Combination forecast on tourism demands based on IOWA operator

Fang Lu, Xinchun Zhu, Zhiqiang Peng

School of Management and Science and Engineering, Hunan University of Technology,Zhuzhou,412007, Hunan, China

Xinmin Zhou*

School of Information, Hunan University of Commerce, Changsha 410205, Hunan, China
**Corresponding author*

Abstract
Accurate tourism demand forecasting can reduce the occurrence of the unreasonable tourism investment planning, scenic spots crowded and the decline of tourism enterprise's service quality in busy season, etc. In order to improve the prediction accuracy of tourism demand, this paper analyzes the disadvantage of the general combination forecast and introduces the concept of IOWA (Induced Ordered Weighted Averaging) operator. Two kinds of single forecasting models, that is ARIMA and GM (1, 1), on tourism demand are firstly established, after then, and then the combination forecasting model in tourism demand based on the IOWA operator is established, which is concerned with two single prediction models' prediction accuracy at each time period. In the end, the combination forecasting model is applied to the forecasting of inbound tourist to Hunan Province. The results show that the proposed combination forecasting model can narrow the prediction error effectively and improve the prediction accuracy, which can be as an effective tool for tourism demand forecasting.
Key words: TOURSIM DEMAND, IOWA OPERATOR, COMBINATION FORECASTING, PREDICATION ACCURACY

Introduction

As the surge of the consume demand of people travel, tourism industry has got an unprecedented development. It has become an important part in the national economy and an important pillar industry. Travel demand is the fundamental basis of the formation of the tourism market and travel demand management and forecast is a very important link in tourism planning [1]. At present, due to the limitation of the prediction methods employed by the members of the tourism service supply chain, predicted results on tourism demand are inaccurate, thus leading to unreasonable tourism investment planning, crowded scenic spots and low service quality in on-season, even large refund. The existence of these problems has seriously restricted the willingness to travel and the development of tourism.

Since literature [2] put forward the combination forecast theory at first time, it made many scholars improve the computation accuracy of combination forecast methods and apply these combination forecast methods into various areas, such as economy, social life. Since combination forecast methods utilized comprehensively information provided by various single prediction methods, and then got the predicted value with the appropriate weighting combination of various forecast methods, it could obtain higher accuracy than single forecast method [2]. Suo Ruixia, Wang Fulin established a combined forecast model on energy consumption based on divorced coefficient method, and showed the model could be an effective tool to predict energy consumption by empirical analysis [3]. Li Bingxiang proposed a nonlinear combination forecasting method based on fuzzy neural network, and concluded it had the superiority on the prediction of the listed companies' financial crisis by comparison analysis [4]. Ng had combined time series and regression analysis to construct a combination forecast model, and applied it to predict the tender price index. The results showed that the combination forecast model could get the better predicted value than a single forecast method [5]. Lei Kewei, Fang Tianhong analyzed the linear and nonlinear law of tourism statistics and proposed a combination forecast model based on BP neural network and ARIMA model, and had predicted the number of inbound tourist arrivals in Shan Xi Province by this combination forecast model, which improved the accuracy of tourism demand forecasting[6].

From above, there are few literatures adopting the combination forecast models to predict the travel demand. And most of the combined forecast models used the same algorithm with weighted single forecast methods at different time points. However, in reality, prediction accuracy of each single forecast method at different time points is not the same, which will cause the fluctuation of prediction accuracy [7]. So, this paper firstly introduces IOWA (Induced Ordered Weighted Averaging) operator; Secondly, selects ARIMA theory and Gray theory with high accuracy in single forecast method on travel demand; lastly, constructs and applies the combination forecast model by applying IOWA to the results of ARIMA and Gray forecast results.

Model Construction

Definition 1 [8]: Set $f_w : \mathrm{R}^n \to \mathrm{R}$. f_W is n-dimensional function and $f_W(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} w_i b_i$. b_i is the i-th number when the sequence of $(a_1, a_2, \ldots, a_n)$ is arranged by descending order. $W = (w_1, w_2, \ldots, w_n)^T$ is the weighted vector of f_W and meet the $\sum_{i=1}^{n} w_i = 1\ (w_i \geq 0, i = 1, 2, \ldots, n)$. f_w is called n-dimensional OWA (Ordered Weighted Averaging) operator. And the OWA operator is a value of an orderly weighted average when the sequence $(a_1, a_2, \ldots, a_n)$ is arranged by descending order. $w_i\ (i = 1 \ldots n)$ is only related with the position of arranged sequence $(a_1, a_2, \ldots, a_n)$, having nothing to do with the value of $a_i\ (i = 1 \ldots n)$.

Definition 2 [9]: Set $\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \ldots, \langle v_n, a_n \rangle$ is two arrays with n numbers, and set $f_w(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \ldots \langle v_n, a_n \rangle) = \sum_{i=1}^{n} w_i a_{v-index(i)}$. f_W is n-dimensional IOWA (Induce Ordered Weighted Averaging) operator generated by the sequence of $(v_1, v_2, ..., v_n)$. $v_i (i = 1 \ldots n)$ is the induced value of $a_i (i = 1 \ldots n)$ and $v - index(i)\ (i = 1 \ldots n)$ represents i-th number after arranging the sequence $(v_1, v_2, ..., v_n)$ by descending order. Definition 2 shows IOWA operator is the orderly weighted average value to the corresponding sequence $(a_1, a_2, ..., a_n)$ when the reduced sequence of $(v_1, v_2, ..., v_n)$ is arranged by descending order.

Set

$$\alpha_{it} = \begin{cases} 1 - \left|(\mathrm{X}_t - X_{it})/\mathrm{X}_t\right|, & and \quad \left|(\mathrm{X}_t - X_{it})/\mathrm{X}_t\right| < 1 \\ 0, \left|(\mathrm{X}_t - X_{it})/\mathrm{X}_t\right| \geq 1 \end{cases} \quad (1)$$

Among them, X_t represents the actual value in the t-th time. X_{it} represents predicted value of the i-th forecast method at t-th time. $a_{it}\ (a_{it} \in [0,1])$ represents the prediction accuracy of the i-th forecast method at the t-th time and a_{it} is called the induction value of X_{it}. The prediction precision of n kinds of singe prediction in t-th time and the predicted value of the corresponding sample interval constitute the two dimensional array with n numbers: $\langle a_{1t}, X_{1t} \rangle, \langle a_{2t}, X_{2t} \rangle, \ldots, \langle a_{nt}, X_{nt} \rangle$

The prediction accuracy of n kinds of forecast methods is arranged in descending order at t-th time and $a - index(it)$ represents the subscript of i-th number in arranged prediction accuracy sequence.

Set

$$f_w\left(\langle a_{1t}, X_{1t}\rangle, \langle a_{2t}, X_{2t}\rangle, \ldots, \langle a_{nt}, X_{nt}\rangle\right) = \sum_{i=1}^{n} w_i X_{a-index(it)} \quad (2)$$

Where f_W is called a predictive value based on IOWA combination. From the above modeling construction process, the fundamental difference between IOWA operator combination forecast model and the traditional combination forecast model is that the weighting coefficient is only related with prediction accuracy of single forecast method at each time point, having nothing with each forecast method.

Let $e_{a-index(it)} = X_t - X_{a-index(it)}$. So the error sum of squares from 1-th to m-th time can be expressed as:

$$S = \sum_{t=1}^{m}\left(X_t - \sum_{i=1}^{n} w_i X_{a-index(it)}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left(\sum_{t=1}^{m} e_{a-index(it)} e_{a-index(jt)}\right) \quad (3)$$

S represents the error sum of squares. The combination forecast model based on IOWA operator, which minimizes the error sum of squares, can be expressed as :

$$\min S(w) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left(\sum_{t=1}^{m} e_{a-index(it)} e_{a-index(jt)}\right) \quad (4)$$

where, $\sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1,2,...,n$

Let

$$\overline{a_1}(T) = \frac{1}{T}\sum_{t=1}^{T-1} a_i(m-t) \quad (T = 1,2,...,m) \quad (5)$$

Where, $\overline{a_1}(T)$ is the average prediction accuracy from the T-th to m-th time periods, and the larger value $\overline{a_1}(T)$ is, the more accurate prediction value in forecast models. By the principle of predicting continuity, the IOWA forecasting model of two time periods based on IOWA operator can be expressed as :

$$\hat{X}_{m+T} = \left(\langle \overline{a_1}(T), X_{1,m+T}\rangle, \ldots, \langle \overline{a_n}(T), X_{2,m+T}\rangle\right) = \sum_{i=1}^{n} w_i * X_{\bar{a}(T)-index(i),(m+T)} \quad (T = 1...m) \quad (6)$$

Where, $\bar{a}(T) - index(i)$ is the subscript of the i-th value in the sequence of $\{\overline{a_n}(T)\} (T = 1..m)$

Combination forecast on tourism demands of Hunan province based on IOWA operator

The steps of combination forecast are as follows:

Step 1: Determine the tourism demand data based on the actual situation;

Step 2: Establish greater than or equal to two kinds of single forecast models on travel demand, and then predict the next phase of travel demand according to the established forecast model (6) respectively; (This paper chooses ARIMA method and gray theory, which have better prediction accuracy.)

Step 3: Analyze the prediction results using step 2. If the predictions are normal, then go to Step 4, or else, analyze the causes, and come back to step 2 ;

Step 4: Construct combination forecast model using the results of step 2, and forecast the tourism demand in Hunan province by the combination forecast model;

Step 5: Analyze the results and make a conclusion.

In step 3, the judgment steps of abnormal prediction results are as follows:

Firstly, make a clear definition for the abnormal prediction results according to the experience. Secondly, the prediction result is normal if results are in a certain error range based on historical data. Otherwise, analyze the cause and give solutions to this "abnormal." Lastly, re-predict when it is needed.

Two kinds of single tourism forecast methods

Time series analysis is a method commonly used in the enterprise, including exponential smoothing, ARIMA, the moving average method and so on. The ARIMA is a typically research method in the quantitative study of tourism demand, and the model can fully consider historical data and its sequence rule. So, ARIMA can overcome the random interference in travel demand very well [10]. What's more, due to the tourism market embrace the complex and uncertain characteristics, the main factors for the development of the tourism market show significant gray nature [11]. The Gray theory has obvious advantages on uncertainty and the absent data and information. So the paper chooses gray prediction model GM (1, 1) and ARIMA model in step 2.

Table 1 is the number of annual tourism demand during 2000-2012 in Hunan Province (Data source: Statistical Yearbook 2000-2012, Hunan Province).

Table 1. Tourist arrivals in Hunan Province from 2001 to 2012

Year	Tourist arrivals (10,000)	Year	Tourist arrivals (10,000)
2000	4695.40	2007	10897.47
2001	5035.52	2008	12829.97
2002	5756.62	2009	16065.03
2003	5970.11	2010	20398.03
2004	6486.34	2011	25328.29
2005	7180.98	2012	30524.60
2006	9195.31		

Literature [10] firstly used SPSS 20.0 to construct ARIMA model on tourist arrivals in Hunan Province; then the residual error sequence was confirmed to white noise sequence by model detecting; thirdly, The form of the ARIMA was identified as ARIMA (0,2,0) with R^2 =0.995. This model had good explanatory power and could explain 99.5% changes in tourism demands.

Literature [11] calculated the parameters in GM(1,1) by Excel and concluded a = -0.1915, u = 2145.80. Therefore the model is:

$$X^{(1)}(k+1)=15898.93e^{0.1915k}-11203.53$$

$$X^{(0)}(k)=X^{(1)}(k+1)-X^{(1)}(k)$$

And $X^{(0)}(k)$ is the predicted value on the number of tourists.

The above two methods are adopted to forecast the number of tourism arrivals respectively in Hunan Province during 2005-2012. The prediction accuracy of two kinds of models is shown in Table 2 according to the formula (1).

Table 2. Results of two kinds of single model of tourism demand forecast

Year	Actual value (10,000)	Grey predicted value	ARIMA predicted value	Grey prediction accuracy	ARIMA prediction accuracy
2005	7180.98	7220.66	7444.04	0.9945	0.9634
2006	9195.31	8744.94	8317.09	0.9510	0.9045
2007	10897.47	10591.00	11651.11	0.9719	0.9308
2008	12829.97	12826.76	13041.10	0.9998	0.9835
2009	16065.03	15534.49	15203.94	0.9670	0.9464
2010	20398.03	18813.82	19741.56	0.9223	0.9678
2011	25328.29	22785.42	25172.50	0.8996	0.9939
2012	30524.60	27595.43	30700.02	0.9040	0.9943

From Table 2, it is in 2006 that ARIMA model has the minimum prediction accuracy and the prediction accuracy is only 90.45%, the forecast accuracy in 2007 and 2009 is 95% or less, but in other years it has a good accuracy; The prediction accuracy of Gray GM (1, 1) model is under 90% in 2011, and the accuracy is only 92.23%, 90.40%in 2010, 2012 respectively, but the error of the prediction accuracy is less than 5% in other years. Although the two models have relatively good prediction accuracy, but the deviation of prediction accuracy based on a single model is still very big, and the prediction accuracy is not stable. To meet the more accurate prediction requirements on tourism demands. a new combination forecast model with more accurate and stable prediction is constructed in this paper.

Combination forecast of tourism demand

It can be seen from the above that ARIMA model and GM (1, 1) model have their own advantages and disadvantages. So we should combine their advantages of these two kinds of models. An actual example is explained according to the steps of combination forecast using the prediction results from 2008 to 2012 as the sample and test sets.

Y_t is the predictive value of combination forecasting model based on IOWA operator. It can be obtained by equation (2):

$$Y_t = IOWA(\langle a_{1t}, X_{1t}\rangle, \langle a_{2t}, X_{2t}\rangle, \dots, \langle a_{nt}, X_{nt}\rangle) = \sum_{i=1}^{n} w_i X_{a-index(it)}$$

$$(t = 1, 2, \dots, 5)$$

Where $a-index(it)$ is the subscript of i-th largest prediction accuracy at t-th time. Thus, when t = 1,

$Y_1 = IOWA(\langle a_{11}, X_{11}\rangle, \langle a_{21}, X_{21}\rangle) = (\langle 0.9835, 13041.10\rangle, \langle 0.9998, 12826.76\rangle)$
$= 12826.76w_1 + 13041.10w_2$

when t = 2,

$Y_2 = IOWA(\langle a_{12}, X_{12}\rangle, \langle a_{22}, X_{22}\rangle) = (\langle 0.9464, 15203.94\rangle, \langle 0.9670, 15534.49\rangle)$
$= 12826.76w_1 + 13041.10w_2$

when t = 3,

$Y_3 = IOWA(\langle a_{13}, X_{13}\rangle, \langle a_{23}, X_{23}\rangle) = (\langle 0.9678, 19741.56\rangle, \langle 0.9223, 18813.82\rangle)$
$= 19741.56w_1 + 18813.82w_2$

Similarly, we can obtain the combined forecasting value when t = 4, 5. Put the above results into the formula (3) and (4), and get the optimal model:

$$\min S(w_1, w_2) = 767478.56w_1^2 + 2756940.03w_1w_2 + 18341997.91w_2^2$$

$$s.t.\begin{cases} w_1 + w_2 = 1 \\ w_1, w_2 \geq 0 \end{cases}$$

Solutions of the above mentioned model BY MATLAB and the optimal weights coefficient of combination forecast model are:

$w_1 = 1, \quad w_2 = 0$

To show the effectiveness of the proposed prediction model based on IOWA operator, this paper forecasts tourist arrivals in Hunan Province during 2008-2012 using three methods constructed by this paper, and select MAE(the average absolute error), RMSE (the percentage of the mean square error), MAPE (the mean absolute percentage error) as indicators of the effectiveness evaluation. The results are shown in Table 4 according to Table 3.

Table 3. Evaluation formulas

Index	MAE	RMSE	MAPE
formula	$\frac{\sum \lvert y_t - \hat{y}_t \rvert}{N}$	$\sqrt{\frac{\sum (y_t - \hat{y}_t)}{N}}$	$\frac{\sum \left\lvert \frac{y_t - \hat{y}_t}{y_t} \right\rvert}{N}$

Where $\hat{y}_t$ represents the predicted value, y_t indicates the actual value, N denotes the value of the number of test data. The smaller the value of MAE, RMSE and MAPE are, the higher of the prediction accuracy is.

Table 4. Comparison of forecast models

Evaluation indicators		MAE	RMSE	MAPE
single forecast	ARIMA	411.98	504.39	0.0228
	GM (1, 1)	1518.00	1888.78	0.0615
combination forecast	w_1=1,w_2=0	304.29	391.79	0.0155

Comparing three methods in table 4, the prediction results of the combination forecast model based on IOWA are more stable, the prediction error is smaller and the prediction accuracy is higher. The mean absolute percentage error of the combined forecast model based on IOWA is 1.55%, and the corresponding values of two single forecast models are 2.28% and 6.15% respectively. Both the average absolute error and the percentage of the mean square error of the combined forecast model based on IOWA and prediction are smallest among three models. In comparison, the combination forecast based on IOWA has better forecast effects, and it not only overcome the shortcomings of single forecast model, but also improve accuracy and stability of the prediction.

Conclusions

Because there are many influence factors on tourism demand forecast decided by the characteristic of tourism, the tourism demand forecast is more complex and uncertain. The traditional single tourism demand forecast methods cannot meet the prediction accuracy and stability of tourism demand forecast, and single forecast method don't take into account the changes of prediction accuracy at different time. So to improve the prediction accuracy and stability, the paper constructs the combination forecast model based on IOWA operator and applies the model to predict tourists' arrivals in Hunan province. The results show that the stability and accuracy of the combination forecast model has improved greatly; the maximum of average absolute error with l2.28% and 6.15% in single forecast models falls to less than 2%; and MAE and RMSE indicators are also better than single forecast models.

But there is still a further development space for combination forecast on tourism demand. This paper only establishes the combination forecast model by ARIMA and GM (1, 1), not including other forecast models. Although the prediction results are more accurate and stable by the combination forecast models based on IOWA operator, we should analyze the combination forecast models using other forecast models.

Acknowledgements

This work was supported by National Natural Science Foundation of China (NO:71201053) and National Social Science Foundation of China(NO: 13CJY007).

References

1. Song H Y (2012) Tourism supply chain management. Routledge, UK, p.p.25-45.
2. Bate J M, Granger C W J (1969) The Combination of Forecasts. OR, 20(4), p.p.451-468.
3. Suo Ruixia, Wang Fulin (2010) The application of combination forecast model in the energy consumption forecast. *The practice and understanding of mathematics*, 40(18), p.p.80-85.
4. Li Bingxiang (2005) The research of enterprise financial crisis nonlinear combination forecast method based on fuzzy neural network nonlinear. *Journal of management engineering*, 19(1), p.p. 19-23.
5. Ng S T, Cheung S O, Skitmore M,et al.(2004) An integrated regression analysis and time series model for construction tender price index forecasting. *Construction Management and Economics*, 22(5), p.p.483-493.
6. Lei Kewei, Fang Tianyu (2009) The tourism market demand forecasting of Shanxi province based on combination model. *Journal of northwest normal university (Natural science edition)*, 45(2), p.p. 99-102.
7. Guo R J, Liu Y (2012) Research on Collaborative Forecasting Model Based on CPFR. *Software Engineering and Knowledge Engineering: Theory and Practice*, no.1, p.p. 523-529.
8. Chen Huayou, Liu Chunlin. A new method of combination forecast based on the IOWA operator. *Prediction*, 2003, 22(6), p.p.61-65.
9. Jain C L. (2007) Benchmarking forecasting models. *Journal of Business Forecasting Methods and Systems*, 26(4), p. p.15-20.
10. Xiong Zhibin (2011) The research based on ARIMA and neural network ensemble GDP time series prediction. *Mathematical statistics and management*, 20(2), p.p. 306-314.
11. Zhu Xiaohua, Yang Xiuchun, Cai Yunlong (2005) Tourist passengers forecasting model based on grey system theory (Take China's inbound tourism tourists as an example). *Economic geography*, 25(2), p.p. 232-235.