Knowledge-Based Systems 000 (2017) 1–14

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Angle-based twin parametric-margin support vector machine for pattern classification

Reshma Rastogi (née Khemchandani)[a,*], Pooja Saigal[a], Suresh Chandra[b]

[a] Department of Computer Science, Faculty of Mathematics and Computer Science, South Asian University, Delhi, India
[b] Ex Faculty, Department of Mathematics, Indian Institute of Technology, Delhi, India

ARTICLE INFO

Article history:
Received 3 June 2016
Revised 5 October 2017
Accepted 6 October 2017
Available online xxx

Keywords:
Twin parametric-margin support vector machines
Nonparallel hyperplanes classifiers
Single optimization problem
Multi-category classification
Image segmentation

ABSTRACT

In this paper, a novel angle-based twin parametric-margin support vector machine (ATP-SVM) is proposed, which can efficiently handle heteroscedastic noise. Taking motivation from twin parametric-margin support vector machine (TPMSVM), ATP-SVM determines two nonparallel parametric-margin hyperplanes, such that the angle between their normal is maximized. Unlike TPMSVM, it solves only one modified quadratic programming problem (QPP) with fewer number of representative samples. Further, it avoids the explicit computation of inverse of matrices in the dual and has efficient learning time as compared to other single problem classifiers like nonparallel SVM based on one optimization problem (NSVMOOP).

The efficacy of ATP-SVM is tested by conducting experiments on a wide range of benchmark UCI datasets. ATP-SVM is extended for multi-category classification using state-of-the-art one-against-all (OAA) and binary tree (BT) based multi-category classification approaches. This work also proposes the application of ATP-SVM for segmentation of color images.

1. Introduction

Support vector machine (SVM) has proven to be an effective classification [1,2] and regression tool [3] in the field of machine learning. SVM is based on statistical learning theory and its formulation follows structural risk minimization (SRM) principle [2,4]. The optimization problem of SVM involves the minimization of a convex quadratic function subject to linear inequality constraints. Numerous variations of SVM have been proposed, such as proximal support vector machine (PSVM) [5] and parametric-ν-SVM (par-ν-SVM) [6]. Contrary to SVM, Mangasarian and Wild proposed generalized eigenvalue proximal SVM (GEPSVM) [7] which is a nonparallel hyperplanes classifier (NHC) and generates two hyperplanes instead of one. Twin support vector machine (TWSVM) [8–10] is another binary classifier that generates two nonparallel hyperplanes such that each plane is closer to its own class and is as far as possible from the other class. TWSVM is approximately four times faster than SVM and achieves comparable classification accuracy when compared with SVM [9]. In the last decade, many variations of TWSVM have been proposed for example twin bounded SVM (TBSVM) [11], improved TWSVM (ITWSVM) [12], structural TWSVM (S-TWSVM) [13] etc. Recently, Peng proposed twin parametric-margin SVM (TPMSVM) [14] which can effectively deal with data which has heteroscedastic error structure.

The NHCs solve the two optimization problems independently in the training phase and then their solutions are used collectively to predict the labels in the testing phase. The predicted label of a test pattern depends on its distance from the two hyperplanes, whereas these two distances do not appear simultaneously in any of the two optimization problems. Hence, the training and testing phases are not consistent. To deal with this condition, Shao et al. proposed Nonparallel hyperplane SVM (NHSVM) [15] which determines two nonparallel proximal hyperplanes simultaneously i.e. by solving only one optimization problem. NHSVM is considered to be logically consistent in its predicting and training processes and has improved classification accuracy. Similar to NHSVM, Tian and Ju proposed Nonparallel SVM based on one optimization problem (NSVMOOP) [16] which aims at separating the two classes with the largest possible angle between their decision hyperplanes and also implements the SRM principle. However, NSVMOOP formulation considers the distance of all the training points from both the hyperplanes simultaneously and merges the two optimization problems of TWSVM into a single problem. This results in a QPP which is twice the size of an SVM problem and the learning time of NSVMOOP is more than that of SVM.

In this paper, a novel NHC termed as Angle-based twin parametric-margin support vector machine (ATP-SVM) with single

* Corresponding author.
E-mail addresses: reshma.khemchandani@sau.ac.in (R. Rastogi (née Khemchandani)), pooja.saigal@students.sau.ac.in (P. Saigal), chandras@maths.iitd.ac.in (S. Chandra).

https://doi.org/10.1016/j.knosys.2017.10.008
0950-7051/© 2017 Published by Elsevier B.V.

optimization problem is proposed, which is motivated by TPMSVM and is formulated on the lines of NSVMOOP. The motivation behind the proposed work is to develop a classifier with one optimization problem, similar to NSVMOOP, but with learning time-complexity comparable to TWSVM-based classifiers. The proposed classifier ATP-SVM is able to achieve good classification accuracy and its learning time is less than that of SVM and NSVMOOP. The geometric interpretation of ATP-SVM is different from that of NSVMOOP and TPMSVM.

Most of the NHCs assume that the noise is uniform in the training data or its functional dependency is known beforehand, however this assumption does not always hold true and could lead to poor results. Also, training and testing phases are not rational due to inconsistency in problem formulation and decision rules. The proposed binary classifier ATP-SVM can overcome both the above mentioned limitations. ATP-SVM combines the merits of TPMSVM and NSVMOOP and hence the resulting classifier is efficient in handling the data with unknown noise and generates consistent results. Therefore, the idea of ATP-SVM is to solve a single optimization problem to generate the two parametric-margin hyperplanes, which bound the data so that the respective class patterns lie on either side of corresponding hyperplanes. Since, ATP-SVM can efficiently handle heteroscedastic noise, it is proved to be more robust than NSVMOOP. ATP-SVM efficiently incorporates the strengths of TWSVM (time-efficient and generalization ability) and NSVMOOP (consistent testing and training). In order to increase the separation between the two classes, the angle between the normal vectors to the hyperplanes is maximized. Further, ATP-SVM avoids solving inverse of matrices in the dual which otherwise is a computationally expensive task.

In our proposed work, a training data selection procedure is introduced which identifies the ‘representative patterns’ from the two classes to further improve the training speed of the proposed classifier. Since, ATP-SVM generates parametric-margin hyperplanes which lie on the boundary of the classes, therefore, the data points that lie on or near the periphery of a class have a prominent role in determining the hyperplanes. The selection procedure identifies the representative patterns and train the classifier with these points only. Since, the number of constraints are reduced in the QPP of ATP-SVM, it results in faster learning of the classifier. The proposed classifier is proved to be more robust with good generalization ability and its efficacy is established by conducting numerical experiments on large number of benchmark UCI datasets.

We also propose extension of ATP-SVM and two NHCs with single optimization problem i.e. NSVM [15] and NSVMOOP [16], in multi-category environment using one-against-all (OAA) [17] and tree-based approach [18] like binary tree (BT). This work includes application of ATP-SVM for color image segmentation into two or more regions. When extended to multi-category scenario, ATP-SVM can be used to identify multiple non-overlapping regions in the image. In our work, we have used Berkley segmentation dataset (BSD) [19].

The remaining paper is organized as follows: Section 2 gives an introduction of TPMSVM and NSVMOOP; and explains the notations used in the rest of the paper. Section 3 introduces “Angle-based twin parametric-margin support vector machine” and is followed by its extension in multi-category framework in Section 4. The numerical results on benchmark binary and multi-category UCI and image datasets are given in Section 5. The application of proposed classifier for image segmentation is also discussed in this section. The concluding remarks are given in Section 6.

2. Background

This section briefly introduces two binary classifiers: TPMSVM and NSVMOOP.

2.1. Twin parametric-margin support vector machine (TPMSVM)

Similar to TWSVM [8,9], TPMSVM is a binary classifier that determines two nonparallel parametric-margin hyperplanes by solving two related SVM-type problems [14], each of which is smaller than a conventional SVM [2] or par-ν-SVM [6] problem. The two nonparallel hyperplanes are given by

$$h_+(x) = x^T w_+ + b_+ = 0 \quad \text{and} \quad h_-(x) = x^T w_- + b_- = 0. \tag{1}$$

Let $h_+(x)$ and $h_-(x)$ be referred as positive and negative hyperplanes respectively. Let the two classes are represented by matrices $\mathcal{X}_+$ and $\mathcal{X}_-$, with labels $+1$ and -1 respectively. With m_+ and m_- patterns in $\mathcal{X}_+$ and $\mathcal{X}_-$, their dimensions are given as $(m_+ \times n)$ and $(m_- \times n)$ respectively. Each row of $\mathcal{X}_+$ or $\mathcal{X}_-$ is a vector in n-dimensional real space $\mathcal{R}^n$, that represents feature vector of a data sample. TPMSVM separates the data if and only if

$$\begin{aligned} &\mathcal{X}_i w_+ + b_+ \geq 0, \quad \text{for} \quad \mathcal{X}_i \in \mathcal{X}_+, \\ &\mathcal{X}_i w_- + b_- \leq 0, \quad \text{for} \quad \mathcal{X}_i \in \mathcal{X}_-, \end{aligned} \tag{2}$$

where $\mathcal{X}_i$ represents ith data sample. The primal formulation for the pair of QPPs in TPMSVM is given as follows:

$$\begin{aligned} \min_{w_+, b_+, \xi_+} \quad & \frac{1}{2}\|w_+\|_2^2 + \frac{c_1}{m_-} e_-^T (\mathcal{X}_- w_+ + e_- b_+) + \frac{c_2}{m_+} e_+^T \xi_+ \\ \text{subject to} \quad & \mathcal{X}_+ w_+ + e_+ b_+ \geq 0 - \xi_+, \quad \xi_+ \geq 0, \end{aligned} \tag{3}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_-} \quad & \frac{1}{2}\|w_-\|_2^2 - \frac{c_3}{m_+} e_+^T (\mathcal{X}_+ w_- + e_+ b_-) + \frac{c_4}{m_-} e_-^T \xi_- \\ \text{subject to} \quad & \mathcal{X}_- w_- + e_- b_- \leq 0 + \xi_-, \quad \xi_- \geq 0. \end{aligned} \tag{4}$$

The constants c_1, c_2, c_3, $c_4 > 0$ are trade off factors; e_+, e_- are vectors of ones, in real space, of appropriate dimensions and $\|.\|_2$ represents L_2 norm. The objective function of (3) or (4) controls the complexity of the model and minimizes the sum of projection values of training points on the hyperplane of other class, with parameter c_1. The objective function also minimizes the sum of error, which occurs due to the data points lying on wrong sides of the hyperplanes. The constraints of (3) require that the projection values of positive training samples on the positive hyperplane should be at least zero. A slack vector ξ_+ measures the amount of error due to positive training points. The constraints of (4) can be defined analogously.

2.2. Nonparallel support vector machine based on one optimization problem (NSVMOOP)

Tian and Ju [16] proposed a binary classifier NSVMOOP, that determines the two nonparallel proximal hyperplanes by solving a single optimization problem and aims at maximizing the angle between the normal vectors of the two hyperplanes. On the lines of ITWSVM [12], the formulation of NSVMOOP combines the two QPPs of ITWSVM together into a single QPP such that the angle between the normal to the hyperplanes is maximized. The optimization problem of NSVMOOP is as follows:

$$\begin{aligned} \min_{w_\pm, b_\pm, \eta_\pm, \xi_\pm} \quad & \frac{1}{2}(\|w_+\|_2^2 + \|w_-\|_2^2) \\ & + c_1(\eta_+^T \eta_+ + \eta_-^T \eta_- + e_+^T \xi_+ + e_-^T \xi_-) + c_2(w_+ . w_-), \\ \text{subject to} \quad & \mathcal{X}_+ w_+ + e_+ b_+ = \eta_+, \\ & \mathcal{X}_- w_- + e_- b_- = \eta_-, \end{aligned}$$

Please cite this article as: R. Rastogi (née Khemchandani) et al., Angle-based twin parametric-margin support vector machine for pattern classification, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.008