

Research and Application of KDD Techniques in Automobile Marketing

Gang Xuhan

Dept. of transportation, Northeast Forestry University
Harbin, China
326634099@qq.com

Jiang Bohan

Dept. of transportation, Northeast Forestry University
Harbin, China
jiangbehan@126.com

Ba Xingqiang

Dept. of transportation, Northeast Forestry University
Harbin, China
bxq1218@sina.com.cn

Abstract—Customer Relationship Management (CRM) system which puts client on the central location, plays an important role in automobile marketing field. It can help corporations understand customer's requirements better. Improve the relationship between the customers and the corporations, can enhance the customer's degree of satisfaction, so that it can advance the corporations' competitive strength. The application of KDD technology in CRM, makes CRM system develop into aptitude, helps automobile marketing corporation forecast operation developmental current, helps corporation disentanglement new customers and hold the old customers, and it makes corporations at a more advantaged competitive position. This article based on CRM and the basic principle of KDD(Knowledge Discovery in Database),then analysis the requirements of KDD application in the automobile field and establish potential customers, for example to obtain model by the database of CRM system of one auto sale company. And using this model to test the applicability of the model.

Keywords—component; Automobile marketing ; CRM ; KDD ; Data analysis ; Customer satisfaction

I. OVERVIEW OF AUTOMOTIVE MARKETING

In the past days the scale of automobile production is the decisive factor of deciding the corporation's competitive power. With the development of science and technology, especially the development of information and network, the difference of products' quality between corporations gradually reduced. But the high standard of customers' service has gradually become the essential factor of corporation [1]. Some farsighted automobile market corporations start to pay attention to the system of Customer Relationship Management (CRM) which helps businesses to administer the customers' information. The implementation of CRM system can help corporations understand customers' requirements better and improve the relationship between customers and corporations, and

enhance the customer's satisfaction degree, then advance competitive strength.

Now, the development of the automobile market is undergoing significant changes. The entire automobile field is changing from "Product-centric" to "Customer Focus" gradually [2]; Customers have become the critical factor and source of potential profit. But with customers' information increasingly sophisticated and data's accumulated, the effective management of information has become a serious problem.

At present, the foreign customer have researched topics of relationship management focused on Web-based Customer Relationship Management, knowledge mining in Customer Relationship Management application, such as the sound management of customers, the customer relationship management, ROI quantitative research and customer relationship management, security and data privacy issues [3,4]. The research of CRM which began at the end of 20th in our country has gone through two stages. The first stage is the introduction of the concept of customer relationship management, and the theory and practice of foreign CRM is mainly introduced in this stage. The second stage is the pilot phase of customer relationship management, and a number of CRM systems are developed and applied to the corporation in this stage. [5, 6].

II. KNOWLEDGE DISCOVERY IN THE FIELD OF AUTOMOTIVE MARKETING

Cars are very luxurious items in people's daily lives. How to find and seize the potential customers and

retain the former clients at the same time? We can use knowledge discovery techniques. Now, knowledge discovery technology has been used to tap potential customers [9].

A. Model construction

Construction and application process of potential customers to obtain the following model:

1) In the service database which obtains from each kind of marketing activity collects reorganizes the primary data, sets up the enterprise customer the data warehouse system. Extracts the suitable field composition customer analytical data in the customer data warehouse, provides the data pool for the latent customer gain model.

2) Analysis of data from the selected clients in the implementation of data mining, found that products of interest to potential customers with the feature model[10].

3) Utilizes the latent customer characteristic model in the current customer database to discover that most has the possibility to become the potential customers.

4) Carries on the target-oriented marketing activity on the characteristics of potential customers with an enterprise customer base property to achieve the purpose of acquiring new customers.

The following example is introduced that one modeling method be used in classifying, the cluster and the forecast by example. The name of the method is Decision tree algorithm. Decision tree algorithm uses the "divide and conquer" approach to the problem of the search space into several subsets. Sample space

$$S = \{S_1, S_2, \dots, S_n\}$$

Each sample has many attributes to describe,

$$S_1 = \{a_1, a_2, \dots, a_i\}$$

1) According to the one attribute of the sample, divided it into many kinds the sample space

$$S = \{C_1, C_2, \dots, C_m\} \quad (4-1)$$

The information content position carrying on classified to S

$$I = -\sum_{i=1}^m P_i \log_2 P_i \quad (4-2)$$

In the formula:

$$P_i = C_i / \sum_{i=1}^m C_i$$

C_i ——Quantity of the element

2) Divides into many kinds the sample space according to the other attribute of the sample.

$$S = \{S_1, S_2, \dots, S_v\} \quad (4-3)$$

In the formula:

S_i ——the example set having some attribute

$S_{ij} = S_i C_j$ Together of two times classifies occurring.

The information of mathematic expectation because the sample S divided the subset according to the attribute:

$$E(a_1) = \sum_{i=1}^v \frac{S_{1j} + S_{2j} + \dots + S_{mj}}{S} I(S_{1j}, S_{2j}, \dots, S_{mj}) \quad (4-4)$$

$$I(S_{1j}, S_{2j}, \dots, S_{mj}) = -\sum_{i=1}^m P_{ij} \log_2 P_{ij}$$

$$P_{ij} = S_{ij} / |S_j| \quad (4-5)$$

3) Carries on the classification to sample attribute a_1 , judging the information gain obtains from a_1 :

$$Gain(a_1) = I(S_1, S_2, \dots, S_m) - E(a_1)$$

With this method computation the information gain of each attribute finding the max[Gain(A)] and founding a note.

B. Example analysis

The customer information of CRM system database in some Auto sales company is shown in table 1. Search the entire interior and exterior message related with the service object from the data and choose the data which is suitable for the knowledge discovery application. Then select the data from the customer's sex, the customer age section, the income and the school record.

TABLE I THE CUSTOMER INFORMATION OF CRM SYSTEM DATABASE
IN SOME AUTO SALE COMPANY

Serial number	Sex	Customer age	Education	Income	Whether to buy
1	Female	≤35	Below technical college	Middle	Yes
2	Male	≤35	Master or Doctor	Middle	Yes
3	Female	≤35	Master or Doctor	High	Yes
4	Male	≤35	Undergraduate	Middle	No
5	Male	≤35	Below technical college	High	Yes
71	Male	≤35	Undergraduate	High	Yes
72	Male	36~50	Undergraduate	High	No
73	Female	36~50	Master or Doctor	Middle	Yes
74	Female	36~50	Below technical college	High	No
75	Male	36~50	Undergraduate	High	No
76	Male	36~50	Master or Doctor	Middle	No
131	Male	36~50	Master or Doctor	High	No
132	Male	>50	Undergraduate	Middle	No
133	Male	>50	Below technical college	High	No
146	Male	>50	Undergraduate	Middle	No
147	Female	>50	Below technical college	Middle	Yes
199	Female	>50	Below technical college	Middle	No
200	Female	>50	Below technical college	High	Yes

Using decision tree method to analyze customer data

Because the initial time belongs “yes” and “no” the kind of sample integer respectively is 95 and 105, Therefore, the entropy of the initial value of time:

$$I_{(x)} = -\frac{95}{200} \log_2 \left(\frac{95}{200} \right) - \frac{105}{200} \log_2 \left(\frac{105}{200} \right) = 0.988$$

If the income was selected as the test attribute now the expect information:

$$E_{(x/income)} = \frac{113}{200} \left[-\frac{78}{113} \log_2 \left(\frac{78}{113} \right) - \frac{31}{113} \log_2 \left(\frac{31}{113} \right) \right] + \frac{87}{200} \left[-\frac{20}{87} \log_2 \left(\frac{20}{87} \right) - \frac{67}{87} \log_2 \left(\frac{67}{87} \right) \right] = 0.756$$

If the customer age was selected as the test attribute now the expect information:

$$E_{(x/customerage)} = \frac{71}{200} \left[-\frac{40}{71} \log_2 \left(\frac{40}{71} \right) - \frac{31}{71} \log_2 \left(\frac{31}{71} \right) \right] + \frac{60}{200} \left[-\frac{23}{60} \log_2 \left(\frac{23}{60} \right) - \frac{37}{60} \log_2 \left(\frac{37}{60} \right) \right] + \frac{69}{200} \left[-\frac{10}{69} \log_2 \left(\frac{10}{69} \right) - \frac{59}{69} \log_2 \left(\frac{59}{69} \right) \right] = 0.459$$

If the education was selected as the test attribute now the expect information:

$$E_{(x/education)} = \frac{63}{200} \left[-\frac{40}{63} \log_2 \left(\frac{40}{63} \right) - \frac{23}{63} \log_2 \left(\frac{23}{63} \right) \right] + \frac{63}{200} + \frac{63}{200} \left[-\frac{21}{63} \log_2 \left(\frac{21}{63} \right) - \frac{42}{63} \log_2 \left(\frac{42}{63} \right) \right] + \frac{77}{200} \left[-\frac{41}{77} \log_2 \left(\frac{41}{77} \right) - \frac{36}{77} \log_2 \left(\frac{36}{77} \right) \right] = 0.879$$

If the sex was selected as the test attribute now the expect information:

$$E_{(x/sex)} = \frac{126}{200} \left[-\frac{60}{126} \log_2 \left(\frac{60}{126} \right) - \frac{66}{126} \log_2 \left(\frac{66}{126} \right) \right] + \frac{74}{200} \left[-\frac{23}{74} \log_2 \left(\frac{23}{74} \right) - \frac{51}{74} \log_2 \left(\frac{51}{74} \right) \right] = 0.954$$

Calculate the information gain of each attribute value:

$$Gain(x, sex) = I_{(x)} - E_{(sex)} = 0.988 - 0.954 = 0.034$$

$$Gain(x, income) = I_{(x)} - E_{(x/income)} = 0.988 - 0.756 = 0.232$$

$$Gain(x, customerage) = I_{(x)} - E_{(x/customerage)} = 0.988 - 0.459 = 0.529$$

$$Gain(x, education) = I_{(x)} - E_{(x/education)} = 0.988 - 0.879 = 0.109$$

The influencing factor important size is proportional with its gain's size, so it can determine the size of its

importance. The gain size of the four influencing factors customer age > income > education > sex, therefore the sex should be selected as the test attribute.

III. CONCLUSION

Knowledge discovery is the non-trivial process which digs from large databases to obtain the correct new value and ultimate potential understandable models [11]. The knowledge discovery process and the calculation method had been mentioned in the article and analyzed by some examples. Believe that through the Knowledge Discovery in automotive marketing CRM systems application ,the cars' sales businesses can create features for the enterprise marketing solutions; expand marketing, and bring more benefits for enterprises

The function of knowledge discovery in the automobile marketing domain CRM system is huge. Effectively through a combination of the two can continue to encourage enterprises to enhance the value of individual customers and clients to expand the scale to effectively promote the value of the cars' businesses and strength continues to rise.

REFERENCES

- [1] Clark Glymour,David Madigan,Daryl Pregibon,Padhraic Smyth. Statistical Themes and Lessons for Data Mining[J]. Data Mining and Knowledge Discovery . 1997 (1)
- [2] Y.Feng,Z.H.Wu,Z.M.Zhou.Enhancing Reliability throughout Knowledge Discovery Process. ICDM'06 workshop of Reliability Issues in Knowledge Discovery (RIKD 06) . 2006.
- [3] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision . 2004 (2)
- [4] Li Chunlin,Li Layuan.A Utility-based Two Level Market Solution For Optimal Resource Allocation In Computational Grid. Proc.of the 2005 International Conference on Parallel Processing2005 . 2005
- [5] Hansen J V,Lowry P B,Meservy R D,et al.Genetic programming for prevention of cyberterrorism through dynamic and evolving intrusion detection. Decision Support . 2007
- [6] Dunham H.Data Mining-Introductory and Advanced Top-ics. . 2003.
- [7] Leila Ismail.Market Mechanisms for Trading Grid Resources. . 2008
- [8] Li Chunlin,Li Layuan.A Utility-based Two Level Market Solution For Optimal Resource Allocation In Computational Grid. Proc.of the 2005 International Conference on Parallel Processing2005 . 2005
- [9] Narendra K S,Mukhopadhyay S.Adaptive control of nonlinear multivariable systems using neural networks. Neural Networks . 1994
- [10] Hou Bin; Zha Jing; Liu Junmei; Luo Jian Central China Normal University; School of P.E.; Wuhan; China, "Application of KDD Techniques in Managing Sport Facilities," Proceedings of 2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, 2008-12-21