Expert Systems with Applications 38 (2011) 15273–15285

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Credit card churn forecasting by logistic regression and decision tree

Guangli Nie [a,e], Wei Rowe [c,1], Lingling Zhang [a,b], Yingjie Tian [a], Yong Shi [a,d,*]

[a] Research Center on Fictitious Economy and Data Science, CAS, Beijing 100190, China
[b] Management School, Graduate University of Chinese Academy of Sciences, Beijing 100190, China
[c] Department of Finance, Banking & Law, University of Nebraska at Omaha, NE 68182-0048, USA
[d] College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA
[e] Postdoctoral Programme of Agricultural Bank of China, Beijing 100005, China

ARTICLE INFO

Keywords:
Credit card
Customer churn
Logistic regression
Decision tree
Data mining

ABSTRACT

In this paper, two data mining algorithms are applied to build a churn prediction model using credit card data collected from a real Chinese bank. The contribution of four variable categories: customer information, card information, risk information, and transaction activity information are examined. The paper analyzes a process of dealing with variables when data is obtained from a database instead of a survey. Instead of considering the all 135 variables into the model directly, it selects the certain variables from the perspective of not only correlation but also economic sense. In addition to the accuracy of analytic results, the paper designs a misclassification cost measurement by taking the two types error and the economic sense into account, which is more suitable to evaluate the credit card churn prediction model. The algorithms used in this study include logistic regression and decision tree which are proven mature and powerful classification algorithms. The test result shows that regression performs a little better than decision tree.

1. Introduction

Data mining refers to discover knowledge from a large amount of data. In this paper, we discuss the application of data mining including logistic regression and decision tree to predict the churn of credit card users. The banks can take corresponding actions to retain the customers according to the suggestion of the models.

With today's cost-cutting and intensive competitive pressure, more companies start to focus on Customer Relationship Management (CRM). The unknown future behaviors of the customers are quite important to CRM. Hence, it is of crucial importance to detect the customers' future decision then the company can take corresponding actions early (Glady, Baesens, & Croux, 2008). The customers who stop using the company's products are usually called churners. Finding the churners can help companies retain their customers. Gustafsson, Johnson, and Roos (2005) studied telecommunication services to examine the effects of customer satisfaction and behavior on customer retention. Results indicated a need for CRM managers to more accurately determine customer satisfaction in order to reduce customer churn.

One of the major reasons for this is that it costs less to retain existing customers than to acquire new customers (Roberts, 2000). It costs up to five times as much to make a sale to a new customer as it does to make an additional sale to an existing customer (Dixon, 1999; Floyd, 2000; Slater & Narver, 2000). And, it is becoming more evident that the only way to remain a leader in this industry is to not only be customer-driven but also focus on building long-term relationships.

Due to the development of information technology, many companies have accumulated a large amount of data. Analyzing this data can help the manager make the right marketing decision and pinpoint the right customer to market. Because of the large amount of accumulated data and serious churn related to credit card holders, it is a very good field in which to predict churn.

Several studies have proved the effectiveness of the power of customer retention. A bank is able to increase its profits by 85% due to a 5% improvement in the retention rate (Reichheld & Sasser, 1990). Van den Poel and Larivière (2004) calculated the financial impact of a one percent increase in customer retention rate. The power of the model can stay for a relatively long time. According to the research of Neslin, the churn models in the data typically still perform very well if used to predict churn for a database compiled 3 months after the calibration data (Neslin, Gupta, Kamakura, Lu, & Mason, 2006).

As the economy develops in China, a large amount of credit cards are issued. As of the third quarter of 2008, 132 million cards have been issued in China.[2] But many of the card holders are not active (or called churn holders). With increasing bank competition,

* Corresponding author at: Research Center on Fictitious Economy and Data Science, CAS, Beijing 100190, China. Tel./fax: +86 10 82680698.

E-mail addresses: sdungl@163.com (G. Nie), wrowe@mail.unomaha.edu (W. Rowe), zhangll@gucas.ac.cn (L. Zhang), tianyingjie1213@163.com (Y. Tian), yshi@gucas.ac.cn (Y. Shi).

[1] Tel.: +1 402 554 2812.

[2] http://www.chinavalue.net/Article/Archive/2009/1/20/155619.html.

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.06.028

customers are able to choose among multiple service providers and easily exercise their right of switching from one service provider to another. If banks can predict future behaviors before the customers close their account or stop using the card to pay, they can market to retain these customers.

The main purpose of this paper is not to provide a new data mining algorithm, but to focus on the application of the churn prediction, to provide a framework of understanding the knowledge of the card holders' hidden pattern using the data of Chinese banks. From the data preparation to useful knowledge, the goal is application of churn prediction. In this paper, we introduce a way to complete churn prediction considering profit.

The rest of the paper is organized as follows. The definition of churn and the summary of the algorithms and criteria are introduced in Section 2. The data used in the research is described in Section 3, and the modeling process based on logistic regression and decision tree are presented in Section 4 and 5, respectively. In Section 6, we conclude.

2. Definition of churn and algorithm evaluation criteria

In different application fields, the definitions of churn differ. Churn means the customer shift from one service provider to another (Lu, 2002). Customer churn is defined as the propensity of customers to cease doing business with a company in a given time period (Neslin et al., 2006).

Many of the previous definitions of churn use the behaviors related to product and a threshold fixed by a business rule. Once the transactions of the customer is lower than the threshold, the customer would be regarded as a churner (Glady et al., 2008). Van den Poel and Larivière (2004) regard the customer who closed his accounts as a churner. Buckinx and Van den Poel (2005) define a partial defector as someone with the frequency of purchases below the average and the ratio of the standard deviation of the interpurchase time to the mean interpurchase time above the average. Glady et al. (2008) defines a churner as a customer with less than 2500 Euros of assets (savings, securities or other kinds of products) at the bank. Glady, Baesens, and Croux (2009) claimed that the threshold is not always relevant and one should observe the evolution of the customer activity instead.

In the telecommunications industry, the broad definition of churn is the action that a customer's telecommunications service is canceled which includes both a service provider-initiated churn such as a customer's account being closed because of payment default and a customer-initiated churn. In one study, only customer-initiated churn is considered and it is defined by a series of cancel reason codes. Examples of reason codes are: unacceptable call quality, more favorable competitor's pricing plan, misinformation given by sales, customer expectation not met, billing problem, moving, change in business, etc. (Lu, 2002).

The relationship between churn rate and average lifetime is also studied. If no new customers are acquired then the average lifetime of an existing customer is equal to $1/c$, where c is the annual churn rate (Neslin et al., 2004). Gustafsson et al. (2005) use customer satisfaction (CS_t), affective commitment (AC_t), calculative commitment (CC_t), a situational trigger condition (ST_t), and a reactional trigger condition (RT_t), all in time t, to predict churn in time $t + 1$ (Churn_{t+1}).

In our application, we define that the customer who did not do any transaction with the bank on his own initiative during the observation period (explained later) is a churner.

In order to predict the churn of the customer effectively, it is crucial to build effective models which fulfill some evaluation criteria. To accomplish this, many predictive modeling techniques are available. These data mining algorithms can help to select variables and build models (Hung, Yen, & Wang, 2006). Researchers use a variety of approaches to develop churn models including a combination of estimation techniques, variable selection procedures, time allocations to various steps in the model-building process, and a number of variables included in the model (Neslin et al., 2006). The techniques include GA, Regression, Neural Networks, Decision Tree, Markov Model, Cluster Analysis (Hadden, Tiwari, Roy, & Ruta, 2005), and optimization (Better, Glover, Kochenberger, & Wang, 2008; Mclain & Aldag, 2009).

According to the research of Hadden et al. (2005), regression and decision tree are the two most popular algorithms used in the research and perform well. Neslin et al. (2006) categorized the approaches as "Logit," "Trees," "Novice," "Discriminant," and "Explain." After comparison they found that the Logit and Tree approaches perform the best and result in that firm achieving a relatively good level of predictive ability. The Novice approach is associated with middle-of-the-road predictive performance, while the Discriminant and Explain approaches are associated with lower predictive performance (Neslin et al., 2004). Multiple-criteria quadratic programming approach has been used to credit card analysis and the models perform well (Li, Shi, & He, 2008; Peng, Kou, Shi, & Chen, 2008; Shi, Peng, Kou, & Chen, 2005).

In our research, we also use logistic regression and decision tree which are mature data mining algorithms to build models and predict the churn of credit card users. We will compare the performance of these two algorithms in credit card churn prediction.

After building a predictive model, marketers will use these classification models to predict future behaviors of customers. It is essential to evaluate the performance of the classifiers. Percentage of correctly classified (PCC) and receiver operating curve (ROC) are usually used as criteria. PCC, which computes the ratio of correctly classified cases to the total number of cases to be classified (also known as accuracy), is undoubtedly the most commonly used evaluation metric of a classifier. The ROC is a graphical plot of the sensitivity – i.e. the number of true positives versus the total number of events – and 1-specificity – i.e. the number of true negatives versus the total number of non-events. The ROC can also be represented by plotting the fraction of true positives versus the fraction of false positives (Coussement & Poel, 2008).

Another two criteria related to accuracy are also used; they are top-decile lift and Gini coefficient. Lift is a way to quantify the accuracy of a predictive model. e.g., among the 10% of customers predicted as most likely to churn, what percentage of them actually do relative to the percentage of all customers who churn. Gini coefficient is also related to lift which measures the area between an entry's cumulative lift curve and the random lift curve.

Beside PCC, the two types of errors, i.e. the Type I error which means a customer who did not churn is misclassified as a churner and Type II error which means a customer who churned is misclassified as an un-churner are also studied. The loss caused by Type II error is generally regarded as 5–20 times higher than the loss caused by Type I error (Lee, Chiu, Chou, & Lu, 2006)

Loss function is also used to compare the performance of the classifiers. The loss function is calculated on the basis of life value and life value has been discussed in several papers. Gupta et al. define the value of customer as "the value of a customer as the expected sum of discounted future earnings" (Gupta, Lehmann, & Stuart, 2004). Glady et al. designed a loss function based on previous definitions (Glady et al., 2009). In this research, we design a misclassification cost measurement to indicate the loss caused by the error of the model.

First, we examine a demo example to show the necessary to consider these two types error respectively. Consider the following misclassification table as shown in Table 1.

The overall error rate of the model is 10%. According to the prediction result of the model, there are 190 customers who may churn in the next period. However only 90 of the 190 are real churners; the remaining 100 are not real churners who are