Application of Data Mining Technology in Tourism Information Recommendation

Wanchun Zhu

Guizhou Minzu University, Guiyang 550025

Key words: Data mining; Tourism; Information recommendation; Application

Abstract: With the continuous improvement of the economy, the tourism industry has gradually become a trend, rich attractions, service diversification has been loved by the majority of tourists. However, with the rapid development of tourism, it is also faced with many difficulties. It is very important to carry out the recommendation according to the characteristics of tourists' preferences and tourist information. It is difficult for tourists to make a choice of tourist attractions and ways in the absence of reliable information. Data mining is a process of decision support, is the deep data analysis method, this paper uses the Apriori algorithm to recommend tourism information for data mining, analysis according to the personal preferences of tourists, to help visitors make a choice. Effectively improve the accuracy of recommendation.

Introduction

With the rapid development of data information and the need to make decisions on complex data, data mining technology emerges as the times require. Although the history of data mining is short, but in 1990, the development of the technology is very fast, has been widely applied to all walks of life, such as finance, retail, manufacturing, education, tourism, etc.. Each industry will have a large amount of data accumulation, the data mining technology can play a huge role in the field. Compared with foreign research, domestic research on this technology is relatively late, at present, is also a study on the algorithm of data mining, the theoretical research and practical application of the primary stage.

In the tourism industry, a lot of tourism industry has seen only a short profit but ignored the development strategy of tourism information resources. The construction of tourism information and tourism business has been most derailed, moreover many tourist destinations will be used to modify this information as a product, still remain in the stage of simple information query and statistics, so the efficiency of information application is very low, the information lost more in-depth digging into their own industry benefits. To sum up, the tourism industry needs the data mining technology to obtain more valuable information for the industry, the tourism industry to improve the level of management, make it more humane, so as to continuously enhance the competitiveness of tourism destinations.

Overview of Data Mining Technology

Data mining is the process of refining the unknown, hidden but valuable information for people from the tedious, vague, incomplete and random data. When we regard the original data as the beginning of forming system, the knowledge we saw not only can be used for information searching, making decision etc., but also can be used for the protection of data itself. Therefore, we can say...
that this technology has integrated with many subjects and it has got extensive application in database, data statistics and artificial intelligence etc., aspects.

The characteristics of data mining mainly include the following aspects:

1. The amount of data needing to be processed is very huge. The IDC data, ending to 2012 has a number of over 10 times of that in 2004.
2. As long as the users provide the vague inquiry request, the data mining technology can obtain the users’ favorite things as possible as it can.
3. Building the data model through this technology and when the data has slight change, this technology can generate fast response, which can increase the users’ decision-making efficiency.
4. The data mining technology not only can build the data model of the original database, what the more important is that the data model will update along with the change of database.
5. Data statistics is the premise for data mining technology to form its discipline, generally, this technology is not taking analysis of all the data, but when one limiting value appears, the corresponding discipline will appear, which confirms to the majority of the users.

1.2 Various Methods of Data Mining

The common used data mining methods is beyond count and the traditional data mining methods include: random-sampling technology, multivariate statistical analysis and big data statistical prediction technology; visualization technology expressing the content in forms and data graph; decision-making tree for prediction by a series of disciplines classification and the involved algorithms include CART, CHAID etc; neural network based on the biological evolutionism viewpoint and genetic algorithm.

1.2 Problems and Improvement of Various Methods

Although the data mining technology has many advantages, in the practical operation process, there are still some problems, which also has provided the development direction for the future of data mining technology.

1. The basic and key point of this technology is the dimension and quantity of data besides, how to take the variate selection and expansion is also the difficult problem needing urgent solution.
2. The knowable data is of huge quantity and complex and the data will inevitably hide the trend of changes, therefore, we should give consideration to the study and analysis of this trend for data mining.
3. It is inevitable to involve the intimate data for data mining, therefore, we should give consideration to data safety.
4. Generally, the result got by data mining is not accurate, therefore, the analysis should be taken by combining with the related professional knowledge.

In this paper, it has selected the Apriori algorithm in association rule and the testing method adopted by this algorithm is the iteration method layer by layer. Besides, through scanning the database for many times to fine the frequency-item sets, so that to decrease the time of scanning database.

The system flowchart of data mining is the primary premise for finishing the application of this technology. Generally, one complete flowchart includes the following steps:

Setting of Business Object

Only when the data mining technology is applied to the practical and concrete field, can it achieve its greatest value. Knowing clearly the significance of data mining and concluding the common problems is the beginning of implementation process, besides, the final result cannot be predicted, however, the research problems are predictable[1]. Therefore, before the implementation, the purpose and demand should be set.
Collection and Preprocessing of data

After setting the business object, we need to collect data. The source of data can be the transaction processing system or the data warehouse. The following work is the ‘decontamination’ of data so as to achieve the generalized processing and make the data be more abundant than before.

Data Mining

This step is the core part of the whole process. The main function of this part is to extract the valuable and easy-to-follow part of the data which has been processed before. The final purpose is to describe and predict, so that to achieve the effect of providing prediction trend.

Not all the mining data is valuable, therefore, it is necessary to take effective analysis and evaluation of the mining result. So the quickest way of that how to analyze the result got is valuable or not is to take evaluation of the pattern and the get the most suitable and best pattern.

Result Description

This step expresses the technology by making use of the visible technology and knowledge and displays the knowledge pattern in huge database by using the data mining tools. In this process, the especially important point is to provide one straightaway result for users to accept, in which way, it not only saves the time, but also increases the efficiency.

---

**Figure 1** The general process of data mining

---

**Overview of Tourism Information Recommendation**

With the fast development of tourism, people’s requirement for tourism service also has certain increase, besides, the requirement for the service of tourism bodies become more and more strict, the reason of which is that most of the agencies still stay on the stage of traditional running pattern, ,which has single programs and the tourism routine and scenic spots are boring. In terms of these problems, this paper has taken the data mining by combining the professional knowledge and techniques and taking analysis according to tourists’ individual habits and interests to help the tourists make the decision. Recommend the tourism scenic spots and service by obtaining valuable information from the huge database and try to satisfy more potential tourists as possible as it can.
The tourism information recommendation can be made to recommend the scenery for tourists according to their hobbies and interests, so that to help them to plan the itinerary.

With the intelligent development of tourism, the tourism information recommendation has become the hot research subject and it is mainly the analysis of the tourism information and users information\(^2\). At present, the quantity of China’s experts engaging in tourism information is of small number, Professor Du Junping from Beijing University of Posts and Telecommunications, is relatively professional in this aspect. His proposition mainly is taking multi-directional research and analysis by making use of the tourism data management decision to support system. The foreign researchers explore the tourists’ hobbies and interest mainly by the gravity model and time sequence model. Since the 1990, the cross-over study of the artificial intelligence etc. disciplines has been taken, besides, it has been applied to the model analysis and has caused great influence.

**Application of Data Mining Technology in Tourism Information Recommendation**

Currently, some researchers combine the data mining technology with OLAP and apply it to the tourism information recommendation, for which, they will take multi-directional and multi-dimensional data mining. Besides, the relevant individual have put forward that use the data itself to describe the fields and user model besides, introduce the personalized service into the system structure so that to achieve the purpose of personalized tourism information recommendation by using the special-purpose language of user model.

The Apriori algorithm selected in this paper belongs to one-dimensional association rule and its association process mainly concludes two parts: the first part is selecting the item-sets which is higher than the set value out; the second part is finding the rule of selecting the item-sets that higher than users’ lowest trust degree from the selected item-sets.

Steps of Apriori algorithm item-sets selection as follows:

Input: B, the limit of minimum support degree: little

Output: frequent item-sets M in B

\[
\begin{align*}
M_1 &= \text{search\_usually\_1\_collection}(B); \\
&\text{for (n = 2; } M_{n-1} \neq \emptyset; \\
&\text{n++)} \\
&\quad \{ \\
&\quad \quad P_n = \text{apriori\_gen}(M_{n-1},\text{little}); \\
&\quad \quad \text{for every turn } x \in B \{/ \text{can}\ B\ /\ number \\
&\quad \quad \quad P_x = \text{subset}(P_n,x); \\
&\quad \quad \quad \text{for every name } p \in P_x \\
&\quad \quad \quad \quad p.\text{number}++; \\
&\quad \quad \} \\
&\quad \quad \text{M}_n = \{p \in P_n \mid p.\text{number} \geq \text{little}\} \\
&\quad \} \\
&\text{return } M = \bigcup_n M_n; \\
\end{align*}
\]

The application of tourism information recommendation process can be described as:

Procedure apriori(Mn-1: usually (n-1)-collection; little: support)
for every collection m1 \in Mn-1
for every collection m2 \in Mn-1
If (m1[1]=m2[1]) \land \ldots \land (m1[n-2]=m2[n-2]) \land (m1[n-1]<m2[n-2]) then {

---

96
p= m1 m2;
if has unusually_subset(p,mn-1) then
    delete p;
else add p to Pn;
}
return Pn;

In the systematical model of tourism information intelligent recommendation, the function of step 3, 4, 5 is taking classification. Taking the attribute selected by step 2 as the decision attribute and the other scenic spots as the target attribute, the concrete recommendation model showed as the following figure\[3\]. For the tourism scenic spots recommendation, it has used five classifiers and because the used target attributes in scenic spots recommendation have a number of over 60, the four classifiers are used as the scenic recommendation and the left one is used for attribute classification, so as to increase the working efficiency of classification. If the attribute value has a large quantity, we can choose many classifiers to reduce the quantity of target attribute value and do not calculate the probability value.

![Data Mining Model](image)

Figure 2 Data Mining Model

People have never stopped the study of data mining and they have not only got achievement theoretically, but also got the predicted effect in practice, besides, for the tourism information recommendation, the common used tools include:

Bayes Network

The Bayes network advocated by researchers is for forming a probability model. For the got network structure, each node concludes one group of father node, which has influence on the node, besides, as long as the father node is given, the value of this node can be got. The probability condition of the corresponding node is showed by the decision-making tree. If the possible value of the father node is known, then its corresponding node value can be got. The model-building time of this method is long, so it is suitable for the environment with small change scope.

Classification

Classification refers to reflecting the data-concentrated region to the designated aspect and what it reflects is the common feature of the same type of things and the different performance features of different things. The most common method is the classification method of decision-making tree and this method refers to that building the decision-making tree in the practical example, besides, the classification also includes statistics and rough set etc. methods.

Artificial Neural Network
Neural network can take processing of the huge and complex data, besides, it takes the mathematical model of self-learning as the basis and it can complete the complex data analysis, which cannot be completed by human and computer. The structure of neural network system is similar to human’s neuron, which is constituted with neural units, which is called node and produces correlation through the linking nodes of network, besides, as long as the data is input, the model of data working can generate.

The workflow of decision-making tree is similar to the shape of tree, besides, the tree trunk embodies the testing result and the node of leaves embodies the distribution situation of classification. This method is the recursive form from top to bottom, so the node on the top is the tree root. This method provides one rule-based method, which will produce different results in different environment. Adopting this method can make the data rule be visible, meanwhile, the result got is easy to understand. The advantages of this method are that it is of high accuracy and consumes less time.

**Correlation Rule**

Correlation rule refers to that find the potential correlation between the known data and the main work of this method is to confirm the database of one thing, so as to find the interrelation of complex data and then form the reliability.

**Conclusion**

This paper has introduced the related knowledge of data mining and has chosen one classical Apriori algorithm to take data mining of the tourism information recommendation and take analysis according to tourists’ individual hobbies and interest, finally help them to make decision. Recommend the tourism scenic spots and service by obtaining valuable information from the huge database and try to satisfy more potential tourists as possible as it can. After the analysis, we can know that the data mining technology has great development potential in the future tourism industry.

**Bibliography**


**Introduction**

Wanchun Zhu (1978-) male, Li people, Qinglong Guizhou people, Doctor of management, Associate Professor of Guizhou Minzu University, research direction is Tourism economics