Research Article

Construction and Simulation of the Market Risk Early-Warning Model Based on Deep Learning Methods

Yuchen Lei$^{1}$ and Yinghui Li$^{2}$

$^{1}$School of Finance, Shandong University of Finance and Economics, Jinan 250000, China
$^{2}$Henan Finance University, Zhengzhou 450046, China

Correspondence should be addressed to Yinghui Li; liyinghui@hafu.edu.cn

Received 21 January 2022; Accepted 23 February 2022; Published 24 March 2022

To address the problem of low efficiency of existing forecasting models for market risk warning, a market risk early-warning model based on improved LSTM is suggested utilizing the whale optimization algorithm (WOA) to optimize the number of hidden layer neurons and time step parameters of long short-term memory. The proposed market risk early-warning model is validated by using 40 real estate companies as the research subjects and 20 relevant variables such as gross operating income, net profit asset growth rate, and total asset growth rate as indicators. The results demonstrate that the proposed model’s prediction accuracy for market risk is greater than 96% and that when compared to the standard CNN and LSTM models, the suggested model’s prediction accuracy for corporate finance from 2012 to 2019 is increased by 14% and 12%, respectively, and the prediction accuracy for corporate finance in 2020 is improved by 22% and 7%, respectively, which has certain practical application value and superiority.

1. Introduction

In recent years, affected by the trade war between China and the USA and the global coronavirus pandemic, the real estate industry across the country has been severely impacted and caused varying degrees of bubbles and other problems, putting the real estate market and the entire national economy in jeopardy. To ensure a healthy and stable development, macroregulation based on early-warning information provided by the real estate market is crucial. To this end, based on Internet big data, Jiang et al. suggested a support vector machine-based (SVM) real estate market risk early-warning model [1]. China A-share listed real estate companies in 2019 are used as the study subjects. Using the random forest algorithm to select five important feature dimensions, current ratio, equity financing ratio, operating income, current liability ratio, and receivables’ turn, the financing risk prediction of real estate companies is achieved by collecting the relevant companies’ financial information from 2010 to 2019 and supplementing the risk sample data from 2005 to 2010. Using Philadelphia as the study object, Junchi et al. suggested an improved regression tree (BRT) for merging urban data, including metadata and image data, with home features to estimate the market value of Philadelphia housing at the projected level [2]. Alvarez et al. proposed to forecast house values, using publicly available information on geography, city characteristics, traffic, and real estate for sale by a tree-based incremental learning model and allowing for early warning of real estate risk. Using massive datasets for training and incremental learning to deliver accurate price projections on a daily basis, the model’s prediction accuracy was enhanced [3]. García-Magarínó and Lacuesta analyzed and predicted the possible buying and selling behavior in the real estate market based on agent’s simulation tool, by taking Spanish real estate as a research object and simulating real estate transactions, which can effectively warn the market risk in the real estate industry [4]. Zhou et al. assessed the real estate market’s internal and external environments and a PSO-SVM model-based, and the real estate risk early-warning model was proposed, which accurately predicts cyclical real estate risk in Beijing and has good early-warning performance [5].
Based on the DEA-Malmquist method, Chen et al. predicted corporate assets by analyzing the inventory manifestation of the Chinese real estate industry from 2005–2015, concluding that there may be zombie enterprises and the risk of future unemployment [6]. Kamara et al. proposed a new hybrid neural network model with CNN attention (CNNA) and bidirectional LSTM (BLSTM)-based modules to extract features to tackle the Day-of-Market (DOM) prediction problem [7]. According to the estimated distribution of the characteristics, confidence intervals for the four properties in the dataset were derived from percentile Bootstrap confidence intervals (CI) or percentile bias-corrected accelerations’ (BCA) Bootstrap CI. Finally, proposed method’s superiority to the DOM prediction problem was demonstrated and the prediction accuracy reached 87% by conducting experiments on the dataset of a well-known real estate agency in Shanghai. By investigating the association between financial stability and real estate price volatility in China and utilizing detrended cross-correlation analysis, Liu et al. proved the interrelationship between financial stability and the real estate market [8]. We use multiple fractals’ asymmetric detrended cross-correlation analysis (MF-ADCCA) to assess scalar features of the correlation between financial stability and estates’ price volatility to achieve monitoring and early warning of that. According to the above related research results, it is clear to observe that deep learning-based early-warning models have advantages in real estate market risk warning and can predict real estate market risk more accurately, with an overall prediction accuracy of about 80%, but its prediction accuracy still needs to be improved. Therefore, this research provides an enhanced LSTM real estate market risk early-warning model based on the LSTM model that utilizes WOA algorithm to maximize the number of hidden neurons and time step for increasing prediction accuracy.

2. Basic Methods

2.1. Introduction to LSTM Networks. LSTM is a temporal recurrent neural network that uses a “gate” structure to overcome the difficulties of gradient disappearance and long-term reliance in recurrent neural network (RNN) [9]. Its basic structure is shown in Figure 1, which consists of input gate, output gate, and forget gate.

In Figure 1, $x_t$ denotes the network input at moment $t$, $h_{t-1}$ and $C_{t-1}$ denote network output and cell state output at $t$-1 moment, $\delta$ denotes sigmoid function, and the mathematical expression is shown as (1), $\tanh$ denotes activation function, and the mathematical expression is shown as (2), and $\odot$ and $\oplus$ denote the Hadamard product and summation, respectively:

\[
s(x) = \frac{1}{1 + e^{-x}}, \quad (1)
\]

\[
\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (2)
\]

The network output at moment $t$ is as follows [10, 11]. Input gate:

\[
i_t = \text{sigmoid}(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i). \quad (3)
\]

Forget gate:

\[
f_t = \text{sigmoid}(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f). \quad (4)
\]

Cell state:

\[
C_t = f_tC_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c). \quad (5)
\]

Output gate:

\[
o_t = \text{sigmoid}(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o). \quad (6)
\]

Network output:

\[
h_t = o_t \tanh(C_t). \quad (7)
\]

In which, $W$ and $b$ are the relative weight coefficient matrices and bias vectors.

The LSTM model is highly efficient, but it is difficult to find the best combination of parameters due to the large number of model parameters and the large amount of computational resources required to combine the relevant parameters, which in turn leads to poor model prediction performance [12]. Therefore, this study employs the whale optimization approach to improve the model prediction performance by optimizing the LSTM model parameters.

2.2. LSTM Network Improvements

2.2.1. An Introduction to the WOA Algorithm. The WOA algorithm is an optimization algorithm presented by Seyedali Mirjalili et al. to model humpback whale hunting behavior [13]. The algorithm uses a hypothetical method to represent the optimal solution, mathematically expressed as follows [14]:

\[
\overrightarrow{D} = |\overrightarrow{C} \times \overrightarrow{X} - \overrightarrow{X}(t)|, \quad (8)
\]

\[
\overrightarrow{X}(t + 1) = \overrightarrow{X} + \overrightarrow{A} \cdot \overrightarrow{D}, \quad (9)
\]

where $t$ is the current iteration, $\overrightarrow{A}$ and $\overrightarrow{C}$ denote the coefficient vector, $\overrightarrow{X}$ and $\overrightarrow{X}^*$ denote the position vectors of the
current solution and the optimal solution, respectively, denotes taking the absolute value, and \( \cdot \) denotes the element product. When the iteration produces an optimization solution, \( X^* \) will update, and then, \( \bar{A} \) and \( \bar{C} \) can be calculated by (8) and (9):

\[
\bar{A} = 2\bar{a} \cdot \bar{r} - \bar{a}, \quad \bar{C} = 2 \cdot \bar{r},
\]

(10)

in which the linearity drops from 2 to 0 during the \( \bar{a} \) operation, which denotes that random vectors take values in the range [0,1].

As illustrated in Figure 2, the WOA algorithm’s search mechanism incorporates a shrinking encircling mechanism and spiral updating position. The specific calculation method of the spiral updating position is to prioritize the distance between the whale’s position \((X, Y)\) and its prey \((X^*, Y^*)\) to simulate the whale’s spiral movement by establishing the spiral equation [15]:

\[
X(t + 1) = \bar{D}^t \cdot e^{bt} \cdot \cos(2\pi l) + X^*, \quad \text{if } p < 0.5,
\]

\[
\bar{D}^t \cdot e^{bt} \cdot \cos(2\pi l) + X^*, \quad \text{if } p \geq 0.
\]

(11)

where \( e^{bt} \) denotes logarithmic spiral, \( b \) denotes constants, and \( l \) denotes random numbers taking values in the range [-1,1].

Assuming that the probability of choosing one of the search mechanisms is 0.5, then [16]

\[
X(t + 1) = \begin{cases} 
\bar{X}^* (t) - \bar{A} \cdot \bar{D}, & \text{if } p < 0.5, \\
\bar{D}^t \cdot e^{bt} \cdot \cos(2\pi l) + X^*, & \text{if } p \geq 0.
\end{cases}
\]

(12)

In addition, the WOA algorithm can search for targets based on random shapes, variables, and vectors. This search mechanism is consistent with \( \bar{A} > 1 \), emphasizes exploration, and allows the execution of global searches, as modeled below [17]:

\[
\bar{D} = |\bar{C} \cdot \bar{X}_{\text{rand}} - \bar{X}|,
\]

(13)

\[
X(t + 1) = \bar{X}_{\text{rand}} - \bar{A} \cdot \bar{D},
\]

(14)

where \( \bar{X}_{\text{rand}} \) is the position vector in the whale population.

The WOA algorithm has fast convergence speed and strong search capability [18], so it is used in this paper to optimize the LSTM parameters.

2.2.2. WOA-Based LSTM Parameter Optimization. The prediction accuracy in LSTM networks is mainly affected by the number of hidden layer’s neurons \( m \) and the time step \( c \) [19], so the optimization of LSTM parameters by WOA is mainly for \( m \) and \( c \). Currently, the number of \( m \) is usually determined as an approximate range according to the empirical (13), and the value of \( c \) is usually set empirically [20]:

\[
m = \sqrt{(a + \beta)} + q,
\]

(15)

where \( a \) and \( \beta \) denote the number of output and input layer nodes and \( q \) is a constant taking values between [0,10].

The optimization process of the LSTM network parameters \( m \) and \( c \) by using WOA is shown in Figure 3.

3. Market Risk Early-Warning Model Based on Improved LSTM

Based on the above improved LSTM model, the market risk early-warning model and its prediction process are designed in this study as Figure 4. The specific operation is as follows.

Step 1: (data collection): collect relevant factor variables affecting the early warning of enterprise market risk, which is preprocessed by one-hot encoding and normalization

Step 2: divide the data into training and test sets according to a certain ratio

Step 3: create and train an LSTM model; then, store the LSTM model with the best prediction accuracy

Step 4: use the WOA algorithm to optimize the number of hidden layer’s neurons \( m \) and time step \( c \)

Step 5: construct the WOA-LSTM model for prediction and output the results

4. Simulation Experiments

4.1. Experimental Environment Construction. The proposed model and comparison model are constructed in MATLAB 2019 for this experiment, running on a 64 bit Windows 7 Professional system with Intel(R) Xeon(R) E5-2620V3
2.40GHz CPU, Tesla K80 GPU, and 8G memory, and the data are preprocessed by using SPSS software.

4.2. Data Sources and Preprocessing

4.2.1. Data Sources. In this experiment, 40 real estate companies listed on the Shanghai Stock Exchange from 2012 to 2020 are used as research subjects, among which 5 companies are in financial crisis and the remaining 35 companies are financially healthy. For the crisis sample, if the sample crisis time period is $T$, the study period for this experiment is $T-1$ years [21]. The listed companies selected for this experiment include five real estate companies such as Songjiang Group and Yin Yi Group. For the normal sample, 35 real estate companies, such as China Fortune Land Development and NACITY PROPERTY SERVICE GROUP, were selected while ensuring the same study period. Combining the current situation of real estate enterprises in China and related literature [22, 23], the relevant variables selected for this experiment are indicated in Table 1.

4.2.2. Data Preprocessing. Among the above variables, different variables have different degrees of influence on the prediction of corporate financial market risk, while variables that have less influence on the prediction results add data dimension and reduce the running speed of the model [24, 25]. Therefore, to solve this problem, this experiment uses factor analysis to analyze the variables and achieve a reduction in data dimension and increase the running speed of the model by removing factors with low commonality. The findings of factor analysis on the following 23 variables are reported in Table 2. The factor commonality of interest coverage multiple, operating income growth rate, and net profit growth rate is less than 0.5, indicating that the information on the impact of corporate financial market risk cannot be basically extracted from these variables. Therefore, the three variables were removed from this experiment, and 20 variables were finally obtained.

In order to expand the features, the variables were treated in this experiment by one-hot encoding. First, expand the data discrete features to the Euclidean space, and then, encode them using one-hot to obtain continuous features. Considering that different variables have different data magnitudes, all data...
magnitudes are normalized in this study to facilitate the analysis. Finally, the data from 2012 to 2019 were divided into a training set and a test set in the ratio of 4:1, and the four quarterly values for 2020 were predicted.

4.3. Evaluation Indexes. The evaluation indexes for this experiment are mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination ($R^2$), which are calculated as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t|,$$

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2,$$

where $y_t$ is the actual value, $\hat{y}_t$ is the predicted value, and $N$ is the number of data points.
\[ RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}, \tag{18} \]

\[ R^2 = 1 - \frac{\sum (y - \bar{y})^2}{\sum (y - \bar{y})^2}, \tag{19} \]

In (19), \( R^2 \) takes a range of \((0,1)\), and the larger the value, the better the model performance.

4.4. Experimental Results

4.4.1. Model Validation

(1) Operating Margin Forecast Results. Taking the operating margin of Centralcon Holding as an example, the training set is used to train the improved LSTM model, and the training results are compared to the test set, as shown in Figure 5. The anticipated values are consistent with the change trend of the actual values, and the overall fitting effect is good, indicating that the proposed algorithm has good prediction effect.

The experience enters the indices except the operating profit margin into the prediction model and constructs the prediction model with the operating profit margin as the output for quantitatively analyzing the prediction performance of the proposed model. The model's prediction performance is shown in Table 3. From the table, the suggested model's prediction accuracy is 98%, showing that it has a high prediction accuracy and can better forecast the impact of each index on the operating margin.

(2) Predicted Results for Each Variable. Using quarterly data from 2012 to 2019 as model inputs and various indexes (operating margin as an example) as model outputs, Table 4 represents the predictive performance of the model, and its fit and iteration plots are shown in Figure 6. The revised LSTM model has an excellent prediction effect, attaining a prediction accuracy of 96%, and the overall fitting effect between the predicted and observed values is good, as shown by the prediction results.

4.4.2. Comparison of Models. The studies evaluated the prediction effect of the proposed model with CNN and LSTM models for each index for each quarter from 2012 to 2019 to ensure that the proposed model is effective, with the finding displayed in Figure 7. From Figure 7, the suggested algorithm outperforms the comparison algorithm in all indexes, and the prediction accuracy is likewise greater, which is improved compared with the CNN model and LSTM model, respectively. This indicates that the proposed model can enhance its prediction accuracy by improving the LSTM algorithm, thus improving the prediction performance to some extent.
For further verification of the validity of the proposed model, the experiments compare its prediction effects with the CNN model and the LSTM model for four quarters in 2020, and the results are shown in Figure 8. As shown in the table, the proposed model outperforms the comparison model in all performance metrics, and the proposed model improves the prediction accuracy by 22% compared to the CNN model and 7% compared to the LSTM model. The reason is that the upgraded LSTM model optimizes the LSTM network parameters using the WOA algorithm, which improves the model’s global optimization capabilities and prediction performance. This shows that the model described in this study has certain effectiveness and superiority to identify and warn market risks in advance and take corresponding measures according to the risks to ensure the healthy operation of the enterprise.

5. Conclusion

In summary, the predictive accuracy of the improved LSTM-based market risk warning model can be improved by optimizing the number of hidden layer’s neurons and the time step of the LSTM using the WOA algorithm, and the prediction accuracy can reach more than 96%, resulting in high-precision early warning of market risks. The proposed model improves the prediction accuracy of corporate finance from 2012 to 2019 to different degrees, by 14% and 12%, respectively. In addition, it improves the prediction accuracy of corporate finance in 2020 by 22% and 7%, respectively. And compared with the traditional CNN and LSTM models, it has certain superiority. The innovation of this study is to use WOA algorithm to improve LSTM and change the previous LSTM parameter optimization method, so as to better improve the prediction accuracy of the algorithm, which is also an innovation of this research [26].

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References


