Knowledge-Based Systems 109 (2016) 17–24

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A generalized Mitchell-Dem'yanov-Malozemov algorithm for one-class support vector machine

Ming Zeng[a,b], Yu Yang[a,b,*], Junsheng Cheng[a,b]

[a] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China
[b] College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, PR China

ARTICLE INFO

Article history:
Received 23 November 2015
Revised 10 June 2016
Accepted 13 June 2016
Available online 15 June 2016

Keywords:
One-class support vector machine
Generalized Mitchell-Dem'yanov-Malozemov algorithm
Minimum norm problem
Geometric algorithm

ABSTRACT

The dual problem of one-class support vector machine (OCSVM) can be interpreted as a minimum norm problem associated with the reduced convex hull. Based on this geometric interpretation, a generalized Mitchell-Dem'yanov-Malozemov (GMDM) algorithm is proposed for OCSVM. The GMDM algorithm finds the minimum norm point in the reduced convex hull of training samples and employs such a point to construct the separating hyper-plane. Numerical experiments are conducted to compare the proposed geometric algorithm with some existing algorithms such as two modified sequential minimal optimization algorithms and the generalized Gilbert algorithm. The experimental results show that the GMDM algorithm exhibits better performance in terms of computational efficiency while achieving comparable classification accuracies to other algorithms.

1. Introduction

Over recent years, support vector machines (SVM) [1] have drawn a lot of attention. Some fruitful methods have also been developed based on the standard SVM [2-5]. Originating from SVM, one-class support vector machine (OCSVM) was firstly proposed by Schölkopf et al. [6] to estimate the boundary of a high-dimensional distribution. The basic idea of OCSVM is to find the hyper-plane that separates training samples from the origin with the maximum margin. The decision boundary in the input feature, corresponding to the separating hyper-plane in the feature space, is regarded as the description of the sample distribution. OCSVM naturally inherits the advantages of SVM, e.g., the utilization of the kernel trick, the robustness to outliers (achieved by adjusting a regularization parameter ν) and the sparseness of the solution. Over recent years, OCSVM has been successfully used in various one-class classification problems, such as computer image [7,8], network security [9,10] and process monitoring [11].

By using the Lagrangian function, the optimization problem in OCSVM that really needs to be solved is not the primal maximum margin problem but its Lagrangian dual problem [6]. The dual problem is formulated as a quadratic programming (QP) problem, which can be solved by using standard algebraic QP algorithms. When dealing with large-scale datasets, however, OCSVM with a standard algorithm will encounter the same computational problem as SVM. The reason is that the standard algorithms generally have the cubic time complexity and at least the quadratic space complexity [12,13]. To reduce computational burden, Schölkopf et al. [6] introduced the well-known sequential minimal optimization (SMO) algorithm [12] to solve the dual problem in OCSVM. The SMO algorithm is a fast optimization algorithm initially proposed for SVM. On the one hand, the SMO algorithm breaks the original QP problem into a series of smallest possible QP sub-problems, and each one can be solved analytically in an efficient way since only two variables are involved in the sub-problem [12]. On the other hand, the SMO algorithm adopts a clever caching scheme, with only the linear space complexity [12]. By using these schemes, the SMO algorithm can often reduce the computational complexity of the original QP problem. Becuase of its higher computational efficiency relative to the standard algorithms, the SMO algorithm has become one of the most popular algebraic algorithms for OCSVM. Nevertheless, Keerthi et al. [14] pointed out an important source of inefficiency caused by the way that the SMO algorithm maintains and updates a single threshold value. Using clues from the Karush-Kuhn-Tucker (KKT) conditions of the dual problem, two threshold parameters are employed to derive two modified SMO algorithms that perform more efficiently than the original one [14]. The two modified algorithms are referred to as SMO-M1 and SMO-M2, respectively.

From a geometric point of view, the dual problem in the soft margin OCSVM can be viewed as the nearest point problem (NPP) whose goal is to find the nearest point to the origin from the

* Corresponding author.
E-mail addresses: zeng_m@126.com (M. Zeng), yangyu@hnu.edu.cn (Y. Yang), signalp@tom.com (J. Cheng).

http://dx.doi.org/10.1016/j.knosys.2016.06.015
0950-7051/

reduced convex hull (RCH) of training samples [15]. As the distance from the nearest point to the origin is exactly the norm of the nearest point (essentially a vector), the NPP is actually equivalent to the minimum norm problem (MNP) whose goal is to find the point with the minimum norm in the RCH of training samples. In other words, the dual problem in the soft margin OCSVM can be interpreted as a Minimum Norm Problem associated with the Reduced Convex Hull (MNP-RCH). There are two classical algorithms to solve an MNP, namely the Gilbert algorithm [16] and the Mitchell-Dem'yanov-Malozemov (MDM) algorithm [17]. Both of them are iterative algorithms that approach the optimal minimum norm point iteratively. At each iteration, the Gilbert algorithm selects one sample point to update the current minimum norm point in the convex hull (CH) of training samples, while the MDM algorithm selects two. Therefore, it can be said that both of the two classical MNP algorithms are initially proposed to solve a Minimum Norm Problem associated with the Convex Hull (MNP-CH). The sample selection strategy either in the Gilbert algorithm or in the MDM algorithm guarantees that the new minimum norm point at each iteration is always a feasible solution to the MNP-CH. As for an MNP-RCH, however, such a guarantee cannot be given if the original sample selection strategy is still used. Consequently, the Gilbert algorithm and the MDM algorithm are not directly suitable for the soft margin OCSVM. One key step of the Gilbert algorithm is to locate the point (actually an extreme point of the CH of training samples) with the minimum projection onto the direction of the current minimum norm point [15]. In an MNP-RCH, however, such a point is not simply a single sample point anymore, which leads to the difficulty of directly applying the Gilbert algorithm to the MNP-RCH. Recently, Zeng et al. [15] developed a generalized version of the Gilbert algorithm. Locating the minimum projection point using a linear combination of several specific sample points, the generalized Gilbert (GG) algorithm can directly handle the soft margin OCSVM. The numerical experiments suggest its potential superiority over the SMO algorithm in terms of computational efficiency.

In this paper, we focus on the generalized version of the MDM algorithm. We analyze the original update strategy in the MDM algorithm, and then develop a new one to ensure that the new minimum norm point is always a feasible solution to an MNP-RCH. By incorporating this new update strategy into the MDM algorithm, a generalized Mitchell-Dem'yanov-Malozemov (GMDM) algorithm is proposed. The GMDM algorithm is essentially developed to solve an MNP-RCH. Thus, it can be directly applied to the soft margin OCSVM. Throughout the paper, we describe the GMDM algorithm in the form of its linear version, but all the comments made on the linear version are true for its corresponding kernel version. Common choices for the kernel function are Gaussian, polynomial and sigmoid kernels, etc. In this work, only the Gaussian kernel is considered. The Gaussian kernel is given in the form $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j)\rangle = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/2\sigma^2)$, where σ is the Gaussian width parameter and $\Phi(\ \cdot\ )$ is the mapping function that maps samples from the input space X (generally the Euclidean space) to the feature space F (possibly a very high-dimensional space).

The remaining part of the paper is organized as follows. Section 2 reviews the concept of an MNP-RCH. The GMDM algorithm is proposed in Section 3. In this section, the classical MDM algorithm is also introduced to promote a better understanding of the proposed algorithm. Section 4 gives numerical experiments and Section 5 concludes the paper.

2. MNP-RCH

Assume $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{l}$ are training samples, where l is the number of training samples. Let $\boldsymbol{I}$ denote the index set $\{1, 2, \cdots, l\}$. The RCH of the sample set $\boldsymbol{X}$ can be written as

$$\mathrm{RCH}(\boldsymbol{X}, \mu) = \left\{ \sum_{i=1}^{l} \alpha_i \boldsymbol{x}_i \,\middle|\, \sum_{i=1}^{l} \alpha_i = 1, 0 \le \alpha_i \le \mu \right\} \tag{1}$$

where α_i $(i \in \boldsymbol{I})$ are the combination coefficients and $\mu \in [1/l, 1)$ is the reduction factor. Any point in the RCH can be definitely written in the above form. In the special case $\mu = 1$, the formula in Eq. (1) turns into the CH of the sample set $\boldsymbol{X}$.

The basic goal of an MNP-RCH is to find the point with the minimum norm in the RCH of training samples. According to Eq. (1), an MNP-RCH can be formulated as follows

$$\begin{aligned} &\min_{\alpha \in \mathbb{R}^l} \left\| \sum_{i=1}^{l} \alpha_i \boldsymbol{x}_i \right\| \\ &\text{s.t.} \sum_{i=1}^{l} \alpha_i = 1, 0 \le \alpha_i \le \mu, i = 1, 2, \cdots, l \end{aligned} \tag{2}$$

Square the objective function and multiply it by a constant 1/2, then generate an equivalent form of the MNP-RCH, i.e.,

$$\begin{aligned} &\min_{\alpha \in \mathbb{R}^l} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\ &\text{s.t.} \sum_{i=1}^{l} \alpha_i = 1, 0 \le \alpha_i \le \mu, i = 1, 2, \cdots, l \end{aligned} \tag{3}$$

When the case $\mu = 1/(\nu l)$ holds, the optimization problem in Eq. (3) is exactly the dual problem in the soft margin OCSVM. In other words, the dual problem can be actually interpreted as an MNP-RCH. The separating hyper-plane in the MNP-RCH can be constructed by only using the optimal minimum norm point. More details refer to [15]. Furthermore, it can be proved that this separating hyper-plane is theoretically the same as the one that is constructed in the primal soft margin OCSVM [15].

3. GMDM algorithm

3.1. Review of the MDM algorithm

Let $\boldsymbol{w}$ denote the minimum norm point in the CH of training samples until current iteration. The current minimum norm point $\boldsymbol{w}$ is always known through the coefficients α_i $(i \in \mathbf{I})$, i.e., $\boldsymbol{w} = \sum_{i \in \boldsymbol{I}} \alpha_i \boldsymbol{x}_i$. At each iteration, the classical MDM algorithm selects two sample points to update $\boldsymbol{w}$ [17]. One is the sample point $\boldsymbol{x}_A$ that has the minimum projection onto the direction of $\boldsymbol{w}$, and the other is the sample point $\boldsymbol{x}_B$ $(\alpha_B > 0)$ that has the maximum projection onto the same direction. Define $\bar{\boldsymbol{w}} = \boldsymbol{w} + \alpha_B(\boldsymbol{x}_A - \boldsymbol{x}_B)$. The new minimum norm point $\boldsymbol{w}^{\text{new}}$ is then chosen to be the point with the minimum norm on the line segment joining $\boldsymbol{w}$ and $\bar{\boldsymbol{w}}$, i.e., $\boldsymbol{w}^{\text{new}} = (1-q)\boldsymbol{w} + q\bar{\boldsymbol{w}}$ where $q = \min(1, \langle \boldsymbol{w}, \boldsymbol{w} - \bar{\boldsymbol{w}}\rangle / \|\boldsymbol{w} - \bar{\boldsymbol{w}}\|^2)$. The basic principle of the MDM algorithm is illustrated in Fig. 1.

One key step of the MDM algorithm is to find the minimum projection point $\boldsymbol{x}_A$ and the maximum projection point $\boldsymbol{x}_B$ $(\alpha_B > 0)$. This goal is achieved through scanning over inner products between all the sample points $\boldsymbol{x}_i$ $(i \in \boldsymbol{I})$ and the current minimum norm point $\boldsymbol{w}$. Clearly, either of $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ could always be an extreme point of the CH of training samples. The case $\langle \boldsymbol{w}, \boldsymbol{x}_A \rangle - \langle \boldsymbol{w}, \boldsymbol{x}_B \rangle = 0$ holds if and only if $\boldsymbol{w}$ reaches the theoretically optimal solution [18]. Note that the minimum projection point is also required in the Gilbert algorithm [15]. Besides such a point, the MDM algorithm needs to find an extra maximum projection point that contributes to the representation of the current minimum norm point.