Expert Systems With Applications 194 (2022) 116491

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Generalized multi-view learning based on generalized eigenvalues proximal support vector machines

Xijiong Xie[a,*], Yujie Xiong[b]

[a] The School of Information Science and Engineering, Ningbo University, Zhejiang 315211, PR China
[b] The School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, PR China

ARTICLE INFO

Keywords:
Multi-view learning
Generalized eigenvalue proximal support vector machines
Multi-view co-regularization
Consistency information

ABSTRACT

Multi-view learning based on generalized eigenvalue proximal support vector machines has brought enormous success by mining the consistency information of two views. Nevertheless, it only aims to handle two-view cases and cannot handle generalized multi-view learning cases (above two views). It also omits the complementarity information among views. In this paper, two generalized multi-view extensions of generalized eigenvalue proximal support vector machines are presented which take advantage of the multi-view co-regularization term to mine the consistency information and the weighted value to mine complementarity information. Experimental results performed on synthetic and real world datasets demonstrate that they can provide higher performance than the relevant two-view classification algorithms.

1. Introduction

Support vector machines (SVMs) are an outstanding supervised classification method (Shawe-Taylor & Sun, 2011) that is on account of the large margin criterion and structural risk minimization. SVMs gain a best classification hyperplane by resolving a quadratic programming problem (QPP).

Lately generalized eigenvalue proximal support vector machines (GEPSVMs) (Mangasarian & Wild, 2006), improved generalized eigenvalue proximal support vector machines (IGEPSVMs) (Shao, Deng, Chen and Wang, 2013) and twin support vector machines (TSVMs) (Jayadeva et al., 2007) as non-parallel hyperplane classifiers have been proved to be very effective. They produce a relevant hyperplane for each class which is close to one class and as far as possible from the other classes. In contrast with the conventional SVMs, GEPSVMs and IGEPSVMs gain two non-parallel hyperplanes by resolving a pair of generalized eigenvalue problems and standard eigenvalue decomposition problems, respectively. GEPSVMs were developed to the semi-supervised learning framework based on the manifold regularization term (Chen et al., 2014; Yang et al., 2009). L_1 norm and L_q norm GEPSVMs with an effective alternating algorithm (Sun et al., 2018; Yan et al., 2018) were proposed for classification. A proximal classification method based on the consistency (Shao, Deng and Chen, 2013) was presented to prevent the possible singular cases in GEPSVMs. GEPSVMs were also applied in the regression (Khemchandani et al., 2013) and clustering problems (Yang et al., 2015).

Multi-view learning (MVL) (Zhao et al., 2017) pays attention to multi-model datasets. For instance, a web page includes the text information in itself and hyperlink information. Images could be represented by numerous features such as HOG, LBP and SIFT. Prevenient MVL methods straightforwardly employ the connected views for learning. Nevertheless, these methods not only produce the curse of dimensionality problem, but also omit the statistical feature of each view. Despite that each view can perform efficiently for a given learning case, improvements could be gained by integrating the manually produced views. The success of MVL methods is usually on account of two main criterions: consistency and complementarity criterions. The consistency criterion enforces predictions of all views to be as identical as possible while the complementarity criterion takes advantage of the complementary information shared by all views.

Co-training class methods, co-regularization class methods and margin consistency class methods are three categories of existing MVL methods. Co-training class methods alternatively maximize the common agreement on two views to ensure the consistency. co-EM (Nigam & Ghani, 2000) and co-training with the robustness (Sun & Jin, 2011) are parts of the classes. Co-regularization class methods integrate multi-view co-regularization term to the objective function or multi-view constraint so as to minimize the discrepancy between the predictions of two views. Typical two-view learning methods based on SVMs and TSVMs are two-view SVMs (Farquhar et al., 2006), sparse multi-view SVMs (Sun & Shawe-Taylor, 2010) and multi-view TSVMs

* Corresponding author.
E-mail addresses: xiexijiong@nbu.edu.cn (X. Xie), xiong@sues.edu.com (Y. Xiong).

https://doi.org/10.1016/j.eswa.2021.116491
Received 12 May 2020; Received in revised form 30 December 2021; Accepted 31 December 2021
Available online 15 January 2022

(MvTSVMs) (Xie & Sun, 2015). Some researchers studied their theoretical analysis with Rademacher complexity (Bartlett & Mendelson, 2002; Sun & Shawe-Taylor, 2010). Lately, PAC-Bayes bounds (Sun et al., 2017) were used to analyze MVL based on SVMs. Margin-consistency class methods are on account of maximum entropy discrimination (MED). Sun and Chao (2013) presented a multi-view version of MED by integrating two-view classification parameters to a independent entropy, and then they (Chao & Sun, 2016a) developed another multi-view MED which gets a compromise on two related entropies for two views. Some improvements about them were proposed in Chao and Sun (2016b, 2016c) and Mao and Sun (2016).

Lately, multi-view least squares support vector machines (Houthuys et al., 2018) gained the consistency by minimizing multiplication cross terms of the errors from all views. Tang, Tian, Liu et al. (2018), Tang, Tian, Zhang et al. (2018) proposed privileged two-view SVMs which mine the complementary information of two views with privileged information, multi-view nonparallel support vector machines (Tang, Li et al., 2018) and coupling privileged kernel method for MVL (Tang et al., 2019). Our prior work of MVL mainly contains regularized multi-view least squares twin support vector machines (Xie, 2018), multi-view GEPSVMs (MGSVMs) and multi-view IGEPSVMs (MIGSVMs) (Sun et al., 2019). However, MGSVMs and MIGSVMs only obey the consistency criterion and omit the complementarity criterion. They are only constructed for two-view cases and cannot handle generalized MVL cases.

In this paper, inspired by these shortcomings, we build generalized multi-view generalized eigenvalue proximal support vector machines (GMGSVMs) and generalized multi-view improved generalized eigenvalue proximal support vector machines (GMIGSVMs) which take advantage of the multi-view co-regularization to mine the multi-view consistency information and adopt the weighted value to fully make use of the multi-view complementary information.

Our contributions are listed below:

(1) Since multi-view GEPSVMs and multi-view IGEPSVMs only handle two-view learning cases and omit the complementarity criterion, we present two novel generalized multi-view learning models called GMGSVMs and GMIGSVMs as their generalized multi-view extensions. GMGSVMs and GMIGSVMs not only own the power of handling multiple views, but also fully make use of the consistency and complementarity criterion among views.

(2) So as to efficiently resolve the optimization of our models, an alternating algorithm is constructed to gain the optimal solution.

(3) We conduct experiments on five datasets. The experimental results demonstrate the effectiveness of the proposed methods.

The following parts of this paper are as follows. Section 2.2 introduces related work about MGSVMs and MIGSVMs. Section 3 thoroughly describes our proposed GMGSVMs and GMIGSVMs. After reporting experimental results in Section 4, we provide conclusions and future work in Section 5.

2. Related work

Before reviewing the work of MGSVMs and MIGSVMs, we first give the detail introduction of Rayleigh quotient which is significant for MGSVMs.

2.1. Rayleigh quotient

For a given real symmetric matrix $S \in R^{d\times d}$ and a non-zero real vector $w \in R^d$, the Rayleigh quotient (Parlett, 1998) $RQ(S,w)$ is defined as

$$RQ(S,w) = \frac{w^\top S w}{w^\top w}, \tag{1}$$

which reaches its minimum (maximum) value γ_1 (γ_d) when w is w_{min} (w_{max}) which is the corresponding eigenvector.

For two real symmetric matrices $H \in R^{d\times d}$, $G \in R^{d\times d}$, and a given non-zero real vector $w \in R^d$, the generalized Rayleigh quotient $Q(H,G,w)$ is defined as

$$RQ(H,G,w) = \frac{w^\top H w}{w^\top G w}, \tag{2}$$

which reaches its minimum (maximum) value γ_1 (γ_d) when w is equal to w_{min} (w_{max}). Here w_{min} (w_{max}) is the eigenvector of the generalized eigenvalue problem $Hw = \lambda G w$ corresponding to the minimum (maximum) eigenvalue.

2.2. MGSVMs

Consider a two-view binary classification problem, given n two-view examples with the label y_i $(i = 1,2,\dots,n) \in \{+1,-1\}$. Matrix $A_1' \in R^{n_1\times d_1}$ denotes the examples belonging to class +1 of view 1, and matrix $A_2' \in R^{n_1\times d_2}$ denotes the examples belonging to class +1 of view 2. Matrix $B_1' \in R^{n_2\times d_1}$ denotes the examples belonging to class −1 of view 1, and matrix $B_2' \in R^{n_2\times d_2}$ denotes the examples belonging to class −1 of view 2. Two hyperplanes for two views are constructed below.

$$\begin{aligned}&\text{view } one: \ w_1^{+\top} x_1 + b_1^+ = 0, \ w_1^{-\top} x_1 + b_1^- = 0,\\ &\text{view } two: \ w_2^{+\top} x_2 + b_2^+ = 0, \ w_2^{-\top} x_2 + b_2^- = 0,\end{aligned} \tag{3}$$

where x_1 and x_2 denote two views of x.

Define

$$\begin{aligned}&A_1 = [A_1' \quad e], \ A_2 = [A_2' \quad e],\\ &B_1 = [B_1' \quad e], \ B_2 = [B_2' \quad e],\\ &G_1 = A_1^\top A_1, \ H_1 = B_1^\top B_1, \ v_1 = \begin{bmatrix} w_1^+ \\ b_1^+ \end{bmatrix}, \ u_1 = \begin{bmatrix} w_1^- \\ b_1^- \end{bmatrix},\\ &G_2 = A_2^\top A_2, \ H_2 = B_2^\top B_2, \ v_2 = \begin{bmatrix} w_2^+ \\ b_2^+ \end{bmatrix}, \ u_2 = \begin{bmatrix} w_2^- \\ b_2^- \end{bmatrix}.\end{aligned} \tag{4}$$

For the simple description, e is the appropriate vector with all elements being 1 in the whole paper.

The optimization problems of MGSVMs (Sun et al., 2019) are given as follows

$$\begin{aligned}\min_{v_1\neq 0, v_2\neq 0} &\frac{v_1^\top G_1 v_1 + v_2^\top G_2 v_2 + c_1(\|v_1\|^2 + \|v_2\|^2)}{v_1^\top H_1 v_1 + v_2^\top H_2 v_2}\\ &+ \frac{c_2\|A_1 v_1 - A_2 v_2\|^2}{v_1^\top H_1 v_1 + v_2^\top H_2 v_2},\end{aligned} \tag{5}$$

$$\begin{aligned}\min_{u_1\neq 0, u_2\neq 0} &\frac{u_1^\top H_1 u_1 + u_2^\top H_2 u_2 + c_1(\|u_1\|^2 + \|u_2\|^2)}{u_1^\top G_1 u_1 + u_2^\top G_2 u_2}\\ &+ \frac{c_2\|B_1 u_1 - B_2 u_2\|^2}{u_1^\top G_1 u_1 + u_2^\top G_2 u_2},\end{aligned} \tag{6}$$

where c_1, c_2 are non-negative regularization parameters, and $\|v_1\|^2$, $\|v_2\|^2, \|u_1\|^2$, $\|u_2\|^2$ are Tikhonov norm regularization terms. The first optimization problem enforces +1 class hyperplane of view 1(2) to be close to the relevant +1 class and as far as possible from the relevant −1 class at the same time. The multi-view co-regularization terms $c_2\|A_1 v_1 - A_2 v_2\|^2$ and $c_2\|B_1 u_1 - B_2 u_2\|^2$ are considered as minimizing the distance discrepancies of two views.

By definition of $v = [v_1^\top \quad v_2^\top]^\top, u = [u_1^\top \quad u_2^\top]^\top$, the above optimization problems are equivalent to

$$\begin{aligned}\min_{v\neq 0} &\frac{v^\top \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} v + c_1\|v\|^2}{v^\top \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} v}\\ &+ \frac{c_2 v^\top [A_1 \quad -A_2]^\top [A_1 \quad -A_2] v}{v^\top \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} v},\end{aligned} \tag{7}$$

$$\min_{u\neq 0} \frac{u^\top \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} u + c_1\|u\|^2}{u^\top \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} u} + \frac{c_2 u^\top [B_1 \quad -B_2]^\top [B_1 \quad -B_2] u}{u^\top \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} u}, \tag{8}$$

define

$$N_1 = \begin{bmatrix} (1+c_2)G_1 & -c_2A_1^\top A_2 \\ -c_2A_2^\top A_1 & (1+c_2)G_2 \end{bmatrix} + c_1 I, \quad N_2 = \begin{bmatrix} (1+c_2)H_1 & -c_2B_1^\top B_2 \\ -c_2B_2^\top B_1 & (1+c_2)H_2 \end{bmatrix} + c_1 I,$$
$$M_1 = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}, \quad M_2 = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix},$$

and the optimization problems (7) and (8) can be further changed to

$$\min_{v\neq 0} \frac{v^\top N_1 v}{v^\top M_1 v}, \tag{9}$$

$$\min_{u\neq 0} \frac{u^\top N_2 u}{u^\top M_2 u}. \tag{10}$$

Their optimal solutions are gained by solving the following generalized eigenvalue problems, respectively.

$$N_1 v = \lambda M_1 v, N_2 u = \lambda M_2 u. \tag{11}$$

Suppose that they are gained, given a test example x, compute the perpendicular distances of x from two hyperplanes for each view

$$\begin{aligned} &\text{view one}: \; d_{11} = \frac{|x_1^\top w_1^+ + b_1^+|}{\|w_1^+\|}, \; d_{12} = \frac{|x_1^\top w_1^- + b_1^-|}{\|w_1^-\|}, \\ &\text{view two}: \; d_{21} = \frac{|x_2^\top w_2^+ + b_2^+|}{\|w_2^+\|}, \; d_{22} = \frac{|x_2^\top w_2^- + b_2^-|}{\|w_2^-\|}. \end{aligned} \tag{12}$$

Then the final decision function for the example x is provided by

$$\hat{y} = \mathrm{sign}(d_{12} + d_{22} - d_{11} - d_{21}), \tag{13}$$

where $\hat{y}$ is the two-view fused result.

2.3. MIGSVMs

The global representations of the hyperplanes and matrices are provided in Eqs. (3) and (4). Then the optimization problems of MIGSVMs (Sun et al., 2019) can be described by

$$\begin{aligned} \min_{v,\alpha} \;& \frac{1}{\alpha} v^\top \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} v - \frac{1}{\alpha} c_1 v^\top \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} v \\ & + \frac{1}{\alpha} c_2 v^\top [A_1 \quad -A_2]^\top [A_1 \quad -A_2] v \\ s.t. \;& \|v\|^2 = \alpha, \quad \alpha > 0, \end{aligned} \tag{14}$$

$$\begin{aligned} \min_{u,\alpha} \;& \frac{1}{\alpha} u^\top \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} u - \frac{1}{\alpha} c_1 u^\top \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} p \\ & + \frac{1}{\alpha} c_2 u^\top [B_1 \quad -B_2]^\top [B_1 \quad -B_2] u \\ s.t. \;& \|u\|^2 = \alpha, \quad \alpha > 0, \end{aligned} \tag{15}$$

where c_1 and c_2 are non-negative regularization parameters.

Define

$$W_1 = \begin{bmatrix} (1+c_2)G_1 - c_1H_1 & -c_2A_1^\top A_2 \\ -c_2A_2^\top A_1 & (1+c_2)G_2 - c_1H_2 \end{bmatrix},$$

$$W_2 = \begin{bmatrix} (1+c_2)H_1 - c_1G_1 & -c_2B_1^\top B_2 \\ -c_2B_2^\top B_1 & (1+c_2)H_2 - c_1G_2 \end{bmatrix},$$

and the optimization problems (14) and (15) are equivalent to

$$\begin{aligned} \min_{v,\alpha} \;& \frac{1}{\alpha} v^\top W_1 v \\ s.t. \;& \|v\|^2 = \alpha, \quad \alpha > 0, \end{aligned} \tag{16}$$

$$\begin{aligned} \min_{u,\alpha} \;& \frac{1}{\alpha} u^\top W_2 u \\ s.t. \;& \|u\|^2 = \alpha, \quad \alpha > 0. \end{aligned} \tag{17}$$

Then norm regularization terms were used to minimize the norm of the hyperplane parameter v and u motivated by IGEPSVMs, and then the optimization problems of MIGSVMs can be further written as

$$\begin{aligned} \min_{v,\alpha} \;& v^\top W_1 v + c_3\|v\|^2 \\ s.t. \;& \|v\|^2 = \alpha, \quad \alpha > 0, \end{aligned} \tag{18}$$

$$\begin{aligned} \min_{u,\alpha} \;& u^\top W_2 u + c_3\|u\|^2 \\ s.t. \;& \|u\|^2 = \alpha, \quad \alpha > 0, \end{aligned} \tag{19}$$

where c_3 is a regularization parameter. The optimal solutions of (18) and (19) are gained by solving the following eigenvalue problems

$$(W_1 + c_3 I)v = \lambda_1 v, \; (W_2 + c_3 I)u = \lambda_2 u. \tag{20}$$

For the simple description, I is an identity matrix of proper dimensions in the whole paper.

Suppose that the hyperplane parameters of MIGSVMs are gained, given a test example x, compute the perpendicular distances of x from two hyperplanes for each view

$$\begin{aligned} &\text{view one}: \; d_{11} = \frac{|x_1^\top w_1^+ + b_1^+|}{\|w_1^+\|}, \; d_{12} = \frac{|x_1^\top w_1^- + b_1^-|}{\|w_1^-\|}, \\ &\text{view two}: \; d_{21} = \frac{|x_2^\top w_2^+ + b_2^+|}{\|w_2^+\|}, \; d_{22} = \frac{|x_2^\top w_2^- + b_2^-|}{\|w_2^-\|}. \end{aligned} \tag{21}$$

Then the final decision function for the example x is provided by

$$\hat{y} = \mathrm{sign}(d_{12} + d_{22} - d_{11} - d_{21}), \tag{22}$$

where $\hat{y}$ is the two-view fused result.

3. Our proposed methods

MGPSVMs and MIGPSVMs are regarded as the first attempting of GEPSVMs and IGEPSVMs for multi-view learning. In this section, we develop them to handle generalized MVL cases with the consistency and complementary criterions.

3.1. Linear GMGSVMs

The optimization problems for linear GMGSVMs are written as

$$\begin{aligned} \min_{\rho_i, z_i^+} \;& \frac{\sum_{i=1}^T \rho_i \|A_i z_i^+\|^2 + c_1 \sum_{i=1}^T \|z_i^+\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|A_i z_i^+ - A_j z_j^+\|^2}{\sum_{i=1}^T \|B_i z_i^+\|^2} \\ & + \frac{\lambda}{2} \sum_{i=1}^T \rho_i^2 \\ \text{s.t.} \;& \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{23}$$

$$\min_{\rho_i, z_i^-} \frac{\sum_{i=1}^T \rho_i \|B_i z_i^-\|^2 + c_1 \sum_{i=1}^T \|z_i^-\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|B_i z_i^- - B_j z_j^-\|^2}{\sum_{i=1}^T \|A_i z_i^-\|^2} + \frac{\lambda}{2}\sum_{i=1}^T \rho_i^2$$
$$\text{s.t.} \quad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \tag{24}$$

where c_1, c_2, ρ_i, λ are the coefficients, z_i^+ and z_i^- are classifier vectors for the ith view. T is the number of views.

For introducing the mechanism of linear GMGSVMs, some points are provided below.

(1) The above optimization problems enforce $+1(-1)$ class hyperplane of each view to be close to the relevant $+1(-1)$ class, and as far as possible from the relevant $-1(+1)$ class at the same time. The weighted value ρ balances the norm regularization terms of multiple views and takes advantage of the complementarity information among them.

(2) The $c_2 \sum_{i=1}^T \sum_{i<j}^T \|A_i z_i^+ - A_j z_j^+\|^2$ and $c_2 \sum_{i=1}^T \sum_{i<j}^T \|B_i z_i^- - B_j z_j^-\|^2$ could be considered as minimizing the distance discrepancies of all views and are designed to maintain the consistency among distinct views. $\frac{\lambda}{2}\sum_{i=1}^T \rho_i^2$ can enforce the sub-optimization problem of ρ_i to be the classical QPP. By setting ρ_i to be a fixed value, the optimization sub-problem is formulated as

$$\min_{z_i^+} \frac{\sum_{i=1}^T \rho_i \|A_i z_i^+\|^2 + c_1 \sum_{i=1}^T \|z_i^+\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|A_i z_i^+ - A_j z_j^+\|^2}{\sum_{i=1}^T \|B_i z_i^+\|^2}, \tag{25}$$

$$\min_{z_i^-} \frac{\sum_{i=1}^T \rho_i \|B_i z_i^-\|^2 + c_1 \sum_{i=1}^T \|z_i^-\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|B_i z_i^- - B_j z_j^-\|^2}{\sum_{i=1}^T \|A_i z_i^-\|^2}. \tag{26}$$

Define

$$N_1 = \begin{pmatrix} \rho_1 A_1^\top A_1 + c_1 I + c_2(T-1)A_1^\top A_1 & \cdots & -c_2 A_1^\top A_T \\ \vdots & \ddots & \vdots \\ -c_2 A_T^\top A_1 & \cdots & \rho_T A_T^\top A_T + c_1 I + c_2(T-1)A_T^\top A_T \end{pmatrix},$$
$$N_2 = \begin{pmatrix} \rho_1 B_1^\top B_1 + c_1 I + c_2(T-1)B_1^\top B_1 & \cdots & -c_2 B_1^\top B_T \\ \vdots & \ddots & \vdots \\ -c_2 B_T^\top B_1 & \cdots & \rho_T B_T^\top B_T + c_1 I + c_2(T-1)B_T^\top B_T \end{pmatrix},$$
$$M_1 = \begin{pmatrix} B_1^\top B_1 & 0 & \cdots & 0 \\ 0 & B_2^\top B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & B_T^\top B_T \end{pmatrix}, M_2 = \begin{pmatrix} A_1^\top A_1 & 0 & \cdots & 0 \\ 0 & A_2^\top A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & A_T^\top A_T \end{pmatrix}, \tag{27}$$
$$z^+ = \begin{pmatrix} z_1^+ \\ \vdots \\ z_T^+ \end{pmatrix}, z^- = \begin{pmatrix} z_1^- \\ \vdots \\ z_T^- \end{pmatrix}.$$

The above optimization problems are equivalent to

$$\min_{z^+ \neq 0} \frac{z^{+\top} N_1 z^+}{z^{+\top} M_1 z^+}, \tag{28}$$

$$\min_{z^- \neq 0} \frac{z^{-\top} N_2 z^-}{z^{-\top} M_2 z^-}. \tag{29}$$

Their optimal solutions could be gained by resolving the following generalized eigenvalue problems

$$N_1 z^+ = \lambda_1 M_1 z^+, \tag{30}$$

$$N_2 z^- = \lambda_2 M_2 z^-. \tag{31}$$

For an example $x = (x_1 \ \dots \ x_T)$, if $\sum_{i=1}^T \rho_i |[x_i, 1] z_i^+| \leq \sum_{i=1}^T \rho_i |[x_i, 1] z_i^-|$, it is given to class $+1$, otherwise class -1.

3.2. Optimization of GMGSVMs

In the optimization of GMGSVMs, variables ρ, z_i^+ and z_i^- are demanded to be optimized, and at present there is no straightforward method to gain the global optimal solution. Thus an alternating algorithm motivated by Luo and Tseng (1992) is built, which continuously renews these variables and resolves each sub-optimization problem until convergence.

The optimization of z_i^+, z_i^-. ρ is predetermined as a random vector, in general, $\rho_i = \frac{1}{T}, i = 1, \dots, T$. Then the optimal solution is gained by resolving the generalized eigenvalue problems (30) and (31).

The optimization of ρ_i. Leveraging the gained z_i^+, z_i^-, the optimization of ρ_i could be written as a QQP.

$$\min_{\rho} \ \rho^\top p + \rho^\top q + \frac{\lambda}{2}\|\rho\|^2$$
$$\text{s.t.} \ \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \tag{32}$$

where $p = \frac{[\|A_1 z_1^+\|^2, \dots, \|A_T z_T^+\|^2]^\top}{\sum_{j=1}^T \|B_j z_j^+\|^2}$, $q = \frac{[\|B_1 z_1^-\|^2, \dots, \|B_T z_T^-\|^2]^\top}{\sum_{j=1}^T \|A_j z_j^-\|^2}$, $\rho = [\rho_1, \dots, \rho_T]^\top$. The coordinate descent algorithm employed selects only a pair of elements ρ_n and ρ_m to be renewed and the others are unchanged. By integrating the Lagrange method and the constraint, the updating criterion is as follows

$$\begin{cases} \rho_n^* = \frac{\lambda(\rho_n + \rho_m) + (p_m - p_n) + (q_m - q_n)}{2\lambda}, \\ \rho_m^* = \rho_n + \rho_m - \rho_n^*. \end{cases} \tag{33}$$

The gained ρ_n^* and ρ_m^* probably cannot obey the non-negative constraint. Therefore, if $\lambda(\rho_n + \rho_m) + (p_m - p_n) + (q_m - q_n) < 0$, make $\rho_n^* = 0$, and the same for ρ_m^*. The ceasing criterion for the algorithm is the discrepancy of current ρ and prevenient ρ. View Algorithm 1 for the pseudo-code of linear GMGSVM. The sub-problems (30) and (31) of the algorithm have the optimal analytical solutions. The sub-problem (32) of the algorithm is convex. Thus, the algorithm is guaranteed to converge the local optimum (Shen et al., 2018; Xu & Yin, 2017).

Algorithm 1 Linear GMGSVM

1: **Input:** Data $A_i, B_i, i = 1, \dots, T$, model parameters ($c_1$, c_2, λ).
2: Gain N_1, N_2, M_1 and M_2 based on the equation (27).
3: Predetermine $\rho_i^0 = 0, \rho_i^s = \frac{1}{T}, i = 1, \dots, T, s = 1$.
4: while E=$\|\rho^s - \rho^{s-1}\| > \delta$ do
5: Based on the equations (30) and (31), resolve the generalized eigenvalue problems and gain the values p and q.
6: Based on the updating criterion in the equation (33), renew the parameter ρ_i.
7: s=s+1.
8: End while
9: Gain the classifiers based on the equations (30) and (31), respectively.
10: **Output:** For a test example $x = (x_1 \ \cdots \ x_T)$, if $\sum_{i=1}^T \rho_i |[x_i, 1] z_i^+| \leq \sum_{i=1}^T \rho_i |[x_i, 1] z_i^-|$, it is given to class $+1$, otherwise class -1.

3.3. Non-linear GMGSVMs

In this section, we develop the non-linear case of GMGSVMs. The kernel-produced hyperplanes can be written as:

$$K(x_i^\top, S_i^\top) w_i^+ + b_i^+ = 0, \quad K(x_i^\top, S_i^\top) w_i^- + b_i^- = 0, \tag{34}$$

where K is the kernel function represented by $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. $\phi(\cdot)$ is a nonlinear mapping from a low-dimensional feature space to a latent high-dimensional one. S_i means training examples of the ith view, and $S_i = (A_i'^{\top}, B_i'^{\top})^{\top}$ where A_i' and B_i' denote positive examples and negative examples of the ith view, respectively. Define

$$A_i = (K(A_i', S_i^{\top}), e), B_i = (K(B_i', S_i^{\top}), e),$$
$$z_i^+ = \begin{pmatrix} w_i^+ \\ b_i^+ \end{pmatrix}, \; z_i^- = \begin{pmatrix} w_i^- \\ b_i^- \end{pmatrix}. \tag{35}$$

The optimization problems for non-linear GMGSVMs are given as

$$\begin{aligned} &\min_{\rho_i, z_i^+} \frac{\sum_{i=1}^T \rho_i \|A_i z_i^+\|^2 + c_1 \sum_{i=1}^T \|z_i^+\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|A_i z_i^+ - A_j z_j^+\|^2}{\sum_{i=1}^T \|B_i z_i^+\|^2} \\ &+ \frac{\lambda}{2} \sum_{i=1}^T \rho_i^2 \\ &\text{s.t.} \quad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{36}$$

$$\begin{aligned} &\min_{\rho_i, z_i^-} \frac{\sum_{i=1}^T \rho_i \|B_i z_i^-\|^2 + c_1 \sum_{i=1}^T \|z_i^-\|^2 + c_2 \sum_{i=1}^T \sum_{i<j}^T \|B_i z_i^- - B_j z_j^-\|^2}{\sum_{i=1}^T \|A_i z_i^-\|^2} \\ &+ \frac{\lambda}{2} \sum_{i=1}^T \rho_i^2 \\ &\text{s.t.} \quad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{37}$$

where c_1, c_2, ρ_i, λ are non-negative regularization parameters, T is the number of views.

The global optimization process is as identical as the linear GMGSVMs. For an example $x = (x_1 \; \dots \; x_T)$, if $\sum_{i=1}^T \rho_i |[K(x_i, S_i^{\top}), 1] z_i^+| \leq \sum_{i=1}^T \rho_i |[K(x_i, S_i^{\top}), 1] z_i^-|$, it is given to class $+1$, otherwise class -1.

3.4. GMIGSVMs

The optimization problems for GMIGSVMs can be written as

$$\begin{aligned} &\min_{\rho_i, z_+} \sum_{i=1}^T \rho_i \frac{\|A_i z_i^+\|^2}{\alpha} + c_1 \sum_{i=1}^T \sum_{i<j}^T \frac{1}{\alpha} \|A_i z_i^+ - A_j z_j^+\|^2 \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^T \rho_i^2 - \sum_{i=1}^T \frac{c_2 \|B_i z_i^+\|^2}{\alpha} \\ &\text{s.t.} \quad \sum_{i=1}^T \|z_i^+\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{38}$$

$$\begin{aligned} &\min_{\rho_i, z_-} \sum_{i=1}^T \rho_i \frac{\|B_i z_i^-\|^2}{\alpha} + c_1 \sum_{i=1}^T \sum_{i<j}^T \frac{1}{\alpha} \|B_i z_i^- - B_j z_j^-\|^2 \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^T \rho_i^2 - \sum_{i=1}^T \frac{c_2 \|A_i z_i^-\|^2}{\alpha} \\ &\text{s.t.} \quad \sum_{i=1}^T \|z_i^-\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{39}$$

where c_1, c_2, α, ρ_i, λ are non-negative regularization parameters, z_i^+ and z_i^- are classifier vectors for the ith view. T is the number of views.

Through simple transforms, the above optimizations can be stated as

$$\begin{aligned} &\min_{\rho_i, z^+} \frac{z^{+\top}}{\alpha} \begin{bmatrix} \rho_1 A_1^{\top} A_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \rho_T A_T^{\top} A_T \end{bmatrix} z^+ + \frac{c_1}{\alpha} z^{+\top} \\ &\begin{bmatrix} (T-1) A_1^{\top} A_1 & \cdots & -A_1^{\top} A_T \\ \vdots & \ddots & \vdots \\ -A_T^{\top} A_1 & \cdots & (T-1) A_T^{\top} A_T \end{bmatrix} z^+ \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^T \rho_i^2 - c_2 \frac{z^{+\top}}{\alpha} \begin{bmatrix} B_1^{\top} B_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & B_T^{\top} B_T \end{bmatrix} z^+ \\ &\text{s.t.} \quad \|z^+\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{40}$$

$$\begin{aligned} &\min_{\rho_i, z^-} \sum_{i=1}^T \frac{z^{-\top}}{\alpha} \begin{bmatrix} \rho_1 B_1^{\top} B_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \rho_T B_T^{\top} B_T \end{bmatrix} z^- + \sum_{i=1}^T \frac{c_1}{\alpha} \\ &z^{-\top} \begin{bmatrix} (T-1) B_1^{\top} B_1 & \cdots & -B_1^{\top} B_T \\ \vdots & \ddots & \vdots \\ -B_T^{\top} B_1 & \cdots & (T-1) B_T^{\top} B_T \end{bmatrix} z^- \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^T \rho_i^2 - \sum_{i=1}^T \frac{c_2}{\alpha} z^{-\top} \begin{bmatrix} A_1^{\top} A_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & A_T^{\top} A_T \end{bmatrix} z^- \\ &\text{s.t.} \quad \|z^-\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1. \end{aligned} \tag{41}$$

Define

$$\begin{aligned} &M_1 = \\ &\begin{pmatrix} \rho_1 A_1^{\top} A_1 + c_1 (T-1) A_1^{\top} A_1 - c_2 B_1^{\top} B_1 & \cdots & -c_1 A_1^{\top} A_T \\ \vdots & \ddots & \vdots \\ -c_1 A_T^{\top} A_1 & \cdots & \rho_T A_T^{\top} A_T + c_1 (T-1) A_T^{\top} A_T - c_2 B_T^{\top} B_T \end{pmatrix}, \\ &M_2 = \\ &\begin{pmatrix} \rho_1 B_1^{\top} B_1 + c_1 (T-1) B_1^{\top} B_1 - c_2 A_1^{\top} A_1 & \cdots & -c_1 B_1^{\top} B_T \\ \vdots & \ddots & \vdots \\ -c_1 B_T^{\top} B_1 & \cdots & \rho_T B_T^{\top} B_T + c_1 (T-1) B_T^{\top} B_T - c_2 A_T^{\top} A_T \end{pmatrix}, \\ &z^+ = \begin{pmatrix} z_1^+ \\ \vdots \\ z_T^+ \end{pmatrix}, z^- = \begin{pmatrix} z_1^- \\ \vdots \\ z_T^- \end{pmatrix}, \rho = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_T \end{pmatrix}. \end{aligned} \tag{42}$$

The optimization problems of GMIGSVMs can be simplified to the following forms

$$\begin{aligned} &\min_{\rho, z^+} z^{+\top} M_1 z^+ + \frac{\lambda \|\rho\|^2}{2} + c_3 \|z^+\|^2 \\ &\text{s.t.} \quad \|z^+\|^2 = \alpha, \alpha > 0, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{43}$$

$$\begin{aligned} &\min_{\rho, z^-} z^{-\top} M_2 z^- + \frac{\lambda \|\rho\|^2}{2} + c_3 \|z^-\|^2 \\ &\text{s.t.} \quad \|z^-\|^2 = \alpha, \alpha > 0, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^T \rho_i = 1, \end{aligned} \tag{44}$$

where norm regularization terms are designed to minimize the norm of the hyperplane variable z^+ and z^-. When ρ is fixed, the optimal solutions of the optimization problems (43) and (44) can be gained by resolving the following eigenvalue problems

$$(M_1 + c_3 I) z^+ = \lambda_1 z^+, \tag{45}$$

$$(M_2 + c_3 I)z^- = \lambda_2 z^-. \tag{46}$$

For a test example $x = (x_i \ \dots \ x_T)$, if $\sum_{i=1}^{T} \rho_i |[x_i, 1]z_i^+| \leq \sum_{i=1}^{T} \rho_i |[x_i, 1]z_i^-|$, it is given to class +1, otherwise class −1.

3.5. Optimization of GMIGSVM

Similarly with the optimization of GMGSVMs, the optimization of GMIGSVMs tends to be as follows.

The optimization of z^+, z^-. ρ is predetermined as a random vector, in general, $\rho_i = \frac{1}{T}, i = 1, \dots, T$. Then the optimal solution is gained by resolving the eigenvalue problems (45) and (46).

The optimization of ρ_i. Leveraging the gained z^+, z^-, the optimization of ρ can be written as a QQP.

$$\begin{aligned} &\min_{\rho} \ \ \rho^\top p + \rho^\top q + \frac{\lambda}{2}\|\rho\|^2 \\ &\text{s.t.} \ \ \rho_i \geq 0, \sum_{i=1}^{T} \rho_i = 1, \end{aligned} \tag{47}$$

where $p = [\|A_1 z_1^+\|_2^2, \dots, \|A_T z_T^+\|_2^2]^\top, q = [\|B_1 z_1^-\|_2^2, \dots, \|B_T z_T^-\|_2^2]^\top$. We also use the coordinate descent algorithm and the updating criterion is

$$\begin{cases} \rho_n^* = \frac{\lambda(\rho_n+\rho_m)+(p_m-p_n)+(q_m-q_n)}{2\lambda}, \\ \rho_m^* = \rho_n + \rho_m - \rho_n^*. \end{cases} \tag{48}$$

The gained ρ_n^* and ρ_m^* probably cannot obey the non-negative constraint. Therefore, if $\lambda(\rho_n + \rho_m) + (p_m - p_n) + (q_m - q_n) < 0$, make $\rho_n^* = 0$, and the same for ρ_m^*. The ceasing criterion for the algorithm is the discrepancy of current ρ and prevenient ρ. View Algorithm 2 for the pseudo-code of linear GMIGSVMs. The sub-problems (45) and (46) of the algorithm have the optimal analytical solutions. The sub-problem (47) of the algorithm is convex. Thus, the algorithm is guaranteed to converge the local optimum (Shen et al., 2018; Xu & Yin, 2017).

Algorithm 2 Linear GMIGSVMs

Input: Data $A_i, B_i, i = 1, \dots, T$, model parameters (c_1, c_2, λ).
Gain M_1 and M_2 based on the equation (42).
Predetermine $\rho_i^0 = 0, \rho_i^s = \frac{1}{T}, i = 1, \dots, T, s = 1$.
while E=$\|\rho^s - \rho^{s-1}\| > \delta$ do
Based on the equations (45) and (46), resolve the eigenvalue problems and gain the values p and q.
Based on the updating rules in the equation (48), renew the parameter ρ_i.
s=s+1.
End while
Gain the classifiers based on the equations (45) and (46), respectively.
Output: For a test example $x = (x_1 \ \cdots \ x_T)$, if $\sum_{i=1}^{T} \rho_i |[x_i, 1]z_i^+| \leq \sum_{i=1}^{T} \rho_i |[x_i, 1]z_i^-|$, it is given to class +1, otherwise class −1.

3.6. Non-linear GMIGSVMs

In this section, we develop the non-linear case of GMIGSVMs. Define

$$\begin{aligned} &A_i = (K(A_i', S_i^\top), e), B_i = (K(B_i', S_i^\top), e), \\ &z_i^+ = \begin{pmatrix} w_i^+ \\ b_i^+ \end{pmatrix}, \ z_i^- = \begin{pmatrix} w_i^- \\ b_i^- \end{pmatrix}. \end{aligned} \tag{49}$$

Table 1
Real-world datasets.

Name	Attributes	Numbers	Classes
Corel	89	68,040	99
Handwritten digit dataset	649	2000	10
Movies dataset	3276	617	17
Caltech-101	3766	9146	101

The optimization problems for non-linear GMIGSVMs can be written as

$$\begin{aligned} &\min_{\rho_i, z_i^+} \sum_{i=1}^{T} \frac{\|A_i z_i^+\|^2}{\alpha} + c_1 \sum_{i=1}^{T} \sum_{i<j}^{T} \frac{1}{\alpha}\|A_i z_i^+ - A_j z_j^+\|^2 \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^{T} \rho_i^2 - \sum_{i=1}^{T} \frac{c_2\|B_i z_i^+\|^2}{\alpha} \\ &\text{s.t.} \ \sum_{i=1}^{T} \|z_i^+\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^{T} \rho_i = 1, \end{aligned} \tag{50}$$

$$\begin{aligned} &\min_{\rho_i, z_i^-} \sum_{i=1}^{T} \frac{\|B_i z_i^-\|^2}{\alpha} + c_1 \sum_{i=1}^{T} \sum_{i<j}^{T} \frac{1}{\alpha}\|B_i z_i^- - B_j z_j^+\|^2 \\ &+ \frac{\lambda}{2\alpha} \sum_{i=1}^{T} \rho_i^2 - \sum_{i=1}^{T} \frac{c_2\|A_i z_i^-\|^2}{\alpha} \\ &\text{s.t.} \ \sum_{i=1}^{T} \|z_i^-\|^2 = \alpha, \\ &\qquad \rho_i \geq 0, \sum_{i=1}^{T} \rho_i = 1, \end{aligned} \tag{51}$$

where $c_1, c_2, \alpha, \rho_i, \lambda$ are non-negative regularization parameters, z_i^+ and z_i^- are classifier vectors for the ith view. T is the number of views.

The global optimization process is as identical as the linear GMIGSVMs. For a test example $x = (x_1 \ \dots \ x_T)$, if $\sum_{i=1}^{T} \rho_i |[K(x_i, S_i^\top), 1] z_i^+| \leq \sum_{i=1}^{T} \rho_i |[K(x_i, S_i^\top), 1]z_i^-|$, it is given to class +1, otherwise class −1.

In short, GMGSVM solves two generalized eigenvalue problems and owns time complexity $O(s(\sum_{i=1}^{T} d_i + T)^3)$ (linear case) and $O(s(n+1)^3 T^3)$ (non-linear case). GMIGSVM solves standard eigenvalue decomposition problems and owns time complexity $O(s(\sum_{i=1}^{T} d_i + T)^2)$ (linear case) and $O(s(n+1)^2 T^2)$ (non-linear case).

4. Experiments

In this section, we validate the performance of our proposed GMGSVM and GMIGSVM on one synthetic dataset and four real-world datasets. The four real-world datasets are corel, handwritten digit dataset from UCI machine learning repository, movies dataset[1] and caltech-101.[2]Detail information about them is listed in Table 1.

4.1. Experimental setup

The best parameters are determined by the ten-fold cross validation for all the methods in the set $[2^{-3}, 2^{-2}, \dots, 2^3]$, and the classification average and the standard deviation are gained by executing these algorithms for five times. We first set $\rho_i = \frac{1}{3}$. The algorithm ceases the iteration when the discrepancy of the objective function is less than $\delta = 10^{-3}$. Two-view classification methods MGSVM and MIGSVM are used for comparison.

[1] http://membres-lig.imag.fr/grimal/data.html.
[2] http://www.vision.caltech.edu/Image_Datasets/Caltech101.

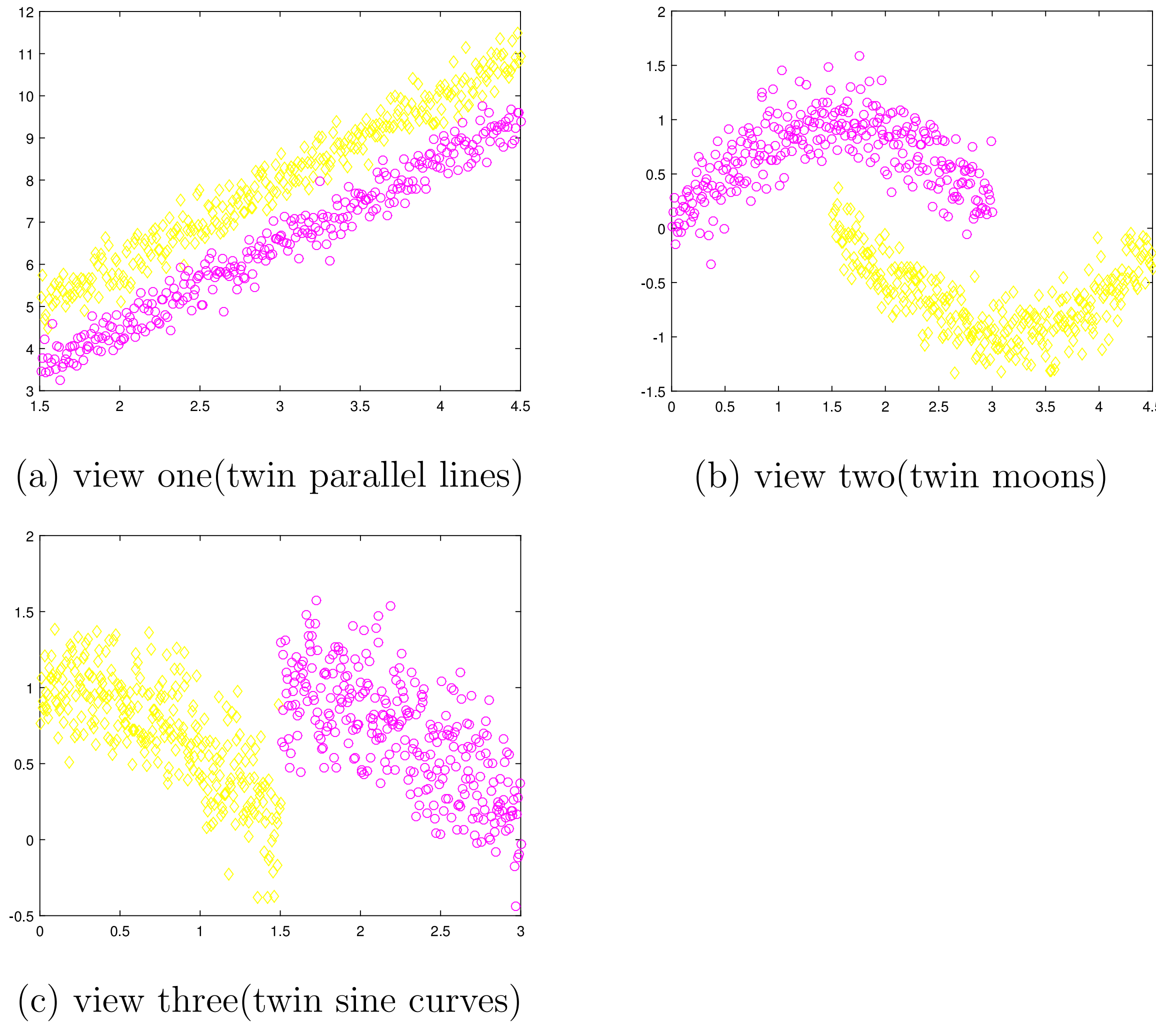

Fig. 1. Synthetic data.

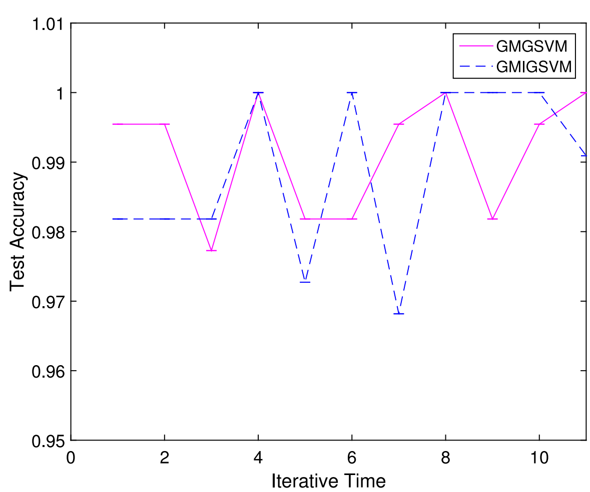

Fig. 2. The influence of the weight ρ for the performance.

4.2. Synthetic data

The synthetic dataset is generated by twin moons, twin sine curves and twin parallel lines in Fig. 1. The data on a moon are connected randomly to data on a line and a sine. It includes 300 examples (half of them are positive examples and half of them are negative examples).

Table 2
Performance (%) on synthetic data.

Method	GMGSVM	MGSVM1	MGSVM2	MGSVM3
Accuracy	**99.91**(0.20)	74.64(13.27)	97.36(0.75)	88.45(3.21)
Running time	0.0566	0.0303	0.0284	0.0285
Method	GMIGSVM	MIGSVM1	MIGSVM2	MIGSVM3
Accuracy	99.82(0.25)	94.64(2.41)	98.55(0.81)	96.27(2.75)
Running time (s)	0.0083	0.0124	0.0144	0.0124

80 examples are used for the training and the remaining are used for the test. Table 2 shows the experimental results with the linear kernel. The best performance of all methods is labeled by the bold. We can observe that GMGSVM and GMIGSVM own higher accuracies than the relevant two-view classification methods, which indicates the fact that our proposed methods can take advantage of the consistency and complementarity criterions of multiple views. To study the influence of the weight ρ for the performance, Fig. 2 plots the iterative number and the corresponding accuracy. The initial weight ρ can make all views share the same weight. Other weights ρ may generate the better performance than the one of the initial ρ. Thus the learning of the weight ρ is needed.

4.3. Movies dataset

This dataset describes the identical movies with two aspects. It is widely used for co-clustering cases whose purpose is to find the type of

Table 3
Performance (%) on movies dataset.

Method	GMGSVM	MGSVM1	MGSVM2	MGSVM3
Accuracy	**62.07**(6.45)	60.69(9.32)	57.93(5.67)	60.69(9.63)
Running time (s)	0.009	0.004	0.0045	0.0033
Method	GMIGSVM	MIGSVM1	MIGSVM2	MIGSVM3
Accuracy	**62.07**(5.45)	60.69(7.55)	60.00(4.63)	57.93(9.25)
Running time	0.0061	0.0020	0.0044	0.0036

Table 4
Performance (%) on corel dataset.

Method	GMGSVM	MGSVM1	MGSVM2	MGSVM3
Accuracy	70.99(2.51)	67.18(8.72)	65.21(9.30)	62.96(12.14)
Running time (s)	0.1096	0.0290	0.0304	0.0361
Method	GMIGSVM	MIGSVM1	MIGSVM2	MIGSVM3
Accuracy	**71.27**(1.68)	70.85(2.20)	65.92(6.64)	69.86(4.91)
Running time (s)	0.0335	0.0117	0.0119	0.0155

the movies by mining the information from the two views (keywords and actors).

Two classes are selected for multi-view classification. The numbers of properties of the three views are 1398, 1878 and 1000 (decreasing from 1878 to 1000 with PCA). From the experimental results with the linear kernel in Table 3, we can observe that our methods GMGSVM and GMIGSVM perform the same and better than the relevant two-view classification methods.

4.4. Corel

This image dataset extracts the features from a corel image gathering. Four feature sets and the relevant dimensionality are as follows, color histogram (thirty-two), color histogram layout (thirty-two), color moments (nine) and co-occurrence texture (sixteen).

222 and 80 examples are chosen for the training and test, respectively. We choose three views of them for the experiment. From the experimental results with the polynomial kernel of degree two listed in Table 4, we can observe that the performances of our methods GMGSVM and GMIGSVM are better than the relevant two-view classification methods.

4.5. Handwritten digit dataset

This dataset is involved with the features of handwritten digits (zero-nine) constructed from a gathering of Dutch utility maps. It has 2000 images with binary forms (200 images for each digit) totally. Six feature sets can be considered as six views which are FOU, FAC, KAR, PIX, ZER and MOR.

Among them, FOU, FAC and KAR are chosen for experiments randomly. For each digit, we randomly choose 40 examples for training, and the remaining for test. We choose four digit pairs (1,7), (2,6), (4,8) and (0,9) for the experimental validation. From the experimental results the linear kernel in Table 5, we can observe that the performances of our methods GMGSVM and GMIGSVM are better than the other MVL methods, and GMIGSVM performs best. To sum up, we can conclude that our methods can mine the consistency and complementarity information of multiple views in contrast with the relevant two-view classification methods in the pairwise pattern.

4.6. Caltech-101

The dataset consists of 9146 images totally which owns 101 classes, and an additional background/clutter class. Each class consists of 40~800 images on average. Among them, we randomly select two classes for generalized multi-view classification. We adopt 80 examples as the training set and 790 examples for test. The experimental results with the polynomial kernel of degree two are listed in Table 6.

In short, our methods GMGSVM and GMIGSVM own higher accuracies than the relevant two-view classification methods, which indicates the fact that they can fully take advantage of the weighted value to obey the complementarity criterion, integrate multi-view regularization terms to obey the consistency criterion and fully take advantage of information among multiple views for higher classification performance, while MGSVM and MIGSVM only obey the consistency criterion between two views and can deal with two-view classification case.

5. Conclusion and future work

In this paper, we propose two generalized multi-view extensions of generalized eigenvalues proximal support vector machines which can handle generalized MVL cases. The weighted value exploits the complementarity information among different views. We also present an efficient alternating algorithm for the optimization of the classifier parameters and weighted value. Experimental results on real-world datasets show that GMGSVM and GMIGSVM are superior to other multi-view classification algorithms in classification performance. Future works contain the extensions of GMGSVM and GMIGSVM to the semi-supervised learning.

CRediT authorship contribution statement

Xijiong Xie: Conceptualization, Methodology, Writing – original draft. **Yujie Xiong:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by Ningbo University talent project 421703670 as well as programs sponsored by K.C. Wong Magna Fund in Ningbo University. It is also supported by the NSFC of China 61906101 and 62006131.

Table 5
Performance (%/s) on handwritten digit dataset.

Method	Two classes			
	(1,7)	(2,6)	(4,8)	(0,9)
GMGSVM(fou kar fac)	98.19(0.26)/0.1137	99.31(0.26)/0.0951	98.87(0.17)/0.1031	99.56(0.36)/0.1159
GMIGSVM(fou kar fac)	**98.63**(0.93)/0.0266	**99.44**(0.26)/0.0240	**98.94**(0.47)/0.0231	**99.94**(0.14)/0.0279
MGSVM1(fou kar)	98.00(0.52)/0.0309	95.25(0.81)/0.0317	97.75(1.28)/0.0385	97.50(0.31)/0.0405
MGSVM2(fou fac)	95.12(1.10)/0.0340	98.00(1.14)/0.0329	92.62(9.09)/0.0442	99.50(0.17)/0.0441
MGSVM3(kar fac)	97.06(0.57)/0.0340	98.12(0.80)/0.0528	95.81(3.66)/0.0504	99.44(0.26)/0.0317
MIGSVM1(fou kar)	97.63(0.90)/0.0125	96.56(2.01)/0.0118	97.75(0.41)/0.0129	98.31(1.30)/0.0113
MIGSVM2(fou fac)	96.31(1.30)/0.0134	99.06(0.54)/0.0118	97.19(1.38)/0.0116	99.06(1.23)/0.0120
MIGSVM3(kar fac)	97.38(0.87)/0.0109	98.75(0.96)/0.0123	98.12(0.44)/0.0130	99.31(0.41)/0.0121

Table 6
Performance (%) on Caltech-101.

Method	GMGSVM	MGSVM1	MGSVM2	MGSVM3
Accuracy	54.00(0.63)	52.96(1.45)	52.00(0.84)	53.49(1.74)
Running time (s)	0.1015	0.0297	0.0331	0.0341
Method	GMIGSVM	MIGSVM1	MIGSVM2	MIGSVM3
Accuracy	**54.68**(2.26)	53.95(0.86)	51.82(0.64)	53.77(2.32)
Running time (s)	0.0277	0.0072	0.0108	0.0113

References

Bartlett, P., & Mendelson, S. (2002). Rademacher and Gaussian complexities:risk bounds and structural results. *Journal of Machine Learning Research*, *3*, 463–482.

Chao, G., & Sun, S. (2016a). Alternative multi-view maximum entropy discrimination. *IEEE Transactions on Neural Networks and Learning Systems*, *27*, 1445–1456.

Chao, G., & Sun, S. (2016b). Consensus and complementarity based maximum entropy discrimination for multi-view classification. *Information Sciences*, *367*, 296–310.

Chao, G., & Sun, S. (2016c). Multi-kernel maximum entropy discrimination for multi-view learning. *Intelligent Data Analysis*, *20*, 481–493.

Chen, W., Shao, Y., Xu, D., & Fu, Y. (2014). Manifold proximal support vector machine for semi-supervised classification. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *40*, 623–638.

Farquhar, J., Hardoon, D., Shawe-Taylor, J., & Szedmak, S. (2006). Two view learning:SVM-2K, theory and practice. In *Advances in neural information processing systems, Vol. 18* (pp. 355–362).

Houthuys, L., Langone, R., & Suykens, J. (2018). Multi-view least squares support vector machines classification. *Neurocomputing*, *22*, 78–88.

Jayadeva, Khemchandani, R., & Chandra, S. (2007). Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 905–910.

Khemchandani, R., Karpatne, A., & Chandra, S. (2013). Generalized eigenvalue proximal support vector regressor. *Expert Systems with Applications*, *38*, 13136–13142.

Luo, Z., & Tseng, P. (1992). On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, *72*, 7–35.

Mangasarian, O., & Wild, E. (2006). Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 69–74.

Mao, L., & Sun, S. (2016). Soft margin consistency based scalable multi-view maximum entropy discrimination. In *Proceedings of the 25th international joint conference on artificial intelligence* (pp. 1839–1845).

Nigam, K., & Ghani, R. (2000). Analyzing the effectiveness and applicability of co-training. In *Proceedings of the 9th international conference on information and knowledge management* (pp. 86–93).

Parlett, B. (1998). *The symmetric eigenvalue problem*. Philadel-phia: SIAM.

Shao, Y., Deng, N., & Chen, W. (2013). A proximal classifier with consistency. *Knowledge-Based Systems*, *49*, 171–178.

Shao, Y., Deng, N., Chen, W., & Wang, Z. (2013). Improved generalized eigenvalue proximal support vector machine. *IEEE Signal Processing Letters*, *20*, 213–216.

Shawe-Taylor, J., & Sun, S. (2011). A review of optimization methodologies in support vector machines. *Neurocomputing*, *74*, 3609–3618.

Shen, Y., Xu, H., & Liu, X. (2018). An alternating minimization method for robust principal component analysis. *Optimization Methods & Software*, *34*, 1251–1276.

Sun, S., & Chao, G. (2013). Multi-view maximum entropy discrimination. In *Proceedings of the 23rd international joint conference on artificial intelligence* (pp. 1706–1712).

Sun, X., Chen, Y., Shao, Y., Li, C., & Wang, C. (2018). Robust nonparallel proximal support vector machine with lp-norm regularization. *IEEE Access*, *6*, 20334–20347.

Sun, S., & Jin, F. (2011). Robust co-training. *International Journal of Pattern Recognition and Artificial Intelligence*, *25*, 1113–1126.

Sun, S., & Shawe-Taylor, J. (2010). Sparse semi-supervised learning using conjugate functions. *Journal of Machine Learning Research*, *11*, 2423–2455.

Sun, S., Shawe-Taylor, J., & Mao, L. (2017). PAC-Bayes analysis of multi-view learning. *Information Fusion*, *35*, 117–131.

Sun, S., Xie, X., & Dong, C. (2019). Multiview learning with generalized eigenvalue proximal support vector machines. *IEEE Transactions on Cybernetics*, *49*, 688–697.

Tang, J., Li, D., Tian, Y., & Liu, D. (2018). Multi-view learning based on nonparallel support vector machine. *Knowledge-Based Systems*, *158*, 94–108.

Tang, J., Tian, Y., Liu, D., & Kou, G. (2019). Coupling privileged kernel method for multi-view learning. *Information Sciences*, *481*, 110–127.

Tang, J., Tian, Y., Liu, X., Li, D., Lv, J., & Kou, G. (2018). Improved multi-view privileged support vector machine. *Neural Networks*, *106*, 96–109.

Tang, J., Tian, Y., Zhang, P., & Liu, X. (2018). Multiview privileged support vector machines. *IEEE Transactions on Neural Networks and Learning Systems*, *29*, 3463–3477.

Xie, X. (2018). Regularized multi-view least squares twin support vector machines. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *48*, 3108–3115.

Xie, X., & Sun, S. (2015). Multi-view twin support vector machines. *Intelligent Data Analysis*, *19*, 701–712.

Xu, Y., & Yin, W. (2017). A globally convergent algorithm for nonconvex optimization based on block coordinate update. *Journal of Scientific Computing*, *72*, 700–734.

Yan, H., Ye, Q., Zhang, Y., Yu, D., & Xu, Y. (2018). L1-Norm GEPSVM classifier based on an effective iterative algorithm for classification. *Neural Processing Letters*, *48*, 273–298.

Yang, Z., Guo, Y., & Li, C. (2015). Local k-proximal plane clustering. *Neural Computing and Applications*, *26*, 199–211.

Yang, X., Pan, Z., & Chen, S. (2009). Semi-supervised proximal support vector machine via generalized eigenvalues. *Pattern Recognition and Artificial Intelligence*, *22*, 349–353.

Zhao, J., Xie, X., Xu, X., & Sun, S. (2017). Multi-view learning overview:Recent progress and new challenges. *Information Fusion*, *38*, 43–54.