Knowledge-Based Systems 35 (2012) 111–119

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM 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 16 September 2011
Received in revised form 20 April 2012
Accepted 20 April 2012
Available online 28 April 2012

Keywords:
Customer relationship management
Behavior analysis
Customer churn prediction
Customer purchase prediction
Support vector machine
Multiple kernel learning

ABSTRACT

In the customer-centered marketplace, the understanding of customer behavior is a critical success factor. The big databases in an organization usually involve multiplex data such as static, time series, symbolic sequential and textual data which are separately stored in different databases of different sections. It poses a challenge to traditional centralized customer behavior prediction. In this study, a novel approach called collaborative multiple kernel support vector machine (C-MK-SVM) is developed for distributed customer behavior prediction using multiplex data. The alternating direction method of multipliers (ADMM) is used for the global optimization of the distributed sub-models in C-MK-SVM. Computational experiments on a practical retail dataset are reported. Computational results show that C-MK-SVM exhibits better customer behavior prediction performance and higher computational speed than support vector machine and multiple kernel support vector machine.

1. Introduction

Today business is evolving from the product-centered to a customer-centered environment [1,2]. The in-depth understanding of customer behavior is a critical success factor to build long term, profitable relationships with specific customers in global competitive marketplace [3]. Therefore, customer behavior prediction is a crucial means of analytical customer relationship management (CRM) [4,5].

Yan et al. proposed a framework of customer behavior prediction for customer retention and profit maximization in telecommunications [6]. This framework consists of five components: churn prediction, churn reason prediction, offer acceptance prediction, revenue estimate, and collection risk estimate. Kim et al. proposed a methodology to identify the propensity of a specific customer to buy a product to enhance the one-to-one marketing [7]. Prinzie et al. used the durable acquisition sequence information and duration information to propose a Next-Product-to-Buy model for cross-selling [8].

In summary, customer behavior prediction mainly involves customer churn prediction and customer purchase prediction. Customer churn prediction is a part of loyalty programs. It aims at identifying the customers who are prone to switch at least some of their purchases from one company, and assisting to companies in improving intervention strategies to convince these customers to stay [2,4,9–13]. Customer purchase prediction aims at predicting whether or not the customer is prone to purchase or repeatedly purchase the product, or predicting the product group from which the customer is prone to purchase his next product [4,14–19]. Therefore, customer purchase prediction is regarded as an important basis for direct marketers to target personalized advertise and promotion activities to specific classes of customers [4,14].

Data is becoming one of the top priorities for information services executives. From the data mining perspective, customer behavior prediction can be regarded as a classification problem that is one of the most common tasks in data mining [4,7]. Data mining techniques such as artificial neural network (ANN) [9], support vector machine (SVM) [1,11,13], Bayesian network [15,16], and ensemble learning [7] are widely used to predict customer behavior and bring potentially useful decision information.

All above studies make contributions to centralized customer behavior prediction. With the growth of database technologies, the limitations of traditional centralized customer behavior prediction arise. The databases within an organization which collect and store internal and external data increase dramatically [20]. Internal data refer to data generated from systems within an organization, such as customer demographic data, transactional data, product-based data, customer review and complaint data [20]. External data refer to data that is not generated by systems within an organization, such as government census data, industry benchmark

* 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 © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.04.023

data, customer psychographic data and economic data [20,21]. Internal and external data involve multiple types. For example, customer demographic and psychographic data are static data; customer transactional data and economic data are multivariate time series data [3,22]; product acquisition data is symbolic sequential data [8,18,19,23]; customer review and complaint data are textual data [2]. In customer behavior prediction, the time series data are usually transformed into static data through aggregation [33,35], and the symbolic sequential and textual data are usually neglected [2]. Moreover, multiplex data are usually separately stored in different databases of different sections [24]. Traditional centralized customer behavior prediction mainly uses static data [3,23]. It faces the challenge of integrating multiple distributed data sources and multiple types of data to reach the combined prediction results.

With the emergence of advanced computing technologies such as sensor networks and cloud computing, lots of efforts have been made on distributed learning in recent years [24,25]. Support vector machine (SVM) is a state-of-art machine learning approach [26,27]. Multiple kernel support vector machine (MK-SVM) is a popular topic in kernel methods [28–31,42,44,45]. It aims at learning the optimal kernel function by optimizing the combination of multiple heterogeneous basic kernels [31,45], multiple basic kernels with different feature subsets [28] or hyperparameters [44]. MK-SVM has been solved by an optimization method such as semi-infinite linear program, quadratically-constrained quadratic program, gradient-based approaches and block coordinator descent approach [28–31,42,44,45]. MK-SVM is well suited to integrating multiplex data. The topics of distributed SVM have been discussed extensively in recent years [32,38,39]. However, little research has been done on distributed MK-SVM. At the best of our knowledge, it is the first paper on improved MK-SVM for distributed customer behavior prediction using multiplex data.

The alternating direction method of multipliers (ADMM) was first proposed by Glowinski and Marrocco in 1975 [36] and Gabay and Mercier in 1976 [37]. Boyd et al. provided a detail discussion of ADMM [25]. They discussed the applications of ADMM to many convex optimization problems such as Basis Pursuit and Lasso. Moreover, Boyd et al. considered two ways including splitting across examples and splitting across features to solve the convex optimization problems in a distributed manner. Hence, ADMM is a powerful algorithm for distributed convex optimization and has the potential for solving MK-SVM in a distributed way.

In this study, a novel approach called collaborative multiple kernel support vector machine (C-MK-SVM) is developed for distributed customer behavior prediction using multiplex data. In comparison with traditional customer behavior prediction, the major contributions of this study are summarized as follows. Firstly, a framework of distributed customer behavior prediction is proposed to integrate multiple distributed data sources and multiple types of data to improve the prediction accuracy. Secondly, in this framework, a collaborative MK-SVM (C-MK-SVM) approach is developed to model multiple feature subsets and multiple sample subsets in a decomposition-coordination manner. In C-MK-SVM, ADMM is applied to the global optimization of the sub-models with multiple basic kernels in the local processors.

This paper is organized as follows. In the next section, the fundamentals of SVM are explained. In Section 3, the proposed method C-MK-SVM is specified. A framework of C-MK-SVM for distributed customer behavior prediction is developed in Section 4. The computational experiments are reported in Section 5. In Section 6, the conclusions are presented.

2. Fundamentals of support vector machine

Given a labeled training set $G = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i \in \Re^m$ and $y_i \in \{+1, -1\}$, SVM constructs the optimal hyperplane in the feature space:

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \cdot \phi(\mathbf{x}) + b \tag{1}$$

where $\phi(\mathbf{x}) : \Re^m \mapsto \Re^{m'}$ with $m \ll m'$ is a nonlinear function which maps the training samples from the input space to a higher-dimensional feature space.

SVM optimizes with respect to $\mathbf{w}$ and b so as to maximize the margin between the positive and negative class and minimize the empirical error:

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

$$\text{s.t.} \begin{cases} y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0, \quad i = 1, \ldots, n \end{cases} \tag{3}$$

where C is a regularization parameter.

The Lagrangian dual of the primal problem is

$$L(\mathbf{w}, \xi, b; \boldsymbol{\alpha}) = J(\mathbf{w}, \xi, b) - \sum_{i=1}^{n}\alpha_i\{y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) + \xi_i - 1\} - \sum_{i=1}^{n}\mu_i\xi_i \tag{4}$$

where α_i is the Lagrange multiplier, and $\boldsymbol{\alpha}$ is used to denote the vector of all Lagrange multipliers.

The Lagrange multipliers $\boldsymbol{\alpha}$ are the solutions of the following dual quadratic program:

$$\max \ L(\mathbf{w}, \xi, b; \boldsymbol{\alpha}) = \left\{\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)\right\} \tag{5}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} y_i\alpha_i = 0 \\ 0 \leqslant \alpha_i \leqslant C, \quad i = 1, \ldots, n \end{cases} \tag{6}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function which replaces the computation of the nonlinear map $\phi(\mathbf{x})$. The Gaussian kernel function is

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right). \tag{7}$$

The classification function of SVM is

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right). \tag{8}$$

3. Collaborative multiple kernel support vector machine

In this section, MK-SVM with the unweighted combination of basic kernels are briefly discussed, the model of C-MK-SVM is formulated, and the training algorithm of C-MK-SVM using ADMM are presented.

3.1. MK-SVM with the unweighted sum of baisc kernels

When multiplex data is used, the independent variables of an observation i are represented by a vector $\mathbf{x}_i = \{\mathbf{x}_i^{k_1} | k_1 = 1, \ldots, m_1\}$ where k_1 indexes the feature subset using a type of data, m_1 is the number of the feature subsets and the number of basic kernels as well. In the feature subset k_1, $\mathbf{x}_i^{k_1} \in \Re^{M_{k_1}}$ is a sub-vector where M_{k_1} is the number of attributes in k_1. The independent variables