Prediction Model of the Development Trend of Chinese Development Finance Based on CEEMD–LSSVM

Jiechen Wang, Huixia Liu
Email: 459274083@qq.com, Email: lhuix369@nwu.edu.cn

School of Economics and Management Northwest University Xi'an, China, School of Economics and Management Northwest University Xi'an, China

Abstract—Development finance can achieve the development goals of the government in the market through the combination of financing and government organization advantages. The prediction of development finance development trend is of great significance for the national and local governments to make relevant decisions. In this work, the core data of development finance, the balance of medium and long-term loans, are selected as the research object, which are nonlinear and non-stationary and have a small sample size, making them difficult to accurately predict. A combination model based on complete ensemble empirical mode decomposition (CEEMD)–least squares support vector machine (LSSVM) is proposed to analyze the development trend of development finance in China for improving the prediction accuracy and effectively explaining development finance loans fluctuations. First, the loan balance could be decomposed into several intrinsic mode component, which solves the nonlinear and non-stationary problems, based on CEEMD. On this basis, LSSVM was used to separately forecast each component. Finally, the development trend of developmental finance was obtained by reorganizing the forecast results of each component. The prediction results of the CEEMD–LSSVM model are compared with support vector machine, BP, and so on, which show that the proposed model has higher accuracy, to verify the effectiveness of the model.

Keywords-development finance, trend prediction, complete ensemble empirical mode decomposition, least squares support vector machine.

1 Introduction

Development finance, an important aspect of the financial system, promotes market and system construction and economic growth by supporting weak links and fields of economic development, such as infrastructure. Development finance has played an important role in supporting national key construction and promoting urbanization and regional coordinated development. In contrast with the optimization of social fund use efficiency of general commercial banks to promote economic growth, development finance realizes the government's development goals in the market through the combination of financing and government organization coordination advantages, especially in underdeveloped areas (JOHN Z, 1983) [1]. The change law of its development trend will have a significant impact not only on

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2326749

regional economic development but also on the relevant decisions of local governments. Therefore, the accurate prediction of the development direction and degree of development finance is helpful in making effective decisions on the changes of development finance.

More in-depth studies in the field of development finance have been conducted across the world. Johnson et al. (1982) [2] found that development finance is a strategic tool for national economic development based on a large number of empirical studies. This tool promotes the optimal allocation of financial funds and plays a significant role in promoting economic development. Yoshino (1984) [3], a Japanese scholar, focused on the Japanese economic development model since the 20th century and analyzed the main functions of development finance. The author concluded that the main goal of development finance is to support industries with weak competitiveness, thereby realizing the optimization and adjustment of the industrial structure. Matic (2009) [4] found that the funds of the Development Bank meet the requirements of suitability, availability, and sufficiency to a great extent through a large number of statistics. This form of financial resources plays a more significant role in regional economic development. Domestic research on development finance started late. Chen (2004, 2010) [5-6] systematically analyzed the role of development finance in urban infrastructure construction, alleviating the capital bottleneck constraints of urbanization and promoting the development of urbanization and the impact of the unique development model of development finance on China's sustainable development. Li Huibin (2009) [7]and Li Xiguang (2015) [8] focused on the role of development finance in promoting China's economic growth and analyzed the mechanism of development finance affecting economic growth. They also quantitatively analyzed the relationship between development finance and economic growth through statistical model.

Scholars have widely acknowledged the role of development finance in promoting economic growth throughout the research results across the world and further subdivided and deepened the theoretical and empirical research in this field. However, scholars rarely include the prediction of the development trend of development finance in China. On the one hand, this situation may be because of the many problems to be studied in the reform practice of China's development finance. On the other hand, this situation may also be because of the difficulty in obtaining relevant financial data and conducting quantitative analysis (Liu Xizhang,2015) [9]. The realization of the government's development goals in the future can no longer meet the requirements by relying solely on the market and commercial finance, especially for developing countries (Chen, 2009) [10]. China Development Bank (CDB), which is the only development financial institution in China, has always supported economic development and strategic adjustment of economic structure, mainly serving major medium and long-term development strategies of the national economy by carrying out medium and long-term credit, investment, and other financial services. According to the above analysis, this work takes the loan balance of CDB as the research object and proposes the CEEMD–LSSVM combination prediction model based on advanced artificial intelligence algorithm. The development trend of development finance is evaluated and predicted by using the nonlinear signal analysis advantages of the CEEMD algorithm and the good learning and prediction ability of the LSSVM model. Finally, the quantitative analysis of its change trend in the future will provide a certain data basis for the optimization of China's economic development policy.

2 Related technology

2.1 Complementary Ensemble Empirical Mode Decomposition (CEEMD)

EMD, a classical signal decomposition method, often has the problem of mode aliasing. Wu solves this problem by adding auxiliary noise and proposes an improved EMD method, namely, ensemble EMD (EEMD). Based on the advantage of uniform distribution of white noise, he manually adds white noise to the original signal to achieve continuous signal distribution. The EEMD decomposition requires several additions of auxiliary white noise, and the influence of artificial noise is offset by means of the mean method; accordingly, each component has antinoise characteristics. Yeh (2010) [11] added positive and negative paired auxiliary noises to eliminate the residual auxiliary noise in the reconstructed signal on the basis of the EEMD method. The computational efficiency is improved due to the low order of auxiliary noise added by Yeh. This decomposition method is called complementary EEMD (CEEMD). CEEMD not only effectively solves the mode aliasing problem of EMD but also retains the advantages of EMD in processing nonstationary signals, such as self-adaptability and binary filtering characteristics. CEEMD decomposition includes the following three steps:

1) Add n groups of positive and negative paired auxiliary white noises. Two sets of sets are generated based on the original signal:

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} S \\ N \end{bmatrix} \tag{1}$$

where M_1 and M_2 are the signals after adding auxiliary noise, S is the original signal, and N is the auxiliary noise.

2) CEEMD decomposes each signal in the set, and each signal obtains a set of intrinsic mode function (IMF) components, in which the jth IMF component of the ith signal is expressed as I_{ij}. The CEEMD decomposition process includes the following three steps:

a) The local maximum and local minimum of the original signal x(t) are located, and the local maximum and minimum sequences with third-order spline function are interpolated to obtain the upper envelope sequence $x_{max}(t)$ and lower envelope sequence $x_{min}(t)$ of the original signal x(t).

b) The sum of $x_{max}(t)$ and $x_{min}(t)$ at each time is averaged to obtain the instantaneous average value m(t):

$$m(t) = (x_{\max}(t) + x_{\min}(t))/2 \quad . \tag{2}$$

c) The class anomaly sequence I(t) can be obtained by subtracting the instantaneous average value m(t) from the original signal x(t):

$$I(t) = x(t) - m(t) \quad . \tag{3}$$

We must judge whether all types of anomaly value sequences I(t) are internal modular functions. If the cumulative number of maximum and minimum points in I(t) is equal to or at most one difference from the number of intersections of abscissa axis, and the instantaneous average value m(t) is constant and equal to zero, then this type of anomaly sequence I(t) is an intrinsic modular function. Otherwise, we treat I(t) as the original signal and repeat the above-mentioned steps until the anomaly sequence meets the definition of intrinsic modulus function.

According to the above-mentioned method, the first intrinsic modulus function $I_1(t)$ can be obtained, and the first component can be separated from the original signal. Specifically, the residual value sequence $r_1(t)$ can be obtained by subtracting $I_1(t)$ from the original signal:

$$r_1(t) = x(t) - I_1(t) \tag{4}$$

Next, we take the residual value sequence $r_1(t)$ as a new original signal, repeat the above-mentioned steps, and extract the n intrinsic modular functions $I_n(t)$ in turn. When the sequence of residual values $r_n(t)$ is a monotone sequence, the CEEMD decomposition process ends. The decomposed intrinsic modulus function components and residual value sequences are reorganized to obtain the following original signals:

$$x(t) = \sum_{i=1}^{n} I_i(t) + r_n(t) \tag{5}$$

3) The following decomposition results can be obtained by combining multiple components:

$$I_j = \frac{1}{2n}\sum_{i=1}^{2n} I_{ij} \tag{6}$$

where I_j represents the jth IMF component corresponding to the CEEMD decomposition.

2.2 Least squares support vector machine (LSSVM)

SVM is an intelligent learning algorithm with high generalization ability. This mechanism can solve nonlinear problems. Furthermore, this model is often used in the prediction, classification, recognition, and so on. LSSVM is an improved SVM method proposed by Suykens (2002) [12] on the basis of SVM. LSSVM uses the square sum error loss function as the objective function instead of the insensitive loss function in SVM. The quadratic programming problem of SVM is transformed into a linear relationship problem, which simplifies the objective function and improves the learning speed, using the solvability advantage of equality constraints.

Given a training set $\{x_i, y_i\}$, i =1, 2……, m with m data, LSSVM is based on the sequential minimum optimization principle (SMO). The optimization decision function in this high-dimensional space is by mapping the sample space to the high-order space with a nonlinear function φ(x):

$$y(x) = w^T \phi(x) + b \tag{7}$$

where w is the weight vector, and b is the deviation.

Its essence is to solve the following quadratic programming problem:

$$\begin{cases} \min J(w,\xi) = \|w\|^2 + \frac{1}{2}\gamma\sum_{i=1}^{n}\xi_i^2 \\ s.t. y_i = w^T\phi(x) + b + \xi_i \end{cases}, \tag{8}$$

where J is the optimization function of LSSVM, γ is the penalty coefficient, and ξ_i is a relaxation variable.

The following can be obtained by substituting Lagrange function L into Equation (8):

$$L(w,b,\xi,\alpha) = \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma\xi_i^2 - \sum_{i=1}^{n}\alpha_i\left[w^T\phi(x) + b + \xi_i - y_i\right], \tag{9}$$

where α_i (i=1, 2, ……, n) is the Lagrange multiplier.

According to Mercer's condition, parameters w, b, ξ and α_i in Equation (9) can be derived, and variables w and ξ can be eliminated. a_i and b can be calculated, and the following LSSVM regression prediction model can be obtained:

$$y(x^*) = \sum_{i=1}^{n}\alpha_i K(x,x^*) + b, \tag{10}$$

Where x* is a new set of prediction points, K(•) is a kernel function, and $K(x_i, x_j)=\varphi(xi)^T\varphi(x_j)$.

In this work, the commonly used radial basis kernel function is selected, and its expression is as follows:

$$K(x,x_i) = e^{-\|x-x_i\|^2/2\sigma^2}, \tag{11}$$

where σ is the width of the radial basis kernel function.

2.3 Prediction model establishment method based on CEEMD–LSSVM

The time series of the loan balance of CDB is nonlinear and non-stationary. The original sequence contains not only high-frequency variable components but also slow changing regular components and linear components. If the original signal is directly trained by LSSVM regression, then the prediction results of the model are prone to large errors. LSSVM regression training is needed for components with different properties in the original signal to establish a high-precision prediction model. Therefore, this work proposes a prediction method based on the combined CEEMD and LSSVM model. The CEEMD–LSSVM prediction model algorithm is divided into three parts: CDB loan balance decomposition, component prediction, and CDB loan balance reorganization. The prediction flow chart of the CEEMD–LSSVM model is shown in Figure 1, and the specific steps are as follows:

1) The CEEMD method is used to decompose the loan balance time series of CDB to obtain the IMF component and trend term r of each eigenmode function. Non-stationary original time series can be decomposed into multiple stationary eigenmode function sequences through CEEMD decomposition. Each stationary component can be predicted and analyzed through LSSVM model to improve the accuracy of development finance development trend prediction.

2) According to the frequency, the IMF component is divided into high-frequency, low-frequency, and residual terms to form three component modules. The input and output samples are substituted into the LSSVM model based on the LSSVM model. The model prediction values of each IMF component are obtained by using the good nonlinear learning and prediction ability of the LSSVM model.

3) The predicted values of each component are superimposed to obtain the final predicted value of CDB's loan balance compared with the actual data and the predicted values of other models, calculate the error, and analyze the accuracy of the model.

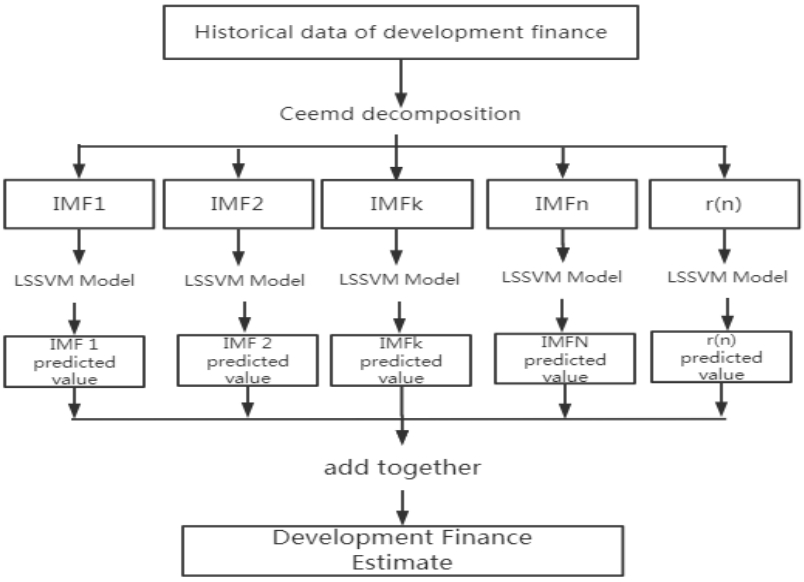

Figure 1 Algorithm flow chart of the CEEMD–LSSVM prediction model

3 Empirical analysis

3.1 Data and evaluation criteria

Since 1998, the National Development Bank has implemented the national policy of expanding domestic demand and stimulating investment, and its support has expanded from major national key projects to urban infrastructure. This work uses the annual loan balance of CDB as the measurement index of development finance to effectively analyze and predict the future trend of development finance, based on the panel data from 1998 to 2018. The data come from the financial statements published by CDB. Figure 2 shows the loan balance curve of CDB. The proportion of development finance loan balance in the loan balance of national financial institutions is selected as the research object to fully reflect the role and change trend of development finance in China's economic development, with reference to the research method of Lin Yifu (2008) [13]. Figure 3 shows the proportion curve of CDB's loan balance in the loan balance of national financial institutions (the data are from EPS database). During the prediction, 21 data from 1998 to 2018 are used as training samples, and the loan data and its proportion of CDB from 2019 to 2021 are selected as the prediction object. This work uses MATLAB R2018 to process data.

The mean square error (MSE) and multi correlation coefficient (MCC) are used as evaluation criteria to evaluate the prediction performance of the model. MSE reflects the statistical characteristics of error, that is, the dispersion of samples. The smaller the MSE value, the higher the prediction accuracy. MCC is an indicator to measure the correlation between the predicted and the original values. The closer the complex correlation coefficient is to one, the closer the correlation between the predicted and the original values is.

We set y and y* as the true and predicted values, respectively. The expressions of MSE and MCC are as follows:

$$MSE = \sum_{i=1}^{n} (y_i - y_i^*)^2 / n \text{,} \tag{12}$$

$$MCC = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(y_i^* - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 (y_i^* - \bar{y})^2}} \text{.} \tag{13}$$

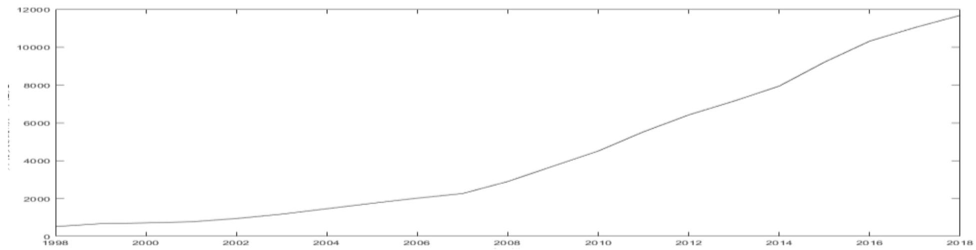

Figure 2 Trend of the loan balance of CDB from 1998 to 2018

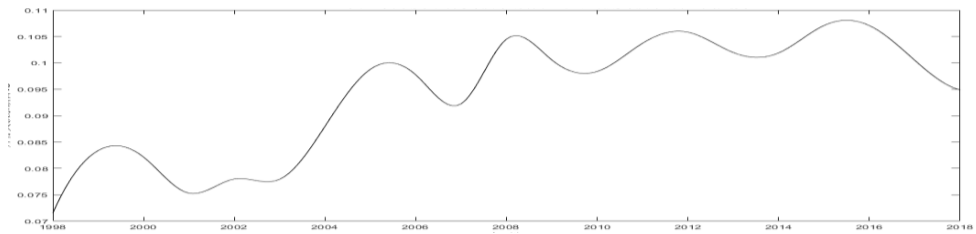

Figure 3 Trend of the loan balance of CDB in proportion to the national financial loan balance from 1998 to 2018

3.2 Analysis of model results

1) CEEMD decomposition: CEEMD can decompose the complex time series of the loan balance proportion of CDB into modular function component IMF and trend term component r containing different scale information. Figure 4 shows the CEEMD decomposition results, including two IMF components and one trend component. In addition, Figure 4 depicts that the frequency of imf1 and IMF2 in the stationary sub-sequence of the loan balance proportion series of CDB after CEEMD decomposition and to the trend term r shows a gradual decreasing trend. Modulus function implies strong economic significance and can be used to explain the internal characteristics hidden in the time series of loan balance of CDB. The trend item is the trend of the original data, that is, it can be used to describe the long-term trend of the proportion of loan balance of CDB.

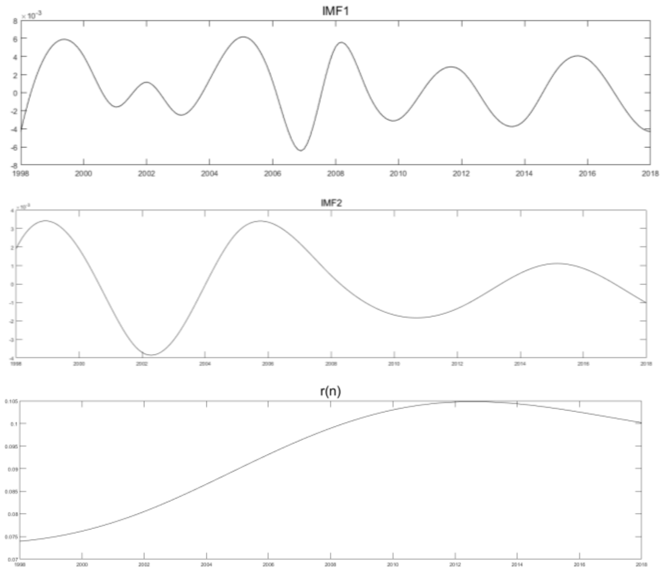

Figure 4 CEEMD breakdown results of the loan balance proportion of CDB

2) Composition Component prediction: The data of IMF components and trend items at different times are regarded as time series {x(t),t=1,2,L,……,n} , and its prediction model can be described as follows:

$$x(t)=\varphi\left[x(t-1),x(t-2),\cdots x(t-p)\right],$$

Where Φ represents the nonlinear function, and p denotes the prediction window width (the prediction window width in this work is three, that is, the data at the first three times are used to predict the data at the fourth time). The input and output samples during the prediction of each IMF component are shown in Table 1.

Table 1 List of input and output vectors

Number of samples	**Input vector**	**Output vector**
First sample	$x(1),\cdots x(p-1),x(p)$	$x(p+1)$
Second sample	$x(2),\cdots x(p),x(p+1)$	$x(p+2)$
……	……	……
Nth sample	$x(n-p),\cdots x(n-2),x(n-1)$	$x(n)$

The original component data are normalized to improve the calculation speed and prediction accuracy. The formula is as follows: $x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$,

Where χ_{max} and χ_{min} represent the maximum and minimum values in the component, $\chi_i \in [0,1]$, respectively.

The input and output vectors of each component from 1998 to 2018 are substituted into the LSSVM model as a training set for learning. Table 2 is the statistical table of the training effect parameters of each component. Figure 5 shows the distribution results of the actual and regression values of each component normalized sequence. The average complex correlation coefficient of each component regression result is 0.9994, and the average mean square error is 0.0001, indicating that the fitting accuracy between the calculated and the actual values of the model is high. The normalized time series of each component from 2019 to 2021 are predicted, and the normalized prediction value is transformed into the original prediction value of each component according to the normalization formula. The results are shown in Table 3.

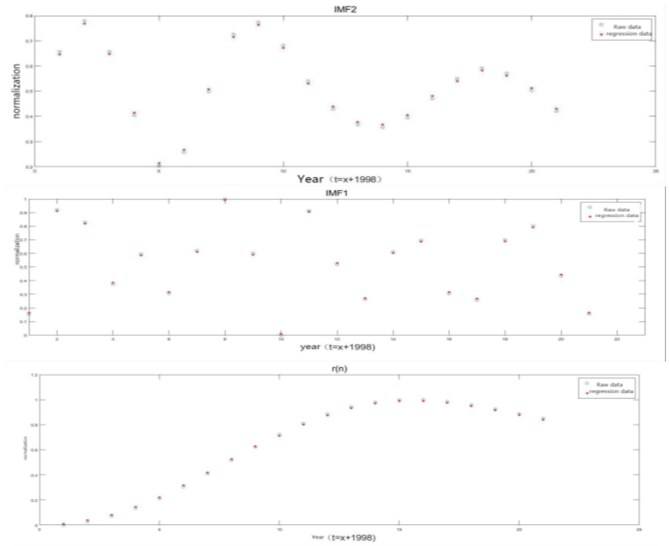

Figure 5 Statistical chart of the LSSVM regression results of each component

Table 2 Statistics of the LSSVM training effect parameters of each component

IMF	RSE	MCC
IMF1	0.0001	0.9996
IMF2	0.0002	0.9987
r	0.0001	0.9999

Table 3 Statistics of the predicted values of each component from 2019 to 2021

IMF	2019	2020	2021
IMF1 normalized predicted value	0.4212	0.4103	0.3994
IMF1 actual predicted value	−0.001	−0.0011	−0.0013
IMF2 normalized predicted value	0.4193	0.4085	0.3977
IMF2 actual predicted value	−0.0007	−0.0008	−0.0009
R normalized predicted value	0.6404	0.6284	0.6205
R actual predicted value	0.0937	0.0933	0.0931

3) Development trend forecast of development finance

The regression and prediction results of each component are superimposed with equal weight to obtain the final regression and prediction value of the loan balance of CDB. The SVM and BP neural network models are used to train the time series of loan balance of CDB to verify the effectiveness of the CEEMD–LSSVM method. The prediction results of the CEEMD–LSSVM model and the other two models are shown in Table 4, and the MSE and MCC values of the three prediction methods are provided. The error indexes of the three models are shown in Table 5.

The comparison result of the prediction results and error indicators shows that the predicted value of the CEEMD–LSSVM model is closer to the actual value of the proportion of loan balance of CDB in most years. The MSE value of the CEEMD–LSSVM model is the smallest, and the MCC value is the largest, indicating that the prediction accuracy of this model is higher than that of the other two common artificial intelligence models.

Table 4 Comparison of the predicted values of the CEEMD–LSSVM, SVM, and BP models

Time	Actual Value	CEEMD–LSSVM	SVM	BP	Time	Actual Value	CEEMD–LSSVM	SVM	BP
1998	0.072	0.072	0.082	0.072	2010	0.098	0.098	0.091	0.099
1999	0.084	0.083	0.091	0.083	2011	0.103	0.104	0.094	0.104
2000	0.082	0.082	0.091	0.082	2012	0.105	0.106	0.096	0.105
2001	0.076	0.075	0.085	0.080	2013	0.102	0.102	0.092	0.105
2002	0.078	0.078	0.088	0.078	2014	0.102	0.102	0.092	0.105
2003	0.078	0.078	0.088	0.079	2015	0.107	0.107	0.097	0.105
2004	0.088	0.088	0.091	0.088	2016	0.107	0.107	0.097	0.104
2005	0.099	0.099	0.091	0.090	2017	0.100	0.100	0.092	0.100
2006	0.098	0.098	0.091	0.090	2018	0.095	0.095	0.091	0.095
2007	0.092	0.092	0.091	0.092	2019		0.092	0.089	0.093
2008	0.104	0.105	0.094	0.103	2020		0.091	0.087	0.091
2009	0.100	0.101	0.091	0.100	2021		0.091	0.086	0.090

Table 5 Error indicators of the CEEMD–LSSVM, SVM, and BP models

Method	RSE	MCC
CEEMD–LSSVM	0.0001	0.9999
SVM	0.0012	0.7765
BP	0.0005	0.9921

The regression value of the total loan balance of CDB from 1998 to 2018 and the prediction value from 2019 to 2021 can be obtained by using the CEEMD–LSSVM prediction method (Figure 6). The figure shows that the prediction accuracy is also high. Therefore, the CEEMD–LSSVM model proposed in this work has better prediction ability and is more suitable for the prediction of the development trend of development finance.

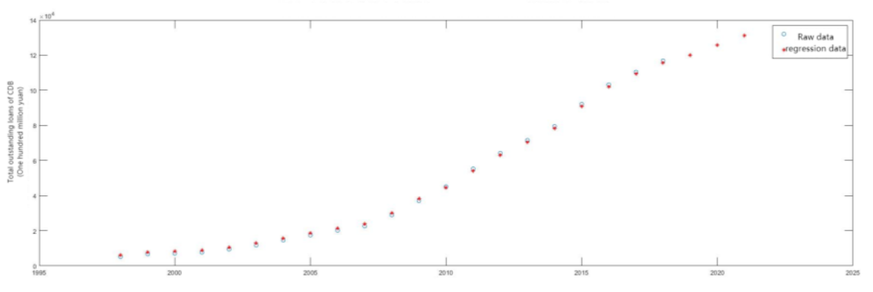

Figure 6 Forecast results of the total loan balance of CDB

4 Conclusion

This work introduces CEEMD signal processing method into development finance prediction and decomposes the original nonstationary development finance time series into stable modular function components. The accuracy of development financial prediction model is improved by integrating the CEEMD and LSSVM model. The comparison result of the actual and the predicted values of SVM and BP models showed that the prediction model has high accuracy. These results have certain guiding significance for local governments to adjust development goals and policies closely related to development finance.

According to the prediction results of the CEEMD–LSSVM model, the balance of development financial loans will continue to increase in the short term. Nonetheless, the growth rate will decrease, and the proportion of development financial loans in the balance of national financial loans will slightly decrease and gradually stabilize. This phenomenon is mainly because the main task of CDB is to support large and medium-sized construction projects in infrastructure, basic industries, and pillar industries. At present, China's economy has changed from high-speed growth stage to high-quality development stage. CDB has also actively responded to the national economic structure adjustment and changed the regional business structure. For a long time, the eastern region of China has occupied most of the development financial resources, while the central and western regions have occupied less. After 2018, the growth rate of GDP and per capita GDP in some eastern regions exceeded that in the western region. Therefore, the regional structure of development finance should be optimized, the regional economic development difference among eastern, central, and Western China should be reduced, and a balanced economic development should be achieved.

References

[1] JOHN Z.Government, Market and Growth: Financial System and the Politics of Industrial Change [M]. Ithaca: Cornell University Press, 1983:35-36.

[2] Johnson C. MITI and the Japanese miracle: the growth of industrial policy, 1925-1975[J]. Journal of Comparative Economics, 1982, 6(4):436-439.

[3] Yoshino. Tax system and financial investment and financing [M]. Tokyo: University of Tokyo Press, 1984.198-202.

[4] Matic B. Financing Regional Development Through Development Banks[J].Interdisciplinary Management Research ,2009,5:749-759.

[5] Chen Yuan. Development finance and China's urbanization development [J]. Economic research, 2010,45 (07): 4-14.

[6] Chen Yuan. Giving play to the role of development finance to promote China's sustainable economic and social development [J]. Management world, 2004 (07): 1-5.

[7] Li Huibin, Gao Jinlong, Cao Guohua. Analysis and empirical test of the impact of Development Finance on economic growth [J]. Exploration of economic issues, 2009 (09): 105-109.

[8] Li Xiguang. Empirical Study on the impact of Development Finance on China's economy [J]. Investment research, 2015,34 (11): 148-160.

[9] Liu Xizhang, Li fufu. Empirical Study on the impact of Development Finance on economic growth [J]. Journal of Dalian University of Technology (SOCIAL SCIENCE EDITION), 2015,36 (04): 73-77.

[10] Chen Yuan. Development finance and China's economic and social development [J]. Economic science, 2009 (04): 5-14Wu Z, Huang N E. ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD[J]. Advances in Adaptive Data Analysis, 2009, 01(01):1-41.

[11] Yeh J R, Shieh J S, Huang N E. COMPLEMENTARY ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOVEL NOISE ENHANCED DATA ANALYSIS METHOD[J]. Advances in Adaptive Data Analysis, 2010, 2(02):1-15.

[12] Suykens J A K, Gestel T V, Brabanter J D, et al. Least Squares Support Vector Machines[J]. International Journal of Circuit Theory & Applications, 2002, 27(6):605-615.

[13] Lin Yifu, sun Xifang. Banking structure and economic growth [J]. Economic research, 2008, 43 (09): 31-45.