SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

August 2022, Vol. 65 182102:1–182102:18
https://doi.org/10.1007/s11432-021-3319-4

Unlabeled data driven cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization

Xiaohui YANG[1], Zheng WANG[1], Jian SUN[2] & Zongben XU[2*]

[1]*Henan Engineering Research Center for Artificial Intelligence Theory and Algorithms, School of Mathematics and Statistics, Henan University, Kaifeng* 475004*, China;*
[2]*National Engineering Laboratory for Big Data Analytics, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an* 750041*, China*

Received 1 January 2021/Revised 8 April 2021/Accepted 12 July 2021/Published online 25 July 2022

Abstract Sparse representation-based classification (SRC) has been widely used because it just relies on simple linear regression ideas to do classification, and it does not need learning. However, the performance of SRC is limited by needing sufficient labeled samples per class and the sensitivity to class imbalance. For tackling these problems, an improved SRC model is constructed in this paper. For alleviating the problem of insufficient labeled samples, an unlabeled data-driven inverse projection sparse representation-based classification model is constructed to achieve effective and stable representation and recognition results. The $L_{1/2}$ and $S_{1/2}$ regularizations are introduced to capture the sparsity of 1-D and 2-D, and to make the model have good statistical properties. Moreover, the cost-sensitive strategy is integrated into the model's classification criteria to improve the imbalance of class distribution adaptively, especially for multiclass imbalanced data. A solver utilizing the mixed Gauss-Seidel and Jacobian ADMM algorithm is developed to obtain the approximate solution. Experiments on common public test databases show that the proposed model achieves competitive results compared with the latest published results and some deep-learning algorithms.

Keywords unlabeled-data driven, 1/2 regularization, cost-sensitive, inverse projection, sparse representation-based classification

Citation Yang X H, Wang Z, Sun J, et al. Unlabeled data driven cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization. Sci China Inf Sci, 2022, 65(8): 182102, https://doi.org/10.1007/s11432-021-3319-4

1 Introduction

Class imbalance [1] and small labeled samples are challenging problems in the field of pattern recognition. The techniques to deal with class imbalance mainly include data level [2–4] and algorithm level [5–7]. Among them, the cost-sensitive method of algorithm level is superior to other methods [8, 9], especially the cost-sensitive deep learning method has achieved good results [10,11]. For insufficient labeled samples problem, few-shot learning methods based on transfer learning or data augmentation are commonly used technologies. In transfer learning, pre-training is carried out in the source domain, and then the weights learned are fine-tuned in the target domain with small data [12, 13]. Data augmentation is another technique to expand the size of the dataset by creating modified versions of the dataset [14,15]. However, these methods still fundamentally rely on a large number of labeled samples or need to learn the effective representation of data with the help of complex network structures. It is a meaningful and interesting work how to design a cost-sensitive data representation method that does not require learning, and does not rely on a large number of labeled samples.

Wright et al. [16] proposed sparse representation-based classification (SRC), which was a sparse coding technique based on sufficient labeled samples per class without learning. SRC has made remarkable success in pattern recognition fields, such as face recognition [17,18] and tumor recognition [19,20]. However,

* Corresponding author (email: zbxu@mail.xjtu.edu.cn)

the classification performance of SRC is degraded when facing problems with class imbalance and few labeled samples per class. Our previous work proposed inverse projection representation-based classification (IPRC) [21], inverse space sparse representation-based classification (ISSRC) [22], pseudo-full-space representation based-classification (PFSRC) [23], and low-rank sparse pseudo-full-space representation-based classification (LR-S-PFSRC) [24]. They all presented the inverse projection sparse representation-based classification model which can effectively alleviate the insufficient labeled samples problem. Inverse projection sparse representation expands representation space via accessible unlabeled data and explores available information among existing samples especially for unlabeled data, which frees labeled samples from quantity needs. Moreover, the classification criterion designed to match inverse projection sparse representation restricts attention to the coefficients before unlabeled samples rather than the coefficients before labeled samples, which shows inverse projection sparse representation-based classification is less sensitive to the number of labeled samples. In addition, inverse projection sparse representation considers different regularizations according to the data prior, which is crucial to the insufficient labeled samples problem.

However, these inverse projection sparse representation-based classification models still have the following limitations. First, the L_1 regularization which characterizes 1-D sparsity of data does not yield sufficiently sparse solution [25] and the nuclear norm [26] which characterizes 2-D sparsity (low rank) of data [27] may yield a matrix with a much higher rank than the real one [28,29]. Second, the matched classification decision criterion directly assumes all misclassifications contribute equal losses, which ignores the class imbalance problem.

With the deepgoing research on the L_1 regularization, Xu et al. [25] proposed the $L_{1/2}$ regularization which had better sparsity than the L_1 regularization. The $L_{1/2}$ regularization also has many excellent statistical properties, such as unbiasedness [30] and Oracle properties [31]. And, the $L_{1/2}$ regularization can completely replace the L_p $(0 < p < 1)$ regularization [25] for the sparsity problem. In addition, Xu et al. [32] proposed a fast solver for the $L_{1/2}$ regularization. According to the $L_{1/2}$ regularization theory, Rao et al. [27] extended the $L_{1/2}$ norm to a matrix and used the $S_{1/2}$ norm and the $L_{1/2}$ norm to describe the 2-D sparsity and 1-D sparsity of matrix, respectively. The author designed an alternating threshold iterative algorithm inspired by the alternating direction method of multipliers (ADMM) framework. ADMM has attracted much attention because it mainly deals with convex constrained optimization problems, while the algorithm cannot guarantee a convergent solution when facing optimization problems with more than two variables. Lu et al. [33] proposed a mixed Gauss-Seidel and Jacobian ADMM (M-ADMM), which can deal with multivariate optimization problems that ADMM may not solve. In addition, the cost-sensitive method offers an effective solution for the class imbalance problem, where it employs different penalties for different types of misclassification. Elkan [34] described the misclassification cost by cost matrix. By introducing "unequal cost" information and constructing a cost matrix, the imbalance of class distribution can be alleviated to a certain extent.

Inspired by these studies, this paper proposes a cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization ($S_{1/2}$-$L_{1/2}$-PFSRC-CS) to alleviate class imbalance and insufficient labeled samples. The proposed model is an extension of SRC, which aims to obtain the best possible representation by using existing data and its inherent characteristics. Similar to SRC [16], for face recognition, the studies in this paper are confined to human frontal face recognition, and assume faces have been performed detection, cropping, and normalization. Also worth noting is that $S_{1/2}$-$L_{1/2}$-PFSRC-CS focuses on limited labeled samples and representation without learning. Here, limited labeled samples mean that there are few labeled samples and others are unlabeled. The main differences between the proposed $S_{1/2}$-$L_{1/2}$-PFSRC-CS and the related studies [16, 21–24] are as follows. (1) $S_{1/2}$-$L_{1/2}$-PFSRC-CS focuses on a completely opposite projection way to [16], uses unlabeled data to expand the representation space to alleviate the impact of insufficient labeled data and can select different representation spaces according to the inherent characteristics of a represented object. (2) Our previous work [21–24] proposed an inverse projection, where Refs. [21, 22] were for tumor recognition and Refs. [23, 24] were for face recognition. Compared with [21–24], $S_{1/2}$-$L_{1/2}$-PFSRC-CS further induces the sparsity and low rank of data through the $L_{1/2}$ regularization and the $S_{1/2}$ regularization. (3) Refs. [16, 21–24] did not consider the misclassification cost caused by class imbalance data. The contributions of this paper are as follows.

(1) An inverse projection sparse representation with 1/2 regularization ($S_{1/2}$-$L_{1/2}$-PFSR) is constructed to alleviate insufficient labeled samples problem. The model driven by unlabeled data exploits the $L_{1/2}$ regularization and the $S_{1/2}$ regularization to further enhance the sparsity.

(2) A sample and robust classification decision criterion, minimum misclassification cost, is designed

to match the proposed model and complete the classification.

(3) The $S_{1/2}$-$L_{1/2}$-PFSR combined with minimum misclassification cost is called $S_{1/2}$-$L_{1/2}$-PFSRC-CS, whose optimization, convergence performance, role of unlabeled data, sparsity performance and classification performance are fully verified on face databases and tumor databases.

The remainder of the paper is summarized as follows. Section 2 mainly includes the theoretical framework of $S_{1/2}$-$L_{1/2}$-PFSR. The optimization of the proposed model is discussed in Section 3. Following that, Section 4 presents $S_{1/2}$-$L_{1/2}$-PFSRC-CS. Section 5 conducts extensive experiments on face databases and tumor databases. Finally, Section 6 comes to conclusions.

2 Construction of $S_{1/2}$-$L_{1/2}$-PFSR model

2.1 Sparse representation

Suppose $X = [x_1, x_2, \ldots, x_{s_c}] \in \mathbb{R}^{d\times s_c}$ is a labeled samples set, $Y = [y_1, y_2, \ldots, y_m] \in \mathbb{R}^{d\times m}$ is an unlabeled samples set. In SRC, each unlabeled sample $y_l \in \mathbb{R}^d$ can be linearly represented by all labeled samples:

$$y_l = \varphi_{l,1}x_1 + \cdots + \varphi_{l,i}x_i + \cdots + \varphi_{l,s_c}x_{s_c} = \sum_{i=1}^{s_c} \varphi_{l,i}x_i = X\varphi_l, \tag{1}$$

where $\varphi_l \in \mathbb{R}^{s_c}$ is representation coefficients. Without causing confusion, the projection way and representation space of SRC are called positive projection and positive projection representation space.

The inverse projection is opposite to positive projection. Suppose $x_i \in X$ is a labeled sample, $\kappa(x_i)$ is its corresponding representation space. Then x_i is linearly represented by $\kappa(x_i)$ as $x_i = \kappa(x_i)b_i$, where b_i is inverse projection representation coefficients. $\kappa(x_i)$ can be different inverse projection representation space according to the inherent characteristics of labeled data. Pseudo-full-space [23] and inverse space [21] are both inverse projection representation space, where inverse space is a special case of pseudo-full-space. Pseudo-full-space $V_i = \{X, Y\} - \{x_i\}$ $(i = 1, \ldots, s_c)$ is composed of all labeled samples and unlabeled samples except the labeled sample x_i itself. Inverse space Y is composed of all unlabeled samples. It is analyzed and verified in detail from [23] that inverse projection can extract the information hidden in unlabeled samples and reduce dependence on the quantity of labeled samples per class, which plays a crucial role in solving insufficient labeled samples problem.

2.2 $S_{1/2}$-$L_{1/2}$-PFSR model

In this subsection, to further enhance the sparsity, sparse regularization and low-rank regularization are introduced into the inverse projection sparse representation.

For traditional sparsity problems, the L_1 regularization is the optimal convex approximation of the L_0 regularization, while Xu et al. [25] have revealed that the $L_{1/2}$ regularization produces better sparse solutions than the L_1 regularization. In addition, the $L_{1/2}$ regularization is the sparsest and most robust among the L_p $(1/2 \leqslant p < 1)$ regularization, and the L_p $(0 < p < 1/2)$ regularization has similar properties to the $L_{1/2}$ regularization. For traditional low-rank problems, the rank function is often relaxed as the nuclear norm. The nuclear norm is the L_1 norm of the singular value vector. Motivated by the superior properties of the $L_{1/2}$ norm, the $S_{1/2}$ norm is proposed to replace the rank function and reveal the subspace structure [27].

Suppose $X = [x_1, x_2, \ldots, x_{s_c}] = [X_1, \ldots, X_j, \ldots, X_c] \in \mathbb{R}^{d\times s_c}$ is a labeled samples set, $X_j = [X_{s_{j-1}+1}, \ldots, X_{s_j}] \in \mathbb{R}^{d\times(s_j - s_{j-1})}$ are the j-th class samples, $j = 1, \ldots, c$ is the number of class, $Y = [y_1, y_2, \ldots, y_m] \in \mathbb{R}^{d\times m}$ is an unlabeled samples set. For each labeled sample $x_i \in X$, choose its largest inverse projection representation space, pseudo-full-space $V_i = \{X, Y\} - \{x_i\}$. Pseudo-full-space utilizes existing available samples rather than constructs auxiliary labeled samples. The inverse projection representation of x_i by pseudo-full-space fully explores the complementary information embedded in existing available samples. Futhermore, the inverse projection sparse representation with 1/2 regularization can also be expressed as seeking pseudo-full-space representation based on the $S_{1/2}$ regularization and the $L_{1/2}$ regularization, namely, $S_{1/2}$-$L_{1/2}$-pseudo-full-space representation ($S_{1/2}$-$L_{1/2}$-PFSR).

$$\min_W \|W\|_{S_{1/2}}^{1/2} + \lambda_1 \|W\|_{L_{1/2}}^{1/2},$$

$$\text{s.t. } X = VW,\ W_{i,i} = 0, \tag{2}$$

where λ_1 is a regularization parameter, $V = [V_1, \ldots, V_{s_c}]$ is the pseudo-full-space matrix, W is representation coefficient matrix, $W_{i,i}$ represents the i-th element of vector W_i, and $W_{i,i} = 0$ means x_i do not exist in V_i. $\|W\|_{S_{1/2}}$ denotes the $S_{1/2}$ norm of W as a low-rank regular constraint to induce structure information, where $\|W\|_{S_{1/2}} = (\sum_{n=1}^{r} \delta_n^{1/2})^2$ and $\{\delta_n\}_{n=1}^{r}$ represents singular values of the matrix W. $\|W\|_{L_{1/2}}$ denotes the $L_{1/2}$ norm of W as a sparse regular constraint to induce sparse information, where $\|W\|_{L_{1/2}} = (\sum_{i=1}^{s_c+m} \sum_{j=1}^{s_c} |W_{i,j}|^{1/2})^2$.

When samples are mildly or grossly corrupted, the labeled data matrix can be rewritten as $X = X_0 + E = VW + E$, in which X_0 is the clean labeled data matrix and E is the error matrix. Thus, a more reasonable $S_{1/2}$-$L_{1/2}$-PFSR is rewritten as

$$\begin{aligned} &\min_{W,E} \|W\|_{S_{1/2}}^{1/2} + \lambda_1 \|W\|_{L_{1/2}}^{1/2} + \lambda_2 \|E\|_1, \\ &\text{s.t. } X = VW + E,\ W_{i,i} = 0, \end{aligned} \tag{3}$$

where λ_2 is a regularization parameter. $S_{1/2}$-$L_{1/2}$-PFSR is mainly for data that contains sparity, subspace structure, and complementarity information between samples. If there is only sparsity and no subspace structure in data, two regularizations of $S_{1/2}$-$L_{1/2}$-PFSR are degenerated into sparse regularization. If there is also no complementary information in data, the representation space of $S_{1/2}$-$L_{1/2}$-PFSR is degenerated into inverse space because labeled data from other classes may lead to interference information rather than complementary information. Then, $L_{1/2}$-inverse-space representation ($L_{1/2}$-ISR), a special case of $S_{1/2}$-$L_{1/2}$-PFSR, is constructed.

$$\min_{A} \|A\|_{L_{1/2}}^{1/2}, \quad \text{s.t. } X = YA, \tag{4}$$

where A is representation coefficient matrix, and $\|A\|_{L_{1/2}} = (\sum_{i=1}^{m} \sum_{j=1}^{s_c} |A_{i,j}|^{1/2})^2$. For the $L_{1/2}$ regularization, similar to [25], it can be analyzed that the $L_{1/2}$-ISR has many promising properties.

Theorem 1. The $L_{1/2}$-ISR possesses sparsity, unbiasedness, and Oracle properties.

Proof. Fan et al. [30] has already proved the sparsity and unbiasedness of the $L_{1/2}$ regularizer. Knight et al. [31] studied the asymptotic normal property of the L_1 and the L_1 type regularizers essentially proved that the L_p $(0 < p < 1)$ regularizer has the Oracle property. Therefore, Theorem 1 can be directly derived from [30, 31].

3 Optimization of $S_{1/2}$-$L_{1/2}$-PFSR model

In this section, inspired by the M-ADMM [33], we propose an alternating threshold iterative algorithm to solve the proposed model.

Considering that the objective function and constraint conditions in (3) are inseparable with respect to low rank and sparse constraints, auxiliary variables Z and J need to be introduced for optimization. Eq. (3) can be converted into an equivalent optimization problem as

$$\begin{aligned} &\min_{W,Z,J,E} \|Z\|_{S_{1/2}}^{1/2} + \lambda_1 \|J\|_{L_{1/2}}^{1/2} + \lambda_2 \|E\|_1, \\ &\text{s.t. } X = VW + E,\ W = Z,\ W = J,\ W_{i,i} = 0,\ Z_{i,i} = 0,\ J_{i,i} = 0. \end{aligned} \tag{5}$$

However, Eq. (5) is a multivariable nonconvex minimization problem. Although M-ADMM algorithm is commonly used to solve multivariable convex optimization problems, its idea can be extended to solve nonconvex optimization problems, just like the optimization idea of ADMM can be extended to the solution of nonconvex optimization problems [27]. Based on this, this paper designs an alternating threshold iterative algorithm to solve (5).

3.1 Existence of orthogonal matrices

It is necessary to divide the four variables W, Z, J, E into two super blocks to ensure the effective alternating iteration of the optimization algorithm. According to the variable division method of M-ADMM [33], we first verify (5) contains orthogonal matrices. For the three constraints in (5), $X =$

$VW+E$, $W=J$, and $W=Z$, they can be converted into a joint constraint:

$$\begin{bmatrix} V \\ I \\ I \end{bmatrix} W + \begin{bmatrix} 0 \\ -I \\ 0 \end{bmatrix} Z + \begin{bmatrix} 0 \\ 0 \\ -I \end{bmatrix} J + \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} E = \begin{bmatrix} X \\ 0 \\ 0 \end{bmatrix}. \tag{6}$$

Then, Eq. (5) can be converted into an equivalent optimization problem,

$$\begin{aligned} &\min_{W,Z,J,E} \|Z\|_{S_{1/2}}^{1/2} + \lambda_1 \|J\|_{L_{1/2}}^{1/2} + \lambda_2 \|E\|_1, \\ &\text{s.t.} \begin{bmatrix} V \\ I \\ I \end{bmatrix} W + \begin{bmatrix} 0 \\ -I \\ 0 \end{bmatrix} Z + \begin{bmatrix} 0 \\ 0 \\ -I \end{bmatrix} J + \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} E = \begin{bmatrix} X \\ 0 \\ 0 \end{bmatrix}, \ W_{i,i}=0, \ Z_{i,i}=0, \ J_{i,i}=0. \end{aligned} \tag{7}$$

In (7), the coefficient matrices of variables Z, J, and E are orthogonal to each other. Hence, it is verified that Eq. (7) contains orthogonal matrices. Therefore, the variables in (5) can be divided into two super blocks $\{Z, J, E\}$ and $\{W\}$. During the optimization algorithm, two super blocks $\{Z, J, E\}$ and $\{W\}$ are alternately iterated.

3.2 Construction of projection operators

Next, three projection operators need to be constructed to ensure the constraint conditions $W_{i,i}=0$, $Z_{i,i}=0$, and $J_{i,i}=0$ in the coefficient matrices.

For the restriction $W_{i,i}=0$, $W \in \mathbb{R}^{(s_c+m)\times s_c}$, let $\Omega = \mathbb{R}^{(s_c+m)\times s_c}$ is a closed convex set. Doing a projection operator projects W into Ω,

$$P_\Omega(W) \in \{W \in \mathbb{R}^{(s_c+m)\times s_c} \mid W_{i,i}=0, \ i=1,\ldots,s_c\}, \quad (P_\Omega(W))_{q,i} = \begin{cases} W_{q,i}, & \text{if } q \neq i, \\ 0, & \text{if } q=i, \end{cases} \ W_{q,i} \in W, \tag{8}$$

where $(P_\Omega(W))_{q,i}$ is a component of $P_\Omega(W)$. Then the restriction $W_{i,i}=0$ is converted to $W=P_\Omega(W)$.

Similar to $W_{i,i}=0$, for the restriction $Z_{i,i}=0$, $Z \in \mathbb{R}^{(s_c+m)\times s_c}$ and $J_{i,i}=0$, $J \in \mathbb{R}^{(s_c+m)\times s_c}$, two other projection operators will be done to project Z and J into Ω, respectively.

$$P_\Omega(Z) \in \{Z \in \mathbb{R}^{(s_c+m)\times s_c} \mid Z_{i,i}=0, i=1,\ldots,s_c\}, \ (P_\Omega(Z))_{q,i} = \begin{cases} Z_{q,i}, & \text{if } q \neq i, \\ 0, & \text{if } q=i, \end{cases} \ Z_{q,i} \in Z, \tag{9}$$

where $(P_\Omega(Z))_{q,i}$ is a component of $P_\Omega(Z)$. Then the restriction $Z_{i,i}=0$ is converted to $Z=P_\Omega(Z)$.

$$P_\Omega(J) \in \{J \in \mathbb{R}^{(s_c+m)\times s_c} \mid J_{i,i}=0, i=1,\ldots,s_c\}, \ (P_\Omega(J))_{q,i} = \begin{cases} J_{q,i}, & \text{if } q \neq i, \\ 0, & \text{if } q=i, \end{cases} \ J_{q,i} \in J, \tag{10}$$

where $(P_\Omega(J))_{q,i}$ is a component of $P_\Omega(J)$. Then the restriction $J_{i,i}=0$ is converted to $J=P_\Omega(J)$.

3.3 Optimization

Based on the existence of orthogonal matrices and the constructions of three projection operators for $W_{i,i}=0$, $Z_{i,i}=0$, and $J_{i,i}=0$, Eq. (5) can be solved. In order to facilitate the optimization, we can transform (5) to

$$\begin{aligned} &\min_{W,Z,J,E} \frac{1}{2}\|Z\|_{S_{1/2}}^{1/2} + \frac{\lambda_1}{2}\|J\|_{L_{1/2}}^{1/2} + \lambda_2 \|E\|_1, \\ &\text{s.t. } X = VW + E, \ W = Z, \ W = J, \ W_{i,i}=0, \ Z_{i,i}=0, \ J_{i,i}=0. \end{aligned} \tag{11}$$

The augmented Lagrange function of (11) is

$$L(W,Z,J,E,Y_1,Y_2,Y_3,\mu) = \frac{1}{2}\|Z\|_{S_{1/2}}^{1/2} + \frac{\lambda_1}{2}\|J\|_{L_{1/2}}^{1/2} + \lambda_2\|E\|_1 + \langle Y_1, X-VW-E\rangle + \langle Y_2, W-Z\rangle$$

$$+\langle Y_3, W-J\rangle+\frac{\mu}{2}(\|X-VW-E\|^2+\|W-Z\|^2+\|W-J\|^2),$$

where Y_1, Y_2, and Y_3 are Lagrangian multipliers, and $\mu>0$ is a penalty parameter. The alternating threshold iterative algorithm takes the following form:

$$\begin{cases} Z^{k+1}=\arg\min\limits_{Z}\dfrac{1}{2}\|Z\|_{S_{1/2}}^{1/2}+\langle Y_2^k, W^k-Z\rangle+\dfrac{\mu}{2}\|W^k-Z\|^2,\\ J^{k+1}=\arg\min\limits_{J}\dfrac{\lambda_1}{2}\|J\|_{L_{1/2}}^{1/2}+\langle Y_3^k, W^k-J\rangle+\dfrac{\mu}{2}\|W^k-J\|^2,\\ E^{k+1}=\arg\min\limits_{E}\lambda_2\|E\|_1+\langle Y_1^k, X-VW^k-E\rangle+\dfrac{\mu}{2}\|X-VW^k-E\|^2,\\ W^{k+1}=\arg\min\limits_{W}\langle Y_1^k, X-VW-E^{k+1}\rangle+\langle Y_2^k, W-Z^{k+1}\rangle+\langle Y_3^k, W-J^{k+1}\rangle\\ \qquad\quad+\dfrac{\mu}{2}(\|X-VW-E^{k+1}\|^2+\|W-Z^{k+1}\|^2+\|W-J^{k+1}\|^2),\\ Y_1^{k+1}=Y_1^k+\mu(X-VW^{k+1}-E^{k+1}),\\ Y_2^{k+1}=Y_2^k+\mu(W^{k+1}-Z^{k+1}),\\ Y_3^{k+1}=Y_3^k+\mu(W^{k+1}-J^{k+1}). \end{cases}\tag{12}$$

The optimization of $S_{1/2}$-$L_{1/2}$-PFSR model is outlined in Algorithm 1, whose major computation is dominated by Steps 1, 2, and 4. For Step 1, the SVD of matrices is required to compute in which its total complexity is $O((s_c+m)(s_c)^2)$. For half-thresholding operator in Step 2, the complexity is $O(s_c(s_c+m))$. The computational complexity of the inverse operation is $O((s_c+m)^3)$ in Step 4. The overall computational complexity in all iterations is $O(k(s_c+m)^3)$, where k is the total number of iterations.

Algorithm 1 Optimization of $S_{1/2}$-$L_{1/2}$-PFSR model

Input: Labeled sample set $X=[x_1,x_2,\ldots,x_{s_c}]$ and unlabeled sample set $Y=[y_1,y_2,\ldots,y_m]$.

Initialize: $Z^0=J^0=W^0=Y_2^0=Y_3^0=0$, $E^0=Y_1^0=0$, $\rho>1$, $\mu^0>0$, $\mu_{\max}>0$, $k=0$.

While not converged do

Step 1. Update Z : $Z^{k+1}=U^kH_{\frac{1}{\mu^k}}(D^k)(V^k)^{\mathrm{T}}$, $W^k+\frac{Y_2^k}{\mu^k}\approx U^kD^k(V^k)^{\mathrm{T}}$,

and $(P_\Omega(Z^{k+1}))_{q,i}=\begin{cases}Z_{q,i}^{k+1}, & \text{if } q\neq i,\\ 0, & \text{if } q=i,\end{cases}\quad Z_{q,i}^{k+1}\in Z^{k+1}$.

Step 2. Update J : $J^{k+1}=H_{\frac{\lambda_1}{\mu^k}}(W^k+\frac{Y_3^k}{\mu^k})$ and $(P_\Omega(J^{k+1}))_{q,i}=\begin{cases}J_{q,i}^{k+1}, & \text{if } q\neq i,\\ 0, & \text{if } q=i,\end{cases}\quad J_{q,i}^{k+1}\in J^{k+1}$.

Step 3. Update E : $E^{k+1}=S_{\frac{\lambda_2}{\mu^k}}(X-VW^k+\frac{Y_1^k}{\mu^k})$.

Step 4. Update W : $W^{k+1}=(V^{\mathrm{T}}V+2I)^{-1}(V^{\mathrm{T}}(X-E^{k+1}+\frac{Y_1^k}{\mu^k})-\frac{Y_2^k}{\mu^k}-\frac{Y_3^k}{\mu^k}+J^{k+1}+Z^{k+1})$

and $(P_\Omega(W^{k+1}))_{q,i}=\begin{cases}W_{q,i}^{k+1}, & \text{if } q\neq i,\\ 0, & \text{if } q=i,\end{cases}\quad W_{q,i}^{k+1}\in W^{k+1}$.

Step 5. Update Y_1,Y_2,Y_3 : $\begin{cases}Y_1^{k+1}=Y_1^k+\mu^k(X-VW^{k+1}-E^{k+1}),\\ Y_2^{k+1}=Y_2^k+\mu^k(W^{k+1}-Z^{k+1}),\\ Y_3^{k+1}=Y_3^k+\mu^k(W^{k+1}-J^{k+1}).\end{cases}$

Step 6. Update μ : $\mu^{k+1}=\min(\rho\mu^k,\mu_{\max})$.

Step 7. Update k : $k=k+1$.

End while

Output: An optimal solution can be obtained.

Since Eq. (4) is a special case of (3), its optimal solution can be obtained in the same way. Eq. (4) can be optimized by introducing an auxiliary variable so there is no need to verify the existence of the orthogonal matrix. Meanwhile, the $L_{1/2}$-ISR projects labeled samples onto inverse space composed of unlabeled samples so there is no need to construct projection operators in optimization.

Eq. (5) is a multivariable nonconvex minimization problem, if there exists a convex approximation, and then the convergence speed of the convex approximation can be characterized by the following convergence bound theorem [33]. Similar to [33], in order to give the theorem, suppose that the convex approximation

of (5) is as follows:

$$\min_{x_{B_1},x_{B_2}} f(x),\ \text{s.t. } \Phi_{B_1}x_{B_1}+\Phi_{B_2}x_{B_2}=d, \tag{13}$$

where $f(x)$ is convex, $x_{B_1}=\{Z,J,E|Z_{i,i}=0,J_{i,i}=0\}$ and $x_{B_2}=\{W|W_{i,i}=0\}$.

Theorem 2 (convergence bound theorem)**.** In optimization algorithm of (13), assume that $\widehat{f}^k \in S_{\{L_i,P_i\}_{i=1}^n}(f,x^k)$ [33] with $P_i \succeq L_i \succeq 0$, $\frac{1}{2}\|\Phi_{B_1}x_{B_1}\|^2$ is $\{L_i'\}_{i\in B_1}$-smooth [33], $\frac{1}{2}\|\Phi_{B_2}x_{B_2}\|^2$ is $\{L_i'\}_{i\in B_2}$ -smooth [33], $G_i \succeq L_i'-\Phi_i^{\mathrm{T}}\Phi_i$, $i\in B_1$ in $\frac{1}{2}\|\Phi_i x_i+\sum_{j\in B_1,j\neq i}\Phi_j x_j^k+\Phi_{B_2}x_{B_2}^k-d+\frac{T^k}{\mu^{(k)}}\|^2+\frac{1}{2}\|x_i-x_i^k\|_{G_i}^2+e_i^k$, $i\in B_1$ with $e_i^{k\prime}s$ satisfying $\sum_{i\in B_1}e_i^k=-\|\Phi x^k-d+\frac{T^k}{\mu^{(k)}}\|^2$, and $G_i \succ L_i'-\Phi_i^{\mathrm{T}}\Phi_i$, $i\in B_2$ in $\frac{1}{2}\|\Phi_i x_i+\sum_{j\in B_2,j\neq i}\Phi_j x_j^k+\Phi_{B_1}x_{B_1}^{k+1}-d+\frac{T^k}{\mu^{(k)}}\|^2+\frac{1}{2}\|x_i-x_i^k\|_{G_i}^2+e_i^k$, $i\in B_2$ with $e_i^{k\prime}s$ satisfying $\sum_{i\in B_2}e_i^k=0$. For any $K>0$, let $\overline{x}^k=\sum_{k=0}^K\gamma^{(k)}x^{k+1}$ with $\gamma^{(k)}=(\mu^{(k)})^{-1}/\sum_{k=0}^K(\mu^{(k)})^{-1}$. Then

$$f(\overline{x}^k)-f(x^*)+\langle\Phi^{\mathrm{T}}T^*,\overline{x}^K-x^*\rangle+\frac{\mu^{(0)}\alpha}{2}\|\Phi\overline{x}^k-d\|^2\leqslant\left(\sum_{j=1}^2\|x_{B_j}^*-x_{B_j}^0\|_{H_j^0}^2+\|T^*-T^0\|_{H_3^0}^2\right)\Big/2\sum_{k=0}^K(\mu^{(k)})^{-1},$$

where T is Lagrange multiplier, $\overline{x}^k$ is a weighted sum of $x^{k\prime}s$, and α, H_1^0, H_2^0, H_3^0 are as follows: $\alpha=\min\{\frac{1}{2},\frac{\delta_{\min}^2(\mathrm{Diag}\{\Phi_i^{\mathrm{T}}\Phi_i+G_i,\,i\in B_2\}-\Phi_{B_2}^{\mathrm{T}}\Phi_{B_2})}{2\|\Phi_{B_2}\|_2^2}\}$, $H_1^0=\mathrm{Diag}\{\frac{1}{\mu^{(0)}}L_i+\Phi_i^{\mathrm{T}}\Phi_i+G_i,i\in B_1\}-\Phi_{B_1}^{\mathrm{T}}\Phi_{B_1}$, $H_2^0=\mathrm{Diag}\{\frac{1}{\mu^{(0)}}L_i+\Phi_i^{\mathrm{T}}\Phi_i+G_i,i\in B_2\}$, $H_3^0=(\frac{1}{\mu^{(0)}})^2I$.

4 Cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization

In this section, minimum misclassification cost is presented to match $S_{1/2}$-$L_{1/2}$-PFSR and improve classification performance in class imbalance problem.

An unlabeled sample y_k is classified into the optimal class $\phi(y_k)$ with the minimum misclassification cost, namely,

$$\phi(y_k)=\arg\min_{v\in\{1,2,\ldots,c\}}\mathrm{loss}(y_k,v)=\arg\min_{v\in\{1,2,\ldots,c\}}\left(\sum_{u=1}^c P_{u,v}\mathrm{CII}_{k,u}\right), \tag{14}$$

where $P\in\mathbb{R}^{c\times c}$ is a cost matrix, $P_{u,v}\in P$ ($P_{u,v}\geqslant 0$) denotes the penalty that the real class is class u and the predicted class is class v. $\mathrm{CII}_{k,u}$ is category importance index (CII), which can be calculated by

$$\mathrm{CII}_{k,u}=\frac{1}{n_u}\sum_i\frac{\delta_u(\{|b_{i,k}|\}_{i=1,\ldots,s_c})}{\|\{b_{i,k}\}_{k=1,\ldots,m}\|_1}, \tag{15}$$

where $u=1,\ldots,c$, $k=1,\ldots,m$, $b_{i,k}$ is representation coefficients before y_k, δ_u is a vector whose entries are 0 except those associated with the u-th class, and n_u denotes the number of u-th class labeled samples. $\mathrm{CII}_{k,u}$ can be regarded as the probability that unlabeled sample y_k belongs to class u. The larger the $\mathrm{CII}_{k,u}$ is, the higher the probability is.

A cost-sensitive method is called $S_{1/2}$-$L_{1/2}$-PFSRC-CS which integrates $S_{1/2}$-$L_{1/2}$-PFSR and minimum misclassification cost, whose algorithm and framework are illustrated in Algorithm 2 and Figure 1, respectively. Similarly, another algorithm is called $S_{1/2}$-$L_{1/2}$-PFSRC which integrates $S_{1/2}$-$L_{1/2}$-PFSR and maximum category importance index. In addition, two other classification algorithms, $L_{1/2}$-ISRC-CS which integrates $L_{1/2}$-ISR and minimum misclassification cost and $L_{1/2}$-ISRC which integrates $L_{1/2}$-ISR and maximum category importance index, are also considered in experiments.

The proposed cost-sensitive method depends on the cost matrix P. Here, the property of the cost matrix is further discussed. When the cost matrix P is selected as the diagonal elements are 0 and the remaining elements are 1, we can prove that the predicted class obtained by the minimum misclassification cost is the same as the class obtained by the maximum category importance index. That is, $S_{1/2}$-$L_{1/2}$-PFSRC is a special case of $S_{1/2}$-$L_{1/2}$-PFSRC-CS. Namely, according to (14) and (15), the simple analysis

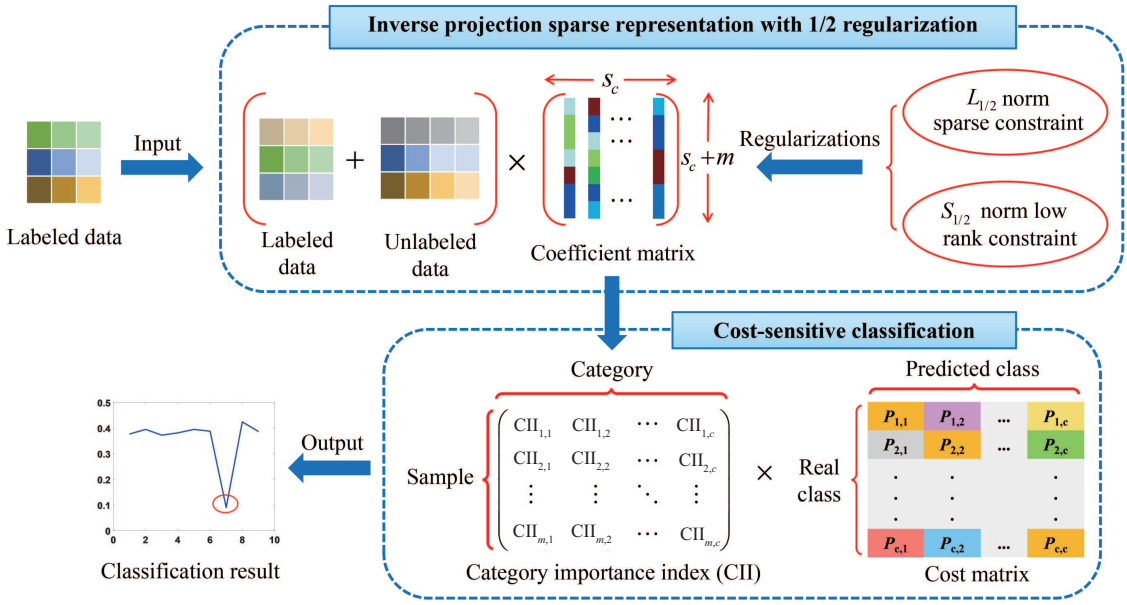

Figure 1 (Color online) Cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization.

Algorithm 2 Cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization

Input: Labeled sample set $X = [x_1, x_2, \ldots, x_{s_c}]$ and unlabeled sample set $Y = [y_1, y_2, \ldots, y_m]$.
1: The inverse projection sparse representation with 1/2 regularization is realized by Eq. (12).
2: The CIIs of each unlabeled sample y_k to all classes are calculated by Eq. (15).
3: The misclassification costs of each unlabeled sample y_k to all classes are calculated by Eq. (14).
Output: Get the class of y_k according to the minimum misclassification cost.

process is as follows:

$$\begin{aligned}\phi(y_k) &= \arg\min_{v\in\{1,2,\ldots,c\}} \sum_{u=1}^{c} P_{u,v}\mathrm{CII}_{k,u} = \arg\min_{v\in\{1,2,\ldots,c\}} \sum_{u\neq v} \mathrm{CII}_{k,u} \\ &= \arg\min_{v\in\{1,2,\ldots,c\}} (\mathrm{Constant} - \mathrm{CII}_{k,v}) = \arg\max_{v\in\{1,2,\ldots,c\}} (\mathrm{CII}_{k,v}).\end{aligned}$$

5 Experiments and discussions

In this section, in order to verify the effectiveness of the proposed model, experiments are carried out on face databases including Extended Yale B database [35], CMU PIE database [36], and AR database [37], as well as tumor databases including DLBCL database [38], Leukemia database [39], and 9_Tumors database [40]. According to the inherent characteristics of data, $S_{1/2}$-$L_{1/2}$-PFSRC-CS is applied to face databases, and $L_{1/2}$-ISRC-CS is applied to tumor databases. For face databases, we independently run all the methods five times and take the means as the final results. For tumor databases, we run the ten-fold cross validation thirty times and take the means as the final results. Since a certain class of 9_Tumors database only contains 2 samples, we run the five-fold cross validation thirty times and take the means as the final results. Before the tumor recognition, 200 genes will be pre-selected by the between-groups to the within-groups sum of squares (BW) [41]. All experiments are carried out using MATLAB R2016a on a 3.30 GHz machine with 4.00 GB RAM. This section mainly verifies the convergence, role of unlabeled data, sparsity, effectiveness for class imbalance problem as well as comparison with the latest published results and some deep-learning algorithms. Since the inverse projection sparse representation-based classification methods [21–24] have been proven to be superior to some classic classifiers, here $S_{1/2}$-$L_{1/2}$-PFSRC-CS is only compared with $S_{1/2}$-$L_{1/2}$-PFSRC, LR-S-PFSRC [24], PFSRC [23], and LRSRC [18]. And $L_{1/2}$-ISRC-CS is only compared with $L_{1/2}$-ISRC, ISSRC [22], IPRC [21], and SRC [16].

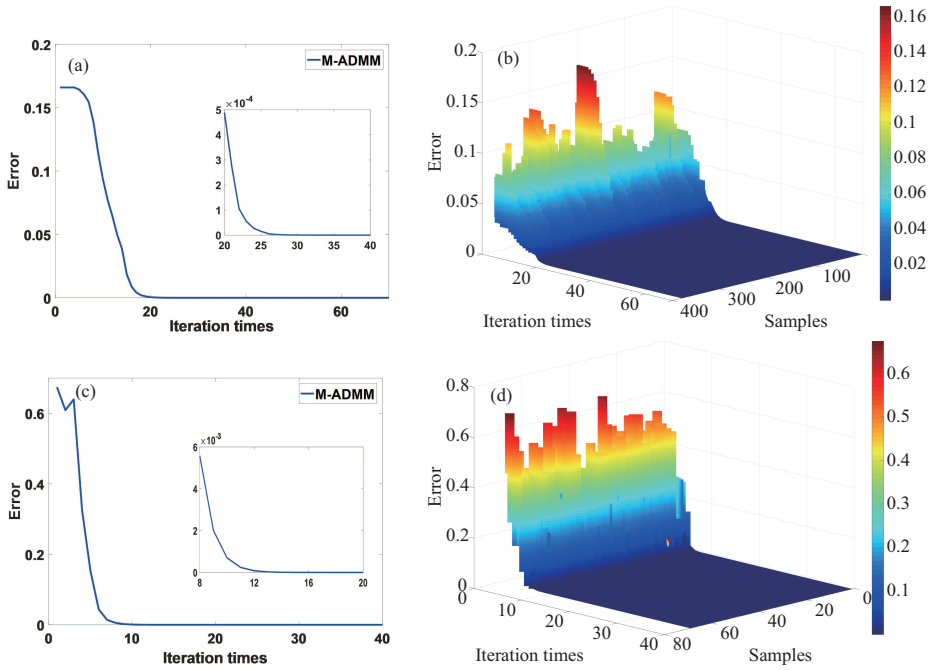

Figure 2 (Color online) (a), (b) Convergence analysis of the $S_{1/2}$-$L_{1/2}$-PFSR; (c), (d) convergence analysis of the $L_{1/2}$-ISR.

5.1 Databases

Face databases: the Extended Yale B database contains 38 subjects and each individual has about 64 images with various laboratory-controlled lighting conditions. All images are manually cropped and resized to be 32×32 pixels in experiments. The CMU PIE database contains 68 subjects with different poses, different illuminations, and different expressions. All experiments are conducted on a near frontal pose subset, namely C07, which contains 68 subjects and each individual has about 24 images. All images are manually cropped and resized to be 64×64 pixels in experiments. The AR database contains 126 subjects and each individual has about 26 images which are composed of 2 sections. There are 3 images with sunglasses, 3 images with scarves, 3 images with expressions, 3 images with illuminations, and 1 natural image for each section. A subset containing 50 male subjects and 50 female subjects is chosen for experiments.

Tumor databases: the DLBCL database contains diffuse large B cell lymphoma and follicular lymphoma, including 77 samples. Each sample has 5469 genes. The Leukemia database contains acute myelogenous leukemia, acute lymphoblastic leukemia, and mixed-lineage leukemia, including 72 samples. Each sample has 11225 genes. The 9_Tumors database contains 9 tumor types, such as NSCLC, colon, and breast, including 60 samples. Each sample has 5726 genes.

5.2 Convergence analysis

In Subsection 3.3, $S_{1/2}$-$L_{1/2}$-PFSR model is optimized. Here, the corresponding convergence is verified by experiments. Figures 2(a) and (b) show the iterative error of $S_{1/2}$-$L_{1/2}$-PFSR on Extend Yale B database. As shown in Figure 2(a), the iterative error curve shows a downward trend and tends to 0. In Figure 2(b), the iterative error trend goes from red to blue expresses the error goes from high to low, and the overall error trend eventually tends to 0. All these verify the convergence and effectiveness of optimization.

Figures 2(c) and (d) show the iterative error of $L_{1/2}$-ISR on the DLBCL database. Figure 2(c) is the iterative error curve, and Figure 2(d) is the iterative error trend graph. It can be seen from Figures 2(c) and (d) that the iterative error decreases and tends to 0 with the increase of iterations. The solution gradually becomes stable and converges to the numerical solution, which verifies that the optimization is convergent.

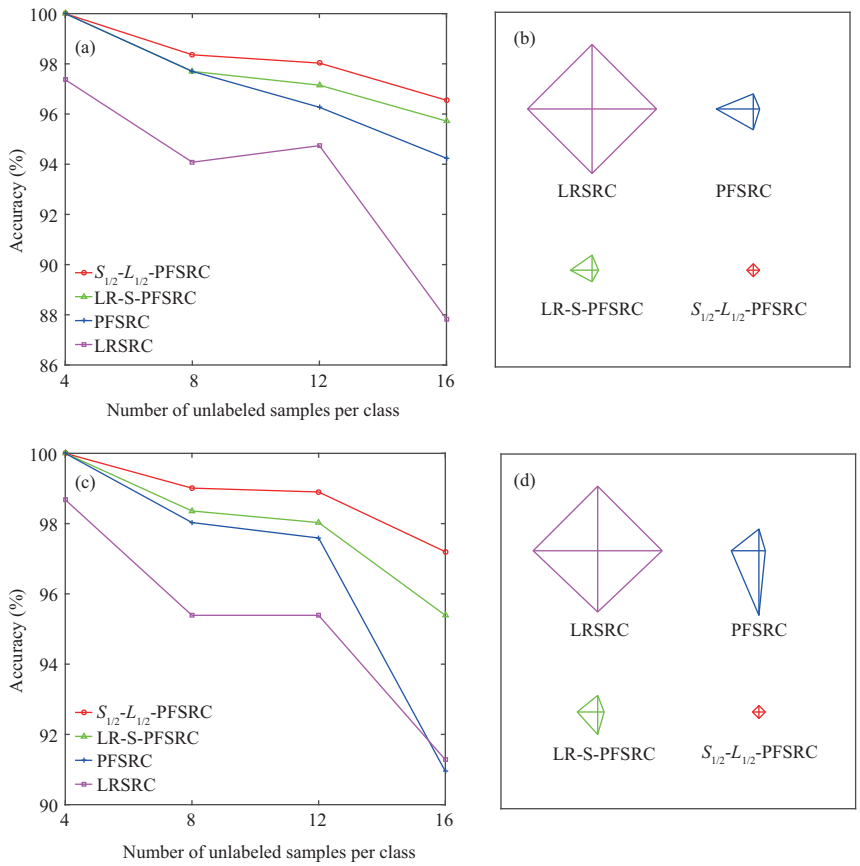

Figure 3 (Color online) Comparison of accuracy and rose figure of RSI with increasing number of unlabeled samples. (a) and (b) are 6 labeled samples for each class; (c) and (d) are 8 labeled samples for each class.

5.3 Role of unlabeled data

In order to illustrate the role of unlabeled data, for $S_{1/2}$-$L_{1/2}$-PFSRC, we fix labeled samples of each class with increasing unlabeled samples of each class on Extend Yale B database, as well as we select unobstructed images as labeled samples and occluded images as unlabeled samples on AR database. For $L_{1/2}$-ISRC, on DLBCL and Leukemia databases, about 25% of each class are selected as labeled data, and the remaining about 75% are selected as unlabeled data. We select the classification accuracy and relative stability indicator (RSI) [24] as evaluation indicators.

For $S_{1/2}$-$L_{1/2}$-PFSRC, on the Extended Yale B database, the labeled samples are fixed to 6 and 8 for each class, and the number of unlabeled samples for each class is 4, 8, 12, and 16, respectively. The feature dimension is set to 200 dimensions. Figure 3 shows the classification accuracy and rose figure of RSI as the number of unlabeled samples gradually increases. As can be seen from Figures 3(a) and (c), with the increase of unlabeled samples, classification accuracy curves of all methods show a downward trend. Compared to the positive projection method LRSRC, the classification accuracy of inverse projection methods (LR-S-PFSRC and $S_{1/2}$-$L_{1/2}$-PFSRC) are higher, which means unlabeled data play the role of auxiliary representation to help classification. $S_{1/2}$-$L_{1/2}$-PFSRC outperforms PFSRC due to low rank and sparse constraints. In addition, $S_{1/2}$-$L_{1/2}$-PFSRC is better than LR-S-PFSRC because the $S_{1/2}$ regularization and the $L_{1/2}$ regularization can further induce data sparsity. For RSI, the more concentrated the rose figure is, the more stable the classification stability of the method is. According to Figures 3(b) and (d), the classification stability of $S_{1/2}$-$L_{1/2}$-PFSRC is the best among all four methods.

In order to further verify that $S_{1/2}$-$L_{1/2}$-PFSRC can make full use of information between unlabeled samples, on the AR database, 3 images are randomly selected from 14 unobstructed images of each class as labeled samples, and the remaining 12 occluded images of each class are taken as unlabeled samples. The feature dimension is set to 200. The classification accuracy of LRSRC is 35.13%, PFSRC is 74.17%, LR-S-PFSRC is 76.20%, and $S_{1/2}$-$L_{1/2}$-PFSRC is 77.97%. It can be seen that $S_{1/2}$-$L_{1/2}$-PFSRC achieves competitive results with the highest classification accuracy among all four methods. Then, 2 natural

Table 1 Comparison of classification accuracy on AR database (%)

	LRSRC	PFSRC	LR-S-PFSRC	$S_{1/2}$-$L_{1/2}$-PFSRC
Sunglasses	32.83	73.00	83.33	**86.83**
Scarves	31.67	67.17	74.33	**76.83**

Table 2 Comparison of classification accuracy on tumor databases (%)

	SRC	IPRC	ISSRC	$L_{1/2}$-ISRC
DLBCL	81.08	84.28	85.45	**91.80**
Leukemia	75.80	80.22	77.49	**89.11**

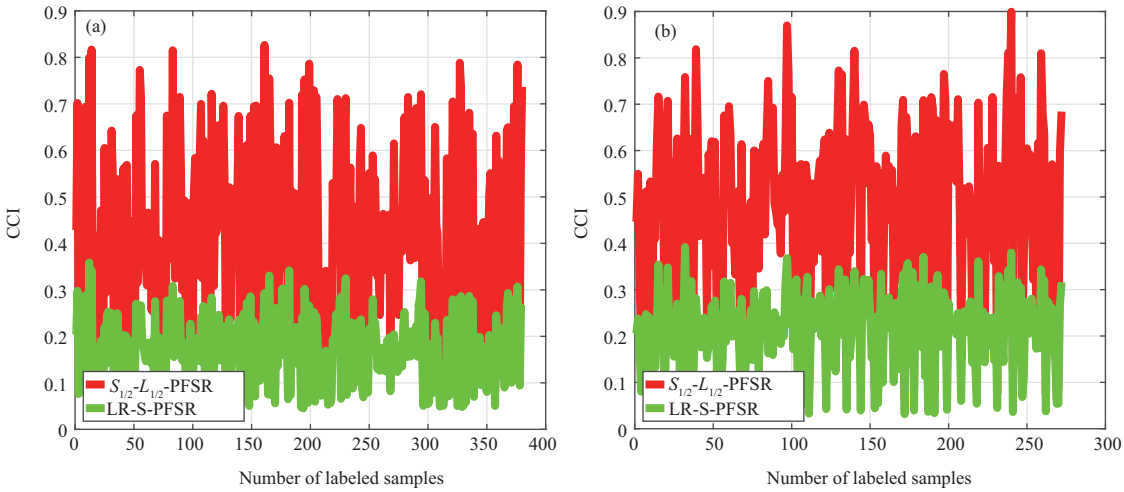

Figure 4 (Color online) Comparison of CCI values among LR-S-PFSR (green curve) and $S_{1/2}$-$L_{1/2}$-PFSR (red curve). (a) Extended Yale B database; (b) CMU PIE database.

images of each class are regarded as labeled samples, as well as 6 sunglasses images of each class and 6 scarf images of each class are regarded as unlabeled samples, respectively. The feature dimension is set to 150. As can be seen from Table 1, among these two different types of unlabeled samples, $S_{1/2}$-$L_{1/2}$-PFSRC have achieved better recognition results. For occluded images that are difficult to recognize, the proposed method can still achieve ideal classification performance, and its recognition results achieve significant improvement compared with the traditional positive projection method LRSRC.

For $L_{1/2}$-ISRC, on the DLBCL database, 14 samples of one class are randomly selected as labeled samples and the remaining 44 are unlabeled samples. 5 samples of another class are randomly selected as labeled samples and the remaining 14 are unlabeled samples. On the Leukemia database, each class of data is randomly divided into four parts, where one of them is selected as labeled samples and the other three parts are unlabeled samples. One can see that no matter which database, the number of unlabeled samples is about three times that of labeled samples. For each experiment, we independently run all the methods 300 times and take the means as the final results. Experimental results are shown in Table 2. It can be seen from Table 2 that the inverse projection methods (IPRC, ISSRC, and $L_{1/2}$-ISRC) driven by unlabeled data are better than SRC even though there are few effective labeled samples. Moreover, $L_{1/2}$-ISRC which applies the $L_{1/2}$ regularization outperforms both IPRC and ISSRC.

5.4 Sparsity analysis

For the sparsity of $S_{1/2}$-$L_{1/2}$-PFSR, the sparsity analysis is carried out by "sparseness" and category concentration index (CCI) [23]. For sparse representation model, the sparseness is defined as the number of non-zero elements in sparse representation coefficients of the same scale. The value range of CCI is $(0, 1]$. The larger the CCI is, the better the sparse representation is. The closer the CCI is to 1, the more concentrated the coefficients are in a certain class.

For face databases, since $S_{1/2}$-$L_{1/2}$-PFSR has sparse and low-rank constraints in representation coefficients, the sparsity can only be measured by CCI. Figure 4 shows the CCI values corresponding to LR-S-PFSR (green curve) and $S_{1/2}$-$L_{1/2}$-PFSR (red curve), respectively. From Figure 4, $\text{CCI}_{S_{1/2}\text{-}L_{1/2}\text{-PFSR}} >$

Table 3 Comparison of sparseness on tumor databases

	DLBCL	Leukemia	9_Tumors
ISSR	533.35	466.21	574.23
$L_{1/2}$-ISR	**289.80**	**261.88**	**232.47**

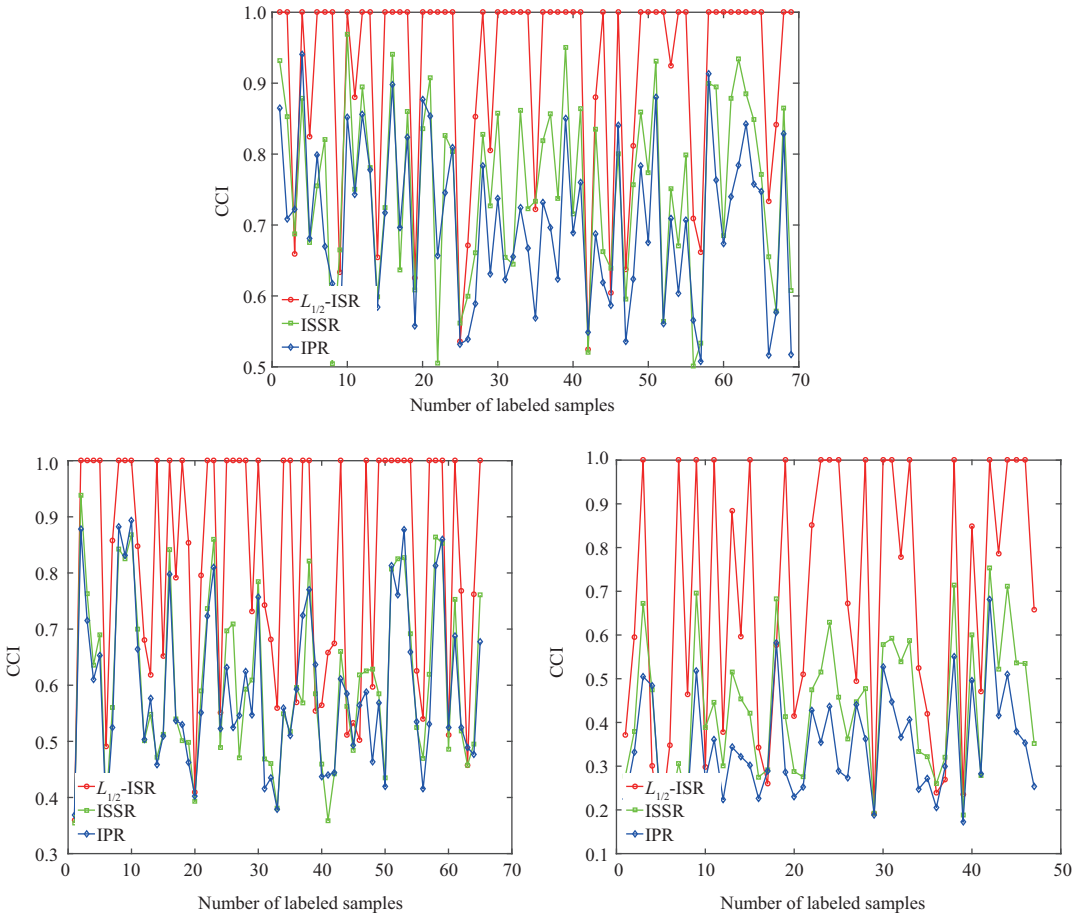

Figure 5 (Color online) Comparison of CCI values among IPR (blue curve), ISSR (green curve), and $L_{1/2}$-ISR (red curve). (a) DLBCL database; (b) Leukemia database; (c) 9_Tumor database.

$\text{CCI}_{\text{LR-S-PFSR}}$ is satisfied for most of samples. Experimental results show that $S_{1/2}$-$L_{1/2}$-PFSR has better sparse representation performance than LR-S-PFSR, because the representation coefficients induced by the $S_{1/2}$ norm are more low rank than the nuclear norm and the $L_{1/2}$ norm can further induce the sparsity of matrix.

For tumor databases, the sparsity of $L_{1/2}$-ISR is verified by comparing the sparseness and CCI. Table 3 shows the sparseness of ISSR and $L_{1/2}$-ISR. It can be seen from Table 3 that $L_{1/2}$-ISR has a higher sparseness than IPR and ISSR. Experimental results show that $L_{1/2}$-ISR selects the unlabeled samples that have high correlation with the represented object to represent, and improves the representation ability of the model. Figure 5 shows $L_{1/2}$-ISR has the largest CII values for most samples, which shows that $L_{1/2}$-ISR has better sparse representation performance than others. All these demonstrate the $L_{1/2}$ regularization has promising properties, such as sparsity, unbiasedness, and Oracle properties.

5.5 Effectiveness analysis of alleviating class imbalance problem

In order to verify the effectiveness of $S_{1/2}$-$L_{1/2}$-PFSRC-CS for the class imbalance problem, we first determine the optimal misclassification cost by experiments, and then we compare the classification performance of the proposed method with other methods when facing class imbalance problem.

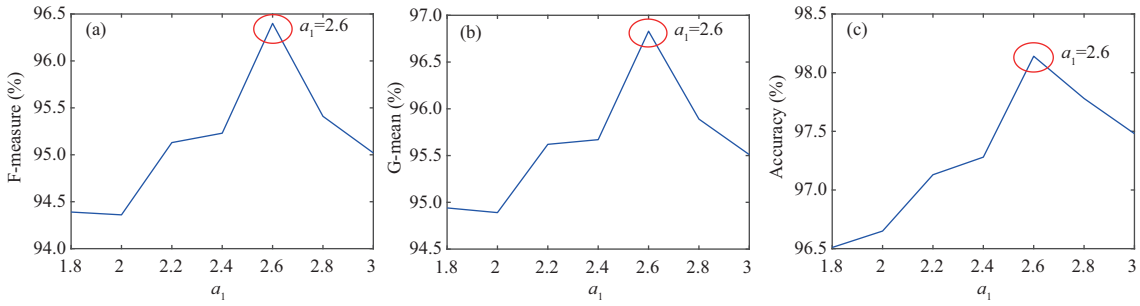

Figure 6 (Color online) (a) F-measure, (b) G-mean, and (c) accuracy of $a_1 \in [1.8, 3]$ on DLBCL database.

5.5.1 *Optimal misclassification cost analysis*

For face databases, since the number of samples for each class on Extended Yale B database and CMU Multi-PIE database is balanced, there is no class imbalance problem. Therefore, we consider the special cost matrix, that is, the diagonal elements are 0 and the remaining elements are 1. Furthermore, $S_{1/2}$-$L_{1/2}$-PFSRC-CS is equivalent to $S_{1/2}$-$L_{1/2}$-PFSRC. Therefore, $S_{1/2}$-$L_{1/2}$-PFSRC is used instead of $S_{1/2}$-$L_{1/2}$-PFSRC-CS in subsequent experiments.

For tumor databases, the DLBCL database contains 2 tumor types, where the size of the majority class and minority class is 58 and 19, respectively. Since the cost matrix of binary class multiplied by a number has the optimal decision invariance, one cost can be set to 1 and the other cost is greater than 1. The Leukemia database contains three tumor types, where the size of each class is 28, 24, and 20, respectively. Due to the class imbalance, we consider different misclassification costs. Assume the cost matrix P_{DLBCL} and P_{Leukemia} are as follows. Here, the a_1, b_1, b_2, and b_3 are determined by experiments.

$$P_{\text{DLBCL}} = \{P_{u,v}\}_{\text{DLBCL}} = \begin{bmatrix} 0 & a_1 \\ 1 & 0 \end{bmatrix}, \quad P_{\text{Leukemia}} = \{P_{u,v}\}_{\text{Leukemia}} = \begin{bmatrix} 0 & b_1 & b_2 \\ 1 & 0 & b_3 \\ 1 & 1 & 0 \end{bmatrix}.$$

Figure 6 shows the F-measure, G-mean, and classification accuracy of $a_1 \in [1.8, 3]$ on DLBCL database, respectively. It can be seen from Figure 6 that when a_1 takes a value in a small neighborhood centered on 2.6, $L_{1/2}$-ISRC-CS achieves competitive classification results. So, for simplicity, let $a_1 = 2.6$ in subsequent experiments. Figure 7 shows the classification accuracy of b_2, $b_3 \in [1, 2.1]$ as $b_1 = 1.2, 1.5, 1.8$ on Leukemia database. One can see that no matter whether b_1 takes any value of 1.2, 1.5 or 1.8, the classification accuracy of $L_{1/2}$-ISRC-CS is competitive when $b_2 = 1.8$, $b_3 = 1.2$. So, for simplicity, let $b_1 = 1.5$, $b_2 = 1.8$, $b_3 = 1.2$ in subsequent experiments. Since the 9_Tumors database contains 9 tumor types, the cost matrix P_{9_Tumor} is determined heuristically by referring to classification accuracy and experiment experience.

5.5.2 *Classification performance analysis*

In order to verify the classification performance of $S_{1/2}$-$L_{1/2}$-PFSRC on multiclass balance face databases, we compare $S_{1/2}$-$L_{1/2}$-PFSRC with LRSRC, PFSRC, and LR-S-PFSRC, as well as selecting the classification accuracy and RSI as evaluation indicators. In order to verify the classification performance of $L_{1/2}$-ISRC-CS on multiclass imbalance tumor databases, we compare $L_{1/2}$-ISRC-CS with IPRC, ISSRC, and $L_{1/2}$-ISRC, as well as selecting F-measure, G-mean, classification accuracy, error rate, and error reduction rate as evaluation indicators.

(a) Multiclass balance face databases. On multiclass balance face databases, for the Extended Yale B database, 50 face images of each class are randomly selected as samples. For the CMU PIE database, 20 face images of each class are randomly selected as samples. For each face database, 20% of the samples are regarded as labeled samples, and the remaining 80% are regarded as unlabeled samples.

Figure 8 shows the classification accuracy and rose figure of RSI on Extended Yale B database and CMU PIE database. Figures 8(a) and (c) are classification accuracy curves. Figures 8(b) and (d) are the rose figures of RSI. In Figures 8(a) and (c), the higher the classification accuracy curve is, the better the method is. In Figures 8(b) and (d), the more concentrated the rose figure of RSI is, the stronger the classification stability of the method is. As can be seen from Figure 8, with the increase of feature

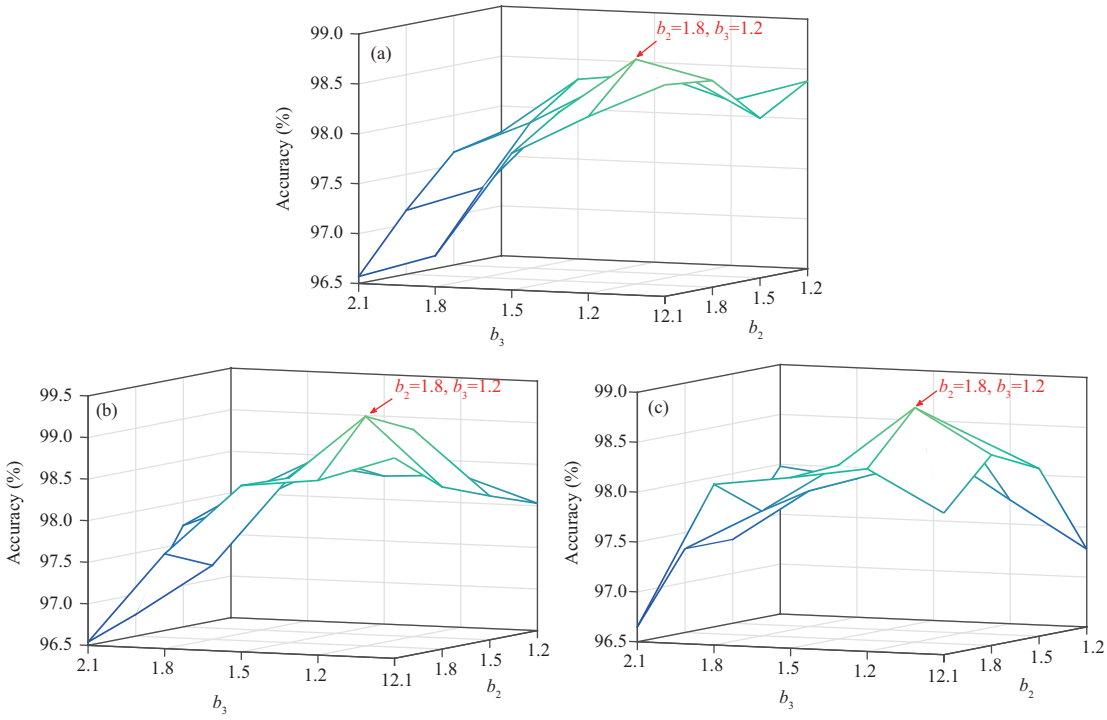

Figure 7 (Color online) Accuracy of b_2, $b_3 \in [1, 2.1]$ on Leukemia database as (a) $b_1 = 1.2$, (b) $b_1 = 1.5$, and (c) $b_1 = 1.8$.

dimension, the classification accuracy of all methods is improved. $S_{1/2}$-$L_{1/2}$-PFSRC outperforms all other methods in a different dimension. Meanwhile, for dimension changes, the rose figure of RSI of $S_{1/2}$-$L_{1/2}$-PFSRC is more concentrated than the other three methods, which indicates that $S_{1/2}$-$L_{1/2}$-PFSRC is less sensitive to selection of feature dimension. One can see that $S_{1/2}$-$L_{1/2}$-PFSRC with about 200 dimensions tends to be stable in Extended Yale B and CMU Multi-PIE databases. So, we use the feature of dimensionality 200 to evaluate the running time of $S_{1/2}$-$L_{1/2}$-PFSRC, LR-S-PFSRC, and LRSRC. Table 4 provides the average running time of each tested method. As a result, $S_{1/2}$-$L_{1/2}$-PFSRC achieves better classification results by spending less running time.

(b) Binary class imbalance tumor database. For binary class imbalance tumor database, Figures 9(a) and (b) show F-measure, G-mean, classification accuracy, and box plots for an error rate of different methods on DLBCL database. It can be seen from Figure 9(a) that the F-measure and G-mean of IPRC, ISSRC, and $L_{1/2}$-ISRC have no significant changes, which indicates that these three methods have similar classification performance for the class imbalance problem. The F-measure of $L_{1/2}$-ISRC is slightly higher than that of IPRC and ISSRC because the $L_{1/2}$ regularization can induce more sparse representation coefficients to make the representation more discriminative. Experimental results show that only using the $L_{1/2}$ regularization cannot alleviate the class imbalance problem. The F-measure of $L_{1/2}$-ISRC-CS is obviously superior to others, which indicates that $L_{1/2}$-ISRC-CS significantly improves the classification accuracy of the minority class. The reason is that $L_{1/2}$-ISRC-CS imposes different costs on different misclassifications, which can alleviate the class imbalance problem. Moreover, $L_{1/2}$-ISRC-CS achieves the largest G-mean among all four methods, which further shows that $L_{1/2}$-ISRC-CS is effective for the class imbalance problem. Finally, the classification accuracy of $L_{1/2}$-ISRC-CS also outperforms the other methods from Figure 9(a). Figure 9(b) shows the box plots for the error rate of different methods on the DLBCL database. It can be seen from Figure 9(b) that the median line of $L_{1/2}$-ISRC-CS is the lowest median among all four methods. And the distance between the upper quartile and the lower quartile of box plots shows that the error rate distribution of $L_{1/2}$-ISRC-CS is relatively concentrated. As shown in Figures 9(a) and (b), all the four evaluation indicators indicate that $L_{1/2}$-ISRC-CS is better than others.

In order to compare IPRC, ISSRC, $L_{1/2}$-ISRC, and $L_{1/2}$-ISRC-CS more intuitively, error rate and error reduction rate are chosen to compare classification performance. According to the error reduction rate (ERR), $\mathrm{ERR} = (\mathrm{ER}_1 - \mathrm{ER}_2)/\mathrm{ER}_1 \times 100\%$, ER_1 is the error rate of other recognition results on the same method, ER_2 is the error rate of highest recognition result on the same method, and ERR is represented

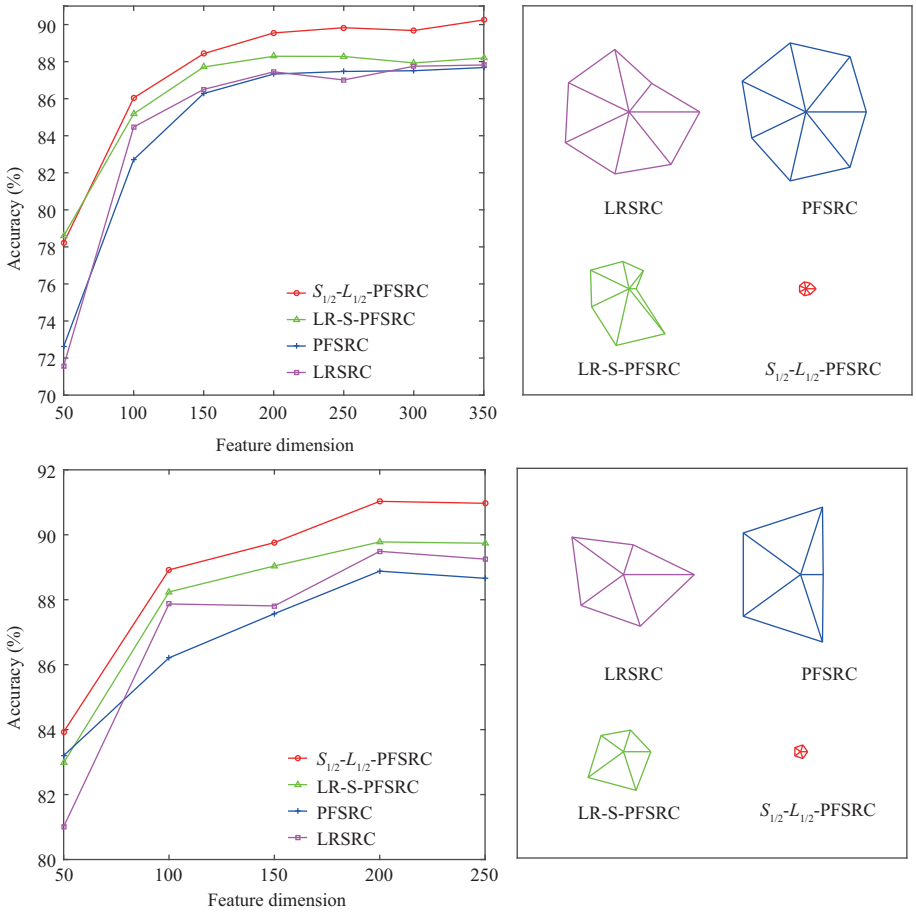

Figure 8 (Color online) Accuracy and rose figure of RSI versus feature dimension of all the methods. (a) and (b) are Extended Yale B database; (c) and (d) are CMU PIE database.

Table 4 The average running time of different methods on face databases (s)

	LRSRC	LR-S-PFSRC	$S_{1/2}$-$L_{1/2}$-PFSRC
Extended Yale B	70.32	15.40	**9.28**
CMU PIE	34.70	7.44	**4.53**

by ↓. Table 5 shows the error rate and ERR of all methods on DLBCL database. It can be seen from Table 5 that $L_{1/2}$-ISRC-CS has the lowest error rate among the all four methods, and the ERR obviously reflects the relative advantages of $L_{1/2}$-ISRC-CS.

(c) Multiclass imbalance tumor databases. For the multiclass imbalance tumor databases, Leukemia and 9_Tumors databases, the classification performance of IPRC, ISSRC, $L_{1/2}$-ISRC, and $L_{1/2}$-ISRC-CS is compared. Table 6 shows the classification accuracy, error rate, and ERR of all methods. And Figures 9(c) and (d) show box plots for the error rate of the four methods. It can be seen from Table 6 that the classification accuracy of $L_{1/2}$-ISRC-CS is significantly higher than others on multiclass imbalance tumor databases. $L_{1/2}$-ISRC-CS still has the lowest error rate among all four methods on multiclass imbalance tumor databases. Also, the ERR reflects the obvious advantages of $L_{1/2}$-ISRC-CS. It can be seen from Figures 9(c) and (d) that box plots for the error rate of $L_{1/2}$-ISRC-CS has the lowest median line and the shortest distance between the upper quartile and the lower quartile of box plots among all four methods. From Table 6 and Figures 9(c) and (d), it can be concluded that the classification performance of $L_{1/2}$-ISRC-CS is still better than others on multiclass imbalance tumor databases.

5.6 Comparison with state-of-the-art methods

The proposed method is further compared with the latest published results and some deep-learning algorithms. For $S_{1/2}$-$L_{1/2}$-PFSRC, on the AR database, we select nautral image of each subject as

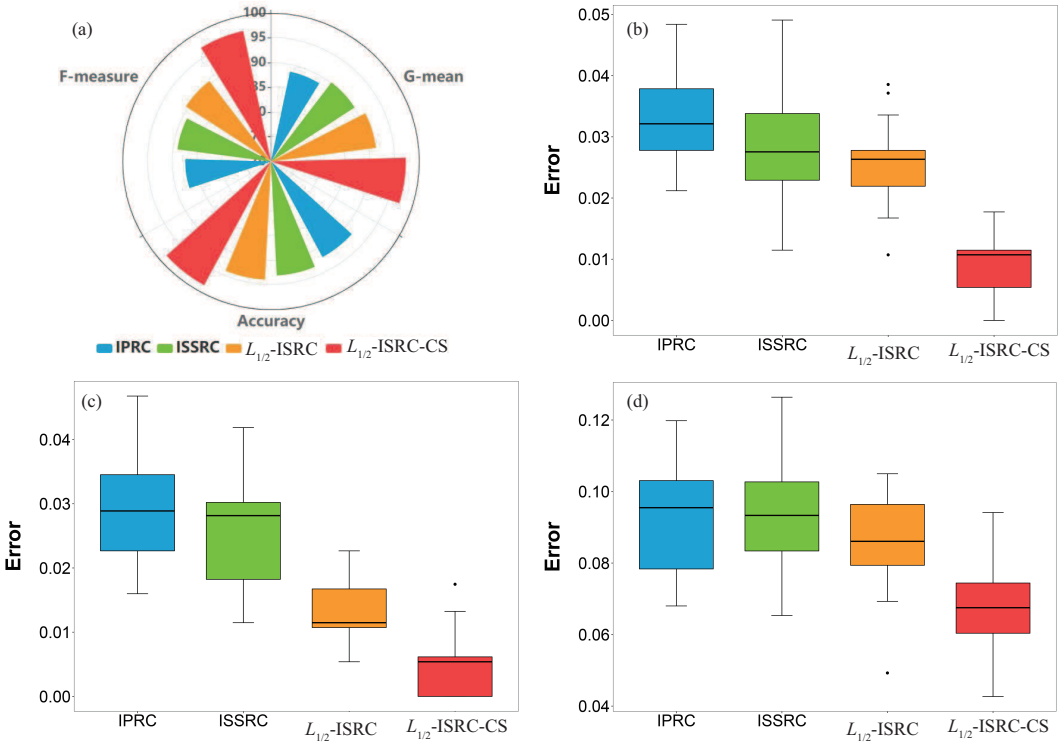

Figure 9 (Color online) (a) Comparison of F-measure, G-mean, accuracy and (b) box plots for error rate on DLBCL database; (c) box plots for error rate on Leukemia database and (d) 9_Tumors database.

Table 5 Comparison of different classification methods on DLBCL database (%)

	IPRC	ISSRC	$L_{1/2}$-ISRC	$L_{1/2}$-ISRC-CS
Error rate	8.06	6.93	6.08	**1.56**
ERR	↓80.67	↓77.51	↓74.38	–

Table 6 Comparison of different classification methods on tumor databases (%)

		IPRC	ISSRC	$L_{1/2}$-ISRC	$L_{1/2}$-ISRC-CS
Leukemia	Accuracy	93.96	93.95	97.42	**99.01**
	Error rate	6.04	6.05	2.58	**0.99**
	ERR	↓83.6	↓83.64	↓61.59	–
9_Tumors	Accuracy	76.04	76.19	79.46	**85.20**
	Error rate	23.96	23.81	20.54	**14.80**
	ERR	↓38.24	↓37.85	↓27.94	–

labeled data, and images with expression changes and illumination changes are chosen as unlabeled data respectively. Table 7 gives the classification results of different methods, including LBP [42], P-LBP [43], and PCANet [44], where our algorithm achieves competitive results when facing different testing conditions. As we know, PCANet is served as a simple but highly competitive deep learning baseline for object recognition. Because of deep learning, PCANet has higher accuracies than both LBP and P-LBP. However, compared to PCANet, $S_{1/2}$-$L_{1/2}$-PFSRC is on par in classification performance but with no learning. In addition, we also compare the proposed method with FDDL [45], JEDL [46], and ADDL [47]. 20 images per class are randomly chosen as labeled samples and the rest are unlabeled samples. The classification results are averaged more than 10 times runs. Table 8 shows experimental results, where one can observe that $S_{1/2}$-$L_{1/2}$-PFSRC is superior to other comparison methods.

For $L_{1/2}$-ISRC-CS, its performance is compared with the latest published results on DLBCL, Leukemia, and 9_Tumors databases. The compared methods are selected from [19–22,48–50], where MSRC-SVD [19], MSRC-NMF [19], and MSRC-SNMF [19] are all based on MSRC and their difference is the way that they extract the metasamples of gene expression data using SVD, NMF, and SNMF, respectively. The classification results are exhibited in Table 9 [19–22, 48–51]. As can be seen from Table 9, $L_{1/2}$-ISRC-CS

Table 7 Classification accuracy of different methods on AR database (%)

	LBP [42]	P-LBP [43]	PCANet-1 [44]	PCANet-2 [44]	$S_{1/2}$-$L_{1/2}$-PFSRC
Expression	81.33	80.33	85.67	85.00	**87.33**
Illumination	93.83	97.50	98.00	99.50	**99.67**

Table 8 Classification accuracy of different methods on AR database (%)

	FDDL [45]	JEDL [46]	ADDL [47]	$S_{1/2}$-$L_{1/2}$-PFSRC
Accuracy	95.60	96.20	97.00	**97.50**

Table 9 Classification accuracy with the latest published results on tumor databases (%)

DLBCL		Leukemia		9_Tumors	
Method	Accuracy	Method	Accuracy	Method	Accuracy
MLP-D [48]	96.24	MSRC-NMF [19]	95.83	MRSRC-SVD [20]	60.00
MSRC-SNMF [19]	97.40	IPRC [21]	96.90	MSRC-SVD [19]	63.33
SRC-LatLRR [49]	97.40	MRSRC-SVD [20]	97.22	SRC-LatLRR [49]	66.67
Integrated ISSRC [22]	97.50	SRC-LatLRR [49]	98.61	IPRC [21]	66.67
IPGSRC [50]	98.00	Integrated ISSRC [22]	98.61	LLR+SR [51]	66.75
$L_{1/2}$-ISRC-CS	**98.44**	$L_{1/2}$-ISRC-CS	**99.01**	$L_{1/2}$-ISRC-CS	**85.20**

leads to competitive results in identifying different types of tumors.

6 Conclusion

In this paper, the cost-sensitive inverse projection sparse representation-based classification with 1/2 regularization is proposed based on the inherent characteristics of data. The proposed method obtains the sparsity of 1-D and 2-D data by introducing the $L_{1/2}$ regularization and the $S_{1/2}$ regularization. The cost matrix is incorporated into the decision criterion to improve classification performance. Moreover, the proposed method is effective and stable even if there are few labeled samples or class imbalance. There remain some interesting questions. How to further optimize the model, such as integrating feature coding and discriminative classification into one model similar to [52], combining between the interclass and intraclass properties of dictionary atoms and coding coefficients similar to [53], or combining with deep learning methods when encountering big data or wild datasets because of limitations of SRC type methods.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 41771375), Open Fund of Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education (Grant No. IPIU2019010), and Natural Science Foundations of Henan Province (Grant Nos. 202102310087, 222300420417).

References

1 Yang Q, Wu X D. 10 challenging problems in data mining research. Int J Info Tech Dec Mak, 2006, 05: 597–604
2 He H B, Garcia E A. Learning from imbalanced data. IEEE Trans Knowl Data Eng, 2009, 21: 1263–1284
3 Liu X Y, Wu J X, Zhou Z H. Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern B, 2009, 39: 539–550
4 Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol, 2002, 3: 1–21
5 Batista G E A P A, Prati R C, Monard M C. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl, 2004, 6: 20–29
6 Liu J F, Hu Q H, Yu D. A weighted rough set based method developed for class imbalance learning. Inf Sci, 2008, 178: 1235–1256
7 Wang Y, Hu Q H, Zhou Y C, et al. Local bayes risk minimization based stopping strategy for hierarchical classification. In: Proceedings of IEEE International Conference on Data Mining, New Orleans, 2017. 515–524
8 Zangeneh V, Shajari M. A cost-sensitive move selection strategy for moving target defense. Comput Secur, 2018, 75: 72–91
9 Zhao P L, Zhang Y F, Wu M, et al. Adaptive cost-sensitive online classification. IEEE Trans Knowl Data Eng, 2019, 31: 214–228
10 Khan S H, Hayat M, Bennamoun M, et al. Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans Neural Netw Learn Syst, 2018, 29: 3573–3587
11 Chung Y A, Lin H T, Yang S W. Cost-aware pre-training for multiclass cost-sensitive deep learning. 2015. ArXiv:1511.09337
12 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, 2017. 1126–1135
13 Munkhdalai T, Yu H. Meta networks. In: Proceedings of the 34th International Conference on Machine Learning, 2017. 2554–2563

14 Schwartz E, Karlinsky L, Shtok J, et al. Delta-encoder: an effective sample synthesis method for few-shot object recognition. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, 2018. 2850–2860
15 Kwitt R, Hegenbart S, Niethammer M. One-shot learning of scene locations via feature trajectory transfer. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 78–86
16 Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell, 2009, 31: 210–227
17 Deng W H, Hu J N, Guo J. Extended SRC: undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell, 2012, 34: 1864–1870
18 Du H S, Hu Q P, Qiao D F, et al. Robust face recognition via low-rank sparse representation-based classification. Int J Autom Comput, 2015, 12: 579–587
19 Zheng C H, Zhang L, Ng T Y, et al. Metasample-based sparse representation for tumor classification. IEEE ACM Trans Comput Biol Bioinf, 2011, 8: 1273–1282
20 Gan B, Zheng C H, Liu J X. Metasample-based robust sparse representation for tumor classification. ENG, 2013, 05: 78–83
21 Yang X H, Tian L, Chen Y M, et al. Inverse projection representation and category contribution rate for robust tumor recognition. IEEE ACM Trans Comput Biol Bioinf, 2020, 17: 1262–1275
22 Yang X H, Wu W M, Chen Y M, et al. An integrated inverse space sparse representation framework for tumor classification. Pattern Recogn, 2019, 93: 293–311
23 Yang X H, Liu F, Tian L, et al. Pseudo-full-space representation based classification for robust face recognition. Signal Process-Image Commun, 2018, 60: 64–78
24 Yang X H, Wang Z, Wu H, et al. Stable and compact face recognition via unlabeled data driven sparse representation-based classification. 2021. arXiv:2111.02847
25 Xu Z B, Zhang H, Wang Y, et al. $L_{1/2}$ regularization. Sci China Inf Sci, 2010, 53: 1159–1169
26 Fazel M. Matrix rank minimization with applications. Dissertation for Ph.D. Degree. Stanford: Stanford University, 2002
27 Rao G, Peng Y, Xu Z B. Robust sparse and low-rank matrix decomposition based on $S_{1/2}$ modeling (in Chinese). Sci Sin Inform, 2013, 43: 733–748
28 Candés E J, Recht B. Exact matrix completion via convex optimization. Found Comput Math, 2009, 9: 717–772
29 Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev, 2010, 52: 471–501
30 Fan J Q, Peng H. Nonconcave penalized likelihood with a diverging number of parameters. Ann Statist, 2004, 32: 928–961
31 Fu W J, Knight K. Asymptotics for lasso-type estimators. Ann Statist, 2000, 28: 1356–1378
32 Xu Z B, Chang X Y, Xu F M, et al. $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. IEEE Trans Neural Netw Learn Syst, 2012, 23: 1013–1027
33 Lu C, Feng J S, Yan S C, et al. A unified alternating direction method of multipliers by majorization minimization. IEEE Trans Pattern Anal Mach Intell, 2018, 40: 527–541
34 Elkan C. The foundations of cost-sensitive learning. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence, 2001. 973–978
35 Georghiades A S, Belhumeur P N, Kriegman D J. From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Machine Intell, 2001, 23: 643–660
36 Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression database. IEEE Trans Pattern Anal Machine Intell, 2003, 25: 1615–1618
37 Martinez A M, Benavente R. The AR Face Database. CVC Technical Report 24, Purdue University. 1998
38 Shipp M A, Ross K N, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med, 2002, 8: 68–74
39 Armstrong S A, Staunton J E, Silverman L B, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet, 2002, 30: 41–47
40 Staunton J E, Slonim D K, Coller H A, et al. Chemosensitivity prediction by transcriptional profiling. In: Proceedings of National Academy of Sciences of the United States of America, 2001. 10787–10792
41 Dudoit S, Fridlyand J, Speed T P. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc, 2002, 97: 77–87
42 Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell, 2006, 28: 2037–2041
43 Tan X Y, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process, 2010, 19: 1635–1650
44 Chan T H, Jia K, Gao S H, et al. PCANet: a simple deep learning baseline for image classification? IEEE Trans Image Process, 2015, 24: 5017–5032
45 Yang M, Zhang L, Feng X C, et al. Sparse representation based fisher discrimination dictionary learning for image classification. Int J Comput Vis, 2014, 109: 209–232
46 Zhang Z, Li F Z, Chow T W S, et al. Sparse codes auto-extractor for classification: a joint embedding and dictionary learning framework for representation. IEEE Trans Signal Process, 2016, 64: 3790–3805
47 Zhang Z, Jiang W M, Qin J, et al. Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier. IEEE Trans Neural Netw Learn Syst, 2018, 29: 3798–3814
48 García V, Salvador Sánchez J. Mapping microarray gene expression data into dissimilarity spaces for tumor classification. Inf Sci, 2015, 294: 362–375
49 Gan B, Zheng C H, Zhang J, et al. Sparse representation for tumor classification based on feature extraction using latent low-rank representation. Biomed Res Int, 2014, 2014: 1–7
50 Yang X H, Jiang X Y, Tian C X, et al. Inverse projection group sparse representation for tumor classification: a low rank variation dictionary approach. Knowl-Based Syst, 2020, 196: 105768
51 Kolali Khormuji M, Bazrafkan M. A novel sparse coding algorithm for classification of tumors based on gene expression data. Med Biol Eng Comput, 2016, 54: 869–876
52 Zhang Z, Li F Z, Zhao M B, et al. Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification. IEEE Trans Image Process, 2016, 25: 2429–2443
53 Li Z M, Zhang Z, Qin J, et al. Discriminative fisher embedding dictionary learning algorithm for object recognition. IEEE Trans Neural Netw Learn Syst, 2020, 31: 786–800