Electronic Commerce Research and Applications 17 (2016) 99–111

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

The gamma CUSUM chart method for online customer churn prediction

Ssu-Han Chen

Department of Industrial Engineering and Management, Ming Chi University of Technology, 84 Gungjuan Rd., Taishan Dist., New Taipei City 24301, Taiwan, ROC

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form 3 April 2016
Accepted 3 April 2016
Available online 4 April 2016

Keywords:
Customer churn
Gamma CUSUM chart
Hierarchical Bayesian mixture model
Online customers
Inter-arrival time
Recency

ABSTRACT

Customer churn in online firms is difficult to manage because customers are so fickle. The ability to detect churn in the early stage is something every online firm would wish to achieve. It represents both a potential revenue source and a cost-saving benefit. Churn prediction models attempt to organize customer behaviors, transactions and demographics to reduce the possibility of churn within a given time. However, most current methods depend on high-dimensionally static data analysis and the model parameters are estimated based on the massively customers. A dynamic and customized prediction model at the individual level cannot be achieved. This study proposes a novel mechanism based on the gamma CUSUM chart in which only inter-arrival time (IAT) and recency need to be collected, so that the customized parameters can be estimated for the purpose of individual monitoring. The data in this study are from an online dating website in Taiwan. The gamma CUSUM chart is compared with the exponential CUSUM chart of Gan (1994), CQC-v of Xie et al. (2002) and CQC of Chan et al. (2000). The results show that the accuracy rate (ACC) for a gamma CUSUM chart is 5.2% higher and the average time to signal (ATS) is about two days longer than required for the best CQC-v.

1. Introduction

Customer relationship management (CRM) has gained increasing attention in modern business management. It helps companies find target customers, retain customers and explore customer values (Berson and Smith, 1999), thereby improving their competitive advantage. Maintaining existing customers is one of the important strategies to enhance corporate profitability. Long-term customers have high and stable spending power (O'Brien and Jones, 1995) and can produce a word-of-mouth effect (Reichheld and Teal, 1996). They can even spontaneously recruit new customers (Oliver, 2010). So they are considered to be the most precious asset of companies. In a mature market, business mainly comes from existing customers, and retaining existing customers can create more value than developing new ones (Coussement and Van den Poel, 2008a; Zorn et al., 2010).

Previous studies show the importance of keeping existing customers. New customers do not immediately bring benefits to revenue (Zeithaml et al., 1996). The success rate for retaining existing customers is 60%, which is double the success rate for developing new ones (Kotler, 2001). The cost of acquiring new customers can be as much as twelve times higher than keeping existing customers (Torkzadeh et al., 2006). Reducing the customer churn rate by 5% can produce an improvement of more than 25% in profits (Rechinhheld and Sasser, 1990; Peppers and Rogers, 2000). Coyles and Gokey (2005) have also shown that businesses can generate ten times the value by halting customer if they respond to small changes in consumer behavior.

Customer churn is a marketing-related term which means customers defect to another supplier or purchase less (Coussement and Van den Poel, 2008b). As existing customers are an important source of business profits, being able to identify customers who show signs that they are about to leave can create more profits for businesses. This is especially important for online customers, as the phenomenon of customer churn appears to be very rapid and difficult to grasp (Peng et al., 2013). If companies cannot take measures to retain customers before their status deteriorates, the customers may never come back, resulting in wasted investment and loss of future earnings. A timely retention strategy can keep customers, and it is the best way to retain customers.

Constructing a prediction mechanism to monitor customer churn is an important step for business development, and it has become a popular topic in the last ten years (Coussement and Van den Poel, 2008a; Tsai and Lu, 2009; Huang et al., 2010; Verbeke et al., 2012; Coussement and De Bock, 2013; Faris, 2014). In general, researchers use the information on customers' previous behaviors from a database, such as background demographics, transaction records and interactions. All this information is quantified into variables and used to create prediction models to predict the likelihood that customers might be lost in the future.

E-mail address: ssuhanchen@mail.mcut.edu.tw

http://dx.doi.org/10.1016/j.elerap.2016.04.003
1567-4223/

However, the previous research studies have conducted static data analysis of customer churn at some particular cut-off time. This makes it necessary to convert longitudinal data into static data through aggregation or rectangularization (Chen et al., 2012). The most significant drawback of static data analysis is that it cannot provide dynamic monitoring of customer churn. Given the extremely volatile nature of online markets, it would seem to be desirable to adopt a more dynamic approach. In addition, the results of repeated static data analysis are displayed in massive tables, and engineers need to make considerable efforts to analyze the predictive differences at every cut-off time before compiling reports for administrators. Most importantly, the parameters or weights in the prediction models are based on the average behavior of customers, so that strictly speaking they cannot be customized to generate prediction models that function effectively at the individual level.

A *gamma cumulative sum* (CUSUM) *control chart* is different from the previous churn prediction mechanisms in that it can simultaneously achieve longitudinal analysis and visualization for monitoring purposes. In the past, a few studies have used control charts to predict customer churn (Pettersson, 2004; Jiang et al., 2007; Samimi and Aghaie, 2008), and the present author believes that the under-use of this approach is mainly due to the difficulty of selecting analytical fields.

Many previous studies analyze the newspaper, cable TV and financial industries in which the customers usually sign contracts for a number of years. After signing their contracts, customers seldom contact the companies unless they are complaining or asking for service changes. In addition, some studies analyze the customer churn of industries selling tangible products. However, customers may not return to make another purchase for a long time after purchasing a tangible product, resulting in the lack of personal shopping history and the difficulties in building customization parameters and monitoring mechanisms.

This study analyzes Taiwan's "Internet industry" in which companies and customers have no signed contracts, trading of tangible products or face-to-face transactions. It is related to the virtual features of the Internet. Customer visits are produced by the website content, and if the service has a persistent appeal, customers will keep coming back and browsing. Taiwanese people spend four hours online per day on average, and the amount of data generated by online activities is sufficient to build a customized prediction model for customer churn.

The model proposed in this study has the following characteristics:

(1) *Visualized management by exception*. This study employs control charts, as they can additional provide dynamic monitoring of customer behaviors and present the monitoring process in a friendly graphical interface rather than just generate static reports. Control charts can immediately issue warnings when customer behaviors deviate from the previously active status. This is management by exception, in which attention is given only when necessary, further reducing the burden on administrators.

(2) *Longitudinal data analysis*. Unlike the previous models in which churn prediction is based on tens of characteristics of customer behavior, this study performs modeling based only on the *inter-arrival time* (IAT) and the longitudinal data of recency. IAT defines the time difference between two consecutive events while recency analyzes the time difference between the cut-off time and the last event. The two variables are complementary in determining customer status. IAT can usually be used to show the trail of historical behavior, especially when the last event and the cut-off time are far apart. In contrast, recency only shows the behavior at the present, and cannot detect past behavior. As the two variables are both the data of a time interval, it is feasible to combine the two. The time interval data of the combined variables are converted into time-series data which can be used in longitudinal analysis of customer login.

(3) *Customized monitoring mechanism*. Another feature of the method developed in this study is the combination of control charts and a mixture Bayesian hierarchical model. This model uses gamma distributions to estimate the active and inactive login behaviors of an individual customer and then uses inverse gamma distributions to capture heterogeneity, thereby defining the gamma CUSUM parameters for each customer and implementing model customization and automation of parameter estimates.

The rest of this paper is organized as follows. In Section 2, related research of customer churn, CUSUM charts and mixture Bayesian hierarchical models is surveyed. In Section 3, our research methodology is described and explained. In Section 4, a sensitivity analysis is conducted to find the best parameters of the method and experimental results are presented. Concluding remarks and further suggestions are discussed in Section 5.

2. Literature review

In this section, some related research is surveyed, which is used to build the bases of proposed method. The survey of the customer churn techniques makes it clear that the area of longitudinal prediction has not yet been adequately investigated. The review of CUSUM charts helps in the selection of a suitable chart which offers a good fit for time-interval datasets. Finally, the idea of handling customer heterogeneity is inspired by several mixture Bayesian hierarchical models.

2.1. Prediction models for customer churn

Customer development is difficult and costly. Attracting new customers requires advertising and promotion expenses, in addition to identity authentication and credit checks. There are reactive and proactive approaches to managing customer churn (Burez and Van den Poel, 2007). The former passively creates incentives to retain customers as they terminate contracts or service relationships. The latter constantly monitors customer status and prevents potential losses from happening by giving immediate incentives.

In general, a proactive approach has more potential benefits because analysis after the occurrence produces delays in responding, and it is often more expensive to retain the customer who have already decided to leave (Chen et al., 2012). In the longer-term, advances in information technology mean that all business will gradually shift from being reactive on the basis of data management to being proactive, focusing on exploring data. By being proactive, before the deterioration of customer behavior, reminders can be sent out to take retention measures as early as possible.

Table 1 summarizes customer churn prediction models reported in the literature in recent years. The distinctive characteristics of each study in terms of the references, fields, variables, data types, methods, and churn definitions are provided. In the columns of variables, the main categories of variables are listed, and the corresponding number of items is in brackets. In the column of data types, the symbol C is cross-sectional, while L is longitudinal analysis.

Table 1 shows that the issue of predicting customer churn has been widely discussed in various fields. These include