

Research Article

A Novel Support Vector Machine with Globality-Locality Preserving

Cheng-Long Ma and Yu-Bo Yuan

Institute of Metrology and Computational Science, China Jiliang University, Hangzhou, Zhejiang 310018, China

Correspondence should be addressed to Yu-Bo Yuan; ybyuan@ecust.edu.cn

Received 25 April 2014; Accepted 27 May 2014; Published 17 June 2014

Academic Editor: Shan Zhao

Copyright © 2014 C.-L. Ma and Y.-B. Yuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Support vector machine (SVM) is regarded as a powerful method for pattern classification. However, the solution of the primal optimal model of SVM is susceptible for class distribution and may result in a nonrobust solution. In order to overcome this shortcoming, an improved model, support vector machine with globality-locality preserving (GLPSVM), is proposed. It introduces globality-locality preserving into the standard SVM, which can preserve the manifold structure of the data space. We complete rich experiments on the UCI machine learning data sets. The results validate the effectiveness of the proposed model, especially on the Wine and Iris databases; the recognition rate is above 97% and outperforms all the algorithms that were developed from SVM.

1. Introduction

In the past decades, support vector machine (SVM) [1] was thought to be a powerful tool for classification tasks. It can separate different classes by hyperplanes, which are determined by optimal directions and support vectors, while the optimal directions are obtained by maximizing the margins between each two classes. SVM and its variants [2–5] have been successfully applied to many research areas such as face detection and recognition [6], speech recognition [7], text classification [8], and image retrieval [9].

It is well known that SVM is an optimization problem. The optimal solution can be found by solving a quadratic programming problem. Because the objective function is convex, the global minimum solution is guaranteed. However, the traditional SVM solution is susceptible to class distribution, which means it is nonrobust to data samples. So, in order to overcome this shortcoming, Zafeiriou et al. [10] proposed a minimum class variance support vector machine (MCVSVM), which was inspired by optimization Fisher's discriminant ratio. By taking the manifold structure of the data space into consideration, Wang et al. [11] introduced within-class locality preserving into SVM and proposed the minimum class locality preserving variance support vector machine (MCLPV-SVM). Besides, since SVM deals with a

subset of data points (support vectors) rather than the entire data set, so to some extent, SVM solution is based on "local" characteristics of the data; therefore, Xiong and Cherkassky [12] incorporated global discriminant information into SVM and proposed the SVM+LDA. Analogously, Khan et al. [13] presented a novel SVM+NDA (nonparametric discriminant analysis) model for classification; it fused some partially global information and local information. In particular, both SVM+LDA and SVM+NDA can cope with the small sample problem, which benefited from the construction method of the models.

According to the above analysis, none of the mentioned methods takes the manifold structure of the data space into consideration, except the MCLPV-SVM method, but MCLPV-SVM loses some discriminant information. In the basic learning algorithms, the locality of learning data set should be considered; in recent years, many excellent publications showed the importance of locality, such as [14–16]. It is also important for finding the clusters in high dimensional data set, such as [17, 18]. Recently, discriminant locality preserving projections (DLPP) [19–21] is proposed, and it is keen to find the subspace which can best discriminate different classes by maximizing the locality preserving between-class distances while minimizing the locality preserving within-class distances. So, DLPP can not only preserve

local structure, but also implicate discriminant information. Inspired by DLPP, and considering that the mean sample can reflect the characteristics of data structure, which is also center-invariant in each class [22], this paper proposed a novel learning algorithm, called support vector machine with globality-locality preserving (GLPSVM). It introduced globality and locality preserving ability into the SVM. The proposed method preserves intrinsic manifold structure of the data space, takes the class distribution into consideration, and obtains a robust solution.

In summary, this paper is organized as follows. Section 2 gives a brief review of SVM. Section 3 gives the proposed method, including derivation, solving, and analysis. The experimental results are given in Section 4. Finally, conclusions are in Section 5.

2. A Brief Review of SVM

Given a set of pairwise samples $\mathbf{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathcal{R}^m$ is a sample point in m -dimensional space and $y_i \in \{+1, -1\}$ is the corresponding label. The direct way to separate these samples into two classes is to find a separating hyperplane.

For the linearly separable case, the SVM model is as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, n. \end{aligned} \quad (1)$$

By transforming this optimization problem into its corresponding dual problem, the optimal discriminant vectors can be found through

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad (2)$$

where α_i and \mathbf{x}_i are the dual variable and data sample (called support vector), respectively. The support vectors are crucial for classification since removing these points may change the solution of SVM. In SVM, the separable directions are decided by the optimal discriminant vectors obtained through (2). So, if we project the data into a feature space spanning by the optimal discriminant vectors, then these data will be separable in the feature space.

Usually, in real world applications, we need to deal with the multiclassification cases, such as face recognition [6] and text categorization [8]. In such cases, we need to extend SVM to multiclass SVM. The general approach is to code the classes according to a certain strategy, like one-against-all (OAA) or one-against-one (OAO) [23]. The OAA coding approach compares data in a single class with all the samples in others classes to generate the decision boundary; in this case, c -many decision boundaries are built for c -many classes. The OAO strategy generates decision boundaries from all possible pair of classes, which obtains $c(c-1)/2$ -many decision boundaries. Comparatively, the OAO can obtain more discriminant vectors than OAA but will cost more computational time.

3. The Proposed GLPSVM

In this section, we will propose a novel support vector classifier which takes the class distribution into consideration, and a robust solution is expected. Firstly, we will introduce the definition of globality-locality preserving.

3.1. Globality-Locality Preserving (GLP). Discriminant locality preserving projections (DLPP) [19–21] is a powerful method for extracting the manifold structure of data samples. Given a set of samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathcal{R}^{m \times n}$, where \mathbf{x}_i is a sample point in m -dimensional space, and $Y = \{y_i\}_{i=1}^n \in \mathcal{R}^{1 \times n}$ is the corresponding labels. Let $y_i \in \{+1, -1\}$ and $C = \{C_+, C_-\}$, and then all $y_i = +1$ label samples belonging to C_+ and the others belonging to C_- . Thus, we have $n = n_1 + n_2$, where n_1 is the number of samples in C_+ and n_2 is the number of samples in C_- . Suppose that $\mathbf{T} = \{\mathbf{t}_i\}_{i=1}^n \in \mathcal{R}^{d \times n}$ is the low-dimensional feature projections of \mathbf{X} and DLPP tries to maximize an objective function as follows:

$$J = \frac{\sum_{i,j=1}^2 (\mathbf{m}_i - \mathbf{m}_j) B_{ij} (\mathbf{m}_i - \mathbf{m}_j)^T}{\sum_{s=1}^2 \sum_{i,j=1}^{n_s} (\mathbf{t}_i^s - \mathbf{t}_j^s) W_{ij}^s (\mathbf{t}_i^s - \mathbf{t}_j^s)^T}, \quad (3)$$

where \mathbf{m}_i and \mathbf{m}_j represent the mean vectors of the projected samples in the i th and j th class, respectively. W_{ij}^s and B_{ij} are elements of the within-class weight matrix \mathbf{W} and the between-class weight matrix \mathbf{B} defined as

$$\begin{aligned} W_{ij}^s &= \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_1}\right), & \mathbf{x}_i \in \mathcal{N}_{k_1}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_{k_1}(\mathbf{x}_i) \\ 0 & \text{otherwise,} \end{cases} \\ B_{ij} &= \begin{cases} \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{\sigma_2}\right), & \mathbf{u}_i \in \mathcal{N}_{k_2}(\mathbf{u}_j) \text{ or } \mathbf{u}_j \in \mathcal{N}_{k_2}(\mathbf{u}_i) \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (4)$$

where $\mathcal{N}_{k_1}(\mathbf{x})$, $\mathcal{N}_{k_2}(\mathbf{u})$ denote local neighbors of \mathbf{x} and \mathbf{u} , respectively, k_1 is the sample neighborhood, and k_2 is the mean sample neighborhood. The parameters σ_1 and σ_2 are empirically determined, and \mathbf{u}_i is the mean vector of samples in the i th class. Suppose $\mathbf{w} : \mathbf{X} \rightarrow \mathbf{T}$ is a mapping from high-dimensional data space to low-dimensional feature space; that is, $\mathbf{T} = \mathbf{w}^T \mathbf{X}$. Then, the objection function (3) can be rewritten as follows:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{U} \mathbf{H} \mathbf{U}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w}}, \quad (5)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$, \mathbf{H} and \mathbf{L} are the Laplacian matrices [24, 25], $\mathbf{H} = \mathbf{D}_b - \mathbf{B}$, $\mathbf{B} = [B_{ij}]_{i,j=1}^2$, \mathbf{D}_b is a diagonal matrix and its elements are column (or row) sums of \mathbf{B} , $\mathbf{L} = \mathbf{D}_w - \mathbf{W}$,

\mathbf{W} is a block diagonal matrix, that is, $\mathbf{W} = \text{diag} \{ [W_{ij}^s]_{i,j=1}^{n_s} \}_{s=1}^2$, and \mathbf{D}_w is also a block diagonal matrix; each block of \mathbf{D}_w is a diagonal matrix and its elements are the column (or row) sums of each block of \mathbf{W} .

Formula (5) is also called locality preserving discriminant ratio (criterion); that is, DLPP is keen to find the feature subspace via maximizing this ratio, which means simultaneously maximizing the locality preserving between-class distance and minimizing the locality preserving within-class distance.

On the other hand, Huang et al. [22] had another view of the weight matrix of DLPP, and they believe that the mean sample is center-invariant in the same class, and it can reflect the characteristics of data structure, so to some extent it decides the accuracy in classification tasks. In this paper, we preserve the structure information of mean samples, which can to a large extent make up the loss of global information when only local structure is preserved. In a word, we define locality preserving matrix and globality preserving matrix as follows:

(i) locality preserving matrix: $\mathbf{Z}_w = \mathbf{X}\mathbf{L}\mathbf{X}^T$,

(ii) globality preserving matrix: $\mathbf{Z}_b = \mathbf{U}\mathbf{H}\mathbf{U}^T$.

3.2. Derivation of GLPSVM. Now, we give the proposed extension of SVM, called GLPSVM. For the linearly separable data, GLPSVM can be described as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T (\lambda \mathbf{Z}_w + \mathbf{Z}_b + \beta \mathbf{I}) \mathbf{w} \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, n, \end{aligned} \tag{6}$$

where $\beta \mathbf{I}$ represents the regularization matrix, which is added to cope with small sample problems. This model not only maximizes the margin of the separating hyperplane, but also minimizes the scatter of the data in discriminant directions, which benefited from taking both the locally manifold structure of the data space and globality manifold structure of the mean sample space into consideration. Here, λ is an empirically determined key parameter which controls the tradeoff.

According to the model, we can see that the optimal discriminant directions are no longer the same as classical SVM. It is because that we introduce the obtained GLP to the optimization model of SVM. The classification performance of the proposed method will be shown in Section 4.

3.3. Solution to the GLPSVM. Similar to SVM, the proposed model can be viewed as a quadratic optimization problem. Lagrange's method of undetermined multipliers is used to solve this problem. Suppose α_i ($i = 1, 2, \dots, n$) is positive Lagrange multipliers, and let $\Delta \triangleq \lambda \mathbf{Z}_w + \mathbf{Z}_b + \beta \mathbf{I}$, and then the corresponding Lagrangian is

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \Delta \mathbf{w} - \sum_{i=1}^n \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1]. \tag{7}$$

Taking derivatives with respect to \mathbf{w} and b , respectively, we obtain

$$\begin{aligned} \mathbf{w} &= \Delta^{-1} \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i, \\ \sum_{i=1}^n y_i \alpha_i &= 0. \end{aligned} \tag{8}$$

Hence, we have the following dual problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \Delta^{-1} \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^n y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, n. \end{aligned} \tag{9}$$

Suppose α^* is the optimization solution of this dual problem. Then, the optimal discriminant vectors \mathbf{w}^* can be found as

$$\mathbf{w}^* = \sum_{i=1}^n y_i \alpha_i^* \Delta^{-1} \mathbf{x}_i. \tag{10}$$

So, the corresponding decision surface is

$$\begin{aligned} g(\mathbf{x}) &= \text{sign}(\mathbf{w}^{*T} \mathbf{x} + b) \\ &= \text{sign} \left(\sum_{i=1}^n y_i \alpha_i^* \mathbf{x}_i^T \Delta^{-1} \mathbf{x} + b^* \right). \end{aligned} \tag{11}$$

Finally, the corresponding optimal bias b^* can be calculated as

$$b^* = \frac{1}{n_{sv}} \sum_{i=1}^{n_{sv}} \left(y_i - \sum_{j=1}^n y_j \alpha_j^* \mathbf{x}_j^T \Delta^{-1} \mathbf{x}_i \right), \tag{12}$$

where n_{sv} is the number of support vectors.

As can be seen, in linearly separable case, GLPSVM is required to obtain a completely accurate decision hyperplane. However, in real world applications, the decision hyperplane no longer needs to be completely accurate, so we extend the GLPSVM to soft margin situations.

3.4. Soft Margin GLPSVM. Reference [13] proposed the soft margin method for SVM, to cope with cases when we do not need to obtain a completely accurate decision hyperplane; that is, we permit an error tolerability within limits. Then, the soft margin GLPSVM can be described as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T (\lambda \mathbf{Z}_w + \mathbf{Z}_b + \beta \mathbf{I}) \mathbf{w} + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n, \end{aligned} \tag{13}$$

where C is a predefined positive real number and larger values of C correspond to higher penalty assigned to errors. $\xi = [\xi_1, \xi_2, \dots, \xi_n]$ is slack variables which reflect the degree of misclassification. Apparently, the soft margin GLPSVM is also a quadratic optimization problem, and we can solve it in the same way as standard GLPSVM.

TABLE 1: The effects of parameter C and sample neighborhood k_1 on classification performance.

k_1	10	20	30	40	50	60	70	80
$C = 0.1$	0.8207	0.8048	0.8008	0.8008	0.7888	0.7769	0.7888	0.7809
$\sigma =$	2	1	2	0.5	0.5	2	1.5	2
$C = 1$	0.8884	0.8884	0.8964	0.8805	0.8765	0.8008	0.8207	0.8048
$\sigma =$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$C = 10$	0.9323	0.9482	0.9044	0.9203	0.9323	0.9203	0.9084	0.8884
$\sigma =$	2	0.5	2.5	0.5	2.5	1	0.5	0.5
$C = 100$	0.9402	0.9283	0.9402	0.9442	0.9402	0.9323	0.9482	0.9363
$\sigma =$	1.5	2	1.5	1.5	2.5	0.5	1	0.5

TABLE 2: The effects of trade-off parameter λ on classification performance; let $\sigma_1 = \sigma_2 = 2$.

λ	0.2	0.4	0.8	1	1.2	1.4	1.6	2
$C = 0.1$	0.8048	0.8048	0.7849	0.8008	0.8088	0.7928	0.7928	0.8008
k_1	40	30	30	10	50	20	90	10
$C = 1$	0.9441	0.7849	0.8167	0.8127	0.7968	0.7968	0.8008	0.8008
k_1	10	10	20	20	10	30	10	10
$C = 10$	0.9203	0.8964	0.9004	0.9203	0.8845	0.8964	0.8884	0.8645
k_1	30	30	10	20	10	10	10	10
$C = 100$	0.9283	0.9442	0.9243	0.9482	0.9203	0.9323	0.9283	0.9163
k_1	50	20	20	20	30	20	20	10

TABLE 3: Database information for comparative analysis.

Database	Number of sample	Attribute	Number of class
Breast	699	9	2
Heart	270	13	2
Pima	768	8	2
New Thyroid (NT)	215	5	3
Wine	178	13	3
Iris	150	4	3
Glass	214	9	6

3.5. *Effective Algorithm for GLPSVM.* Note that the objection function of classical SVM is $(1/2)\mathbf{w}^T\mathbf{w}$; however, in this paper, we use $(1/2)\mathbf{w}^T\Delta\mathbf{w}$ for replacement. In fact, if we define some relational expressions as follows:

$$\begin{aligned} \mathbf{w}^r &= \Delta^{1/2}\mathbf{w}, \\ \mathbf{x}_i^r &= \Delta^{-1/2}\mathbf{x}_i, \quad \forall i = 1, 2, \dots, n, \end{aligned} \tag{14}$$

and then the GLPSVM model is equivalent to

$$\begin{aligned} \min_{\mathbf{w}^r} \quad & \frac{1}{2}\|\mathbf{w}^r\|^2 \\ \text{s.t} \quad & y_i(\mathbf{w}^{rT}\mathbf{x}_i^r + b) \geq 1, \quad \forall i = 1, \dots, n. \end{aligned} \tag{15}$$

Then, we can find that the solution to GLPSVM can be solved by standard SVM software package, but the optimal discriminant vectors are different. Since we introduce globality-locality preserving to SVM, the optimal discriminant directions in GLPSVM can preserve the intrinsic

manifold structure of the data in low-dimensional feature space. Besides, matrices $\Delta^{1/2}$ and $\Delta^{-1/2}$ can be calculated through the eigenvalue decomposition of the matrix Δ ; the interested readers can refer to literature [13] for more details.

4. Performance Evaluation

4.1. *The Parameters' Influence on the Performance.* In the proposed model, there are totally six parameters; they are the neighborhood parameters k_1, k_2 and the heat kernel parameters σ_1, σ_2 in locality preserving matrix and globality preserving matrix, respectively, and a trade-off parameter λ along with the regularization parameter C . In this section, to show the parameters' influence on the performance, we do experiments on a binary ionosphere database. We select 30% of samples of each class for training, the rest of samples are for testing, and all the samples are normalized before the experiment.

For the parameters setting, the regularization parameter C is selected from the set $\{0.1, 1, 10, 100\}$, the heat kernel parameter σ is selected from the set $\{0.5, 1, 1.5, 2, 2.5\}$, the neighborhood parameter k_1 is selected from the set $\{10, 20, 30, 40, 50, 60, 70, 80\}$, and the trade-off parameter λ is set to $\{0.2, 0.4, 0.6, 0.8, 1, 1.2, 2\}$.

Firstly, we set the trade-off parameter to be 0.2 and use the same heat kernel parameter ($\sigma_1 = \sigma_2 = \sigma$) to see the effect of the parameters C and k_1 . Table1 shows classification accuracy under different settings of these three parameters. We can find that the regularization parameter C plays an important role in classification performance. Besides, appropriate parameter selection can provide better classification results.

TABLE 4: Classification accuracy for comparative analysis.

Database	SVM	SVM + LDA	MCVSVM	MCLPV_SVM	GLPSVM
Breast C = 1	0.8613 ± 0.0103	0.9572 ± 0.0063	0.9596 ± 0.0052	0.9658 ± 0.0045 k = 40 t = 2	0.9686 ± 0.0046 k ₁ = 40, k ₂ = 1 σ ₁ = σ ₂ = 2
Heart C = 10	0.6912 ± 0.0352	0.8176 ± 0.0205	0.8235 ± 0.0186	0.8294 ± 0.0338 k = 50 t = 2	0.8471 ± 0.0243 k ₁ = 50, k ₂ = 1 σ ₁ = σ ₂ = 2
Pima C = 10	0.6723 ± 0.0296	0.7689 ± 0.0158	0.7663 ± 0.0144	0.7689 ± 0.1023 k = 80 t = 2	0.7841 ± 0.0277 k ₁ = 80, k ₂ = 1 σ ₁ = σ ₂ = 2
NT C = 100	0.6923 ± 0.0000	0.8965 ± 0.0157	0.8569 ± 0.0148	0.9496 ± 0.0144 k = 10 t = 0.5	0.9512 ± 0.0170 k ₁ = 10, k ₂ = 2 σ ₁ = σ ₂ = 0.5
Wine C = 100	0.5714 ± 0.0275	0.9490 ± 0.0228	0.8163 ± 0.0126	0.9694 ± 0.0219 k = 20 t = 2	0.9796 ± 0.0207 k ₁ = 20, k ₂ = 2 σ ₁ = σ ₂ = 2
Iris C = 1	0.9467 ± 0.0172	0.9713 ± 0.0139	0.9733 ± 0.0147	0.9333 ± 0.0140 k = 10 t = 0.5	0.9753 ± 0.0124 k ₁ = 10, k ₂ = 2 σ ₁ = σ ₂ = 0.5
Glass C = 100	0.6508 ± 0.0179	0.7996 ± 0.0132	0.7967 ± 0.0098	0.9172 ± 0.2222 k = 4 t = 0.5	0.9402 ± 0.0219 k ₁ = 4, k ₂ = 4 σ ₁ = σ ₂ = 0.5

Next, we will explore the effect of the trade-off parameter λ . Table 2 presents the classification accuracy of GLPSVM under different trade-off parameter λ and different regularization parameter C . Here, we give the highest accuracy with its corresponding sample neighborhood parameter k_1 . It can be seen that the λ also plays an important role in classification results and small λ may be more appropriate than large λ .

4.2. Comparative Analysis. In this subsection, comparative experiments are conducted to test the ability of the proposed GLPSVM. We compare it with SVM, SVM+LDA, MCVSVM, and MCLPV_SVM on seven different databases selected from the well-known UCI database. The seven databases include three binary databases and four multiclass databases. In the multiclass tasks, one-against-one coding strategy is employed. The detailed information of these selected databases is shown in Table 3. For all these databases, 30% of samples in each class are randomly selected for training, the remaining samples are used for testing, and all the data samples are normalized before experiment.

Table 4 gives the classification accuracy of SVM, SVM+LDA, MCVSVM, MCLPV_SVM, and GLPSVM under different regularization parameter, neighborhood parameter, and heat kernel parameter, where the highest average classification accuracy is presented. Here, the average classification accuracy is obtained through repeating the operation 20 times. It can be seen from Table 4 that the proposed GLPSVM has always the highest accuracy, especially on the Wine and the Iris databases. The accuracy is more than 97%, far more than the other algorithms.

5. Conclusions

In this paper, a new extension of SVM was proposed, which was called support vector machine with globality-locality preserving (GLPSVM). It took the intrinsic manifold structure of the data space into consideration. Besides, the soft margin GLPSVM was also presented. The effective algorithm of GLPSVM showed that the model could be solved through transferring it to the standard SVM model and using the standard SVM software package for solving, which would greatly improve the implementation efficiency. Finally, experimental results on real world databases validated that the proposed method could have better performance than SVM, SVM+LDA, MCVSVM, and MCLPV_SVM.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research has been supported by the National Natural Science Foundation under Grant (no. 61001200).

References

- [1] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.
- [2] Y. B. Yuan and T. Huang, "A polynomial smooth support vector machine for classification," in *Advanced Data Mining and*

- Applications*, vol. 3584 of *Lecture Notes in Artificial Intelligence*, pp. 370–377, Springer, Berlin, Germany, 2005.
- [3] Y. B. Yuan and C. Li, “A new smooth support vector machine,” in *Computational Intelligence and Security*, vol. 3801 of *Lecture Notes in Artificial Intelligence*, pp. 392–397, Springer, Berlin, Germany, 2005.
- [4] Y. B. Yuan, “Canonical duality solution for alternating support vector machine,” *Journal of Industrial and Management Optimization*, vol. 8, no. 3, pp. 611–621, 2012.
- [5] Y. B. Yuan, W. Fan, and D. Pu, “Spline function smooth support vector machine for classification,” *Journal of Industrial and Management Optimization*, vol. 3, no. 3, pp. 529–542, 2007.
- [6] Y. Li, S. Gong, J. Sherrah, and H. Liddellb, “Support vector machine based multi-view face detection and recognition,” *Image and Vision Computing*, vol. 22, no. 5, pp. 413–427, 2004.
- [7] A. Ganapathiraju, J. E. Hamaker, and J. Picone, “Applications of support vector machines to speech recognition,” *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2348–2355, 2004.
- [8] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” *The Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.
- [9] S. Tong and E. Chang, “Support vector machine active learning for image retrieval,” in *Proceedings of the 9th ACM International Conference on Multimedia*, pp. 107–118, October 2001.
- [10] S. Zafeiriou, A. Tefas, and I. Pitas, “Minimum class variance support vector machines,” *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2551–2564, 2007.
- [11] X. Wang, F. L. Chung, and S. Wang, “On minimum class locality preserving variance support vector machine,” *Pattern Recognition*, vol. 43, no. 8, pp. 2753–2762, 2010.
- [12] T. Xiong and V. Cherkassky, “A combined SVM and LDA approach for classification,” in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '05)*, pp. 1455–1459, August 2005.
- [13] N. M. Khan, R. Ksantini, I. S. Ahmad, and B. Boufama, “A novel SVM+NDA model for classification with an application to face recognition,” *Pattern Recognition*, vol. 45, no. 1, pp. 66–79, 2012.
- [14] L. Zhang, “Locally regressive projections,” *International Journal of Software and Informatics*, vol. 7, no. 3, pp. 435–451, 2013.
- [15] L. Zhang, C. Chen, J. Bu, and Z. Chen, “Locally discriminative coclustering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1025–1035, 2012.
- [16] L. Zhang, C. Chen, J. Bu, and Z. Chen, “Active learning based on locally linear reconstruction,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, no. 10, pp. 2026–2038, 2011.
- [17] Q. Q. Gu, Z. Li, and J. Han, “Linear discriminant dimensionality reduction,” in *Machine Learning and Knowledge Discovery in Databases*, vol. 6911 of *Lecture Notes in Computer Science*, pp. 549–564, 2011.
- [18] Q. Q. Gu and J. Han, “Clustered Support Vector Machines,” in *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, pp. 307–315, 2013.
- [19] W. Yu, X. Teng, and C. Liub, “Face recognition using discriminant locality preserving projections,” *Image and Vision Computing*, vol. 24, no. 3, pp. 239–248, 2006.
- [20] L. Yang, W. Gong, X. Gu, W. Li, and Y. Liang, “Null space discriminant locality preserving projections for face recognition,” *Neurocomputing*, vol. 71, no. 16–18, pp. 3644–3649, 2008.
- [21] X. Gu, W. Gong, and L. Yang, “Regularized locality preserving discriminant analysis for face recognition,” *Neurocomputing*, vol. 74, no. 17, pp. 3036–3042, 2011.
- [22] S. Huang, D. Yang, F. Yang et al., “Face recognition via globality-locality preserving projections,” *Computer Vision and Pattern Recognition*, 2013, <http://arxiv.org/abs/1311.1279>.
- [23] C. W. Hsu and C. J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [24] X. F. He and P. Niyogi, “Local preserving projections,” *Neural Information Processing Systems (NIPS)*, 2003.
- [25] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: a geometric framework for learning from labeled and unlabeled examples,” *Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.