Customer Segmentation Using CLV Elements

Mitra Bokaei Hosseni, Mohammad Jafar Tarokh

Information Technology Group, Industrial Engineering Department, K. N. Toosi University of Technology, Tehran, Iran.
Email: mitra.bokai@sina.kntu.ac.ir, mjtarokh@kntu.ac.ir

Received June 22nd, 2011; revised July 28th, 2011; accepted August 4th, 2011.

ABSTRACT

To have an effective customer relationship management, it is essential to have information about the different segments of the customers and predict the future profit of them. For this reason companies can use customer lifetime value that consists of three factors—current value of customers, potential value, and customer churn. Potential value of customers focuses on the cross-selling opportunities for current customers. Therefore, cross selling models are built on the total customers of the database that is not interesting. To overcome this, we presented a framework that estimates the current value and churn probability for the customers and then segmented them based on these two elements and select the most profitable segment for the cross-selling models. In this study we predict the customer churn base on logistic regression as a case study on the insurance database.

Keywords: CRM, Customer Lifetime Value, Churn Prediction, Cross-selling, Logistic Regression

1. Introduction

Life insurance is become one of the popular insurances in recent years. It is divided into many categories which each of them delivers different services to the customer. Customers agree to pay for the insurance in different ways. General way to pay for the agreed amount is monthly payment. Therefore, it is important for the insurance company to know about its customers and their payment styles to manage its relationships with them. The insurance company must have the information about which customers are likely to leave the company or are likely to pay their loans not in the defined time. It also must understand its loyal customers and use different marketing strategies to retain them.

Some customers repeatedly switch providers, or “churn”, it is also obvious in insurance industry. Different methods for churn rate measurement were mentioned in the literature. Companies can use data mining techniques to identify the characteristics of the customers who will remain loyal or the churners. In insurance industry rapid customer churn is a significant problem due to the competitive environment of this industry. Wu et al. used decision rules and data mining to investigate the potential customers for an existing or new insurance product [1]. These methods enable companies to invest in customers who will produce the most profit for the company.

Many life time value models were presented in order to evaluate the customer value in its lifecycle. Each of them had specific characteristics that were suitable for special industry. The usage of these models also depends on the available data about that industry. Customer value can be identified by three factors, current value, potential value and churn rate and also by socio-demographic data from the customers. Up-selling, cross-selling, and customer retention is defined as the three core activities for increasing the customer value [2].

Current cross-selling models were being built on the total customer database of the organization. This leads to a high overhead for building cross-selling models because the whole database contains also the data from the unprofitable customers and churners. Also the current LTV models needed data from different products that a customer had for a period of time to calculate the potential value of the customers. Some organizations lack from these kinds of data. To overcome these problems mentioned, it could be interesting to build a cross-selling model on the loyal customers with appropriate current value.

LTV is based on the understanding of the behavior of the profitable customers of the organization. For this reason, this study proposed a model to understand the profitable segments of the customers based on the LTV components and socio-demographic data of customers. First we evaluated the current value of each customer based on the transactional data. Then the churn rates were evaluated and loyal customers were predicted with the use of logistic regression. The customers were segmented
based on the current value and customer loyalty that were calculated. After analyzing and selecting the profitable segments organizations can use the cross-selling strategies to develop the relationship between customers and the organization. For this reason, customers’ data should be collected for evaluating the life time value for each customer.

In this paper we presented a new model for understanding the customer behavior based on the customer lifetime value. We built a cross-selling model on the loyal customers with appropriate current value. We also examined this model to data that was collected from an insurance company. Logistic regression was selected to predict the customer behavior based on these data. This paper is organized as follows. Section 2 presents an overview about life time value literature and related works and also review of mathematical model we used for prediction. Section 3 specifies the model and section 4 evaluates the model based on the real data collected from the insurance company. Section 5 includes the conclusion of the research.

2. Literature Review

This section is divided into three parts, first we review the CRM dimensions and mention which dimension we are going to focus. Second, we review the LTV definitions and models. Third, we describe the mathematical models for prediction we use for this paper.

2.1. CRM Dimensions for Customer Behavior Classification

To manage the different segments of customers managers can use customer relationship management (CRM) as a leading business strategy in a competitive environment, while retention of the current customers in a competitive environment is vital for survival of the companies. CRM pursues long term relationship with profitable customers. There are different definitions in the literature for the customer relationship management, that we mention some of them here. Ling & Yen believe that CRM comprises a set of processes and enabling systems supporting a business strategy to build long term, profitable relationships with specific customers [3]. Swift defined CRM as an enterprise approach to understand and influence customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability [3]. Parvatiyar and Sheth defined CRM as a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply chain functions of the organization to achieve greater efficiencies and effectiveness in delivering customer value [4]. Kincaid viewed CRM as the strategic use of information, processes, technology, and people to manage the customer's relationship with your company (Marketing, Sales, Services, and Support) across the whole customer life cycle [5]. Lawson-Body & Limayen believe that CRM refers to all business activities directed toward initiating, establishing, maintaining, and developing successful long-term relational exchanges and it is the set of methodologies and tools that help an enterprise manage customer relationships in an organized way [6]. All these definitions emphasized on the importance of customer acquisition and retention through business intelligence to provide value to the organization and customers. It is implied in the literature that customer acquisition is more expensive than customer retention because the lack of information on new customers makes it difficult to target the appropriate customers. Therefore, precise evaluations of customer value and customer segmentation are critical for successful CRM. Customer relationship management also identifies the suitable products for a special segment.

Increased customer retention and loyalty, higher customer profitability, creation value for the customer, customization of the products and services, and lower process, higher quality products and services are mentioned as the potential benefits of CRM [7,8]. Marketers believe that 80% of the profits are produced by to 20% of profitable customers and 80% of the costs are produced by top 20% of unprofitable customers. This rule is called the 80/20 rule that the marketers use it for customer profitability evaluation [9,10].

By these definitions it may seem that CRM is only useful for managing the relationships between businesses and customers. A closer examination revealed that CRM is also applicable to business-to-business environments. CRM helps smooth the process when various representatives of seller and buyer companies communicate and collaborate [11].

CRM is used to identify the most profitable customers and allocate the resources to this segment of customers to achieve more profit. The four dimensions of the CRM are essential efforts to gain customer insight [12]. CRM dimensions are listed below:

- Customer identification: CRM begins its work with this cycle. It is also known as customer acquisition. In this cycle company seeks its target customers based on the organization and marketing strategy. This target is assumed profitable for the organization. In this cycle organization analyses the target customer and segments the target customer. Analyzing the target customer involves seeking the profitable segments of customers through analysis of
customers’ underlying characteristics, whereas customer segmentation involves the subdivision of an entire customer base into smaller customer groups or segments, consisting of customers who are relatively similar within each specific segment [13].

- Customer attraction: after identifying the appropriate customer segment, organization must attract this segment by offering services and products and allocating resources to this segment. Direct marketing is one of the techniques for customer attraction.

- Customer retention: this part of CRM refers to customer satisfaction that is the significant factor for customer loyalty. In this cycle companies can use one-to-one marketing strategies that refer to personalization of services or products for each customer which need understanding the customer behavior. Loyalty programs and churn analysis are the other elements of customer retention strategies. Customer retention is a vital part of CRM because of the high costs for identifying and attracting the new customers.

- Customer development: this cycle consist of customer lifetime value, up selling and cross selling and market basket analysis. Customer lifetime value is discussed in the next part of this section. Cross selling analysis refers to finding the optimal product to offer to a given customer [14] and up selling analysis is focused on selling more—or a more expensive version—of the products that are currently purchased by the customer [15].

This paper focuses on the life time value models to develop the customer relationship by calculating the customer current value and predict the characteristics of the churners by prediction models such as logistic regression. This paper also suggests the profitable segment based on the churn probability and current value for the cross selling strategies. The current value of the customers and prediction of churners were based on the data collected from the existing customers of the insurance company.

2.2. Customer Life Time Value (CLTV)

Customer life time value is known as customer value, customer equity, and customer profitability. Berger and Nasr defined LTV as the net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm [16]. Gupta and Lehmann defined LTV as the present value of all future profits generated from a customer [17]. Hwang et al. defined the LTV as the sum of the revenues gained from the company’s customers over the lifetime of transactions after the deduction of the total cost of attracting, selling, and servicing customers, taking into account the time value of money [18]. Gladys et al. defined the CLV as the present value of future cash flows yielded by the customer’s product usage, without taking into account previously spent costs [19]. Basic model for LTV is based on the definition of Hwang et al. and is represented in Equation (1).

$$LTV = \sum_{i=1}^{n} \frac{(R_i - C_i)}{(1 + d)^{t_i - 0.5}}$$

where $i$ is the period of cash flow from customer transaction, $R_i$ the revenue from the customer in period $i$, $C_i$ the total cost of generating the revenue $R_i$ in period $i$, and $n$ is total number of periods of projected life of the customer under consideration. The calculation above is the most basic model that ignores the fluctuations of sales and costs. Berger and Nasr have proposed LTV calculation model that reflects the fluctuations of sales and costs that is represented in Equation (2) [17].

$$LTV = \sum_{i=1}^{n} \pi(t) \times \frac{1}{(1 + d)}$$

where $\pi(t)$ is the function of customer profits according to time $t$.

Hwang et al. suggested a new LTV model of individual customer considering churn rate of a customer. This model is represented in Equation (3).

$$LTV = \sum_{i=1}^{n} \pi_r(t_i)(1 + d)^{t_i - 0.5} + \sum_{i=1}^{N_i} \pi_f(t_i) + B(t_i)$$

where $t_i$ is service period index of customer $i$, $N_i$ is the total service period of customer $i$, $d$ is the interest rate, $E(i)$ is expected service period of customer $i$, $\pi_r(t_i)$ is the past profit contribution of customer $i$ at period $t_i$, $\pi_f(t_i)$ is the future profit contribution of customer $i$ at period $t_i$, and $B(t_i)$ is the potential benefit from customer $i$ at period $t_i$.

Customer segmentation methods using LTV can be classified into three categories: 1) segmentation by using only LTV values, 2) segmentation by using LTV components and 3) segmentation by considering both LTV values and other information [20]. The first category uses the equations above and the data collected from the organization to calculate the customer lifetime value. The second category uses LTV components—current value of customers, potential value and customer loyalty. And the third one uses both three LTV components and also the socio-demographic data from customers and product or transaction information. In this paper we use the third category to understand the most profitable segment of our customers.

2.3. Logistic Regression

In our suggested model we used binary logistic regression...
model for predicting customer churn. Binary logistic regression is most useful when you want to model the event probability for a categorical response variable with two outcomes. The model is represented in Equation (4).

\[ F(z) = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}} \quad (4) \]

The variable \( z \) represents the exposure to some set of independent variables, while \( f(z) \) represents the probability of a particular outcome, given that set of explanatory variables. The variable \( z \) is a measure of the total contribution of all the independent variables used in the model and is known as the logit. The variable \( z \) is usually defined as the Equation (5).

\[ z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (5) \]

where \( x(i) \) is the socio-demographic data about the customer \( i \). Our goal is to achieve the intercept and other coefficients to predict whether a new customer with specific socio-demographic information will churn or not.

3. Research Model

Currently cross-selling models are being built on the total customer database that is not profitable for the company and doesn’t lead to the appropriate result. To overcome this, it could be interesting to build a cross selling model on the loyal customers with appropriate current value. LTV is based on the understanding of the behavior of the profitable customers of the organization. For this reason we propose a model to understand the profitable segment of the customers based on the LTV components that is mentioned in the previous section. First phase of this model evaluates the current value of each customer based on the transactional data of each customer. The second phase evaluates the churn rate and predicts loyal customers using logistic regression model. The third phase segments the customers based on the current value and customer loyalty that calculated in previous phases. After analyzing and selecting the profitable segment company can use the cross selling strategies to develop the relationship between customers and the organization. The research model is illustrated in the Figure 1.

4. Model Evaluation

4.1. Data Collection and Description

The raw data of this study consists of life insurance data of a private insurance company in Iran and collected in 2009. The dataset is composed of 11695 records and 12 data fields. This data set consists of 8 types of life insurance and the customers have different payment styles. Data fields and their value variances are described in the Table 1. We used data from one insurance type to evaluate the proposed model. The data from this insurance type consists of 1651 records. In this dataset some customers refuse to pay the specified amount in the contract in the specified time periods. Therefore it is essential for the company to know about its loyal customers and allocate its resources to these customers.

4.2. Phase 1: Calculating the Current Value

Current value is a profit gained from a customer during a period of time. In this paper we assume the time period from the contract date of each customer to 2009. Therefore we calculate the cumulative value of the customer from the past to present. The current value is calculated from a simple calculation as follows [19]:

Customer Value = (average amount asked to pay – Cumulative amount in arrears)/total service period.

Table 2 presents the minimum, maximum and mean of the current values for 1651 customers.

4.2. Phase 2: Calculating Customer Loyalty and Churn Rate

Customer loyalty is derived from the customer satisfaction.

Table 1. Sample of data fields and their values in the dataset.

<table>
<thead>
<tr>
<th>Field</th>
<th>Minimum value</th>
<th>Maximum value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life asset</td>
<td>0</td>
<td>600,000,000</td>
</tr>
<tr>
<td>Die asset</td>
<td>0</td>
<td>1,000,000,000</td>
</tr>
<tr>
<td>Asset incremental percentage</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Insurance incremental percentage</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>14</td>
<td>88</td>
</tr>
<tr>
<td>Insurance time</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Payment type</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Number of delays in payment</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Amount of money that is not paid</td>
<td>0</td>
<td>200,052,000</td>
</tr>
</tbody>
</table>
Table 2. Calculated current value for one insurance type.

<table>
<thead>
<tr>
<th>Current value</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2,227,541</td>
<td>53,751,000</td>
</tr>
</tbody>
</table>

Loyal customers are defined as the customers that are more likely to continue their relationship with the organization. Customer loyalty can be achieved from the following equation:

Customer Loyalty = 1 – Churn rate

Churners are customers that have a relationship with a company but will go to the competitor in the near future. In the insurance company, experts defined the churners as customers who have delays in payments. Also it is very popular that most people have two or more different insurances from one insurance company. Therefore, it is important for an organization in a competitive environment to understand the characteristics of its loyal customers due to the high costs of identifying and acquiring new customers and introduce its new services to the loyal customers. With the use of prediction models and the current customers’ data—both socio demographic and transactional data—companies can predict the behavior of the new customers and target new customer segments. In this paper we evaluate logistic regression as a prediction model. The interest is to segment the customers into churners and non churners. Therefore the result is a binary probability of churn.

Logistic regression can be used to predict a dependent variable on the basis of continuous or categorical independents and to determine the effect size of the independent variables on the dependent; to assess interaction effects; and to understand the impact of covariate control variables. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). Goodness of fit tests such as the likelihood ratio test is available as indicators of model appropriateness, as is the Wald statistic to test the significance of individual independent variables. The insurance data set has some variables that can be considered as categorical variables, such as payment type or insurance time. These variables are represented in Table 3.

The independent variables are the dataset fields and the calculated regression coefficients and the intercept is represented in Figure 2. Figure 2 shows the coefficients (B), the standard errors associated with the coefficients, the Wald chi-square statistics, associated p-values (Sig), and odds ratio (Exp(B)). Variables with Sig > 0.05 are not statistically significant. The probability of churn can be calculated by substituting the coefficients for each variable in the Equation (6).

\[
\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k \quad (6)
\]

where \( p \) is the probability of churn.

Figure 3 compares the odd ratio for all the independent variables.

Table 3 represents four segments of customers based on their current value and the churn probability. We briefly discuss the strategies for each segment.

**Segment I:** This segment consists of loyal customers, but company has not succeeded in gaining profit from...

Copyright © 2011 SciRes.
Figure 3. Exp(B) for the regression variables.

Table 4. Segmentation with current value and loyalty.

<table>
<thead>
<tr>
<th>Current value</th>
<th>Churn probability</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Segment II</td>
<td>II</td>
<td>IV</td>
</tr>
<tr>
<td>Low</td>
<td>Segment I</td>
<td>I</td>
<td>III</td>
</tr>
</tbody>
</table>

them. Customers in this segment may have larger opportunities for the up-selling strategies.

Segment II: this segment can be regarded as unattractive. These customers have high churn probability, and their current value is low.

Segment III: this segment consists of loyal customers with high current value. These customers are the best targets for cross-selling strategies.

Segment IV: this segment is the least important one for the up/cross selling activities. Customers have high current value and the churn probability is high. They are churners that have low possibility for cross-selling activities.

Data collected from the insurance company is the witness for this claim. Table 5 shows the result of the collected data for each segment. In segment four, only one customer is reference for cross-selling example. Although this segment has high current value but it seems that it’s better for companies not to investigate on these customers and allocate the resources to other segments that has higher potential for up/cross-selling activities.

In the two first phases we calculated and discussed the current value and customer churn probability for a type of life insurance. In this phase we present a segmentation model based on the current value and the customer churn and we evaluated this model based on the real data collected from the insurance company. The dispersal of the customer data in the four segments are represented in the Table 5. As illustrated in Table 5, most of customers are in the second segment that is not interesting at all for the company. Although this segment includes most of the customers but it has the least number of cross-selling and up-selling activities. This shows benefits of segmenting the customers before offering cross-selling and up-selling to all customers of the company.

5. Conclusions

The main objective of CRM is to have an effective relationship with different segments of the customers. For this reason it is essential to have enough information about the current customers and their behavior. Based on the information about these current customers, companies can identify the most profitable segment of these customers. Identifying the most profitable segments can help the company to manage a different relationship with this segment, and also can help the company to use the socio-demographic features of the profitable segment as a selective condition for the new customer targeting.

In this paper we focused on the relationship management of the current customers and suggested a segmentation model to optimize the selection of the most desirable segment by the company and implement the cross-selling models based on this desirable segment. This approach is more effective than implementing the cross-selling models on all the current customers of the company. For this reason we segmented the customers based on the current value of the customers and the churn rate of the customers. We used data from the insurance company. The current value for each customer was calculated based on the transaction data. To predict the customer loyalty, we used prediction models such as logistic regression. This model used both customer’s socio-demographic data and transaction data about the number of delays for paying the life insurance fees. The results of these models showed coefficients for the prediction variables. With the use of these models and current values of customers, companies can segment their customers and select the desirable segment to announce them about the other services of the company.

Table 5. Customer dispersal through four segments.

<table>
<thead>
<tr>
<th>Current value</th>
<th>Churn probability</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Segment II</td>
<td>926 customers</td>
<td>299 customers</td>
</tr>
<tr>
<td></td>
<td>Cross-selling:</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>Segment I</td>
<td>311 customers</td>
<td>115 customers</td>
</tr>
<tr>
<td></td>
<td>Up-selling:</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Cross-selling:</td>
<td>9</td>
<td>3</td>
</tr>
</tbody>
</table>

REFERENCES


