ARTICLE IN PRESS

Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Structural regularized projection twin support vector machine for data classification

Xinjun Peng [a,b,*], Dong Xu [a]

[a] Department of Mathematics, Shanghai Normal University, Shanghai 200234, PR China
[b] Scientific Computing Key Laboratory of Shanghai Universities, Shanghai 200234, PR China

ARTICLE INFO

Article history:
Received 4 January 2013
Received in revised form 3 March 2014
Accepted 5 March 2014
Available online xxxx

Keywords:
Binary classification
Support vector machine
Twin projections
Kernel trick
Cluster-based structural information

ABSTRACT

Projection twin support vector machine (PTSVM) seeks two projection directions for two classes by solving two smaller-sized quadratic programming problems (QPPs), such that the projected samples of one class are well separated from those of the other one in its respective subspace. However, it only simply considers the prior class-based structural information in the optimization problems. In this paper, a structural regularized PTSVM (SRPTSVM) classifier for binary classification is presented. This proposed SRPTSVM focuses on the cluster-based structural information of the corresponding class in each optimization problem, which is vital for designing a good classifier in different real-world problems. This SRPTSVM is extended to a nonlinear version by the kernel trick. Experimental results demonstrate that SRPTSVM is superior in generalization performance to other classifiers.

1. Introduction

In the past decade, support vector machine (SVM) [1–3] has become a hot topic of research in machine learning. Based on the structural risk minimization principle [2,3], classical SVM finds the maximal margin between two classes by solving a quadratic programming problem (QPP) in the dual space. Within a few years after its introduction SVM has already outperformed most other systems in a wide variety of applications [4–7]. However, one of the main challenges in classical SVM is the large computational cost of QPP. The long training time of QPP not only causes classical SVM to take a long time to train on a large database, but also prevents it from locating the optimal parameter set from a very fine grid of parameters over a large span. Another shortcoming of SVM is it usually pays more attention to the separation between classes than the prior structural information within classes in data. In fact, for different real-world problems, different classes may have different underlying data structures.

Recently, a class of classifiers, called nonparallel hyperplane classifiers, have been developed. Jayadeva et al. [8] proposed a twin support vector machine (TWSVM) classifier for binary classification in the light of the generalized eigenvalue proximal support vector machine (GEPSVM) [9]. TWSVM aims at generating a pair of nonparallel planes such that each plane is as close as possible to the corresponding class and is at least one far from the other class. To this end, it solves a pair of smaller-sized QPPs, instead of a large one in traditional SVM, making the learning speed of TWSVM be approximately four times faster than that of SVM in theory [8]. Some extensions to TWSVM include the least squares TWSVM (LSTWSVM) [10], smooth TWSVM (STWSVM) [11], nonparallel-plane proximal classifier (NPPC) [12], v-TWSVM [13], twin-hypersphere SVM

* Corresponding author at: Department of Mathematics, Shanghai Normal University, Shanghai 200234, PR China. Tel.: +86 21 64324866.
E-mail address: xjpeng999@gmail.com (X. Peng).

http://dx.doi.org/10.1016/j.ins.2014.03.129
0020-0255/

Please cite this article in press as: X. Peng, D. Xu, Structural regularized projection twin support vector machine for data classification, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.03.129

(THSVM) [14], twin parametric-margin SVM (TPMSVM) [15], twin support vector regression (TSVR) [16], and twin parametric insensitive support vector regression (TPISVR) [17].

Different from TWSVM which improves GEPSVM by seeking a hyperplane for each class using a SVM-type formulation, a multi-weight vector projection support vector machine (MVSVM) [18] was proposed by seeking a pair of weight vectors, such that the samples of one class are closest to its class meanwhile the samples of the other classes are separated as far as possible. The weight vectors of MVSVM can be found by solving a pair of eigenvalue problems. Further, [19] proposed a projection twin support vector machine (PTSVM) classifier in the light of MVSVM. Instead of solving two eigenvalue problems in MVSVM, PTSVM still solves two related SVM-type problems, similar to that in TWSVM, to obtain the two projection directions. In addition, PTSVM can generate multiple projection axes by using the recursive algorithm, i.e., recursive PTSVM [19]. More recently, a regularization framework for PTSVM by adding a maximum-margin regularization term for each problem [20] was discussed, RPTSVM for short, to overcome the difficulty of PTSVM, in which is the possible singularity of the matrices appeared in its primal problems.

As the relation between the structural information of data and SVM, it is desirable that an SVM classifier be adaptable to the discriminant boundaries to fit the structures in the data, especially for increasing the generalization capacities of the classifier. Fortunately, some algorithms have been developed to focus more attention on the structural information than SVM recently. They provide a novel view in which to design a classifier, that is, a classifier should be sensitive to the structure of data distribution [21]. These algorithms can be mainly divided into two kinds of approaches. The first one is manifold assumption-based, which assumes that the data actually lie on a sub-manifold in the input space. A typical model is Laplacian SVM (LapSVM) [22]. LapSVM constructs a Laplacian graph for each class on top of the local neighborhood of each point to form the corresponding Laplacian (matrix) to reflect the manifold structure of individual-class data. They are then embedded into the traditional framework of SVM as additional manifold regularization terms. The second approach is cluster assumption-based [23], which assumes that the data contains clusters. For instance, structured large margin machine (SLMM) [21], ellipsoidal kernel machine (EKM) [24], minimax probability machine (MPM) [25], and maxi-min margin machine (M^4) [26]. However, the computational cost of these approaches is larger than classical SVM. More recently, Xue et al. proposed a structural regularized SVM (SRSVM) [27]. This SRSVM embeds a cluster granularity into the regularization term to capture the data structure.

For PTSVM and RPTSVM, it can be found that they also embed the class granularity-based structural information [27] into the optimization problems, i.e., the covariance matrix of each class is introduced to minimize the within-class variance. However, this structural information is too rough for real-world problems, which makes PTSVM and RPTSVM cannot find the reasonable projection for each class, then reduce the generalization performance. To overcome this shortcoming, we present an improvement version for PTSVM and RPTSVM in this paper, called the structural regularized PTSVM (SRPTSVM) classifier. This SRPTSVM respectively embeds the data structures of two classes into the optimization problems based on the cluster granularity [27]. That is, in the pair of optimization problems of SRPTSVM, it considers a cluster-based structural regularization term, i.e., it makes the variance of the projected samples of each cluster in one class as small as possible with a weight according to the size of this cluster. Further, SRPTSVM constructs a set of new constraints for each optimization problem based on cluster granularity. In addition, a pair of maximum margin regularization terms in the two optimization problems are introduced to overcome the possible singularity and to implement the structural risk minimization principle as RPTSVM. This SRPTSVM is extended to the nonlinear case by the kernel trick and Woodbury matrix identity [28]. These differences make the pair of axes of our SRPTSVM depicts better projection directions. The experiment results show that this SRPTSVM obtains the better generalization.

The rest of this paper is organized as follows: Section 2 briefly introduces the structural granularities of data, MVSVM, and PTSVM. Section 3 presents the proposed SRPTSVM, including the linear, nonlinear, and recursive SRPTSVM. Experimental results both on the toy and real-world problems are given in Section 4. Some conclusions and remarks are drawn in Section 5.

2. Background

In this paper, the training samples are denoted by a set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{l}$, where $x_i \in \mathcal{X} \subset \mathcal{R}^m$ and $y_i \in \{1, 2\},\ i = 1, \ldots, l$. For simplicity, we use $\mathcal{I}_k,\ k = 1, 2$ to denote the sets of index i such as $y_i = k,\ k = 1, 2$, use the set $\mathcal{I}$ to denote all point indices, i.e., $\mathcal{I} = \mathcal{I}_1 \bigcup \mathcal{I}_2$, and use the matrices $C \in \mathcal{R}^{m\times l},\ A \in \mathcal{R}^{m\times l_1}$ and $B \in \mathcal{R}^{m\times l_2}$ to represent all training points, and points belonging to classes ± 1, respectively, where $l_k = |\mathcal{I}_k|$. In addition, we also use $V_{\mathcal{P}}$ to represent the submatrix of V consisting of columns indexed by the index set $\mathcal{P}$, such as $C_{\mathcal{I}_1} = A$ and $C_{\mathcal{I}_2} = B$.

2.1. Structural granularity

Let $\mathcal{S}_1, \ldots, \mathcal{S}_t$ be a partition of $\mathcal{D}$ according to some relation measure, where the partition characterizes the whole data in the form of some structures such as cluster, and $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_t = \mathcal{D}$. Here $\mathcal{S}_i,\ i = 1, \ldots, t$ is called *structural granularity* [27]. In general, four granularity layers can be differentiated:

Global granularity: The granularity refers to the dataset $\mathcal{D}$. With this granularity, the whole data are characterized or enclosed by a single ellipsoid with center μ and covariance matrix Σ obtained by minimizing its volume [24]:

Please cite this article in press as: X. Peng, D. Xu, Structural regularized projection twin support vector machine for data classification, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.03.129