MMFE: Multitask Multiview Feature Embedding

Qian Zhang*, Lefei Zhang†, Bo Du†, Wei Zheng*, Wei Bian‡, and Dacheng Tao‡
*Beijing Samsung Telecom R&D Center, Beijing 100028, China.
{q0712.zhang, w0209.zheng}@samsuang.com
†School of Computer, Wuhan University, Wuhan 430079, China.
{zhanglefei, remoteking}@whu.edu.cn
‡QCIS and FEIT, University of Technology, Sydney, NSW 2007, Australia.
{wei.bian, dacheng.tao}@uts.edu.au

***Abstract*—In data mining and pattern recognition area, the learned objects are often represented by the multiple features from various of views. How to learn an efficient and effective feature embedding for the subsequent learning tasks? In this paper, we address this issue by providing a novel multi-task multiview feature embedding (MMFE) framework. The MMFE algorithm is based on the idea of low-rank approximation, which suggests that the observed multiview feature matrix is approximately represented by the low-dimensional feature embedding multiplied by a projection matrix. In order to fully consider the particular role of each view to the multiview feature embedding, we simultaneously suggest the multitask learning scheme and ensemble manifold regularization into the MMFE algorithm to seek the optimal projection. Since the objection function of MMFE is multi-variable and non-convex, we further provide an iterative optimization procedure to find the available solution. Two real world experiments show that the proposed method outperforms single-task-based as well as state-of-the-art multiview feature embedding methods for the classification problem.**

***Keywords*—Multiview, Multitask, Dimension reduction, Classification.**

I. Introduction

Recently, a huge number of algorithms have been studied to learn a unified low-dimensional feature embedding from data which represented by multiple distinct feature sets [1], [2]. It is known from these researches that the different views from different feature spaces have particular physical meanings and statistical properties, thus the expected learning algorithms should address the redundancy as well as the complementary property among the multiview features [3], [4], [5]. There are three groups of representative methods for multiview learning, i.e., co-training, multiple kernel learning, and subspace learning [2]. In this paper, we focus on the subspace learning based approaches, which aim to find a latent subspace shared by multiple features. Since we believe that the efficient feature representation is the key engine for the subsequent learning tasks , the discriminative power of feature representation determines the upper boundary of learning performance by the subsequent classifier. Although the original feature dimensionality for each view is high, the intrinsical dimensionality of such multiview data is lower than that of any input view, thus the "curse of dimensionality" is correspondingly reduced by the subspace representation.

Originally, the straightforward approach for multiview subspace learning is to directly concatenate all the different feature vectors together into a long vector and then apply a certain feature dimensionality reduction (DR) algorithm, e.g., PCA, LDA, LPP, and NPE [6]. However, it is not appropriate mainly because the different statistical properties are not duly considered and the complementary information of different features is not well explored [7]. Alternatively, as the pioneer work of multiview feature embedding, the distributed spectral embedding (DSE) [8] learns a common low-dimensional embedding which is close to the individual embedding of each view as much as possible. After that, some multiview subspace learning methods set out to find a latent space by learning multiple view-specific projections, e.g., multiview spectral embedding (MSE) [9], generalized multiview analysis [10], and multiview patch alignment [11]. In addition, there are also some multiview metric learning approaches such as the semi-supervised multiview learning [12]. However, most of the existing multiview subspace learning algorithms have at least one of the following problems: (1) the multiview embedding of test samples can not be directly predicted due to the model requires a number of training samples or the suggested feature mapping is nonlinear and implicit, (2) the final embedding is learned by the single task of the joint matrix while the intrinsic correlation between different views has been neglected and the generalization ability for the whole task has been reduced, and (3) the complementary property of different views has not been optimally explored since each view actually has its specific statistical and physical meanings.

In order to solve the above problems, in this paper, we propose a multitask multiview feature embedding (MMFE) algorithm for multiview subspace learning. Recently, a least-squares framework for component analysis has been proposed to formulate the component analysis methods and overcomes the small sample size problem in subspace learning [13]. Based on this framework, we firstly stack the multiple features into a whole matrix, and the low-dimensional feature embedding and the projection matrix are distinctly obtained by the low-rank matrix approximation. It is also known that multitask learning is a machine learning technique that learns several tasks simultaneously for better performance by capturing the intrinsic correlation between different tasks [14], [15], in this paper, we further enhance our subspace learning framework into the multitask version to boost the performance. In addition, the MMFE simultaneously enforces the ensemble manifold regularization (EMR) [16] on the embedding, since it is conditionally optimal for intrinsic manifold approximation of multiple views. Furthermore, since there is no closed-form solution for MMFE, we propose an efficient MMFE optimization procedure to solve the proposed optimization.

1550-4786/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDM.2015.82

The remainder of this paper is organized as follows. In Section II, we provide the objective formulation of MMFE in detail. Section III proposes the efficient MMFE optimization procedure. Then, the experimental results are reported in Section IV, followed by the conclusion in Section V.

II. Problem Formulation

Suppose we have V views (tasks), for each view (task), $X^{(v)}$, $Y^{(v)}$, and $U^{(v)}$ denote the original feature matrix, low-dimensional representation, and projection matrix, respectively. Then, by combining all of the views (tasks) together, X, Y, and U denote the multiview feature matrix, multiview feature embedding, and multitask projection matrix, respectively. By these definitions, the full objective function of MMFE algorithm is shown in Eq. (1), which will be detailedly discussed in the following three subsections, i.e., the multiview low-rank approximation, the multitask learning scheme, and the ensemble manifold regularization.

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

A. Multiview Low-Rank Approximation

The proposed MMFE algorithm is based on the least-squares component analysis framework. Suppose the multiview data $X=\{x_1,x_2,\cdots,x_N\}$, in a least-squares point of view, denote the desires low dimensional feature representation as $Y=\{y_1,y_2,\cdots,y_N\}$, principal components analysis minimizes the following reconstruction error by using the optimal orthogonal basis under the least-squares framework [13]:

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

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

If define $\Phi=[\phi_1,\phi_2,\cdots,\phi_d]$, ($\Phi^{\mathrm{T}}\Phi=I$), thus we have $Y=\Phi^{\mathrm{T}}X$. Then, Eq. (2) has its matrix formulation:

$$\varepsilon \;=\; \sum_{i=1}^{N}\left\|x_i-\Phi(\Phi^{\mathrm{T}}x_i)\right\|^2=\left\|X-\Phi(\Phi^{\mathrm{T}}X)\right\|^2. \tag{3}$$

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

$$\arg\min_{\Phi}\ \left\|X-\Phi(\Phi^{\mathrm{T}}X)\right\|^2, \text{s.t. } \Phi^{\mathrm{T}}\Phi=I. \tag{4}$$

Now we turn to the case of multiview feature embedding. For each sample, we have $x_i=\{x_i^{(1)}\in\mathbb{R}^{m_1},\cdots,x_i^{(V)}\in\mathbb{R}^{m_V}\}$. For the consideration of multiview feature concatenation, the new feature of each sample is denoted by $x_i=[x_i^{(1)},x_i^{(2)},\cdots,x_i^{(V)}]^{\mathrm{T}}$, $m=\sum_{i=1}^{V}m_i$, and the multiview feature matrix is denoted by $X=[X^{(1)},X^{(2)},\cdots,X^{(V)}]^{\mathrm{T}}$. Followed by these definitions, the low-rank matrix approximation of the multiview data can be simply unified into Eq. (5):

$$\arg\min_{U,Y}\ \|X-UY\|^2, \text{s.t. } U^{\mathrm{T}}U=I. \tag{5}$$

B. Multitask Learning Scheme

Multitask learning, aiming to jointly train multiple related problems, can improve performance by exploiting the shared information among all the problems [17]. This motivates the development of multitask learning [18], [19], [20] and it has been demonstrated that multitask learning is superior to single-task learning. For better exploitation of multiple features in the learned subspace, we resort to multitask learning to mine the shared information of different features. It must be noted that most existing multitask methods tackle classification problems, in which each task has label data. In this paper, we consider multitask feature embedding problem, where the data in each task are unlabeled. It aims to learn a task with other related task at the same time by learning a shared low-dimensional representation.

In this paper, we consider low-rank approximation of multiview data and each view data as a task. Thus, Eq. (5) can be rewritten as follows:

$$\begin{aligned}\arg\min_{U,U^{(v)},Y,Y^{(v)}}\ & \|X-UY\|^2+\sum_{v=1}^{V}\left\|X^{(v)}-U^{(v)}Y^{(v)}\right\|^2 \\ \text{s.t. } & U^{\mathrm{T}}U=I,\ {U^{(v)}}^{\mathrm{T}}U^{(v)}=I.\end{aligned} \tag{6}$$

When it comes to multitask feature embedding, we are going to learning a shared low-dimensional representation among related tasks, in which we perform low-rank approximation for both multiview data and each view data. And the shared low-dimensional representation can be seen as a new feature embedding, where the data distribution from each view are similar to each other. Then, we obtain:

$$\begin{aligned}\arg\min_{U,U^{(v)},Y,Y^{(v)}}\ & \|X-UY\|^2+\sum_{v=1}^{V}\left\|X^{(v)}-U^{(v)}Y^{(v)}\right\|^2 \\ & +\alpha\sum_{v=1}^{V}\left\|Y-Y^{(v)}\right\|^2 \\ \text{s.t. } & U^{\mathrm{T}}U=I,\ {U^{(v)}}^{\mathrm{T}}U^{(v)}=I,\end{aligned} \tag{7}$$

in which α is a regularization parameter. It balances the low-rank approximation of input data and the shared low-dimensional representation across all multiple features.

C. Ensemble Manifold Regularization

So far, the complementary property of different views has not been properly considered since the objective functions provided in the aforementioned subsections deal with all the views equally. Therefore, in MMFE, we further employ the ensemble manifold regularization [16] to regularize the feature embedding 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 and $W^{(v)} \in \mathbb{R}^{N\times N}$ is the relation matrix weighted by the heat kernels [21]. According to the patch alignment framework [6], the data manifold structure in the v-th feature space can be preserved as much as possible by the following optimization:

$$\arg\min_{Y} \sum_{i\neq j} W_{ij}^{(v)} \|y_i - y_j\|^2 = \arg\min_{Y} \ \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}). \quad (8)$$

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:

$$\arg\min_{Y,\beta} \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}), \text{s.t. } \sum_{v=1}^{V} \beta_v = 1,\ \beta > 0, \quad (9)$$

in which r is a scale parameter to control the weights of multiple features. It is obviously 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. (9), which indicates each feature has been regularized to a particular role in the multiview feature embedding.

By combining Eqs. (7) and (9), we have the objective function of MMFE algorithm as provided in Eq. (1). From this equation, we further note that the proposed MMFE 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 (MFE).

III. Solution

Our objective function of the MMFE in (1) is a multivariable and non-convex problem, and it cannot be easily solved. To overcome this problem, we develop an efficient MMFE optimization procedure by an iterative way. By fixing other variables, we alteratively optimize only one variable. Thus, our objective function can be decomposed into the following sub-optimization problems.

The first step is fixing $U^{(v)}$, Y, $Y^{(v)}$, and β, solving U. Then, the objective function (1) is reduced to

$$\arg\min_{U} \ \|X - UY\|^2, \text{s.t. } U^{\mathrm{T}}U = I. \quad (10)$$

Thus, the Lagrangian function of (10) is:

$$\begin{aligned} L(U,\Lambda) &= \|X - UY\|^2 + \Lambda(U^{\mathrm{T}}U - I) \\ &= \mathrm{tr}(X^{\mathrm{T}}X) + \mathrm{tr}(Y^{\mathrm{T}}Y) - 2\mathrm{tr}(U^{\mathrm{T}}XY^{\mathrm{T}}) + \Lambda(U^{\mathrm{T}}U - I), \end{aligned} \quad (11)$$

in which Λ is the Lagrangian multiplier. Taking the derivative *w.r.t* U to 0, we have:

$$U = XY^{\mathrm{T}}\Lambda^{-1}. \quad (12)$$

By setting $Z = XY^{\mathrm{T}}$, we have $U = Z\Lambda^{-1}$. Putting it into the constraint $U^{\mathrm{T}}U = I$, then we obtain:

$$\Lambda^{-1}Z^{\mathrm{T}}Z\Lambda^{-1} = I \implies \Lambda = (Z^{\mathrm{T}}Z)^{\frac{1}{2}}. \quad (13)$$

By performing SVD decomposition on Z, we get $Z = GDV^{\mathrm{T}}$. Substituting it into $U = Z\Lambda^{-1}$, U can be optimized as follows:

$$\begin{aligned} U = Z\Lambda^{-1} &\implies U = Z(Z^{\mathrm{T}}Z)^{-\frac{1}{2}} \\ &\implies U = Z(VDG^{\mathrm{T}}GDV^{\mathrm{T}})^{-\frac{1}{2}} \\ &\implies U = (GDV^{\mathrm{T}})(VD^{-1}V^{\mathrm{T}}) \\ &\implies U = GV^{\mathrm{T}}. \end{aligned} \quad (14)$$

The second step is fixing U, Y, $Y^{(v)}$, and β, solving $U^{(v)}$. The objective function (1) is equivalent to:

$$\begin{aligned} \arg\min_{U^{(v)}} \ &\left\|X^{(v)} - U^{(v)}Y^{(v)}\right\|^2, \\ \text{s.t. } &{U^{(v)}}^{\mathrm{T}}U = I. \end{aligned} \quad (15)$$

In the same manner with the first step, we can obtain

$$U = G^{(v)}{V^{(v)}}^{\mathrm{T}}. \quad (16)$$

The third step is fixing U, $U^{(v)}$, $Y^{(v)}$, and β, solving Y. The objective function (1) converts to the following one problem.

$$\begin{aligned} \arg\min_{Y} \ \|X - UY\|^2 + \alpha \sum_{v=1}^{V} \left\|Y - Y^{(v)}\right\|^2 \\ + \lambda \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}). \end{aligned} \quad (17)$$

Taking the derivative *w.r.t* Y to 0, we can get:

$$\begin{aligned} &Y + \alpha VY + \lambda YL - U^{\mathrm{T}}X - \alpha \sum_{v=1}^{V} Y^{(v)} = 0 \\ \Longrightarrow &Y = (U^{\mathrm{T}}X + \alpha \sum_{v=1}^{V} Y^{(v)})((1+Va)I + \lambda L)^{-1}. \end{aligned} \quad (18)$$

The fourth step is fixing U, $U^{(v)}$, Y, and β, solving $Y^{(v)}$. The objective function (1) is rewritten as

$$\arg\min_{Y^{(v)}} \left\|X^{(v)} - U^{(v)}Y^{(v)}\right\|^2 + \alpha \left\|Y - Y^{(v)}\right\|^2 \quad (19)$$

Taking the derivative *w.r.t* $Y^{(v)}$ to 0, we have:

$$\begin{aligned} &Y^{(v)} + \alpha Y^{(v)} - {U^{(v)}}\mathrm{T}X^{(v)} - \alpha Y = 0 \\ \Longrightarrow &Y^{(v)} = ({U^{(v)}}^{\mathrm{T}}X^{(v)} + \alpha Y)/(1+a). \end{aligned} \quad (20)$$

The fifth step is fixing U, $U^{(v)}$, Y, and $Y^{(v)}$, solving β. The objective function (1) is formulated as the following optimization problem

$$\begin{aligned} \arg\min_{\beta} \ &\sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}}), \\ \text{s.t. } &\sum_{v=1}^{V} \beta_v = 1,\ \beta > 0. \end{aligned} \quad (21)$$

By using a Lagrangian multiplier η, we can get the Lagrangian function of (21)

$$L(\beta_v, \eta) = \sum_{v=1}^{V} (\beta_v)^r p^{(v)} - \eta(\sum_{v=1}^{V} \beta_v - 1), \quad (22)$$

in which $p^{(v)} = \mathrm{tr}(YL^{(v)}Y^{\mathrm{T}})$. Taking the derivative *w.r.t* β_v to 0, we have:

$$r(\beta_v)^{r-1} p^{(v)} - \eta = 0 \implies \beta_v = (\frac{\eta}{rp^{(v)}})^{\frac{1}{r-1}}. \quad (23)$$

Substituting $\sum_{v=1}^{V} \beta_v = 1$ into Eq. (23), we can get

$$\beta_v = (rp^{(v)})^{\frac{1}{1-r}} / \sum_{v=1}^{V} (rp^{(v)})^{\frac{1}{1-r}}. \quad (24)$$

We summarize the proposed efficient MMFE optimization procedure as in Algorithm 1.

Algorithm 1 Optimization procedure for MMFE algorithm

Input: Multiview feature matrix $X = [X^{(1)}, X^{(2)}, \cdots, X^{(V)}] \in \mathbb{R}^{m \times N}$ and single feature matrix $X^{(v)} \in \mathbb{R}^{m_v \times N}$, dimensionality of embedded feature space d, regularization parameters α and λ, scale parameter r, maximal iteration number *Iter* and threshold ξ.

Output: Feature embedding Y, $Y^{(v)}$ and projection matrix U, $U^{(v)}$

Initialization

- Randomly initialize $U^{(v)}$,
- Randomly initialize Y and $Y^{(v)}$,
- Initialize $\beta = [1/V, 1/V, \cdots, 1/V]$.

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

- SVD decomposition $Z_t = XY_t^{\mathrm{T}}$, i.e., $Z_t = G_t D_t V_t$, Update $U_{t+1} = G_t V_t^{\mathrm{T}}$
- SVD decomposition ${Z^{(v)}}_t = X^{(v)} {Y^{(v)}}_t^{\mathrm{T}}$, i.e., ${Z^{(v)}}_t = {G^{(v)}}_t {D^{(v)}}_t {V^{(v)}}_t$, Update ${U^{(v)}}_{t+1} = {G^{(v)}}_t {V^{(v)}}_t^{\mathrm{T}}$
- update $Y_{t+1} = (U_{t+1}^{\mathrm{T}} X + \alpha \sum_{v=1}^{V} {Y^{(v)}}_t)((1 + Va)I + \lambda L)^{-1}$
- update ${Y^{(v)}}_{t+1} = ({U^{(v)}}_{t+1}{}^{\mathrm{T}} X^{(v)} + \alpha Y_{t+1})/(1 + a)$
- update Let $p_{t+1}^{(v)} = \mathrm{tr}(Y_{t+1} L^{(v)} Y_{t+1}^{\mathrm{T}})$, update $(\beta_v)_{t+1} = (rp_{t+1}^{(v)})^{\frac{1}{1-r}} / \sum_{i=1}^{V} (rp_{t+1}^{(v)})^{\frac{1}{1-r}}$
- Calculate Obj_{t+1} by Eq. (1), break iteration if $|Obj_{t+1} - Obj_t| < \xi$

End

IV. Experiments

In this section, we provide experimental results on the classification of MSRA-MM images [22] and NUS-WIDE dataset [23], to test the performance of our proposed MMFE algorithm and compare it to the single-task-based models and state-of-the-art multiview feature embedding methods. For all the experiments, we apply the individual-class accuracy as the measurement for performance evaluation. To compare the effectiveness of our MMFE algorithm, we also show the performance of the following unsupervised methods: (1) LPP [24], and (2) NPE [25] which directly perform on the stacked multiview feature matrix; and the state-of-the-art multiview subspace learning algorithms: (3) DSE [8], (4) MSE [9], and (5) MSLR [26]. The detailed parameters setting of these algorithms are as follows. For all of the algorithms, we firstly normalize each view to the range [-1, 1], respectively. In LPP, we set *k*=10 and *t*=5, while in NPE, we also set *k*=10. In MSE, MSLR, and MMFE, we fix *k*=10, *t*=5, and *r*=10, respectively. In addition, for the two regularization parameters in MSLR (λ_1, λ_2) and MMFE (α, λ), they are tuned by a two-fold cross validation with the same range of $10^{[-5,\ldots,4]}$. The only parameter left is the embedding dimensionality *d*, which is widely recognized as an open question in DR area. In the experiments, we show the classification performance respect to *d* in all the datasets for a comprehensive comparison. For each algorithm, after the feature embedding is obtained, a multi-class one-versus-one support vector machine (SVM) [27] is trained by a set of training samples, then the labels of the test samples are predicted by the SVM model and the individual-class accuracies are reported.

A. MSRA-MM images dataset

The MSRA-MM images dataset is collected from Microsoft Live Search, including images and videos [22]. The image dataset contains 68 representative classes and 65443 images in all, in which each class contains around 1000 samples. A set of features have been provided, including: the block-wise color moment, color correlogram, edge distribution histogram, HSV color histogram, RGB color histogram, wavelet texture, and face features. In our experiments, we use six views among them as indicated in TABLE I.

TABLE I. Multiple features in the MSRA-MM images dataset.

View index	Feature description	Dimensionality
1#	Block-wise color moment	225
2#	Color correlogram	144
3#	Edge distribution histogram	75
4#	HSV color histogram	64
5#	RGB color histogram	256
6#	Wavelet texture	128

In the experiment, we select 8 classes (i.e., party, cat, panda, earth, dogs, snakes, cartoon, and background) out of the 68 classes to test the classification performance. In the classification step, we randomly select 200 samples per class for training, and use the rest of samples as test samples for accuracy evaluation. The comparison of individual-class accuracies of all the feature embedding algorithms is plotted in Fig. 1. We learn that nearly all of the individual-class accuracies are less than 0.5 (except the class cartoon) because this dataset is extreme challenging and in fact a portion of samples are not high relevant to the assigned label information (images with indicators 1 and 0) [28]. Despite all that, it can be observed that the proposed MMFE algorithm obvious outperforms its competitors in most of the individual accuracies, including party, cat, earth, dogs, and background. For the remaining three classes, the MMFE algorithm also reaches the comparable results (in fact, all the curves are overlapped in these classes).

B. NUS-WIDE dataset

The NUS-WIDE dataset is a real-world web image database from National University of Singapore [23]. The

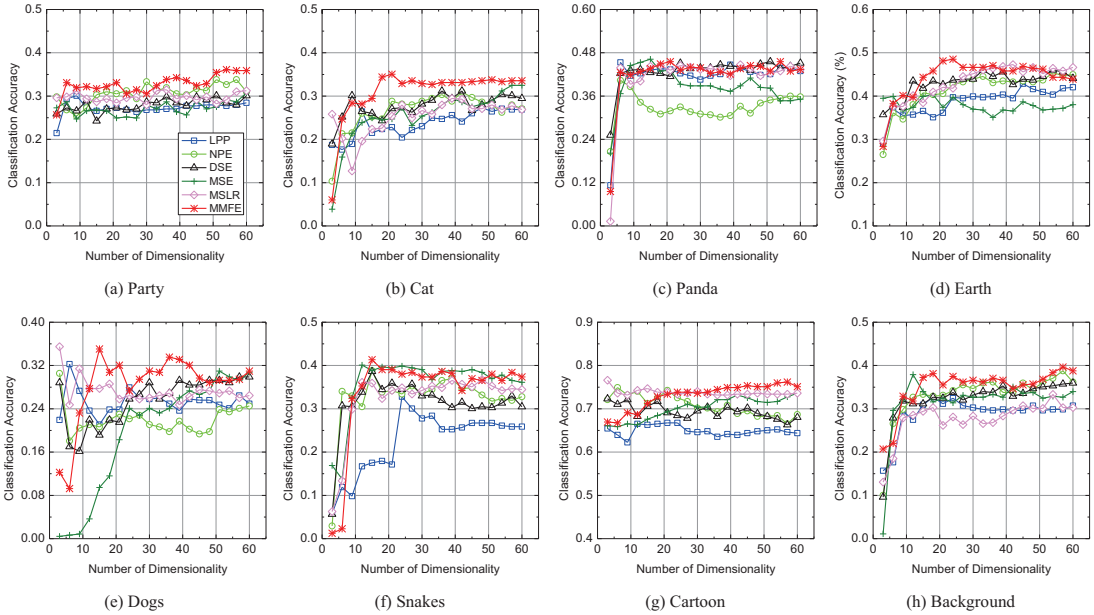


Fig. 1. Classification accuracy with regard to the number of subspace feature dimensionality in the MSRA-MM images dataset.

dataset includes: (1) 269648 images and the associated tags from Flickr, (2) six types of low-level features extracted from these images, and (3) ground-truth for 81 concepts that can be used for evaluation. To represent each image, six types of low-level features have been extracted as multiple views, i.e., the color histogram, block-wise color moment, color correlogram, edge direction histogram, and wavelet texture, the feature dimensionality of these views are given in TABLE II. Along with the NUS-WIDE dataset, a smaller dataset, i.e., NUS-WIDE-LITE, is also available. 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.

TABLE II. MULTIPLE FEATURES IN THE NUS-WIDE DATASET.

View index	Feature description	Dimensionality
1#	Color histogram	64
2#	Block-wise color moment	225
3#	Color correlogram	144
4#	Edge direction histogram	73
5#	Wavelet texture	128

In the experiment, 8 categories of images are interested in, i.e., the animal, buildings, clouds, flowers, grass, person, sky, and water. Since we only deal with the single label based classification, we further remove the images with zero label or more than one labels. Thus, there are 4283 images for training and 4375 images for test. In the classification step, we randomly select 100 samples per class for training, and use all of the test samples for accuracy evaluation.

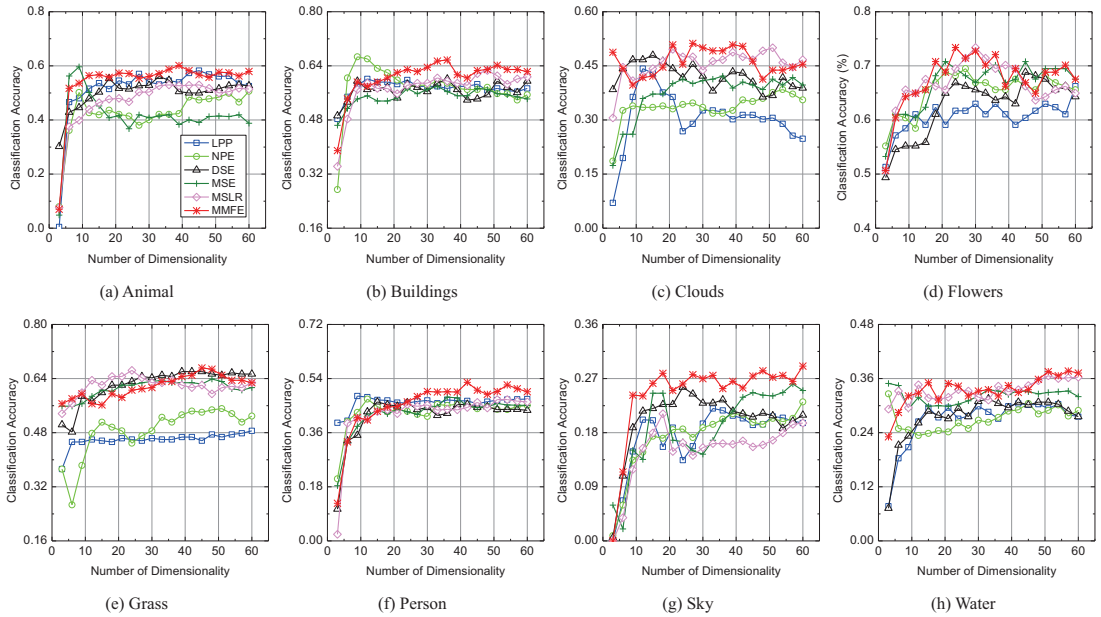
Finally, we investigate the classification results of NUS-WIDE dataset, as shown in Fig. 2. NUS-WIDE dataset is also a very challenging dataset since the concepts differ greatly and the images have large variances in scale, color, and shape. Even for the focused 8 categories classification problem, only 4 of individual-class accuracies reach 0.6 (i.e., the animal, buildings, flowers, and grass), while the accuracies of the sky and water are no more than 0.4. Also from this figure, many curves are overlapped and the observed accuracy improvement appears to be fewer than the MSRA-MM dataset reported above. However, we still experimentally find that the proposed MMFE algorithm performs a little better than or comparable to others.

V. CONCLUSION

In this paper, to deal with the multiview learning, we propose the MMFE algorithm to simultaneously reduce the feature dimensionality of multiview data while learn the enhanced feature embedding. The proposed algorithm is based on the framework of least-squares component analysis, by further introducing the multitask learning scheme and ensemble manifold regularization into our MMFE, an efficient and effective feature embedding is uncovered for the subsequent learning tasks. The effectiveness of the proposed method has been verified by classification results on the MSRA-MM images and NUS-WIDE datasets. For future work, the proposed method will be extended to more challenging tasks such as cross-domain and cross-media data mining and pattern recognition.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China under Grants 61401317 and 91338202, the Australian Research Council Projects FT-130101457 and LP-140100569, the Chancellor's Postdoctoral Research Fellowship of the University of Technology, Sydney, and the Open Project Program of the State Key Lab of CAD&CG (A1413), Zhejiang University.



(a) Animal (b) Buildings (c) Clouds (d) Flowers

(e) Grass (f) Person (g) Sky (h) Water

Fig. 2. Classification accuracy with regard to the number of subspace feature dimensionality in the NUS-WIDE dataset.

REFERENCES

[1] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, DOI: 10.1109/tpami.2015.2417578, 2015.

[2] S. Sun, "A survey of multi-view machine learning," *Neural Comput. & Applic.*, vol. 23, no. 7-8, pp. 2031–2038, 2013.

[3] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *ICDM*, 2014, pp. 110–119.

[4] X. Mei, Z. Hong, D. Prokhorov, and D. Tao, "Robust multitask multiview tracking in videos," *IEEE Trans. Neural Netw. Learn. Syst.*, DOI: 10.1109/tnnls.2015.2399233, 2015.

[5] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, 2012.

[6] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, 2009.

[7] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1559–1572, 2014.

[8] B. Long, P. S. Yu, and Z. Zhang, "A general model for multiple view unsupervised learning," in *SIAM Int. Conf. Data Mining*, 2008, pp. 822–833.

[9] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 60, no. 6, pp. 1438–1446, 2010.

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

[11] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, and Y. Y. Tang, "Group sparse multiview patch alignment framework with view consistency for image classification," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3126–3137, 2014.

[12] J. Yu, M. Wang, and D. Tao, "Semi-supervised multiview distance metric learning for cartoon synthesis," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4636–4648, 2012.

[13] F. D. la Torre, "A least-squares framework for component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1041–1055, 2012.

[14] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in *NIPS*, vol. 19, 2007, pp. 41–48.

[15] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 980–993, 2015.

[16] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, 2012.

[17] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.

[18] T. Evgeniou and M. Pontil, "Regularized multi-task learning," in *ACM SIGKDD*, 2004, pp. 109–117.

[19] P. Yang, K. Huang, and C.-L. Liu, "Multi-task low-rank metric learning based on common subspace," *Neural Information Processing*, vol. 7063, pp. 151–159, 2011.

[20] S. Parameswaran and K. Q. Weinberger, "Large margin multi-task metric learning," in *NIPS*, vol. 23, 2010, pp. 1867–1875.

[21] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[22] M. Wang, L. Yang, and X.-S. Hua, "Msra-mm: Bridging research and industrial societies for multimedia information retrieval," *Microsoft Res. Asian, Tech. Rep. Microsoft*, 2009.

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

[24] X. He and P. Niyogi, "Locality preserving projections," in *NIPS*, vol. 16, 2004, pp. 153–160.

[25] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *ICCV*, vol. 2, 2005, pp. 1208–1213.

[26] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, 2015.

[27] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.

[28] Y. Liu, Y. Liu, S. Zhong, and K. C. Chan, "Semi-supervised manifold ordinal regression for image ranking," in *ACM MM*, 2011, pp. 1393–1396.