Dimensionality Reduction via Representation and Affinity Learning

Miao Qi[1,2], Shuang Lu[1], Xing Huang[1], Lin Yang[1], Chengrong Gong[1], Jianzhong Wang[1,*], Yuanyuan Gu[3]
[1]School of Information Science and Technology
Northeast Normal University
Changchun, China
[2]Key Laboratory for Applied Statistics of MOE
Northeast Normal University
Changchun, China
[3]College of Humanities and information
Changchun University of Technology
Changchun, China

Abstract—**Because graphs are crucial in modeling the data distribution or structure, graph-based dimensionality reduction methods have been widely employed in many machine learning and pattern recognition applications. However, in most existing graph-based dimensionality reduction techniques, the graph is usually specified empirically or learned from raw data, which may deteriorate their performances. In this paper, we propose a novel algorithm named Dimensionality Reduction via Representation and Affinity Learning (DRRAL), which adaptively constructs graph and learns the projection matrix for dimensionality reduction. Our algorithm takes noise and different local structure of each sample into account. Moreover, an efficient optimization strategy based on the iterative updating Augmented Lagrange Multiplier (ALM) and eigenvectors is developed to solve the proposed DRRAL. Extensive experimental results on five databases are carried out to verify the feasibility and effectiveness of proposed approach.**

Keywords-dimensionality reduction; graph construction; adaptive graph

I. INTRODUCTION

With the rapid development of electronic sensors and social media, there is a huge number of high-dimensional data, which can much increases time and space complexities, making the data processing greatly intractable [1-2]. To overcome the problem, dimensionality reduction is designed to reduce the dimension by mapping the high-dimensional data into low-dimensional space, yet retains most of the intrinsic information content in the raw data [3]. Many researchers have proposed various techniques for dimensionality reduction, which can be generally classified as linear and nonlinear algorithms. Conventional linear dimensionality reduction algorithms include Principal Component Analysis (PCA) [4], Linear Discriminant Analysis (LDA) [5] and so on. Nevertheless, since data are not always linear distributed, these linear dimensionality reduction algorithms can't show efficiency all the time. To discover the intrinsic manifold structure of data, many non-linear dimensionality reduction algorithms were developed, such as Isometric Feature Mapping (ISOMAP) [6], Locally Linear Embedding (LLE) [7], Laplacian Eigenmaps (LE) [8]. These algorithms are all based on manifold learning [9] and assume that sampled data lie in a low dimensional nonlinear manifold. However, because there is an implicity of nonlinear map in the above approaches, they usually can't readily yield the test samples' images in the embedding subspaces, In other words, they may suffer from the "out-of-sample" [10-12] problem.

Recently, many manifold learning algorithms which use graph (or affinity matrix) as a tool to preserve some properties of the raw data and show flexibility and effectiveness have been developed. In cases where graph-based dimensionality reduction algorithms have become increasingly popular, graph construction plays an important role in the data structure or distribution. But, the ideal graph is difficult to discover in practice and how to construct a "good" graph become a key issue in learning tasks. The most commonly used way to construct graph is specifying the graphs empirically, e.g., Isometric Projection (IsoProject) [13], Neighborhood Preserving Embedding (NPE) [14] and Locality Preserving Projections (LPP) [15]. These algorithms all utilize some parameters, which are selected empirically [16] or via neighborhood optimization [17]. The improper parameter values may give downturn to effectiveness of the algorithms. What's more, when we set the same neighbor parameter value for all the samples, we may ignore the different local structure of each sample.

To alleviate the harassment of the classical approaches, some novel approaches based on adaptive graph have been put forward for dimensionality reduction recently. Reference [18] developed a Graph-optimized Locality Preserving Projections (GoLPP) algorithm which achieves empirical superiority to traditional LPP, but the graph obtained from GoLPP usually loses required sparsity because its entropy regularization. To achieve a sparse graph, Zhang et al. proposed an algorithm termed Graph Optimization for Dimensionality Reduction with Sparsity Constraints (GODRSC) [19], which achieved state-of-the-art performance. However, due to GODRSC neglected the

*Corresponding author. E-mail address: wangjz019@nenu.edu.cn.
This work is supported by National Natural Science Foundation of China (Nos. 61403078, 61672150, 41671379, 61602221 and 61403400) and Research Fund of Jilin Province Science and Technology Development Project (Nos. 20160204047GX,20170204018GX)

similarity among input data, it cannot guarantee that the nearby high dimensional samples are still close to each other in low-dimensional subspace [20]. Besides, the Dimensionality Reduction with Adaptive Graph (DRAG) algorithm [21] was put forward as a compromise between LPP and GoLPP. Though these algorithms avoid the arbitrary and capricious characteristic of choosing parameters, integrating the graph construction and dimensionality reduction process into a unified framework increases the time complexity of the iterative alternate solution and makes them cannot be directly and conveniently applied to large scale.

Inspired by the success of compressed sensing [22], there's another way to construct graph, i.e., learning it through data representations, e.g. Sample-dependent Graph (SG) [16], Sparsity Preserving Projections (SPP) [23], Low Rank Representation (LRR) [24], Least Squares Regression (LSR) [25] and Nonparametric Discriminant Multi-manifold Learning (NDML) [26]. These algorithms firstly get data representations. Subsequently, the data representations are utilized to calculate the affinity matrix of an undirected graph. At last, the original data are projected into a low dimensional space based on the graph. But the meaning of this kind of affinity is not completely same as the original definition to some extent[27].

In order to solve the limitations of aforementioned algorithms, we propose a novel algorithm termed Dimensionality Reduction via Representation and Affinity Learning (DRRAL) to perform dimensionality reduction task though a two-step process. DRRAL uses the model of [27] and simultaneously learning data representations and their affinity graph in a unified optimization framework to construct graph structures in the first step. We utilize the graph to perform dimensionality reduction task in the second step. Furthermore, extensive classification and clustering experimental results are provided for demonstration.

II. Related works

In this section, we review several algorithms, which utilize data representations to construct graph and are closely related to our proposed method.

Inspired by the success of compressed sensing, many recent works are proposed, which aim to construct the affinity matrix through obtaining the relationship of data representations $R \in \mathrm{R}^{n\times n}$, such as Low Rank Representation (LRR) [24] and Least Square Regression (LSR) [25].

These methods provides a set of n samples $X=\{x_1, x_2, \ldots, x_n\}$, $x_i \in \mathrm{R}^D$ and express each data point x_i as a linear combination of all other data points, i.e., $x_i = \sum_{j\neq i} R_{ij} x_j$. They measure the similarity between points x_i and x_j by utilizing a representational coefficient $(|R_{ij}|+|R_{ji}|)/2$. LRR and LSR's regularizations of R leads to different affinity graphs. The goal of the low-rank LRR, is

$$\min \ \operatorname{rank}(R) \qquad s.t.\ X = XR. \tag{1}$$

where rank(R) is the rank of R.

In [24], instead of rank(R), the nuclear norm is employed . Thus, LRR can be replaced

$$\min \ \operatorname{rank} ||R||_* \qquad s.t.\ X = XR. \tag{2}$$

where $||R||_*$ is the constrained norm of R, which can be regarded as the sum of all the singular values.

LSR minimized the Frobenius-norm of R as

$$\min \ \operatorname{rank} ||R||_F \qquad s.t.\ X = XR. \tag{3}$$

Besides, in [27], Guo analyzed if the data points' representations have a smaller distance, then they may have a bigger probability to be in a same cluster. So he proposed the following constraint based on the assumption

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n} ||R_i - R_j||_F^2 \ A_{ij} \qquad s.t. \quad X = XR. \tag{4}$$

where A is an affinity matrix, in which A_{ij} measures the similarity between data points x_i and x_j. More details can be found in [27].

III. Dimensionality Reduction via Representation and Affinity Learning (DRRAL)

As described previously, graph construction plays an important part in the process of graph-based learning methods. If we determine graph artificially and empirically, we may learn a "bad" graph, which may seriously decrease the performance of methods. Thus, in this subsection, we attempt to construct a meaningful affinity graph in a unified objective function by [27] and utilize the affinity graph to learn projection matrix though an operation like LPP [15].

A. Learning Adaptive Affinity Graph

Given a set of data points $X=\{X_1, X_2, \ldots, X_n\}$ sufficiently sampled from c independent subspaces $\{S_{s=1}^{c}\}$. To settle the data to the actual segmentation of data, we use a hypothesized permutation matrix Γ, that is $X^*=X\Gamma=[X^1, X^2, \ldots, X^c] \in \mathrm{R}^{d\times n}$, where X^s is a collection of n_s data points from the s^{th} subspace Ss with $n=\sum_{s=1}^{c} n_s$. We can use a linear combination of the items in X as $X=XR$ to make the data be self-represented. According to [27], we naturally have the following constrain

$$\min_{\forall i \ A_i^T 1=1, Ai\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{n} ||R_i - R_j||_F^2 \ A_{ij} . \tag{5}$$

where $A \in \mathrm{R}^{n\times n}$ is defined as the data representations' affinity graph, it's based on the distance between their representations R_i and R_j and its element A_{ij} means the probability of the data points X_i and X_j from the same cluster. To constrain the probability property of A_i, we guarantee that $A_i^T 1 = 1$ and $A_i \geq 0$.

However, we can't use (5), because it shall leads to a bad situation, just only the nearest data (or the nearest data representation) is assigned as the neighbor of X_i (or Ri) with probability 1 and all the others with probabilities 0. Besides, to avoid the trivial solution, we minimize $\|A\|_F^2$ similarly to R. Based on the above concerns together[27], we have the following transformation

$$\begin{aligned} &\min\ \lambda_1 \|R\|_F^2 + \lambda_2 tr(RL_A R^T) + \lambda_3 \|A\|_F^2 \\ &s.t.\ X = XR; \forall i\ \ A_i^T 1 = 1; A_i \geq 0. \end{aligned} \tag{6}$$

where D_A can be regarded as a diagonal matrix where the i^{th} diagonal element is $\sum_j A_{ij}$. Moreover, L_A is A's Laplacian matrix, $L_A = D_A - A$. λ_1, λ_2 and λ_3 are three non-negative terms.

In the real life, there is no nearly clean data and the independent subspaces assumption. It is desirable to introduce the noise term $\|E\|_F^2$, then the task can be ultimately formulized as follows[27]

$$\begin{aligned} &\min\ \|E\|_F^2 + \lambda_1 \|R\|_F^2 + \lambda_2 tr(RL_A R^T) + \lambda_3 \|A\|_F^2 \\ &s.t.\ X = XR + E; \forall i\ \ A_i^T 1 = 1; A_i \geq 0. \end{aligned} \tag{7}$$

Although (7) is not jointly convex in the single variable (A, E and R), it is convex with respect to each of them when the other is fixed. To obtain the minimal solution of (7), the Augmented Lagrange Multiplier (ALM) with Alternating Direction Minimizing (ADM) strategy [28] becomes an effective and efficient solver [27]. To separate our objective function and apply ALM-ADM on (7), we make the variable Q to replace R as the additional constraint ($Q=R$) acts as the additional constraint. Then the augmented Lagrangian function of (8) $L\{\forall i\ A_i^T 1 = 1; A_i \geq 0\}$ can be formulized as follows

$$\begin{cases} \|E\|_F^2 + \lambda_1 \|R\|_F^2 + \lambda_2 tr(RL_A R^T) + \lambda_3 \|A\|_F^2 \\ \quad + \Phi(Z_1, X - XR - E) + \Phi(Z_2, Q - R) \end{cases} \tag{8}$$

where $\Phi(Z, C) = \frac{\mu}{2}\|C\|_F^2 + \langle Z, C \rangle$, Z_1 and Z_2 are the Lagrangian multipliers, μ is a positive penalty parameter.

Analogously, after the Lagrangian multipliers were solved, we need to deal with additional four variables (A, R, E and Q. The solver iteratively updates one variable at a time by keeping the others invariant. The more specific solutions of the subproblems and this part of the convergence behavior are shown in [27].

B. Learning Projection

To fulfill DRRAL we take advantage of the model in LPP [15] and seek a mapping matrix W to acquire a low-dimensional out data set Y, $Y = \{y_1, y_2, \ldots, y_n\}$ in R^d (d<<D), which preserves some certain properties of raw samples. We assume that there is a transformation $y_i = w^T x_i$, $y_i \in R^d$ such that y_i is more convenient to be distinguished in the projective subspace. Then, eventual dimensionality reduction task can be performed in (9)

$$\begin{aligned} \frac{1}{2}\sum (y_i - y_j)^2 A_{ij} &= \frac{1}{2}\sum (w^T x_i - w^T x_j)^2 A_{ij} \\ &= w^T X(D - A) X^T w \\ &= w^T XLX^T w \end{aligned} \tag{9}$$

where A is the affinity matrix obtained by (7). D is an order diagonal weight matrix and its entries are column (or row, since A is symmetric) sums A, $D_{ii} = \sum_j A_{ij}$, and $L = D - A$ is the Laplacian matrix. The bigger D_{ii} corresponds to the more significant y_i.

When neighboring points x_i and x_j are mapped far apart from each other, it's better to make the criterion function triggers a heavy penalty. In other words, we hope such a condition, that's if x_i and x_j are close, then y_i and y_j are also close. Such that (9) can be written as

$$\min w^T XLX^T w \qquad s.t. w^T XLX^T w = 1. \tag{10}$$

where the transformation matrix W consists of the d smallest eigenvectors w, i.e. $W = [w_1, w_2, \ldots, w_d]^T$.

Next, we need to deal with the constricted objective function (10) and it can be transformed into the generalized eigenvalue problem[15]

$$XLX^T w = \lambda XDX^T w \tag{11}$$

where λ is the eigenvalue and w represents the corresponding eigenvector. The first d eigenvectors from $w_1, w_2, \ldots, w_d$ (d<<D) are associated with the d smallest eigenvalue. Finally, the dimensionality reduction is formalized as follows:

$$Y = W^T X \qquad W = [w_1, w_2, \ldots, w_d]^T \tag{12}$$

where W is a $D \times d$ projection matrix. Y is a low-dimensional out data set and will be employed as the new features.

IV. EXPERIMENTS

We compared the proposed DRRAL's performance with several classical approaches including LPP [15], NPE [14], GoLPP [18], SGLPP [16], SPP [23], GODRSC [19], LSR-NPE [25], LRR-NPE [24], DRAG [21] and NDML [26] in the experiments.

A. Classification Experiment

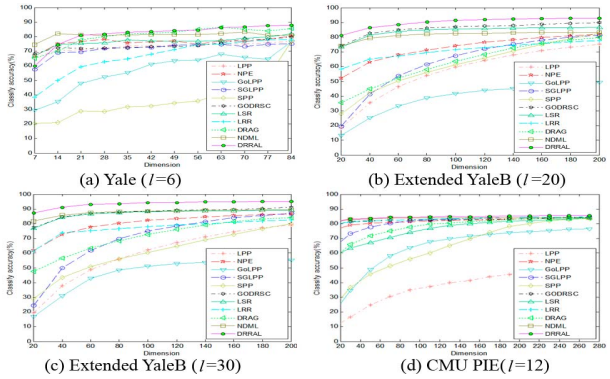
In this subsection, we validated our proposed approach DRRAL on image classification task on four widely used image databases, i.e., Yale [29], Extended YaleB [30] and CMU PIE [31]. We summarize the detailed information of these databases in the Table 1.

TABLE I. THE DETAILED INFORMATION OF ALL DATABASES

Database	Class	Size	Dimension	*l*	*t*
Yale	15	165	1,024	6	5
Extended YaleB	38	2,414	1,024	20/30	44/34

CMU PIE	24	1,632	1,024	12	12

To improve the forecasting accuracy, the experimental data were normalized in advance and we randomly select l samples of each individual to be train set and the rest t samples are for testing in each database. The random sample selection is repeated 10 times for reliability. In all methods involving k nearest neighbor strategy for graph construction, the k value is tuned from the set {3, 5, 7, 9}. For other parameters in the comparative methods, we set their values according to [14-16] [18-19] [23-26]. In our proposed DRRAL, we tune the parameters λ_1, λ_2 and λ_3 from $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$. In addition, to solve the small sample issue, all methods on the used databases keeps 98% image energy by adding a PCA processing phase. The final classification rate is computed by averaging of the all results.



(a) Yale (l=6) (b) Extended YaleB (l=20)

(c) Extended YaleB (l=30) (d) CMU PIE(l=12)

Figure 1. Classification accuracy of various algorithms with different dimensions on different databases.

TABLE II. THE BEST CLASSIFICATION RATES AND STANDARD DEVIATIONS OBTAINED BY VARIOUS METHODS

Methods	Yale (l=6)	Extended YaleB (l=20)	Extended YaleB (l=30)	CMU PIE (l=12)
LPP	76.79±0.03 (84)	75.24±0.18 (200)	79.14±0.19 (200)	50.77±0.13 (280)
NPE	80.53±0.03 (84)	82.00±0.09 (200)	86.77±0.08 (200)	84.03±0.02 (280)
GoLPP	76.40±0.14 (84)	49.16±0.12 (200)	55.09±0.13 (200)	76.50±0.18 (280)
SGLPP	75.07±0.05 (84)	81.39±0.20 (200)	86.85±0.20 (200)	84.36±0.06 (280)
SPP	76.40±0.15 (84)	82.46±0.17 (200)	79.96±0.16 (200)	83.50±0.19 (280)
GODRSC	79.64±0.03 (84)	90.01±0.05 (200)	91.02±0.04 (200)	84.60±0.01 (280)
LSR-NPE	81.87±0.04 (84)	86.61±0.04 (200)	89.16±0.04 (200)	83.73±0.08 (280)
LRR-NPE	77.82±0.12 (84)	76.63±0.06 (180)	82.47±0.06 (200)	84.55±0.01 (280)
DRAG	86.53±0.07 (63)	79.67±0.15 (200)	84.11±0.12 (200)	83.90±0.08 (280)
NDML	83.07±0.02 (70)	83.22±0.03 (200)	89.27±0.02 (200)	84.14±0.01 (220)
DRRAL	**88.00±0.08 (84)**	**92.91±0.04 (200)**	**95.15±0.02 (200)**	**85.65±0.01 (280)**

Note that the number in the parentheses is the feature dimension corresponding to the best result.

The performances of comparative approaches are shown in Fig. 1 and Table 2. We can see that as the size of features dimension increased, the performance of all approaches shows the trend with increase. Additionally, it's clear that the proposed DRRAL is superior to other approaches on all databases. Through dimensionality reduction with the adaptive graph, we obtain refined data which contains more valuable information. To be specific, DRRAL respectively has about 5%, 3%, 4% and 1% improvement for the classification task in these human face data sets with different sizes of training set compared to the second best approach. It directly validates that our proposed DRRAL's effectiveness, which utilizes adaptive graph to dimensionality reduction and better depicts the data's quality.

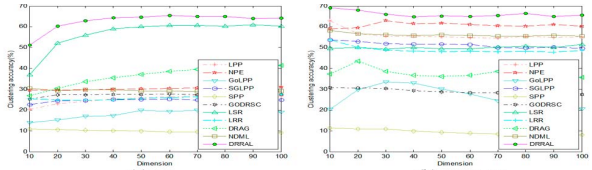
B. Clustering Experiment

In this subsection, the clustering performance of the proposed DRRAL algorithm is compared with other approaches on two classical databases including AR [32] and Columbia Object Image Libraray (COIL 20) [33]. We summarize the detailed information of these databases in the Table 3.

TABLE III. THE DETAILED INFORMATION OF ALL DATABASES

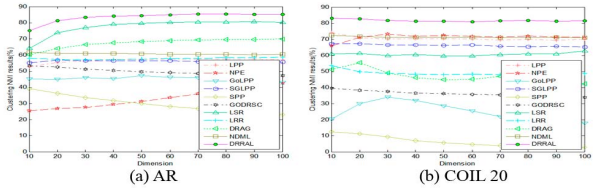
Database	Class	Number of per class	Size	Dimension
AR	100	14	1,400	1,024
COIL 20	20	72	1,440	1,024

We use k-means to cluster the low-dimensional features in clustering experiment due to its simplicity and effectiveness. Here, it should be noted that since the performance of k-means method is a little sensitive to the initialization, we repeat the clustering 10 times with random initialization and finally report the average best results. For the sake of evaluating the clustering performance of different dimensionality reduction algorithms in a sound manner, here, we utilize the clustering accuracy and Normalized Mutual Information (NMI) as two metrics[34].



(a) AR (b) COIL 20

Figure 2. Clustering accuracy of the variation on the databases with different dimensions.



(a) AR (b) COIL 20

Figure 3. NMI results of the variation on the databases with different dimensions.

TABLE IV. THE BEST CLUSTERING ACCURACY, NMI RESULTS AND STANDARD DEVIATIONS OBTAINED BY VARIOUS METHODS

Methods	Clustering Accuracy		NMI	
	AR	*COIL 20*	*AR*	*COIL 20*
LPP	26.98±0.02 (90)	63.04±0.03 (10)	57.61±0.02 (80)	74.16±0.01 (10)
NPE	30.99±0.01 (100)	61.80±0.01 (50)	42.39±0.06 (100)	72.49±0.02 (50)
GoLPP	20.00±0.02 (80)	33.52±0.05 (30)	45.87±0.01 (70)	33.97±0.06 (30)
SGLPP	25.28±0.01 (60)	53.73±0.01 (10)	56.63±0.00 (20)	67.18±0.01 (20)
SPP	10.89±0.01 (10)	11.42±0.01 (10)	39.11±0.05 (10)	12.62±0.04 (10)
GODRSC	27.69±0.01 (60)	30.88±0.01 (10)	53.31±0.02 (10)	39.50±0.02 (10)
LSR-NPE	60.72±0.07 (60)	51.55±0.01 (100)	80.55±0.05 (90)	62.74±0.01 (100)
LRR-NPE	27.46±0.01 (100)	53.68±0.02 (10)	58.53±0.01 (100)	53.68±0.02 (10)
DRAG	41.53±0.05 (100)	43.49±0.02 (20)	69.78±0.01 (100)	55.52±0.04 (20)
NDML	29.81±0.00 (30)	58.32±0.01 (10)	60.91±0.00 (30)	72.37±0.00 (10)
DRRAL	**65.40±0.04 (60)**	**69.00±0.01 (10)**	**85.39±0.03 (60)**	**82.91±0.01 (10)**

Note that the number in the parentheses is the feature dimension corresponding to the best result.

We illustrate the curve of clustering accuracy and NMI results with the variation of dimensions, as shown in Fig. 2 and Fig. 3. Furthermore, all of the best clustering accuracies and NMI results of these methods on AR and COIL 20 databases are presented in Table 4. From these results, we can obtain following main observations:

Firstly, the methods which use adaptive graphs to perform dimensionality reduction including DRAG and the proposed DRRAL can consistently outperform the other comparative approaches in most cases. This implies that adaptive automatically-optimized graphs are more effective than traditional pre-specified graphs, which use the discriminative information during its iteration and can help improve the subsequent clustering performance. Exceptionally, though GoLPP also construct optimized graphs, the graph obtained by GoLPP usually loses sparsity because its entropy regularization, which leads to an inferior results.

Secondly, different metrics bring the slightly different results of clustering experiments, it's because they are from different viewpoints to perform evaluating task. But the overall ranking of each method is nearly consistent and our DRRAL is superior to other methods all the time in the two evaluation systems.

At last, in our experiments, DRRAL achieves more excellent performance than the compared methods and it is can be seen that our DRRAL obtains a wonderful stability from standard deviations. What's more, the feature dimension corresponding to the best result in proposed DRRAL mostly is lower than compared methods, which further justifies that our method's effectiveness for dimensionality reduction.

C. Parameter Analysis

These are mainly three parameters to influence the performance of our proposed DRRAL, including λ_1, λ_2 and λ_3. Due to the space limit, we only use the classification results on Yale database(l=6) to report the parameter sensitivity of our algorithm here. The certain parameter is chosen by a "grid-search" strategy [35] from $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ and fix other parameters. The results in the Fig. 4 (a)-(c) clearly show that DRRAL is robust to the parameters to some extent. Furthermore, the performance of DRRAL would be slightly improved by tuning λ_1, λ_2 and λ_3, but for the simplicity of DRRAL's model, we set λ_1, λ_2 and λ_3 equivalent in the practical experiments. Here, we introduce the parameter t and set $t=\lambda_1=\lambda_2=\lambda_3$. We also explore the sensitivity results of t on the classification performance in terms of different numbers of dimension in the Fig. 4 (d) and it is clear that DRRAL is still efficient and robust through the setting.

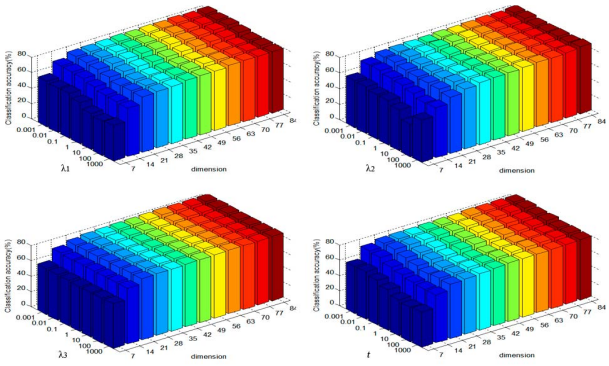


Figure 4. The sensitivity results of our algorithm on Yale database.

V. CONCLUSIONS

In this paper, we develop a novel algorithm termed Dimensionality Reduction via Representation and Affinity Learning (DRRAL) to perform dimensionality reduction task. The idea behind DRRAL is to utilize the model in [27] to construct graph and learn the projection matrix to perform dimensionality reduction task. Compared with the existing dimensionality reduction methods, this paper gets more sparse and flexible graph to perform dimensionality reduction task, which takes different local structure of each sample into consideration and obtains a better subset of features to depicts the data's quality. Furthermore, extensive experiments have been conducted on several publicly available high-dimensional databases to demonstrate the superior performance of our method in comparison with other relevant ones.

REFERENCES

[1] Chao-Ton Su and Yu-Hsiang Hsiao, "Multiclass MTS for Simultaneous Feature Selection and Classification," IEEE Trans. Knowledge and Data Engineering, Vol. 21, pp.192-205, 2009.

[2] L.P. Wang, Y.L. Wang, C. Qing, "Feature selection methods for big data bioinformatics: a survey from the search perspective," Methods, vol.111, pp.21-31, 2016

[3] Xiuju Fu and L.P. Wang, "Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance," IEEE Trans. System, Man, Cybern, Part B-Cybernetics, vol.33, no.3, pp. 399-409, 2003.

[4] I. T. Jolliffe, "Principal component analysis," New York: SpringerVerlag, 1986.

[5] AM. Martinez, "PCA versus LDA," IEEE Trans. PAMI. 2001, pp. 228-233.

[6] J. B. Tenenbaum, "A global geometric framework for nonlinear dimensionality reduction," Science 290 (5500), pp. 2319-2323.

[7] S. T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science 290 (2000), pp. 2323-2326.

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

[9] C. Bregler and S. M. Omohundro, "Nonlinear manifold learning for visual speech recognition," Proceedings of the Fifth International Conference on Computer Vision, 1995.

[10] R. Meese and K. Rogoff, "The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification NBER chapters," International Finance Discussion Papers, 1982, pp. 67-112.

[11] M. B. Nagarajan, M. B. Huber and T. Schlossbauer, "Classification of small, lesions on dynamic breast MRI: Integrating dimensionality reduction and out-of-sample extension into CADx methodology," Artificial Intelligence in Medicine, 2014.

[12] Y. Bengio, J. F. Paiement and P. Vincent, "Out-of-sample extensions for LLE, IsoMap, MDS, Eigenmaps, and Spectral Clustering," Advances in Neural Information Processing Systems, 2004, pp. 177-184.

[13] D. Cai , X. He and J. Han, "Isometric projection," 2006(Dec 1).

[14] X. He, D. Cai and S. Yan, "Neighborhood Preserving Embedding," Tenth IEEE International Conference on Computer Vision. IEEE, 2005, pp. 1208-1213.

[15] X. He and P. Niyogi, "Locality preserving projections," Proc. NIPS, 2003, pp. 153-160.

[16] B. Yang and S. Chen, "Sample-dependent graph construction with application to dimensionality reduction," Neurocomputing, 2010, pp. 301-314

[17] O. Samko, A. Marshall and P. Rosin, "Selection of the optimal parameter value for the Isomap algorithm," Pattern Recognition Letters, 2006, pp. 968-979

[18] L. Zhang, L. Qiao and S. Chen, "Graph-optimized locality preserving projections," Pattern Recognition, 2010, pp. 1993-2002

[19] L. Zhang, S. Chen and L. Qiao, "Graph optimization for dimensionality reduction with sparsity constraints," Pattern Recognition, 2012, pp. 1205-1210.

[20] J. Wang, R. Zhao and Y. Wang, "Locality Constrained Graph Optimization for Dimensionality Reduction," Neurocomputing, 2017.

[21] L. Qiao, L. Zhang and S. Chen, "Dimensionality reduction with adaptive graph," Frontiers of Computer Science, 2013, pp. 745-753.

[22] D. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, 52(4), 2006, pp. 1289‐1306.

[23] L. Qiao, S. Chen and X. Tan, "Sparsity preserving projections with applications to face recognition," Pattern Recognition, 2010, pp. 331-341.

[24] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, pp. 171‐184.

[25] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," In Proceedings of European Conference on Computer Vision, 2012, pp. 347-360.

[26] B. Li, J. Li and X. Zhang, "Nonparametric discriminant multi-manifold learning for dimensionality reduction," Intelligent Computing Theory, 2014, pp. 121-126.

[27] X. Guo, "Robust subspace segmentation by simultaneously learning data representations and their affinity matrix," International Conference on Artificial Intelligence. AAAI Press, 2015, pp. 3547-3553.

[28] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," In Proceedings of Advances in Neural Information Processing Systems (NIPS), 2011, pp. 612-620.

[29] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[30] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

[31] http://vasc.ri.cmu.edu/idb/html/face/

[32] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[33] http://www.cs.columbia.edu/CAVE/coil-20.html

[34] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, 1999, pp. 788-91.

[35] Göçken M, Özçalıcı M, Boru A and Dosdoğru AT, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," Neural Computing and Applications, pp.1-16, 2017