Study on Passenger Traffic Volume Forecast of High-speed Railway

Hai-bo MU and Li SUN

School of Traffic & Transportation Engineering, Lanzhou Jiaotong University, Lanzhou Gansu, China

Keywords: Passenger traffic volume of high-speed railway, Partial least-squares regression, Grey forecasting model, IOWA combination forecast.

Abstract. To improve the forecasting accuracy of the future passenger traffic volume of high-speed railway in our country, a combinational forecasting method based on induced ordered weighted averaging (IOWA) operator is proposed, and the optimal combined forecasting model based on IOWA operator is established by regarding the sum of error squares as a criterion. Finally, the passenger traffic volume of high-speed railway during the 13th five-year is predicted by using IOWA combination model. It indicates that the proposed method in this article can effectively improve the prediction accuracy, and it is suitable for predicting our country’s passenger traffic volume of high-speed railway.

Introduction

As a new mode of transportation, high-speed railway is regarded as the best means of transport for passengers to solve the busy trunk among big cities. The prediction of passenger traffic volume of high-speed railway plays an important role in the country's economic development pattern and resource allocation, and the operation and management of the railway enterprise. There are many methods for passenger traffic volume forecast \[1\]. Although numerous methods can be used for prediction, any single forecasting method is difficult to obtain satisfactory results. Some researchers put forward combination forecasting method based on Induced Ordered Weighted Averaging Operator (IOWA) \[2\], and some scholars applied IOWA combination forecasting model to passenger traffic volume forecast \[3\].

In this paper, IOWA operator is adopted to give weight values according to prediction accuracy of the single prediction method in each point within the sample interval from high to low, the sum of error squares is regarded as a criterion and therefore, a new combination forecasting model is established. Finally, the validity of the proposed model and method is verified by the applied results of China’s passenger traffic volume forecast of high-speed railway.

Combination Forecasting Model for Passenger Traffic Volume of High-speed RAILWAY based on IOWA Operator

Definition of IOWA Operator

IOWA is a method of information integration, which is between maximum operator and minimum operator \[4\], and the conventional weighted arithmetic average operator is its special case. IOWA operator is defined as follows \[5\]:

Suppose that \(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \ldots, \langle v_m, a_m \rangle\) is \(m\) two-dimensional array, and \(f_w\) is defined as

\[
f_w = \left(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \ldots, \langle v_m, a_m \rangle\right) = \sum_{i=1}^{m} v_i \cdot a_{i-index(i)}
\]

where \(f_w\) is said \(m\)-dimensional induced ordered weighted averaging operator which is produced by \(v_1, v_2, \ldots, v_m\), shorthand for IOWA operator, \(v_i\) is called the induced value of \(a_i\). In this formula:
\[ W = (\omega_1, \omega_2, \ldots, \omega_m) \] is the weighted vector that is associated with IOWA, and satisfies that \( \sum_{i=1}^{m} \omega_i = 1, \omega_i \geq 0, i = 1, 2, \ldots, m \). \( \nu - \text{index}(i) \) is a subscript of the \( i \)th large number of \( v_1, v_2, \ldots, v_m \) with descend order.

The definition mentioned above indicates that IOWA operator is to ordered weighted average number of \( a_1, a_2, \ldots, a_m \) corresponding with induced value \( v_1, v_2, \ldots, v_m \) from big to small order later. Corresponding weight value \( \omega_i \) has nothing to do with the size and position of \( a_i \), but to the location of induced value.

**Modeling Ideas and Steps for Combination Forecasting Model**

Specific modeling steps for combination forecasting model are as follows\(^6\):

1) Calculate prediction accuracy \( a_i \)

\[
a_i = \begin{cases} 
1 - \frac{|(x_i - x_t)|}{x_i} & \text{if } (x_i - x_t)/x_i < 1 \\
0 & \text{if } (x_i - x_t)/x_i \geq 1 
\end{cases} \quad i = 1, 2, \ldots, m; t = 1, 2, \ldots N \tag{2}
\]

Where \( a_i \) is Prediction accuracy in the \( t \)th year of the \( i \)th kind of prediction method; \( x_i \) is the predictive value in the \( t \)th year of the \( i \)th kind of prediction method; \( x_t \) is the real value of passenger traffic volume of high-speed railway in each year.

Regarding the prediction accuracy \( a_i \) as induced value of the predictive value \( x_t \), thus \( m \) kinds of single forecasting model in the \( t \)th year’s prediction accuracy and the corresponding predicted value within the range of sample constitute \( m \) two-dimensional array \( \{a_{i1}, x_{i1}\}, \{a_{i2}, x_{i2}\}, \ldots, \{a_{im}, x_{im}\} \).

2) Calculate the IOWA operator value of combination forecasting model

Suppose that \( W = (\omega_1, \omega_2, \ldots, \omega_m) \) is weighted coefficient of all kinds of forecasting methods in the combination forecast, and the prediction accuracy sequence \( a_{i1}, a_{i2}, \ldots, a_{im} \) of \( m \) kinds of single forecasting method is ordered from large to small. Suppose that \( a - \text{index}(it) \) is a subscript of the \( i \)th large prediction precision. According to the Eq. (1), we have

\[
f_{\text{IOWA}} \left( \{a_{i1}, x_{i1}\}, \{a_{i2}, x_{i2}\}, \ldots, \{a_{im}, x_{im}\} \right) = \sum_{i=1}^{m} \omega_i x_{a - \text{index}(it)} \tag{3}
\]

Eq. (3) is called IOWA operator combination forecast value that is produced by the prediction accuracy sequence \( a_{i1}, a_{i2}, \ldots, a_{im} \) in the \( t \)th year.

3) Calculate the forecast error of combination forecasting model.

\[
e_t = x_t - f_{\text{IOWA}} = x_t - \sum_{i=1}^{m} \omega_i x_{a - \text{index}(it)} \tag{4}
\]

4) Establish optimal combined forecasting model based on IOWA operator by minimizing the sum of error squares as the criterion.
\[
\min S(\omega) = \sum_{t=1}^{N} e_t^2 = \sum_{t=1}^{N} \left( x_t - \sum_{i=1}^{m} \omega_i x_{a-index(t)} \right)^2
\]

s.t.
\[
\sum_{i=1}^{m} \omega_i = 1
\]
\[
\omega_i \geq 0, i = 1, 2, \ldots, m
\]

5) Calculate the weighted coefficient of single forecasting model within the combination forecasting model according to above optimal programming model.

\[
W = (\omega_1, \omega_2, \ldots, \omega_m)^T
\]

6) Generate the calculated weighted coefficient into Eq. (3) of IOWA operator, and the predictive value of passenger traffic volume of high-speed railway can be obtained in each year for combination forecasting model.

7) Predict the passenger traffic volume of high-speed railway in the future years.

By solving the model shown in Eq. (5), we can obtain IOWA optimal weighted coefficient of combination forecast within the range of sample, represented by \( \omega' = (\omega'_1, \omega'_2, \ldots, \omega'_m)^T \). Based on the principle of prediction consistency, it can be used in IOWA combination forecast for the forecasting range \([N+1, N+2, \ldots]\)

\[
f_{IOWA'} \left( \langle a_1, x_1 \rangle, \langle a_2, x_2 \rangle, \ldots, \langle a_m, x_m \rangle \right) = \sum_{i=1}^{m} \omega'_i x_{a-index(i)}, t = N+1, N+2, \ldots
\]

The size of the prediction accuracy sequence \( a_1, a_2, \ldots, a_m \) for the forecasting range \([N+1, N+2, \ldots]\) can be reflected by \( k \) recently period average fitting precision \( \frac{k}{N-k+1} \sum_{t=1}^{N} a_{it} \) within the range of sample with the \( i \)th kind of prediction method.

In order to reflect the effectiveness of the combination forecasting model based on IOWA established in this paper, in accordance with the prediction effect evaluation principle, the following three error index is selected as evaluation index, in which \( x_t \) is the original data sequence and \( \hat{x}_t \) is the prediction results.

1) The sum of error squares: \( SSE = \sum_{t=1}^{N} (x_t - \hat{x}_t)^2 \)

2) The mean square error: \( MSE = \frac{1}{N} \sqrt{\sum_{t=1}^{N} (x_t - \hat{x}_t)^2} \)

3) The average absolute error: \( MAE = \frac{1}{N} \sum_{t=1}^{N} |x_t - \hat{x}_t| \)

### Prediction of Passenger Traffic Volume of High-speed Railway Based on Combination Forecasting Model

In this paper, the main influence factors including: gross domestic product (GDP), residents' consumption level (RCL), population (P), high-speed rail mileage (HM), civil aviation passenger traffic (CAPT), civil car ownership (CCO) and the number of domestic tourism (NDT). We select data of high-speed rail passenger traffic volume (HPTV) and the seven affecting factors mentioned above from year 2008 to 2015 in our country as shown in Table 1.

357
Table 1. The original sample data.

<table>
<thead>
<tr>
<th>Year</th>
<th>HPTV</th>
<th>GDP</th>
<th>RCL</th>
<th>P</th>
<th>HM</th>
<th>CAPT</th>
<th>CCO</th>
<th>NDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>734</td>
<td>319515.5</td>
<td>8707</td>
<td>132802</td>
<td>672</td>
<td>19251</td>
<td>5099.61</td>
<td>1712</td>
</tr>
<tr>
<td>2009</td>
<td>4651</td>
<td>349081.4</td>
<td>9514</td>
<td>133450</td>
<td>2699</td>
<td>23052</td>
<td>6280.61</td>
<td>1902</td>
</tr>
<tr>
<td>2010</td>
<td>13323</td>
<td>413030.3</td>
<td>134091</td>
<td>672</td>
<td>132802</td>
<td>672</td>
<td>19251</td>
<td>5099.61</td>
</tr>
<tr>
<td>2011</td>
<td>28552</td>
<td>489300.6</td>
<td>134735</td>
<td>2699</td>
<td>23052</td>
<td>6280.61</td>
<td>1902</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>38815</td>
<td>540367.4</td>
<td>135404</td>
<td>5133</td>
<td>26769</td>
<td>7801.83</td>
<td>2103</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>52962</td>
<td>595244.4</td>
<td>136072</td>
<td>9356</td>
<td>29317</td>
<td>9356.32</td>
<td>2641</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>70378</td>
<td>643974.0</td>
<td>136782</td>
<td>16456</td>
<td>3261</td>
<td>10933.09</td>
<td>3262</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>96139</td>
<td>685505.8</td>
<td>137462</td>
<td>19838</td>
<td>35397</td>
<td>12670.14</td>
<td>3611</td>
<td></td>
</tr>
</tbody>
</table>

Calculation of Passenger Traffic Volume of High-speed Railway Based on IOWA Operator

Let \( y \) denote passenger traffic volume of high-speed railway, and \( x_1, x_2, x_3, x_4, x_5, x_6, x_7 \) denote the seven influencing factors respectively. According to Partial least-squares regression (PLS) analysis method, we make correlation analysis for the data shown in Table 1, and the correlative coefficient between passenger traffic volume of high-speed railway and the seven influencing factors can be obtained as shown in Table 2.

Table 2. Correlation coefficient.

<table>
<thead>
<tr>
<th>Variable</th>
<th>( y )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( x_4 )</th>
<th>( x_5 )</th>
<th>( x_6 )</th>
<th>( x_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_2 )</td>
<td>0.98</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_3 )</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_4 )</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_5 )</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_6 )</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>( x_7 )</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
</tr>
</tbody>
</table>

In order to verify the effectiveness of combination forecasting model based on IOWA operator, the forecasting results obtained by PLS model and grey GM (1, 1) model from year of 2009 to 2015 are combined to establish combination forecasting model based on IOWA operator, and the optimal model can be obtained as follows:

\[
\begin{align*}
\min S(\omega_1, \omega_2) &= (4651 - 6572\omega_1 - 14084\omega_2)^2 + (13323 - 17623\omega_1 - 19778\omega_2)^2 \\
&+ (28552 - 27775\omega_1 - 27321\omega_2)^2 + (38815 - 39005\omega_1 - 40652\omega_2)^2 \\
&+ (52962 - 54775\omega_1 - 50809\omega_2)^2 + (70378 - 74572\omega_1 - 76920\omega_2)^2 \\
&+ (96139 - 92299\omega_1 - 108019\omega_2)^2 \\
\text{s.t.} & \\
&\omega_1 + \omega_2 = 1 \\
&\omega_1 \geq 0, \omega_2 \geq 0
\end{align*}
\]

The optimal weighted coefficient of combination forecasting model based on IOWA can be calculated by using MATLAB optimization toolbox, the value of \( \omega_1 \) and \( \omega_2 \) is 0.8997 and 0.1003 respectively. According to the predicted results obtained by IOWA model, PLS model and grey GM (1, 1) model, three major indicators namely SSE, MSE and MAE are calculated and listed in table 3.
Table 3. Analysis of the predicted results.

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>SSE</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS model</td>
<td>6.4×10^7</td>
<td>1143.22</td>
<td>2782.29</td>
</tr>
<tr>
<td>GM(1,1) model</td>
<td>3.2×10^8</td>
<td>2549.54</td>
<td>5298.57</td>
</tr>
<tr>
<td>IOWA combination forecast</td>
<td>5.5×10^7</td>
<td>1060.43</td>
<td>2353.86</td>
</tr>
</tbody>
</table>

Prediction of High-speed Rail Passenger Traffic Volume during the 13th Five-year

In order to predict high-speed rail passenger traffic volume during the 13th five-year planning, we have to predict the related data of influence factors. According to the situation of our country, we assume that the mileage during the 13th five-year is increased uniformly. The other influence factors are predicted by the grey GM (1, 1) model. We adopt PLS model and grey forecasting model to predict the passenger traffic volume firstly, and then in the light of the optimal combination weighted values $\omega_1 = 0.8997, \omega_2 = 0.1003$, we predict passenger traffic volume with IOWA combination forecasting model, and the results are shown in Table 4.

Table 4. Predicted value with different models Units: ten thousand people.

<table>
<thead>
<tr>
<th>Year</th>
<th>PLS model</th>
<th>GM(1, 1) model</th>
<th>IOWA combination forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>121278</td>
<td>213024</td>
<td>130483</td>
</tr>
<tr>
<td>2018</td>
<td>137192</td>
<td>299151</td>
<td>153441</td>
</tr>
<tr>
<td>2019</td>
<td>154344</td>
<td>420100</td>
<td>181007</td>
</tr>
<tr>
<td>2020</td>
<td>172896</td>
<td>589950</td>
<td>214739</td>
</tr>
</tbody>
</table>

Since combination forecasting model is established on the basis of single forecasting model, we found that combination forecasting model has obvious advantages. According to the results of IOWA combination forecast model, by 2018, China's high-speed rail passenger traffic volume will reach 1.53441 billion people, and it will reach 2.14739 billion people by the end of the 13th five-year in keeping current trend of economic growth and social development of the whole society, and high-speed railway passenger demand presents rising trend in the next few years.

Conclusions

We present a quantitative description of our country’s passenger traffic volume of high-speed railway based on IOWA operator combination forecasting method. On the basis of all required data from year 2008 to 2015, taking PLS model and grey forecasting model as single forecasting model, the combination forecasting model based on IOWA operator has been established, and its prediction accuracy has been verified. We found that prediction effect of combination forecasting model is better than any other single forecasting models, and the IOWA combination model established in this paper has higher prediction accuracy. Furthermore, its predicting results are more credible. Finally, our country’s passenger traffic volume of high-speed railway during the 13th five-year has been predicted in accordance with IOWA combination forecasting model. The analytical results in this paper show that the proposed method in this article can effectively improve the prediction accuracy, and it is suitable for predicting our country’s passenger traffic volume of high-speed railway.

Acknowledgement

This research is supported by National Nature Science Foundation of China (Nos. 61563029, 71671079, 71361018, 71571090).
References


