

Application of Data Mining Technology in the Tourism Product's Marketing CRM

Shenglei PEI^{1, a}

¹ School of Physics and Electronic Information Engineering, Qinghai University for Nationalities, Xining 810007, China

^a psl211@163.com

Abstract—To carry out Customer Relationship Management (CRM) is necessary to improve enterprises' core competition ability. CRM focus on customer resource management by using information technology, in order to assist managers to make practical decisions. And data mining and data warehouse technology are powerful tools to improve the rightness and accuracy of CRM. As a simple and efficient classification technology decision tree is a member of data mining family. This paper referred the advanced ideas and systematic theory research of CRM. And with the combination of the actuality of tourism industry, we built a CRM system for tourism industry based on data warehouse and data mining technology. [1]We analyzed CRM system structure and functional modules with application program, and introduced the characteristics of data storage integration and information processing and analysis.

Keywords—Customer relationship management; Data mining; CRM in tourism industry

I. INTRODUCTION

Facing the cruel market competition, all enterprises spare no effort to get as many customers as possible and maintain a long-term steady relationship with the customers. The aim of Customer Relationship Management (CRM), the new management idea, is to improve the relationship between enterprise and the customer; its core is "understand customers, listen to customers", whose goal can be summarized as "attract potential customers, improve the existing customers' satisfaction and loyalty, and reduce customer churn". In one word, the ultimate goal of all is to improve profitability and competitiveness. [2] As one of the three pillars, for tourism travel agencies to win from the market the importance of implementing CRM is self-evident. However, over a long period, the CRM data in the database is growing and people want to be able to provide a higher level of data analysis functions to convert data to be processed into useful information and knowledge automatically and intelligently. Data mining just provides us with effective method to solve these problems. It finds the potential relationship between data and provides decision support for us through the analysis of customer needs. This article will discuss the data mining technology in the travel product sales briefly.

II. THE MAIN METHOD OF DATA MINING

There are many approaches of data mining and in actual application of these methods are often linked together to complete a mining project. The most commonly used methods include decision trees, concept description method, association rule mining method, neural network, etc.

1) Decision tree

Decision Tree is a technique frequently used. It mainly uses a series of rules to divide and establish tree chart. It can be used for data analysis, as well as be used as a prediction. Algorithms commonly used are CART, ID3, CHAID, PUBLIC, C4.5, etc.

2) Concept description

Concept Description is the most basic form of the describe type of data mining; it is made up of characteristic and comparison. [3]It mainly collects, analyzes and compares data related to a kind of object and these data are summarized according to the relevant features. Concept description digging generally adopt the method of oriented database and concept description results vary in forms, such as a pie chart, bar chart, curve and multidimensional data cube and multidimensional table, including cross table, etc.

3) Association rule mining algorithm

The main object of the association rules is transactional database to describe the rules of data relationship. Association rules data set A to data set B means "most database records meeting the condition of data set A also meet the data set B conditions". Another example, "if there are 73% of consumers access both /picture/image1 page and /picture/image2 page", using association rules it can be expressed as: Image 1 page - image2, confidence level being 73%.

III. DATA MINING PROCESS IN CRM

Data mining process in CRM is shown in figure 1.

(1) Data sorting

Because the source of the data collected representation method differ from enterprise to enterprise, data is chaotic. [5]Before data mining, it must first organize the data. Data sorting mainly includes four processes, namely data cleaning, data integration, data selection, data exchange. Data cleansing' goal is to eliminate the noise in the raw data and inconsistent data to make a variety of data can be combined together and makes the data into a unified or suitable for the form of mining by retrieving and analyzing related data in the database.

(2) Data mining

Data mining stage is to perform operations on data

already selected by data mining method already set to extract useful information from these data.

(3) Assessment and show

This stage is divided into two sub-processes, pattern evaluation and knowledge representation. Pattern evaluation refers to identify really interesting knowledge model according to the scale of interest. [6] Knowledge show refers to use visualization and the knowledge representation techniques to provide customers with the mining knowledge.

IV. THE APPLICATION EXAMPLES

Table 1 is the basic information of the customer and their corresponding monthly average consumption quota of an enterprise product and we will use the decision tree algorithm for customer profitability analysis in the following.

(1) Data sorting

In order to facilitate the operation the data should be processed first.

(2) Carry out data mining by using decision tree algorithm

The key point of using decision tree algorithm is to calculate the information gain and look for a branch node. [7] Formula to calculate the information gain is:

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A)$$

$Gain(A)$ represents information gain of attribute A ;

$I(S_1, S_2, \dots, S_m)$ is the expectations of the element information, m in which says the number of attribute values. Attributes in the table above describing results are consumption quota, i.e. the higher the consumption quota is, the more valuable the customers are; there are two values 0, 1. m is of 2. $S_1 = 0, S_2 = 1$. $E(A)$ is the expectation of attribute A . And:

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i \log_2(p_i)$$

P_i is frequency of attribute value

$i, p_1 = 5/8$ (frequency of 1), $p_2 = 3/8$ (frequency of 2), so

$I(S_1, S_2) = -5/8 \log_2(5/8) - 3/8 \log_2(3/8) = 0.9544$. And:

$$E(A) = \sum_{j=1}^v I(S_{ij}, \dots, S_{mj}) (S_{ij}, \dots, S_{mj}) / S$$

Among them, algorithm $I(S_{1j}, \dots, S_{mj})$ is with properties of gender as an example, the gender value also has two values, 1 and 0; consumption quote = 1, gender = 1, $S_{11} = 3$, similarly

$S_{21} = 1, S_{12} = 2, S_{22} = 2$, so

$$I(S_{11}, S_{21}) = I(3, 1) = -3/4 \log_2(3/4) - 1/4 \log_2(1/4) = 0.8112, I(S_{12}, S_{22}) = I(2, 2) = 1$$

$$E(gender) = 4/8 I(S_{11}, S_{21}) + 4/8 I(S_{12}, S_{22}) = 0.9056$$

$$Gain(gender) = I(S_1, S_2) - E(gender) = 0.9544 - 0.9056 = 0.0488$$

$$Gain(Age) = 0.2657$$

$$Gain(Marital\ status) = 0.0157$$

$$Gain(Income) = 0.5032$$

$$Gain(Distance) = 0.90565$$

According to the information gain value, the decision tree is drawn in figure 2.

According to the method above, the final decision tree is shown in figure 3.

V. SUMMARY

CRM system based on data mining technology can make better use of customer information, quickly and efficiently get valuable knowledge thus make the enterprise to realize efficient management and operation. The application of data mining technology in CRM research has made many achievements, the enterprise is more and more aware of the importance of CRM and data mining technology has also made a vigorous development. But a lot of research is still stay in theoretical and lack of practice; many theories need to be tested and perfected in practice.

ACKNOWLEDGEMENTS

Sponsored by "The Ministry of Education "Chunhui Plan" cooperative research project (Z2012104), Qinghai Province International Science and Technology Plan funded Project(2013-H-811)"

REFERENCES

- [1] N.Dayasindhu. Embeddedness, Knowledge Transfer, Industry Cluster and Global Competitiveness: A Case Study of the India Software Industry [J]. Technovation, 2002 (22).
- [2] Maryann P.Feldman, Johanna Francis. Creating a Cluster while Building a Firm: Entrepreneurs and the Formation of Industrial Cluster [J]. Regional Studies, 2005 (1).
- [3] Stansfield C A, Rickert J E. The Recreational District[J]. Journal of Leisure Research, 1970 (4).
- [4] Pamela J.McRae-Williams. Wind and Tourism: Cluster Complementarity and Regional Development; New Zealand Tourism and Hospitality Research Conference [J]. Wellington, 2004.
- [5] Marina Novelli, Birte Schmitz. Networks, Clusters and Innovation in Tourism: A UK Experience [J]. Tourism Management, 2006 (27).

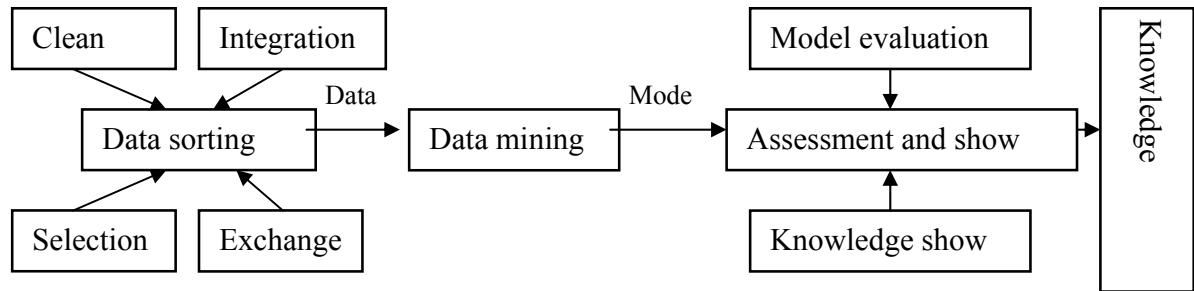


Figure1 Data mining process in CRM

Table 1 The initial basic customer information

Gender	Age	Marital status	Income	Distance	Quota
Female	35	Married	5000	3	800
Male	26	Unmarried	4000	7	500
Female	30	Married	3000	8	300
Male	25	Married	2000	1	400
Female	45	Married	6000	5.5	600
Male	20	Unmarried	1000	2.5	100
Male	50	Married	4000	2	500
Female	38	Married	8000	18	900

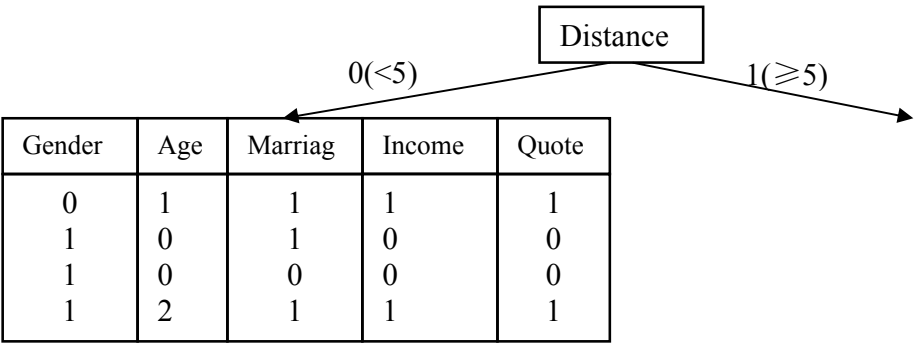


Figure 2 Decision tree with distance as branch node

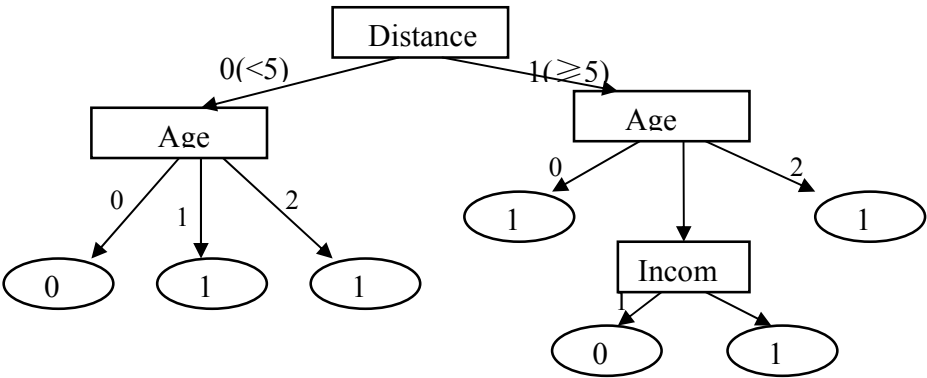


Figure 3 The final decision tree