Forecast of Fiscal Revenue Based on Fruit Fly Optimization Algorithm Optimized Support Vector Machine

Xinyu Liu[1, a], Junyan Liu[2], Yi Zhang[3], Yuxin Sun[3], Zhiyang Wang[3], Yuqi Wu[4], Haodong Hong[4], Depei Zhang[3]

[1]Tourism College of Hainan University, Haikou, Hainan, 570228

[2]College of Forestry, Hainan University, Haikou, Hainan, 570228

[3]School of Economics, Hainan University, Haikou, Hainan, 570228

[4]Management School of Hainan University, Haikou, Hainan, 570228

[a]1393239448@qq.com

Keywords: Fruit fly optimization algorithm; Support vector machine; fiscal revenue; forecasting model;

Abstract: Fiscal revenue is not only an important source of national revenue, but also the basis for reflecting the country's economic conditions. The trend of fiscal revenue determines the development trend of the local economy and market prospects, and is particularly important for the development of my country's financial industry. For this reason, this article applies support vector machines to the forecast of fiscal revenue. In order to solve the shortcomings of time-consuming, labor-intensive and inefficient selection of model parameters of support vector machines, this paper proposes a method to optimize the selection of the penalty parameters and kernel parameters of the support vector machine using the Fruit fly optimization algorithm of global optimization, and establishes a prediction model based on the Fruit fly optimization algorithm to optimize the support vector machine. The simulation results show that the model optimized by Fruit fly optimization algorithm has higher prediction accuracy and meets the demand for prediction accuracy.

1. Introduction

With the vigorous development of science and technology in recent years, big data has been widely used in various industries, and data mining technology has gradually improved. Fiscal revenue is very important to the country and region, and it is an important foundation to ensure the normal operation of the country. Therefore, it is worth studying how to apply data mining technology in fiscal revenue forecasting.

Literature [1] analyzes and forecasts my country's fiscal revenue by using a time series model. The forecast results show that as the forecast period extends, the forecast error will gradually increase, but compared with other forecasting methods, in the short term The accuracy of its prediction is relatively high. Literature [2] combines the Lasso regression model with the GRNN neural network model. Taking Haixi Prefecture's 1994-2016 local fiscal revenue as an example, the Lasso variable selection method is used to determine the main indicators that affect local fiscal revenue. Then the selected index values are used as the input of the GRNN neural network, and the output of the network is the corresponding local fiscal revenue. A prediction model based on the Lasso-GRNN neural network is established to predict the local fiscal revenue. Literature [3] selected 8 variables and used a multiple regression model to predict the fiscal revenue of Hebei Province, and obtained the size of the contribution rate of these factors to the fiscal revenue, and gave reasonable suggestions.

Based on this, this paper establishes a support vector machine prediction model based on Fruit fly optimization algorithm optimization. Taking the fiscal revenue of Jinan from 1999 to 2017 as an example, the fiscal revenue of Jinan is predicted and analyzed.

2. Fruit fly optimization algorithm optimization support vector machine

2.1 Basic Principles of Support Vector Machine

Support vector machines show greater advantages in applications with less sample data and more influencing factors, and are widely used in various predictions, regressions, and classifications. The basic principle of the support vector machine is to map the trained data set to the high-dimensional feature space through nonlinear mapping $\phi(x)$, so that the nonlinear function estimation problem in the low-dimensional feature space is transformed into the linear function estimation problem in the high-dimensional feature space. The calculation process is as follows.

1) Suppose the linear function is shown in Equation 1.

$$f(x) = \omega^T \phi(x) + b \quad \left(\omega \in R^{nh}, b \in R\right) \tag{1}$$

In the formula: x is the input variable; ω is the weight vector; b is the offset; R^{nh} is the high-dimensional feature space.

2) Then quote the insensitive loss function on ε as shown in Equation 2.

$$\left|y - f(x)\right| \bullet \varepsilon = \max\left\{0, \left|y - f(x)\right| - \varepsilon\right\} \tag{2}$$

3) According to the principle of structural risk minimization, the function estimation problem is transformed into finding the $f(x)$ that minimizes the following risk function, as shown in Equation 3.

$$R_{reg} = \frac{1}{2}\omega^T \omega + \gamma R_{emp}^{\varepsilon}[f] \tag{3}$$

In the formula: $\omega^T \omega = \|\omega\|^2$ is the complexity of the model; $R_{emp}^{\varepsilon}[f]$ is the experience risk; γ is the set penalty parameter.

4) Solving the risk minimization problem of formula (3) is equivalent to solving the optimization problem of formula 4.

$$\begin{gathered}
J = \min_{\alpha,\alpha^*}\left[\frac{1}{2}\sum_{j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)K\left(x_i, x_j\right)\right] \\
-\varepsilon\sum_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{l} y_i\left(\alpha_i - \alpha_i^*\right) \\
\text{s.t.}\begin{cases}\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0 \\ \alpha_i, \alpha_i^* \in [0, y]\end{cases}
\end{gathered} \tag{4}$$

5) By using the quadratic programming method in the optimization theory to solve the equation (4), the parameters,, can be obtained α_i, α_i^* then the offset b can be obtained using the KKT condition, and the estimated equation of the support vector machine regression equation is shown in equation 5.

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)K\left(x_i, x\right) + b \tag{5}$$

The samples (x_i, y_i) of the corresponding coefficients $\left(\alpha_i - \alpha_i^*\right) \neq 0$ in Equation 5 are all support vectors, and K(x_i, x) is the kernel function that satisfies the Mercer condition, which is defined as $K(x_i, x) = \phi(x)^T \phi(x_i)$. In this paper, the RBF kernel function is selected, and the insensitive loss function ε is taken as 0.01. After substituting the kernel function, the formula 6 can be obtained.

$$K\left(x_i, x\right) = \exp\left(-\frac{\left\|x - x_i\right\|}{2\sigma^2}\right) \tag{6}$$

Where σ is the width of the kernel function parameter.

2.2 Basic Principles of Fruit Fly Optimization Algorithm

Fruit fly optimization algorithm is an evolutionary algorithm that seeks global optimization based on the foraging behavior of fruit flies to characterize complex phenomena and then get inspired. The main principle is that fruit flies perceive the taste concentration of food around them according to their sense of smell. By comparing the concentration, the fruit flies approach the position with the best concentration, and then find the position of the food based on keen vision and fly to that position. The specific principle of the algorithm is shown in the following seven steps.

1) Initialize the fruit fly population, initialize the position of the fruit fly population X_axis and Y_axis.

2) Give the individual fruit flies the random distance and direction to search for food by smell, so that the fruit flies can judge the direction and distance of food according to the sense of smell, as shown in Equations 7 and 8.

$$X_i = X\_axis + Random \quad Value \tag{7}$$

$$Y_i = Y\_axis + Random \quad Value \tag{8}$$

3) Since the food location is unknown, it is necessary to estimate the distance D between the location of each fruit fly and the origin, and then calculate the taste concentration determination value S_i. S_i is the reciprocal of D. The specific calculation formulas are shown in Equation 9 and Equation 10.

$$D = sqrt\left(X_i^2 + Y_i^2\right) \tag{9}$$

$$S_i = 1 / D \tag{10}$$

4) The taste concentration determination value S_i is substituted into the taste concentration determination function (Function) to obtain the taste concentration (Smell) of the individual position of the fly.

5) Find the fruit flies with the highest taste concentration in the fruit flies population, that is, solve the optimal value. This article is to find the penalty factor C and the kernel function σ in the SVM, as shown in equations 11 and 12.

$$smell = Function\left(S_i\right) \tag{11}$$

$$\left[bestsmell, bestIndex\right] = \left(smell\right) \tag{12}$$

6) Keep the best taste concentration value and the x and y coordinates. At this time, the fruit fly colony uses vision to fly to this position. As shown in equations 13-15.

$$smellbest = bestsmell \tag{13}$$

$$X\_axis = X\left(bestIndex\right) \tag{14}$$

$$Y\_axis = Y\left(bestIndex\right) \tag{15}$$

7) Enter iterative optimization, repeat steps 2 to 5, and judge whether the taste concentration is better than the taste concentration of the previous iteration, if yes, proceed to step 6, otherwise continue the loop iteration until the maximum number of iterations is reached to end the loop.

The basic process of Fruit fly optimization algorithm optimization support vector machine is shown in Figure 1.

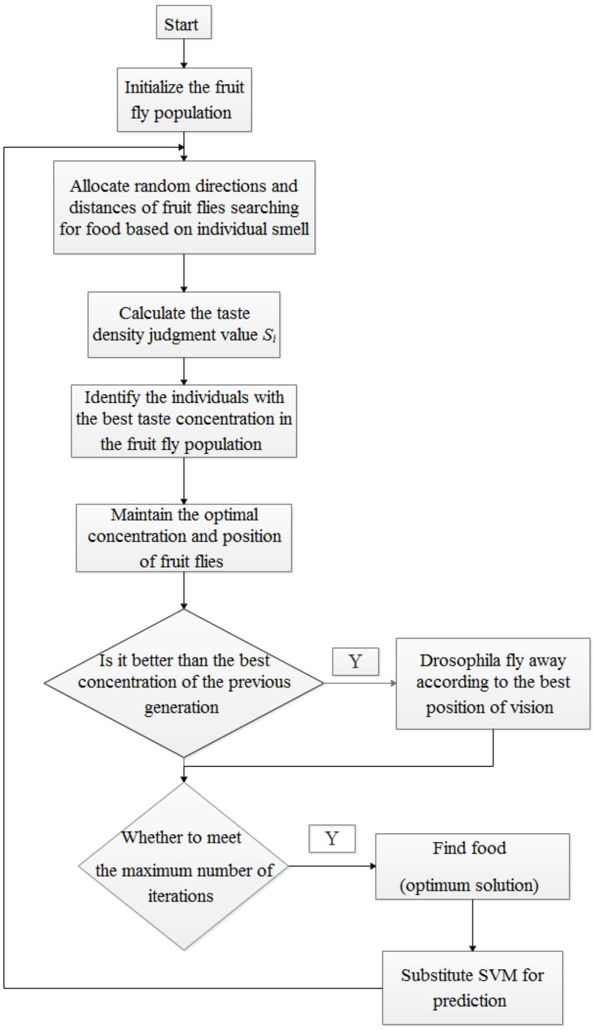

Fig 1 FOA-SVM flowchart

3. Fiscal revenue forecast

3.1 Selection of model parameters

Fiscal revenue is often determined by a combination of multiple factors. This article comprehensively considers the main influencing factors and indicators, and selects the 6 most important factors affecting fiscal revenue by consulting a large number of documents. They are the total retail sales of consumer goods X1, the per capita disposable income of urban residents X2, the per capita consumption expenditure of urban residents X3, the total investment in fixed assets of the whole society X4, the gross regional product X5, the output value of the tertiary industry X6, and the fiscal revenue Y. This paper uses these six influencing factors as the input of the forecasting model, and fiscal revenue as the output of the model. The six influencing factors and the symbols and units of fiscal revenue are shown in Table 1.

Tab. 1 Symbol description of Jinan's fiscal revenue and influencing factors

symbol	meaning	unit
X_1	The total retail sales of social consumer goods	Ten thousand yuan
X_2	Per capita disposable income of urban residents	yuan
X_3	Per capita consumption expenditure of urban residents	yuan
X_4	Total investment in fixed assets of the whole society	Ten thousand yuan
X_5	GDP	Ten thousand yuan
X_6	Tertiary industry output value	Ten thousand yuan
Y	Revenue	Billion

The simulation parameters of the prediction model in this paper are set as follows: For the Fruit fly optimization algorithm, the maximum number of iterations is set to 100, the number of population individuals is 20, the random initial position of the fruit fly is [0,1], and the random position of looking for food is [-10,10]. For support vector machines, the penalty factor is in [0, 100], and the value range of the RBF kernel function is [0, 100].

3.2 Model prediction results

This article quotes the financial revenue data from 1999 to 2017 corresponding to these six influencing factors in the Jinan Statistical Yearbook. A total of 14 sets of data from 1999 to 2012 are selected to train the prediction model of this article, and a total of 5 sets of data from 2013 to 2017 are used. Using the trained Fruit fly optimization algorithm to optimize the support vector machine prediction model, the financial revenue of Jinan was predicted. The Fruit fly optimization algorithm optimization support vector machine and the support vector machine's prediction value for Jinan's fiscal revenue are shown in Table 2, and the specific prediction curves are shown in Figure 2.

Tab. 2 Forecast results of Jinan's fiscal revenue

years	Actual value/ billion	FOA-SVM	Error/ %	SVM	Error/ %
2013	48.207	47.391	1.8	46.513	3.5
2014	54.312	54.639	0.6	55.738	2.6
2015	61.431	62.909	2.4	64.256	4.6
2016	64.121	62.279	2.9	60.154	6.2
2017	67.721	67.983	0.4	70.562	4.2

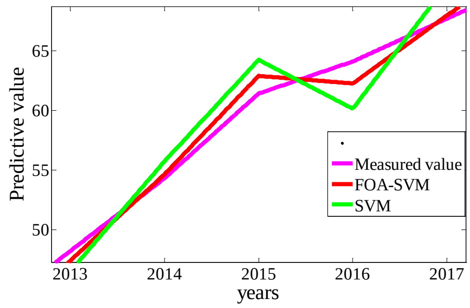

Fig 2 The forecast curve of Jinan's fiscal revenue

According to the prediction results in Table 1, it can be concluded that the Fruit fly optimization algorithm optimization support vector machine prediction model has a maximum prediction error of 2.9%, a minimum error of 0.4%, and an average error of 1.62% for Jinan's fiscal revenue; The highest forecast error of support vector machine for Jinan's fiscal revenue is 6.2%, the lowest error is 2.6%, and the average error is 4.22%. The prediction accuracy of the support vector machine model optimized by the Fruit fly optimization algorithm is significantly higher than that of the support vector machine prediction model.

4. Conclusion

This paper uses Fruit fly optimization algorithm to optimize the support vector machine to predict the fiscal revenue of Jinan. The main conclusions are shown in the following two points.

1) In this paper, the Fruit fly optimization algorithm optimized support vector machine algorithm is used to establish a forecasting model based on Jinan's fiscal revenue. The results show that the forecasting accuracy is high. This method can provide certain reference ideas for other related fields.

2) Compared with the prediction model of a single support vector machine, the prediction model established by the Fruit fly optimization algorithm optimized support vector machine has a greatly improved convergence speed and prediction accuracy.

References

[1] Zheng Penghui, Shan Rui, Chen Jing. The application of time series analysis in my country's fiscal revenue forecasting [J]. Journal of Chongqing University of Arts and Science (Natural Science Edition), 2008(02):15-18.
[2] Jiang Feng, Zhang Ting, Zhou Yanling. Local fiscal revenue forecast based on Lasso-GRNN neural network model [J]. Statistics and Decision, 2018, 34(19): 91-94.
[3] Zhao Guanghua, Liu Wei. The application of multiple regression models in regional economic forecasting [J].China Business, 2009(13): 180-181.