Vip Customer Segmentation Based on Data Mining in Mobile-communications Industry

ZHANG Yihua
Department of Information management
Business Administration college, Jimei University
Xiamen, Fujian, P.R.China,361021
yward@jmu.edu.cn

***Abstract*—The data mining is widely used in various industries with its powerful data processing and information mining capabilities. The CRM system has been used popularly, especially in the communications industry. The important component of the CRM system is how to segment customers, which is one of the bases to do the other decisions. An efficient and simple method of the mobile VIP customer segmentation makes the most use of the Customer Value and Customer Behavior model. It extracts the related data from a sample survey data to be the basis of the analytical process, then implements the operation of the mobile VIP customer segmentation and rule discovery, which has some practical values.**

***Index Terms*—data mining, VIP, customer segmentation, customer value, customer behavior**

I. INTRODUCTION

With the rapid development of the mobile market, mobile services are popular among consumers. In results, the mobile customers gradually present the characteristics of segmentation and diversification. Especially for the high-value VIP customers who have different consumption patterns and value-added service requirements, it is necessary to divided them into appropriate targets to provide effective differentiated services, then, on this basis, we can take differentiated marketing and customer service strategies, and provide a strong basis for the relevant departments to make business and service market strategies. Also, it is an effective way for the mobile operators to effectively improve customer loyalty, strengthen customer relationships management, and maintain a market leadership. [1~3]

II. THE THEORY OF CUSTOMER SEGMENTATION[4]

Customer segmentation (Customer Segmentation) was proposed by the American scholar Wendell • Smith in the 50's mid 20th century, which is such a behavior that divides the existing customers of the enterprise into different groups in accordance with a certain standard.[5] It is the product of the modern marketing concept, which is the new development of the marketing theories and strategies in the Western developed countries after World War II. Xuzanne Donner considers that the right customer segmentation can effectively reduce costs, meanwhile, get stronger and more profitable market penetration.

The so-called customer segmentation, mainly means that an enterprise intends to classify their customers in accordance with the value, needs and preferences and other comprehensive factors among the clear strategy, the business model and the focus market, then, it can provide targeted products, services and marketing patterns .The consumers who belong to the same sub-group are similar to each other, while those belong to different sub-groups are regarded as different. For example, those who brought about the most profit or the least have their own groups. The same sub-group of consumers can be regarded as similar for a variety of reasons, probably for their living areas, their incomes or their thinking and behavior.

When the theory of customer segmentation is applied in customer classification, the first problem to be solved is how to determine the customer segmentation variables, that is to determine to segment the customers by what variables. Different segmentation variables often produce different results. The common types of segmentation are: (1) geographic segmentation; (2) population subdivision; (3) mentality segmentation; (4) behavior segmentation; (5) according to the different interests which the customers pursuit from the products; (6) User status segmentation; (7) use rate segmentation. [6]

III. CUSTOMER SEGMENTATION

The prerequisite for the success of CRM is the ability to segment the customers of the enterprise based on the customer's business value, especially for the high-value VIP group. It can identify the customers by segmentation and target the different VIP customers, then provide effective differentiated services.[7~9]

A. Customer Segmentation based on data mining process

It is an effective and easy to implement method to use a customer value matrix to achieve customer segmentation. This paper effectively integrates the commonly-used clustering and decision-making algorithm of data mining and applies it to VIP customer segmentation .The concrete analysis is illustrated in Figure 2.

B. Cluster Analysis

Clustering is the process to group the physical or abstract objects and to classify the similar objects into a class. [10] Cluster analysis is to classify the physical or abstract objects into several groups, within each group, there are high similarity between objects, whereas the similarity between groups is relatively low. There are thee following several steps according

to the cluster analysis application of mobile client: identification of sub-themes, data preparation, selection clustering model and Cluster analysis release.

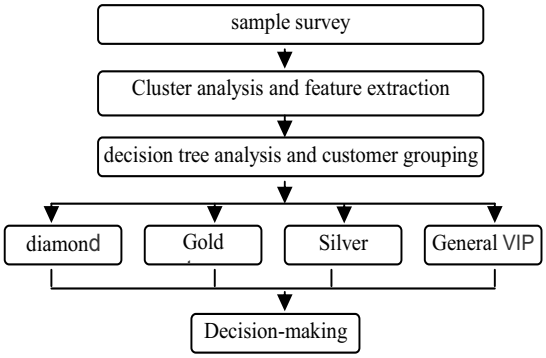

Figure 1. VIP Customer Segmentation Process

1) Determine the sub-theme

Objective of this study is the customer segmentation based on the value of big customers, and consumption features are the main features, while the natural properties are secondary characteristics. So it needs to extract the data which is related to the basic information of the customers and the customer's behavior.

2) Data understanding

There are 230 survey questionnaires were distributed in this sample survey, and Valid 220 are valid. The contents of the questionnaire include consumer behavior characteristics of mobile customers, demand characteristics, the mobile services satisfaction ratings for the mobile services, and the attitude towards the various services. These data are completed or the answer sorting from the customers, which is of high reliability.

3) Data preparation

Data can be divided into two categories: original data (which can be directly extracted from the basic information of the customer, such as gender, age, etc.) and statistics data (such as: gender proportion, total telephone charges, etc.).

4) Data Preprocessing

As the mobile industry has a great customer volume, in the design of the questionnaire and the survey ,it is important to require the research staff to ask the customers to fill out complete information and process the missing values and abnormal values ,such as: empty customer attribute data ; eliminate the customer who" don't know" the research problems. These questionnaires have been discarded while recycled, and then after data preprocessing, it obtained 138 finally.

5) Modeling

In the original k-means algorithm, we noticed two issues: (1) in the k-means algorithm, the choice of initial cluster centers is arbitrarily chosen; this will lead to unstable clustering results. (2)Within one iteration, each data object assigned to the class of the nearest cluster center and the time complexity of this process is O (nkd). It is a great time when the data is larger. In view of the above two problems, this research adopt a modified k-means cluster analysis method:

a) Select K initial cluster centers: Conduct J times data searches to set S, get K similar cluster centers. Search (S, K, J) input data sampling frequency of J, the data set S, the number of classes K; output K initial cluster centers.

b) Reduce the number of iterations: Use the triangle trilateral idea that both sides of the triangle is greater than the third side to reduce the number of each iteration of the calculation, which simplifies the comparison process calculation

6) Customer Segmentation Results

a) Select the K value of 10: Select the telephone detailed bills and the total consumption from the VIP customer survey sample data, after the cleaned-up process , resulting in 138 customer records. Analyze these consumptions, then it can subdivide them into 10 customer group by using the improved K-means algorithm in Table I.

We can see from this result: Class 8 customers have a high proportion of long distance, class 1 and 4 have a higher proportion of roaming. Class 7 has a high proportion of both long distance and roaming.

b) Based on customer value matrix, select the 220 sample whose K value is 4 after the cleaned-up process, and do clustering analysis of the 138 customer records

TABLE I. CUSTOMER SEGMENTATION RESULTS (K = 10)

cluster No	Amount	Total amount	Rent of month	local call	l.d.tel.	ramble	increment	l.d.tel. rate	Ramble rate
1	12	419.64	106.01	95.83	38.47	150.48	22.76	9.17%	35.86%
2	22	265.95	26.86	91.46	45.83	44.57	39.72	17.23%	16.76%
3	11	336.15	40.98	152.94	53.09	56.19	38.61	15.79%	16.72%
4	12	397.15	67.66	158.62	53.37	105.52	38.79	13.44%	26.57%
5	28	140.85	25.26	57.49	25.92	27.52	27.75	18.40%	19.54%
6	15	457.40	42.83	171.27	89.98	60.84	47.33	19.67%	13.30%
7	5	1161.10	201.55	113.53	287.08	370.68	50.32	24.72%	31.92%
8	6	550.71	92.38	216.69	120.23	84.08	31.70	21.83%	15.27%
9	13	600.62	101.51	200.64	66.58	117.11	61.83	11.09%	19.50%
10	14	705.73	63.61	243.73	120.59	129.58	114.45	17.09%	18.36%

Assumpt Vl, V2, V3, V4 to represent the diamond card customers, Gold, and Silver customers, the general VIP customers. Cluster the customers into four clusters, and the number of each cluster, the number of contributions to the company's revenues and its ratio shown in Table II.

TABLE II. CUSTOMER SEGMENTATION RESULTS (K = 4)

Customer cluster	amount of customer	Amount rate	Telephone fee	income ratio
V4	42	30.43%	186.18	10.52%
V1	28	20.29%	706.61	39.91%
V3	38	27.54%	361.30	20.41%
V2	30	21.74%	516.50	29.17%

7) Model application and marketing strategy

According to the above "select the K value of 10," the results of the actions taken are as follows:

a) For class 7,8,10 customers, it can design long-distance preferential packages to encourage customers to make more long-distance calls.

b) For class 2, 3, 5 customers, it can encourage them to continue to use the original Package.

c) For class 1, 4, 7 with higher roaming charges, it can recommend them to use business travel cards or packages, and provide high-quality "easy boarding", "VIP area" and so on.

d) While for class 1O class with a high value-added cost, it can recommend them to use value-added service package subscription packages.

It is the main and the most direct method to adopt big customer segmentation, customer value analysis and demand analysis to administrate and serve the big customers. We can provide personalized, differentiated service for the big customers by a large customer-oriented marketing (CRS) integration of external and internal resources, and the implementation of One to One Marketing. The above simple strategies do more than enhance the degree of customer care from the mobile operators, while more effectively promoting their own brand image, customer perception and loyalty.

At present, under the knowledge economy environment of the increasingly fierce competition, customer relationship is not longer a simple telephone care, a message or a decision will influence the existing client and the potential customers of the enterprise, and decide the fate of the companies. How to retain the old customers and to develop new customers is crucial for any business.

IV. DISCOVERY RULES WITH DECISION TREE

Decision tree is a simple knowledge representation; it will gradually classify cases, representing different categories. As the classification rules are more intuitive and therefore it is easier to understand, this approach generally limited to the classification task. Decision tree is similar to the flow chart of the tree structure, where each internal node represents an attribute in the test, each branch represents a test output, and each leaf node represents a class or class distribution. Decision tree provides a display similar to under what conditions will get the certain value of these Rules. [11 ~ 12]

A. Application of decision tree for ulterior segmentation

Regard the "customer class" as the main attribute in the design, select 7 attributes which represents the features of the clients : family income, position, personal income, gender, occupation, age, educational level, and then according to the decision tree construction process to get the Figure 2. (I class customers - Diamond card customer, Ⅱ class customers - Gold customers, Ⅲ class customers - Silver customers, Ⅳ class customers - General VIP customers).

B. Rule Discovery

The decision tree produced a total of 26 information as a part of the following (in case I, type of customer):

1) IF principalship<=1 AND sex<=1 AND age<=55 AND occupation>2 AND age>42 AND educational level<=3.

2) IF principalship<=1 AND sex<=1 AND age<=55 AND occupation>2 AND age>51 AND educational level>3.

C. Marketing Strategy

According to the above analysis of the results of class I, it can be drawn: no matter how the customer is educated, family income and sex, aged between 42 to 55 year-old customers, all belong to class I. As class I is currently an important source of profits, so businesses must attract and retain the customer relationship and apply marketing strategy to the age of 42 to 55 year-old clients, but for the customers outside of the range, it needs to carry out selective promotions.

V. CONCLUSION

T Through analysis of the communications industry's current situation and the problems of the traditional customer relationship management, it boldly applied the data mining technology to the customer relationship management of mobile communications industry, and through practical application and research, it proves its effectiveness and practicality in the mobile industry and for provided strong technical support for the management and decision-making of the operators.

It needs to continuously improve customer service satisfaction, loyalty, business use viscosity to achieve the transformation from the mobile communications experts to the mobile information experts. An analysis of how to improve the management and service of the big customer has a direct practical significance of "development of new big customers, win-back of the lost big clients, stabilization of existing big customers and a maximization of the customer value".

REFERENCES

[1] Rui Xu, Donald Wunsch II. Survey of Clustering Algorithm[J]. IEEE Transactions on Neural Networks, 2005, 16(3), P645~678.

[2] Raymond T. Ng and Jiawei Han. CLARANS: A Method for Clustering Objects for Spatial Data Mining[J]. IEEE transactions on knowledge and data engineering, 2002,14(5), P1003~1016.

[3] Liu Xiaodong, Chen Zhiping, Customer Segmentation based on Data Mining Technology[J], Journal of Shenzhen Institute of Information Technology Vol.6 No.4 Dec. 2008).

[4] Yan Liping, The Analysis and Design of Telecom Customer Segmentation Model Based on Data Mining[J], Science Mosaic,2007. 5, P89.

[5] Marcus,Claudio. A Practical Yet Meaningful Approach to Customer Segmentation[J].Journal of Consumer Marketing, 1998, 15(2).

[6] Xiao Sheng, Li Aimei, Commercial Customer Segments and Different Marketing Ideas [J]. P&T Enterprise Management, 2004,11 (7).

[7] Huang Xiaoyan, Based on Data Mining of Telecommunications Customer Segments, SCIENCE & TECHNOLOGY INFORMATION TECHNOLOGY 2007, No.29, P65 ~ 66.

[8] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques. Machinery Industry Press, 2001.

[9] M. Kantardzic, Data mining - Concepts, Models, Methods and Algorithms. Tsinghua University Press, .2003.

[10] Zhang Xiaohang, Customer Segmentation based on Clustering algorithm[J], P&T Enterprise Management , 2005; (12),P74~77.

[11] Guo Xiujuan, Summary of Data Mining Methods[J], Journal of Jilin Architectural and Civil Engineering Institute, Vol. 21, No. 1, March 2004:,P50~53.

[12] HAN Jiang,Data Mining--a New Area with Best Development Potential[J], Journal of Suzhou Vocational University,Vol.15,No.1, February 2004,P56~58.

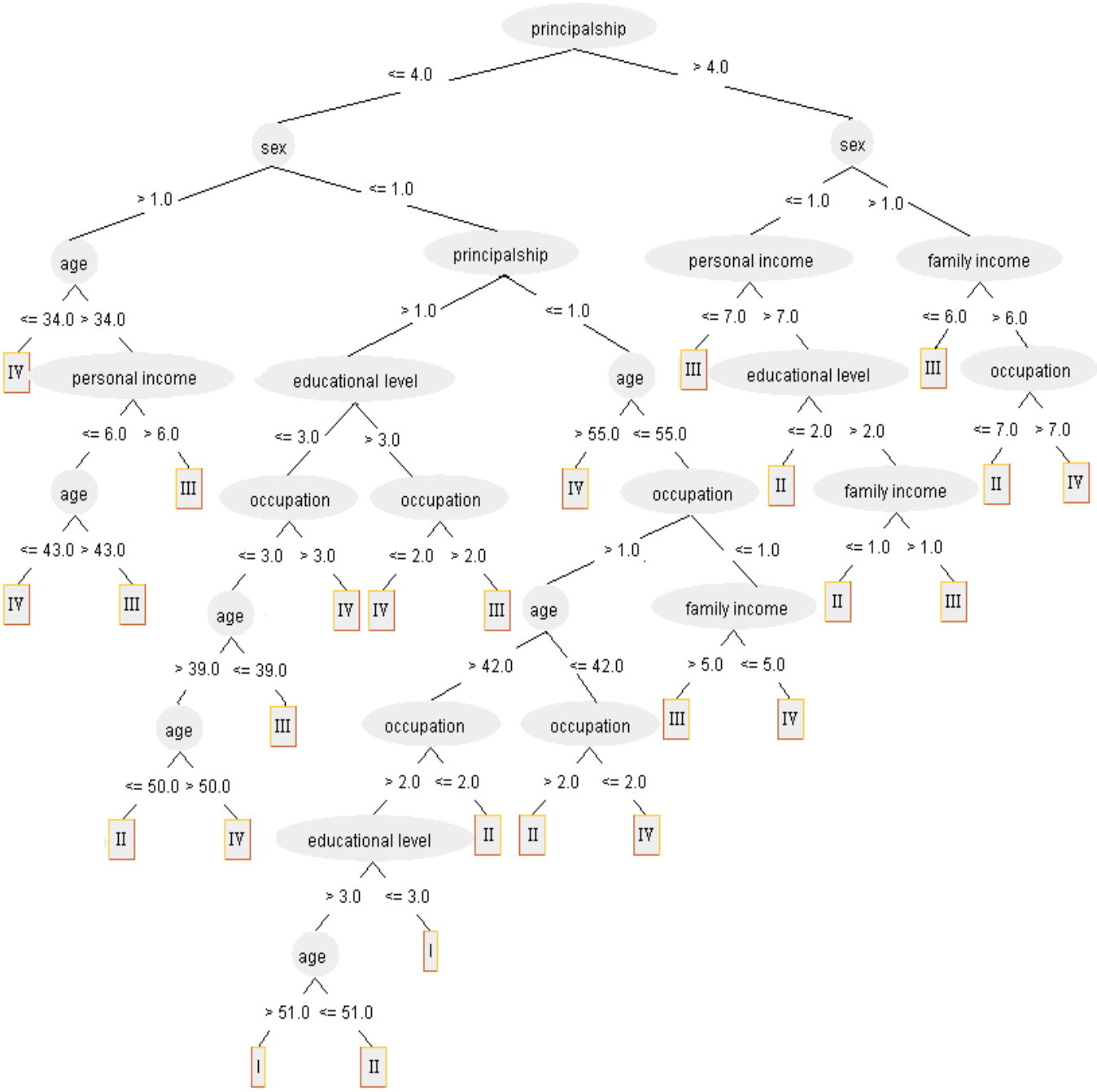

Figure 2. The decision tree from the analysis of the customer attributes