AN ANALYSIS AND FORECAST OF TOURISM DEMAND BASED ON WEB DATA MINING IN CHINA

[1,2]GUOJING FAN, [3]ZHIJUN LIU

[1]Hunan Agriculture University, 410128, Changsha, China

[2]gannan Normal University, 341000, Ganzhou, China

[3]zhejiang University, 310058, Hangzhou, China

E-mail: [1]fguojing@163.com , [3]liuzhijun@zju.edu.cn

ABSTRACT

This article explored into the advantages of Web data application in forecasting tourism demand in China. The authors constructed a mapping relation between Web information organization and tourism demand, and elaborated the methods of using Web data mining Carries to analyze and forecast tourism demands and the principles of constructing the knowledge library based on Web data analysis with clarified procedures of analysis and forecast. The authors also did an empirical analysis by taking Shanghai city as a case in point.

Keywords: *Web, Data, Tourism, Forecasting, Knowledge Library*

1. INTRODUCTION

Tourist destination, scenic spots and tourism enterprise not only need to understand the tourists' reality demand, also they need to analysis and grasp of the tourists' potential demand. According to the immovability of tourism products, tourist destination mainly provide the tourists the reception service, while the operator is often in the visitors arrived through the tourist survey to collect the tourist information, so that, to a certain extent, we can understand the practical needs of tourists, but it is difficult to master potential tourists related material and potential demand characteristic, and to go against the development of new market, to develop tourist market range. And the Internet for travel demand research provides a new way [1]-[3]. This article ,through to travel Web information source of deep analysis, which is based on Web data of travel demand analysis and forecast mechanism, provides a new travel demand analysis and prediction method, in order to afford better and faster service for tourism destination[4].

2. TOURIST INFORMATION ANALYSIS APPROACH

Visitors get tourism information source to basically have two kinds: internal information source and external information source. The Internal information source is mainly refers to the information which stored in long-term memory of tourists, closely related to the relevant information about tourists personal experience, education, knowledge. It is the original information for tourist to make a decision. The External information source is which refers to the information that the user cannot get from internal information source to obtain a more effective and sufficient information and consumption direction .Our research is mainly about the external information source that it is available in a variety of forms. What's more, (particularly the Internet) under the environment of online information; search has become an important channel for people to search information and a research hot spot to study [5].

Along with the development of information technology and economic level unceasing enhancement, the Internet not only has been used by more and more ordinary people, but also becomes an important means for the tourists to make travel information in tourism auxiliary decision-making. The study found that in the process of using the Internet, most of the visitors will not directly reserve in the Internet, but they will be through the network search tourist information, then according to the online information combination, stroke after through the travel agency for reservation. CCTV investigation and consultation center of Beijing made by young Internet users "Internet users tourism consumption survey results show that 21.8% of Internet users

have logged on a tourist site, 43.2% of the netizens said they will be logged in tourism website; Among those who had logged in tourism website of netizens, 3.8% of them have already had and travel related Internet booking, 54.2% of the netizens said will try to make Internet booking. Cen ChengDe (2007) ,who had made a study about our country young tourists network information search behavior ,found that about 62.6% of college students have got travel information through the "related web" obtain, in university students obtain travel information in various ways in the first. Visible Internet in tourist information acquisition in the position gradually improves [6].

3. THE WEB DATA APPLIED TO TOURISM MARKET ANALYSIS AND FORECAST ADVANTAGE

The unique advantages of internet make it a main method for market research and marketing enlarge, and is widely applied in many industries. Using Internet to travel demand survey has large amount of information for collection, and the speed is quick, and convenient as well as saving money, and many other advantages. Compared to the traditional market research and network research, the advantage of the Web data mining is obviously. At present, the Internet is a huge, distribution of a wide range of global information service center, including news, advertisement, consumer information, etc all kinds of tourist information. Not only that, Web also contains rich structured data, such as dynamic change information and links to Web page access and use of information, and so on, this is Web data mining provides abundant resources. Potential customers to interested tourist information click on the record can be used as a potential source of tourists tourism to the consumption of the analysis of the characteristics of important data basis, these records can be displayed and where the potential customers (who) to any destination (where) when (when) to click. From the tourist market research perspective, compared to the traditional information source, the use of Web data as information source has very obvious advantages (see Table 1).

Table 1: The Web Data And Market Research Data Comparison

Data acquisition ways	Data acquisition cost	Data number	Cycle	Tourist	Reliability	Extensive	Feedback
The web data	Low cost	In large quantities	Long	Active	Need to be validated	Content widely	In time
The market research dat	High cost	Investment related	Short	Passive	High reliability	Not widely	Slow

Data source: guo wei(2011).

4. TRAVEL WEB INFORMATION ORGANIZATION AND THE ESTABLISHMENT OF THE TOURISM MARKET MAPPING FRAMEWORK

Although travel Web information source it has many advantages, there are still some problems, such as numerous tourism information, fast growth speed and jumbled in content is difficult to extract, etc., to obtain travel demand, Web raw data is difficult to analyze ,and is more difficult to analysis and track the process the tourism market and potential tourists demand. Therefore, we should make full use of Web data source in tourism market analysis and forecast, according to the research target and the demand to set the mapping relationship between the tourism markets and obtain travel Web data. Thus we can structure data mode from Web information source for tourist market and tourist analysis [7].

At present, tourism Web information source common uses tree data structure. Figure 2 is the diagram of tourist Web information source and tourism market mapping relationship. In figure left users access to Web connection Web information source (server), the links are "destination", "product class" and "tourism product information". Corresponding tourist market and tourist travel demand analysis, we should establish two kinds of Web data organization model with purpose shall form regional tourism market and products (the) level of the tourism market.

Tourist destination have one or more tourism products that has rich degree, awareness, scale, concentration, the competition, and the existing market conditions constituting the tourism destination market analysis data set; The characteristics of Tourism enterprises (the) visibility and price ,and so on are constituting the tourism market analysis data set. These two kinds of data are each relatively independent, but internal data have crossed. These two data set form the framework of the expansion tourist market and tourist demand analysis organization [8].

5. BASED ON WEB DATA MINING TOURIST MARKET ANALYSIS AND FORECAST

（1）Based on Web data tourist market analysis and forecast knowledge base construction.

In Web data and the mapping relationship tourism market as the foundation, can be conducted based on Web data mining travel demand analysis, including tourists demand based on the destination and the scenic area tourist market analysis and forecast. The overall analysis and prediction of destination market is to study the market structure and development trend of the simulation study, which can help us to grasp the destination tourist market structure and the development tendency; the product level market analysis is from the perspective of tourism products micro destination tourism development trend and market structure analysis. According to the travel demand should be the analysis of the affecting factors, destination tourism product characteristics and life cycle to build tourist market analysis and forecast knowledge base, its organization is divided into two levels, namely facing the whole Web data and facing to tourism products. Facing Web information source of knowledge base will organize and classify Web data, formed tourist market analysis and forecast of the whole frame, For tourism product knowledge base content, from the point of view of tourism demand, according to the development of tourism demand situation, using the application of tourism product life cycle theory, this paper analyzes tourists travel demand concerns the focus and hotspot [9].

（2）The analysis rule of Web Information source.

In the Web Tourism Information source, the characteristics of Web information can display the user the focus of the destination products. Such as the awareness of the products in public, the value and function of tourism, etc. Based on the analysis of the characteristics of unit information, we can determine the subset of the related PKD (Product Knowledge Division), and according to the related subset of PKD, it makes the following content analysis in element information more effective.

The analysis of the characteristics of Web information element involves the requirements status and the life cycle of tourism products. According to the constructing way of analysis of the knowledge library, we define the following rules to determine the destination market of tourist and the state of the travel demand.

The overall analysis of the tourist destination market relies on tourism products. Tourists to the destination attention can reflect by the whole tourism product click. Based on the statistical analysis of the characteristics of tourism product information, we can understand the whole situation of tourism destination.

The demanding analysis of the products depends on the choice of PKD. The efficiency and accuracy of travel demanding analysis and the choice of the PKD have a lot of dependence, so the determine and perfecting of PKD subset has a great meaning. First, determine the subset of the initial state. With the increase of quantity of Web data and the continuous adjustment of methods, and the feedback of analysis results, we can use cluster analysis methods such as PKD gradually perfecting, so as to provide a reference for travel demand analysis.

（3）Based on the tourism market analysis and forecast process Web data.

Choose the suitable tourist information server for analysis. Save the users' accessing logs to the local database (Web server) after redundant processing through the download filter. Analysis and process the Web log data by the use of Web data mining method. It is the core of Web data disposal to analysis the base support analysis process of knowledge.

Firstly, locate the information of products and determine the PKD by using the based knowledge, which face the source of information. Then take it as a information content for analyzing. Here are two mainly ways: one is to click the rate statistics, the other is interactively process this information content (tourism BBS) processing. It is the main content of Web data analysis according to statistics processing of the frequency, clicking on various travel information by visitors. The statistical result is foundation of the analysis of a tourist market trend, the requirement of tourism product, and the overall tourist market. Interactive analysis method requires users to judge classification information by

people which has structured, and the product of a hand make conclusive summary, to help market researchers and decision-making departments adjust product function or marketing strategy, adapt to the travel demand and the change of the market. And then to the destination tourist market field survey data for reference, the use of statistics, probability theory and econometrics, principle and method of the statistics and model analysis, so as to find out the relationship between the two kinds of data and law. According to the situational results of tourism market, compared with the specific condition of the realistic tourism market, and the advantages and disadvantages of assessed method, we will corrected the model until we finally find the best method, then output the corresponding results.

（4） Analysis and prediction of the Shanghai travel demand.

According to the paper Web informational source, it is known that choosing appropriate data is very important. The relevant network provides Shanghai more than 30 scenic tourist information; it is the typical tourism information source. Therefore in the study, we choose to cooperate with the relevant network. We acquire travel Web data in Shanghai by using the relevant network platform .The research data records the collection of 3.04 million Web data in January, 2004 to December, 2005.

First, the analysis of Shanghai travel demanding season. We can get the seasonal change of travel Web record data in Shanghai; according to the data (see Figure 1). Its longitudinal axis said the monthly used proportion in all the year round produced, we can see that there are four obvious peaks in 2004 and April 2005 and September, in other words, April and September are tourist season Shanghai. Meanwhile, it is the preparatory stag for travel information on "May Day" and National golden week. According to the general rule of tourism decision-making collection of tourism information is before traveling, so the Web data are basically in agreement with the Shanghai tourism rule. Averagely moving the Web data, it can be seen the difference about a month, from the figure 4 Web data and the tourism actual demand in Shanghai, that is the Web data is a month or so earlier than actual demand. Web data peak appears in every April and September, and obviously Shanghai tourist season is in every two phases: April to May and September to October.
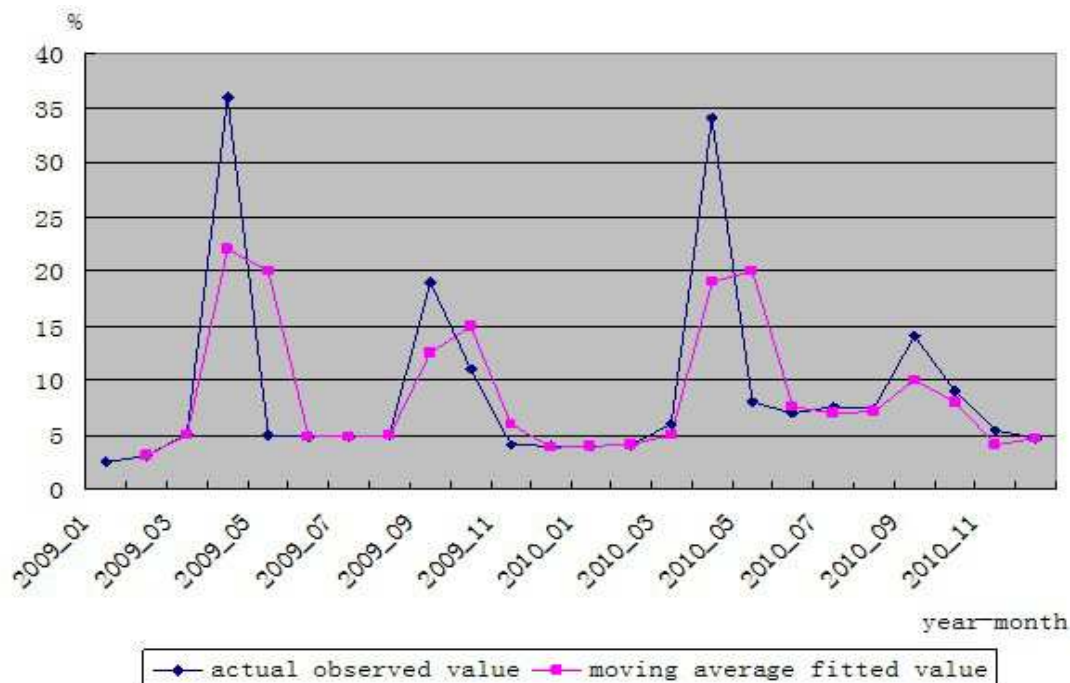

Figure 1: 2009.01-2010.12 the Web date change

Second, the forecast of Shanghai tourism demand. According to the actual situation of the domestic tourists, We select several indexes as the main factors ,which influence domestic travel demand of Shanghai, such as tourism source selection population (province, autonomous region, municipality directly under the central government) (R), the per capita GDP (S), the per capita tourism consumption of the source visitors in Shanghai (L),the ticket price from the tourism source capital city to Shanghai (F).and take other tourism source travel Web record (P) and the penetration of Internet users (W) as a correction, the establishment of the fixed travel demand gravity model:

$$T_{it+1} = G\,\frac{R_{it}^{a}S_{it}^{\beta}}{(L_{it}+F_{it})^{\gamma}}\,M_{it}^{\lambda} \tag{1}$$

T_{t+1} : the total domestic tourists of Shanghai in t + 1 year (ten thousand people), among them:

$$T_{t+1} = \sum_{i=1}^{n} T_{it+1}$$

T_{it+1} : the number of tourists from i province to Shanghai (ten thousand people);

R_{it} : the population of i province in t year(ten thousand people);

S_{it} : the per capita GDP of i province in t year (yuan);

M_{it} :the travel preference degree of i province in t year .

$$M_{it} = {P_{it}}\big/{W_{it}},\ P_{it}:$$

The rate of clicking in i province in t year (the proportion of the province clicks accounted for the Shanghai general clicks); for i province t annual network penetration (the number of Internet users in the province accounted for the proportion of the population);

W_{it} : the average consumption of visitors in i province t annual in Shanghai (yuan);

: F_{it} the air ticket price in i province t annual to Shanghai (provinces city, autonomous regions, municipalities directly under the central government capital to Shanghai ticket price, the unit: Yuan);

: $G \,、\alpha\,、\beta\,、\rho\,、\gamma\,、\lambda$ a parameter.

Using multiple regression modeling method, the type one take logarithmic available type 2:

$$\lg T_{it+1} = \lg G + \alpha\,\lg R_{it} + \beta\,\lg P_{it} + \rho\,\lg P_{it} - \gamma\,\lg(L_{it}+F_{it}) - \lambda\,\lg W_{it} \tag{2}$$

Using multiple regression method, get equation corresponding parameters. Parameter estimation value into the gravity model, and final model for type 3:

$$T_{it+1}' = 6.56 + 0.148 R_{it}' + 0.318 (P_{it}/W_{it})' + 2.305 (L+F)'_{it} \tag{3}$$

Inspection:

$$F = 28.81 \rangle F_{0.05}(2.76)$$ regression model significantly.

According to the model, we can get the Shanghai travel demand predicted value (see Table 2).

Table 2: The Comparison About Forecast Value And The Actual Value

year	2009	2010	2011
predictive value	4389.8	5858.6	6808.5
actual value	4150	5987.9	6950

6. CONCLUSION

According to the features of Web data, such as advanced and timely acquisition, so in the study of tourism market demand, we can use Web monthly data to judge the seasonal changes of tourists in advance. When there is some significant activities like festivals and commemorative activities or some critical incidents like scenic natural disasters, Web data will appear obvious anomaly, and its rate such as attention degree will also appear abnormal (in general condition the data will improve), it is particularly apparent for the scenic area. So we can use Web data's abnormal monitoring to do the early warning, then make the tourist destination and the enterprise can respond preparation and sales promotion in advance. At the same time we can also use Web data apace, conveniently and accurately understanding the market demand's characteristics and trends, then does the tourism demand forecasting of tourist destination and tourism enterprise, for provide a foundation to the tourist destination and the tourist enterprise.

Web data application prospects: web data will bring greater economic benefits to tourism enterprises. Tourism enterprises will be more emphasis on web data. Web data will be applied to the broader field of tourism. Everyone will be more aware of the importance of web data.

REFRENCES:

 [1] Daniel Guttentag, "Virtual reality:Applications and implications for tourism", *Tourism Management*, Vol. 31, 2010, pp. 637-651.

[2] Büyük zkan G,Ergün B., "Intelligent system applications in electronic tourism", *Expert Systems with Applications,* Vol. 38, No. 6, 2011, pp. 6586-6598.

[3] Jiao Guo,Pei Yan, "Analysis and Design of Urban Tourism Information System Based on WebGIS", *Energy Procedia*, 2011, Vol. 13, 2011, pp. 3794-3799.

[4] Shi Weidong,Lu Yang,Li Zhu, "SHARC:A scalable3D graphics virtual appliance delivery framework in cloud", *Journal of Network and Computer Applications*, Vol. 34, 2011, pp. 1078-1087.

[5] Steven F Illum,Stanislav H Ivanov,Yating Liang, "Using virtual communities in tourism research", *Tourism Management*, Vol. 31, 2010, pp. 335-340.

[6] Robertson, R. A., "Recreational displacement and overall satisfaction: A study of central Iowa licensed boaters", *Journal of Leisure Research*, Vol. 26, 2009, pp. 174-181.

[7] Henrique Souto, Nuno Gomes, Rui Carvalho " Development of a GIS for the "Celebration of Coastal Culture" " *Journal of Coastal Conservation*；Vol. 16, 2012, pp. 431-437.

[8] Esteban Ruiz-Ballesteros, Miguel A. Gual , "Recreational displacement and overall satisfaction: A study of central Iowa licensed boaters", *Human Ecology* Vol.40, 2012, pp. 847-862.

[9] Chia-Lin Hsu, Kuo-Chien Chang, Mu-Chen Chen"The impact of website quality on customer satisfaction and purchase intention: perceived playfulness and perceived flow as mediators", *Information Systems and e-Business Management,* Vol. 10, 2012, pp. 549-570.