Pattern Recognition 60 (2016) 296–305

Contents lists available at ScienceDirect

Pattern Recognition

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

Structural nonparallel support vector machine for pattern recognition

Dandan Chen [a,b], Yingjie Tian [b,c,*], Xiaohui Liu [d]

[a] College of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China
[d] Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, UK

ARTICLE INFO

Article history:
Received 23 September 2015
Received in revised form
7 March 2016
Accepted 29 April 2016
Available online 17 May 2016

Keywords:
Structural information
Nonparallel support vector machine
Alternating direction method of multipliers
Pattern recognition

ABSTRACT

It has been widely accepted that the underlying structural information in the training data within classes is significant for a good classifier in real-world problems. However, existing structural classifiers do not balance structural information's relationships both intra-class and inter-class. Combining the structural information with nonparallel support vector machine (NPSVM), we design a new structural nonparallel support vector machine (called SNPSVM). Each model of SNPSVM considers not only the compactness in both classes by the structural information but also the separability between classes, thus it can fully exploit prior knowledge to directly improve the algorithm's generalization capacity. Furthermore, we apply the improved alternating direction method of multipliers (ADMM) to SNPSVM. Both our model itself and the solving algorithm can guarantee that it can deal with large-scale classification problems with a huge number of instances as well as features. Experimental results show that SNPSVM is superior to the other current algorithms based on structural information of data in both computation time and classification accuracy.

1. Introduction

Support vector machine (SVM) [1,2], as an efficient large margin classifier derived from statistical learning theory [3], has already reached many achievements in a wide variety of problems in machine learning [4–7]. According to the central idea of traditional SVM for classification, two parallel support hyperplanes are constructed following the principle that the band between the two hyperplanes separates positive and negative data points and the margin between the two hyperplanes should be maximized [8], thus the final decision separating hyperplane is located in the middle of the two parallel hyperplanes [9]. The nonlinear case of the SVM can be successfully solved by introducing kernel trick into the dual quadratic programming problems (QPPs) [10,11].

However, there are two key issues that limit effectiveness of the traditional SVM. The first one is that the final decision separating hyperplane is located unbiasedly right in the middle of two parallel hyperplanes, which loses the prior information within the classes. For this reason, many scholars have moved on to the classifiers with the nonparallel hyperplanes. The second issue is that the standard SVM does not effectively take advantage of the prior data distribution information within classes ignoring the fact that a good classifier should be sensitive to the structure of the data distribution [9]. For this consideration, methods of extracting the data distribution information within classes and techniques of putting these information into the classic classifiers are emerging in machine learning.

Over the last decade, lots of nonparallel hyperplane classifiers have been proposed and one of the most popular is the twin support vector machine (TWSVM). It seeks two nonparallel proximal hyperplanes such that each hyperplane is closer to one of the two classes and is at least one distance from the other. So far, there are four representative TWSVMs, that is, original TWSVM [12], twin bounded SVM (TBSVM) [13], improved TWSVM (ITSVM) [14] and nonparallel SVM (NPSVM) [15,16]. Despite that there are many improved versions of TWSVMs [17–19], they can be classified as one of the above four types. Since NPSVM is the most successful one among all these nonparallel classifiers, it is chosen as the foundation of our improved model.

At the same time, researchers have proposed many improved algorithms to modify the performance of standard SVM by embedding the structural information into the models [20]. The methods that extract the structural data distribution information within classes can be roughly divided into two categories: The first assumes that the data actually lie on a submanifold in the input space and a typical paradigm is Laplacian support vector machine (LapSVM) [21]. The second assumes that the data contains clusters,

* Corresponding author at: Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China.
E-mail addresses: chendandan1@126.com (D. Chen), tyj@ucas.ac.cn (Y. Tian), xiaohui.liu@brunel.ac.uk (X. Liu).

http://dx.doi.org/10.1016/j.patcog.2016.04.017
0031-3203/

which has yielded a class of effective structural large margin classifiers, such as minimax probability machine (MPM) [22], maxi-min margin machine (M^4) [8], structural large margin machine (SLMM) [9]. The most successful one is SLMM, but it is limited by its high computational complexity to solve a sequential second-order cone programming (SOCP). Consequently, structural regularized support vector machine (SRSVM), introduced by Xue et al. [23], naturally integrates the prior structural information within classes into SVM, without destroying the classical framework of SVM.

Either the nonparallel classifiers or the structural machines can extract and make use of hidden information in the training data from different perspectives, yet learning with a single one of them is likely insufficient. In this paper, by combining the idea of SRSVM with NPSVM [15], we propose a novel structural nonparallel support vector machine (SNPSVM), which inherits their strengthes, and meanwhile overcomes their weaknesses.

Our main contributions are summarized as follows:

(1) The framework of SNPSVM is established, which further improves the performance of NPSVM by making the most of the data distribution information. Specifically, SNPSVM not only digs potential information within each class efficiently but also subtly make up the inadequacy of the original SRSVM model. Furthermore, SNPSVM has advantages over existing structural twin SVMs [24–26], because it inherits all the advantages that NPSVM has over other twin SVMs such as its elegant dual form and avoiding the matrix inversion.

(2) It has been proved in another perspective that our SNPSVM is preferable to structural twin SVMs [25]. As a model for comparison, SNPSVM* is built, which is the same as SNPSVM except the embedding mode of the structural information. Based on Ward's linkage clustering (the same clustering technology of SRSVM, STWSVM for comparative purpose), SNPSVM embeds the data structures within every class into each optimization problem of NPSVM, rather than considering prior information within only one class for one model as the STWSVM and SNPSVM* do. Since each model of all kinds of twin SVMs is built by the samples in both positive and negative class, this design seems to be more reasonable, which has been experimentally verified (SNPSVM preforms better than STWSVM and SNPSVM*).

(3) In order to improve the training efficiency of our model, we adopt an improved version of the *alternating direction method of multipliers* (ADMM) [27] for SNPSVM, which is a simple but powerful algorithm that is well suited to distributed convex optimization and in particular to large-scale problems in applied statistics and machine learning. For both the SNPSVM model and the ADMM algorithm, parallel computing is feasible, which also allows us to model for large-scale classification problems. In fact, the L_1-NPSVM [28] itself can deal with large-scale classification problems with a huge number of instances as well as features.

(4) Experiments on the synthetic and real-world data sets show that our model not only achieves remarkable classification performance with higher accuracies, but also possesses powerful generalization ability.

The rest of this paper is organized as follows. Section 2 gives a brief review of the SRSVM and NPSVM. Section 3 defines the framework of SNPSVM and its dual form and discusses the kernelization for the nonlinear versions. Section 4 derives the *alternating direction method of multipliers* for SNPSVM. Section 5 shows the experimental results and analyzes the efficiency of SNPSVM by comparing with other classifiers. Some conclusions are drawn in Section 6.

2. Background

In this section, we briefly introduce the SRSVM and NPSVM. Given a training sample set of input-output pairs $\{(x_1, y_1), \dots, (x_n, y_n)\}$, $x_i \in \mathbb{R}^m$, $i = 1, \dots, n$ drawn from an unknown distribution. This paper takes the binary classification as example, that is, $y_i \in \{1, -1\}$, $i = 1, \dots, n$.

2.1. SRSVM

After clustering, two sets of d_1 and d_2 clusters are obtained, respectively, in the two classes. $P = P_1 \cup \cdots P_i \cup \cdots \cup P_{d_1}$ denotes clusters belonging to positive class, and $N = N_1 \cup \cdots N_j \cup \cdots \cup N_{d_2}$ denotes clusters belonging to negative class. The SRSVM model can be formulated as:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}\|w\|^2 + \frac{\lambda}{2} w^T \Sigma w \\ \text{s. t.} \quad & y_i(w^T w_i + b) \ge 1, \quad i = 1, 2, \dots, n \end{aligned} \tag{1}$$

where $\Sigma = \Sigma_{P_1} + \cdots + \Sigma_{P_{d_1}} + \Sigma_{N_1} + \cdots + \Sigma_{N_{d_2}}$, Σ_{P_i} and Σ_{N_j} are the covariance matrices corresponding to the ith and jth clusters in the two classes, d_1 and d_2 are the numbers of clusters in positive and negative class, respectively. λ is a nonnegative parameter that regulates the relative importance of the structural information.

2.2. Nonparallel support vector machine (NPSVM)

The nonparallel support vector machine model seeks the two nonparallel hyperplanes by solving two convex QPPs

$$\begin{aligned} \min_{w_+, b_+, \eta_+^{(*)}, \xi_-} \quad & \frac{1}{2}\|w_+\|^2 + C_1 \sum_{i=1}^{p} (\eta_i + \eta_i^*) + C_2 \sum_{j=p+1}^{p+q} \xi_j \\ \text{s. t.} \quad & (w_+ \cdot x_i) + b_+ \le \varepsilon + \eta_i, \quad i = 1, \dots, p, \\ & -(w_+ \cdot x_i) - b_+ \le \varepsilon + \eta_i^*, \quad i = 1, \dots, p, \\ & (w_+ \cdot x_j) + b_+ \le -1 + \xi_j, \quad j = p+1, \dots, p+q, \\ & \eta_i, \eta_i^* \ge 0, \quad \xi_j \ge 0, \quad i = 1, \dots, p, \quad j = p+1, \dots, p+q \end{aligned} \tag{2}$$

and

$$\begin{aligned} \min_{w_-, b_-, \eta_-^{(*)}, \xi_+} \quad & \frac{1}{2}\|w_-\|^2 + C_3 \sum_{i=p+1}^{p+q} (\eta_i + \eta_i^*) + C_4 \sum_{j=1}^{p} \xi_j \\ \text{s. t.} \quad & (w_- \cdot x_i) + b_- \le \varepsilon + \eta_i, \quad i = p+1, \dots, p+q, \\ & -(w_- \cdot x_i) - b_- \le \varepsilon + \eta_i^*, \quad i = p+1, \dots, p+q, \\ & (w_- \cdot x_j) + b_- \ge 1 - \xi_j, \quad j = 1, \dots, p, \\ & \eta_i, \eta_i^* \ge 0, \quad \xi_j \ge 0, \quad i = p+1, \dots, p+q, \quad j = 1, \dots, p \end{aligned} \tag{3}$$

where x_i $(i = 1, \dots, p)$ are positive inputs, and x_j $(j = p+1, \dots, p+q)$ are negative inputs, $C_i \ge 0$ $(i = 1, \dots, 4)$ are penalty parameters, $\xi_+ = (\xi_1, \dots, \xi_p)^\top$, $\xi_- = (\xi_{p+1}, \dots, \xi_{p+q})^\top$, $\eta_+^{(*)} = (\eta_+^\top, \eta_+^{*\top})^\top = (\eta_1, \dots, \eta_p, \eta_1^*, \dots, \eta_p^*)^\top$, $\eta_-^{(*)} = (\eta_-^\top, \eta_-^{*\top})^\top = (\eta_{p+1}, \dots, \eta_{p+q}, \eta_{p+1}^*, \dots, \eta_{p+q}^*)^\top$ are slack variables.

3. Structural nonparallel support vector machine (SNPSVM)

In this section, we propose a novel nonparallel classifier, termed as SNPSVM. Firstly, SNPSVM adopts some clustering techniques to capture the data distribution within classes. Secondly, it embeds a term that minimizes the compactness of clusters inside each class into the objective function of NPSVM, which leads to further maximizing the margin in the sense of incorporating the data structures.