Expert Systems with Applications 39 (2012) 8975–8985

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Pattern selection for support vector regression based response modeling

Dongil Kim, Sungzoon Cho *

Seoul National University, 599 Gwanangno, Gwanak-gu, Seoul 151-744, Republic of Korea

ARTICLE INFO

Keywords:
Response modeling
Support vector regression
Pattern selection
Training complexity

ABSTRACT

Two-stage response modeling, identifying respondents and then ranking them according to their expected profit, was proposed in order to increase the profit of direct marketing. For the second stage of two-stage response modeling, support vector regression (SVR) has been successfully employed due to its great generalization performances. However, the training complexities of SVR have made it difficult to apply to response modeling based on the large amount of data. In this paper, we propose a pattern selection method called Expected Margin based Pattern Selection (EMPS) to reduce the training complexities of SVR for use as a response modeling dataset with high dimensionality and high nonlinearity. EMPS estimates the expected margin for all training patterns and selects patterns which are likely to become support vectors. The experimental results involving 20 benchmark datasets and one real-world marketing dataset showed that EMPS improved SVR efficiency for response modeling.

1. Introduction

A response model identifies customers who are likely to respond and the amount of profit expected from each customer using customer databases consisting of demographic data and purchase history for the purpose of direct marketing. With this model, marketers are able to decide who to contact within a limited marketing budget. A well-targeted response model can increase profit, while a mis-targeted response model not only decreases profit but also worsens the relationship between the company and customers (Blattberg, Neslin, & Kim, 2008; Gönül, Kim, & Shi, 2000; Shin & Cho, 2006).

Usually, a response model employs a classification model to predict the likelihood to respond of each customer. Then, those likelihoods are directly used to sort the predicted respondents. However, as pointed out in KDD98 Cup (1998), there may be an inverse correlation between the likelihood to respond and the dollar amount to spend to some marketing datasets (Kim, Lee, & Cho, 2008; Wang, Zhou, Yang, & Yeung, 2005). In this case, profit may not be maximized because some low-spending customers are top-ranked, while some high-spending customers may be low-ranked. Therefore, an additional effort to maximize the profit should be added to the conventional response modeling. Two-stage response modeling (see Fig. 1), identifying respondents at the first stage and then ranking them according to expected profit at the second stage, was proposed to overcome this problem (Kim et al., 2008). In the first stage, conventional classification response models can be directly applied to predict desirable respondents based on their likelihood to respond. However, for the second stage, a new model is needed to estimate the purchase amount of respondents.

Support vector regression (SVR) is one possible solution for use in the second stage of the two-stage response modeling. Even though SVR has an ability to train nonlinear patterns with great generalization performances (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Smola & Schölkopf, 2002; Vapnik, 1995), one drawback of this algorithm is the training complexity. The training complexity of SVR is strongly correlated to the number of training patterns, as is that of Support Vector Machines (SVM): $O(n^3)$ of the training time complexity and $O(n^2)$ of the training memory complexity, where n is the number of training patterns. The training time of SVR is expensive, and occasionally, SVR does not work in a limited memory space for large datasets. In addition, response modeling datasets usually consist of very large training patterns, sometimes including billions of transactions from millions of customers. In addition, data analysis for a marketing campaign includes the construction of various models with different samples of a dataset to verify multiple marketing actions. Moreover, SVR contains an additional hyper-parameter which requires that the SVM classifier, ε, be set empirically. Hence, response modeling with SVR consists of a repeated modeling process with a very large dataset and including parameter searching processes. The training complexity of SVR must be reduced for use in practical two-stage response modeling.

To overcome this training complexity problem, decomposition methods, such as Chunking, SMO, SVM^{light}, SOR and LIBSVM, have been proposed in order to divide the original optimization problem into a series of smaller problems (Platt, 1999). However, the

* Corresponding author. Tel.: +82 2 883 4913; fax: +82 2 889 8564.
E-mail addresses: dikim01@snu.ac.kr (D. Kim), zoon@snu.ac.kr (S. Cho).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.026

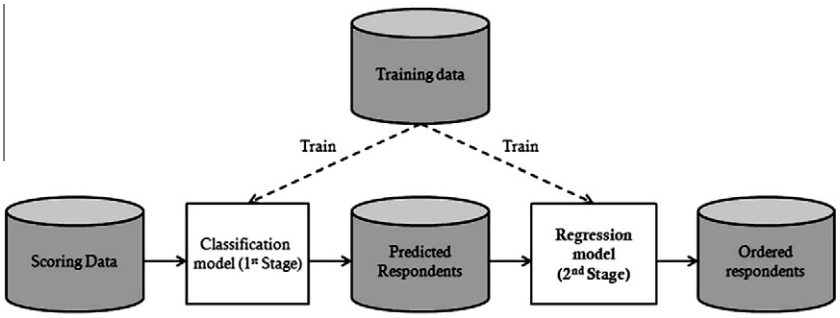

Fig. 1. Concept of the two-stage response model.

training complexities of these methods are still strongly correlated with the number of training patterns (Shin & Cho, 2007). Other research studies have focused on the pattern selection method, which relies on the assumption that there are important or informative patterns among the redundance patterns and noise patterns in the original dataset. However, most pattern selection studies using methods such as SVM-KM (Almeida, Braga, & Braga, 2000), NPPS (Shin & Cho, 2007), a cross-training based method (Bakir, Bottou, & Weston, 2005) and a linear SVM based method (Joachims, 2006) have focused their efforts on classification problems, rather than regression problems. For regression problems, HSVM (Wang & Xu, 2004), a k-NN based method (Sun & Cho, 2006) and an ε-tube based method (Kim & Cho, 2006) have been proposed. However, there have been some drawbacks with the use of these methods. HSVM is basically useful for time-series problems. Hence, an additional effort is needed to conduct the partitioning part of HSVM for non-time-series problems. Moreover, HSVM and the k-NN based method have cut-off parameters that directly and empirically determine the number of patterns. The ε-tube based method tends to produce inferior performances when it is applied to high-dimensional datasets. We previously proposed another method in which the likelihood of becoming a support vector is estimated with binary voting (Kim & Cho, 2008). However, we found that this method resulted in information loss. Also, our previous method cannot control the trade-off between training complexity and model performance. In real-world applications, a parameter to control the trade-off within a reasonable bound is needed.

Since the training complexity of SVR is highly related to the number of training patterns, and the size of the response modeling dataset is large, the pattern selection method is preferred to reduce the training complexity of SVR for response modeling. For example, if the size of training dataset is reduced to 70%, the training time complexity of SVR decreases by 34.3% of original dataset. The pattern selection method for response modeling should consider the following. First, since response modeling aims to maximize profit, the pattern selection method reduces the number of training patterns with minimum loss of generality. Second, the pattern selection method should be able to adapt to high dimensional and highly nonlinear problems. In addition, since users do not know the appropriate number of selected training patterns, the pattern selection method automatically determines the number of patterns selected. Finally, the pattern selection method should be able to control the trade-off between SVR performance and training time to adapt to various types of marketing circumstances and modeling processes.

In this paper, we propose a new pattern selection method called Expected Margin based Pattern Selection (EMPS) to reduce the training complexity of SVR for two-stage response modeling. For SVR, the most important patterns are Support Vectors (SVs), which affect the construction of a regression model. However, before training, there is no way to identify which training patterns will become the SVs. Using the fact that SVs are always located outside the ε-tube (see Fig. 2), the training patterns with a margin greater than ε should be selected. With multiple bootstrap learning, we estimated the expected margins of all training patterns and selected the patterns with a margin greater than ε. EMPS automatically determines the number of patterns selected according to a parameter α which allows EMPS to control the trade-off between the training complexity and model performance. The experiments were conducted on 20 benchmark datasets, as well as one real-world response modeling dataset. Unlike the previous research in two-stage response modeling, which implemented only the second stage of the two-stage response modeling, we implemented and experimented with both the first and the second stage of the two-stage response modeling involving classification models and SVR with EMPS.

The remainder of this paper is organized as follows. In Section 2, we explain the main idea of EMPS and state the algorithm along with a brief review of SVR. Also, the experimental results of EMPS for 20 datasets are included. In Section 3, we introduce a real-world marketing dataset, as well as the experimental results. In Section 4, we summarize the results and conclude the paper with remarks on the limitations of this study and the directions for future research.

2. Support vector regression and expected margin based pattern selection

2.1. Support vector regression

For a brief review of SVR, consider a regression function $f(\mathbf{x})$ to be estimated with training patterns $\{(\mathbf{x}_i, y_i), i = 1, \ldots, n\}$ as follows:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b, \quad \text{with} \quad \mathbf{w}, \mathbf{x} \in \mathbb{R}^d, b \in \mathbb{R}$$
$$\text{where} \quad \{(\mathbf{x}_1, y_1), \ldots (\mathbf{x}_n, y_n)\} \subset \mathbb{R}^d \times \mathbb{R}. \tag{1}$$

According to the SRM principle, the generalization accuracy is optimized by the flatness of the regression function. Since the flatness is

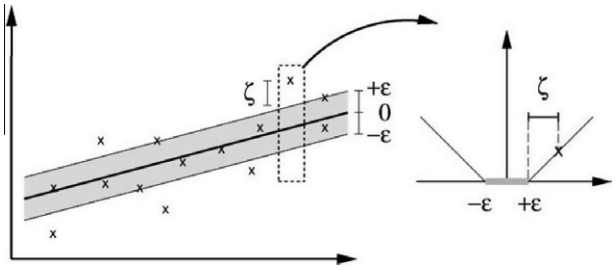

Fig. 2. ε-Tube based on the margin of training patterns and the ε-loss function of SVR (Smola & Schölkopf, 2002).