Research Article

Enterprise Financial Risk Prevention and Control and Data Analysis Method Based on Blockchain Technology

Julin Liu¹,² and Ken Chen³

¹School of Economics and Management, Hunan Institute of Science and Technology, Yueyang 414000, Hunan, China
²Department of Accounting, Hunan University of Finance and Economics, Changsha 410205, Hunan, China
³Business School, The University of Sydney, Sydney 2006, Australia

Correspondence should be addressed to Ken Chen; kche9534@uni.sydney.edu.au

Received 27 April 2022; Revised 18 May 2022; Accepted 26 May 2022; Published 23 June 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Julin Liu and Ken Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of engineering technology and computer networks, artificial neural networks, which mimic human brain neural networks, are being used in financial market forecasting to improve the accuracy of stock predictions and are making significant progress. Therefore, there is a great need to actively investigate the method of financial data analysis based on blockchain technology. The purpose of this paper is to investigate the neural network method of financial data analysis based on blockchain technology. Shanghai Index and Shenzhen 200 Index are chosen as experimental data, which are divided into two subsets: training and test samples. The BP model is constructed based on blockchain technology to analyze MARE, RMSRE, MSPEE, RMSPE, and MARE errors. The results show that the mean absolute error rate (MARE), RMSPE, and MSPEE of training samples of blockchain-based BP model are 0.0056, 0.0787, and 0.0085, respectively. Blockchain-based BP model plays an important role in solving financial data analysis problems.

1. Introduction

With the continuous enhancement of the operation of the socialist market economy system, the financing of funds plays an increasingly important role [1]. The market for the national (regional) economic development and financial stability of the building has made an invaluable contribution [2]. At the level of modern economics, financial markets level to promote rapid economic development of a country or region to some extent. Reference [3]. However, it is very difficult to predict and judge the trend of financial market operation [4]. Financial business contains a lot of information processing tasks, and the effective use of information plays a very important role [5, 6].

Due to the rapid development of information science and computer network technology, interdisciplinary research has promoted interdisciplinary integration [7]. Many people use many other research methods to study financial market predictions, but this is very difficult [8]. The imperfection of the existing financial market theory limits its prediction space [9]. Global financial markets are linked to the particularity of the forecast time span. This connection can be challenging for the financial market forecasts [10]. Many improved methods and technologies as well as new specialties have been used in the prediction of financial markets [11]. Among them, artificial neural networks that simulate the human brain have strong self-learning capabilities. It is often used by everyone in imitating the financial system and forecasting financial markets. And it is loved in stock market analysis [12].

The research on the financial neural network model has been carried out by many scholars. Mar established a fuzzy neural network model and an ARIMA time series model, respectively, and conducted single-step and multi-step forecasting studies on the stock price of listed company SAIPA. The comparison of prediction results shows that the fuzzy neural network is better than the traditional ARIMA model. Reference [13]. Ganesh studied that in the field of financial engineering, when using artificial neural network technology to make predictions, there is no basis for
selecting input variables [14]. Alam uses neural networks to predict the daily maximum and minimum stock prices of Brazilian power distribution companies. In his research, he used correlation to analyze the choice of input variables, and tested different experimental methods based on experience. The best connection point is a hidden layer, 5 hidden layer neurons [15]. Tamal uses his own network model structure and main components to predict the value of an exchange. The results of his research found that his proposed model outperformed his competitors. In addition, the principal component analysis method selects 20 accounting variables, which can more accurately predict stock prices [16]. Ahmed uses artificial neural network structure to study the stock market and price. The richer the technical data collected in different periods, the better the prediction performance of the model. However, the predictive performance of the model will change with changes in parameter variables and indices [17]. Smith used the ANN model to study market securities in Taiwan based on experience, and then obtained 10 input variables [18]. The data of these studies are not comprehensive, and the results of the studies are still open to question, so they cannot be recognized by the public and thus cannot be popularized and applied.

The innovation of this paper is that the CSI 200 index is selected as the experimental data, which is divided into two subsets: training samples and test samples. And built is a BP model based on genetic algorithm, used for MARE, RMSRE, MSPEE, RMSPE, and MARE error analysis. The article summarizes the description of some concepts of the financial market, so as to have a clear understanding of the financial market.

2. Proposed Method

The financial market is a dynamic system, and the choice of control variables can improve the accuracy of model prediction. By studying the statistical characteristics of certain financial market observation indicators (trading rate, stock index, technical indicators, etc.), you can find the operating rules of financial indexes [19–21]. The main variables in financial markets are listed below.

Closing price refers to the last transaction price on the trading day. If there is no transaction on that day or there is no transaction under the circumstances of market closure, the closing price of the most recent trading day shall be used as the closing price of that day. The historical time series data of closing price can be used to predict the trend of stock index; the opening price (opening price) refers to the first transaction price on the trading day, which is determined by the rule of maximum transaction amount. If there is no transaction for a certain period of time (usually 30 minutes) or several consecutive days, the opening price of the latest trading day shall be used as the opening price of that day; the highest price refers to the highest price generated in all transaction records of a certain security from the opening to closing time on the trading day. If there is no transaction on that day due to market closure or trading suspension, the lowest price recorded on the most recent trading day shall be used; volume is the total volume of transactions counted in a specific period, and the smallest unit is hand. It reflects the supply-demand relationship of a certain stock and can judge the general trend of the market based on it; the transaction amount is the total transaction amount in a specific period. According to this, we can roughly analyze the market fluctuation trend and the flow direction of funds; use the closing price of a certain security on the same day to subtract from the closing price of the previous day, with a positive sign (+) for up and a negative sign (−) for down, otherwise it is flat, indicating the change direction of the price.

The moving average shows the ups and downs of securities in the past, objectively reflects the state of the stock market, and provides relevant reference information for stock market participants to identify trends and make trading decisions. It is drawn by connecting the average values of different periods (5 days, 6 days, 10 days, 20 days, 30 days, 60 days, and 120 days). The calculation formula is

\[
MA_n = \frac{\sum_{i=1}^{n} P_i}{n},
\]

\[
MACD = EMA \text{ (Close, 12)} - EMA \text{ (Close, 26)}.
\]

Momentum index measures the change of closing price in each time, reflecting the relationship between price and supply and demand.

\[
MTM = C - C_n,
\]

\[
MTMMA = \frac{\sum_{i=1}^{m} MTM_i}{m},
\]

wherein C is the closing price of the day, \( C_n \) is the closing price of \( n \) days ago, and \( n \) is the setting parameter, the value range is generally between 6 and 14 days, usually 10 days, and the parameter \( m \) is generally set to 6 days.

The stochastic shock indicator is a kinetic energy indicator, mainly fluctuating between 0 and 100. The indicator is mainly composed of two lines, the % K fast line and the % D slow line. It has bullish/bearish divergence signals, bullish/bearish trend line crossing, and overbought/oversold regional functions. The BOLL indicator is composed of three rail lines (Brin upper rail, Brin middle rail, and Brin lower rail). Bollinger upper rail and Bollinger lower rail constitute the pressure line and support line of the stock price. Bollinger Middle Rail is the average price of Bollinger Upper and Lower Rails. RSI refers to Wilder’s research on futures trading based on pioneering and measuring the strength of both parties. Comparing the ratio of the rising average closing index to the sum of rising and falling over a period, analyzing the market’s intentions and future market trends, can be used as a “barometer” or “alarm” technical indicator for the financial market [22–24]. The calculation formula is


The calculation formula is

\[ \text{PSY} = \frac{\text{number of rising days}}{N} \times 100\%. \quad (5) \]

In the formula, \( N \) is generally set to 12 days, the maximum is no more than 24, and the longest is no more than 26.

### 3. Theoretical Basis

Biology has stimulated the study of artificial nerves to some extent to simulate the actual physiological structure of the human neural network and the thinking function of the human brain, so as to obtain simulations of information about the human brain.

The input vector of the artificial neuron is as follows:

\[ P = [P_1, P_2, P_3, \ldots, P_N]^T. \quad (6) \]

The weight vector of the neuron is as follows:

\[ W = [W_{1,1}, W_{1,2}, W_{1,3}, \ldots, W_{1,R}]^T. \quad (7) \]

Its matrix form is

\[ n = WP + b. \quad (8) \]

Output of neurons is as follows:

\[ a = f(WP + b), \]

where

\[ a = \frac{1}{1 + e^{-n}}. \quad (9) \]

Through coding, it has unique advantages in dealing with some non-numerical concept problems. Using random search technology and using multiple search points at the same time, it has good parallelism. The value of the target function is directly used as the search information, which increases the flexibility of the search process. This is the advantage of the blockchain technology. The natural mechanism is used to display complex phenomena and facilitate the application of genetic computation. Blockchain technology (BT) is widely used in engineering optimization, financial forecasting, machine learning system, and other fields due to its unique problem-solving ability, strong search ability, and scalability [25–28].

Coding is based on the process that blockchain technology knows nothing about the solved problem itself, and establishes a corresponding relationship between the feasible solution data of the problem to be solved and the chromosomes that the blockchain technology can handle. Although the requirements of coding are not strict, the coding method greatly affects the implementation of crossover and mutation operations.

### 4. Experiments

#### 4.1. Experimental Preparation

#### 4.1.1. Subjects

Basic data for this experiment selected for model training and testing are the Shanghai and Shenzhen 200 indexes. The Finance site provides data from Yahoo. The representative main component index is the Shanghai and Shenzhen 200 index. It reflects the overall performance of the Shanghai and Shenzhen stock markets. As shown in Table 1, neural network input variables are the ordinary stock market variable, and the output variable is the closing price of the stock index the next day.

#### 4.1.2. Sample Collection

The selected moving average, index moving average, and triangle moving average are divided into four-time scales of 15 days, 16 days, 20 days, and 25 days, respectively. Due to the calculation of the 20-day simple moving average of the closing price, the closing price data of 20 trading days before today is required. Therefore, the sample collection period of this paper is from June 27, 2015 to December 21, 2018. Excluding the influence of holidays and other factors, there are 3620 groups of stock price and trading volume data.

#### 4.1.3. Data Structure

Technical index data is constructed by using time series data such as stock price and trading volume. According to each index formula, technical index value is obtained by adding, subtracting, multiplying, and dividing. The experiment data of 3600 rows and 54 columns are obtained and divided into training samples and test samples. The first 3500 sets of experimental data are used for training, the last 100 sets of data are test sets, and the closing price of stocks in 100 trading days is predicted in one step.

#### 4.1.4. Experimental Data Processing

All data obtained were analyzed using spss22 software. Sample data were subjected to a normal test; data corresponding to a normal distribution were subjected to an analysis of differences between the two groups, expressed as mean ± standard deviation; data not corresponding to a normal distribution were subjected to an analysis of differences between the two groups, expressed as Q50 (q25, Q75) (Q: quartile, quartile). The chi-square test was used to analyze the differences between the two groups.
An independent sample $t$-test was used to analyze differences between the two groups. Paired analysis was used to measure the correlation between the two samples. First, the normality of the data was checked, and if the data were normal, the difference between the paired groups was analyzed using the paired sample $t$-test; if the data were not normal, the difference between the paired groups was analyzed using the Wilcoxon test. The Wilcoxon test is a nonparametric test, meaning that it does not depend on data belonging to the family of probability distributions for any particular parameter.

### 4.2. Establishment of Experimental Model

PCA is principal component analysis technology, also known as principal component analysis technology, which aims to use the idea of dimensionality reduction to convert multiple indicators into a few comprehensive indicators. The basic idea of PCA is to project high-dimensional spatial data into low-dimensional space. This is our expression of the original data, which makes those variables with large differences play a greater role in classification. Another method of data dimensionality reduction is to select variables, that is, under the supervision of a certain performance index, selecting those variables which have greater effect on classification from the original data variables. This blockchain technology is implemented through special operations such as copying and crossover. After the blockchain technology uses the appropriate fitness function to “guide”, the next generation can get the high-quality genes of the previous generation. It leads to high-quality individuals who can survive with a relatively high probability. Obviously, the initial variable can be selected under the supervision of a certain performance index. Another way to reduce the data dimension is to select variables, that is, under the supervision of a certain performance indicator, select those variables that have a greater impact on the classification from the original data variables.

When using PCA for dimensionality reduction, the average of 100 experiments, and after repeating the experiment the results are obtained.

### 4.2.2. The Algorithm’s Model Efficiency Score

In order to evaluate the effectiveness of the model, good estimation methods are needed, as well as estimation criteria to measure the generalizability of the model. The root mean square error (MSE) is very similar to the mean absolute error. The only difference is that MSE is the square of the average difference between the initial value and the predicted value. The advantage of MSE is that the gradient is easier to calculate, whereas the mean absolute error requires sophisticated linear programming tools. As we use the square of the error, the impact of the larger error becomes more obvious, so the model can now focus more on the larger error. It is common to use the root mean square absolute error (RMSE) and the number of influencing factors for comparison and testing. MSE is a measure of the deviation between the actual value and the predicted value, so the lower these values are, the better the predicted result.

### 5. Discussion

#### 5.1. Data Analysis of BT-BPNN Model

As shown in Table 3, the number of variables in the input layer in the neural network is 34 and the number of output layers is 1, so the value range of the number of hidden layer nodes is [10, 14]. Gradually increase the hidden layer. The number of neurons in the layer, the mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) in Section 4 are used as performance indicators. Each experiment takes the average of 100 experiments, and after repeating the experiment the results are obtained.

As shown in Table 3, at the same training time, when the number of hidden layer neurons is 12, the prediction error is the smallest. Its MAE is 22.46, MSE is 1006.65, and RMSE is 27.20; at the same training time when the number of hidden layer neurons is 10, its MAE is 23.67, MSE is 1013.26, and RMSE is 23.56; at the same learning time, when the number of hidden layer neurons is 11, its MAE is 23.35, MSE is 1053.35, and RMSE value is 34.22; at the same learning time,
when the number of hidden layer neurons is 13, its MAE is 33.01, MSE is 1029.99, and RMSE value is 35.03. When the number of hidden layer neurons is 14, its MAE is 37.89, MSE is 1684.76, and RMSE value is 46.07. With the same learning time, when the number of neurons of the hidden layer is 14, its MAE is 37.89, MSE is 1684.76, and RMSE value is 46.07.

The neurons receive information from the outside world and weight it as parameters of the transfer function. As the hidden layer must reflect the nonlinear relationship between the input and output layers, it is more appropriate to choose the sigmoid tangent as the transfer function of the hidden layer, while the transfer function of the output layer is a purely linear transfer function. The Levenberg–Marquardt learning algorithm is applied, the learning rate is 0.02, the accuracy of the learning target is set to 1e-6, and the learning frequency is 100.

As shown in Figure 1, BPNN has 44 neural network input dimensions when one neuron participates, 2 neural network input dimensions when 2 neurons participate, and neural network input when 3 neurons participate the dimension is 19, and the input dimension of the neural network when 4 neurons participate is 16.

Use the neural network toolbox of MATLAB R2018a software to write the BPNN model running program. BP neural network training sample input variable input train is a matrix of 3600 rows and 54 columns, and training sample output vector is only the next day in Shanghai and Shenzhen 200 index closing price. The training sample input BPNN model learning and training, BPNN model simulation results are consistent with the training samples.

As shown in Figure 2, the red curve is the actual data of the closing price of the Shanghai and Shenzhen 200 Index, and the blue curve is the predicted fitting data of the training sample. Through the comparison between the two sets of different data, in the training results, the data results obtained by the BPNN model are more perfect than the other. After the BPNN model performs learning training on the training samples, it generates a higher-quality prediction function, and the training samples have smaller fitting errors.

5.2. Neural Network Blockchain Technology Based on Analysis of Financial Data. BPNN, the trained model, is used to predict the next 100 days Study CSI closing price index 200, and outputs the result of the prediction error of the test sample.

As shown in Figure 3, the prediction effect of the BPNN model in the training sample is satisfactory, and the prediction accuracy is higher when the gap between the predicted value and the actual value is small.

Based on the standardized data matrix, the MATLAB principal component analysis Principe program is called to calculate the principal component coefficients, principal component scores, and eigenvalues of the original data matrix. The principal component contribution rate and the cumulative contribution rate are calculated based on the characteristic values. The larger the eigenvalue, the larger the variance contribution rate.

As shown in Figure 4, the eigenvalues of the CSI 200 Index after the principal component analysis fluctuate below the 20 range, the eigenvalue contribution rate fluctuates within the range of 40, and the cumulative contribution rate of the eigenvalues is 40–80 due to its cumulative attributes fluctuation within the range, explaining that the relationship between the three has a certain proportional relationship and is positively related.

Investigate the blockchain technology’s ability to optimize variables and improve the prediction effect of neural
networks on stock index. The prediction effects of BT-BP neural network prediction model, IBT-BP neural network model, and PCA-BPNN with BP neural network prediction model and principal component analysis as preprocessing methods are compared, respectively. The BP neural network training algorithm uses the LM algorithm, and the number of trainings is fixed at 1000.

As shown in Figure 5, the performance of analysis and neural combined model (PCA-BPNN), combination of blockchain technology, and the neural network model combining can be obtained. The prediction accuracy of BP is greatly affected by the input variables on the collection neural network model (BPNN). Principal component analysis or blockchain technology optimization of input variables will improve the prediction accuracy of BP and is better than the prediction effect of BP model.

6. Conclusions

The research background of this paper is that with the rapid development of information science and technology and computer networks, interdisciplinary research promotes interdisciplinary integration. The trading volume is the total trading volume in a specific period of time, which reflects the supply and demand of a certain stock, and can be used to judge the overall trend of the market; the transaction amount is the total transaction amount in a specific time period. Artificial neural networks that simulate human brain neural networks are used in financial market forecasting to improve stock forecasting accuracy and achieve significant development. Therefore, it is very necessary to actively carry out the research of financial data analysis method based on blockchain technology. There is a problem of uneven data distribution in this study, which leads to some randomness in the conclusion of the experiment.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author(s) declare that there are no conflicts of interest.

Acknowledgments

This work was supported by National Social Science Fund: Research on global supply chain disruption risk management of high-tech enterprises in China, 20BJY006.
References


