Customer Segmentation Based on Survival Character

Guozheng Zhang
College of Business, Houzhou Dianzi University
Hangzhou, P.R.China
(E-mail: guozhengzhang@gmail.com)

***Abstract*—"Customer Segmentation" is an increasingly pressing issue in today's over-competitive commercial arena. More and more literatures have researched the application of data mining technology in customer segmentation, and achieved sound effectives. But most of them segment customer only by single data mining technology from a special point, rather than from a systematical framework. Furthermore, one of the key purposes of customer segmentation is customer retention. Although previous segment methods may identify which group needs more care, it is unable to identify customer churn trend for adopting different action. This paper focus on proposing a customer segmentation framework based on data mining and constructs a new customer segmentation method based on customer survival character. The new customer segmentation method constituted of two steps. Firstly, using K-means clustering arithmetic cluster customer into different segments which has the similar survival function (churn trend) inside. Secondly, the method uses survival analysis to predict each cluster's survival/hazard function to test the validity of clustering and identify customer churn trend. The method mentioned above was applied in a dataset from China Telecommunications Company, which result proposes some useful management measures and suggestions. This paper also suggests some propositions for further research.**

Keywords-Customer segmentation; Data mining; Survival character; Survival function; Survival analysis

I. INTRODUCTION

Over the past decade, there has been an explosion of interest in customer relationship management (CRM) by both academics and executives[15]. Organizations are realizing that customers have different economic value to the company, and they are subsequently adapting their customer offerings and communications strategy accordingly. Currently research demonstrates that the implementation of CRM activities generates better firm performance when managers focus on maximizing the value of the customer[5]. A deeper understanding of customers has validated the value of focusing on them. Customer segmentation is one of the core functions of CRM. Customer segmentation is the base of how to maximize the value of customer[12]. Again and again firms find that the Pareto principle holds true, with 20% of the customer base generating 80% of the profits. Both researchers and managers need to evaluate and select customer segmentations in order to design and establish different strategies to maximize the value of customers.

Generally, customer segmentation methods mostly include experience description method, traditional statistical methods, and non-statistical methods [3, 9, 12]. Non-statistical methods mainly are arisen application of data mining technology in segmentation [1, 18, 13]. Jaesoo Kim etc (2003) researched the application of ANN in tour customer segmentation[7]. Fraley, C. and Raftery, A.E (2002) researched the application of clustering approaches in customer segmentation[4]. These literatures use single data mining technology analyses single business issue, and have got some good effectives. But these applications have two obvious shortages: ①These applications of data mining technology in customer segmentation are separate, and didn't construct a systematical framework for guiding segmentation issues. ②One of key purpose of customer segmentation is customer retention, previous segmentation methods may be able to position which segment need more

1-4244-1312-5/07/$25.00 © 2007 IEEE

care, but it is unable to identify customer churn trend for take effective actions for different customers cluster.

The purpose of this article is to propose a framework for customer segmentation and constructs a new customer segment model for new customer retention way. The framework presents the whole process of segmentation, which include data acquiring, issue mapping, segmentation model selecting, data mining arithmetic selecting, and function analysis. It gives out the way to guiding data mining technology used in customer segmentation. The new customer segment model is called Customer Segmentation Method Based on Survival Character. Its effectiveness lies not only in that it identifies key customer segments, but also in that it is able to predict the customer hazard/survival probability (churn trend). Enterprises can make effective customer retention action based on hazard/survival probability (churn trend) of each customer segment. And after segmentation, each customer's hazard/survival probability (churn trend) in one segment is approximately same level, so when calculating customer value, we can use the customer retention rate instead of customer hazard/survival probability is reasonable.

II. CUSTOMER SEGMENTATION FRAMEWORKS

As practitioners are enthusiastically seeking out groups of profitable customers whose loyalty is steady, some academics are beginning to question whether segments are actually stable entities and more fundamentally whether they really exist at all[8]. The customer segmentation based on data mining can solve this problem because the ways could study from new information that input afterward and get new rules. It provides completely support to the dynamic management process of customer acquiring, customer keeping and customer value increasing, customer satisfaction and customer loyalty promoting.

Building the mapping relationship between the conception attribute and the customer is the key step of segmentation method based on data mining. Customer data contain dispersive and continues attribute. Setting each customer attribute as a dimension and setting each customer as a particle, the whole customers in enterprise can form a multidimensional space, which has been defined as the attribute space of the customer.

The mapping relationship between customers attributes and conception category can be constructed by analytic method, or by sample learning method. Analytic method analyzes the attribute character of each conception category should have, subsequently constructs the mapping relationship between attributes space and conception space. But much mapping relationship between attributes space and conception space weren't clear, it need use sample learning method constructing the mapping relationship [1]. Sample learning method automatically generalize the mapping relationship between attributes space and conception space by applying data mining technology on known conception category in enterprise database. The data mining process is called sample learning.

Assume $B=\{G_1,G_2,\cdots,G_n\}$, we can confirm a group conception category by B, $L=\{L_1,L_2,\cdots,L_p\}$, $C'\subseteq C$, C' is known customer category.

The three steps of customer segmentation framework based on data mining：

(1) Ensuring mapping route. $p:C\rightarrow L$, set $\forall c\in C'$, if $c\in L_i$, then $p(c)=L_i$.

(2) $\forall c\in C$, confirming which conception category c belong to by seeking the value of $p(c)$.

(3) Rule creation and Function analysis. As showed in figure.1. The segmentation model applied data mining technology such as Association rule, Neural network, Decision tree, etc, as its method foundation. This segmentation model firstly segment customer according to the mapping relationship, subsequently processing various kind of business application.

1) Segmentation rule creation

Sort the customers with the Customer Segment Model and the Segment Function. After training the segment model, we get the segment rule or the network segment rule. We can effectively segment new customers based on the trained model.

2) Function Analysis

Function analysis includes customer value analysis, credit analysis and promotion analysis, etc, based on the foundation of the mapping relationship between customer and concept. Further on, new function requirement will be bring forward to CRM with the developing of the management practice. The new demand functions would be added to the conceptual dimension and reconstruct the mapping relationship with the customer characteristic.

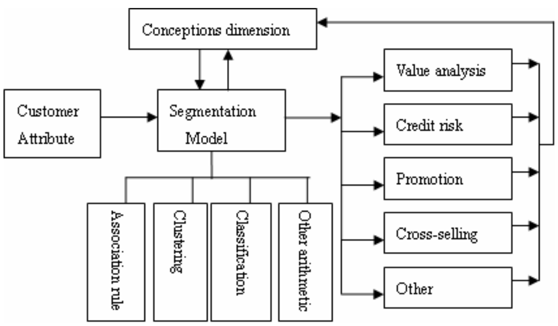

Figure 1. *C*ustomer Segmentation F*ramework* Based on Data Mining

III. CUSTOMER SEGMENTATION MODEL BASED ON CUSTOMER SURVIVAL CHARACTER

A. Survival Analyses

Survival analysis is a clan of statistical methods for studying the occurrence and timing of events [6]. From the beginning, survival analysis was designed for longitudinal data on the occurrence of events. Keeping track of customer churn is a good example of survival data. The best observation plan is prospective. We begin observing a set of customers at some well-defined point of time (called the origin of time) and then follow them for some substantial period of time, recording the times at which customer churns occur. It's not necessary that every customer experience churn (customers who are yet to experience churn are called censored cases, while those customers who already churned are called observed cases).Survival analysis was designed to handle survival data, and therefore is an efficient and powerful tool to predict customer hazard/survival (churn trend).

B. Segmentation based on customer survival character

The method consists of cluster technology and survival analysis, as shown in figur2. The customer character for clustering is extracted according to industry and customer behavior attributes. Take the telecommunication industry as example, customer's communication hours, times and expense in recent several months are widely used for churn forecast and they are explainable exactly. So this paper uses these behavior attributes as segmentation variables in empirical analysis.

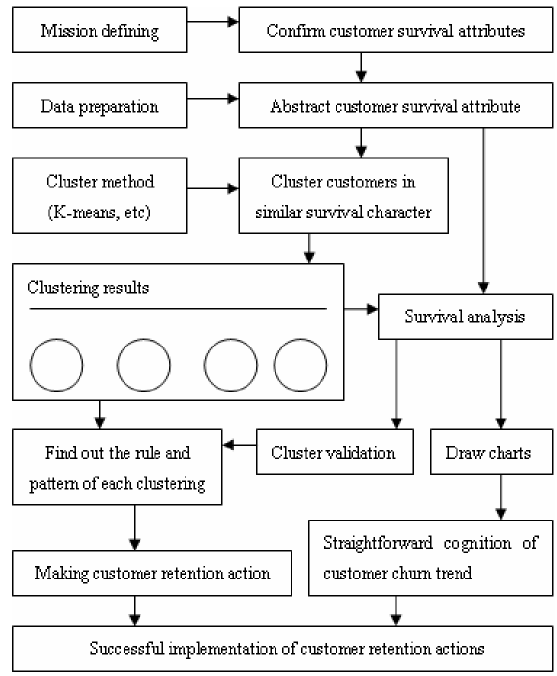

Figure 2. Customer Segmentation based on survival character

C. Steps

(1) Mission defining. Confirming customer attributes for mission.

(2) Data preparation and Data abstraction. Abstracting necessary customer behavior attributes from data warehouse.

(3) Clustering. Using K-means to segment customer based on similar hazard/survival possibility. After clustering, each customer has a new attribute named cluster number.

(4) Survival analysis. It just needs three necessary attribute: Months with service, Customer number and clustering number. Survival analysis have two purpose: ①Segmentation's quality can be measured by homogeneity in subdivisions and heterogeneity among subdivisions[14]. It can make comparer between different clusters to test clustering performance;

②Drawing survival function curve for straightforward cognition of customer churn trend.

(5) Find out the rule and pattern of each clustering.

(6) Successful Implementation of customer retention action

IV. CASE: CUSTOMER SEGMENTATION IN TELECOM

A. Data Selected and Filtering

A china telecom firm, which throngs more and more customers, but on the other hand, strives with stronger competition, is our studying case. The company wants to make decision to satisfy customers, and prevent customer churn. In collaborate with the company, it supplies with us a part of the data, aiming to predict feature of customers churn. Firstly customer data are selected and filtering, and delete some insignificant records, such as register without transaction records. In the end, we select 1000 records from data warehouse, each record including256 attributes.

Attributes list as follow: customer's basic information (name, gender, register date), churn flag, customer's transaction records from first month to sixth month after register（such as total numbers each month, total fee, the number of calling in and calling out each month, fee in every month, roaming about each month, the number of note, the number of calling in and calling out in working day, the number of transactions in one net and the time, the number of transactions among nets and the time）, the number of customers' consultation, etc..

In the selected dataset, 27.4% customers occur with churn among all customers, namely 72.6% shares with our service. In this paper, we don't list all of attributes, a part of attributes have been listed as showed in Table1.

B. clustering customer based on similar survival character

A K-means cluster analysis was performed. According to former literatures with the prediction in customers churn, we selected and filtering 196 attributes to cluster, and omitting some of attributes, like worthless attributes or inapplicable in K-means, especially like sparse datum [10]. Generally, the parameter must be appointed in the beginning with K-means. Now according with advanced experience, the parameter K is set as 2-6.

The studying uses the software of SPSS, the result finally arrived at four clustering .Parts of clustering results as showed in Table2

C. Verifying the clustering

The aim to estimate the survival and hazard function help us obtain customer survival/churn information. Necessary variable: Months with service, Churn flag, Customer ID. Firstly, these cluster were been pairwise compared. As showed in Table3, their difference is distinct (all sig. <0.10, almost all sig. <0.05). And we get survival function(churn trend) , as shows in Figure3。

TABLE I. CUSTOMERS' BEHAVIOR ATTRIBUTES

Attributes	Explains
tenure	Months with service
age	Age in years
marital	Marital status
address	Years at current address
ed	Level of education
employ	Years with current employer
retire	Retired
gender	Gender
reside	Number of people in household
Fields Name	description
Churn flag	churning flag
Total_disc_fee	Total discount in the past six months
Total_Times_1	Total number of transaction in first month
Total_Times_2	Total number of transaction in second month
Total_Times_3	Total number of transaction in third month
…	…
Total_Times_6	Total number of transaction in sixth month
Total_Duration_1	Total duration in first month
Total_Duration_2	Total duration in second month
Total_Duration_3	Total duration in third month
…	…
Total_Duration_6	Total duration in sixth month
…	…

TABLE II. CLUSTERING RESULTS

Customer ID	Churn flag	Cluster ID	Number of customers in each cluster
1	1	cluster2	Cluster1:194 Cluster2:264 Cluster3:269 Cluster4:273
2	1	cluster4	
3	0	cluster4	
4	0	cluster2	
5	1	cluster3	
6	0	cluster2	
7	1	cluster4	
…	…	…	

TABLE III. PAIRWISE COMPARISONS

Cluster ID	Compared cluster ID	Wilcoxon (Gehan) Statistic	Sig.
1	2	18.640	.000
	3	37.154	.000
	4	2.949	.086
2	1	18.640	.000
	3	5.515	.019
	4	9.222	.002
3	1	37.154	.000
	2	5.515	.019
	4	27.229	.000
4	1	2.949	.086
	2	9.222	.002
	3	27.229	.000

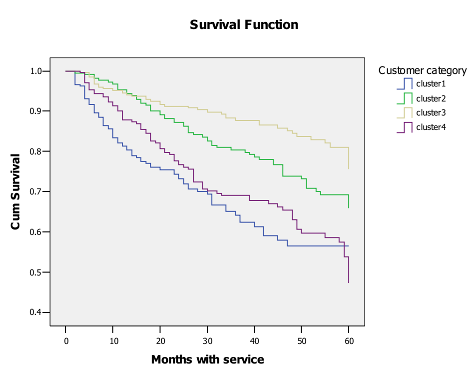

Figure 3. survival function

V. MANAGEMENT APPLICATIONS

With the analysis to all kinds of features, we can find the rule and pattern as showed in Table4, and in the end find relevant market tactics.

TABLE IV. CLUSTER FEATURES

class	
1	More fees, more long-distance transactions, highest churn
2	Normal fees, second lowest churn
3	Little transaction each month and more note, lowest churn
4	More transactions each month, more call in and less call out, second highest churn

Clustering1: these customers share important similar features, such as more fees, more long-distance transactions and highest churn. As these customers expend more, and their expectation is high, we must hold up them with more resource, so as to decrease the churn. And as showed in Figure3, the churn trend is much quicker in the beginning phase that latter phase, so firms must adopt retention actions as soon as possible.

Clustering2: generally, these customers hold average expenditure each month, and every numeric displays equilibration, not higher or lower. The churn is low comparatively. Relatively, these customers satisfy with the company's service, and they hope to share with the company's service. The company must launch into appropriate resource to hold them, for they are the foundation of the customers.

Clustering3: these customers' expenditure is lowest comparatively, and the probability of churn also is also lowest comparatively. They satisfy present services provided with the company, but contribution is lower than average level. Company needn't distribute any resource on them.

Clustering4: these customers have more spending, with more call in and lower call out. They are valuable customers, but their churn is also high. Company must allocate some resources to retain them and encourage them to call out more.

VI. CONCLUSIONS

A key role of marketing is to identify the customers or segments with the greatest value-creating potential and target them successfully with corresponding marketing strategies to reduce the risk of these high lifetime value customers defecting to competitors[2]. Segmenting customer is the basic work of data mining according to known historic segmentation information. The training data used to construct segment forecast mode can be historic data or exogenous data that gain from experience or survey.

For an enterprise, how to use data mining, and how to practice enterprise's tactics should we use in determine segmentation? To answer this question, this paper proposes a segmentation framework based on data mining and constructs a segmentation methods based on customer survival character. The segmentation framework describes the commonly form of customer segment, it can guide for application of data mining in customer segment. Customer segment based on customer survival character clustering customer with similar churn trend, each clustering have its unique churn trend, so enterprise can make retention action according to this. And via observe the sharp of survival function curl, enterprises can get straightforward cognition of customer churn trend.

Effective segmentation can help companies increase revenue by acquiring and retaining high value customer at low cost. It can also help in aligning cost-to-serve to customer value, perhaps reducing overall marketing, sales and service costs. The segment methods of this paper is testing in the telecomm industry, it may be used in other industry such as finance service etc. Therefore, the future researches may focus on testing this segmentation model in other service industry.

ACKNOWLEDGEMENTS

The Research is supported by the Shanghai Shuguang Project "Research of Customer Value Analysis and Management Strategy Based on Data Mining Technology" under the grant No: 05SG38.

The Research is also supported by the Young Teacher Research Start-up Project "Research of Customer Lifetime Value Management System Basded on Data Mining Technology" of Houzhou Dianzi University.

REFERENCE

[1] Agnes Nairn, and Paul Bottomley (2003): Cluster analysis procedures in the CRM era. International Journal of Market Research, Vol. 45 Quarter 2

[2] Andrew Banasiewicz, Acquiring high value, retainable customers, *Database Marketing & Customer Strategy Management*, 2004, Vol. 12, 1, 21–31

[3] E. H. Sub, K .C. Noh, C .K.Suh (1999): Customer list segmentation using the combined response model. Expert Systems with Applications,17(2): 89-97

[4] Fraley, C. and Raftery, A. E. (2002): "Model-Based Clustering, Discriminant Analysis, and Density Estimation". Journal of the American Statistical Association 97:611-631.

[5] Gupta, Sunil, Donald R. Lehmann, and Jennifer A. Stuart (2004): Valuing Customers. *Journal of Marketing Research*, , 41 (February), 7–18.

[6] Hosmer, JR. DW, and Lemeshow S(1999): Applied Survival Analysis: Regression Modeling of Time to Event Data. New York: John Wiley & Sons,.

[7] Jaesoo Kim, et al.(2003): Segmenting the market of West Australian senior tourists using an artificial neural network. Tourism Management, 24(l):25-34

[8] Jon Kleinberg, Christos Papadimitriou, Prabhakar Raghavan (2004): Segmentation Problems. Journal of the ACM, Vol. 51, No. 2, March, pp. 263–280.

[9] Lei-da Chen, Khalid S. Soliman, En Mao, Mark N. Frolick (2000): Measuring user satisfaction with data warehouse: an exploratory study. Information & Management, 37(3): 103--110.

[10] Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2006). Introduction to Data Mining. Posts & Telecom Press,pp310-320, Beijing.

[11]Per Vagn Freytag, et al(2001): Business to Business Market Segmentation Industrial Marketing Management，30(6):4 73486

[12]Yun Chen, Guozheng Zhang, Dengfeng Hu and Shanshan Wang (2006), Customer Segmentation in Customer Relationship Management Based on Data Mining, *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management,* Volume 207, pp.288-293

[13]Verhoef P.C., Spring P.N., Hoekstra J.C. The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. Decision Support Systems, 2 003,34(4):471-481

[14] Wedel, M., W.A. Kamakura and U. Bockenholt (2000): "Marketing data, models and decisions," International Journal of Research in Marketing 17(2-3) 203-208.

[15] Werner Reinartz, Manfred Krafft, And Wayne D. Hoyer(2004): The Customer Relationship Management Process: Its Measurement and Impact on Performance, *Journal of Marketing Research*, 293 Vol. XLI (August), 293–305）