

Article

Intelligence in Tourism Management: A Hybrid FOA-BP Method on Daily Tourism Demand Forecasting with Web Search Data

Keqing Li ¹ , Wenxing Lu ^{1,2,*}, Changyong Liang ^{1,2} and Binyou Wang ¹

¹ School of Management, Hefei University of Technology, Hefei 230009, China; keqinglhfut@gmail.com (K.L.); cyliang@hfut.edu.cn (C.L.); wbylrzzy@163.com (B.W.)

² Ministry of Education Key Laboratory of Process Optimization and Intelligent Decision-making, Hefei University of Technology, Hefei 230009, China

* Correspondence: luwenxing@hfut.edu.cn; Tel.: +86-180-5604-4985

Received: 17 May 2019; Accepted: 5 June 2019; Published: 11 June 2019



Abstract: The Chinese tourism industry has been developing rapidly for the past several years, and the number of people traveling has been increasing year by year. However, many problems still beset current tourism management. Lack of effective management has caused numerous problems, such as tourists stranded during tourist season and the declining service quality of scenic spots, which have become the focus of tourists' attention. Network search data can intuitively reflect the attention of most users through the combination of the network search index and the back propagation (BP) neural network model. This study predicts the daily tourism demand in the Huangshan scenic spot in China. The filtered keyword in the Baidu index is added to the hybrid neural network, and a BP neural network model optimized by a fruit fly optimization algorithm (FOA) based on the web search data is established in this study. Different forecasting methods are compared in this paper; the results prove that compared with other prediction models, higher accuracy can be obtained when it comes to the peak season using the FOA-BP method that includes web search data, which is a sustainable means of practically solving the tourism management problem by a more accurate prediction of tourism demand of scenic spots.

Keywords: tourism management; hybrid method; fruit fly optimization algorithm; neural network; web search data; forecast of daily tourism demand; optimization method

1. Introduction

The Chinese tourism industry has grown along with development of the Chinese economy. According to statistics, the number of inbound and domestic tourists in China are increasing year by year and the tourism industry is developing rapidly [1]. Such rapid growth has resulted in tourism management problems requiring urgent solutions, including forecasting tourism demand especially when large numbers of tourists travel to scenic areas for short-term visits. Management is under considerable pressure, and any negligence can cause serious public safety problems. For instance, on 2 October 2013, many tourists were stuck at the entrance of Jiuzhaigou Valley because of overcrowding. To prevent this from happening again and ensure tourism develops healthily and sustainably, the forecast of tourist flow, especially short-term forecasting, is an important research direction.

Forecasting methods used in the past include an econometric model [2], a time series model [3,4], artificial neural network and support vector machine, and hybrid methods [5]. Most of these techniques are based on historical data, but the long lag period of the predicted values often leads to problems. At present, an increasing number of scholars regard network search as an important and leading source

of research data and timely information [6]. When people search for information on the Internet, their search record can reflect their concerns.

This study aims to propose an effective short-term tourism demand forecasting method that can effectively forecast daily tourist flow on the basis of web search data and back propagation (BP) neural network optimized by a fruit fly optimization algorithm (FOA). Meanwhile, the hybrid neural network contains selected web search data in order to optimize the prediction effect of the model, which is proved to be effective in this study.

This paper proceeds as follows: Section 2 gives the background of this study. Section 3 proposes the process by which the model is built to forecast the tourist flow using network search data. Section 4 provides the process of empirical study. Section 5 presents the result and evaluation of our study. Finally, Section 6 discusses the contributions of this work and implications for further research.

2. Background

2.1. Research on Tourism Demand Forecasting

In recent years, a number of research on the prediction of tourist flow and various prediction methods have been proposed, including an econometric model, a time series model, artificial neural network and support vector machine, and hybrid methods [5]. Econometric models [7,8], which are widely used for forecasting tourism demand, analyze the causal relationships between dependent variables (tourism demand) and explanatory variables (influencing factors). A time series model is always based on historical data, and many scholars use this model to research tourism demand forecasting [9]. With the development of computer technology, artificial intelligence methods such as artificial neural network and support vector machine are being applied in the research on tourism demand forecasting [10,11]. Some scholars propose hybrid methods by combining an econometric model and artificial intelligence method or time series models [12]. No single model works in every situation because different models have their own applicable occasions [5]. Genetic algorithm is used to optimize neural network to improve prediction accuracy [13]. A vector error correction model (VECM) was used for forecasting the tourism demand of Jeju Island [14]. Neural networks (NN) are used to improve the accuracy of the Grey–Markov (GM) forecasting model [15]. An autoregressive integrated moving average (ARIMA) model is used to predict the urban residents' future travel rate in five years [16].

The researches of the above scholars focus on the tourism demand forecasting in a long-term forecasting such as several years and months. Daily forecasting is rarely studied by scholars, and deserves further study, and which is concerned in this study.

2.2. Forecasting with Network Search Data

Internet search is an important channel for people to search for information. With the popularity of mobile Internet, people can more conveniently inquire information through search engines [6]. In recent years, the use of network data has also provided a new data source and analysis basis for social science. Ginsberg et al. [17] used the Google search engine to determine whether the online search index of flu-related keywords was highly correlated with the number of people with influenza in the same period. The researchers successfully predicted the trend of influenza outbreak and proved that the search index has a certain predictive ability for the epidemic. The method of network search index as a prediction tool quickly spread in different research directions. Valuable research achievements that have used this method include predictions on a film box office [18], consumer confidence index [19], unemployment [20], and stock market [21].

Baidu is the largest search engine in China and has the most users. Research also shows that when studying consumer behavior, Baidu search has higher predictive power than Google search [22].

3. Methodology

This study proposes a complete modeling process—from selecting the keywords, getting the Baidu index, analyzing the correlation between daily Baidu index and actual total tourist flow, and choosing the index with the most correlation, to building forecasting models and evaluating the corresponding performance. The steps are shown as follows:

(1) Select the keywords. This step depends on the process by a Chinese traveler who is considering the plan to visit a scenic spot. Before traveling, he/she may search for information about destination, weather, strategy, price, hotel, and the like on the Internet. The traveler defines the keywords related to his/her destination.

(2) Obtain the corresponding keywords from Baidu index for correlation analysis. The Baidu index is based on the search volume of netizens in Baidu. It takes keywords as the statistical object to scientifically analyze and calculate the weighted sum of the search frequency of each keyword in Baidu web search.

(3) Considering the lag in web search, set the lag time and analyze the correlation between the Baidu index and actual total tourist flow. The most relevant lag period will be selected.

(4) The improved FOA algorithm is used to optimize the BP neural network. A hybrid FOA-BP model is established in this study in order to predict daily tourism demand.

(5) Determine the parameters in the model in order to train the model.

(6) Evaluate the accuracy. Genetic algorithms–back propagation (GA-BP) neural network and particle swarm optimization–back propagation (PSO-BP) neural network are selected as the benchmark models. The mean absolute percentage error (MAPE) is selected as the evaluation standard.

Current research on keyword selection method has not reached a consensus yet. At present, the three main methods of keyword selection [6] are as follows: technical method, direct method, and scope method. With the technical method, all possible keywords should be brought into the research scope via high-performance computer technology. The direct method determines the keyword directly by subjective experience. The scope method initially determines the range of a choice of words and then selects keywords within the range.

3.1. Back Propagation Neural Network

The neural network of error BP training algorithm, or BP network, is a multilayer feedforward network with hidden layers, which systematically solve the problem of learning the connection weight of hidden units in multilayer networks. If the number of input nodes of the network is M and the number of output nodes is L , then the neural network can be regarded as the mapping from M -dimensional Euclidean space to L -dimensional Euclidean space. This mapping is highly nonlinear. The structure diagram of the BP network is shown in Figure 1.

Its basic principle is the gradient maximum drop method, and the central idea is to adjust the weight (W_{ij} , W_{ki}) to minimize the total error of the network. Gradient search technology is adopted to minimize the error mean square of the actual output value and the expected output value of the network. In the network learning process, the error (e_k) propagates back and corrects the weight coefficient. The vulnerability to many external factors, such as weather, weekend, holidays, and so on, causes tourist traffic to be highly nonlinear. Traditional statistical models cannot easily show these complex nonlinear features [23]. Owing to the strong nonlinear mapping ability of the BP neural network, it is selected as the prediction model of this study.

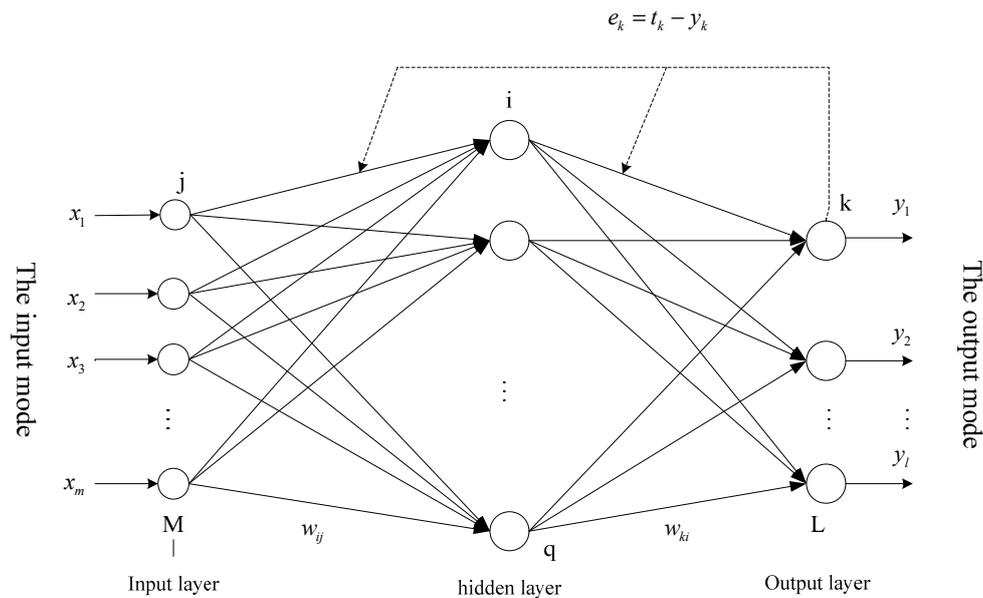


Figure 1. Back propagation neural network.

3.2. Fruit Fly Optimization Algorithm

There are several intelligent algorithms before such as genetic algorithms (GA) proposed in 1975 by Holland [24], and the particle swarm optimization (PSO) proposed in 1995 by Eberhart [25]. Fruit fly optimization algorithm (FOA) is a new swarm intelligent algorithm proposed in 2011 by Wen-Tsao Pan [26], and at present the research is still in the initial stage [27]. The algorithm is simple in process, with few control parameters and easy to implement. Some scholars have successfully applied it to structural engineering design optimization problems [28], wireless sensor network layout [29], resource-constrained project scheduling problems [30] and other fields.

Standard fruit fly optimization algorithm steps are as follows:

Step 1: Initialize population size, termination criterion, and the fruit fly swarm location (a_o, b_o) .

Step 2: Foraging with smell. Individual flies use their sense of smell to find random distances and directions for food. Generate the i th location of fruit fly randomly as (a_i, b_i) :

$$\begin{cases} a_i = a_o + \text{RandomValue} \\ b_i = b_o + \text{RandomValue} \end{cases} \quad (1)$$

where *RandomValue* means random distances and directions.

Step 3: First, the distance from the origin (Dist) is estimated, and then the determination value of smell concentration (S) is calculated, which is the reciprocal of the distance:

$$\text{Dist}_i = \sqrt{(a_i^2 + b_i^2)} \quad (2)$$

$$S_i = \frac{1}{\text{Dist}_i} \quad (3)$$

Step 4: Substitute S_i into the smell concentration determination function to find the smell concentration:

$$\text{Smell}_i = \text{Function}(S_i) \quad (4)$$

Step 5: Find the fruit flies with the lowest concentration of smell (strive for the minimum):

$$\text{bestSmell} = \min(\text{Smell}_i) \quad (5)$$

Step 6: The fruit fly swarms fly towards this position using vision, and record the best smell concentration and position:

$$\begin{cases} i^* = \operatorname{argmin}\{i : \text{Function}(S_i) = \text{bestSmell} \\ a_o = a_{i^*} \\ b_o = b_{i^*} \end{cases} \quad (6)$$

Step 7: Iterate to see if the new Smell is better than the previous one, terminate the search progress when reaching the termination criterion. Otherwise, repeat the steps 2–6 above.

The fruit fly optimization algorithm is shown in Figure 2:

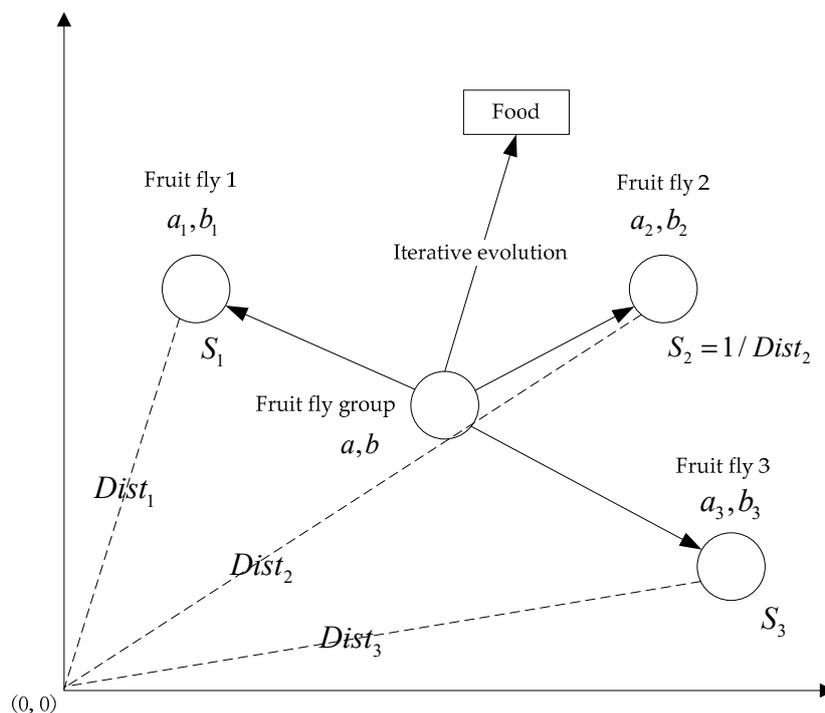


Figure 2. Fruit fly optimization algorithm.

3.3. The Hybrid Fruit Fly Optimization Algorithm-Back Propagation Model with Web Search Data

The hybrid FOA-BP model optimizes the connection weight and threshold of the neural network to improve the generalization ability and learning performance of the BP network, so as to improve the overall search efficiency of the initial BP neural network.

The optimization process is actually the change of fruit flies' position. Meanwhile the learning process of BP neural network is actually the updating process of weights and thresholds. Therefore, it can be considered that the smell concentration value of each fruit fly group corresponds to the weight value and threshold value in the BP network iteration process, the number of fruit fly population depends on the number of parameters to be optimized and the neural network output error of training samples is used as the smell concentration determination function. In the foraging process, the position change of the fruit fly can minimize the error of the network. At the end of each iteration, the fruit fly with the best smell concentration is regarded as the current globally optimal fruit fly. When the training process is repeated repeatedly until the error meeting the requirements or the number of pre-set iterations is reached, the search terminates. At this time, the set of weights and thresholds obtained are the final results.

The whole process in this study is shown in Figure 3 as follows:

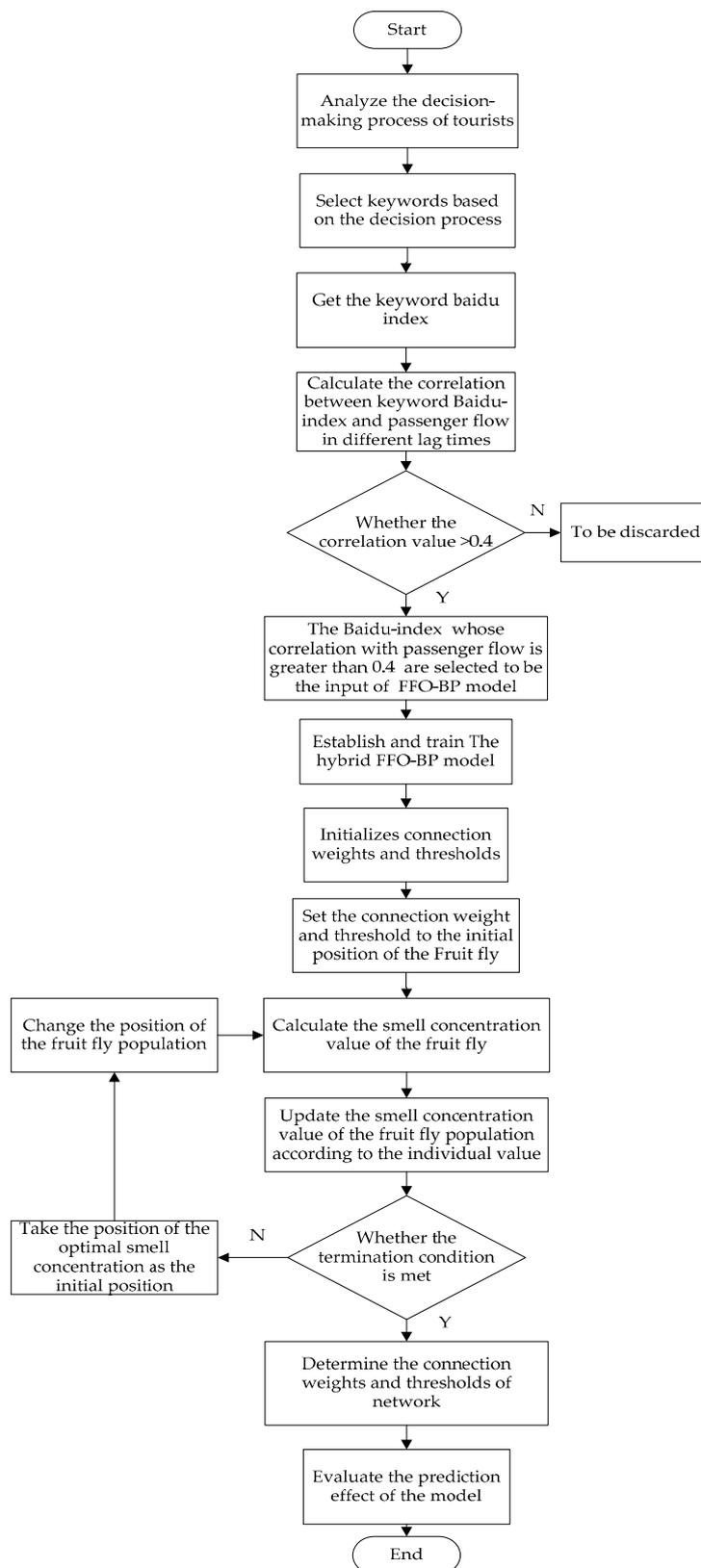


Figure 3. The flow chart.

4. Empirical Study

4.1. Data

This paper takes the Mount Huangshan scenic area as an example of a famous Chinese scenic spot. Mount Huangshan is listed as a United Nations Educational Scientific and Cultural Organization (UNESCO) world natural and cultural heritage site in 1990, and was selected as one of the first global geoparks in 2004. In 2017, there were 3.3687 million visitors from around the world.

The daily historical data selected are from 2015 to 2017, and all data were obtained from the research project in cooperation with the Huangshan scenic area and the network search index (i.e., Baidu index) from the large data in the Baidu search engine. Baidu's massive Internet behavior data are based on a data-sharing platform. The search index is based on the search volume of netizens in Baidu, and it takes keywords as the statistical object to scientifically analyze and calculate the weighted sum of the search frequency of each keyword in the Baidu search engine.

4.2. Keyword Selection

This study comprehensively used the direct method and scope method based on the tourism decision-making process and the related information that visitors would want to focus on before traveling to their destination. This study selected relevant keywords, including "destination", "destination + guide", "destination + travel guides", "destination + tickets", "destination + weather", based on the tourism destination, strategy, ticket price, scenic spots, weather, and accommodation, among many other factors. Finally, 50 initial keywords related to the decision-making process were selected. On the basis of the Huangshan scenic area, this study chose "Huangshan", "Huangshan tourism guide", "Huangshan tickets", "Huangshan guide", "Huangshan weather", and "Huangshan accommodation" from 50 initial keywords as benchmark keywords according to its correlation with passenger flow and queries the corresponding Baidu index. At the same time, a keyword-mining tool (<http://tool.chinaz.com/>) was employed to verify that the above six keywords are the top keywords. To verify the correlation between keywords and actual number of tourist flow, this study analyzed the correlation between six keywords from Baidu index and the actual number of total daily tourist flow to the Huangshan scenic spot. Table 1 shows that the correlation analysis results with a confidence level of 0.01.

Table 1. Results of correlation analysis.

Keywords	Correlation
Huangshan	0.471
Huangshan tourism guide	0.551
Huangshan weather	0.102
Huangshan guide	0.417
Huangshan accommodation	0.323
Huangshan tickets	0.449

The table reveals that the four keywords with high correlation degrees were "Huangshan", "Huangshan tourism guide", "Huangshan ticket", and "Huangshan tourism guide", with correlations of 0.471, 0.551, 0.449, and 0.417, respectively. Meanwhile, the correlation between the Baidu index of "Huangshan weather" and "Huangshan accommodation" and the total passenger flow was relatively low (less than 0.4). The four keywords with high correlations from the Baidu index were selected as input variables of the prediction model.

4.3. Data Preprocessing

The FOA-BP neural network model in this study was established using MATLAB R2016a software. To improve the prediction accuracy and stabilize the data before using the BP neural network for

training prediction, the original data sequence of the total population was normalized to [0, 1] by mapminmax function. The formula is as follows:

$$Y_t = \left(\frac{X_t - X_{\min}}{X_{\max} - X_{\min}} \right) \tag{7}$$

where X_t is the passenger flow on day t in the original one-year data series, X_{\min} and X_{\max} are the minimum and maximum values of the original sequence, respectively.

4.4. Selection of Input Variables

Data sets in this paper include actual traffic data of the Huangshan scenic spot from 2015 to 2017. To contain the characteristics of the whole data set, data from 2015–2016 were selected as the training set and the 2017 actual tourist flow data were selected as the test set. The tourist flow prediction model was built based on the FOA-BP neural network containing web search data. The past total number of people, weather, weekends, and official holidays were selected as input variables [23], and the Baidu index of relevant keywords was used as the input variable. Weather, weekend, and official holidays were added into the model as dummy variables.

(1) Daily total number of tourists in the past.

Past total number of tourists had four corresponding rules: by date, by week, by total number of tourists last week, and by total number of tourists the week before last. The correlation analysis results of past total number of tourists and target total number of tourists are shown in Table 2.

Table 2. Results of correlation analysis.

Corresponding Rules	By Date	By Week	By Total Number of Tourists Last Week	By Total Number of Tourists the Week before Last Week
Correlation	0.416	0.376	0.258	0.171

Note: “By date” means the rule is that “2017-01-01” corresponds to “2016-01-01”.

Therefore, the past total number of tourists corresponded by date is selected as the input variable X_1 .

(2) Weather.

Weather was added into the model in the form of the dummy variable X_2 :

$X_2 = \begin{cases} 1 \\ 0 \end{cases}$, 1 represents severe weather such as blizzard, heavy snow, moderate snow, heavy rain, heavy rain, thundershowers, and showers; 0 represents non-severe weather such as sunny, cloudy, and drizzle.

(3) Weekend.

Weekend was added into the model in the form of the dummy variable X_3 :

$X_3 = \begin{cases} 1 \\ 0 \end{cases}$, 1 represents weekend; 0 represents workday.

(4) Official holiday

Official holiday was added into the model in the form of the dummy variable X_4 :

$X_4 = \begin{cases} 1 \\ 0 \end{cases}$, 1 represents official holiday; 0 represents ordinary day.

(5) Baidu index of keywords

Through the previous analysis, we selected four keywords which have the highest correlation: “Huangshan”, “Huangshan tourism guide”, “Huangshan ticket”, and “Huangshan tourism strategy.”

Given the lag period between searching information on the Internet and going to travel, this study respectively analyzed the correlation between the Baidu index of the four keywords with a lag period of one day, two days, and three days to a week, and the actual total number of tourists. As shown in Table 3, the Baidu index of keyword with a lag period of two days has the highest correlation with the actual total number of tourists, and was thus selected as the input variable of the model.

Table 3. Results of the correlation analysis of different lag periods.

Lag Period	Correlation with the Total Number of People			
	Huangshan	Huangshan Tourism Guide	Huangshan Guide	Huangshan Tickets
1	0.511	0.629	0.554	0.544
2	0.518	0.638	0.606	0.560
3	0.459	0.581	0.567	0.498
4	0.404	0.512	0.496	0.432
5	0.371	0.458	0.411	0.379
6	0.309	0.408	0.351	0.304
7	0.257	0.372	0.291	0.253

4.5. Building the Model

The tourist flow in peak season (from April to October) presents strong nonlinear characteristics [23]. Compared with the traditional time series prediction model, BP neural network can deal with these complex nonlinear relations well. Usually, the three-layer structure of the BP neural network is enough to reflect the complex nonlinear relationship. Thus, the current study set up a three-layer structure of the FOA-BP neural network. Nine hidden layer nodes and one output layer node were chosen based on many experiments. Previous experiments show that using a sigmoid function as the activation function between the hidden layer and the output layer can have the good prediction effect. As a type of elastic algorithm, trainrp has the advantage of fast convergence speed and small footprint compared with functions such as trainlm, trainscg, and traingd. The trainrp function can likewise achieve better prediction effect, and was selected as the training function of the model in this study.

5. Empirical Results and Evaluation

The GA-BP model, and PSO-BP model were selected as the benchmark model in this study. The prediction result of the model mentioned above were compared with the benchmark model. The input variables of the model we proposed included the past total number of tourists, weather, weekend, official holiday, and Baidu index of keywords (from 0 to 4 keywords). The input variables of the benchmark model only included weather, weekend, and official holiday. In evaluating the prediction effect of the model, MAPE indicator was selected. The formula is as follows:

$$MAPE = \frac{1}{n} \left(\sum \left| \frac{Y - X_t}{Y} \right| \right) \times 100, t \in 1, 2, \dots, n \tag{8}$$

where Y represents the actual number, and X_t represents the predicted value.

The comparative experiments in this study were divided into two categories: the prediction results of different models including/excluding the web search data.

Table 4 shows the accuracy of different models excluding the web search data.

Tables 5–7 show the accuracy of different models including the web search data.

According to the predicted results in the tables above, GA-BP, PSO-BP and FOA-BP models including the web search data had a better accuracy than the benchmark models excluding the web search data; the average accuracy of the whole year was improved to a certain extent. Also, more keywords make the result more accurate. From the tables above it can be seen that the average accuracy from April to October was more accurate than in the whole year, especially in June, July and August. Meanwhile, there were also some limitations; the accuracy in January, February, March, November

and December was not good enough, but the results were still better than the benchmark models. One of the reasons could be that the actual value is small; usually there are only 1000 to 2500 tourists, which likely leads to a high numerical deviation.

Three models including the web search data have an approximate accuracy for the whole year, however, from the result shown in tables above, the accuracy from April to October of the FOA-BP model is better than that of the GA-BP and PSO-BP models.

The time from April to October is considered the peak season of the Huangshan scenic spot traditionally [23]. The accuracy of the predicted value in this period has more significance for the tourism management of the scenic spot because there are many more tourists in this period. The predicted result and the actual value are shown in Figure 4. Compared with January, February, March, November and December, there are more tourists in this period. As is shown in this study, the FOA-BP model including the web search data achieved the better predicted value compared to benchmark models. As a result, more advice can be provided for the management of scenic spots, such as increasing the corresponding staff to ensure the work efficiency of scenic spots to avoid the occurrence of stranded tourist events.

Table 4. Results of different models excluding the web search data.

Month	MAPE (%)		
	GA-BP	PSO-BP	FOA-BP
1	53.63	58.62	63.41
2	50.99	52.98	49.49
3	46.60	45.09	42.64
4	27.63	32.51	37.52
5	43.63	43.11	36.43
6	32.01	35.99	32.61
7	22.73	20.69	19.23
8	16.83	22.18	18.29
9	31.61	26.93	29.31
10	30.34	28.98	26.10
11	39.00	33.54	36.30
12	57.18	59.43	61.07
The average from April to October	29.23	29.93	28.58
The average for the whole year	37.57	38.25	37.61

Table 5. Results of GA-BP model including the web search data with 1–4 keywords as input.

GA-BP	MAPE (%)			
	One Keyword	Two Keywords	Three Keywords	Four Keywords
1	58.74	45.78	53.94	53.06
2	40.46	33.93	36.54	47.37
3	41.24	41.04	40.12	36.74
4	32.96	26.47	27.64	30.75
5	38.63	35.89	34.30	40.13
6	30.41	30.88	28.06	32.49
7	23.02	23.59	20.68	18.67
8	16.34	17.39	15.49	14.85
9	41.06	34.07	37.21	28.74
10	34.75	28.04	31.45	29.67
11	28.55	25.30	24.92	26.98
12	51.54	32.45	44.47	44.59
The average from April to October	30.91	28.08	27.85	27.84
The average for the whole year	36.36	31.19	32.81	33.47

Table 6. Results of PSO-BP model including the web search data with 1–4 keywords as input.

PSO-BP	MAPE (%)			
	One Keyword	Two Keywords	Three Keywords	Four Keywords
1	54.29	58.31	63.13	43.36
2	40.67	39.47	37.68	50.33
3	46.53	39.31	37.86	37.54
4	31.58	32.96	31.35	30.35
5	36.50	35.75	36.64	34.44
6	35.22	30.59	24.43	25.51
7	18.98	19.16	13.91	15.95
8	26.94	16.74	24.56	20.42
9	34.88	33.52	41.86	36.81
10	30.32	29.89	27.84	26.36
11	30.10	43.27	30.74	31.76
12	43.96	46.77	55.70	48.98
The average from April to October	31.44	28.21	28.71	27.06
The average for the whole year	36.32	35.35	35.46	33.39

Table 7. Results of FOA-BP model including the web search data with 1–4 keywords as input.

FOA-BP	MAPE (%)			
	One Keyword	Two Keywords	Three Keywords	Four Keywords
1	58.66	65.00	65.62	72.91
2	42.04	40.59	39.05	36.40
3	42.12	36.50	36.16	36.72
4	34.03	29.35	30.43	29.57
5	41.84	32.51	33.67	33.90
6	30.78	23.98	22.72	24.59
7	19.58	16.24	13.50	11.80
8	16.32	19.39	16.26	16.42
9	32.04	34.78	39.36	35.71
10	28.48	29.65	29.64	27.49
11	29.11	26.93	28.30	26.00
12	58.25	53.15	52.34	52.14
The average from April to October	28.90	26.45	26.48	25.60
The average for the whole year	36.03	33.94	33.81	33.55

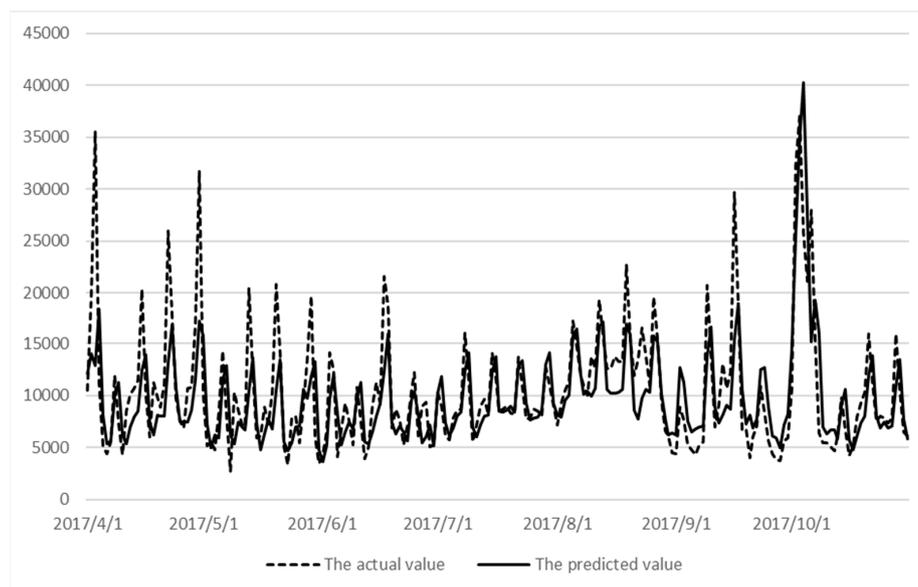


Figure 4. The actual and predicted value.

In general, the hybrid FOA-BP method on daily tourism demand forecasting with web search data proposed in this study can effectively improve the tourist flow prediction accuracy in peak season. Furthermore, the hybrid method proposed was better than other benchmark models with regard to the peak season, which proves the validity of the method in short-term daily tourism demand forecasting.

6. Conclusions and Implications

The rapid development of tourism in recent years has become an important part of the Chinese economy. Thus, the problem of tourism management has become more and more pressing. The prediction of tourist flow, especially short-term tourist flow during peak season, is crucial for tourism management departments. The management department needs to effectively predict future tourism demand, so as to maintain the sustainable development of the scenic spot and avoid damage caused by excessive tourists. A hybrid FOA-BP model is established in this study, which is proved to obtain a more accurate prediction compared with other intelligent algorithms when used in short-term daily tourism demand forecasting. The hybrid model can effectively help the management department to carry on the sustainable management to the scenic spot. Furthermore, taking the famous tourist destination Huangshan scenic spot as an example, this study discusses the application of the Internet search index in the forecast of short-term tourist flow. Moreover, it establishes a model combined with the Baidu index to predict short-term tourist flow. In the selection of Internet search keywords, the benchmark keywords are selected according to the characteristics of the research objects. The benchmark keywords should be reasonable, operable and as comprehensive and accurate as possible. Combining the direct method and the scope method, this study selects the keywords that are related to the destination tourist attractions and have a high search volume, and from which the keywords with a high correlation degree are selected. Considering the lag period between online search and travel, the network search index of the lag period with the highest correlation between related keywords and total number of people is selected through the correlation degree analysis. Experimental results show that the network search index can greatly improve the prediction effect of the original model and is more effective than the benchmark model. However, there are some limitations in the experiment which deserve further study, such as a more accurate keyword selection method and other application methods of web search data in tourism demand forecasting. Generally speaking, the hybrid FOA-BP method proposed in this paper provides a new view for short-term tourism demand forecasting. The proposed method has good prospects in research and application for tourism management, from which the tourism industry can be healthy and sustainable.

Author Contributions: Data curation, K.L.; formal analysis, K.L. and W.L.; methodology, K.L.; supervision, W.L., C.L. and B.W.; writing—original draft, K.L.

Funding: This research received no external funding.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (NSFC) (71331002, 71771075, 71771077, 71601061) and supported by “the Fundamental Research Funds for the Central Universities” (PA2019GDQT0005).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bureau, A.S. *Anhui Statistical Yearbook*; China Statistics Press: Beijing, China, 2017.
2. Smeral, E.; Witt, S.F.; Witt, C.A. Econometric forecasts: Tourism trends to 2000. *Ann. Tour. Res.* **1992**, *19*, 450–466. [[CrossRef](#)]
3. Rostan, P.; Rostan, A. The versatility of spectrum analysis for forecasting financial time series. *J. Forecast.* **2017**, *37*, 327–339. [[CrossRef](#)]
4. Shang, H.L. Forecasting intraday S&P 500 index returns: A functional time series approach. *J. Forecast.* **2017**, *36*, 741–755.

5. Song, H.; Li, G. Tourism demand modelling and forecasting—A review of recent research. *Tour. Manag.* **2008**, *29*, 203–220. [[CrossRef](#)]
6. Sun, Y.; Lv, B. A Review of Researches on the Correlation between Internet Search and Economic Behavior. *Manag. Rev.* **2011**, *23*, 72–77.
7. Wong, K.K.F.; Song, H.; Chon, K.S. Bayesian models for tourism demand forecasting. *Tour. Manag.* **2006**, *27*, 773–780. [[CrossRef](#)]
8. Song, H.Y.; Witt, S.F. Forecasting international tourist flows to Macau. *Tour. Manag.* **2006**, *27*, 214–224. [[CrossRef](#)]
9. Gil-Alana, L.A.; Cunado, J.; Gracia, F.P.D. Tourism in the Canary Islands: Forecasting using several seasonal time series models. *J. Forecast.* **2008**, *27*, 621–636. [[CrossRef](#)]
10. Palmer, A.; Montañó, J.J.; Sesé, A. Designing an artificial neural network for forecasting tourism time series. *Tour. Manag.* **2006**, *27*, 781–790. [[CrossRef](#)]
11. Hong, W.C. *The Application of Support Vector Machines to Forecast Tourist Arrivals in Barbados: An Empirical Study*; Social Science Electronic Publishing: Rochester, NY, USA, 2006; Volume 23.
12. Song, G.F.; Liang, C.Y.; Liang, Y.; Zhao, S.P.; Management, S.O. Prediction for Tourist Attractions Daily Traffic Based on Modified Genetic Algorithm Optimized BP Neural Network. *J. Chin. Comput. Syst.* **2014**, *20*, 232–238.
13. Huang, H.-C.; Hou, C.I. Tourism Demand Forecasting Model Using Neural Network. *International. J. Comput. Sci. Inf. Technol.* **2017**, *9*, 19–29. [[CrossRef](#)]
14. Song, W.L.; Lee, H. The Study on the Tourism Demand Characteristics and Forecasting of Jeju Island. *Tour. Res.* **2018**, *43*, 111–124. [[CrossRef](#)]
15. Hu, Y.-C.; Jiang, P.; Lee, P.-C. Forecasting tourism demand by incorporating neural networks into Grey–Markov models. *J. Oper. Res. Soc.* **2018**, *70*, 12–20. [[CrossRef](#)]
16. Zong, C.-L.; Wang, L. Prediction of urban residents’ travel rate in China based on ARIMA models. *J. Interdiscip. Math.* **2018**, *21*, 1285–1290. [[CrossRef](#)]
17. Ginsberg, J.; Mohebbi, M.H.; Patel, R.S.; Brammer, L.; Smolinski, M.S.; Brilliant, L. Detecting influenza epidemics using search engine query data. *Nature* **2009**, *457*, 1012. [[CrossRef](#)] [[PubMed](#)]
18. Wang, L.; Jia, J.M. Forecasting box office performance based on online search: Evidence from Chinese movie industry. *Syst. Eng.-Theory Pract.* **2014**, *34*, 3079–3090.
19. Vosen, S.; Schmidt, T. Forecasting private consumption: Survey-based indicators vs. Google trends. *J. Forecast.* **2011**, *30*, 565–578. [[CrossRef](#)]
20. Francesco, D.A. Predicting unemployment in short samples with internet job search query data. *MPRA Pap.* **2009**, *6*, 1–17.
21. Dimpfl, T.; Jank, S. Can Internet Search Queries Help to Predict Stock Market Volatility? *Eur. Financ. Manag.* **2016**, *22*, 171–192. [[CrossRef](#)]
22. Xin, Y.; Pan, B.; Evans, J.A.; Lv, B. Forecasting Chinese tourist volume with search engine data. *Tour. Manag.* **2015**, *46*, 386–397.
23. Liang, C.Y.; Ma, Y.C.; Chen, R.; Liang, Y. The Daily Forecasting Tourism Demand Based on SVR-ARMA Combination Model. *J. Ind. Eng. Eng. Manag.* **2015**, *29*, 122–127.
24. Niazi, M.A. In Memoriam: John Henry Holland—A pioneer of complex adaptive systems research. *Complex Adapt. Syst. Model.* **2015**, *3*, 1–2.
25. Shi, Y.; Eberhart, R.C. Empirical study of particle swarm optimization. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), Washington, DC, USA, 6–9 July 1999; Volume 1943, pp. 1945–1950.
26. Pan, W.T. *Fruit Fly Optimization Algorithm*; Tsang Hai Book Publishing: Taipei, China, 2011.
27. Lin, S.; Dong, C.; Chen, M.; Zhang, F.; Chen, J. Summary of new group intelligent optimization algorithms. *Comput. Eng. Appl.* **2018**, *54*, 1–9.
28. Du, T.S.; Ke, X.T.; Liao, J.G.; Shen, Y.J. DSLC-FOA: Improved fruit fly optimization algorithm for application to structural engineering design optimization problems. *Appl. Math. Model.* **2018**, *55*, 314–339. [[CrossRef](#)]

29. Wu, L. Research on Wireless Sensor Network Layout Based on Improved Fruit Fly Optimization Algorithm. *Microelectron. Comput.* **2016**, *33*, 152–156.
30. Wang, L.; Zheng, X.L. A knowledge-guided multi-objective fruit fly optimization algorithm for the multi-skill resource constrained project scheduling problem. *Swarm Evolut. Comput.* **2018**, *38*, 54–63. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).