Simulation analysis of the prediction model for tourist number in tourist season

Wei Lingli

Sichuan Tourism University, Chengdu, Sichuan 610100, China

Abstract. The investigation is about accurately predicting the tourist number in tourist season. Factors that affecting people's travel demand is currently more and more, and the impact on the tourist number in tourist season is increasingly complex. The traditional prediction models of tourist number in tourist season are mostly static model, with the property weight of to each determine factors are relatively fixed, cannot be flexible changed. But these attributes influence the final number of visitors differently, which led to the traditional model is difficult to form an accurate judgment. In order to avoid these shortcomings, the analysis method based on big data presented in the tourist number prediction in tourist season. Relevant data are collected, to make pre-processing, including data cleaning, data transformation and data integration. Based on the theory of support vector machine, the data attributes are classified, the prediction model of number of visitors in the tourist season is established. By calculating the weight of model attribute classification results, the classification of importance degree of tourist number affected by different properties is obtained. By comparison with the parameters, the forecast results of tourist number in tourist season are obtained. The results show that using the proposed algorithm in this article, the prediction of the tourist number in tourist season can be effectively improved the accuracy, thus providing an accurate basis for decision-making tourism industry and promoting the development of tourism industry more healthy, rapid and sustainable.

Introduction

With the continuous improvement of people's living standards and the continuous development of economy, people in the pursuit of spiritual life is higher and higher [1]. In recent years, people's travel demand also shows a significant increasing trend [2]. In the tourism industry, the prediction of tourist number in tourist season is an important basis for the tourism decision-making [3]. According to the predicting results of tourist number in tourist season, the tour personnel can be provided adequate preparation to ensure quality of tourism and full use of tourism resources for tourism personnel. So it has a very important economic significance [4].

When the domestic tourism industry is still in its infancy, it has been carried out the research of the prediction the number of visitors in tourist season [5]. At this stage, due to the complexity of tourism data can be quickly and accurately analyzed on a computer, which provides powerful protection on the number of tourists in the tourist season [6]. Therefore, the methods to predict the number of tourists in the tourist season have become a major issue to be studied in the tourism industry, and pay more and more attention to scholars. At this stage, the main methods to prediction number of visitors in tourist season are including the prediction method based on data mining algorithms, the prediction method based on gray neural network algorithm and the prediction method based on the static predictive model [7]. One of the most commonly used method is static prediction model. Because the forecasting of the number of visitors in tourist season has a very wide range of applications, therefore subject to the attention by many experts, has become a hot research topic, and with a broad development space [8].

Factors that affect people's travel demand is currently more and more, and the impact on the amount of visitors during tourist season is increasingly complex. The traditional prediction models of tourist number in tourist season are mostly static model, with the property weight of to each determine factors are relatively fixed, cannot be flexible changed. But these attributes influence the

final number of visitors differently, which led to the traditional model is difficult to form an accurate judgment.

In order to avoid these shortcomings, the analysis method based on big data presented in the tourist number prediction in tourist season. Collect relevant data, and pre-processing the collected data, including data cleaning, data transformation and data integration. Based on the theory of support vector machine, the data attributes are classified, the prediction model of number of visitors in the tourist season is established. By calculating the weight of model attribute classification results, the classification of importance degree of tourist number affected by different properties is obtained. By comparison with the parameters, the forecast results of tourist number in tourist season are obtained. The results show that using the proposed algorithm in this article, the prediction of the tourist number in tourist season can be effectively improved the accuracy, thus providing an accurate basis for decision-making tourism industry and promoting the development of tourism industry more healthy, rapid and sustainable.

The optimization method to predict the number of visitors in tourist season

Data pre-processing. Collect relevant data, the database of tourist information in tourist season is established. And the collected data is pre-processed by the following method:

(1) Data cleaning. Inevitably, there will be a large number noise which are irrelevant to the prediction of number of visitors, as well as incomplete information in collected original information. Therefore, data cleansing is needed.

(2) Data integration. During the forecasting process of tourist number, reference values of predictors and related properties need to be identified, and the desired data of attributes need to be selected from database.

(3) Data conversion. Converts the data of a non-Boolean attribute into Boolean data attribute, and enter the data into a new database. And the data was processed by discretization.

By processing the data by the method described above, the data analysis can be more convenient, to provide data support for the prediction of the tourist number in tourist season.

Prediction model establishment. According to the theory of support vector machine, large data was processed for analysis and classification. The prediction model establishment of number of visitors in tourist season was achieved.

It is known that the support vector machine penalty coefficient c determines the degree of compromise between the largest classification interval and the minimum training error. If the number of positive type in the results of data type classification are far less than the number of negative type, then the sum of error of the positive type is small and the sum of error of the negative type is large. That is the penalty imposed on negative type is large, which can easily result in misclassification. Therefore, using the most optimal conditions can be obtained:

$$\alpha_i(y_i(\omega \cdot \phi(x_i) + b) - 1 + \xi) = 0$$
$$\beta_i \xi_i = (C - \alpha_i)\xi_i = 0 \tag{1}$$

When classification result of sample data x_i is correct, then $\alpha_i = 0, \xi = 0$. Assuming that $0 < \alpha_i < C$, then:

$$y_i(\omega \cdot \phi(x_i) + b) - 1 + \xi_i = 0, \xi_i = 0 \tag{2}$$

That shows x_i is on classification interval surface. When $\alpha_i = C$, then:

$$y_i(\omega \cdot \phi(x_i) + b) - 1 + \xi_i = 0, \quad \xi_i \geq 0 \tag{3}$$

With equation above, we can obtain, when $\xi_i \geq 0$, x_i is boundary support vector, the contributions to classification surface are largest; when $0 \leq \xi_i \leq 1$, x_i is between interval surface and hyperplane, the classification is correct, but the importance degree of data concentration is less; when $\xi_i \geq 1$, x_i is error-classifying samples. It would seem:

$$\sum_{i=1}^{N} \alpha_i = \sum_{y_i = +1} \alpha_i = \sum_{y_i = -1} \alpha_i = A, \sum_{y_i = +1} \alpha_i = \sum_{y_i = -1} \alpha_i \tag{4}$$

On the issues of data imbalance in data concentration sample, the problem can be solved by following equation:

$$\min \tau(W,\xi)\frac{1}{2}W^2 + C\lambda\sum_{i=1}^{l}s_i\xi_i$$
$$s.t. \qquad y_i((W\cdot x_i)+b)+\xi_j \geq 1 \tag{5}$$
$$\xi_i \geq 0, i=1,2...,m$$

Among the equation, the weight of sample class y_i is $\lambda \geq 1$, the weight of sample x_i is $s_i > 0$, when $0 < s_i < 1$, it shows that the effect sample data is not important in prediction. When $s_i = 1$, it shows that the effect sample data is general important in prediction. When $= s_i > 1$, it shows that the effect sample data is very important in prediction. According to the principle of support vector machine, the weighted dual equation was obtained:

$$\max L(\alpha) = \sum_{j=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$
$$s.t. \quad 0 \leq \alpha_i \leq C\lambda, \sum_{j=1}^{l}\alpha_i y_i = 0, i=1,2,...m \tag{6}$$

The corresponding discrimination function $f(x)$ can be described by following equation:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{m}\alpha_i y_i K(x_i x)+b\right) \tag{7}$$

By the above method, the model for predicting the tourist number in tourist season can be established.

Weight calculating. Assuming the training set denoted by $S = \{(x_1,y_1,s_1),...,(x_m,y_m,s_m)\}$, and to meet $x_1 \in R^n, y_1 \in \{-1,1\}, \sigma \leq s_i \leq 1$, if weight training points one right index $y_i = 1/-1$ is s_i, y_i means that the degree of importance in predicting the course of the role s_i. $s_i\xi_i$ is used to represent wrong error loss of data points x_i.

In the weight calculation process, individual positive and negative density need to be defined. Density is same category with sample of is defined as positive density $\rho(x_i)^+$, and the density of the sample is different with x_i is negative density $\rho(x_i)^-$, the average sample density for the same category $\rho^+ = \frac{1}{m^+}\sum_{i=1}^{m}\rho(x_i)^+$, if the values of the positive sample density and the negative density are very small, then the sample is isolated great probability point; but if the positive density is small and negative density samples are large, the probability of this sample is a great noise, on the contrary, if the density of the samples are large and negative density is small, it means that this sample is normal samples. Therefore, it can calculate the position of the sample in the collection and importance.

Setting individual sample points density function as the normal distribution density function $p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp^{\frac{1}{2\sigma^2}(x-\mu)^2}$, then:

(1) Support the right of the point value as 1;
(2) In addition to the support point of the value of the right to a normal sample interval $(0,1)$;
(3) The weights noise and isolated points are very small or zero;
Thus, the weighted index can be calculated with following:

$$s_i = \frac{\rho(x_i)^+}{\rho(x_i)^+ + \rho(x_i)^-} \cdot \frac{\rho(x_i)^+}{\rho^+} \tag{8}$$

By the method above, it can complete screening valuable data. By comparison with the standard parameters to obtain the number of visitors forecast results in tourist season.

Experimental results

This paper selects the past three years of visitor numbers information in tourist season as a basis for simulation data. Select 10 tourism landscape as the research object. The basic data can be described by following table.

Table 1 the visitor number data

Travel Resources	2011（in thousand）	2012（in thousand）	2013（in thousand）
Emeishan	340000	420000	510000
Qianshan	270000	330000	420000
Dujiangyan	330000	360000	410000
Terracotta	190000	230000	270000
Forbidden City	1070000	1170000	1310000
Summer Palace	880000	980000	1080000
Pagoda	920000	1040000	1160000
Westlake	1490000	1650000	1760000
Guilin	540000	680000	880000
Lijiang	670000	720000	890000

In this paper, the traditional algorithms and our algorithms are used respectively to predict the number of visitors of 2014 tourist season. The results obtained can be described by following table.

Table 2 Comparison of predict results with different algorithms

Travel Resources	The accuracy of traditional prediction algorithms (%)	The accuracy of the prediction algorithm (%)
Emeishan	88	96
Qianshan	87	97
Dujiangyan	88	95
Terracotta	85	97
Forbidden City	87	94
Summer Palace	84	96
Pagoda	86	95
Westlake	88	97
Guilin	87	97
Lijiang	85	94

Using the proposed algorithm to predict the number of visitors in the tourist season can improve forecast accuracy, and achieved satisfactory results.

Conclusions

This paper presents the number of visitor prediction method in tourist season based on analysis of large data. Collect relevant data, and pre-processing the collected data, including data cleaning, data transformation and data integration, based on support vector machine theory, the classification of the data attributes, prediction model of the number of visitors in tourist season is establish. Computational model attribute classification results weights, it is important to get the degree of impact on the number of visitors to the different attributes of classification, by comparison with the parameters to get the number of visitors forecast results in tourist season. Experimental results show that using the proposed algorithm for the number of visitors forecast in tourist season, which can effectively improve the accuracy of prediction, thus providing an accurate basis for decision-

making tourism industry, to promote healthy, rapid and sustainable development of tourism industry.

References

[1] Liu Jing, Guan Wei. Review on traffic flow forecasting methods [J]. Highway traffic science and technology, 2004, 21(3):82-85.

[2] Shang Ning, Qin Minggui, Wang Yaqin. traffic flow forecasting in a short time in the intersection based on BP neural network [J]. Computer Applications and Software, 2006, 23(2): 32-33.

[3] F.A. Oliveira, C.N. Nobre, L.E. Zárate. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index - Case study of PETR4, Petrobras, Brazil [J]. Expert Systems with Applications, 2013, 40(18), 7596–7606.

[4] S. Aghabozorgi, Y.W. Teh. Stock market co-movement assessment using a three-phase clustering method [J].Expert Systems with Applications, 2013, 41(4), 1301–1314.

[5] Y. Zuo, E. Kita, Stock price forecast using Bayesian network [J]. Expert Systems with Applications, 2012, 39(8), 6729–6737.

[6] Cao Xingfeng. Application research of a neural network in weather prognosis [J]. Bulletin of Science and Technology, 2012, 8(28):56-57.

[7] Ying Shangjun, Fan Ying, Wei Yiming, Wang Binghong. Complexity study of stock market based on cellular automaton—investor psychology and market behaviour [J]. System Engineering Theory and Practice, 2003, (12):19-24.

[8] Wang Huimin, Liu Guoguang. Dynamic correlation analysis of stock yield and volume of business changs [J]. Mathematics in Practice and Theory, 2006, 36(9):70-75.