due to the operators' underestimation of market risks or blind expansion. This caused major financial risk because the closed enterprises had no money to pay back investors. P2P trading volume can reflect the prosperity of the market. If P2P trading volume can be predicted, the law of market changes can be determined and a theoretical basis for government regulation and business operations can be obtained.

In the real world, investors actively post their comments concerning a specific P2P lending platform or the P2P market on the Internet. These comments play a significant role in other investors' decision making because the comments reflect the investor sentiment about the market.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhanyu Ma.

[1]https://baijiahao.baidu.com/s?id=1588320701731305319&wfr=spider&for=pc

In this context, we develop the following research questions:

- Is the sentiment fluctuation extracted from online comments relevant to the trading volume of P2P market?
- Furthermore, is it possible to predict P2P trading volume through the sentiment analysis of these comments?

To answer these questions, we develop a sentiment classification model to calculate the sentiment trend through investors' comments and verify that sentiment change is relevant to the P2P market trading volume. Then, we propose a long short-term memory (LSTM)-based trading volume prediction model to predict P2P trading volume. The contribution of this study is that it develops an efficient and practical method for predicting P2P trading volume. As P2P in China is a growing industry, conventional prediction methods are based on historical trading volume. It is difficult to find an effective prediction method because market changes are rapid and a large amount of information is hidden by operators. We find that the online comments of investors can reflect operating status. However, using this information to predict the market is still a challenge. First, a large number of comments must be processed automatically. In addition, it is important to determine the relationship between investor comments and trading volume. In this aspect, research has been conducted on using online text to predict the stock market [2]–[4] by using the deep learning method. We find that the sentiment of investor comments is related to the trading volume of a P2P platform. Thus, we propose a model that uses the change in investor sentiment to predict P2P trading volume based on a deep neural network (DNN) method.

The remainder of this paper is organized as follows: Section II discusses related work. Section III defines the problem addressed and outlines the framework applied to solve it. Section IV discusses the data utilized, including how they were obtained, preprocessed, and analyzed. Further, the sentiment classification model and classification results are presented. Sections V and VI present the P2P trading volume prediction model and the analysis of experimental results. Finally, the verification of our hypothesis is described in Section VII.

II. RELATED WORK

Numerous scholars have studied the risks of the P2P market to more effectively prevent the potential crisis of the P2P market and avoid large losses. In general, the conventional efforts towards investigating such risks can be categorized into two types according to research objects, i.e., research on borrower default risk and research on online P2P platform default risk.

A. BORROWER DEFAULT RISK

The studies on borrower default risk attempt to evaluate the risk that borrowers cannot repay loans on time based on individual characteristics. These studies are generally divided into two stages. First, through the statistical analysis of the attributes of borrowers and the default status, the main factors that affect default are determined. Then, statistical methods are used to explore the relationship between these influencing factors and the default status and construct a default risk prediction evaluation model. Previous research on the factors that affect the default status [5]–[7] showed that hard information, such as financial indicators, credit ratings, income, education level, family, and marital status, soft information, and social account information have an important impact on loan default. Malekipirbazari *et al.* [8], Lin *et al.* [9], and Xia *et al.* [10] compared different machine learning methods to construct a risk assessment model and showed that machine learning methods are considerably better than conventional credit system level recognition. With the development of deep learning, numerous methods based on DNNs [11]–[13] have been applied to the study of the default risk of borrowers. Duan [14] proposed a decision method based on DNNs. Their work showed that DNNs have more powerful feature extraction capabilities compared to general machine learning methods.

B. PLATFORM DEFAULT RISK

The research on the default risk of online lending platforms analyzes the relationship between the platform default status and the internal and external environmental factors of the platform, such as the platform's attributes, macroeconomic environment, and related policies, for comprehending the complex patterns to assess the risks of prevention platforms. One of the most widely used methods is linear regression [15] owing to its simplicity and interpretability. However, there is considerable noise in an actual P2P online loan environment, and the changes in the risk of the platform are frequently nonlinearly related to influencing factors. To better identify these complex nonlinear mapping relationships, machine learning methods such as decision trees, logistic regression, and support vector machines have been utilized by scholars. Based on the data of China's online lending platform, Yang and Luo [16] proposed an improved SVM (improved AdaBoost SVM) using rating indicators. Experiments showed that this method can effectively predict whether a platform is dangerous based on a limited data set. Liu *et al.* [17] proposed a method for dynamically evaluating the operational risk of a P2P platform based on a short-term multisource regression algorithm. The experimental results showed that the method can provide investors with a dynamic risk assessment and effective platform prompts.

C. DEEP NEURAL NETWORK METHODS FOR FINANCIAL MARKET PREDICTION

There has been a number of efforts for financial markets predicting using deep learning technologies [18]–[22]. For example, Ding *et al.* [21] introduced a novel deep learning framework extract events embedding from unstructured news texts to predict the stock market trend. The Long and Short Time Memory neural network (LSTM) proposed by Gers *et al.* [23] can further capture the long-term dependency in time series Inspired by this, Chen *et al.* [24] proposed an LSTM-based stock market forecasting method.

Recently, many research has studied the impact of investor sentiment changes on financial markets. Oliveira *et al.* [25] used the sentiment and survey indices extracted from Weibo to predict stock market behavior. Li *et al.* [26] used the Harvard Mental Dictionary and the Loughran-McDonald Financial Emotion Dictionary to construct an emotional space and evaluate the impact of emotional volatility on stock price returns in financial news. These studies demonstrate the important role of investor sentiment in the financial market.

In summary, through the analysis of the abovementioned works, we find the following two gaps in existing research on P2P market: 1) The emotional factors of investors are not fully considered. 2) Only a few studies have analyzed the risks of the entire online lending industry. To bridge these gaps, in this work we attempt to predict P2P market volume in this work. Motivated by the application of sentiment features in financial markets such as stock and forex, we extracted sentiment features from investor comments and considered it as an important feature for trading volume prediction. Our study provides a feasible method for the research of the P2P online lending industry, which can provide the relevant regulatory authorities with a theoretical basis for preventing the risks of the P2P online lending industry.

III. RESEARCH FRAMEWORK

A. PROBLEM STATEMENT

We made the following assumption:

Sentiment trends are relevant to the P2P market and predict the P2P trading volume.

To prove the hypothesis, we carried out our study on the Chinese P2P market. Specifically, we established a sentiment classification model and a P2P trading volume prediction model. In the sentiment classification model, which is based on TextCNN [27], we used the word vector of comments as input and the sentiment orientation as output.

In the P2P volume predicting model, we used the first seven days of the trading volume index, sentiment orientation, and day of the week as inputs to predict trading volume and set up a control group to compare the predict results.

We implemented our program in the Python language for reproducibility. Deep learning models were developed based on TensorFlow, which is an open deep learning framework released by Google.

The symbols used in this paper are defined in TABLE 1.

TABLE 1. Symbol definition.

Symbol	Explanation
ND	Time series of the percentage of negative-sentiment comments for each day
TVI	Daily trading volume index
VD	Daily trading volume index (TVI) time series
L	The input of different LSTM models
Accuracy	Percentage of the output of the model is the same as the original label in the test data
PCC	Pearson correlation coefficient
MSE	Mean Squared Error

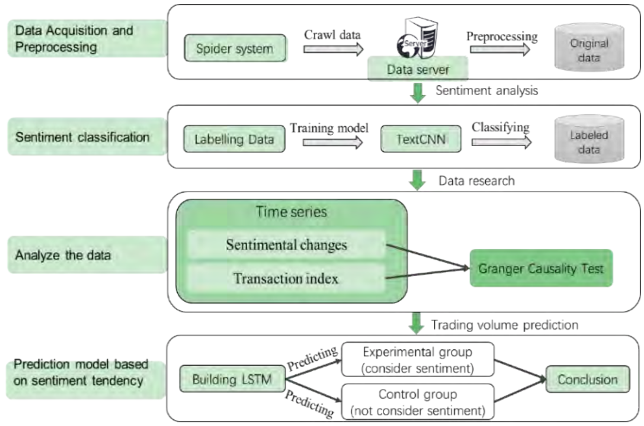


FIGURE 1. Research framework.

B. RESEARCH FRAMEWORK

An overview of our research framework is shown in FIGURE 1.

The following procedure was carried out in this study:

1) Data acquisition and preprocessing: We developed a Python-based crawler script to obtain the comments of all P2P platforms and the daily trading volume index (TVI) from Wangdaizhijia (https://www.wdzj.com/). As all comments were in Chinese text, further preprocessing was required before input to the computer. In all the following experiments, we tokenized each comment and removed the stop words and the words that appeared less than 10 times to build a vocabulary. Then, we employed an unsupervised CBOW Word2Vec [28] model with a dimension of 100 through the entire comments corpus to achieve pretrained word-level embedding.

2) Sentiment classification: We annotated randomly selected data from the dataset with positive and negative labels. Then, we used the annotated dataset to train the TextCNN model and utilized the trained model for sentiment classification on the unlabeled dataset, eventually forming a sentiment dataset.

3) Data analysis: The negative sentiment comments were divided into parts according to date. And we calculate the percentage of negative-sentiment comments each part to generate time series. Then, a Granger causality test was used to analyze the relevance of sentiment changes on the P2P financial market.

4) Sentiment-based P2P trading volume prediction: The negative-sentiment time series was used as a feature to predict trading volume and compared with other prediction models.

IV. DATA ACQUISITION AND SENTIMENT CLASSIFICATION

A. DATA ACQUISITION AND PREPROCESSING

P2P financial platform data sources are generally divided into two categories: Sina Weibo[2] and Baidu Post Bar[3], and all

[2] https://www.weibo.com/
[3] https://tieba.baidu.com/

other social media or P2P information website. We found that the former category has the following characteristics:

1) There are many users, the user groups are involved in a wide range of topics.

2) Texts from Sina Weibo and Baidu Posts Bar are not standardized and include Internet terms, emoji, symbols, links, and pictures, which are not conducive to our data analysis.

3) These websites feature significant advertising and promotional content, and the text contains considerable extraneous content.

Most users of P2P information websites are investors and borrowers who comment and share information about P2P companies. After comparing several websites, including Net Credit Eye[4], Wangdaizhijia and Dailuopan[5], we eventually selected Wangdaizhijia comment data and daily published industry data (the trading volume) as our research object. Wangdaizhijia is China's first authoritative P2P information platform and one of the largest P2P industry portals. Wangdaizhijia collects comprehensive data, updates data in a timely manner, and is convenient to study. Its ranking index has become an important index for investors measuring the comprehensive strength of a platform, affecting users' investment decisions.

We obtained 212,964 Chinese comments between February 2013 and April 2018 from about 6086 platforms. In addition, we collected the daily TVI from January 1, 2016, to March 31, 2018. The TVI reflects the overall change in the volume of the P2P market. To calculate TVI, experts select 20 P2P platforms as component platforms, and the arithmetic average of the daily trading volume of the component platforms is the value of TVI. The selection was based on influence, representativeness, diversity, and stability. For n component platforms the TVI is calculated as

$$TVI = \frac{\sum TradingVolume}{n \times 10000}. \tag{1}$$

After crawling all the data, we preprocessed the raw data, removing special characters and emoji, and then segmented the text. The dataset has 211664 comments. Our text segmentation tool used the lexical analysis application programming interface (API) provided by the Baidu Cloud Engine.[6] We obtained a vocabulary of size 7971 after the word segmentation.

B. SENTIMENT CLASSIFICATION

1) DATA ANNOTATION

We divided comments into two categories: negative and positive. Negative comments include complaints against websites, and customer services; questions or dissatisfaction with P2P businesses and interests; and fraud risks of various companies. The remaining data were labeled as positive.

[4] https://www.tianyancha.com/
[5] http://www.dailuopan.com/
[6] http://ai.baidu.com/tech/nlp/lexical

TABLE 2. Labeled results.

Positive	*Negative*
13,268	6732

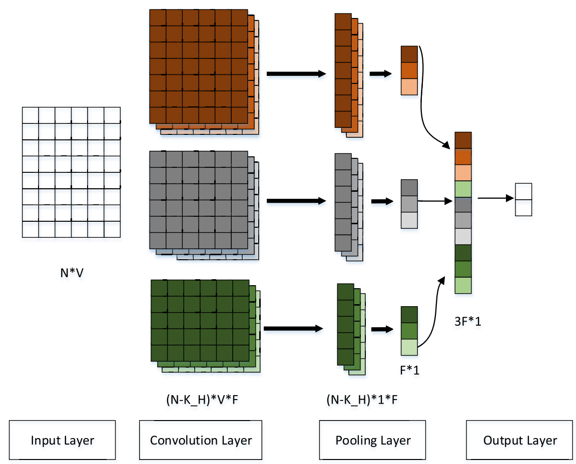


FIGURE 2. Structure of our sentiment classifier model (N denotes the length of the input sentence and K_H is the width of the filter kernel. We use three types of filters with a filter width of 2,3,4, respectively. F denotes the number of feature maps, which is 32 in our experiments).

We randomly selected 20,000 comments from the original dataset for labeling. The dataset distribution is shown in TABLE 2.

2) SENTIMENT CLASSIFIER MODEL

The sentiment classifier employed in this study is based on TextCNN, which is a widely used text classification model that utilizes convolutional neural networks to efficiently extract semantic information. TextCNN has achieved good performance in numerous NLP tasks [29]. FIGURE 2 shows the structure of our model, which consists of a convolutional layer with three types of convolutional kernels, a pooled layer, and a fully connected layer. Detailed parameters of the model are shown in TABLE 3.

TABLE 3. Detailed parameters of TextCNN based model.

Parameter	Description	Value
vocab_dim	Dimensions of Word embedding	100
Batch_size	Batch size	64
lr	Learning rate	0.001
sen_max_len	The maximum length of the sentence	200
num_filters	Number of feature maps	32
n_epoch	Training epochs	60
num_class	Number of target classes	2
optimizer	Optimizer algorithm	Adam
filter_sizes	Filters widths	[2,3,4]
filter_steps	Strides	1

Comments annotated in the manner described above, were applied to train our sentiment classifier model. We split the

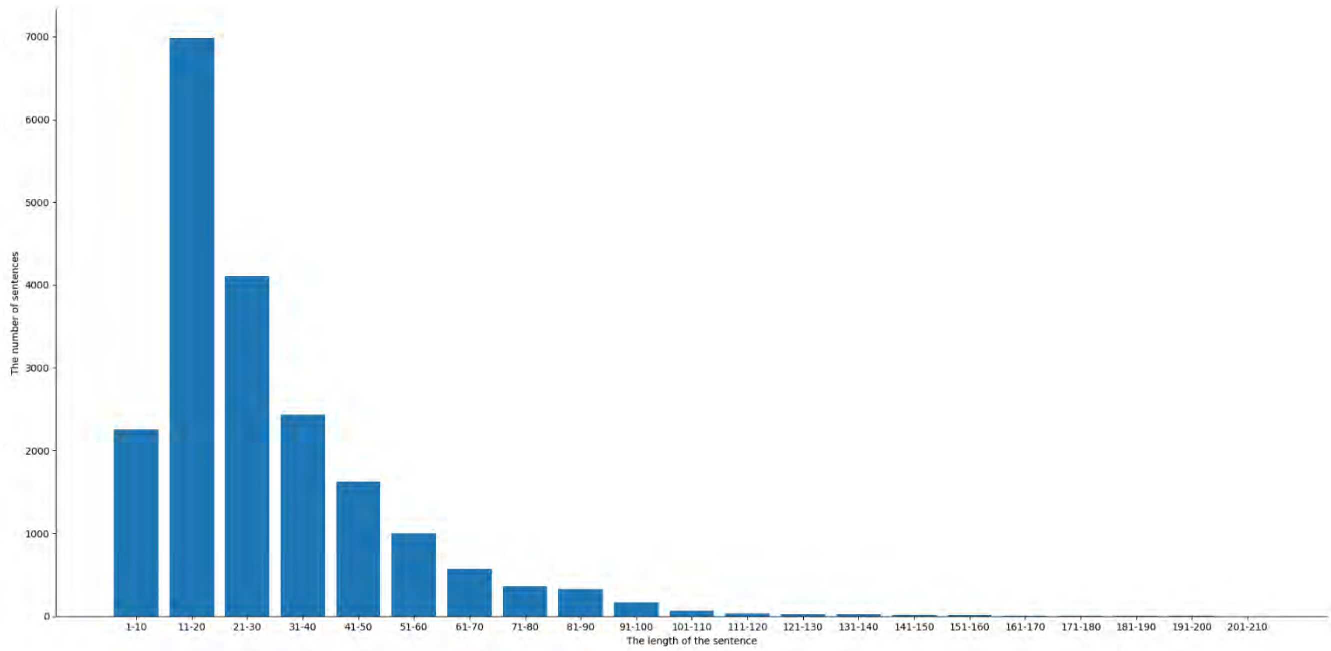


FIGURE 3. Sentence length distribution.

dataset into a training set (90%) with 18000 comments and a test set (10%) with 2000 comments.

FIGURE 3 shows the distribution of sentence length. It can be seen that the distribution of sentence length is uneven, with a maximum length of 220. Thus, we select 220 as the length of the word sequence. A sentence length of less than 220 is padded with 0. After the padding operation, each word in the word sequence is substituted by its index in the vocabulary we built and then converted to a dense vector that we pretrained in data preprocessing. Our model considers these dense vectors as input and returns a predicted value. The model eventually achieves a promising classification result after iterative error backpropagation and optimization.

We use accuracy to evaluate our model, accuracy means that the percentage of the output of the model is the same as that of the original label in the test data. The accuracy is defined as (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{2}$$

TN is the number of true negative cases. *FP* is the number of false positive cases. *FN* is the number of false negative cases. *TP* is the number of true positive cases.

We set training epochs as 60 and the early stopping method was applied to get the model weights with the best classification result. After every 50 iterations (An iteration means that the parameters of the model are updated in a batch size samples training.) of the model training, the model parameters will be stored if it achieves a higher *F1-value* in the validation set. After 3500 iterations we obtain the best model parameters. The accuracy of the model changes with the number of iterations as shown in FIGURE 4. The figure indicates that our model gradually converges after 2000 iterations with good accuracy. TABLE 4 presents the statistic of the result obtained using by our model. We can conclude that our model can automatically and efficiently extract the sentiment orientation of comments. Then, we can employ the sentiment feature to predict the P2P market trading volume.

The results of using the trained model to classify the original dataset are shown in TABLE 5.

TABLE 4. Precision, recall, and F1 measures.

PRECISION	RECALL	F1
0.9560	0.9278	0.9417

TABLE 5. Classification results.

Number of negative-sentiment comments	Number of positive sentiment comments
82,011	129,653

C. RELATIONSHIP BETWEEN SENTIMENT AND TVI

We calculated the percentage of negative comments for each day and generate a time series denoted as *ND*. We denoted the daily TVI in time series as *VD*. The changes in its trends are shown in FIGURE 5 and the *ND* trend in the same time span is shown in FIGURE 6.

It can be seen from the figures that the trading volume index fluctuates periodically. Therefore, in the trading volume prediction model, we need to consider this periodical feature.

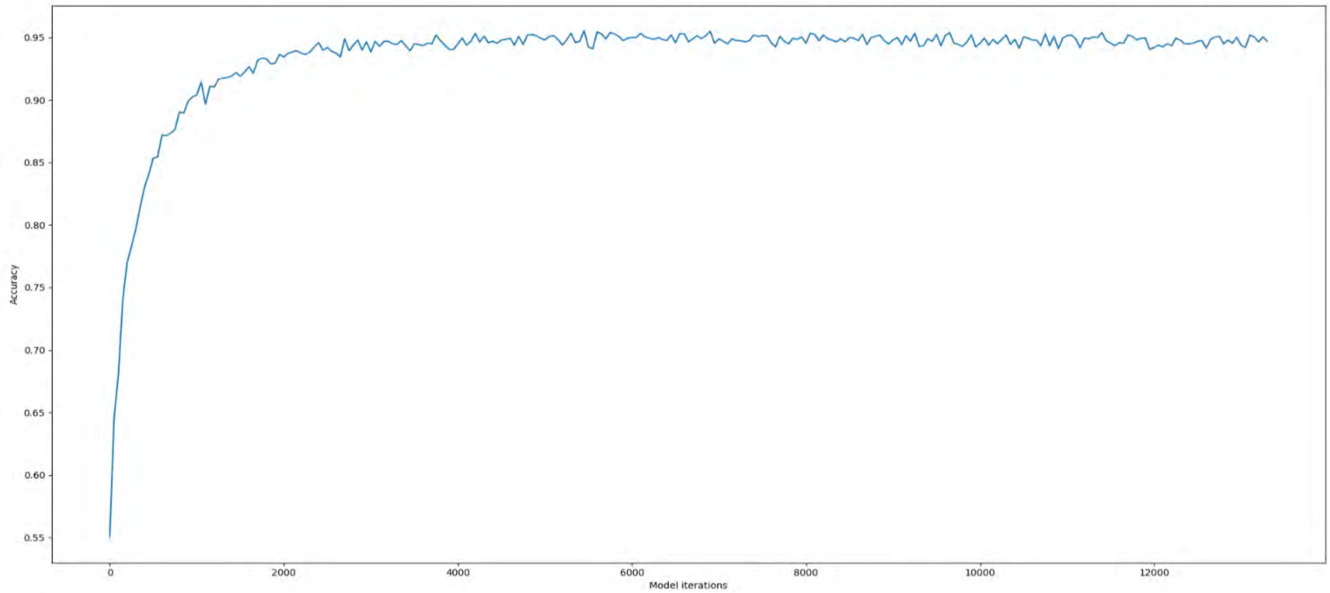


FIGURE 4. The learning curve of the sentiment classifier model (Model iterations represent the number of model parameter updates).

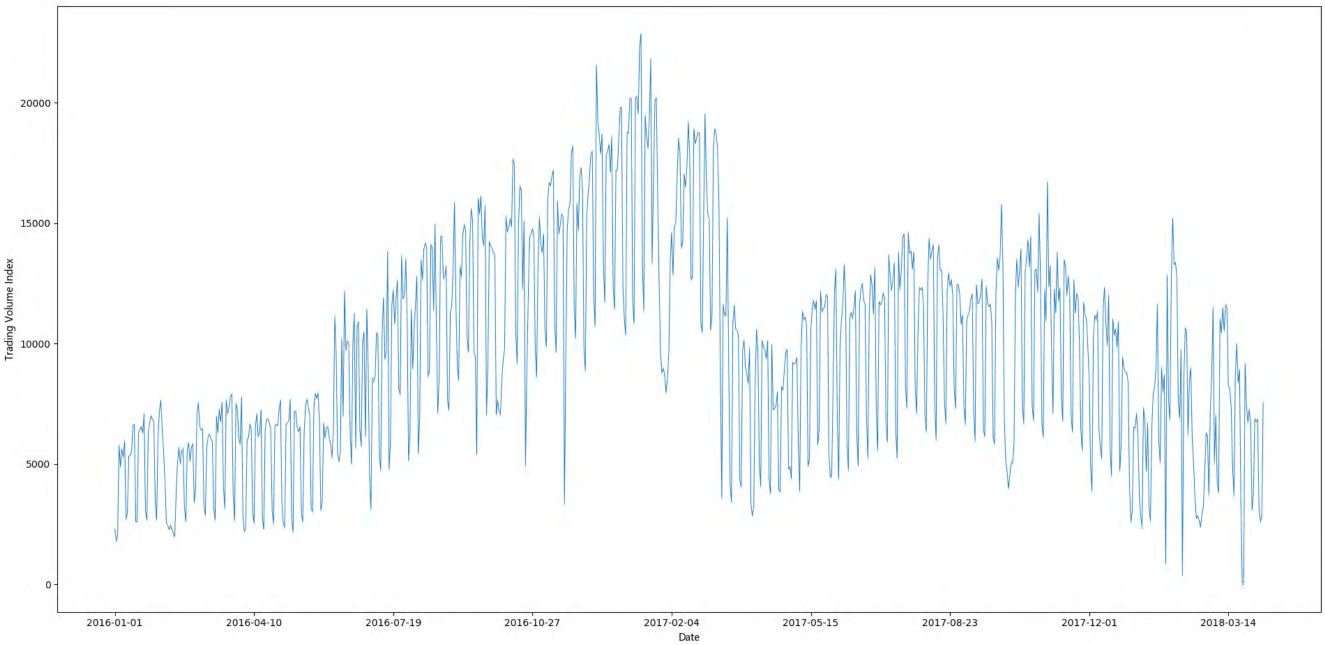


FIGURE 5. Daily trading volume index trends.

To compare ND and TVI, we use the z-score method to normalize the experimental data. $\bar{X}$ is the mean of X and σ is the standard deviation of X. The formula for calculating the Z score is

$$Z = \frac{X - \bar{X}}{\sigma}. \tag{3}$$

The processed data met the standard normal distribution with a mean of zero and a standard deviation of one. After the data were normalized, we found that the Pearson coefficient of VD and ND was -0.2097 and the p-value of this correlation was less than 1%. This correlation seemed to be negative, meaning that trading volume decreased as the proportion of negative sentiments increased. Of course, this relationship needed to be more rigorously tested.

D. GRANGER CAUSALITY TEST

The Granger causality test is a hypothesis test used to examine whether a set of time series can be used to predict another set

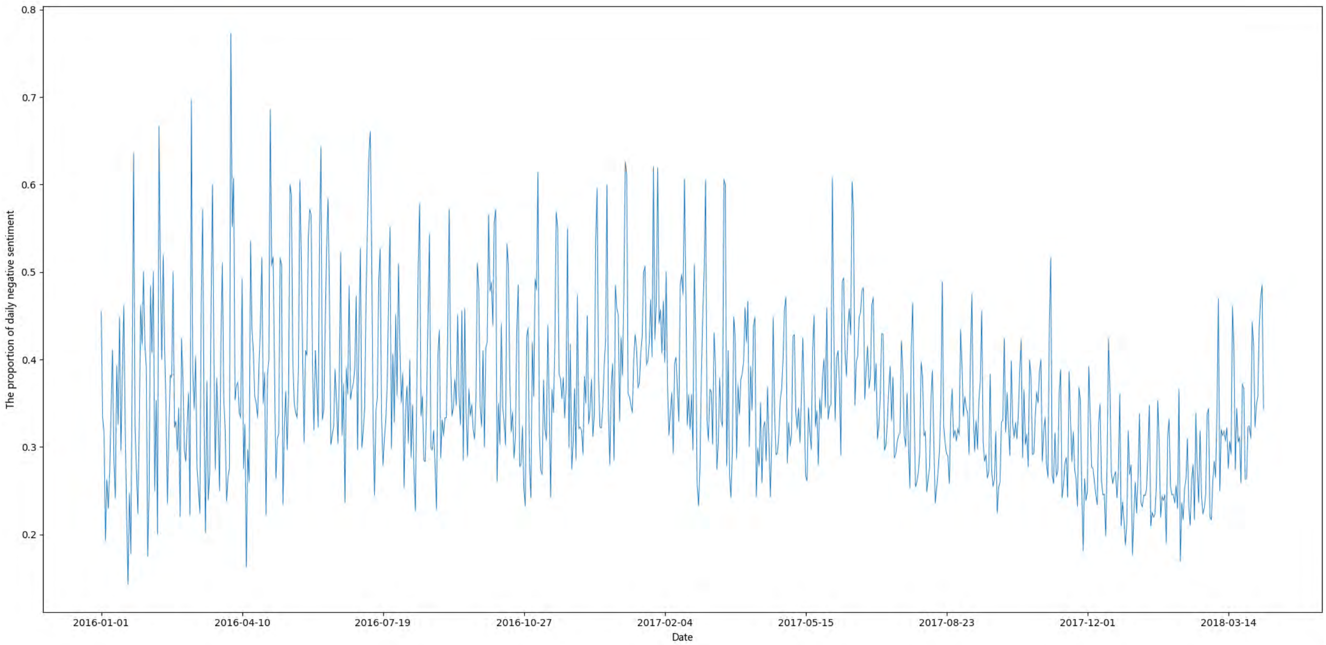


FIGURE 6. Daily negative-sentiment trends.

of time series [30]. If time series X is the Granger cause of time series Y, they have a statistically causal relationship.

The premise of the Granger causality test is the stationary time series. We first performed the Augmented Dickey-Fuller (ADF) test which is under unit root test, because for time series, unit root test is the stationarity test. We used the R language for this calculation. The null hypothesis was that the time series has unit roots; the calculation results are shown in TABLE 6.

TABLE 6. Value of the statistic test.

Name	*VD*	*ND*
T Value	-10.8384	-16.9337

The test statistic was less than 1% of the critical value, -3.96 which indicate that the two time series have no unit roots and thus are stationary time series. Thus, we could perform the Granger causality test. The null hypothesis of a Granger causality is that ND is not a Granger-causal of VD. In Granger causality testing, the p-value is the probability for the statistical model. The results of the Granger causality test are shown in TABLE 7.

All p-values were less than 1% and the null hypothesis was rejected, indicating a causal relationship between negative sentiments and changes in trading volume, as well as confirming that sentiment can be used to predict trading volumes.

TABLE 7. Grainger causality test results.

Lagged value	1	2	3	4	5	6	7
p-value	3.916 e-09	2.261 e-16	1.227 e-13	5.207 e-14	4.827 e-12	9.172 e-07	2.206 e-08

V. P2P TRADING VOLUME PREDICTION BASED ON SENTIMENT TRENDS

LSTM is suitable for processing and predicting time series events. LSTM-based models can be used for solving problems including text summarization, image recognition, and stock prediction. In this study, we built a deep learning model using LSTM. This model was used along with the sentiment trend time series to predict the TVI.

First, we observed cyclical changes in the *VD* fluctuations because P2P platforms usually issue bids (i.e., a borrower or investor issues a loan request or investment request, creating a loan project or investment project) on workdays and have fewer bids on the weekend. Therefore, we added the *WeekDay* variable to the prediction model; *WeekDay* is a feature that is represented by a seven-dimensional one-hot. To measure the effect of sentiment changes on the TVI, we set up a control group to conduct experiments. As the volatility cycle of VD is seven days, the control group used only the first seven days of the TVI and *WeekDay* as a feature, where t indicates the day we want to predict, and VD_{t-n} and $WeekDay_{t-n}$, respectively, represent the TVI and *WeekDay* n days ago. Control group 1 was denoted as L_1:

$$L_1 = \{\left(VD_{t-1}, WeekDay_{t-1}\right), \left(VD_{t-2}, WeekDay_{t-2}\right), \ldots, \left(VD_{t-7}, WeekDay_{t-7}\right)\}.$$

In addition, in order to verify the necessity of adding WeekDay to the model, we set up control group 2, which was

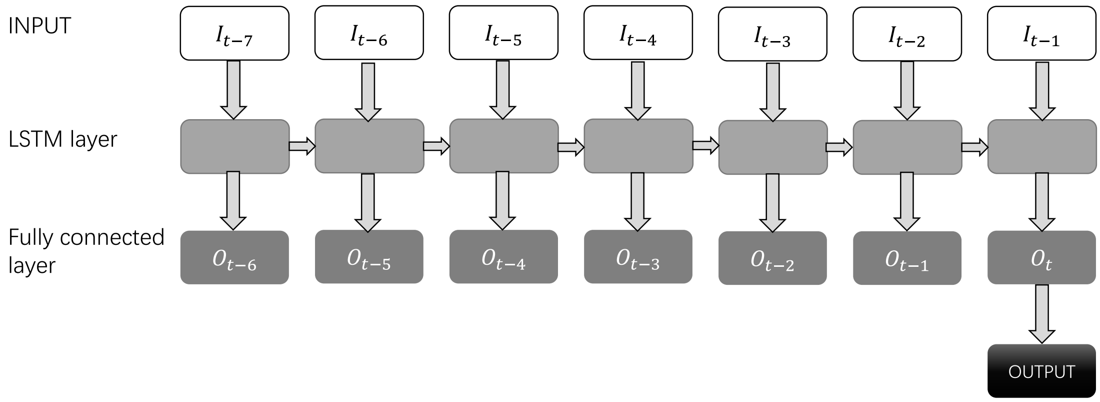


FIGURE 7. Structure of our LSTM model.

denoted as L_2:

$$L_2 = \{(VD_{t-1}), (VD_{t-2},), \ldots, (VD_{t-7})\}.$$

The experimental group incorporated emotion change characteristics, where ND_{t-n} represents ND n days ago, denoted as L_3:

$$L_3 = \{\left(VD_{t-1}, ND_{t-1}, WeekDay_{t-1}\right), \left(VD_{t-2}, ND_{t-2}, WeekDay_{t-2}\right) \ldots, \left(VD_{t-7}, ND_{t-7}, WeekDay_{t-7}\right)\}.$$

The LSTM model is shown in FIGURE 7. It consists of an LSTM layer and a full connection layer, where vector I and vector O represent the input and output of each step of LSTM, and OUTPUT is our final output.

VI. EXPERIMENTAL RESULTS

A. PARAMETER TUNING AND RESULT ANALYSIS

We obtained the TVI between January 1, 2016, and April 8, 2018, we obtained a total of 819 data points after processing. We used the first 600 data points as training data and the last 219 data points as validation data.

Initially, our training model did not consider the WeekDay feature. In the experimental group, the Pearson coefficient of the real and predicted values was only 0.70. After adding the weekday feature, the fitting effect was more significant. We used stochastic gradient descent (SGD) and Adam as the optimization algorithm.

The experimental results show that Adam can achieve better fitting results. We set a learning rate of 0.001 and 800 epochs because the loss of the model does not decrease after 800 training epochs. The overall model parameters are shown in TABLE 8.

After the model training was completed, we used the validation dataset for prediction. The results of the control groups are shown in FIGURE 8 and FIGURE 9 and the experimental group prediction results are shown in FIGURE 10. In these figures, the blue line represents the real value and the red line represents the predicted value.

TABLE 8. P2P trading volume prediction model parameters.

Parameter	Value	Description
N epoch	800	the number of training epochs
Cell Size	10	LSTM neuron size
Learning Rate	0.001	Model learning rate
Time Steps	7	LSTM step length
Batch Size	20	Batch size for each training

Pearson correlation coefficient, R-squared and mean squared error (MSE) were used to measure the prediction results. Pearson correlation coefficient (PCC) is a measure of the linear correlation between two variables X and Y. PCC is defined by (4), where cov is the covariance and σ is the standard deviation. Equation (5) is the formula for MSE, where $\hat{Y}$ is a vector of n predictions, and Y is the vector of observed values of the variable being predicted. R-squared presents the fitness of the predicted value and ground true value; it can be calculated using equation (6), where Y denotes the true value vector, $\hat{Y}$ denotes the predicted value vector, and $\bar{Y}$ denotes the mean number of true value vector. The results are shown in TABLE 9.

$$PCC = \frac{COV(X, Y)}{\sigma_X \sigma_Y}. \tag{4}$$

$$MSE = \frac{1}{n} \sum\nolimits_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}_i\right)^2}. \tag{6}$$

Compared with $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$, it can be seen that the model obtained better results after adding WeekDay. The Pearson correlation coefficient value increases by 11.86%, MES decreases by 23.11%, R2 increases by 40.56%, which also demonstrated that the ability of WeekDay to predict trading volume. In addition, we found that after adding ND, the correlation between the predicted and real values increased by

FIGURE 8. Control group 1 results.

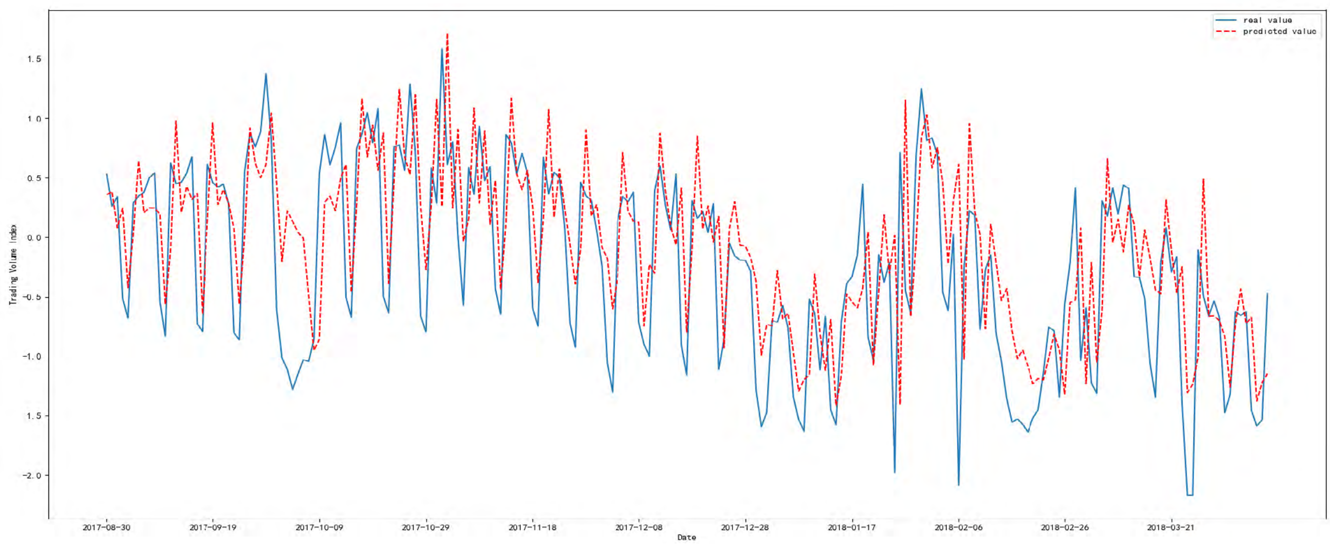


FIGURE 9. Control group 2 results.

TABLE 9. Pearson correlation confection between predicted and ground true value.

Symbol	Pearson correlation coefficient	MSE	R-squared
L_1	0.7355	0.3184	0.5105
L_2	0.6575	0.4141	0.3632
L_3	0.7994	0.2390	0.6324

8.62%, whereas the MSE decreased by 24.9%. R^2 increased by 23.9%. Thus, ND improved the accuracy of predicting P2P financial market trading volume somewhat, indicating that changes in sentiment tend to influence and predict P2P financial market trading volume.

B. COMPARISON OF MODELS

The stationarity of *VD* and *ND* has been proved by the ADF test, and the Granger casual check result has confirmed that they have a statistical causality relation. We consider the vector autoregression (VAR) model as a benchmark to verify the capability of our approach. The VAR model was proposed by Sims [31]; it is an extension of the one-dimensional autoregressive model. The VAR model regresses all lag variables of all variables using all current variables in the model. We also compared our result with the performance of the following methods: DNN, linear regression, random forest, and SVR. All benchmarks used the same training set and testing set. Further details are provided below.

VAR: The VAR model can be formulated with equation (7).

$$Y_t = \sum_{t=1}^{T} \alpha_t X_t + \beta_t Y_t \tag{7}$$

where X denotes a sentiment feature, Y denotes trading volume, and T denotes the number of lag days, which is set as 7 in our experiment.

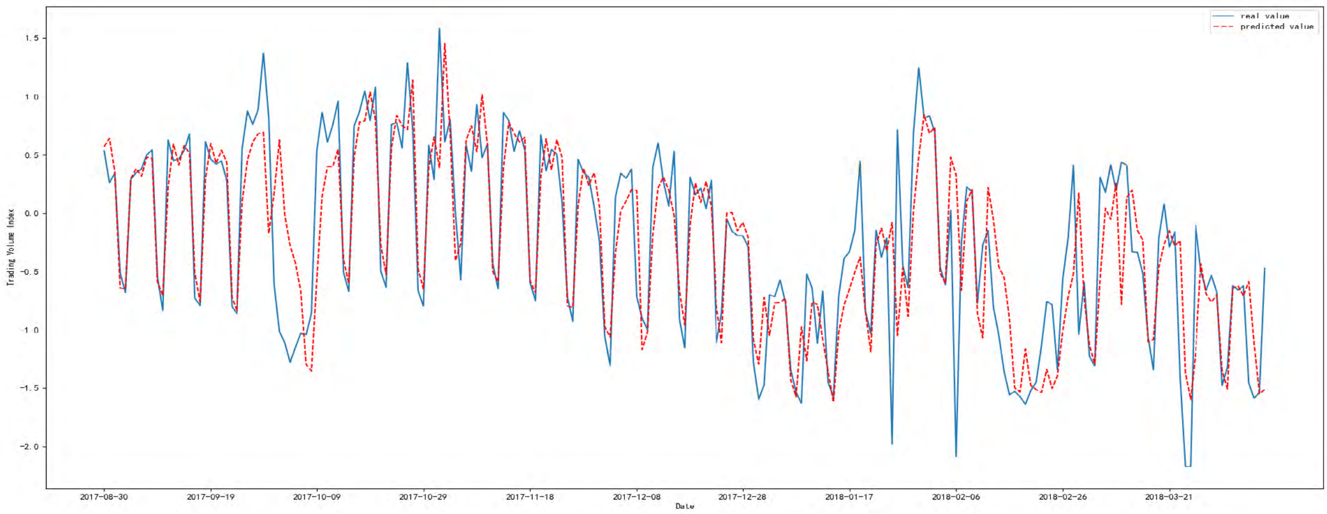


FIGURE 10. Experimental group results.

DNN: We use a fully connected DNN with six layers with sizes of 1000, 500, 300, 100, 10, and 1.

Linear regression: Linear regression is a regression analysis of the relationship between one or more independent variables and dependent variables using a least-squares function called a linear regression equation.

Random Forest: Random forest is a combinatorial classification algorithm for ensemble learning that can be used for classification and regression. We set the number of trees in the forest as 20.

SVR: Support vector regression is a regression version of support vector machine (SVM). We choose Gauss radial based kernel function to map the original data into a high-dimensional space.

The comparative results are shown in TABLE10:

TABLE 10. Experiment results.

	Pearson correlation coefficient	MSE	R-squared
LSTM	0.7994	0.2390	0.6324
VAR	0.6364	0.4083	0.3720
DNN	0.69293	0.3622	0.4430
Linear regression	0.7058	0.3302	0.4922
Random forest	0.6447	0.4054	0.3766
SVR	0.7113	0.3286	0.4945

LSTM has loops that allow the network to use information from previous passes; thus, it can store information as memory. These loops share weights, reducing the parameters and reducing the complexity of the model. DNN, SVR, and other models cannot record the relationship between time series. LSTM is more advantageous than DNN, SVR, and other models in predicting time series. Table 9 shows that LSTM has a better prediction effect than other models. This shows that using an LSTM-based trading volume predicting model is correct.

VII. CONCLUSION

The analysis of P2P financial markets is a very extensive application topic, although we have shown that it is helpful to use changes in investor sentiment to study trading volume and that this data may provide important reference information for government and enterprises.

In this paper, TextCNN was used for sentiment classification. Although the model was optimized, the precision and recall values obtained still have room for improvement. In addition, this study divided sentiments into only two categories: positive and negative. If we could further categorize sentiment, we would better grasp the changes in the P2P market.

Although the predictive value of the volume predicting model was closer to the actual value after the emotional index was added, the change in the peak and trough values could not be accurately predicted. The changes in the P2P financial market are related to many factors; therefore, to further improve the accuracy of the model, we will consider integrating more features into the prediction model.

REFERENCES

[1] A. Bachmann, A. Becker, D. Buerckner, M. Hilker, F. Kock, M. Lehmann, P. Tiburtius, and B. Funk, "Online peer-to-peer lending-a literature review," *J. Internet Banking Commerce*, vol. 16, no. 2, pp. 1–18, 2011.

[2] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.

[3] J. Si, A. Mukherjee, B. Liu, Q. Li, H. Li, and X. Deng, "Exploiting topic based Twitter sentiment for stock prediction," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2013, pp. 24–29.

[4] Z. K. Zhou, J. C. Zhao, and X. Ke, "Can online emotions predict the stock market in China?" in *Proc. 17th Int. Conf. Web Inf. Syst. Eng. (WISE)*, vol. 10041. Shanghai, China: Springer, 2016, pp. 328–342.

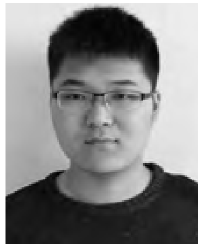
[5] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending," *Appl. Econ.*, vol. 47, nos. 1–3, pp. 54–70, 2015.

[6] R. Ge, J. Feng, B. Gu, and P. Zhang, "Predicting and deterring default with social media information in peer-to-peer lending," *J. Manage. Inf. Syst.*, vol. 34, no. 2, pp. 401–424, 2017.

[7] X. Chen, B. Huang, and D. Ye, "The role of punctuation in P2P lending: Evidence from China," *Econ. Model.*, vol. 68, pp. 634–643, Jan. 2018.

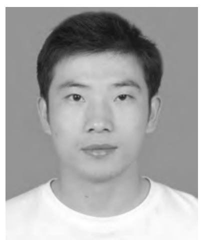
[8] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4621–4631, Jun. 2015.
[9] X. Lin, X. Li, and Z. Zhong, "Evaluating borrower's default risk in peer-to-peer lending: Evidence from a lending platform in China," *Appl. Econ.*, vol. 49, no. 35, pp. 3538–3545, 2017.
[10] Y. Xia, C. Liu, and N. Liu, "Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending," *Electron. Commerce Res. Appl.*, vol. 24, pp. 30–49, Jul./Aug. 2017.
[11] A. Byanjankar, M. Heikkilä, and J. Mezei, "Predicting credit risk in peer-to-peer lending: A neural network approach," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Cape Town, South Africa, Dec. 2015, pp. 719–725.
[12] L.-H. Li, C.-T. Lin, and S.-F. Chen, "Micro-lending default awareness using artificial neural network," in *Proc. 2nd Int. Conf. Multimedia Syst. Signal Process.*, 2017, pp. 56–60.
[13] K. Aleum and S.-B. Cho, "An ensemble semi-supervised learning method for predicting defaults in social lending," *Eng. Appl. Artif. Intell.*, vol. 81, pp. 193–199, May 2019.
[14] J. Duan, "Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction," *J. Franklin Inst.-Eng. Appl. Math.*, vol. 356, no. 8, pp. 4716–4731, May 2019.
[15] Y. Yoon, Y. Li, and Y. Feng, "Factors affecting platform default risk in online peer-to-peer (P2P) lending business: An empirical study using Chinese online P2P platform data," *Electron. Commerce Res.*, vol. 19, no. 1, pp. 131–158, Mar. 2019.
[16] J. Yang and D. Luo, "The P2P risk assessment model based on the improved AdaBoost-SVM algorithm," *J. Financial Risk Manage.*, vol. 6, no. 2, pp. 201–209, 2017.
[17] P. Liu, H. Li, and G. Sun, "P2P lending platform risk observing method based on short-time multi-source regression algorithm," *Comput. Sci.*, vol. 45, no. 5, pp. 97–101, 2018.
[18] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *J. Appl. Math.*, vol. 2014, Mar. 2014, Art. no. 614342. doi: 10.1155/2014/614342.
[19] M. Göçken, M. Özçalici, A. Boru, and A. T. Dosdogru, "Integrating metaheuristics and artificial neural networks for improved stock price prediction," *Expert Syst. Appl.*, vol. 44, pp. 320–331, Feb. 2016.
[20] L. A. Laboissiere, R. A. S. Fernandes, and G. G. Lage, "Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks," *Appl. Soft Comput.*, vol. 35, pp. 66–74, Oct. 2015.
[21] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI)*, Buenos Aires, Argentina, 2015, pp. 2327–2333.
[22] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Knowledge-driven event embedding for stock prediction," in *Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers (COLING)*, Osaka, Japan, Dec. 2016, pp. 2133–2142.
[23] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[24] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE Int. Conf. Big Data*, Santa Clara, CA, USA, Oct./Nov. 2015, pp. 2823–2824.
[25] N. Oliveira, P. Cortez, and N. Areal, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," *Expert Syst. Appl.*, vol. 73, pp. 125–144, May 2017.
[26] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14–23, Oct. 2014.
[27] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882
[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: https://arxiv.org/abs/1301.3781
[29] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Dec. 2016, pp. 649–657.
[30] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.
[31] C. A. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, no. 1, pp. 1–48, 1980.

XIANGLING FU is currently an Associate Professor with the Software School, Beijing University of Posts and Telecommunications (BUPT). She is also the Director of BUPT-China Huarong Group Joint Lab for AI Finance. She has participated in several projects funding by the National Natural Science Foundation of China. Her research interests include social network analysis, deep learning, and text mining.



SHUAI ZHANG is currently pursuing the master's degree with the School of Software Engineering, Beijing University of Posts and Telecommunications, China. His research interests include natural language processing and machine learning.

JINPENG CHEN is currently an Associate Professor with the School of Software Engineering, Beijing University of Posts and Telecommunications, China. His research interests include social network analysis, recommendation systems, data mining methodologies, machine learning algorithm, and information retrieval technique.



TIANXIONG OUYANG is currently pursuing the master's degree with the School of Software Engineering, Beijing University of Posts and Telecommunications, China. His research interests include natural language processing and machine learning.

JI WU (SM'15) has been the Director of Tsinghua-iFlyTek Joint Lab for Speech Technologies, since 2006. He is currently a Professor and the Deputy Director of the Department of Electronic Engineering, Tsinghua University, China. He is also heading the Multimedia Signal and Intelligence Information Processing Lab, Tsinghua University. His research interests include speech recognition, natural language processing, pattern recognition, machine learning, and data mining. He received the Second Prize of the National Science and Technology Progress Award, in 2011, and the First Prize of the Beijing Science and Technology Award, in 2014.