

## Overview of Tourism Data Mining in Big Data Environment

Wenjie Xiao<sup>1,2</sup> and Changguo Xiang<sup>1</sup>

<sup>1</sup>Key laboratory for ecotourism of Hunan province, Jishou University, Zhangjiajie, Hunan, China

<sup>2</sup>Tourism and management project, Jishou University, Zhangjiajie, Hunan, China

314347559@qq.com

**Keywords:** Big data; Data mining; Cloud computing; Tourism; Clustering

**Abstract.** With the advent of the information age, the network is filled with all kinds of information and people in the vast amount of information are in the face of all kinds of confusion, thus data analysis has become a more difficult problem. Big data can solve this problem well and can greatly improve the efficiency of data mining. In order to apply the big data technology to the tourism industry, this paper introduces the concept of big data and the development demand of tourism big data, summarizes the common technology of data mining and the mining technology of tourism big data, and finally gives the application direction of data mining in tourism.

### Introduction

Nowadays, big data has been mentioned by more and more people, and gradually become the focus of social concern. Especially for the information technology companies, big data has already been the core of the study. With high-speed broadband and virtualization technology, combined with a variety of operations, the big data environment comes into being[1]. Although big data are being mentioned by more people, throughout the development of big data areas, big data capacity is still in a relatively low level and has not been given the exact meaning by the industry. However, through the basic understanding of big data, it is not difficult to find big data with many features, such as large-scale computing power, storage capacity of resources and dynamic contraction of data, which greatly improve the possibility of data mining becoming commercialized under big data environment.

### Overview of Big Data

Big data, a huge capacity of data resources, refers to the data involved that is of large amount and unable to use the current software tools to capture, manage, process and organize in a reasonable time. Big data is the data collection composed of a large number of data with complex structure, and numerous types. It is intellectual resources and knowledge service ability, based on the cloud computer data processing and application mode, forming through the integration, sharing and cross-use of data[2].

With the development of the times, the big data age has been quietly coming, but the problem that how much data can be called the big data has no clear definition so far. However, there are some characteristics of big data, which recognized by the industry, that is, a large number, diverse, high-speed, value[3-4]. Today, the rapid development of the Internet and the boom of mobile Internet make the scale of Internet larger, also the user's use of the Internet has broken the time and space limits, and the user can interact with the Internet indiscriminately, as well as write data frequently, which causes the amount of data generated everyday become extremely large.

At present, there are big data problems in all industries, such as medical big data, financial big data, government big data, tourism big data, etc. Big data is a valuable wealth, but how to tap the wealth is one of the largest challenges now. In terms of tourism, China's tourism industry is developing well and the construction of tourism information is developing rapidly. However, in the face of big data, problems such as higher maintenance cost, poorer service experience and less rapid and accurate mining of specified information immediately reveal. The advent of big data age has

brought challenges, but also brought opportunities. Make full use of the original and mass tourism data and mine the mass tourism data fast, conveniently and accurately to provide tourists with good tourism information services, which has become the new direction of tourism big data research[5-6].

### Development Demand of Tourism Big Data

The rapid development of the Internet accelerates the construction of tourism information. Tourism information services develops from the initial traditional artificial mode with paper into electronization and digitalization. Information brings revolutionary changes to the development and value enhance of tourism again and again. From the tourism service platform, initially, hotels, airlines, travel agencies and other tourism service providers moved their own offline business to the Internet, forming a single travel service platform, later One-stop service platform like food, accommodation, travel, shopping, entertainment, etc. boomed at home and abroad, which relied on the support of network technology. The advancement of technology has made the information delivery and service experience better and better, and the tourism service providers have received a generous return from it, at the same time, the tourism has made great contributions to the economic development of the country. Rapid development of tourism is inseparable from the tourism information, and the survival of the fittest of the tourism industry is closely related to high-tech, which has become an important tool and method to strengthen tourism competition[7]. And thus high-tech, advanced platform and business philosophy related with the tourism service have also been changing. The single service platform previously is quietly restructuring now.

Along with the rapid development of China's tourism, China's tourism industry has also undergone tremendous changes that the industry has been expanding, leading to the explosive growth of tourism data, and it has formed a huge mass of information space. It has become a new direction for the application of tourism big data that how to make full use of the massive original tourism data and mining and analyzing the massive tourism data reflecting the passengers' information in rapid, accurate and convenient way. The following figure shows the changes of international tourists.

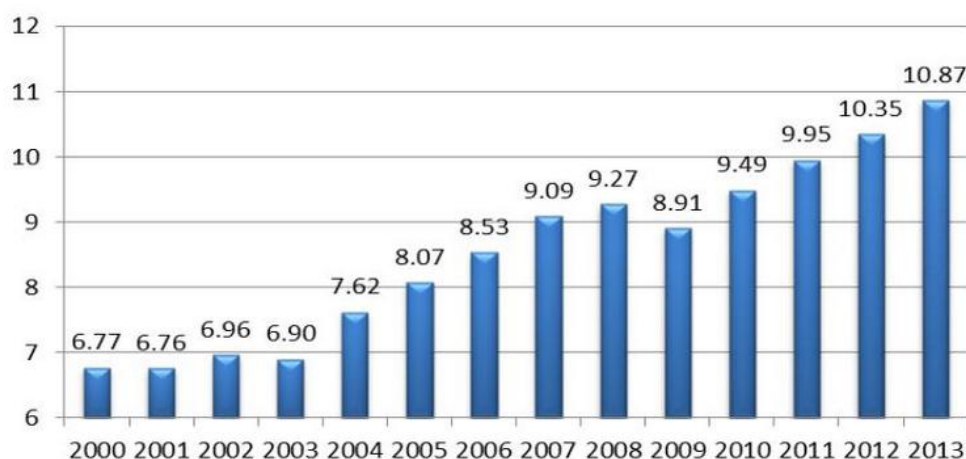


Figure 1. Chart of International tourists changes (one hundred million)

The development of tourism big data is mainly reflected in the following aspects:

1) Tourism big data mining services

Analyze the massive tourism data through dedicated service agencies to provide standard data analysis report and forecasting results, it can also allow the users to create their own model through open interface to analyze the data.

2) Data visualization

The mining results are presented through visualization (graphs, etc.) tools[8-9].

3) Data analysis database

Preserve the commonly used analysis model and the results as the data analysis tool, so that the development personnel can fast carry on the tourism data mining application.

One of the functions of big data is to provide consumers with matching choices and personalized services in each market segment so that they can make better decisions. In the times of big data, tourism begin to shift from sightseeing to experience travel, and tourists' relying on travel agents transforms to relying on the Internet self-help travel, which becomes increasingly apparent, and their demand of the convenience of information access is higher, and consumption patterns are more diversified and personalized[10-11]. Using the terminal interface tool provided by the wisdom of tourism, tourists can obtain fully information of the destination to arrange their own itinerary and custom private tourist routes. Subject to the tourism reform driven by big data, the deep excavation of massive tourism data makes visitors can not only guide their own consumer behavior precisely, but also can make the scenic spot better understand the tourists, and ultimately promote the improvement of overall information level of tourism industry. The following figure shows the changes of tourism industry investment.



Figure 2. Tourism investment scale

## Tourism Data Mining

**Data Mining.** Data mining[3], also known as the knowledge discovery of data, extracts valuable information hidden in the massive, incomplete, fuzzy, noisy and random data, which is one of the hot topics in current research of artificial intelligence and database field. It is a decision support process. Based on machine learning, artificial intelligence, database, pattern recognition, etc., it automatically analyzes large amounts of data to make inductive reasoning and dig out the potential model from it so as to provide decision support for government departments.

Data mining technology mainly includes association analysis, sequence pattern, classification, clustering, and anomaly detection and so on. In the application of tourism data, we can use the association analysis to search the tourism data and find the patterns with high probability of occurrence, or analyze the similarity of tourism data through the clustering and classification of data and store the similar data together, to provide decision-making support for decision-makers. Data mining has higher requirements on the computer, and the following table is the computer hardware configuration of the database.

Table 1 Hardware configuration table

Server Name	Model Number	Server Quantity	Cpu	Memory	Hard Disk
IBM	M4	6	E5-2650	16G	5T
IBM	M4	10	E5-2650	16G	5T

### Application of data mining in the tourism industry

Mine the hidden information with the use of data mining technology to provide valuable information for the tourism industry, which mainly reveals in the following aspects:

Mine valuable tourism information

Analyzing the tourist's log of browsing website and using the data mining technology to find out the common browsing behavior and grasp the tourist target and information which tourists are interested in. Therefore, the tourism management department can adjust and optimize the tourism information website based on this information so that it is in line with the needs of tourists, to provide more qualified travel services, and for specific visitors, to develop personalized service. For example, through data mining analysis, when a visitor browses the website information of Zhangjiajie National Forest Park, he is likely to visit the website information of the ancient city of Phoenix, you can see it is of high possibility that the tourist continues to visit the ancient city of Phoenix after visiting Zhangjiajie National Forest Park, and vice versa. Therefore, the tourism management department can increase the tourist bus between the two scenic spots in corresponding, or recommend the website of the ancient city of Phoenix in the website of Zhangjiajie National Forest Park.

#### Mine potential tourism customers

It can obtain the potential tourism customers through tourism data mining, and the main methods include data clustering and analysis. For the new visitors, it can identify whether the tourists are potential tourists or not through the cluster and analysis of their access records, and then to decide whether to treat them as a potential customer. If it is, recommend the tourism information the customer is interested in to the customer.

#### Optimize the tour route strategy

Through the mining and analysis of tourist routes and destination access, the tourism administration department can provide the best or the most market-potential tourism routes. The management department can reasonably plan the lines with certain relativity and dynamically adjust the structure of the tourism websites so that customers can access to travel information more conveniently. It can effectively increase the viscosity of the website and improve the visit rate of the website.

#### Recommend tourism projects and destinations

Using data mining technology develops new tourism projects and tourist destinations. First, create a tourism database containing all the tourist destinations, and use data mining tools to analyze customer behavior, interests, hobbies, etc., to form a set of tourism service programs for different types of tourism customers. According to customers' preferences, needs and physiology, provide personalized travel advice, recommend tourist routes, provide tourism service projects during the destination, and assist tourists to make the appropriate tourism plans, including the tourist routes in the destination, transportation, accommodation, dining, entertainment, local specialties, sightseeing projects, etc., to improve satisfaction of the tourism customers.

The general process object and process of data mining are as follows:

Table 2 Data mining process object table

User:	Tourism management personnel
Function:	Tourism data mining
Data:	Accumulated data in usual days
Access :	Computer Server
DB size:	TB

## Summary

Tourism big data and data mining technology are developing at an unprecedented speed. In the increasingly fierce tourism market, it will get more valuable information and more market opportunities with tourism data technology. Tourism big data makes the allocation of tourism resources more reasonable, finds potential customers and provides better personalized service to the existing customers. With the development of tourism big data, there will be more data analysis technologies used in the tourism industry.

## Acknowledgements

A Project Supported by Open Fund FY2014 Key Laboratory of Eco-tourism in Hunan  
Project number: JDSTLY1501

## References

- [1] Lanxiang Chen. Using algebraic signatures to check data possession in cloud storage[J]. Future Generation Computer Systems. 2013(7):20-24.
- [2] Yan Zhu, Hongxin Hu, Gail-Joon Ahn, Stephen S. Yau. Efficient audit service outsourcing for data integrity in clouds [J]. The Journal of Systems & Software . 2011 (5):34-38.
- [3] Sangjae Lee, Sungil Jeon, Doyoung Kim. The impact of tour quality and tourist satisfaction on tourist loyalty: The case of Chinese tourists in Korea[J]. Tourism Management . 2010 (5):112-118.
- [4] K Keahey, R Figueiredo, J Fortes. Science Clouds: Early Experiences in Cloud Computing for Scientific Applications [D].Cloud Computing & Its Applications, 2014,17(11):745-748.
- [5] RL Grossman, Y Gu, M Sabala. Compute and Storage Clouds Using Wide Area High Performance Networks[J].Future Generation Computer Systems, 2014, 25(02):179-183.
- [6] R Grossman, Y Gu.Data mining using high performance data clouds: experimental studies using sector and sphere[D].Acm Sigkdd International Conference on Knowledg, 2013,98(07):920-927.
- [7] Jinlin Chen,Keli Xiao. BISC[J]. ACM Transactions on Knowledge Discovery from Data (TKDD). 2010 (3):334-339.
- [8] Chung-Hong Lee,Tzan-Feng Chien. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking[J]. Journal of Information Science. 2013 (4):320-328.
- [9] J Wang, J Wan, Z Liu, P Wang. Data Mining of Mass Storage Based on Cloud Computing[D].Ninth International Conference on Grid & Cloud,2013,257(12):426-431.
- [10]G Piatetsky-Shapiro. Knowledge discovery in databases: 10 years after[J]. Acm Sigkdd Explorations Newsletter[J].2012,1(02):59-61.