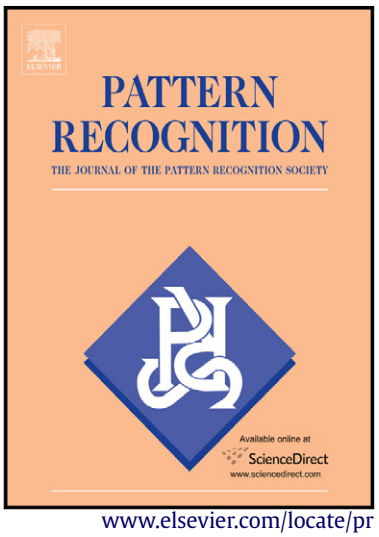


www.elsevier.com/locate/pr

Ensemble Manifold Regularized Sparse Low-Rank Approximation for Multiview Feature Embedding

Lefei Zhang, Qian Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang, Bo Du

PII: S0031-3203(14)00527-5
DOI: http://dx.doi.org/10.1016/j.patcog.2014.12.016
Reference: PR5306

To appear in: *Pattern Recognition*

Received date: 14 August 2014
Revised date: 15 November 2014
Accepted date: 20 December 2014

Cite this article as: Lefei Zhang, Qian Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang, Bo Du, Ensemble Manifold Regularized Sparse Low-Rank Approximation for Multiview Feature Embedding, *Pattern Recognition,* http://dx.doi.org/10.1016/j.patcog.2014.12.016

Ensemble Manifold Regularized Sparse Low-Rank Approximation for Multiview Feature Embedding

Lefei Zhang[a], Qian Zhang[b], Liangpei Zhang[c], Dacheng Tao[d], Xin Huang[b], Bo Du[a,*]

[a] *Computer School, Wuhan University, Wuhan 430072, China*
[b] *Beijing Samsung Telecom R&D Center, Beijing 100028, China*
[c] *State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China*
[d] *Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia*

Abstract

In computer vision and pattern recognition researches, the studied objects are often characterized by multiple feature representations with high dimensionality, thus it is essential to encode that multiview feature into a unified and discriminative embedding that is optimal for a given task. To address this challenge, this paper proposes an ensemble manifold regularized sparse low-rank approximation (EMR-SLRA) algorithm for multiview feature embedding. The EMR-SLRA algorithm is based on the framework of least-squares component analysis, in particular, the low dimensional feature representation and the projection matrix are obtained by the low-rank approximation of the concatenated multiview feature matrix. By considering the complementary property among multiple features, EMR-SLRA simultaneously enforces the ensemble manifold regularization on the output feature embedding. In order to further enhance its robustness against the noise, the group sparsity is introduced into the objective formulation to impose direct noise reduction on the input multiview feature matrix. Since there is no closed-form solution for EMR-SLRA, this paper provides an efficient EMR-SLRA optimization procedure to obtain the output feature embedding. Experiments on the pattern recognition applications confirm the

*Corresponding author
Email address: gunspace@163.com (Bo Du)

effectiveness of the EMR-SLRA algorithm compare with some other multiview feature dimensionality reduction approaches.

Keywords: multiview, feature extraction, low-rank matrix approximation, ensemble manifold regularization, group sparsity

1. Introduction

In many pattern recognition applications, e.g., human identification, image annotation, hyperspectral data classification, the studied objects are often represented by different views of features [1]. For example, a person can be identified by face (frontal or profile), palmprint and gait with information obtained from multiple sources, an image can be characterized by its color, shape and texture, and a pixel in the hyperspectral data can be represented by both spatial and spectral features. Each view of a feature summarizes a specific characteristic of the studied object from different feature spaces, and features for different views are complementary to one another [2]. Although such multiview feature provides more potential discriminative information to distinguish the patterns of different classes, the feature vector of each view is usually lies in a high dimensional feature space and thus combining them together frequently leads to the problem of "curse of dimensionality" [3, 4, 5]. In fact, feature embedding is critical for the final performance as it both reduces the required computational load as well as regularizes the learning problem onto a smaller subset of input features. In this case, there is a need for the feature dimensionality reduction (DR) technologies that can reduce the redundancy among features while preserve the discriminative information that is important for the subsequent given task (e.g., modeling, classification, visualization, and clustering).

In the past decades, a large family of DR methods have been proposed in the attempt to find an appropriate low dimensional subspace of the original high dimensional feature space [6, 7]. Feature DR can be generally categorized into supervised and unsupervised approaches. Practically, the supervised branch often requires a large amount of labeled data, which is an expensive and laborious

task and sometimes even infeasible. In contrast, the unlabeled data are cheap and easy to obtain because a large amount of them can be easily collected from the data set. Hence, this paper focuses on the unsupervised technique which automatically learns the multiview feature embedding without any priori discriminative information. The most representative unsupervised DR method is Principal Component Analysis (PCA) [8, 9], which maximizes the data variance in the projected linear subspace. Since the global linearity of the conventional DR algorithms prohibits their effectiveness for nonlinear distributed data, some manifold learning based DR algorithms have been proposed to overcome this issue, as pioneered by the works of Locally Linear Embedding (LLE) [10], ISOMAP [11], Laplacian Eigenmaps (LE) [12], Hessian eigenmaps (HLLE) [13], and Local Tangent Space Alignment (LTSA) [14]. All the aforementioned linear and nonlinear DR algorithms can be unified to a graph embedding framework [15] or a patch alignment framework [16]. In addition, the matrix factorization based methods are also applied to data feature extraction and DR, e.g., the Regularized Nonnegative Matrix Factorization [17], the Non-negative Patch Alignment Framework [18], and the Regularized Low-Rank Matrix Approximation [19].

The traditional solution for multiview feature embedding is to concatenate the feature vectors from different views together and then apply the feature DR algorithms directly on the concatenated vector. Since this strategy completely ignores the diversity of multiple features, researchers have suggested to consider different views in different ways to achieve a better feature embedding, such as Distributed Spectral Embedding (DSE) [20] and Multiview Spectral Embedding (MSE) [21]. Some more recently effort on the topic of multiview feature embedding including [22, 23, 24, 25, 26]. However, most of the existing multiview feature embedding algorithms share at least one of the following problems: (1) the multiview embedding of test samples can not be directly computed due to the feature mapping is nonlinear and implicit, (2) the complementary property of different views has not been optimally explored since each view actually has its specific statistical and physical meanings, and (3) the robustness of algorithm

is limited because it ignores the fact that the input multiview feature matrix is usually affected by noises from different views.

To address the issues illustrated above, in this paper, we propose an ensemble manifold regularized sparse low-rank approximation (EMR-SLRA) algorithm for multiview feature embedding, based on the framework of least-squares component analysis [4]. As illustrated in Fig. 1, the multiview feature is firstly concatenated into a matrix, the low dimensional feature representation and the projection matrix are obtained by the low-rank approximation of that multiview feature matrix. By considering the complementary property among the multiple features, EMR-SLRA simultaneously enforces the ensemble manifold regularization on the output feature embedding. In order to further enhance its robustness against the noise, the group sparsity is introduced into the objective formulation to impose direct noise reduction on the input multiview feature matrix. Furthermore, since there is no closed-form solution for EMR-SLRA, we propose an efficient EMR-SLRA optimization procedure to obtain the multiview feature embedding. In summary, the main advantages of EMR-SLRA algorithm for multiview feature embedding lie in the following:

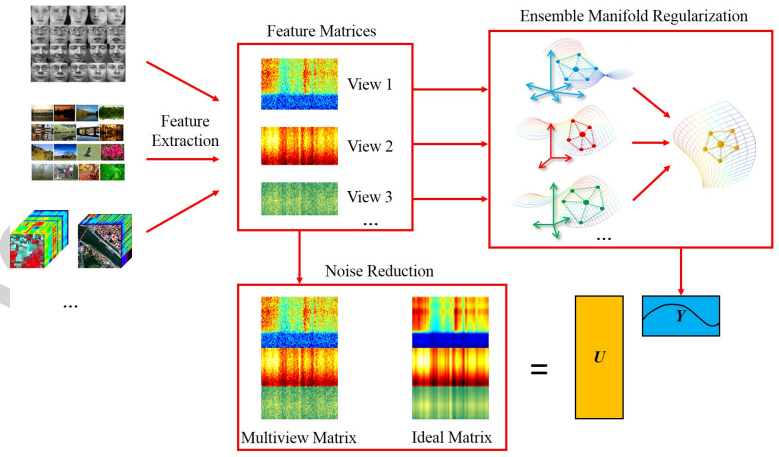


Figure 1: Flowchart of the EMR-SLRA algorithm for multiview feature embedding.

- The least-squares component analysis framework is generalized to the mul-

tiview version, by which the output feature representation of the original multiview data and the projection matrix can be simultaneously learned. Such low-rank approximation of the multiview feature matrix provides an straightforward approach to predict the embedding of test samples, thus the out-of-sample problem can be avoided.

- The ensemble manifold regularization (EMR), which considers the complementarity of different features by combining all the graphs in the multiple feature spaces together, is enforced on EMR-SLRA to look for a physically meaningful embedding that is discriminative for subsequent given task. It is worth noting that the coefficient for each view is automatically optimized adapt to its contribution to the data embedding.

- The group sparsity is introduced to control the error between the ideal and input multiview feature matrices. Although this $\ell_{2,1}$-norm fitting constraint is only defined with respect to the ideal matrix, the effect of noise reduction can be transferred to the projection matrix in the EMR-SLRA algorithm. By this consideration, the robustness of algorithm can be guaranteed even if some views have severely corrupted by noise.

The remainder of this paper is organized as follows. In Section 2, we provide the objective formulation of EMR-SLRA in detail. Section 3 proposes the efficient EMR-SLRA optimization procedure. Then, the experimental results of EMR-SLRA compared with some other multiview feature DR approaches are reported in Section 4, followed by the conclusion in Section 5.

2. Ensemble Manifold Regularized Sparse Low-Rank Approximation

This section presents the EMR-SLRA algorithm in detail, which can be divided into three parts, i.e., the multiview matrix low-rank approximation, the ensemble manifold regularization, and the group sparsity constraint.

2.1. Multiview Matrix Low-Rank Approximation

The proposed EMR-SLRA algorithm is based on the least-squares component analysis framework [4]. Suppose we have a data set $X = \{x_1, x_2, \cdots, x_N\}$ with N samples in the $\mathbb{R}^m$ dimensionality feature space (the data set has been pre-centralized, i.e., $\frac{1}{N}\sum_{i=1}^{N} x_i = 0$), PCA seeks the optimal linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^d$ ($d < m$) by which the data is decorrelated while the variance is maximized. In a least-squares point of view, denote the desires low dimensional feature representation as $Y = \{y_1, y_2, \cdots, y_N\}$, PCA minimizes the following reconstruction error by using the optimal orthogonal basis under the least-squares framework:

$$\varepsilon = \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{d} (x_i e_j) e_j \right\|^2, \tag{1}$$

in which $\{e_j\}_{j=1}^{d}$ is a subset of orthogonal basis of X in $\mathbb{R}^m$.

If define $P = [e_1, e_2, \cdots, e_d] \in \mathbb{R}^{m\times d}$ ($P^{\mathrm{T}}P = I$), thus we have $Y = P^{\mathrm{T}}X$. Then, Eq. (1) has its matrix formulation:

$$\begin{aligned} \varepsilon &= \sum_{i=1}^{N} \left\| x_i - P(P^{\mathrm{T}}x_i) \right\|^2, \\ \Rightarrow \varepsilon &= \left\| X - P(P^{\mathrm{T}}X) \right\|^2. \end{aligned} \tag{2}$$

Therefore, the objective of PCA can be rewritten as follows:

$$\begin{aligned} \arg\min_{P} \quad & \left\| X - P(P^{\mathrm{T}}X) \right\|^2, \\ \text{s.t.} \quad & P^{\mathrm{T}}P = I. \end{aligned} \tag{3}$$

Now we tune to the case of multiview feature embedding. Suppose we have observed V views for each sample, i.e., $x_i = \{x_i^{(1)} \in \mathbb{R}^{m_1}, \cdots, x_i^{(V)} \in \mathbb{R}^{m_V}\}$. For the consideration of multiview feature concatenation, the new feature of each sample is denoted by $x_i = [x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(V)}]^{\mathrm{T}} \in \mathbb{R}^m$, $m = \sum_{i=1}^{V} m_i$, and the multiview feature matrix is denoted by $X = [X^{(1)}, X^{(2)}, \cdots, X^{(V)}]^{\mathrm{T}} \in$

$\mathbb{R}^{m\times N}$. Followed by these definitions, the low-rank matrix approximation of the multiview data can be simply unified into Eq. (4):

$$\begin{aligned} &\arg\min_{U,Y} \ \|X-UY\|^2, \\ &\text{s.t. } U^{\mathrm{T}}U=I, \end{aligned} \tag{4}$$

in which $Y=[y_1,y_2,\cdots,y_N]\in\mathbb{R}^{d\times N}$ and $U\in\mathbb{R}^{m\times d}$ are the multiview feature embedding and the projection matrix to predict the rest of samples, respectively.

2.2. Ensemble Manifold Regularization

While the objective formulation in Eq. (4) achieves to unify the goal of multiview feature embedding into a matrix low-rank approximation framework, the complementary property of different views has not been considered since such feature concatenation deals with all the views equally. In fact, the different views are obtained from different perspectives which have different physical meanings and statistical properties, thus the desires low dimensional feature embedding should be able to encode the correlation among different views by fully account for the complementary property. In EMR-SLRA, we employ the ensemble manifold regularization [27] to regularize our matrix approximation to fit the intrinsic and nonlinear structure of multiview data.

Based on the aforementioned notations, for the v-th view, we denote its undirected graph as $G^{(v)}=\{X^{(v)},W^{(v)}\}$, in which $X^{(v)}$ is the set of vertices (actually it is the feature matrix of the v-th view itself) and $W^{(v)}\in\mathbb{R}^{N\times N}$ is the relation matrix weighted by the heat kernels [28]:

$$W_{ij}^{(v)}=\begin{cases} e^{(-\frac{\left\|x_i^{(v)}-x_j^{(v)}\right\|^2}{t})}, & x_j^{(v)}\in \mathrm{N}(x_i^{(v)}) \quad \text{or} \quad x_i^{(v)}\in \mathrm{N}(x_j^{(v)}) \\ 0, & \text{else} \end{cases} \tag{5}$$

in which $\mathrm{N}(x_i^{(v)})$ is the k-nearest-neighbors of sample $x_i^{(v)}$ and t is the parameter in Gaussian function.

According to our previous patch alignment framework [16], the data manifold structure in the v-th feature space can be preserved as much as possible by the following optimization:

$$\arg\min_{Y} \sum_{i \neq j} W_{ij}^{(v)} \left\| y_i - y_j \right\|^2 = \arg\min_{Y} \ \mathrm{tr}(Y L^{(v)} Y^{\mathrm{T}}). \tag{6}$$

in which $L^{(v)}$ is the Laplacian matrix of the v-th view, i.e., $L^{(v)} = D^{(v)} - W^{(v)}$ and $D^{(v)}$ is a diagonal matrix whose entries are column sums of $W^{(v)}$ [29].

Because of the complementary property of multiple views to each other, different views definitely have different contributions to the multiview feature embedding. The ensemble manifold regularization suggests that the intrinsic manifold can be learned by the optimal linear combination of the pregiven candidates. If we impose a set of nonnegative weights $\beta = [\beta_1, \beta_2, \cdots, \beta_V]$ ($\beta > 0$ and $\sum_{v=1}^{V} \beta_v = 1$) on the optimization (6) over all views, the ensemble manifold regularization of the multiview data X is:

$$\begin{aligned} &\arg\min_{Y,\beta} \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(Y L^{(v)} Y^{\mathrm{T}}), \\ &\text{s.t.} \ \sum_{v=1}^{V} \beta_v = 1, \ \beta > 0, \end{aligned} \tag{7}$$

in which r is a scale parameter to control the weights of multiple features. It is obvious that the larger weight β_v makes more important contribution of v-th view in the embedding Y. Note that both the weight vector and the feature embedding are simultaneously optimized in Eq. (7), which indicates each feature has been regularized ro a particular role in the multiview feature embedding.

By combining Eqs. (4) and (7), we have:

$$\begin{aligned} &\arg\min_{U,Y,\beta} \ \left\| X - UY \right\|^2 + \lambda_1 \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(Y L^{(v)} Y^{\mathrm{T}}), \\ &\text{s.t.} \ U^{\mathrm{T}} U = I, \ \sum_{v=1}^{V} \beta_v = 1, \ \beta > 0. \end{aligned} \tag{8}$$

2.3. Group Sparsity Constraint

Generally, in computer vision and pattern recognition researches, the observed data is more or less suffered from the noise which challenges the robustness of the recognition model. Furthermore, in the discussed multiview feature learning task, the comprehensive noise of the multiview feature matrix is more complicated than the single feature DR since the noises are from different views with different distributions. Therefore, this noise can be transferred from the observed matrix X to the low dimensional feature representation Y by the low-rank approximation (3) and then decrease the robustness of our multiview DR algorithm. To address this issue, in EMR-SLRA algorithm, besides the two low-rank factor matrices U and Y, we also introduce the ideal multiview feature matrix $\widehat{X}$ to alleviate the affect by noises from different views, as shown in Fig. 1. Our main motivation here is to impose a direct noise reduction by the extra consideration of the error between ideal matrix $\widehat{X}$ and input matrix X. Although this constraint is only defined with respect to $\widehat{X}$, the effect of noise reduction can be transferred to U and Y in the EMR-SLRA algorithm. By this consideration, the robustness of EMR-SLRA can be guaranteed even if some views have severely corrupted by noise.

In this paper, we focus on the $\ell_{2,1}$-norm fitting constraint since it is robust against to noise in the data [30, 31, 32]:

$$\arg\min_{\widehat{X}} \left\| \widehat{X} - X \right\|_{2,1}. \tag{9}$$

in which the $\ell_{2,1}$-norm regularizer is defined as [31]:

$$\|X\|_{2,1} = \sum_i \sqrt{\sum_j X_{ij}^2} = \sum_i \|x_{i,:}\|_2 \tag{10}$$

By this way, we replace the input matrix X by the ideal one $\widehat{X}$ in Eqs. (4) and (8), and then put Eq. (9) into Eq. (8) to formulate the objective function

of EMR-SLRA algorithm as follows:

$$\begin{aligned} \arg\min_{U,\widehat{X},Y,\beta} & \left\|\widehat{X}-UY\right\|^2+\lambda_1\sum_{v=1}^{V}(\beta_v)^r\mathrm{tr}(YL^{(v)}Y^{\mathrm{T}})+\lambda_2\left\|\widehat{X}-X\right\|_{2,1}, \\ \text{s.t. } & U^{\mathrm{T}}U=I,\ \sum_{v=1}^{V}\beta_v=1,\ \beta>0, \end{aligned} \tag{11}$$

in which λ_1 and λ_2 are parameters to control the ensemble manifold regularization and the group sparsity constraint, respectively.

From Eq. (11), we further note that the proposed EMR-SLRA algorithm can also deal with the single feature based embedding as well, by the settings of $V = 1$ and $\beta = 1$. Thus the proposed feature DR algorithm degrades to the version of manifold regularized sparse low-rank approximation (MR-SLRA).

To apply our EMR-SLRA into practice, we have to concern the following key problem: how to solve Eq. (11) efficiently. Since there is no closed-form solution for EMR-SLRA, we develop an efficient EMR-SLRA optimization procedure by an iterative way, which is elaborated in the next section.

3. Efficient EMR-SLRA Optimization

The objective function of the EMR-SLRA in (11) is a multi-variable and non-convex problem, and there is no known optimal solution which allows for the simultaneous optimization of all the variables. To overcome this problem, in our developed optimization procedure, we iteratively optimize only one of the variables (U, Y, β, and $\widehat{X}$) by fixing other three. By this way, we decompose the objective function (11) into four sub optimizations as follows:

3.1. Optimize U by fixing Y, β, and $\widehat{X}$

Then the objective function (11) is reduced to:

$$\begin{aligned} \arg\min_{U} & \left\|\widehat{X}-UY\right\|^2, \\ \text{s.t. } & U^{\mathrm{T}}U=I_d. \end{aligned} \tag{12}$$

The Lagrangian function of (12) is:

$$\begin{aligned} L(U,\Lambda) &= \left\|\widehat{X} - UY\right\|^2 + \Lambda(U^{\mathrm{T}}U - I_d) \\ &= \mathrm{tr}(\widehat{X}^{\mathrm{T}}\widehat{X}) + \mathrm{tr}(Y^{\mathrm{T}}Y) - 2\mathrm{tr}(U^{\mathrm{T}}\widehat{X}Y^{\mathrm{T}}) + \Lambda(U^{\mathrm{T}}U - I_d), \end{aligned} \tag{13}$$

in which Λ is the Lagrangian multiplier. The partial derivative of $L(U,\Lambda)$ respect to U is:

$$\begin{aligned} \frac{\partial L}{\partial U} = 0 &\implies -2\widehat{X}Y^{\mathrm{T}} + 2U\Lambda = 0 \\ &\implies U = \widehat{X}Y^{\mathrm{T}}\Lambda^{-1}. \end{aligned} \tag{14}$$

If we set $Z = \widehat{X}Y^{\mathrm{T}}$, thus $U = Z\Lambda^{-1}$, by putting it into the constraint $U^{\mathrm{T}}U = I_d$, then we have:

$$\begin{aligned} U^{\mathrm{T}}U = I_d &\implies \Lambda^{-1}Z^{\mathrm{T}}Z\Lambda^{-1} = I_d \\ &\implies \Lambda = (Z^{\mathrm{T}}Z)^{\frac{1}{2}}. \end{aligned} \tag{15}$$

We further denote the rank of matrix Z is $rank(Z) = d$, by performing the SVD decomposition on Z, we have $Z = GDV^{\mathrm{T}}$. Thus U can be optimized by:

$$\begin{aligned} U = Z\Lambda^{-1} &\implies U = Z(Z^{\mathrm{T}}Z)^{-\frac{1}{2}} \\ &\implies U = Z(VDG^{\mathrm{T}}GDV^{\mathrm{T}})^{-\frac{1}{2}} \\ &\implies U = (GDV^{\mathrm{T}})(VD^{-1}V^{\mathrm{T}}) \\ &\implies U = GV^{\mathrm{T}}. \end{aligned} \tag{16}$$

3.2. Optimize Y by fixing U, β, and $\widehat{X}$

Then the objective function (11) reduces to:

$$\arg\min_{Y} \left\|\widehat{X} - UY\right\|^2 + \lambda_1 \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}). \tag{17}$$

Denoting $L = \sum_{v=1}^{V} (\beta_v)^r L^{(v)}$, Eq. (17) reduces to:

$$\arg\min_{Y} \left\| \widehat{X} - UY \right\|^2 + \lambda_1 \mathrm{tr}(YLY^{\mathrm{T}}). \tag{18}$$

By the derivative of the single-variable function $F(Y) = \left\| \widehat{X} - UY \right\|^2 + \lambda_1 \mathrm{tr}(YLY^{\mathrm{T}})$ respect to Y, we have:

$$\begin{aligned} \frac{dF(Y)}{dY} \; = \; 0 \Longrightarrow \; & Y + \lambda_1 YL - U^{\mathrm{T}} \widehat{X} \; = \; 0 \\ \Longrightarrow \; & Y = U^{\mathrm{T}} \widehat{X} \Psi^{-1}, \end{aligned} \tag{19}$$

in which $\Psi = I + \lambda_1 L$ and I is an identity matrix.

3.3. Optimize β by fixing U, Y, and $\widehat{X}$

Then the objective function (11) is reduced to:

$$\begin{aligned} & \arg\min_{\beta} \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}), \\ & \text{s.t.} \sum_{v=1}^{V} \beta_v = 1, \; \beta > 0. \end{aligned} \tag{20}$$

The Lagrangian function of (20) is:

$$L(\beta_v, \eta) \; = \; \sum_{v=1}^{V} (\beta_v)^r p^{(v)} - \eta (\sum_{v=1}^{V} \beta_v - 1), \tag{21}$$

in which $p^{(v)} = \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}})$ and η is the Lagrangian multiplier. The partial derivative of $L(\beta_v, \eta)$ respect to β_v is:

$$\begin{aligned} \frac{\partial L(\beta_v, \eta)}{\partial \beta_v} \; = \; 0 \; \Longrightarrow \; & r(\beta_v)^{r-1} p^{(v)} - \eta \; = \; 0 \\ \Longrightarrow \; & \beta_v = (\frac{\eta}{rp^{(v)}})^{\frac{1}{r-1}}. \end{aligned} \tag{22}$$

Finally, by considering the constraint $\sum_{v=1}^{V} \beta_v = 1$ into Eq. (22), we have:

$$\beta_v = (rp^{(v)})^{\frac{1}{1-r}} / \sum_{v=1}^{V} (rp^{(v)})^{\frac{1}{1-r}}. \tag{23}$$

3.4. Optimize $\widehat{X}$ by fixing U, Y, and β

Then the objective function (11) reduces to:

$$\arg\min_{\widehat{X}} \left\|\widehat{X} - UY\right\|^2 + \lambda_2 \left\|\widehat{X} - X\right\|_{2,1}. \tag{24}$$

Since the term of $\ell_{2,1}$-norm is non-smooth, the problem in Eq. (24) can be rewritten as follows, according to [31, 33].

$$\arg\min_{\widehat{X}} \left\|\widehat{X} - UY\right\|^2 + \lambda_2 \mathrm{tr}\{(\widehat{X} - X)^{\mathrm{T}} \widehat{W} (\widehat{X} - X)\}, \tag{25}$$

in which $\widehat{W}$ is a diagonal matrix given by:

$$\widehat{W} = \begin{pmatrix} \frac{1}{2\left\|(\widehat{X}-X)_{1,:}\right\|_F} & & \\ & \ddots & \\ & & \frac{1}{2\left\|(\widehat{X}-X)_{d,:}\right\|_F} \end{pmatrix}. \tag{26}$$

Finally, by the derivative of the single-variable function in Eq. (25), $\widehat{X}$ can be optimized by:

$$\widehat{X} = (I_d + \lambda_2 \widehat{W})(UY + \lambda_2 \widehat{W} X)^{-1}. \tag{27}$$

We summarize the proposed efficient EMR-SLRA optimization procedure as in Algorithm 1.

4. Experiments

In this section, we show some experimental results on the pattern recognition applications to compare the EMR-SLRA algorithm with some other multiview feature DR approaches. We focus our multiview feature embedding algorithm on the challenging problems of the text-image retrieval, web image annotation, and hyperspectral image classification. In the reminder of this section, we first provide a brief description of the aforementioned three public data sets and their corresponding input multiple features, and then give the detailed experimental settings of our algorithm and related comparison methods. Finally, we

Algorithm 1 Efficient EMR-SLRA Optimization Algorithm

Input: Multiview feature matrix $X = [X^1, X^2, \cdots, X^V] \in \mathbb{R}^{m \times N}$, dimensionality of embedded feature space d, regularization parameters λ_1 and λ_2, size of k-nearest-neighbors k, scale parameter r, maximal iteration number *Iter* and threshold ξ.

Initialization

- Randomly initialize $Y \in \mathbb{R}^{d \times N}$,
- Initialize $\widehat{X} = X + \text{random}(m, N)$,
- Initialize $\beta = [1/V, 1/V, \cdots, 1/V]$,
- Initialize $\widehat{W}$ by Eq. (26).

Repeat for $t = 1$ **to** *Iter*

- SVD decomposition $Z_t = \widehat{X} Y_t^{\mathrm{T}}$, i.e., $Z_t = G_t D_t V_t$
- Update $U_{t+1} = G_t V_t^{\mathrm{T}}$
- Let $L_t = \sum_{i=v}^{V} (\beta_v)_t^r L^{(v)}$, $\Psi_t = I + \lambda_1 L_t$, update $Y_{t+1} = U_t^{\mathrm{T}} \widehat{X}_t \Psi_t^{-1}$
- Let $p_t^{(v)} = \mathrm{tr}(Y_t L^{(v)} Y_t^{\mathrm{T}})$, update $(\beta_v)_t = (r p_t^{(v)})^{\frac{1}{1-r}} / \sum_{i=1}^{V} (r p_t^{(v)})^{\frac{1}{1-r}}$
- Update $\widehat{X}_{t+1} = (I_d + \lambda_2 \widehat{W}_t)(U_t Y_t + \lambda_2 \widehat{W}_t X)^{-1}$ and $\widehat{W}_{t+1}$ by Eq. (26)
- Calculate Obj_{t+1} by Eq. (11), break iteration if $|Obj_{t+1} - Obj_t| < \xi$

End

Output: Multiview feature embedding Y and projection matrix U

present the pattern recognition performance using the embedded feature and the parameters analyses.

4.1. Data set description

In this paper, we select three benchmark data sets for performance evaluation of our EMR-SLRAalgorithm: the Wiki Text-Image data set [1] [34], the NUS-WIDE-LITE data set [2][35], and the HYDICE hyperspectral data set [3][36].

The Wiki Text-Image data set was collected from the Wikipedia's featured articles collection. At the time of collection (October 2009), it had 2669 articles spread over 29 categories. By considered only 10 most populated ones

[1] http://www.svcl.ucsd.edu/projects/crossmodal/

[2] http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm

[3] https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html

with at least 150 instances per class, the final corpus contain 2,866 multimedia instances, in which 2173 for training and 693 for testing, and each instance contains a single image and at least 70 words. The detailed categories of Wiki Text-Image data set includes art, biology, geography, history, literature, media, music, royalty, sport, and warfare. Some examples of these images are shown in Fig. 2 and the median text length is about 200 words. Along with the data set, the authors had supplied the 128-D SIFT histogram image features and 10-D latent Dirichlet allocation model based text features, these two features serve as the input multiple features for the EMR-SLRA algorithm.

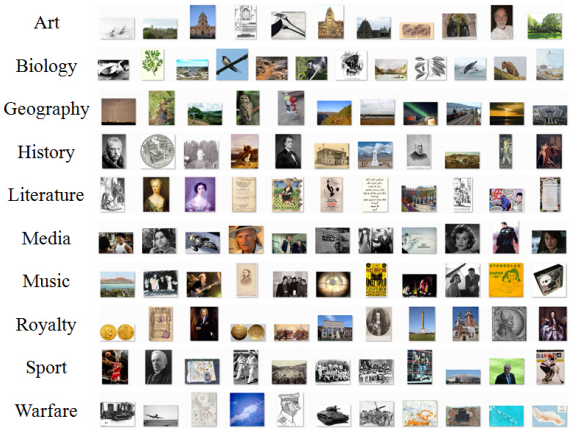


Figure 2: Example images in the Wiki Text-Image data set.

The NUS-WIDE-LITE data set is a subset of NUS-WIDE data set, which is a real-world web image data set from Flickr created by National University of Singapore. This smaller data set is composed of 28807 images for training and 28808 images for testing. Each of the image is combined with a 81-D label vector to indicate its relationship to all the 81 distinct concepts. In our experiment, we only use the object images for performance evaluation. Since we deal with the single label image annotation, we further remove the images with zero label or more than one labels from the object image subset. Then, by discarding the categories of scarce images, we have the following nine classes: birds, boats, flowers, rocks, sun, tower, toy, tree, and vehicle, in which 10600 images for

training and 7094 images for testing. Five visual features already included in the data set distribution serve as the input of EMR-SLRA algorithm: (1) 65-D color histogram, (2) 226-D block-wise color moments, (3) 145-D color auto-correlogram, (4) 74-D edge direction histogram, and (5) 129-D wavelet texture.

The hyperspectral images (HSI) are captured by the special designed hyperspectral sensors in which each pixel has contiguous bands of spectra, as shown in Fig. 3 refer to the hyperspectral digital imagery collection experiment (HYDICE) data cube we use for image classification. This data set was primarily released by Purdue University, and then serves as the most standard HSI for DR and classification. This data set is an urban site from the mall in Washington, DC, which has the size of 280 $\times$ 307 pixels. By delete the water absorption bands, a total of 191 channels in the 0.4-2.4 μm region of the visible and infrared spectra are considered in the experiments. In HSI processing area, the term classification is used to assign all the single pixels in the image to a set of classes, e.g., the roof, grass, tree, road, path, and shadow in the HYDICE data set. The input three kind of features are extracted by the following approaches [37, 38]: (1) 191-D spectral feature, by arranging a pixel's digital number (DN) in all of the spectral bands, (2) 60-D Gabor wavelet based texture feature over the top principal component of HSI (scale=5, direction=12), and (3) 80-D extended morphological profile [39] algorithm based morphological feature (scale element as [0,2,4,6,8]).

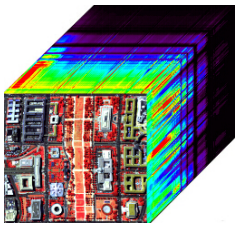


Figure 3: HYDICE hyperspectral data set.

In all of the data sets, we use the aforementioned multiple features as the input for DR algorithm, and then train the multi-class one versus one Support

Vector Machine (SVM) [40, 41] using the selected training samples with the embedded feature. After that, we predict the labels of the test samples and report the accuracy.

4.2. Experimental settings

To compare the effectiveness of the EMR-SLRA with the conventional feature DR methods for multiview feature embedding, we also show the performance of the following unsupervised methods: (1) the best performance of a single view feature (Best); (2) directly stack the multiview feature without DR (Stack); (3) Principal Component Analysis (PCA) [8, 9] of the multiview feature matrix; (4) Locality Preserving Projection (LPP) [29] of the multiview feature matrix; (5) Neighborhood Preserving Embedding (NPE) [42] of the multiview feature matrix, (6) Canonical Correlational Analysis (CCA) [43, 24], (7) Partial Least Squares (PLS) [44, 24], (8) Bilinear Model (BLM) [45, 24], and (9) Multiview Spectral Embedding (MSE) [21]. Note that the algorithms CCA, PLS and BLM were previously designed to deal with the data set that only has two different views, although they can be generalized to the multiview version [24], we have to carefully tune too many coefficients. Therefore, we only perform these three algorithms in the Wiki Text-Image data set which exactly has two views.

The detailed parameters setting of the aforementioned algorithms are as follows. For all of the algorithms, we first normalize each of the single feature matrix to the range [0, 1] individually. In LPP, we set the k-nearest-neighbors parameter as k=10 and the kernel parameter t=1, while in NPE, we also set k=10. The three coefficients in CCA, PLS and BLM are set according to literature [24]. In MSE and EMR-SLRA, the heat kernel parameters are set followed by LPP (k=10 and t=1), and the scale parameter r is fix to 10. The two regularization parameters λ_1 and λ_2 (Eq. (11)) are decided by using cross validation with the same range of $[10^{-5}, 10^{-4}, \cdots, 10^{4}]$. Since the optimal subspace dimensionality d is data dependent, we report the overall accuracy respect to d in all the data sets and compare the detailed class accuracies by keep the same

number of dimensionality for all the methods (i.e., 10 for Wiki Text-Image, 100 for NUS-WIDE-LITE and 50 for HYDICE hyperspectral data set). Finally, in the subsequent classification step, we use the embedded feature to train the SVM with Gaussian radial basis function kernel, again, the parameters C and γ in SVM are tuned by cross validation approach in the ranges of $[1, 10, 50, 100]$ and $[0.1, 1.0, 10, 100]$, respectively.

4.3. Performance on the three data sets

In this subsection, we report the performance of the EMR-SLRA algorithm and comparison methods on the three public data sets, i.e., the Wiki Text-Image data set, the NUS-WIDE-LITE data set, and the HYDICE hyperspectral data set, respectively.

4.3.1. Wiki Text-Image data set

In this subsection, we use the Wiki Text-Image data set to illustrate the effectiveness of EMR-SLRA algorithm in comparing with its competitors including the traditional multiview methods CCA, PLS and BLM. Table 1 shows the number of training and test samples of each category in the Wiki Text-Image data set. In order to simulate a small-sample-size scenario in the classification step, we randomly select 100 training samples per category to train the SVM classifier. The test collection is left unchanged and all the performance accuracies are reported apply to this test collection. The experiments are repeated ten times using the aforementioned independent training samples and fixed test samples, and then the average results are reported in Table 2 and Fig. 4.

Table 2 compares the classification accuracies of all categories and the overall accuracy (OA) and average accuracy (AA) use the best single view feature, multiview feature stacking, and the top 10 most significant features produced by eight DR algorithms. We observe that the proposed EMR-SLRA algorithm performs the highest OA and AA and obtains most of the top classification accuracy values in the individual categories. Although EMR-SLRA does not achieve the best accuracy value for several categories, the classification accu-

Category	training	test
art	100/138	34
biology	100/272	88
geography	100/244	96
history	100/248	85
literature	100/202	65
media	100/178	58
music	100/186	51
royalty	100/144	41
sport	100/214	71
warfare	100/374	104
total	1000/2173	693

Table 1: Number of training and test samples in the Wiki Text-Image data set.

racy is almost close to the best one. In order to further investigate the effect of subspace dimensionality d on the classification performance, Fig. 4 shows classification OA with regard to the embedded feature dimensionality for all feature embedding algorithms and two baseline algorithms. As shown in this figure, the EMR-SLRA performs better than the other eight DR algorithms when d is larger than 6. Moreover, EMR-SLRA performs at the best classification OA when d is larger than 10 and achieves the relative stable and high level of OA record when the d is increased to a larger value, whereas the traditional methods such as LPP and NPE perform unsatisfactorily since they were basically designed for DR of feature in a single feature space.

Algorithm	Best	Stack	PCA	LPP	NPE	CCA	PLS	BLM	MSE	EMR-SLRA
art	0.3823	0.4117	0.3529	0.4411	0.4117	0.4705	0.3823	0.4117	0.4117	0.4411
biology	0.9090	0.8863	0.9090	0.9204	0.8977	0.8977	0.8863	0.8977	0.9204	0.9090
geography	0.6666	0.6770	0.7187	0.6666	0.6875	0.7187	0.7083	0.7185	0.6770	0.7708
history	0.4352	0.4235	0.5058	0.5411	0.5176	0.4941	0.4705	0.5764	0.4941	0.5058
literature	0.6769	0.5846	0.6461	0.6769	0.6769	0.6153	0.6923	0.6769	0.6461	0.6923
media	0.5344	0.4827	0.4310	0.3965	0.3448	0.2068	0.2413	0.1379	0.3101	0.5862
music	0.1372	0.5490	0.5882	0.6666	0.6274	0.7254	0.7450	0.8039	0.6666	0.5882
royalty	0.7073	0.6341	0.6585	0.6585	0.6097	0.6829	0.7073	0.6585	0.6829	0.6829
sport	0.8591	0.8450	0.8872	0.8732	0.8591	0.8873	0.8873	0.8732	0.8732	0.8873
warfare	0.7307	0.7596	0.7788	0.7788	0.7596	0.7211	0.7788	0.8079	0.7788	0.7596
OA	0.6378	0.6522	0.6810	0.6883	0.6695	0.6652	0.6767	0.6883	0.6738	0.7085
AA	0.6039	0.6253	0.6476	0.6620	0.6392	0.6420	0.6499	0.6562	0.6461	0.6823

Table 2: Category specific accuracies for various features in the Wiki Text-Image data set (d=10).

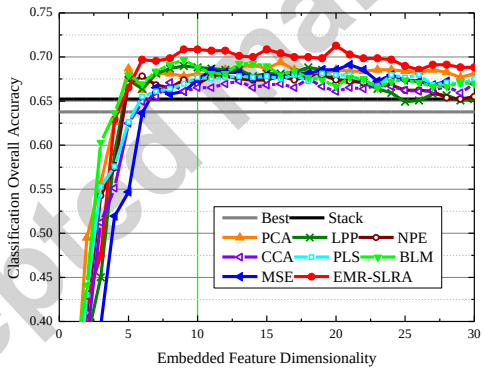


Figure 4: Classification OA with regard to the embedded feature dimensionality in the Wiki Text-Image data set.

4.3.2. NUS-WIDE-LITE data set

In this subsection, we show the effectiveness of EMR-SLRA algorithm in real-world web image annotation use the NUS-WIDE-LITE data set. Similar to the aforementioned Wiki Text-Image data set, we randomly select 200 samples per category from the training database to feed into SVM, and all the test samples are left unchanged for performance evaluation. The number of training

and test samples are indicated in Table 3. Again, ten independent classification experiments are implemented and the the average results are reported in Table 4 and Fig. 5.

Since there are five different views in the NUS-WIDE-LITE data set, besides CCA, PLS and BLM algorithms (which have too many coefficients to be carefully tuned when the number of views is larger than 2), another seven different feature representations based image classification accuracy values are listed in Table. 4. Among them, the feature dimensionality of the best feature and stack feature are 129 (the wavelet texture) and 639, respectively, while for the other DR algorithms we fix the embedded feature dimensionality d to 100. It is obvious that the proposed EMR-SLRA algorithm achieves the best global OA and several of the top category specific accuracies, which is similar to the results in the Wiki Text-Image data set. In addition, the classification OAs with an increase in the number of feature d for all the DR approaches are compared in Fig. 5. From this figure, we learn that the proposed multiview feature embedding algorithm achieves the best OA when the value of d is larger than 20, and the optimal OA stabilizes when the embedded feature dimensionality increases from 40 to 100. The results provided in Table. 4 and Fig. 5 confirm the similar conclusion that EMR-SLRA outperforms the traditional feature DR algorithms for multiview feature embedding.

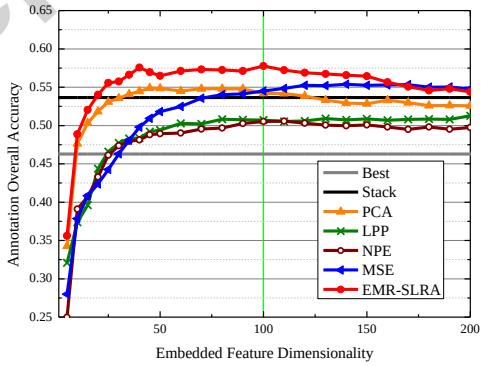


Figure 5: Classification OA with regard to the embedded feature dimensionality in the NUS-WIDE-LITE data set.

Category	training	test
birds	200/897	623
boats	200/798	508
flowers	200/2229	1523
rocks	200/1542	1059
sun	200/800	504
tower	200/633	443
toy	200/717	480
tree	200/1262	805
vehicle	200/1722	1149
total	1800/10600	7094

Table 3: Number of training and test samples in the NUS-WIDE-LITE data set.

Algorithm	Best	Stack	PCA	LPP	NPE	MSE	EMR-SLRA
birds	0.2600	0.4943	0.4687	0.4269	0.3996	0.4526	0.4815
boats	0.3405	0.4603	0.4015	0.3622	0.3661	0.3917	0.4291
flowers	0.5778	0.6447	0.5877	0.5502	0.5219	0.6270	0.6704
rocks	0.4721	0.5278	0.6081	0.5835	0.5571	0.5561	0.6185
sun	0.5753	0.6309	0.6746	0.5992	0.6428	0.6289	0.6428
tower	0.4153	0.4492	0.4582	0.4537	0.4108	0.4898	0.4921
toy	0.1792	0.4333	0.4646	0.4083	0.4500	0.4833	0.4646
tree	0.4658	0.4496	0.4596	0.3776	0.4298	0.4770	0.4944
vehicle	0.5509	0.5535	0.5866	0.6336	0.6057	0.6022	0.6449
OA	0.4628	0.5365	0.5420	0.5126	0.5052	0.5451	0.5777
AA	0.4263	0.5160	0.5233	0.4883	0.4871	0.5232	0.5486

Table 4: Category specific accuracies for various features in the NUS-WIDE-LITE data set (d=100).

Category	training	test
roof	30	2576
grass	30	1980
tree	30	1772
road	30	1568
path	30	1174
shadow	30	1182
total	180	10253

Table 5: Number of training and test samples in the HYDICE hyperspectral data set.

4.3.3. HYDICE hyperspectral data set

The pixel-based image classification task of HYDICE hyperspectral data set is challenging because many pixels of the roof, road, and path are similar in spectral domain (1st view) since they may be made of similar materials [46]. For each round of image classification, we randomly select 30 samples per class for training, and all the rest samples in reference data are treated as test samples (Table 5). Then the average results of ten independent rounds of classification are shown in Table 6 and Fig. 6. In the hyperspectral image classification task, we usually use the kappa coefficient instead of AA to couple with OA for performance evaluation [47]. The similar classification performance can be observed from Table 6 and Fig. 6, in that the proposed EMR-SLRA based feature representation achieves the best performance by the OA and kappa in the HYDICE hyperspectral data set. In summary, the classification performance on the various of data sets in this subsection all demonstrate that the proposed EMR-SLRA is an effective and robust multiview feature embedding algorithm.

Algorithm	Best	Stack	PCA	LPP	NPE	MSE	EMR-SLRA
roof	0.6832	0.7554	0.8839	0.8881	0.8652	0.9262	0.9475
grass	0.9989	0.9919	0.9974	0.9681	0.9959	0.9252	0.9833
tree	0.9599	0.9655	0.9644	0.8617	0.9102	0.9531	0.9633
road	0.9706	0.9770	0.9540	0.9241	0.8686	0.9171	0.9674
path	0.9088	0.9071	0.9335	0.9173	0.9403	0.9378	0.9710
shadow	0.9754	0.9864	0.9129	0.8732	0.9923	0.9814	0.9585
OA	0.8955	0.9153	0.9395	0.9061	0.9220	0.9369	0.9642
Kappa	0.8740	0.8975	0.9265	0.8859	0.9054	0.9233	0.9564

Table 6: Category specific accuracies for various features in the HYDICE hyperspectral data set (d=50).

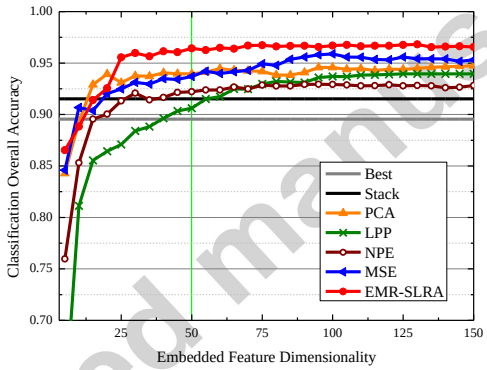


Figure 6: Classification OA with regard to the embedded feature dimensionality in the HYDICE hyperspectral data set.

4.4. Parameters analyses

Finally, we further provide some analyses of several key parameters and the effect of ensemble manifold regularization in the proposed EMR-SLRA algorithm, based on the Wiki Text-Image data set.

4.4.1. Regularization parameters λ_1 and λ_2

The two regularization parameters λ_1 and λ_2 in objective function (11) provide the tradeoff between ensemble manifold regularization and group sparsity constraint in EMR-SLRA. Theoretically, the large λ_1 rises the effect of ensemble manifold regularization but would causes a potential overfiting; the large λ_2 em-

phasizes the sparse level but might removes too many significant features, while the small λ_2 might keeps too many redundant and noisy features that would be introduced into low dimensional feature representation. In our experiments, we search these parameters in the range of $\{10^{-5}, 10^{-4}, \cdots, 10^{4}\}$. Fig. 7 shows the classification OA with respect to λ_1 and λ_2 in the Wiki Text-Image data set, in this experiment, we fix d=10. As it can be seen from this figure that we can obtain a near-optimal performance of our proposed method by set a moderate λ_1 value (e.g., from 10^{-1} to 10^{1}) and a larger λ_2 value (e.g., from 10^{1} to 10^{2}).

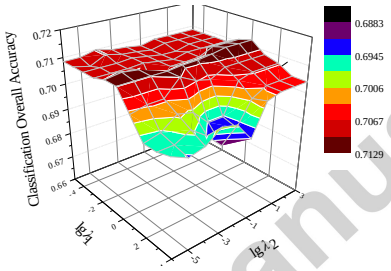
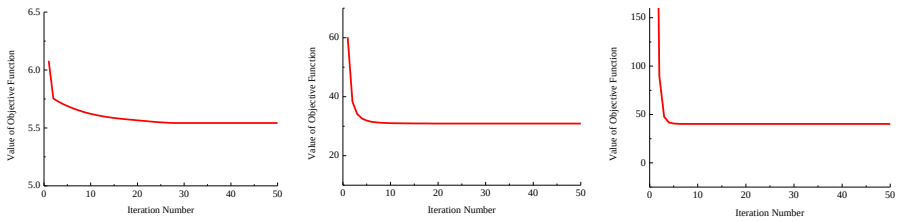


Figure 7: Classification OA with regard to two regularization parameters λ_1 and λ_2 in the Wiki Text-Image data set.

4.4.2. Convergence analysis

To solve the optimization problem of EMR-SLRA algorithm, we develop an iterative procedure as in Algorithm 1. Figs. 8(a) to 8(c) show the objective function value of EMR-SLRA with various of parameter combinations in the Wiki Text-Image data set. From this figure, it is clear that the value of the objective function monotonically decreases at each iteration, and it is also observed that the objective function quickly deceases at the first few rounds of iteration (always less than 10) and becomes stable within 40 iterations. We have also observed the similar trend in the other two data sets, as well as with more parameter combinations. Thus, these results demonstrate that our proposed optimization algorithm is effective and converges rapidly.



(a) (b) (c)

Figure 8: Convergence report of the proposed algorithm in the Wiki Text-Image data set with various of parameter combinations. (a) $\lambda_1 = 0.001, \lambda_2 = 0.1$; (b) $\lambda_1 = 0.01, \lambda_2 = 1$ and (c) $\lambda_1 = 0.1, \lambda_2 = 10$.

4.4.3. Effect of the ensemble manifold regularization

In the EMR-SLRA algorithm, the ensemble manifold regularization term is introduced to explore the complementary property of different views and helps to find an effective feature embedding for pattern classification. In order to confirm the significant contribution of EMR in the EMR-SLRA algorithm, we further compare the performance of our algorithm with its companion version without EMR and show the accuracies of the Wiki Text-Image data set in Fig. 9. In this experiment, all the remained parameters are set as the same as we illustrated above. It is evident from this figure that the imposed EMR improves the accuracies for 6/10 categories, and the overall accuracy and average accuracy have also promoted from 0.6652 to 0.7085 and 0.6365 to 0.6823, respectively.

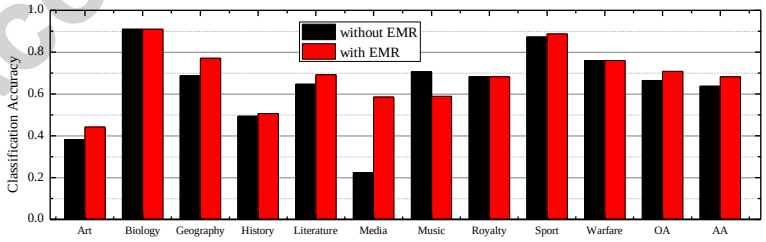


Figure 9: Effect of the ensemble manifold regularization in the Wiki Text-Image data set.

5. Conclusion

In this paper, to reduce the feature dimensionality of multiview data while learn the enhanced feature embedding, we propose the EMR-SLRA algorithm for multiview feature dimensionality reduction. The proposed algorithm is based on the framework of least-squares component analysis, by introducing the ensemble manifold regularization and group sparsity constraint into our algorithm, the complementary property among multiple features has been properly considered and its robustness against the noise has been simultaneously promoted. In order to efficiently solve the optimization of EMR-SLRA, we further develop an iterative procedure to obtain the multiview feature embedding. The effectiveness of the proposed method has been verified by the extensive experimental results on the challenging problems of the text-image retrieval, web image annotation, and hyperspectral image classification. We also investigate the free parameters involved in our algorithm and conclude that these parameters are easy to be tuned to achieve a near-optimal feature embedding. For future work, the proposed method will be extended to more challenging tasks such as cross-domain and cross-media pattern recognition.

Acknowledgment

The authors would like to thank the handing editor and the anonymous reviewers for their careful reading and helpful remarks, which have contributed in improving the quality of this paper. This paper is supported by the National Natural Science Foundation of China under Grants 61401317, 61471274, 91338202 and 91338111.

References

[1] J. Yu, D. Tao, Y. Rui, J. Cheng, Pairwise constraints based multiview features fusion for scene classification, Pattern Recognit. 46 (2) (2012) 483–496.

[2] W. Liu, D. Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process. 22 (7) (2013) 2676–2687.

[3] I. A. Gheyas, L. S. Smith, Feature subset selection in large dimensionality domains, Pattern Recognit. 43 (1) (2010) 5–13.

[4] F. D. la Torre, A least-squares framework for component analysis, IEEE Trans. Pattern Anal. Mach. Intell. 34 (6) (2012) 1041–1055.

[5] J. Ni, Q. Qiu, R. Chellappa, Subspace interpolation via dictionary learning for unsupervised domain adaptation, in: CVPR, 2013, pp. 692–699.

[6] Y. Ma, P. Niyogi, G. Sapiro, R. Vidal, Dimensionality reduction via subspace and submanifold learning, IEEE Signal Process. Mag. 28 (2) (2011) 14–15.

[7] X. Chen, S. Chen, H. Xue, X. Zhou, A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data, Pattern Recognit. 45 (5) (2012) 2005–2018.

[8] H. Hotelling, Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology 24 (6) (1933) 417–441.

[9] I. T. Jolliffe, Principal Component Analysis, Springer, New York, USA, 2002.

[10] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (22) (2000) 2323–2326.

[11] J. B. Tenenbaum, V. de Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (22) (2000) 2319–2323.

[12] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: NIPS, Vol. 14, 2002, pp. 585–592.

[13] D. L. Donoho, C. Grimes, Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data, PNAS 100 (10) (2003) 5591–5596.

[14] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, SIAM J. Sci. Comput. 26 (1) (2004) 313–338.

[15] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.

[16] T. Zhang, D. Tao, X. Li, J. Yang, Patch alignment for dimensionality reduction, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1299–1313.

[17] D. Cai, X. He, J. Han, T. S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1548–1560.

[18] N. Guan, D. Tao, Z. Luo, B. Yuan, Non-negative patch alignment framework, IEEE Trans. Neural Networks 22 (8) (2011) 1218–1230.

[19] Z. Zhang, K. Zhao, Low-rank matrix approximation with manifold regularization, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1717–1729.

[20] B. Long, P. S. Yu, Z. Zhang, A general model for multiple view unsupervised learning, in: SDM, 2008, pp. 822–833.

[21] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man Cybern. Part B Cybern. 60 (6) (2010) 1438–1446.

[22] D. Zhou, C. J. C. Burges, Spectral clustering and transductive learning with multiple views, in: ICML, 2007, pp. 1159–1166.

[23] Z. Zhao, H. Liu, Multi-source feature selection via geometry-dependent covariance analysis, J. Mach. Learn. Res. 4 (2008) 36–47.

[24] A. Sharma, A. Kumar, H. D. III, D. W. Jacobs, Generalized multiview analysis: A discriminative latent space, in: CVPR, 2012, pp. 2160–2167.

[25] Q. Qiu, V. M. Patel, P. Turagay, R. Chellappa, Domain adaptive dictionary learning, in: ECCV, 2012, pp. 631–645.

[26] B. Lin, X. He, C. Zhang, M. Ji, Parallel vector field embedding, J. Mach. Learn. Res. 14 (1) (2013) 2945–2977.

[27] B. Geng, D. Tao, C. Xu, L. Yang, X.-S. Hua, Ensemble manifold regularization, IEEE Trans. Pattern Anal. Mach. Intell. 34 (6) (2012) 1227–1233.

[28] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.

[29] X. He, P. Niyogi, Locality preserving projections, in: NIPS, Vol. 16, 2004, pp. 153–160.

[30] A. Argyriou, T. Evgeniou, M. Pontil, Multi-task feature learning, in: NIPS, Vol. 19, 2007, pp. 41–48.

[31] F. Nie, H. Huang, X. Cai, C. H. Ding, Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization, in: NIPS, Vol. 23, 2010, pp. 1813–1821.

[32] S. Xiang, F. Nie, G. Meng, C. Pan, C. Zhang, Discriminative least squares regression for multiclass classification and feature selection, IEEE Trans. Neural Netw. Learn. Syst. 23 (11) (2012) 1738–1754.

[33] H. Wang, F. Nie, H. Huang, Multi-view clustering and feature learning via structured sparsity, in: ICML, 2013, pp. 352–360.

[34] J. C. Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G. R. Lanckriet, R. Levy, N. Vasconcelos, On the role of correlation and abstraction in cross-modal multimedia retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 36 (3) (2014) 521–535.

[35] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, Nus-wide: a real-world web image database from national university of singapore, in: CIVR, ACM, 2009, pp. 1–9.

[36] D. Landgrebe, Hyperspectral image data analysis, IEEE Signal Process. Mag. 19 (1) (2002) 17–28.

[37] L. Zhang, L. Zhang, D. Tao, X. Huang, On combining multiple features for hyperspectral remote sensing image classification, IEEE Trans. Geosci. Remote Sens. 50 (3) (2012) 879–893.

[38] L. Zhang, L. Zhang, D. Tao, X. Huang, A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images, ISPRS J. Photogramm. 83 (2013) 30–39.

[39] J. A. Benediktsson, J. A. Palmason, J. R. Sveinsson, Classification of hyperspectral data from urban areas based on extended morphological profiles, IEEE Trans. Geosci. Remote Sens. 43 (3) (2005) 480–491.

[40] V. N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Networks 10 (5) (1999) 988–999.

[41] C.-C. Chang, C.-J. Lin, Libsvm: A library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27.

[42] X. He, D. Cai, S. Yan, H.-J. Zhang, Neighborhood preserving embedding, in: ICCV, Vol. 2, 2005, pp. 1208–1213.

[43] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, New York, NY, USA, 2004.

[44] A. Sharma, D. W. Jacobs, Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch, in: CVPR, 2011, pp. 593–600.

[45] J. B. Tenenbaum, W. T. Freeman, Separating style and content with bilinear models, Neural Comput. 12 (6) (2000) 1247–1283.

[46] X. Huang, L. Zhang, An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, IEEE Trans. Geosci. Remote Sens. 51 (1) (2013) 257–272.

[47] L. Zhang, L. Zhang, D. Tao, X. Huang, Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction, IEEE Trans. Geosci. Remote Sens. 51 (1) (2013) 242–256.