Expert Systems with Applications 39 (2012) 3964–3973

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Segmentation of telecom customers based on customer value by decision tree model

Shui Hua Han [a], Shui Xiu Lu [a], Stephen C.H. Leung [b,*]

[a] Department of Management Science, Xiamen University, 361005 Xiamen, China
[b] Department of Management Sciences, City University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Customer value
Customer lifecycle
Loyalty
Credit
Decision tree model

ABSTRACT

The more the telecom services marketing paradigm evolves, the more important it becomes to retain high value customers. Traditional customer segmentation methods based on experience or ARPU (Average Revenue per User) consider neither customers' future revenue nor the cost of servicing customers of different types. Therefore, it is very difficult to effectively identify high-value customers. In this paper, we propose a novel customer segmentation method based on customer lifecycle, which includes five decision models, i.e. current value, historic value, prediction of long-term value, credit and loyalty. Due to the difficulty of quantitative computation of long-term value, credit and loyalty, a decision tree method is used to extract important parameters related to long-term value, credit and loyalty. Then a judgments matrix formulated on the basis of characteristics of data and the experience of business experts is presented. Finally a simple and practical customer value evaluation system is built. This model is applied to telecom operators in a province in China and good accuracy is achieved.

1. Introduction

The telecom industry in China was restructured in 2008 when 3G licenses were finally granted to three mobile operators. Since then, competition has been intensified further. As a result, telecom operators are paying much more attention to high-value customers. The 80/20 rule points out that 80% of the profits come from the top 20% of profitable customers and 80% of the costs are incurred on the top 20% of unprofitable customers (Duboff, 1992; Gloy, Akridge, & Preckel, 1997). However, finding the top 20% customers is the crucial issue for the operators. It is believed that companies who can capture the top 20% customers will win the battle for the market.

Traditionally, experience-based or ARPU (Average Revenue per User) method is widely-used to find the top 20% customers in China's telecom industry. In general, customers whose ARPU is ranked in top 20% are customers whose usage value is in the top 20% bracket. However, such a method considers only the current and the historic profit, but not future revenue and customer lifecycle. So this method cannot effectively discover the real high-value customers. For instance, customer A and customer B have different ARPUs (A is 200, B is 150), and their indirect values (e.g., loyalty, credit, etc.) may also be significantly different (A is 0, B is 50), however they may have the same contribution to the company's profit, i.e., 200 (Fig. 1). On the other hand, a pair of customer A and B may have the same ARPUs, but their costs to the company may be significantly different. Nevertheless, it can be noted that they may have different contribution to the company (A is 200, B is 80) (Fig. 2).

Therefore, it is crucial to establish indexes for customer value. From a business point of view, customer value should be the potential profit from a customer along the customer's lifecycle. To the best of our knowledge, a practical and well designed customer evaluation system based on customer value has not been proposed so far; evaluation of customer value is still an unsolved, if not unaddressed, problem. According to Walter, Ritter, and Gemunden (2001), customer value is defined as customers' net cash flow and prospective profit perceived by the decision-maker. We divide customer value into two parts: direct value, which is used to scale the monetary effect, and indirect value, which is used to scale the non-monetary effect. Direct value includes earnings not only from customer's current value but also from the long-term value. We try to predict the lifecycle of in-net customers based on characteristics data, and then obtain the long-term value. For indirect value, positive and negative samples are used to extract the characteristics data which may affect loyalty, and then based on characteristics data and business experts' experience, a judgment matrix is presented. The purpose of this paper is to evaluate customers' contribution for facilitating enterprise decision-making by building a simple and practical customer value computation system.

The rest of this paper is organized as follows. Section 2 describes an overview of the related work, while Section 3 presents

* Corresponding author. Tel.: +852 3442 8650; fax: +852 3442 0189.
E-mail address: mssleung@cityu.edu.hk (S.C.H. Leung).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.09.034

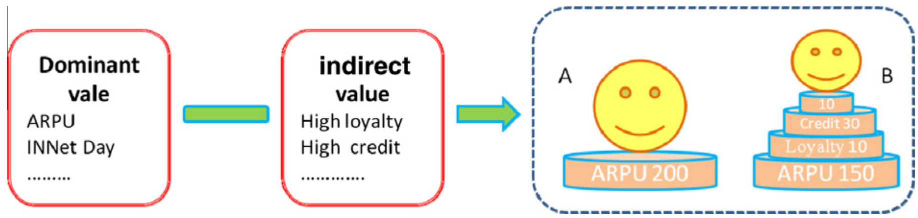

Fig. 1. Customer direct value and implicit values.

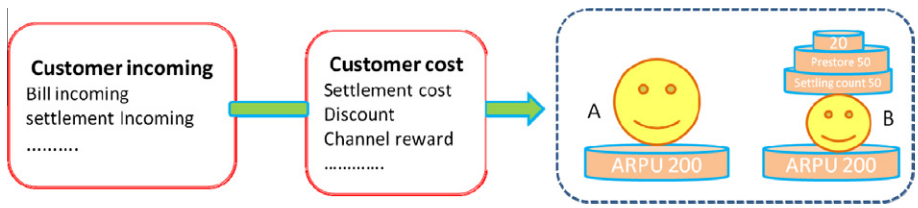

Fig. 2. Customer incoming and customer cost.

the computational model. Section 4 describes the empirical results. Finally, conclusion is provided in Section 5.

2. Related works

2.1. Customer value

According to Berger and Nasr (1998), the basic formula for calculating customer value for customer i at time t for a finite time horizon (T) is as follows:

$$CV_{i,t} = \sum_{\tau=0}^{\tau=T} \frac{\text{Profit}_{i,t+\tau}}{(1+d)^{\tau}}$$

where d is pre-determined discount rate, and $\text{Profit}_{i,t}$ is the profit contributed by customer i in period t. Donkers, Verhoef, and Jong (2007) noted that there are two kinds of models to compute $\text{Profit}_{i,t}$. The first kind is relationship-level models, and the second kind is service-level models. Relationship-level models include status quo model, regression model, customer retention model, Probit model, bagging approach, customer relationship duration and Tobit II model. Service-level models include independent choice models, independent duration models, multivariate choice models and multivariate duration models.

However, there are other kinds of models. Hwang, Jung, and Suh (2004) considered churn rate of a customer in their model to compute customer value. Cheng and Chen (2009) and Liang (2010) used the RFM model to calculate customer value where R refers to recency of the last purchase, F refers to the frequency of purchases and M refers to monetary value of purchases. Chan, Ip, and Cho (2010) and Donkers et al. (2007) applied Markov chain to calculate customer value. In view of skewed distribution, Benoit and Van de Poel (2009) adopted quantile regression to calculate customer value. With an eye to non-contractual setting, Glady, Basen, and Croux (2009) employed Pareto/NBD model to calculate customer value.

A sizable part of extant research on customer value considers the monetary effect, but not non-monetary effect. However, in fact, the non-monetary effect can also influence customer value (Fig. 1). In this paper, we take both monetary effect and non-monetary effect into account to calculate customer value. We also add loyalty and credit as the indexes of customer value.

2.2. Customer segmentation

There are two kinds of methods to segment customers. One is segmentation based on customer value and the other is to apply data mining for customer segmentation.

In general, segmentation strategies based on customer value can be classified into three categories (Kim, Jung, Suh, & Hwang, 2006): (i) segmentation by using only customer value (Zeithaml, Rust, & Lemon, 2001), (ii) segmentation by considering both customer value and other information (e.g., customer value, uncertainty, etc.) (Benoit & Van de Poel, 2009), and (iii) segmentation by using only customer value components (e.g., current value, potential value, loyalty, etc.) (Hwang et al., 2004). In the first method, the list is stored (in descending order) by customer value. Segmentation is by the percentile of the list. In the second method, customers are divided into n-dimensional segment space, where customer value is one of the axes and other information consists of the rest ($n-1$)-dimensional segment space. In the last method, customers are divided into n-dimensional segment space, where components of customer value consist of the n-dimensional segment space.

For data mining approach, there are two categories of segmentation methods; multivariable statistical analysis and the neural network model (Liang, 2010). For example, K-means approach is one kind of multivariable statistical analysis (Chiu, Chen, Kuo, & Ku, 2009). Hung and Tsai (2008) applied neural networks for customer segmentation.

In this paper, we use a segmentation strategy based on customer value as well as data mining technology. We apply data mining technology to extract characteristics of customer data, and then segment customers based on customer value, as derived from historic value, current value, long-term value, loyalty and credit. The segmentation process considers monetary value as well as non-monetary value.

3. Computing customer value based on decision tree

We divide customer value into two parts: direct value and indirect value. Direct value refers to monetary value, which represents decision-makers perception of customers' net cash flow from the beginning of the lifecycle to the end of the lifecycle. It includes historic value, current value and long-term value. Long-term value is