Expert Systems with Applications 41 (2014) 7889–7903

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Dynamic churn prediction framework with more effective use of rare event data: The case of private banking

Özden Gür Ali [a,*], Umut Arıtürk [b]

[a] College of Administrative Sciences and Economics, Business Administration, Koç University, Sarıyer, Istanbul, Turkey
[b] Department of Industrial Engineering, Koç University, Sarıyer, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online 19 June 2014

Keywords:
Dynamic churn prediction
Data mining
Customer retention
Private banking
Customer relationship management
Rare event
Sampling
Training data generation

ABSTRACT

Customer churn prediction literature has been limited to modeling churn in the next (feasible) time period. On the other hand, lead time specific churn predictions can help businesses to allocate retention efforts across time, as well as customers, and identify early triggers and indicators of customer churn. We propose a dynamic churn prediction framework for generating training data from customer records, and leverage it for predicting customer churn within multiple horizons using standard classifiers. Further, we empirically evaluate the proposed approach in a case study about private banking customers in a European bank.

The proposed framework includes customer observations from different time periods, and thus addresses the absolute rarity issue that is relevant for the most valuable customer segment of many companies. It also increases the sampling density in the training data and allows the models to generalize across behaviors in different time periods while incorporating the impact of the environmental drivers.

As a result, this framework significantly increases the prediction accuracy across prediction horizons compared to the standard approach of one observation per customer; even when the standard approach is modified with oversampling to balance the data, or lags of customer behavior features are added as additional predictors.

The proposed approach to dynamic churn prediction involves a set of independently trained horizon-specific binary classifiers that use the proposed dataset generation framework. In the absence of predictive dynamic churn models, we had to benchmark survival analysis which is used predominantly as a descriptive tool. The proposed method outperforms survival analysis in terms of predictive accuracy for all lead times, with a much lower variability. Further, unlike Cox regression, it provides horizon specific ranking of customers in terms of churn probability which allows allocation of retention efforts across customers and time periods.

1. Introduction

Customer retention is an important element of the customer relationship management (CRM) literature e.g. (Bijmolt et al., 2010). Several studies show the economic value associated with customer retention: The costs arising from acquiring a new customer surpass the expenditure to retain an existing one (Dawes & Swailes, 1999); long-term customers buy more and bring in new customers (Reichheld, 1996); for a bank, the longer the customer relationship, the higher the customer's worth (Reichheld & Kenny, 1990).

Accurate churn prediction helps the company to target the retention effort on the right customer at the right time and thereby convince valuable customers to stay. Further, churn probability is an important input to the calculation of the customer lifetime value (CLV) metric.

While customer specific churn rate prediction is well accepted in the literature, the typical approach does not consider its variation over time, i.e., lacks the dynamic aspect. On the other hand, dynamic churn prediction would allow the business to allocate retention and service improvement resources across customers and time (Venkatesan & Kumar, 2004). From the customer relationship management point of view, managers need insights into the reasons and timing of churn to create retention strategies and diagnose how much of the attrition can be controlled (Braun & Schweidel, 2011). Further, earlier detection of would-be-churners is valuable as it provides more time to the company to convince those customers to stay before they make their final decision (Verbeke, Dejaeger, Martens, Hur, & Baesens, 2012). Finally,

* Corresponding author. Tel.: +90 212 3381450; fax: +90 212 3381653.
E-mail address: oali@ku.edu.tr (Ö. Gür Ali).

http://dx.doi.org/10.1016/j.eswa.2014.06.018
0957-4174/

dynamic churn predictions are important for evaluating the aggregate value of the firm's customer base and hence its financial health (Fader & Hardie, 2010).

Recently, researchers have started investigating the impact of customer data preprocessing and representation of longitudinal customer data on the predictive accuracy and identified the need for dynamic churn prediction; i.e., churn probability in time (Crone, Lessmann, & Stahlbock, 2006; Lee, Wei, Cheng, & Yang, 2012). Typically, the churn training data consists of the churn response in one particular time period, while customers who have churned earlier are discarded; and the time series nature of the customer behavior is aggregated away in favor of simplicity and interpretability (Cao, 2010; Lee et al., 2012). The recent studies which focus on how to incorporate the-longitudinal data about the customer's historical behavior as predictors still use one outcome observation per customer (Crone et al., 2006; Lee et al., 2012; Orsenigo & Vercellis, 2010; Prinzie & Van den Poel, 2006). Such training data (a) throws away valuable data regarding customers who churned earlier, and (b) does not incorporate variation in the economic and competitive environment (c) does not facilitate dynamic customer churn prediction, which would allow the company to target retention programs better by selecting the customers as well as the time period.

In this paper, we propose a framework for generating training data that facilitates dynamic churn prediction, by generating multiple observations per customer from different time periods and hence including customers who churned earlier. By including the previously churned customers in the training data, this framework also alleviates the rare event problem associated with churn prediction in many organizations (Burez & Van den Poel, 2009), which poses difficulties for learning classification models (Weiss, 2004). Further, it incorporates the impact of the environmental drivers affecting customer churn; for example, customers may reevaluate the value proposition of their service provider in changing economic conditions and switch to competing institutions whose offerings are more suitable to their changing needs. Training datasets that contain churn response of only one time period do not contain variation in the environmental variables, and cannot provide insights about their impact on churn. Thus, their predictive power is expected to be limited in dynamic business environments. We apply and evaluate the proposed training data generation scheme and dynamic churn prediction approach against current standard approaches on the private banking data from a European bank. Private banking customers are high worth, high status individuals (Lassar, Manolis, & Winsor, 2000). As such, they constitute a very valuable customer segment that is sought after by competing banks and financial institutions where retention of even single individual can have significant impact on the company profitability. We empirically show that regardless of the classification method the proposed framework significantly improves the predictive accuracy of the resulting churn models, enables identification of the impact of the environmental factors, and provides additional insights about customer churn drivers compared with the standard approach even when the data is balanced by oversampling, and even when additional lags of customer behavior features are included. Further, the proposed dynamic churn prediction method - independent binary classifiers – outperforms the current standard approach, survival analysis (Cox regression) for all prediction horizons, and provides lead-time specific ranking of customers in terms of churn probability.

To summarize, our contributions are as follows: First, we introduce a dynamic churn prediction for generating training data from customer records and leverage it for predicting customer churn within a specified horizon with standard classifiers. To our knowledge, this is the first paper to perform dynamic churn prediction. Second, we provide a comparison of the accuracy impact of using single versus multiple observations per customer for churn prediction, both by identifying the theoretical drivers, as well as by empirically evaluating in the private banking customer churn context.

The remainder of this paper is organized as follows: In Section 2 we review the relevant literature for predictive churn modeling, representation of longitudinal customer data and environmental variables, and the rare event problem in churn prediction and associated remedial sampling approaches. The following section describes the proposed dynamic churn prediction framework and the proposed independently trained binary classifiers approach, and provides theoretical justifications. Section 4 describes the private banking application and the experimental design and presents the empirical evaluation results. We conclude with a summary of the results, managerial implications and limitations of the research, and offer future research directions.

2. Relevant literature

There is substantial body of work on churn prediction models. The initial focus of the literature was comparison of classification methods in terms of prediction accuracy. Table 1 provides a sample of papers since 2004. Logistic regression, decision trees, neural networks, support vector machines and survival analysis are the most popular methods e.g. (Buckinx & Van den Poel, 2005; Coussement & Van den Poel, 2008; Karahoca & Karahoca, 2011; Neslin, Gupta, Kamakura, Junxiang, & Mason, 2006). While SVM and decision trees have been used for predictive purposes, survival analysis is used descriptively (Burez & Van den Poel, 2008). Many of the later applications use hybrid models or ensembles that integrate multiple classifiers and/or develop variants of the existing algorithms (Buckinx & Van den Poel, 2005; Chu, Tsai, & Ho, 2007; De Bock & Poel, 2011) which improve prediction accuracy while decreasing interpretability (Verbeke, Martens, Mues, & Baesens, 2011). In their large comparative study Verbeke et al., find that accuracy performance of many classifiers are comparable since different classifiers yield better performances in different contexts and datasets (Verbeke et al., 2012).

Recently, researchers have started investigating the impact of data, preprocessing, sampling and representation of customer data on predictive accuracy independent of the classification algorithm. For example, Crone et al., have shown that attribute scaling, sampling, coding of categorical and continuous attributes have a significant impact on predictive accuracy on the classifier performance of decision trees, neural networks and support vector (Crone et al., 2006). Ballings and Van den Poel (2012) investigated how long an event history should be used for creating customer behavior summaries for churn prediction, while Tsai and Chen explored association rules for feature selection (Tsai & Chen, 2010). Next, we review the work related to the representation of longitudinal customer data and environmental variables; and the rare event problem in churn prediction in detail.

2.1. Representation of longitudinal customer data and environmental variables

Predictive models in customer relationship management rely on both static (non-time varying) characteristics of the customer, such as demographics; as well as dynamic characteristics consisting of customer behavior time series including customer transactions and interactions with the company. Each customer is represented with one observation. Customer demographics are obtained from a data warehouse, while the behavior information comes from transaction (Cao, 2010) databases. Traditionally, the transaction and interaction information in different time windows