ARTICLE IN PRESS
JID: KNOSYS [m5G;December 9, 2016;23:12]
Knowledge-Based Systems 000 (2016) 1–14
Contents lists available at ScienceDirect
Knowledge-Based Systems
journal homepage: www.elsevier.com/locate/knosys

Laplacian total margin support vector machine based on within-class scatter

Huimin Pei[a], Yanyan Chen[b], Yankun Wu[a], Ping Zhong[a,*]

[a] College of Science, China Agricultural University, Beijing, 100083, China
[b] College of Applied Science and Technology, Beijing Union University, Beijing 102200, China

ARTICLE INFO

Article history:
Received 6 July 2016
Revised 21 November 2016
Accepted 8 December 2016
Available online xxx

Keywords:
Support vector machine
Total margin
Within-class scatter
Manifold regularization
Semi-supervised learning

ABSTRACT

Insufficient volume of supervised information is a major challenge for supervised learning. An effective method to handle this problem is semi-supervised learning, which can make full use of the geometric information embedded in unlabeled instances. In this paper, we present a novel laplacian total margin support vector machine based on within-class scatter (LapWCS-TSVM) method to deal with the semi-supervised binary classification problem. The proposed LapWCS-TSVM incorporates the total margin algorithm and the manifold regularization into WCS-SVM to help improve its performance. With the help of kernel trick, the proposed LapWCS-TSVM can be easily generalized to non-linear separable case and solved by the optimization programming of the traditional support vector machine. Experiments conducted on artificial datasets, UCI datasets and face recognition datasets show the validity of the newly proposed algorithm.

1. Introduction

In recent years, semi-supervised learning (SSL) which is a very important branch of machine learning has received considerable attention in many research areas [1–4]. One main reason is that the acquisition of labeled instances is usually difficult but the collection of unlabeled ones is much easier in many practical problems. Different from the traditional supervised learning methods which need the label of each training instance, semi-supervised learning methods exploit abundant amounts of unlabeled training instances and rare labeled ones simultaneously with the objective to improve performance of classifiers. Therefore, the way how to use the information embedded in unlabeled instances effectively is the key to the formulation of a semi-supervised learning method.

The manifold regularization [5,6], which introduces a meaningful regularization term to encode the geometric information of the unlabeled instances and make the smoothness of the classifier along the intrinsic manifold, is one of the most elegant frameworks for the semi-supervised learning. Following the manifold regularization framework, a broad community of semi-supervised learning [7–12] has focussed on support vector machine, which has been extensively studied for supervised learning. Support vector machine (SVM) [13,14] which has been successfully applied in many pattern recognition problems due to its attractive features and wonderful practical performance [15–18], is a promising supervised machine learning technique based on solid mathematical background. However, the performance of the supervised algorithm usually deteriorates because of the lacking of enough supervised information. By adding the manifold regularization term to the traditional SVM, Belkin et al. proposed the laplacian support vector machine (LapSVM) [5] which took the underlying geometric information into full consideration to build a more reasonable classifier.

As shown in previous researches [19–27], the geometric information of instances is an important priori knowledge for classifiers. By incorporating the within-class scatter matrix in the traditional SVMs, a lot of effective SVM-based algorithms [19–21] have been proposed. In addition, Arvanitidis et al. [22] utilized the geometric data information described in intrinsic graphs to introduce a novel classification framework that is based on the combination of the SVM and the graph embedding framework. Further, Iosifidis et al. [23] proposed a general multi-class classification framework by incorporating the geometric information of instances described in both intrinsic and penalty graphs in multi-class SVM. Also, the geometric information of instances was exploited to enhance the performance of ELM-based algorithms [24–27]. And the extreme learning machine (ELM) [28] is known as a fast and efficient algorithm for single-hidden layer feedforward neural(SLFN) networks training. It is notable that most of the mentioned algorithms are also supervised algorithms which only consider the labeled instances for their models training. Therefore, it may be beneficial

* Corresponding author.
E-mail address: zping@cau.edu.cn (P. Zhong).

http://dx.doi.org/10.1016/j.knosys.2016.12.009

Please cite this article as: H. Pei et al., Laplacian total margin support vector machine based on within-class scatter, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.12.009

to introduce the algorithms in the framework of semi-supervised learning.

In this paper, we mainly focus on the support vector machine based on minimum within-class scatter (WCS-SVM) [20]. Instead of the maximization of the margin between two classes, WCS-SVM focuses on the minimization of the within-class scatter and the maximization of the margin simultaneously, which provides significant advantages over the traditional SVM. Note that WCS-SVM is also a supervised algorithm which only exploits the labeled instances to train its model. In this paper, we try to introduce the algorithm WCS-SVM in the framework of semi-supervised learning.

In addition, it is generally known that sparsity is a significant advantage for SVM. However, from the perspective of statistical, SVM suffers the disadvantage of the loss of information contained in the majority of training instances. WCS-SVM is a variant of the traditional SVM, so it suffers the same disadvantage. Intuitively, it is unavoidable that these algorithms are less efficient than those who take the lost information into consideration. To handle this problem, Yoon et al. proposed the total margin support vector machine (TM-SVM) which considers the distance between all instances and the separating hyperplane [29]. The total margin algorithm extends the soft margin algorithm by introducing surplus variables to the correctly classified samples. By maximizing surplus variables additionally, the generalization error bound of SVM can be further improved since all instances information is used to construct the separating hyperplane. Therefore, it may be beneficial to use the total margin algorithm to substitute the soft margin algorithm in WCS-SVM.

In this paper, inspired by the studies above, we develop a semi-supervised learning algorithm by incorporating the manifold regularization and the total margin algorithm into WCS-SVM. The newly proposed algorithm is termed as laplacian total margin support vector machine based on within-class scatter or LapWCS-TSVM for short. By adding the manifold regularization term, the proposed LapWCS-TSVM can exploit the geometry information of the margin distribution embedded in unlabeled instances to construct a more reasonable classifier. Furthermore, our algorithm has the potential to obtain good generalization ability because it introduces the total margin algorithm to replace the soft margin algorithm in WCS-SVM. The proposed LapWCS-TSVM can be easily generalized to non-linear separable case by the well known "kernel trick" [14] and it is effortless to implement since it only requires, with the help of the Representer Theorem [5], simple modification of the optimization problem in the traditional SVM. Experiments results on artificial datasets, UCI datasets and face recognition datasets are rendered to show the reliability and feasibility of our proposed LapWCS-TSVM.

The remainder of this paper is organized as follows. We briefly review the related algorithms TM-SVM and WCS-SVM In Section 2. In Section 3, the manifold regularization is succinctly described and our proposed algorithm LapWCS-TSVM including linear and nonlinear cases are described in detail. Moreover, we compare our proposed LapWCS-TSVM with other related algorithms. Section 4 presents the experimental results. Finally, conclusions are drawn in Section 5.

2. Preliminaries

In this section, we briefly review two related algorithms: the total margin support vector machine (TM-SVM) and the support vector machine based on within-class scatter (WCS-SVM).

2.1. TM-SVM

The total margin support vector machine (TM-SVM) [29], which considers the distances between all instances and the separating hyperplane, has extended the existing soft margin algorithm. It aims to improve the generalization error bound of the traditional support vector machine through exploiting the surplus variables.

Given the training set

$$\mathcal{T}_l = \{(x_1, y_1), \ldots, (x_l, y_l)\}, \tag{1}$$

where $x_i \in R^m$ is the input instance, $y_i \in \{+1, -1\}$ is the corresponding class label. Each instance is supposed to be classified by the hyperplane $w^T\phi(x) + b = 0$ in the feature space, where w is the normal vector of the hyperplane, $\phi(\cdot)$ is a nonlinear mapping function from the original low-dimensional input space into a high-dimensional feature space, b is the bias.

By minimizing the sum of slack variables and maximizing the sum of surplus variables simultaneously, the optimization problem of TM-SVM can be formulated as

$$\begin{aligned} \min_{w,b,\xi,\delta} \quad & \frac{1}{2}\|w\|^2 + C_1 e^T\xi - C_1(1-\beta)e^T\delta \\ \text{s.t.} \quad & Y(Dw + eb) \geq e - \xi + \delta, \\ & \xi \geq 0e, \quad \delta \geq 0e, \end{aligned} \tag{2}$$

where C_1 is the regularization parameter; β is constant meeting the condition $0 < \beta < 1$, which can ensure that at least one of ξ_i and δ_i becomes zero; $\xi \in R^l$ is the slack variable vector which measures the distance between the misclassified instances and the hyperplane; $\delta \in R^l$ is the surplus variable vector which measures the distance between the correctly classified instances and the hyperplane; $Y \in R^{l\times l}$ is a diagonal matrix with diagonal elements $Y_{ii} = y_i$ according to the class of the instance; e is the vector of ones of l dimension; $D = [\phi(x_1), \ldots, \phi(x_l)]^T$.

Using the kernel representation $k(\cdot, \cdot)$, we can obtain the following dual of problem (2):

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\alpha^T H\alpha - e^T\alpha \\ \text{s.t.} \quad & e^T Y\alpha = 0, \\ & C_1(1-\beta)e \leq \alpha \leq C_1 e, \end{aligned} \tag{3}$$

where $\alpha \in R^l$ is the lagrange parameter vector; $H = YKY$ in which $K \in R^{l\times l}$ is the Gram matrix with elements $K_{ij} = k(x_i, x_j)$, k is a suitable kernel function. As typical kernel functions, there are the linear kernel function, the polynomial kernel function (POLY) and the radial basis function kernel function (RBF). Later, the linear kernel (4) and RBF kernel (5) which have been confirmed superior to other kernels [30] will be used in our experiments:

$$k(x_i, x_j) = x_i^T x_j, \tag{4}$$

$$k(x_i, x_j) = \exp(-\sigma\|x_i - x_j\|^2), \tag{5}$$

where σ is the kernel parameter.

The optimal bias b^* can be computed as

$$b^* = \frac{1}{n_1 + n_2}\left((n_1 - n_2) - \sum_{j=1}^{n_1+n_2}\sum_{i=1}^{l} y_i\alpha_i^* K(x_i, x_j)\right),$$

where α^* is the optimal solution of the dual problem (3); n_1 is the number of x_i with $C_1(1-\beta) < \alpha_i^* < C_1$ and $y_i = +1$; n_2 is the number of x_i with $C_1(1-\beta) < \alpha_i^* < C_1$ and $y_i = -1$. If $n_1 = n_2 = 0$, then $b^* = 0$. The final decision function of (2) can be given by

$$f(x) = \text{sign}\left(\sum_{i=1}^{l} y_i\alpha_i^* k(x_i, x) + b^*\right).$$

More details can be seen in [29].

2.2. WCS-SVM

It is well known that the geometric information of instances is a very important priori knowledge for a classifier. By incorporating minimum within-class scatter in Fisher Discriminant Analysis