Tax Forecasting Based on Linear Discriminant Analysis-wavelet Support Vector Regression Algorithm

Zou Rong

Institute of Finance and Public Administration, Anhui University of Finance and Economics, Bengbu 233030, China
zourongg@126.com

Abstract

The accurate prediction for future tax values is the key to make the reasonable tax collection policy. In order to improve the prediction accuracy of wavelet support vector regression, linear discriminant analysis-wavelet support vector regression algorithm is proposed to predict future tax values in the paper. In the linear discriminant analysis-wavelet support vector regression algorithm, the linear discriminant analysis (LDA) is used to optimize wavelet support vector regression algorithm. Tax data of China from 1978 to 2001 are used as our experimental data. The comparison of tax prediction values between the LDA-WSVR model and traditional WSVR model with the 3~7 input nodes respectively shows that the prediction results for future tax of LDA-WSVR are better than those of traditional WSVR.

Keywords: Linear Discriminant Analysis, Wavelet Support Vector Regression, Regression Function, Tax Prediction

1. Introduction

The accurate prediction for future tax values is the key to make the reasonable tax collection policy[1]. It is well-known that support vector regression (SVR) has been used as an efficient tool to solve the complex nonlinear prediction problem[2-5]. In order to improve the prediction accuracy of traditional support vector regression model, some scholars proposed that wavelet support vector regression was used for tax prediction[6-9]. However, wavelet support vector regression has some shortcomings of over-fitting and local extremum, which leads to the poor prediction accuracy of wavelet support vector regression[10-12]. Thus, in the study, linear discriminant analysis-wavelet support vector regression algorithm is proposed to predict future tax values. In the linear discriminant analysis-wavelet support vector regression algorithm, the linear discriminant analysis (LDA) is used to optimize wavelet support vector regression algorithm.

Tax data of China from 1978 to 2001 are used as our experimental data, and the years 1978~2001 are denoted as 1~24 respectively. We use MATLAB 7.1 to develop the tax forecasting model based on linear discriminant analysis-wavelet support vector regression algorithm. Traditional wavelet support vector regression algorithm is compared with the linear discriminant analysis-wavelet support vector regression algorithm to show the superiority of the linear discriminant analysis-wavelet support vector regression algorithm. The LDA-WSVR models with the 3~7 input nodes respectively are trained and tested to find the optimal number of input nodes of the prediction models in the study. It can be seen that the testing results of the LDA-WSVR model and traditional WSVR model with 4 input nodes have the best effects. The comparison of tax prediction values between the LDA-WSVR model and traditional WSVR model with the 3~7 input nodes respectively shows that the prediction results for future tax of LDA-WSVR are better than those of traditional WSVR.

2. The principle of support vector regression

Support vector regression (SVR), based on the statistical learning theory, is a new regression method. Support vector regression, based on the principle of structured risk minimization, seeks to minimize an upper bound of the generalization error instead of the empirical error.

Given a set of training samples \( \{(x_i, y_i)\}_{i=1}^N \) with the input vector \( x_i \) and target value \( y_i \), the nonlinear mapping formula between the input vector and target output of support vector regression is
shown as followings:

\[ f(x) = \omega' \cdot \theta(x) + b \]  

(1)

where \( \omega \) is the weight and \( b \) is bias.

In order to prevent over-fitting and improve the generalization ability, the optimization problem below is created and solved.

Minimize

\[ T(\omega, \xi, \xi^*) = \frac{1}{2} ||\omega||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \]  

(2)

Subject to

\[
\begin{cases}
 y_i - \langle \omega \cdot \theta(x) \rangle + b \leq \epsilon + \xi_i & \xi_i \geq 0 \\
\langle \omega \cdot \theta(x) \rangle + b - y_i \leq \epsilon + \xi_i^* & \xi_i^* \geq 0
\end{cases}
\]

where \( \xi_i, \xi_i^* \) are the positive slack variables; \( C \) denotes the penalty parameter.

Finally, the nonlinear regression of support vector regression is described as followings:

\[ f(x) = \sum_{i=1}^{N} (\beta_k - \beta_k^*) K(x_i, x) + b \]  

(3)

where \( \beta_k, \beta_k^* \) are the Lagrange multipliers; \( K(x_i, x) \) is the kernel function, which can be expressed as followings:

\[ K(x_i, x) = \theta(x_i) \theta(x) \]  

(4)

3. The linear discriminant analysis-wavelet support vector regression algorithm

The wavelet function in the wavelet support vector regression model can be described as followings:

\[ k(x, x') = \prod_{i=1}^{M} \theta \left( \frac{x_i - c_i}{a} \right) \theta \left( \frac{x_i' - c_i'}{a} \right) \]  

(5)

where \( M \) is the sample number.

Then, translation invariant wavelet kernel function can be given as followings:

\[ k(x, x') = \prod_{i=1}^{M} \theta \left( \frac{x_i - x_i'}{a} \right) \]  

(6)

where \( x_i \) denotes the variable vector.

The wavelet kernel of this Morlet mother wavelet is described as followings:

\[ k(x, x') = \prod_{i=1}^{M} \theta \left( \frac{x_i - x_i'}{a} \right) = \prod_{i=1}^{M} \cos \left( 1.75 \cdot \frac{x_i - x_i'}{a} \right) e^{x p \left( -\frac{||x_i - x_i'||^2}{2a^2} \right)} \]  

(7)

This Morlet wavelet kernel function can be used as the kernel of support vector regression. The linear discriminant analysis (LDA) is used to optimize wavelet support vector regression algorithm, and the key function of linear discriminant analysis is described as followings:
\begin{align*}
S_1 &= \sum_{i=1}^{h} N_i (\lambda_i - \lambda)(\lambda_i - \lambda)'
\tag{8} \\
S_2 &= \sum_{i=1}^{h} \sum_{x_i \in X_i} (x_i - \lambda_i)(x_i - \lambda_i)'
\tag{9}
\end{align*}

where $\lambda_i$ denotes the mean for all the samples.

4. Experimental analysis for tax forecasting based on linear discriminant analysis-wavelet support vector regression algorithm

Tax data of China from 1978 to 2001 are used as our experimental data, and the years 1978~2001 are denoted as 1~24 respectively. We use MATLAB 7.1 to develop the tax forecasting model based on linear discriminant analysis-wavelet support vector regression algorithm. In the linear discriminant analysis-wavelet support vector regression algorithm, the linear discriminant analysis (LDA) is used to optimize wavelet support vector regression algorithm. In order to show the superiority of the linear discriminant analysis-wavelet support vector regression algorithm compared with traditional support vector regression algorithm, traditional wavelet support vector regression algorithm is employed to predict the tax values.

As the research results indicate that the number of input nodes of the prediction models including LDA-WSVR and WSVR has a great influence on their testing effects, the LDA-WSVR models with the 3~7 input nodes respectively are trained and tested to find the optimal number of input nodes of the prediction models in the study. The testing effects of the LDA-WSVR model and traditional WSVR model with 3 input nodes are shown in Fig.1, the testing effects of the LDA-WSVR model and traditional WSVR model with 4 input nodes are shown in Fig.2, the testing effects of the LDA-WSVR model and traditional WSVR model with 5 input nodes are shown in Fig.3, the testing effects of the LDA-WSVR model and traditional WSVR model with 6 input nodes are shown in Fig.4, and the testing effects of the LDA-WSVR model and traditional WSVR model with 7 input nodes are shown in Fig.5. The comparison of the mean prediction error between the LDA-WSVR model and traditional WSVR model are shown in Tab.1. It can be seen that the testing results of the LDA-WSVR model and traditional WSVR model with 4 input nodes have the best prediction effects. The comparison of tax prediction values between the LDA-WSVR model and traditional WSVR model with the 3~7 input nodes respectively shows that the prediction results for future tax of LDA-WSVR are better than those of traditional WSVR.
Figure 1. The comparison of the testing effects of the LDA-WSVR model and traditional WSVR model with 3 input nodes.

Figure 2. The comparison of the testing effects of the LDA-WSVR model and traditional WSVR model with 4 input nodes.
Figure 3. The comparison of the testing effects of the LDA-WSVR model and traditional WSVR model with 5 input nodes.

Figure 4. The comparison of the testing effects of the LDA-WSVR model and traditional WSVR model with 6 input nodes.
Figure 5. The comparison of the testing effects of the LDA-WSVR model and traditional WSVR model with 7 input nodes

Table 1. The comparison of the mean prediction error between the LDA-WSVR model and traditional WSVR model with 3~7 input nodes

<table>
<thead>
<tr>
<th>The number of input nodes</th>
<th>MAPE of LDA-WSVR</th>
<th>MAPE of WSVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.0463</td>
<td>0.0562</td>
</tr>
<tr>
<td>4</td>
<td>0.0169</td>
<td>0.0298</td>
</tr>
<tr>
<td>5</td>
<td>0.034</td>
<td>0.0467</td>
</tr>
<tr>
<td>6</td>
<td>0.0413</td>
<td>0.0513</td>
</tr>
<tr>
<td>7</td>
<td>0.457</td>
<td>0.0644</td>
</tr>
</tbody>
</table>

4. Conclusion

In order to improve the prediction accuracy of wavelet support vector regression, linear discriminant analysis-wavelet support vector regression algorithm is proposed to predict future tax values in the study. In the linear discriminant analysis-wavelet support vector regression algorithm, the linear discriminant analysis (LDA) is used to optimize wavelet support vector regression algorithm. The experimental results show that the testing results of the LDA-WSVR model and traditional WSVR model with 4 input nodes have the best prediction effects. The comparison of tax prediction values between the LDA-WSVR model and traditional WSVR model with the 3~7 input nodes respectively shows that the prediction results for future tax of LDA-WSVR are better than those of traditional WSVR.

5. References


