IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. VII (May – Jun. 2015), PP 39-46
www.iosrjournals.org

Semi-Supervised Discriminant Analysis Based On Data Structure

Xuesong Yin*, Rongrong Jiang, Lifei Jiang
Department of Computer Science & Technology, Zhejiang Radio & TV University, Hangzhou, 310030, China

***Abstract**: Dimensionality reduction is a key data-analytic technique for mining high-dimensional data. In this paper, we consider a general problem of learning from pairwise constraints in the form of must-link and cannot-link. As one kind of side information, the must-link constraints imply that a pair of instances belongs to the same class, while the cannot-link constraints compel them to be different classes. Given must-link and cannot-link information, the goal of this paper is learn a smooth and discriminative subspace. Specifically, in order to generate such a subspace, we use pairwise constraints to present an optimization problem, in which a least squares formulation that integrates both global and local structures is considered as a regularization term for dimensionality reduction. Experimental results on benchmark data sets show the effectiveness of the proposed algorithm.*

***Keywords:** Dimensionality reduction; Pairwise constraints; High-dimensional data; Discriminant analysis.*

I. Introduction

Applications in various domains such as face recognition, web mining and image retrieval often lead to very high-dimensional data. Mining and understanding such high-dimensional data is a contemporary challenge due to the curse of dimensionality. However, the current researches [3,4,5,6,13] show that reducing the dimensionality of data has been a feasible method for more precisely charactering such data. Therefore, dimensionality reduction (DR for short) makes sense in many practical problems [2,3,5] since it allows one to represent the data in a lower dimensional space. Based on the availability of the side information, dimensionality reduction algorithms fall into three categories: supervised DR [7,12,28,35], semi-supervised DR [8,16,18,21,22] and unsupervised DR [10,13,14,15,34]. In this paper, we focus on the case of semi-supervised DR. With few constraints or class label information, existing semi-supervised DR algorithms appeal to projecting the observed data onto a low-dimensional manifold, where the margin between data form different classes is maximized. Most algorithms in this category, such as Locality Sensitive Discriminant Analysis [20], Locality Sensitive Semi-supervised Feature Selection [36], and Semi-supervised Discriminant Analysis [26], greatly take the intrinsic geometric structure into account, i.e. local structure of the data, indicating that nearby instances are likely to have the same label. However, global structure of the data, implying that instances on the same structure (typically referred to as a cluster or a manifold) are like to have the same label [1], is ignored in those algorithms. In fact, when mining the data, global structure is the same significance as local structure [1,10,11,29]. Therefore, it is necessary to integrate both global and local structures into the process of DR. The key challenge is how we can incorporate relatively few pairwise constraints into DR such that the represented data can still capture the available class information.

To this end, we propose a semi-supervised discriminant analysis algorithm with integrating both global and local structures (SGLS), which naturally address DR under semi-supervised settings. Given pairwise constraints, the key idea in SGLS is to integrate both global and local structures in semi-supervised discriminant framework so that both discriminant and geometric structures of the data can be accurately captured. The SGLS algorithm has the same flavor as not only supervised DR algorithms, which try to adjust the distance among instances to improve the separability of the data, but unsupervised DR algorithms as well, which make nearby instances in the original space close to each other in the embedding space.

The remainder of the paper is organized as follows. In Section 2, related work on the existing algorithms of semi-supervised DR is discussed. In Section 3, we introduce the general framework of our proposed SGLS algorithm. Section 4 discusses the experimental results on a number of real-world data sets. Finally, we draw conclusions in Section 5.

II. Related Work

Unsupervised DR has attracted considerable interest in the last decade [9,19,23,25,29]. Recently, a few researchers [5,9,13,17] have considered the case that discovers the local geometrical structure of the data manifold when the data lie in a low-dimensional space of the high-dimensional ambient space. Algorithms exploited along this line are used to estimate geometrical and topological properties of the embedding space,

*Corresponding author: Xuesong Yin, yinxs@nuaa.edu.cn

such as Locality Preserving Projections (LPP) [34], Local Discriminant Embedding (LDE) [35], Marginal Fisher Analysis (MFA) [37] and Transductive Component Analysis (TCA) [27]. Different with typical unsupervised DR algorithms, more lately, *K*-means was applied in the dimension-reduced space for avoiding the curse of dimensionality in several algorithms [25,31] which perform clustering and dimensionality reduction simultaneously.

Linear Discriminant Analysis (LDA) [7,12], capturing the global geometric structure of the data by simultaneously maximizing the between-class distance and minimizing the within-class distance, is a well-known approach for supervised DR. It has been used widely in various applications, such as face recognition [5]. However, a major shortcoming of LDA is that it fails to discover the local geometrical structure of the data manifold.

Semi-supervised DR can be considered as a new issue in DR area, which learns from a combination of both the side information and unlabeled data. In general, the side information can be expressed in diverse forms, such as class labels and pairwise constraints. Pairwise constraints include both *must-link* form in which the instances belong to the same class and *cannot-link* form in which the instances belong to different classes. Some semi-supervised DR algorithms [20,26,27,30,36,37] just applied partly labeled instances to maximize the margin between instances form different classes for improving performance. However, in many practical data mining applications, it is fairly expensive to label instances. In contrast, it could be easier for an expert or a user to specify whether some pairs of instances belong to the same class or not. Moreover, the pairwise constraints can be derived from labeled data but not vice versa. Therefore, making use of pairwise constraints to reduce the dimensionality of the data has been an important topic [8,16,33].

III. The SGLS Algorithm

3.1 The Integrating Global and Local structures Framework

Given a set of instances $x_1, \cdots, x_n$ in R^N, we can use a *k*-nearest neighbor graph to model the relationship between nearby data points. Specifically, we put an edge between nodes *i* and *j* if x_i and x_j are close, i.e. x_i and x_j are among *k* nearest neighbors of each other. Let the corresponding weight matrix be *S*, defined by

$$S_{ij} = \begin{cases} \exp(-\dfrac{\left\| x_i - x_j \right\|^2}{t^2}) & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0 & otherwise \end{cases} \tag{1}$$

Where $N_k(x_i)$ denotes the set of *k* nearest neighbors of x_i. In general, if a weighted value S_{ij} is large, x_i and x_j are close to each other in the original space. Thus, in order to preserve the locality structure, the embedding space obtained should be as smooth as possible. To learn an appropriate representation *y* ($y = [y_1, \cdots, y_n]^T$) with the locality structure, it is common to minimize the following objective function [13,34]:

$$\sum_{i,j} \left\| y_i - y_j \right\|^2 S_{ij} = 2y^T Ly \tag{2}$$

where *D* is a diagonal matrix whose diagonal entries are column (or row, since *S* is symmetric) sum of *S*, that is $D_{ii} = \sum_{j=1}^{n} S_{ij}$. $L = D - S$ is the Laplacian matrix.

On the other hand, we expect that the projections $\{a^T x_i\}$ generated by a projection vector $a \in R^d$, which span the final embedding space, approach the sufficiently smooth embeddings $\{y_i\}$ as much as possible. This expectation can be satisfied by the least squares formulation which can incorporate the local structure information via a regularization term defined as in Eq. (2). Mathematically, a projection vector *a* can be acquired by solving the following optimization problem [27]:

$$\min J(a, y) = \left\| X^T a - y \right\|^2 + \lambda_1 y^T Ly \tag{3}$$

where $\lambda_1 > 0$ is the trade-off parameter. Here we assume that the data matrix *X* has been centered.

Taking the derivative of *J(a,y)* with respect to *y* and setting it equals to zero, we get

$$\frac{\partial J(a,y)}{\partial y} = 2(y - X^T a + \lambda_1 L y) = 0 \tag{4}$$

It follows that

$$y^* = (I + \lambda_1 L)^{-1} X^T a \tag{5}$$

where I is the identity matrix of size *n*.

Applying Eq. (5) to Eq. (3), we can eliminate *y* and obtain the following optimization problem with regard to *a*:

$$\begin{aligned} \min J(a) = J(a, y^*) &= y^{*T} (\lambda_1 L)(I + \lambda_1 L) y^* \\ &= a^T X (I + \lambda_1 L)^{-1} (\lambda_1 L) X^T a \\ &= a^T X G X^T a \end{aligned} \tag{6}$$

where *G* is a positive semi-definite matrix of size *n* by *n* and has the following expression:

$$G = (I + \lambda_1 L)^{-1} (\lambda_1 L) \tag{7}$$

We can observe from Eq. (6) that minimizing *J* will bring the smooth subspace where both global and local structures of data are calculated with regard to *a*.

3.2 Discriminat Analysis With Pairwise Constraints

Suppose we are given two sets of pairwise constraints including *must-link* (*M*) and *cannot-link* (*C*). In the previous work [32,38,40,41], such constraints were used for learning an adaptive metric between the prototypes of instances. However, recent research has shown that the distance metric learned on high-dimensional space is relatively unreliable [33,39]. Instead of using constraint-guided metric learning, in this paper our goal is to use pairwise constraints to find a projection vector such that in the embedding space the distance between any pair of instances involved in the *cannot-link* constraints is maximized while that between any pair of instances involved *must-link* constraints is minimized.

If the projection vector *a* is acquired, we can introduce such a transformation:

$$y_i = a^T x_i \tag{8}$$

Under this transformation, we calculate the following sum of the squared distances of the point pairs in *M*:

$$\begin{aligned} d_w &= \sum_{(x_i, x_j) \in M} (a^T x_i - a^T x_j)^T (a^T x_i - a^T x_j) \\ &= a^T S_w a \end{aligned} \tag{9}$$

where S_w is the covariance matrix of the point pairs in *M*:

$$S_w = \sum_{(x_i, x_j) \in M} (x_i - x_j)(x_i - x_j)^T \tag{10}$$

Correspondingly, for the point pairs in *C*, we have

$$d_b = a^T S_b a \tag{11}$$

where S_b has the following expression form:

$$S_b = \sum_{(x_i, x_j) \in C} (x_i - x_j)(x_i - x_j)^T \tag{12}$$

Similar to LDA that seeks directions on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other, we try to maximize d_b and minimize d_w. Thus, we can optimize the following problem to obtain the projection direction:

$$\max_a \frac{a^T S_b a}{a^T S_w a} \tag{13}$$

3.3 The Optimization Framework

So far, we pursue the projection direction *a* from two formulations: (1) we introduce a least squares formulation for finding *a*, which facilities the integration of global and local structures; (2) we obtain the projection vector by maximizing the *cannot-link* covariance and simultaneously minimizing the *must-link* covariance. When the first formulation is considered as a regularization term, we arrive at the following optimization problem:

$$\max_a \frac{a^T S_b a}{(1-\lambda_2)a^T S_w a + \lambda_2 J(a)} \tag{14}$$

where $0 \le \lambda_2 \le 1$ controls the smoothness of estimator. The optimal projection vector *a* that maximizes the objective function (14) is acquired by the maximum eigenvalue solution to the following generalized eigenvalue problem.

$$S_b a = \lambda((1-\lambda_2)S_w + \lambda_2 XGX^T)a \tag{15}$$

In real-world applications, we usually need more projection vectors to span a subspace rather than just one. Let the matrix charactering such a target subspace be $A = [a_1, \cdots, a_d]$ which is formed by the eigenvectors associated with the *d* largest eigenvalues of Eq. (15). Also, it is worth noting that since our method is linear, it has a direct out-of-sample extension to novel sample *x*, i.e. $A^T x$.

Based on the analysis above, we propose to develop a semi-supervised dimensionality reduction algorithm which integrates both global and local structures defined in Eq. (14). The corresponding algorithm, called SGLS is presented in Algorithm 1.

Algorithm 1

Input: X, M, C.

Output: A projection matrix $A \in R^{N \times d}$.

1 Construct an unsupervised graph on all *n* instances to calculate *S* as in Eq. (2) and *L* = *D* – *S*.

2 Present a least squares formulation together with the local structure information to compute matrix *G* as in Eq. (7).

3 Compute S_w and S_b as in Eqs. (10) and (12).

4 Compute a generalized eigenvalue problem Eq. (15) and let the solved eigenvectors be $a_1, \cdots, a_d$ in an order of decreasing eigenvalues.

5 Output $A = [a_1, \cdots, a_d]$.

To get a stable solution of the eigen-problem in Eq. (15), the matrix is required to be non-singular which is not true when the number of features is larger than the number of instances, i.e. $N > n$. In such case, we can apply the Tikhonov regularization technique [24,26] to improve the estimation. Thus, the generalized eigen-problem in this paper can be rewritten as:

$$S_b a = \lambda((1-\lambda_2)S_w + \lambda_2 XGX^T + \beta I_N)a \tag{16}$$

where I_N is the identity matrix of size *N* and $\beta > 0$ is a regularization parameter.

IV. Experiments

In this section, we study the performance of our SGLS algorithm in terms of clustering accuracy on several real-world data sets from different application domains. We set $k = 5$ and $t = 1$ for the construction of the similarity matrix *S* defined in Eq. (1).

4.1 Experimental Setup

In order to verify the efficacy of the proposed algorithm, we compare it with the state-of-the-art, semi-supervised dimensionality reduction algorithms including SSDR [22], SCREEN [33], LADKM [31] and LMDM [16]. We implemented SGLS in the MATLAB environment. All experiments were conducted on a PENTIUM DUAL 3G PC with 2.0 GB RAM. We compared all algorithms on seven benchmark data sets, including Segment and Digits from UCI machine Learning Repository, three document data sets DOC1, DOC2 and DOC3 from the 20-Newgroups data, and two image data sets : USPS handwritten data and Yale Face B (YALEB). The statistics of all data sets are summarized in Table 1.

Table 1: Summary of the test datasets used in our experiment

Data sets	Instance	Dimension	Class
Segment	2309	19	10
Digits	6300	16	10
DOC1	4000	4985	6
DOC2	5200	3132	5
DOC3	6500	6466	9
USPS	7122	256	10
YALEB	3110	2500	10

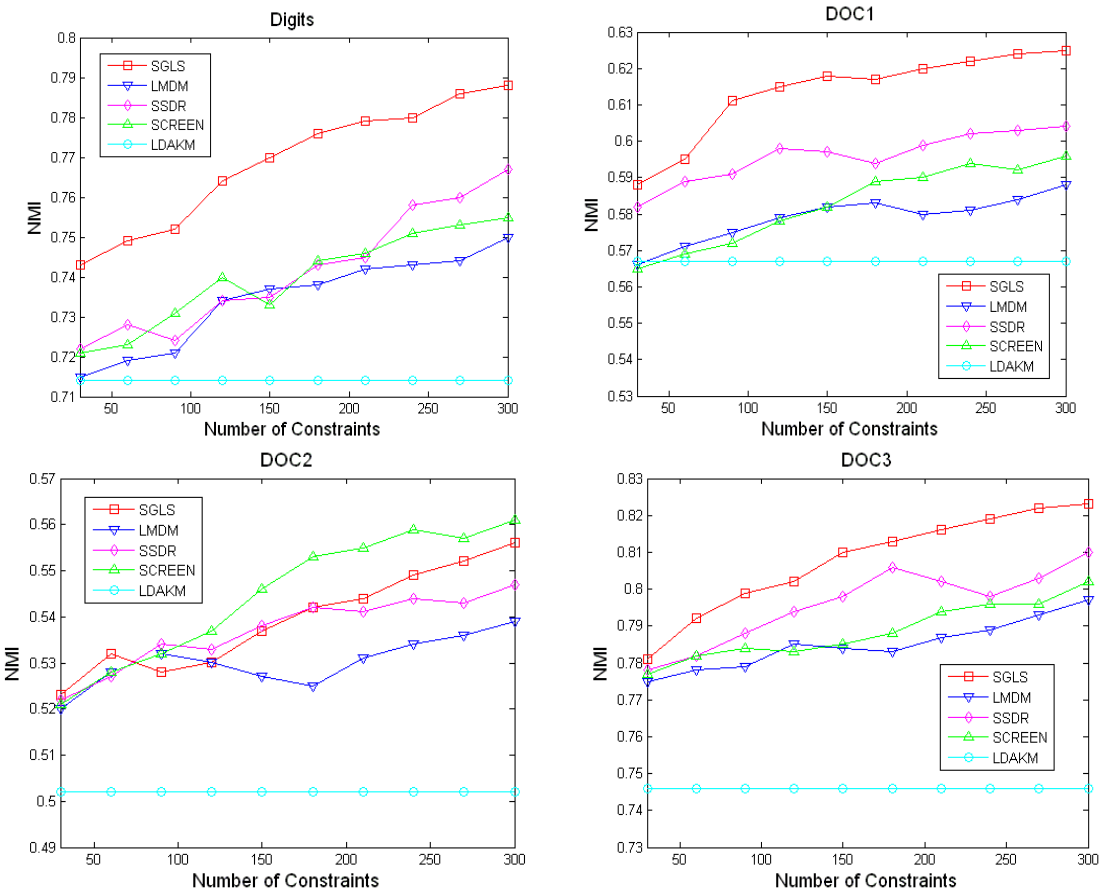
As the unified clustering platform, we use *K*-means algorithm as the clustering method to compare several algorithms after dimensionality reduction. For each data set, we ran different algorithms for 20 times and the comparison was based on the average performance. The parameter in SGLS are always set to $\lambda_1 = 10$ and $\lambda_2 = 0.8$ if without extra explanation. Moreover, for the fair comparison, the dimensionality of all the algorithms is set to *K*-1 (*K* is the number of clustering).

To evaluate the clustering performance, in this paper, we employ normalized mutual information (NMI) as the clustering validation measure, which is widely used to verify the performance of clustering algorithms [8,25,33]. Then, NMI is defined as

$$NMI = \frac{I(X,Y)}{\sqrt{H(X)H(Y)}} \tag{17}$$

where *X* and *Y* denote the random variables of cluster memberships from the ground truth and the output of clustering algorithm, respectively. $I(X,Y)$ is the mutual information between *X* and *Y*. $H(X)$ and $H(Y)$ are the entropies of *X* and *Y* respectively.

4.2 Experimental Results



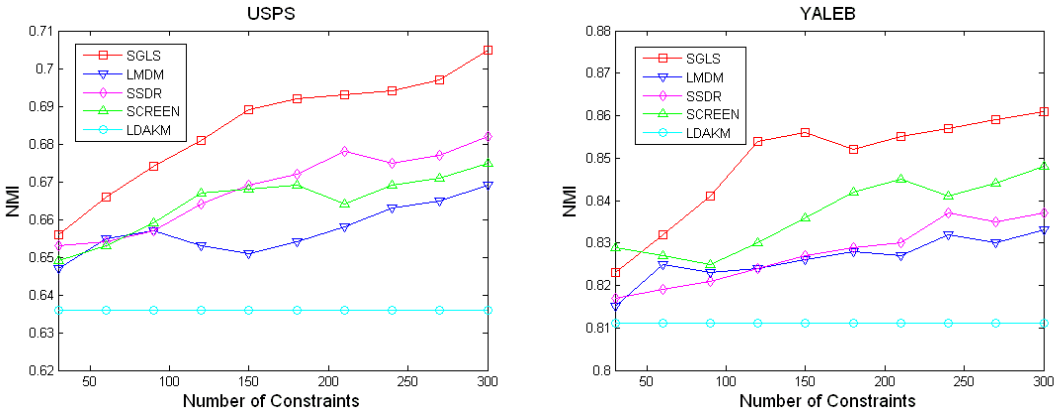


Figure 1 Clustering performance on 6 data sets with different number of constraints

Table 2 presents the NMI results of various algorithms on all seven data sets. Also, Figure 1 shows the clustering performance of standard *K*-means applied to the projected data by different dimensionality reduction algorithms with different numbers of pairwise constraints. As can be seen, our SGLS algorithm achieves the best performance on Digits, DOC1, DOC3, USPS, and YALEB data sets. In fact, the performance of our algorithm is always comparable to that of the other algorithm on the rest of all seven data sets used in this paper. Specifically, we can obtain the following main observations from Table 2 and Figure 1:

(1) The SGLS algorithm significantly outperforms not only LMDM and LDAKM on all of the seven data sets used in the experiment, but SCREEN and SSDR on the six data sets as well. This can be contributed to such a fact that integrating both global and local structures in SGLS may be beneficial.

(2) It is interesting to note that SGLS slightly underperform both SSDR on Segment and SCREEN on DOC2. This implies that in certain case such as Segment, global and local structures may capture similar information so that integrating such both structures seems to hardly help in the dimensionality reduction process.

(3) It is clear from the presented results that LDAKM performs relatively worse that the other algorithms. This can be explained that LDAKM does not employ any side information to pursue the projection direction, which only makes use of the abundant unlabeled instances. Thus, LDAKM has a greatly inferior performance.

(4) Due to the fact that SCREEN and LMDM just pairwise constraints to learn an optimal subspace, their performances are worse that those of SSDR and SGLS respectively. In fact, a large number of the unlabeled data play an important role for semi-supervised dimensionality reduction [8,26,27].

(5) Since SSDR can apply both the pairwise constraints and the unlabeled data to learn a subspace, it can provide a relatively satisfying clustering performance in comparison with LDAKM, SCREEN and LMDM. Compared with our SGLS, however, SSDR has still provided the fairly inferior performance. Actually, SSDR can be considered as the constrained Principal Component Analysis (PCA). Thus, it fails to preserve local neighborhood structures of the instances in the reduced low-dimensional space.

Table 2: NMI comparisons on seven data sets

Data sets	LDAKM	SCREEN	SSDR	LMDM	SGLS
Segment	0.687	0.723	**0.744**	0.712	0.737
Digits	0.714	0.755	0.767	0.750	**0.788**
DOC1	0.567	0.596	0.604	0.588	**0.625**
DOC2	0.502	**0.561**	0.547	0.539	0.556
DOC3	0.746	0.802	0.810	0.797	**0.823**
USPS	0.636	0.675	0.682	0.669	**0.705**
YALEB	0.811	0.848	0.837	0.833	**0.861**

V. Conclusions

In this paper, we propose a novel semi-supervised dimensionality reduction algorithm, called SGLS. For dimensionality reduction and clustering, SGLS integrates both global and local geometric structures so that it can be naturally extended to deal with the abundant labeled data, as graph Laplacian is defined on all the instances. Compared with the state-of-the-art dimensionality reduction algorithms on a collection of benchmark data sets in terms of clustering accuracy, SGLS can always achieve a clear performance gain. Extensive experiments exhibit the advantages of the novel semi-supervised clustering method of SGLS+*K*-means.

Acknowledgments

The research reported in this paper has been partially supported by Natural Science Foundation of Zhejiang Province under Grant No. Y1100349, Public Project of Zhejiang Province under Grant No. 2013C33087, Academic Climbing Project of Young Academic Leaders of Zhejiang Province No. pd2013446.

References

[1]. D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. In Advances in Neural Information Processing Systems 16, 2003, pp. 321-328.
[2]. P. Garcia-Laencina, G. Rodriguez-Bermudez, J. Roca-Dorda. Exploring dimensionality reduction of EEG features in motor imagery task classification, Expert Systems with Applications, 41 (11) (2014)5285-5295.
[3]. Y. Lin, T. Liu, C. Fuh Dimensionality Reduction for Data in Multiple Feature Representations. In Advances in Neural Information Processing Systems 21, 2008, pp. 961-968.
[4]. X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-Supervised Learning using Gaussian Fields and Harmonic Functions, Proc. Twentieth Int'l Conf. Machine Learning, 2003, pp. 912-919.
[5]. X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328-340.
[6]. S. Roweis, L. Saul. Nonlinear dimensionality reduction by locally linear embedding, Science, 290 (5500) (2000) 2323-2326.
[7]. R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Wiley-Interscience, Hoboken, NJ, 2nd edition, 2000.
[8]. X. Yin, T. Shu, Q. Huang. Semi-supervised fuzzy clustering with metric learning and entropy regularization. Knowledge-Based Systems, 2012, 35(11):304-311.
[9]. M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation, 15(6) (2003) 1373-1396.
[10]. J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500) (2000) 2319-2323.
[11]. F. Tsai. Dimensionality reduction techniques for blog visualization, Expert Systems with Applications, 38(3) (2011) 2766-2773.
[12]. J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, Z. Jin, KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 230-244.
[13]. M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Advances in Neural Information Processing Systems 14, 2001, pp. 585-591.
[14]. D. Cai, X. He, J. Han, Document clustering using locality preserving indexing, IEEE Trans. Knowl. Data Eng. 17 (12) (2005) 1624-1637.
[15]. J. He, L. Ding, L. Jiang, Z. Li, Q. Hu. Intrinsic dimensionality estimation based on manifold assumption, Journal of Visual Communication and Image Representation, 25(5)(2014) 740-747.
[16]. S. Xiang, F. Nie, and C. Zhang. Learning a Mahalanobis distance metric for data clustering and classification. Pattern Recognition, 41(12) (2008) 3600 - 3612.
[17]. D. Cai, X. He, J. Han, H.-J. Zhang, Orthogonal laplacianfaces for face recognition, IEEE Trans. Image Process. 15 (11) (2006) 3608-3614.
[18]. Z. Zhang, M. Zhao, T. Chow. Marginal semi-supervised sub-manifold projections with informative constraints for dimensionality reduction and recognition, Neural Networks, 36(2012) 97-111.
[19]. J. Chen, J. Ye, and Q. Li. Integrating Global and Local Structures: A Least Squares Framework for Dimensionality Reduction, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-8.
[20]. D. Cai, X. He, K. Zhou, J. Han, H. Bao, Locality sensitive discriminant analysis, in: International Joint Conference on Artificial Intelligence, 2007, pp. 708-713.
[21]. W. Du, K. Inoue, K. Urahama, Dimensionality reduction for semi-supervised face recognition, in: Proceedings of the Fuzzy Systems and Knowledge Discovery, 2005, pp. 1-10.
[22]. D. Zhang, Z.-H. Zhou, S. Chen, Semi-supervised dimensionality reduction, in: SIAM Conference on Data Mining (SDM), 2007, pp. 629-634.
[23]. X. Yang, H. Fu, H. Zha, J. Barlow, Semi-supervised nonlinear dimensionality reduction, in: Proceedings of the International Conference on Machine Learning, 2006, pp. 1065-1072.
[24]. M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, J. Mach. Learning Res. 1 (1) (2006) 1-48.
[25]. J. Ye, Z. Zhao, H. Liu. Adaptive distance metric learning for clustering. In: Proc. of the Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-7.
[26]. D. Cai, X. He, J. Han, Semi-supervised discriminant analysis, in: Proceedings of the International Conference on Computer Vision, 2007, pp. 1-7.
[27]. W. Liu, D. Tao, and J. Liu. Transductive Component Analysis, in: Proceedings of the International Conference on Data Mining, 2008, pp. 433-442.
[28]. P. Han, A. Jin, A. Kar. Kernel Discriminant Embedding in face recognition, Journal of Visual Communication and Image Representation, 22(7) (2011) 634-642.
[29]. J. Ye. Least Squares Linear Discriminant Analysis, in: Proceedings of International Conference on Machine Learning, 2007, pp. 1087-1093.
[30]. X. Chen, S. Chen, H. Xue, X. Zhou. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data, Pattern Recognition, 45(5) (2012) 2005-2018.
[31]. C. Ding, T. Li. Adaptive dimension reduction using discriminant analysis and K-means clustering. In: Proceedings of the International Conference on Machine Learning, 2007, pp. 521-528.
[32]. E. Xing, A. Ng, M. Jordan, S. Russell. Distance metric learning with application to clustering with side-information. In: Advances in Neural Information Processing Systems 15, 2003, pp. 505-512.
[33]. W. Tang, H. Xiong, S. Zhong, J. Wu. Enhancing semi-supervised clustering: a feature projection perspective. In: Proceedings of the International Conference on Knowledge Discovery and Data Mining, 2007, pp. 707-716.
[34]. X. He, P. Niyogi. Locality preserving projections, In: Advances in Neural Information Processing Systems 16, 2003.
[35]. H. Chen, H. Chang, and T. Liu. Local discriminant embedding and its variants. In: Proceedings of the Internal Conference on Computer Vision and Pattern Recognition, 2005, pp.846-853.
[36]. J. Zhao, K. Lu, X. He, Locality sensitive semi-supervised feature selection, Neurocomputing, 71 (10-12), (2008)1842-1849.

[37]. S. Yan, D. Xu, B. Zhang, H.J. Zhang, Q. Yang, and S. Lin. Graph embedding and extension: A general framework for dimensionality reduction. IEEE Transactions on PAMI, 29(1) (2007) 40-51.
[38]. B. Kulis, S. Basu, I. Dhillon, and R. Mooney. Semi-supervised graph clustering: a kernel approach. In: Proceedings of the International Conference on Machine Learning, 2005, pp.59-68.
[39]. D. Meng, Y. Leung, Z. Xu. Passage method for nonlinear dimensionality reduction of data on multi-cluster manifolds, Pattern Recognition, 46 (8) (2013) 2175-2186.
[40]. M. Bilenko, S. Basu, R.J. Mooney, Integrating constraints and metric learning in semi-supervised clustering, in: Proceedings of International Conference on Machine Learning, 2004, pp. 81–88.
[41]. D.Y. Yeung, H. Chang, Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints, Pattern Recognition 39 (5) (2006) 1007–1010.