European Journal of Operational Research 223 (2012) 461–472

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Decision support

A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data

Zhen-Yu Chen [a,*], Zhi-Ping Fan [a], Minghe Sun [b]

[a] Department of Management Science and Engineering, School of Business Administration, Northeastern University, Shenyang 110819, China

[b] Department of Management Science and Statistics, College of Business, The University of Texas at San Antonio, San Antonio, TX 78249-0632, USA

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 27 June 2012
Available online 4 July 2012

Keywords:
Data mining
Customer relationship management
Customer churn prediction
Support vector machine
Multiple kernel learning

ABSTRACT

The availability of abundant data posts a challenge to integrate static customer data and longitudinal behavioral data to improve performance in customer churn prediction. Usually, longitudinal behavioral data are transformed into static data before being included in a prediction model. In this study, a framework with ensemble techniques is presented for customer churn prediction directly using longitudinal behavioral data. A novel approach called the hierarchical multiple kernel support vector machine (H-MK-SVM) is formulated. A three phase training algorithm for the H-MK-SVM is developed, implemented and tested. The H-MK-SVM constructs a classification function by estimating the coefficients of both static and longitudinal behavioral variables in the training process without transformation of the longitudinal behavioral data. The training process of the H-MK-SVM is also a feature selection and time subsequence selection process because the sparse non-zero coefficients correspond to the variables selected. Computational experiments using three real-world databases were conducted. Computational results using multiple criteria measuring performance show that the H-MK-SVM directly using longitudinal behavioral data performs better than currently available classifiers.

1. Introduction

In markets with intensive competition, customer relationship management (CRM) is an important business strategy. Business firms use CRM to build long term and profitable relationships with specific customers (Coussement and Van den Poel, 2008b; Ngai et al., 2009). An important task of CRM is customer retention. Customer churn is a marketing related term meaning that customers leave or reduce the amount of purchase from the firm. Customer churn prediction aims at identifying the customers who are prone to switch at least some of their purchases from the firm to competitors (Buckinx and Van den Poel, 2005; Coussement and Van den Poel, 2008b). Usually, new customer acquisition results in higher costs and probably lower profits than customer retention (Buckinx and Van den Poel, 2005; Coussement and Van den Poel, 2008a; Zorn et al., 2010). Therefore, many business firms use customer churn prediction to identify customers who are likely to churn. Measures can be taken to assist them in improving intervention strategies to convince these customers to stay and to prevent the loss of businesses (Zorn et al., 2010).

Longitudinal behavioral data are widely available in databases of business firms. How to use the longitudinal behavioral data to improve customer churn prediction is a challenge to researchers. In some methods, longitudinal behavioral data are transformed into static data through aggregation or rectangularization. In this study, frameworks for customer churn prediction are developed. In the framework using ensemble techniques, a novel data mining technique called hierarchical multiple kernel support vector machine (H-MK-SVM) is proposed to model both static and longitudinal behavioral data. A three phase algorithm is developed and implemented to train the H-MK-SVM. The H-MK-SVM constructs a classification function by estimating the coefficients of both static and longitudinal behavioral variables in the training process without transformation of the longitudinal behavioral data. Because of the sparse nature of the coefficients, the H-MK-SVM supports adaptive feature selection and time subsequence selection. Furthermore, the H-MK-SVM can benefit customer churn prediction performance in both contractual and non-contractual settings.

This paper is organized as follows. The next section reviews previous work. Section 3 describes the frameworks for customer churn prediction using longitudinal behavioral data. Section 4 presents the fundamentals of the support vector machine (SVM) and the multiple kernel SVM (MK-SVM). The model formulation and the three phase training algorithm of the H-MK-SVM are presented in Section 5. The computational experiments are described in Section 6 and the computational results are reported in Section 7. Conclusions and further remarks are given in Section 8.

* Corresponding author. Tel.: +86 24 83871630; fax: +86 24 23891569.
E-mail addresses: zychen@mail.neu.edu.cn (Z.-Y. Chen), zpfan@mail.neu.edu.cn (Z.-P. Fan), minghe.sun@utsa.edu (M. Sun).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.06.040

2. Previous work

Recently, the topic of customer churn prediction has been discussed extensively in a number of domains such as telecommunications (Kisioglu and Topcu, 2010; Tsai and Lu, 2009; Verbeke et al., 2011,Verbeke et al., 2011), retail markets (Baesens et al., 2004; Buckinx and Van den Poel, 2005), subscription management (Burez and Van den Poel, 2007; Coussement and Van den Poel, 2008a), financial services (Glady et al., 2009) and electronic commerce (Yu et al., 2010). Many data mining techniques have been successfully applied in customer churn prediction. These techniques include artificial neural networks (ANNs) (Tsai and Lu, 2009), decision trees (Qi et al., 2009), Bayesian networks (Baesens et al., 2004; Kisioglu and Topcu, 2010), logistic regression (Buckinx and Van den Poel, 2005; Burez and Van den Poel, 2007), AdaBoosting (Glady et al., 2009), random forest (Buckinx and Van den Poel, 2005; Burez and Van den Poel, 2007), the proportional hazard model (Van den Poel and Larivière, 2004) and SVMs. Lessmann and Voß (2008) gave a detailed review on this topic.

SVMs have strong theoretical foundations and the SVM approach is a state-of-the-art machine leaning method. SVMs have been widely used in many areas such as pattern recognition and data mining (Schölkopf and Smolla, 2002; Vapnik, 1995, 1998) and have achieved successes in customer churn prediction (Coussement and Van den Poel, 2008a; Lessmann and Voß, 2008, 2009; Verbeke et al., 2011,Verbeke et al., 2011).

Demographic and behavioral attributes have been widely used for customer churn prediction (Buckinx and Van den Poel, 2005). Customer demographic data are static, while longitudinal behavioral data are temporal. Customer demographic data can be directly obtained from the data warehouse of the business firm, while the longitudinal behavioral data of individual customers are usually separately stored in transactional databases (Cao, 2010; Cao and Yu, 2009; Chen et al., 2005). Three typical customer behavioral variables are recency, frequency and monetary variables. Recency is the time period since the customer's last purchase to the time of data collection; frequency is the number of purchases made by individual customers within a specified time period; and the monetary variable represents the amount of money a customer spent during a specified time period (Chen et al., 2005; Buckinx and Van den Poel, 2005).

The temporal nature of customer longitudinal behavioral data is usually neglected in customer churn prediction (Eichinger et al., 2006; Orsenigo and Vercellis, 2010; Prinzie and Van den Poel, 2006a). Usually, the longitudinal behavioral variables are transformed into static variables through aggregation or rectangularization before being included in the prediction model (Cao, 2010; Cao and Yu, 2009; Eichinger et al., 2006; Orsenigo and Vercellis, 2010; Prinzie and Van den Poel, 2006a). The transformation results in the loss of temporal development information with potential discriminative ability. For example, the changing values of recency, frequency, and monetary variables between different time periods may have better customer churn predictive ability than the values of these variables in a fixed time period or their averages in a series of time periods. Furthermore, traditional customer churn prediction models usually use static data with two dimensions. Relatively few models proposed in the literature capture temporal information in longitudinal behavioral data with three dimensions (Prinzie and Van den Poel, 2006a).

Customer longitudinal behavioral modeling has been studied in the fields of financial distress prediction (Sun et al., 2011) and customer acquisition analysis (Prinzie and Van den Poel, 2006b, 2007, 2009). However, relatively little research has focused on longitudinal behavioral data for customer churn prediction. Prinzie and Van den Poel (2006a) incorporated static and a one dimensional temporal variable into a customer churn prediction model. They used a sequence alignment approach to model the temporal variable, clustered customers on the sequential dimension, and incorporated the clustering information into a traditional classification model. This approach is limited to the use of temporal variables with one dimension. Eichinger et al. (2006) proposed a classification approach using sequence mining combined with a decision tree for modeling customer event sequence. Orsenigo and Vercellis (2010) proposed a two stage strategy for multivariate time series classification. In the first stage, a rectangularization strategy using a fixed cardinality warping path is proposed to transform multivariate time series into a rectangular table. In the second stage, a temporal discrete SVM is used for classification. They applied this approach to telecommunication customer churn prediction. Huang et al. (2010) reformulated the rectangular table by adopting each time element of the temporal vector as a predictor. Long historical behavioral time series may lead to the "dimension curse" of the prediction models in this approach. The proportional hazard model (Cox) was extended for customer churn prediction using longitudinal behavioral variables (Van den Poel and Larivière, 2004). The coefficients of static and longitudinal behavioral variables are estimated, and then a threshold of the hazard is set for predicting churners.

The above studies made great contributions to customer churn prediction using longitudinal behavioral data. However, they still have limitations. A transformation may result in the loss of potentially useful structural information embodied in the longitudinal behavioral data or increase the computational cost. Moreover, customer churn prediction involves predictors with multiplex data such as static data, temporal data, event sequential data, textural data, and so on (Coussement and Van den Poel, 2008b; Eichinger et al., 2006; Orsenigo and Vercellis, 2010; Prinzie and Van den Poel, 2006a). Relatively little research has focused on simultaneously modeling multiplex data. In addition, the availability of expanded customer longitudinal behavioral and demographic data raises a new and important question about which variables to use and which time subsequence in the longitudinal behavioral variables to consider (Dekimpe and Hanssens, 2000).

The H-MK-SVM approach proposed in this study is the first attempt of using static and longitudinal behavioral data in a direct manner. The computational results show that this approach outperforms other methods.

3. Frameworks for customer churn prediction

In this section, three frameworks for customer churn prediction using longitudinal behavioral data are discussed. The longitudinal behavioral attributes are used in different ways in these frameworks.

Within the CRM context, demographic and transactional data recorded in the data warehouses of business firms have been widely used for customer churn prediction. Customer demographic and transactional data are organized in terms of entity relationship in relational databases (Cao, 2010; Eichinger et al., 2006). Each customer is treated as an observation and n is used to represent the number of observations in the demographic dataset.

Demographic data can be directly used as static attributes after simple data preprocessing such as feature selection and data cleaning for missing values. In a dataset with m_1 static variables, the static attributes of a customer i is usually represented by the input vector $\mathbf{s}_i = \{s_{ij} | j = 1, \ldots, \boldsymbol{m}_1\}$.

Transactional data are transformed into longitudinal behavioral data, i.e., customer-centered multivariate time series of fixed length. The number of longitudinal behavioral attributes is represented by M and the number of time points in the longitudinal behavioral variables is represented by T. The longitudinal behavioral data are represented by a three-dimensional matrix $\{\mathbf{b}_i | i = 1, \ldots, n\}$. Each $\mathbf{b}_i = \{b_{i\tilde{j}t} | \tilde{j} = 1,$