Knowledge-Based Systems 43 (2013) 123–134

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach

Zhen-Yu Chen *, Zhi-Ping Fan

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

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form 18 December 2012
Accepted 21 January 2013
Available online 29 January 2013

Keywords:
Data mining
Customer relationship management
Customer lifetime value
Dynamic multi-step-ahead prediction
Support vector machine
Multiple kernel learning

ABSTRACT

Customer lifetime value (CLV), as an important metric in customer relationship management (CRM), has attracted widespread attention over the last decade. Most CLV prediction models do not take into consideration the dynamics of the customer purchase behavior and changes of the marketing environment such as the adoption of different promotion policies. In this study, a framework for the dynamic CLV prediction using longitudinal data is presented. In the framework, both the dynamic customer purchase behavior and customized promotions are considered. An improved multiple kernel support vector regression (MK-SVR) approach is developed to predict the future CLV and select the best promotion using both the customer behavioral variables and controlled variable about multiple promotions. Computational experiments using two databases show that the MK-SVR exhibits good prediction performance and the usage of longitudinal data in the MK-SVR facilitate the dynamic prediction and promotion optimization.

1. Introduction

Customer relationship management (CRM) has been widely adopted in the business world [19,33,34]. Customer lifetime value (CLV) has been an important metric for CRM to evaluate marketing decisions over the last decade [6,7,16]. The CLV for a firm is defined as the present value of the future profits obtained from a customer [4,21]. The CLV plays an important role in designing marketing programs and allocating marketing resources [4,31]. In the context of customer retention, the individual-level CLV gives a sense of how much is being lost due to churn [36], and the decrease of the CLV is used as a criterion to define churners in non-contractual settings [20]. In the context of customer development, the CLV can be used as the object to select the promotions for the cross-selling and up-selling [27,45].

The CLV prediction is to predict the future profits obtained from a customer using the historical behavior data extracted from the company's database [16,21]. How to predict a customer's CLV is a challenging problem [30]. The study on CLV prediction models has attracted widespread attention over the last decade [16,21,45]. In general, the CLV prediction models can be classified as one-step methods (relationship-level models) and two-step methods (service-level models) [4,16].

The one-step methods directly predict the total profits of multi-product or multi-service. The techniques applied in the one-step methods include linear regression [16,31], quantile regression [4], neural network [31], random forests [16] and Markov chain models [17].

In two-step methods, the probability of a customer buying a product or service is predicted firstly, and then the probabilities with the margins associated with the product or service are combined to obtain the CLV. The techniques applied in predicting the purchase probabilities include probit model [16,45], Pareto/NBD model [18,21], the hierarchical Bayes model [6] and the proportional hazard model [16].

The above studies made great contributions to the CLV prediction. However, these studies fall into the following limitations. First, most CLV models do not take into consideration the dynamics of the CLV [26,27,33]. Second, the commonly used regression models are not ideally suitable for the CLV prediction [4].

Usually, there are two sources of the dynamics of the CLV which should be considered in the CLV prediction. One source is the changes of the marketing environment such as the uses of different marketing policies, and the other is the dynamics of the customer purchase behavior.

Jain and Singh [26] suggested developing a more "complete" CLV model which considers the factors such as the drivers of customers to purchase and the effect of the marketing activities on these drivers. According to their suggestion, the variables incorporated into the CLV models can be classified as the independent variables (uncontrollable variables) and the controlled variables.

* 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).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.01.022

The independent variables used in the CLV models include the customer demography and purchase behavior such as recency, frequency and monetary (RFM) [18]. The controlled variables are those that drive customers to purchase over time [26,33]. The marketing interventions such as customized promotions [27] and the customer-firm interactions [33,37] are two kinds of controlled variables. Khan [27] proposed a dynamic programming model to investigate the impact of customized promotion on the decision to buy and expenditure and determine the sequence and timing of promotions. Although they pointed out that expected profits over eight weeks were maximized, whether or not the expected profits were computed by a prediction model and how to compute the expected profits were not involved in their article. Rust and Chung [37] suggested that the further work should determine the multiple personalized marketing interventions to maximize the CLV over time.

Chen et al. [10] pointed out that the dynamics of the customer purchase behavior can be described as customer-centered longitudinal data. Longitudinal data are widely available in databases of business firms. For example, the RFM variables are expressed as longitudinal time series; the purchase product variables are expressed as longitudinal symbolic sequences. Longitudinal data of a single customer is a multivariate data stream. Longitudinal data with rich historical information have potential to improve the CLV prediction. Netzer et al. [33] proposed a hidden Markov model to dynamically segment the customer base and determine personally methods of marketing interventions. Reinartz and Kumar 35] applied Pareto/NBD model with longitudinal information to identify factors influencing the profitable lifetime duration. At the best of our knowledge, few models applied longitudinal data to the CLV prediction.

Benoit and Van den Poel [4] pointed out that the commonly used regression models are not ideally suitable for the CLV prediction. In two-step methods, the customers buying a product or service are a small fraction of all customers. It is a challenge to model rare events (imbalanced data) [4,24]. In one-step methods, it is also a challenge for the regression techniques to model the CLV with a strong non-normal distribution [4]. The support vector regression (SVR), proposed by Vapnik [43], exhibits good robustness properties [14]. The SVR has been successfully used for the protein structure prediction in bioinformatics where strong skewed distribution occurs [39,47]. However, until recently, the SVR has not been used to predict the CLV. Selecting the best kernel function and the hyperparameters greatly influence the performance of SVR. Multiple kernel learning (MKL) can adaptively determine the optimal kernel by optimizing the combinations of multiple basic kernels [2,8–10,29]. The MKL methods have been used for supervised classification such as customer churn prediction [10] and customer purchase prediction [9], and regression such as stock price prediction [46]. It is fascinating to investigate the performance of multiple kernel SVR (MK-SVR) on the CLV prediction.

In this study, a framework for the dynamic CLV prediction using longitudinal data is developed. As two sources of the dynamics of the CLV, the dynamic customer purchase behavior and customized promotions are two parallel lines in the framework. An improved MK-SVR approach is proposed to simultaneously model the customer behavioral variables and controlled variables about promotions using multiple different kernels. A multiple phase algorithm is developed to train the MK-SVR, predict the future CLV over some time periods, and determine the dynamic customized promotion policy to maximize the future CLV. Besides, the criteria to evaluate the performance of the dynamic multi-step-ahead prediction and promotion optimization are presented.

This paper is organized as follows. Section 2 describes the preliminaries of the CLV computation, the standard SVR and MKL. Section 3 presents the framework of the dynamic CLV prediction using longitudinal data. The model formulation and the algorithm of the proposed MK-SVR are presented in Section 4. The computational experiments are described in Section 5. The computational results are reported in Section 6. Conclusions are given in Section 7.

2. Preliminaries

In this section, the definition and the computation formula of CLV are given. The theoretical foundations of SVR and MKL are also presented.

2.1. Customer lifetime value

CLV is generally defined as the present value of all future profits obtained from a customer over the whole life of relationship with a company [24]. In theory, the CLV prediction should estimate the total net profit a company can expect from a customer [34]. In practice, most studies predict the CLV over a finite time period such as the future three years [4].

The formula of CLV for customer i at the current time T for a finite time period $\widetilde{T}$ which was first formulated by Berger and Nasr [5] is

$$\mathrm{CLV}_{i,T} = \sum_{\tau=0}^{\widetilde{T}} \frac{\mathrm{Profit}_{i,T+\tau}}{(1+d)^{\tau}} \tag{1}$$

where $\widetilde{T}$ denotes the time period for predicting CLV, $\mathrm{Profit}_{i,T}$ denotes the expected profits obtained from customer i at time T, and d denotes the discount rate. Usually, the discount rate is a pre-determined constant [21]. The weighted average cost of capital (WACC) of the firm can be used as the discount rate [21,24,28]. The prediction of the individual-level CLV is usually used for customer segmentation such as churn prediction and marketing resources allocation. The customers and the promotion policies can be ranked without considering the effect of the discount rate on the CLV. In the following experiments, the discount rate is set as 15% per year or 1.25% per month [28].

In multi-product (multi-service) industries, $\mathrm{Profit}_{i,T}$ can be written as

$$\mathrm{Profit}_{i,T} = \sum_{p=1}^{P} \mathrm{Procut}_{i,p,T} \times \mathrm{Usage}_{i,p,T} \times \mathrm{Margin}_{p,T} \tag{2}$$

where P denotes the number of the product (service) categories, $\mathrm{Procut}_{i,p,T}$ denotes a dummy indicating whether a customer i purchases the product (service) p at the current time T, $\mathrm{Usage}_{i,p,T}$ denotes the amount of products (services) purchased, and $\mathrm{Margin}_{p,T}$ denotes the average profit margin of the product (service) p.

2.2. Support vector regression

The details of SVR can be found in Smola and Schölkopf [38]. Given a training dataset $G = \{(\mathbf{x}_1 y_1), \ldots, (\mathbf{x}_n, y_n)\}$, we construct a linear regression function in a high dimensional feature space

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \tag{3}$$

where $\mathbf{x} \in \Re^m$ is the input vector, $\phi(\mathbf{x})$ is a nonlinear map, $\mathbf{w}$ is the weighting vector and b is the bias.

The weighting vector $\mathbf{w}$ and the bias b are obtained by solving the following quadratic program

$$\min \quad J(\mathbf{w}, b, \xi, \xi^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4}$$

$$\text{s.t.} \quad y_i - (\mathbf{w}^T\phi(\mathbf{x}_i) + b) \leqslant \varepsilon + \xi_i, \tag{5}$$

$$(\mathbf{w}^T\phi(\mathbf{x}_i) + b) - y_i \leqslant \varepsilon + \xi_i^*, \tag{6}$$

$$\xi_i, \xi_i^* \geqslant 0, \quad i = 1, \ldots, n \tag{7}$$