WEIGHTED GRAPH EMBEDDED LOW-RANK PROJECTION LEARNING FOR FEATURE EXTRACTION

*Zhuojie Huang, Shuping Zhao, Lunke Fei, Jigang Wu**

School of Computer Science and Technology,
Guangdong University of Technology, Guangzhou, China, 510006

ABSTRACT

Low-rank based methods have been widely adopted to structure preserving, when the projection matrix is learned for feature extraction. However, some dilemmas still exist that degrade the classification performance: 1) The local structure of the data is ignored; 2) the reconstructed data is not consistent with the original data. To solve those problems, in this paper a weighted graph embedded low-rank projection (WGE_LRP) method is proposed. In WGE_LRP, a novel weighted graph regularization term is proposed, which can learn the local structure of the data based on the similarity of different samples. Meanwhile, an extra global information term is introduced to keep the reconstructed data consistent with the original data. Experimental results show that the proposed method can obtain competitive performance in comparison to the state-of-the-arts.

***Index Terms*—** Feature extraction, dimension reduction, low-rank representation, weighted graph

1. INTRODUCTION

Feature extraction plays a key role in the fields of pattern recognization. The classical methods for feature extraction include principal component analysis (PCA) [1], linear discriminant analysis (LDA), locality preserving projections (LPPs) [2], neighborhood preserving embedding (NPE) [3], sparsity preserving projections (SPP) [4], $etc.$ The above method extract data features by retaining local structure or global structure, and they can be unified into a common graph embedding framework [5].

In recent years, the low-rank-based feature extraction methods, which can perfectly preserve the global structure of data, have received a lot of attention owing to its robustness to noise [6, 7]. Preserving global and local structures during projection learning is very important for feature extraction [8, 9, 10, 11]. However, the original low-rank-based methods such as robust PCA (RPCA) and low-rank representation (LRR) [12] cannot deal with the new samples after the end of the training stage. To solve this problem, some methods have been proposed, including Double LRR (DLRR), Latent LRR (LatLRR) [13], $etc.$ But none of them can reduce the dimensionality of the data. To address this issue, those methods including low-rank preserving embedding (LRPE) and low-rank embedded projection (LRE), similar to the strategy of SPP [4], learn a projection matrix from data.

Unfortunately, those above methods cannot preserve the global structures and local structures at the same time, which leads to their performance was unsatisfying on some special tasks [11]. To solve this problem, Wen $et\ al.$ [14] proposed a method calling low-rank preserving projection via graph regularized reconstruction (LRPP_GRR), which imposed the graph constraint on the reconstruction error of data to capture the local structure of data. However, this model preserves the local structure only using the nearest k samples and ignores the local information of most of the samples, so the local structure cannot be well preserved.

To solve the problems mentioned above, we propose a method named weighted graph low-rank preserving projection (WGE_LRP). Different from the previous methods, WGE_LRP exploits a weighted graph term to preserve local structures, avoiding to chose the optimal number of the nearest samples. Meanwhile, we further introduce a new global structure term to make the reconstructed samples consistent with the original data. By adding these two additional items, local and global information can be better preserved, WGE_LRP can obtain competitive performance on a variety of different databases. In brief, the contributions of this paper are as follows:

- We exploit a weighted graph term to preserve local structures and avoid to chose the number of nearest samples, so that some important local information is not ignored.
- We propose a model to preserve the local structure information according to the weight. Furthermore, a new global structure preserving term is proposed so that reconstruction data is consistent with the original data.

2. RELATED WORK

This section brief introduce the low-rank-based method. The general model of LRR [12] is formulated as follows:

$$\min_{Z} \|Z\|_* \text{ s.t. } X = XZ \tag{1}$$

*Corresponding author: asjgwucn@outlook.com

where $X = [x_1, x_2, \ldots, x_n] \in R^{m\times n}$ denotes a data matrix and each column x_i is a sample. Each element Z_{ij} of representation matrix $Z \in R^{n\times n}$ reveals the 'similarity' between samples x_i and x_j. To learn a more robust projection, Yin $et\ al.$ [15] proposed the following LRR-based feature extraction model:

$$\min_{Z,L} \|Z\|_* + \|L\|_* \text{ s.t. } X = LXZ. \tag{2}$$

In this case, the learned projection matrix $L \in R^{m\times m}$ is able to capture the global structure of original data. To learn the local structure of original data, LRPP_GRR [14] imposes the graph constraint on the reconstruction error of data. The model is formulated as follows:

$$\begin{aligned}\min_{P,Q,Z} \sum_{i,j=1}^{n} \left\|x_i - PQ^T X z_j\right\|_2^2 w_{ij} + \lambda_1\|Z\|_* + \lambda_2\|Q\|_{2,1} \\ \text{s.t. } P^T P = I, X = PQ^T XZ\end{aligned} \tag{3}$$

In Function (3), the matrix $P \in R^{n\times d}$ is the data reconstruction matrix and the matrix $Q \in R^{n\times d}$ is the learned projection matrix. d is the selected number of the projected dimension ($d<m$). Each w_{ij} of graph W can be simply defined as follows:

$$w_{ij} = \begin{cases} 1, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $N_k(x_j)$ denotes a set of k nearest neighbor samples of sample x_j. However, there are two problems with the graph W defined as Equation (4): 1) It is difficult to choose an appropriate number of neighbors; 2) the learning weights among neighbors are the same.

3. THE PROPOSED APPROACH

This section first introduces a novel weighted graph to calculate the learning weight of samples, followed by the proposed WGE_LRP.

3.1. Weighted Graph

There are two problems with the graph W: 1) How to choose a number of neighbors; 2) the learning weights are the same. To solve the two problems, we hope that the more similar the samples, the more local information the reconstructed data can learn from. Cosine similarity keeps the local structure of non-neighbors and pays more attention to the local structure of similar samples, so we use cosine similarity to describe the similarity between samples and normalize the similarity.

For a data matrix $X = [x_1, x_2, \ldots, x_n] \in R^{m\times n}$ with n samples, each element w_{ij} of the weighted graph W can be simply defined as follows:

$$w_{ij} = \frac{cos(x_i, x_j)}{\sum_{k=1}^{n} cos(x_i, x_k)} \tag{5}$$

where w_{ij} represents the learning weight that sample i can learn from sample j and $cos(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\|\|x_j\|}$. If the w_{ij} is larger, more information can be learned obviously [10].

3.2. Model and Solution

- **Model:** The representation matrix Z can represent the similarity between data but it is a low-rank matrix which will ignore many local structures of data. However, the weighted graph matrix W include all the similarity between the difference data, which cannot ignore some local structures of data. So the weighted graph is used to construct graph regularization item to capture the local structure. Additionally, the global consistency item is introduced to keep the reconstructed data consistent with the original data. WGE_LRP is proposed as follows:

$$\begin{aligned}\min_{P,H,Z,E} \sum_{i,j=1}^{n} \left\|x_i - PH^T x_j\right\|_2^2 w_{i,j} + \lambda_1\|Z\|_* + \lambda_2\|H\|_F^2 \\ + \lambda_3\|E\|_{2,1} + \lambda_4 \left\|PH^T X - PH^T XZ\right\|_F^2 \\ \text{s.t. } P^T P = I, X = PH^T XZ + E\end{aligned} \tag{6}$$

In Function (6), similar to LRPP GRR, the matrix $P \in R^{m\times d}$ is the data reconstruction matrix and the matrix $H \in R^{m\times d}$ is the learned projection matrix. d is the selected number of the projected dimension ($d<m$). W is a weight graph that defines as Equation (5). $E \in R^{m\times n}$ is a sparse additive error matrix. $\|\cdot\|_F$ is the matrix Frobenius norm. The first term is the weighted graph term that captures the local structure. The last term is a global structure term that pays more attention to maintaining the consistency of projection data before and after low-rank processing than the second term.

On feature extraction, we can learn the projection H after the objective Function (6) has been solved. From this, dimensionality reduction and feature extraction using the learned H is performed. Given the new samples N, we obtain the features from the N by $\bar{N} = H^T N$.

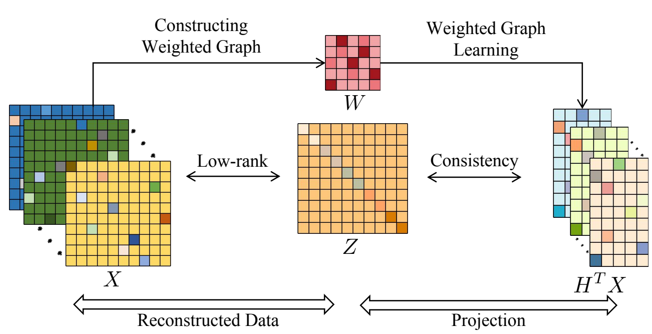


Fig. 1. Flowchart of the proposed method.

Fig. 1. illustrates the proposed method. Given n training samples $X = [X_1, X_2, \ldots, X_c] \in R^{m\times n}$ from c classes, we learn the low-rank coefficient Z and projection matrix H at the same time. Meanwhile, the weighted graph learning and consistency make projection matrix H more discriminative and consistent.

- **Solution:** Now, we use the alternating direction method of multipliers (ADMM) [16] to solve the optimization problem. To make the objection function separable, we first convert the problem into the following formula by imposing two auxiliary variables Y, U:

$$\begin{aligned} \min_{P,H,Z,E,U,Y} \sum_{i,j=1}^{n} \left\| x_i - PH^T x_j \right\|_2^2 w_{i,j} + \lambda_1 \|U\|_* + \lambda_2 \|H\|_F^2 \\ + \lambda_3 \|E\|_{2,1} + \lambda_4 \|PH^T X - PY\|_F^2 \\ \text{s.t. } P^T P = I, X = PY + E, Y = H^T XZ, Z = U \end{aligned} \tag{7}$$

Then, we reformulate Equation (7) into the following augmented Lagrangian function:

$$\begin{aligned} &\min_{P,H,Z,E,U,Y} Tr\left(XDX^T\right) - Tr\left(2XWX^T HP^T\right) \\ &+ Tr\left(H^T X(D+\lambda_4 I)X^T H\right) + \lambda_1 \|U\|_* + \lambda_2 \|H\|_F^2 \\ &+ \lambda_3 \|E\|_{2,1} - 2\lambda_4 Tr\left(H^T XY^T\right) + \lambda_4 Tr\left(YY^T\right) \\ &+ \frac{\rho}{2}\left(\|H^T XZ - Y\|_F^2 + \|Z-U\|_F^2 + \|X - PY - E\|_F^2\right) \\ &+ \langle C_1, Z-U\rangle + \langle C_2, X-PY-E\rangle + \left\langle C_3, H^T XZ - Y\right\rangle \end{aligned} \tag{8}$$

where $D = diag\,(D_{11}, D_{22}, \dots, D_{nn})$ is a diagonal matrix, and $D_{ii} = \sum_j w_{ij}$. $Tr\,(\cdot)$ is the trace operator. $\langle A, B\rangle = Tr\left(A^T B\right)$, C_1, C_2, C_3 are the Lagrangian multipliers. ρ is a positive penalty parameter. By alternately solve (8), we can obtain the solutions of all variables $P, H, Z, U, Y, E, C_1, C_2, C_3$. We set $L = X - E + \frac{C_2}{\rho}$ and the detail solution steps are as follow:

$$\begin{aligned} \mathcal{L}(H) = \min_H -2Tr\left(XWX^T HP^T\right) + Tr\left(H^T X(D+\lambda_4 I)X^T H\right) \\ + \lambda_2 \|H\|_F^2 - 2\lambda_4 Tr\left(H^T XY^T\right) + \frac{\rho}{2}\left(\|H^T XZ - Y + \frac{C_3}{\rho}\|_F^2\right) \end{aligned} \tag{9}$$

$$\mathcal{L}(Z) = \min_Z \|Z - U + \frac{C_1}{\rho}\|_F^2 + \|H^T XZ - Y + \frac{C_3}{\rho}\|_F^2 \tag{10}$$

$$\mathcal{L}(U) = \min_U \lambda_1 \|U\|_* + \frac{\rho}{2}\|Z - U + \frac{C_1}{\rho}\|_F^2 \tag{11}$$

$$\mathcal{L}(E) = \min_E \lambda_3 \|E\|_{2,1} + \frac{\rho}{2}\|X - PY - E + \frac{C_2}{\rho}\| \tag{12}$$

$$\begin{aligned} \mathcal{L}(Y) = \min_Y \frac{\rho}{2} Tr\left(LL^T\right) + Tr\left(YY^T\right) - 2Tr\left(LY^T P^T\right) \\ + \frac{\rho}{2}\|H^T XZ - Y + \frac{C_3}{\rho}\|_F^2 + \lambda_4 Tr\left(YY^T - 2H^T XY^T\right) \end{aligned} \tag{13}$$

$$\mathcal{L}(P) = \min_{P^T P = I} Tr\left(-2XWX^T HP^T\right) + \frac{\rho}{2}\|L - PY\|_F^2 \tag{14}$$

The derivative of (9) with respect to H, we can achieve

$$\begin{aligned} H = \left[2\lambda_2 I + \rho XZZ^T + (X(D + D^T + \lambda_4)X^T)\right]^{-1} \\ \cdot(\rho XZ(Y - \frac{C_3}{\rho})^T + 2XW^T X^T P + 2\lambda_4 H^T X) \end{aligned} \tag{15}$$

by setting it to zero. Z can be calculated by the derivative of (10) with respect to Z ,then Z is obtained

$$Z = (X^T HH^T X + I)^{-1}(X^T H(Y - \frac{C_3}{\rho}) + (U - \frac{C_1}{\rho})). \tag{16}$$

Then U is obtained by using the singular value thresholding (SVT) shrinkage operator [13] as follows:

$$U = \Theta_{\frac{\lambda_1}{\rho}}\left(Z + \frac{C_1}{\rho}\right), \tag{17}$$

where Θ denotes the shrinkage operator. To update E, the $l_{2,1}$-norm is utilized for (12) and E can be obtained as

$$E_{j,:} = \begin{cases} \frac{\|G_{j,:}\|_2 - \frac{\lambda_3}{\rho}}{\|G_{j,:}\|_2} G_{j,:} & if \|G_{j,:}\|_2 > \frac{\lambda_3}{\rho} \\ 0 & otherwise \end{cases}, \tag{18}$$

where $G = X - PY + \frac{C_2}{\rho}$. Y can be calculated by the derivative of $L(Y)$ with respect to Y, then Y is obtained

$$Y = \frac{1}{2\rho + 2\lambda_4}\left[\rho\left(P^T L + H^T XZ\right) + C_3 + 2\lambda_4 H^T X\right], \tag{19}$$

Problem (14) is an orthogonal Procrustes problem and can be simply solved. If

$$SVD(2XWX^T H + \rho LY^T) = USV^T, \tag{20}$$

then P is obtained as $\quad P = UV^T, \quad (21)$

where SVD is the singular value decomposition (SVD) operation. Lagrangian multipliers C_1 , C_2 ,C_3 and parameter ρ are updated by using the following formulas:

$$\begin{aligned} &C_1 = C_1 + \rho(Z - U), C_2 = C_2 + \rho(X - PY - E), \\ &C_3 = C_3 + \rho(H^T XZ - Y), \rho = \min(\mu\rho, \rho_{\max}) \end{aligned} \tag{22}$$

We set $P = \arg\max_{P^T P = I} Tr(P^T(\sum)P)$, where $\sum$ is the data covariance in advance. The detail procedures of the proposed algorithm are summarized in Algorithm 1.

Algorithm 1 WGE_LRP Algorithm [Solving (6)]

Input: data matrix X, parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, projected dimension d.

Output: P, H, Z

Initialization: $H = P, Z = 0, U = 0, Y = 0, C_1 = C_2 = C_3 = 0, \rho{=}1, \mu{=}1.1, \rho_{\max} = 10^8$, where μ and $\rho_{\max}$ are constans.

while not converged **do**

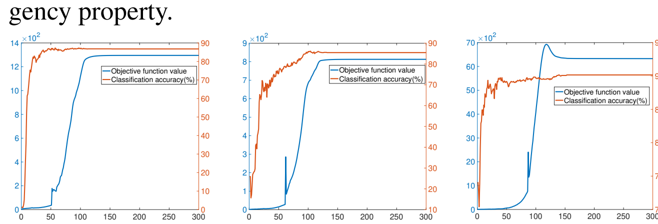
1.Update the variable of H, Z, U, E, Y,P by (15) - (21), respectively.

2.Update C_1, C_2, C_3, ρ by (22) .

end while

3.3. Convergency Analysis

In this sub-section, we experimentally show the convergency of the algorithm through a series of experiments including Yale B [17], AR [18], COIL20 [19]. Fig. 2. shows that the proposed method reaches the stable point in terms of the objective function value within about 150 iterations. These results demonstrate the proposed method has good convergency property.



(a) Yale B (b) AR (c) COIL 20

Fig. 2. Convergency and classification accuracy versus the number of iterations using the proposed method on Yale B, AR and COIL20 databases.

4. EXPERIMENTS AND ANALYSIS

In this section, we first use some figures to vividly show the classification accuracy for these four parameters and then present the strategy used in this paper for parameter selection. Furthermore, we perform the numerical experiment to assess the performance of WGE_LRP.

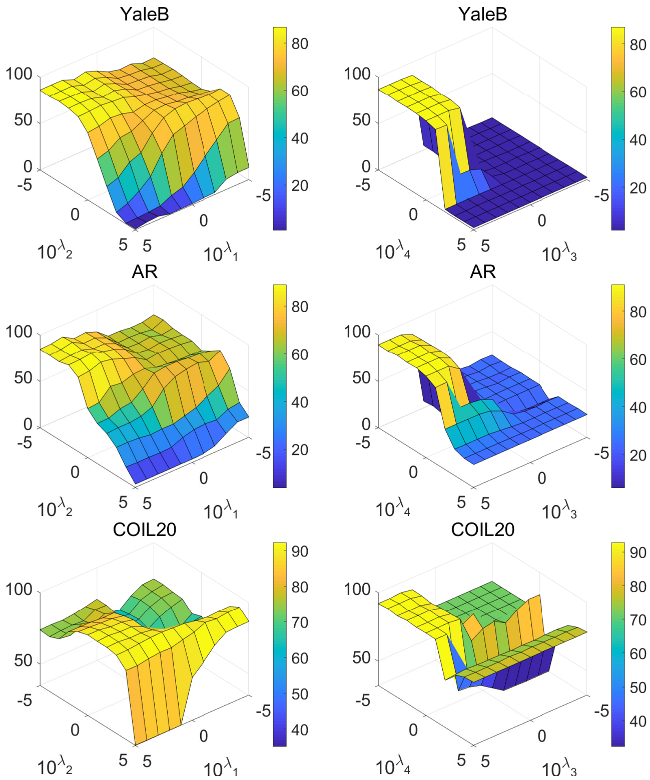


Fig. 3. The relation ship between parameters and classification accuracies of Yale B, AR, COIL20

- **Databases:** Detailed information of three databases are as follows:1)The Yale B [17] database contains 2,414 face images captured from 38 individuals. Each person contains 64 images under different illumination conditions. All images were transformed into gray images with the size of 32×32 in advance. 2)The original AR [18] face database is chosen as a subset that contains 2600 images of 100 subjects. Each color image was transformed into the gray image and then resized to 32×32 in advance. 3)The COIL20 [19] image database contains 20 objects and 1440 gray-scale images in total. We resized each image into a 32×32 matrix by the down-sampling algorithm in advance. In order to improve the efficiency of computation, we further perform PCA on all samples and reduce their dimensions by preserving 98% energy for a fair comparison.
- **Parameter Selection:** We select different combinations of values from the coarse candidate range as values of parameters to perform the proposed method and use some figures to show the relationship between parameters and the classification accuracy. From Fig. 3, we can find that the best combinations of parameters λ_1, λ_2, λ_3 and λ_4 are 10^3, 10^{-5}, 10^3 and 10^{-4} for the extented Yale B database, 10^3, 10^{-5}, 10^3 and 10^{-4} for the AR database, 10^4, 10^1, 10^3 and 10^{-2} for the COIL20 database, respectively.
- **Results and Analysis:** We evaluate our proposed method with some projection learning methods: including PCA [1], LPP [2], OLPP [20], NPE [3], SPP [4], CRP [21], LatLRR [13], LRPP [22], LRPP_GRR [14]. For a fair comparison, we perform all methods 30 times and report their mean classification accuracies (%) with standard deviations. We randomly select samples from each object as training samples (No.) and treat the remaining samples as test samples. We choose the nearest neighbor classifier with Euclidean Distance to obtain their final classification accuracies. All experimental results are reported in the same table, in which bold numbers denote The Best Result.

Table 1. ARR (%) of different methods on the Yale B database

No.	PCA	LPP	OLPP	NPE	SPP	CRP	LatLRR	LRPP	LRPP_GRR	WGE_LRP
10	50.77 (1.28)	56.56 (2.04)	46.44 (1.37)	61.83 (1.71)	84.64 (0.77)	84.89 (0.99)	48.35 (1.40)	50.77 (1.28)	85.03 (0.64)	**86.31 (0.81)**
15	59.35 (1.56)	64.96 (1.92)	56.70 (1.40)	69.75 (1.94)	88.20 (0.78)	88.87 (0.94)	56.14 (1.44)	59.35 (1.56)	88.40 (0.97)	**89.89 (0.84)**
20	65.00 (1.00)	70.52 (1.44)	63.89 (0.95)	75.10 (1.39)	90.08 (0.69)	91.04 (0.64)	61.24 (1.03)	65.00 (1.00)	90.02 (0.79)	**91.85 (0.75)**
25	69.18 (0.98)	74.47 (1.13)	68.09 (1.05)	78.70 (1.08)	91.23 (0.56)	92.00 (0.61)	65.22 (1.06)	69.18 (0.98)	90.69 (0.63)	**92.99 (0.41)**

Table 2. ARR (%) of different methods on the AR database

No.	PCA	LPP	OLPP	NPE	SPP	CRP	LatLRR	LRPP	LRPP_GRR	WGE_LRP
4	26.21 (0.84)	22.89 (1.30)	24.97 (0.84)	24.65 (1.17)	65.72 (1.82)	53.82 (2.15)	26.05 (0.89)	26.21 (0.84)	61.75 (1.40)	**67.71 (1.09)**
6	32.98 (0.89)	30.04 (1.30)	32.29 (0.86)	32.55 (1.50)	73.11 (1.43)	61.61 (1.72)	32.52 (0.87)	32.98 (0.89)	63.67 (1.66)	**75.90 (1.26)**
8	38.13 (1.02)	35.92 (1.13)	37.42 (1.08)	39.06 (1.33)	78.05 (1.49)	67.28 (1.64)	37.42 (0.91)	38.13 (1.02)	68.84 (2.14)	**81.20 (0.84)**
10	42.14 (0.89)	41.93 (1.06)	41.47 (0.89)	44.79 (1.40)	81.44 (1.05)	70.71 (1.35)	41.33 (0.85)	42.14 (0.89)	75.27 (1.39)	**84.37 (1.28)**

Table 3. ARR (%) of different methods on the COIL20 database

No.	PCA	LPP	OLPP	NPE	SPP	CRP	LatLRR	LRPP	LRPP_GRR	WGE_LRP
10	89.60 (0.97)	79.64 (1.31)	89.92 (1.09)	81.55 (1.35)	82.05 (1.29)	85.27 (1.20)	89.43 (0.98)	89.64 (0.92)	65.87 (5.15)	**91.40 (1.10)**
15	93.13 (0.97)	86.41 (1.21)	93.50 (0.94)	88.30 (0.90)	88.56 (1.32)	88.72 (1.33)	93.02 (0.95)	93.13 (0.94)	73.07 (7.25)	**94.60 (0.96)**
20	94.97 (0.93)	90.20 (1.07)	95.21 (0.88)	91.59 (1.05)	91.59 (0.86)	91.66 (1.15)	94.86 (0.92)	94.94 (0.94)	69.92 (8.37)	**96.25 (0.98)**
25	96.64 (0.58)	92.89 (0.82)	96.78 (0.56)	94.07 (0.82)	93.91 (0.80)	94.51 (0.79)	96.55 (0.60)	96.65 (0.59)	71.33 (7.09)	**97.52 (0.63)**

For Yale B, one can see from the Table 1 that the proposed method obtained the highest accuracies compared to the other methods. The result on the AR is reported in Table 2. From Table 2, we can find that all classification accuracy of compared methods is very low, while our method is obtain the highest accuracy. For COIL20, Table 3 shows that the proposed method performs much better than other methods. So our method obtains the best results in three databases.

5. CONCLUSION

In this paper, we presented a novel projection learning method for dimension reduction, in which a novel weighted graph regularization term and a global consistency term are integrated into a low-rank-based method constraint to learn a projection matrix. The proposed method not only captures local and global structures of data but also keeps the consistency of data. Experimental results demonstrate that the proposed method achieves a competitive result than some projection learning methods.

Acknowledgement: This work was supported in part by National Natural Science Foundation of China under Grants numbers of 62106052 and 62072118. It was also supported by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021B1515120010.

6. REFERENCES

[1] Hervé Abdi and Lynne J Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[2] Xiaofei He and Partha Niyogi, "Locality preserving projections," *Advances in neural information processing systems*, vol. 16, no. 16, pp. 153–160, 2004.

[3] Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-Jiang Zhang, "Neighborhood preserving embedding," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. IEEE, 2005, vol. 2, pp. 1208–1213.

[4] Lishan Qiao, Songcan Chen, and Xiaoyang Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[5] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 40–51, 2006.

[6] Xiaolin Xiao, Yongyong Chen, Yue-Jiao Gong, and Yicong Zhou, "Low-rank preserving t-linear projection for robust image feature extraction," *IEEE Transactions on image processing*, vol. 30, pp. 108–120, 2020.

[7] Shuai Yin, Yanfeng Sun, Junbin Gao, Yongli Hu, Boyue Wang, and Baocai Yin, "Robust image representation via low rank locality preserving projection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 15, no. 4, pp. 1–22, 2021.

[8] Shuping Zhao and Bob Zhang, "Learning salient and discriminative descriptor for palmprint feature extraction and identification," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 12, pp. 5219–5230, 2020.

[9] Shuping Zhao, Bob Zhang, and Shuyi Li, "Discriminant and sparsity based least squares regression with l 1 regularization for feature representation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1504–1508.

[10] Jie Wen, Yong Xu, and Hong Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE transactions on cybernetics*, vol. 50, no. 4, pp. 1418–1429, 2018.

[11] Jie Wen, Zheng Zhang, Zhao Zhang, Lunke Fei, and Meng Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE transactions on cybernetics*, vol. 51, no. 1, pp. 101–114, 2020.

[12] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 171–184, 2012.

[13] Guangcan Liu and Shuicheng Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *2011 international conference on computer vision*. IEEE, 2011, pp. 1615–1622.

[14] Jie Wen, Na Han, Xiaozhao Fang, Lunke Fei, Ke Yan, and Shanhua Zhan, "Low-rank preserving projection via graph regularized reconstruction," *IEEE Transactions on Cybernetics*, vol. 49, no. 4, pp. 1279–1291, 2019.

[15] Ming Yin, Shuting Cai, and Junbin Gao, "Robust face recognition via double low-rank matrix recovery for feature extraction," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 3770–3774.

[16] Stephen Boyd, Neal Parikh, and Eric Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011.

[17] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[18] AM MARTINEZ, "The ar face database," *CVC Technical Report*, 1998.

[19] SA NENE, "Columbia object image library (coil-20)," *Technical Report, CUCS-005-96*, 1996.

[20] Deng Cai and Xiaofei He, "Orthogonal locality preserving indexing," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, 2005, pp. 3–10.

[21] Wankou Yang, Zhenyu Wang, and Changyin Sun, "A collaborative representation based projections method for feature extraction," *Pattern Recognition*, vol. 48, no. 1, pp. 20–27, 2015.

[22] Yuwu Lu, Zhihui Lai, Yong Xu, Xuelong Li, David Zhang, and Chun Yuan, "Low-rank preserving projections," *IEEE transactions on cybernetics*, vol. 46, no. 8, pp. 1900–1913, 2015.