

Feature Combination Methods Based on Graphs

ZHENFENG ZHU

Liaocheng University
School of Mathematics Sciences
Liaocheng, Shandong, 252059
CHINA
zhzhf1987@163.com

LIYA FAN*

Liaocheng University
School of Mathematics Sciences
Liaocheng, Shandong, 252059
CHINA
fanliya63@126.com

Abstract: In pattern recognition, graph-based feature combination techniques attract many researchers to study this field. In this paper, we construct a unified framework based on graphs (GF), and derive that FDA, PCA, LPP, DLPP, MFA and MMC are special cases of GF, and then three new algorithms are proposed for GF, which are regularized GF (RGF), GF based on null space (NGF) and GF based on singular value decomposition (GF/SVD). Experiments are made on AVIRIS remote sensing image to illustrate the efficient and effective of our algorithms. The results show that the effects of proposed algorithms are very significantly.

Key-Words: Framework based on graphs (GF), regularized GF, GF based on null space, GF based on singular value decomposition

1 Introduction

Feature combination is one of the most important parts of pattern recognition and machine learning, which makes computation cost decrease and classification performance increase. In recent years, a great amount of efforts have been spent in the research of local or global linear feature combination techniques, such as Fisher discriminant analysis (FDA) [1-9], principal component analysis (PCA) [10], locality preserving projection (LPP) [11-12], discriminant locality preserving projection (DLPP) [13], marginal Fisher analysis (MFA) [14-15], maximum margin criterion (MMC) [13,16-17] and so on. PCA and FDA are global linear subspace learning methods and LPP, DLPP, MFA and MMC are local methods. FDA aims to seek optimal linear directions such that the Fisher criterion of the between-class scatter versus the within-class scatter is maximized. PCA seeks an optimal transformation to map data points to a low-dimension subspace and in which preserve the construction of data's total covariance as possible. LPP aims to preserve the local similarity of data, that is, when adjacent data points are projected to the low-dimensional subspace, the corresponding adjacent property is preserved. DLPP aims at preserving the neighbor relationship of samples which share the same class label. The idea of MFA is to make the distances of data points with neighbor relationship in the same class are close to each other, while data points of the different class with neighbor relationship are far

from each other. MMC aims at getting a feature subspace and the margin between classes is maximized by mapping the original data to that low-dimension feature space.

In this paper, although some of the above methods are not proposed based on graphs, we attempt to make an explanation with graph theory about these methods. We propose a unified framework based on graphs (GF) and show that FDA, PCA, LPP, DLPP, MFA and MMC are special cases of GF by means of different selection of weighting matrixes in GF. In addition, in order to solve the singularity problem of weighting matrixes, we propose three algorithms for GF, regularized GF (RGF), GF based on null space (NGF) and GF based on singular value decomposition (GF/SVD), respectively. We make experiments on AVIRIS remote sensing image to illustrate the efficient and effective of our new methods.

The rest of this paper is organized as follows. GF is proposed in Section 2. FDA, PCA, LPP, DLPP, MFA and MMC are expressed as special cases of GF in Section 3. In Section 4, three new algorithms for solving GF are proposed. Experiments on AVIRIS remote sensing image with new algorithms are made in Section 5, the results demonstrate the effectiveness of proposed algorithms. Section 6 concludes the paper.

2 Framework Based on Graphs (GF)

In this section, we introduce a framework based on graphs for feature combination of high dimensional

*Corresponding author.

data. Given a data set $\{x_1, \dots, x_n\} \subseteq R^m$ with c classes and let $X = [x_1, \dots, x_n] \in R^{m \times n}$. The aim of feature combination is to reduce the dimension of data, that is to find a matrix $A \in R^{m \times d} (d \ll m)$ such that $y = A^T x \in R^d$.

In order to construct a framework based on graphs, we first introduce the constructions of graphs. Generally, a graph is made by using adjacency relationship. We construct graphs by weighting each pairs of data points. By means of different definition methods of weights, we can get two different weight matrixes B and C that are symmetric $n \times n$ matrixes. Let $G_b = \{X, B\}$ denote an intrinsic graph with vertex set $\{x_1, \dots, x_n\}$ and affinity matrix B and $G_c = \{X, C\}$ a constraint graph with the same vertex set as that of G_b and constraint matrix C . The graphs G_b and G_c are defined to characterize certain statistical or geometric properties of the vertices x_1, \dots, x_n . The purpose of graph embedding is reduce the high dimensional vertices into low dimensional space that preserve the similarities which are measured by the edge weights of the matrix B and suppress the similarities which are measured by the edge weights of the matrix C . The general framework of graph embedding is to find the optimal $w = (w_1, \dots, w_n)^T \in R^n$ by maximizing the following objective function (for details, see [18]):

$$\max_w \frac{w^T B w}{w^T C w}. \quad (1)$$

Let $a \in R^m$ be a projection direction and $w_i = a^T x_i, i = 1, \dots, n$, then $w = X^T a$ and the problem (1) becomes

$$\max_a \frac{a^T X B X^T a}{a^T X C X^T a}. \quad (2)$$

We can get the optimal projection direction a as the eigenvector corresponding to the maximum eigenvalue of the following generalized eigen-problem when $X C X^T$ is nonsingular:

$$X B X^T a = \lambda X C X^T a.$$

3 Applications of GF for feature combination

In this section, we show that FDA, PCA, LPP, DLPP, MFA and MMC can be unified into GF by means of the different choices of affinity matrix B and constraint matrix C . Let $A = [a_1, \dots, a_d] \in R^{m \times d}$, $y_i = A^T x_i, i = 1, \dots, n$ and $Y = A^T X \in R^{d \times n}$. Let l_k be the number of samples in the class k and $x_i^{(k)}$ be the i th sample in the class k . Let $X^{(k)} = [x_1^{(k)}, \dots, x_{l_k}^{(k)}], X = [X^{(1)}, \dots, X^{(c)}]$ and

$n = \sum_{k=1}^c l_k$. Let $\mu^{(k)} = \frac{1}{l_k} \sum_{i=1}^{l_k} x_i^{(k)}$ and $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ be the means of the data in class k and all data, respectively.

3.1 FDA Underlying GF (GFDA)

Let S_w, S_b and S_t denote the within-class, between-class and total scatter matrixes, respectively. They are defined by

$$\begin{aligned} S_w &= \sum_{k=1}^c \sum_{i=1}^{l_k} (x_i^{(k)} - \mu^{(k)})(x_i^{(k)} - \mu^{(k)})^T, \\ S_b &= \sum_{k=1}^c l_k (\mu^{(k)} - \mu)(\mu^{(k)} - \mu)^T, \\ S_t &= \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T. \end{aligned}$$

The idea of FDA is to seek d directions by the following Fisher criterion on which the data points of different class are far from each other while data points of the same class are close to each other:

$$\max_a \frac{a^T S_b a}{a^T S_w a}. \quad (3)$$

We can show that the problem (3) is equivalent to

$$\max_a \frac{a^T S_b a}{a^T S_t a}. \quad (4)$$

Let $W^{(k)}$ be an $l_k \times l_k$ matrix with all the elements equal to $1/l_k$ and define an $n \times n$ matrix W by

$$W = \begin{bmatrix} W^{(1)} & 0 & \dots & 0 \\ 0 & W^{(2)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W^{(c)} \end{bmatrix}.$$

If we let $B = W - (1/n)ee^T$ and $C = I - (1/n)ee^T$, where $e = [1, \dots, 1]^T \in R^n$ and $I \in R^{n \times n}$ is an identity matrix, then

$$\begin{aligned} X B X^T &= X(W - (1/n)ee^T)X^T \\ &= \sum_{k=1}^c X^{(k)} W^{(k)} (X^{(k)})^T \\ &\quad - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n x_i^T \\ &= \sum_{k=1}^c l_k \mu^{(k)} (\mu^{(k)})^T - n \mu \mu^T \\ &= \sum_{k=1}^c (\mu^{(k)} - \mu)(\mu^{(k)} - \mu)^T \\ &= S_b, \\ X C X^T &= X(I - (1/n)ee^T)X^T \\ &= \sum_{k=1}^c X^{(k)} (X^{(k)})^T - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n x_i^T \\ &= \sum_{k=1}^c l_k \sum_{i=1}^{l_k} x_i^{(k)} (x_i^{(k)})^T - n \mu \mu^T \\ &= \sum_{k=1}^c \sum_{i=1}^{l_k} (x_i^{(k)} - \mu)(x_i^{(k)} - \mu)^T \\ &= S_t. \end{aligned}$$

Consequently, the problem (4) can be presented as

$$\max_a \frac{a^T X B X^T a}{a^T X C X^T a},$$

which indicates that FDA is a special case of GF.

3.2 LPP Underlying GF (GLPP)

LPP aims to preserve the local similarity of data, that is, when adjacent data points are projected to a low-dimensional subspace, the corresponding adjacent property is preserved [11-12]. The criterion of LPP for seeking an optimal transformation is to minimizing the following objective function

$$\min_{i,j} \sum_{i,j} \|y_i - y_j\|^2 p_{ij}, \quad (5)$$

where $P = [p_{ij}]_{n \times n}$ is a similarity matrix defined as follows

$$p_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}, & x_i \in N_k(x_j) \text{ or} \\ & x_j \in N_k(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $N_k(x_i)$ denotes the set of k nearest neighbor of x_i and $\sigma > 0$ is a given parameter. Due to

$$\begin{aligned} & \sum_{i,j} \|y_i - y_j\|^2 p_{ij} \\ &= \sum_{i,j} (y_i - y_j)^T (y_i - y_j) p_{ij} \\ &= 2 \sum_{i,j} y_i^T y_j p_{ij} - 2 \sum_{i,j} y_i^T y_j p_{ij} \\ &= 2 \text{trace}(Y(D - P)Y^T) \\ &= 2 \text{trace}(A^T X L X^T A), \end{aligned}$$

where $D = \text{diag}\{D_{11}, \dots, D_{nn}\}$ is a diagonal matrix, $D_{ii} = \sum_j p_{ij}$ and $L = D - P$ is the Laplacian matrix, the problem (5) is equivalent to

$$\min_A \text{trace}(A^T X L X^T A). \quad (7)$$

Obviously, $A = 0$ is the optimal solution of the problem (7). But we will lose all of the discriminant information with that optimal transformation. So, we can impose a constraint $\text{trace}(A^T X D X^T A) = 1$ and the problem (7) can be rewritten as

$$\begin{aligned} & \min_A \text{trace}(A^T X L X^T A) \\ & \text{s.t. } \text{trace}(A^T X D X^T A) = 1. \end{aligned} \quad (8)$$

The problem (8) is equivalent to

$$\max_A \frac{\text{trace}(A^T X P X^T A)}{\text{trace}(A^T X D X^T A)}. \quad (9)$$

Let a is a projection vector, then the problem (9) is the following optimization problem:

$$\max_a \frac{a^T X P X^T a}{a^T X D X^T a},$$

which indicates that LPP is a special case of GF.

3.3 DLPP Underlying GF (GDLPP)

DLPP is an improvement of LPP. In LPP, neighbor relationships of data points are just preserved while information of classes is not considered, which may enhance the adjacent relationship of points with different classes. To overcome the limitation, Ma et al. proposed DLPP in [13]. The criterion of DLPP is as follows:

$$\min_A \sum_{i,j} \|y_i - y_j\|^2 (p_{ij} - q_{ij}), \quad (10)$$

where p_{ij} is defined in (6) and

$$q_{ij} = \begin{cases} 1, & x_i \in N_{k_2}^-(x_j) \text{ or } x_j \in N_{k_2}^-(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where $N_{k_2}^-(x_i)$ denotes the set of k_2 nearest neighbor of x_i and share the different class label with x_i . Since

$$\begin{aligned} & \sum_{i,j} \|y_i - y_j\|^2 (p_{ij} - q_{ij}) \\ &= 2 \text{trace}(A^T X (D - P) X^T A) \\ & \quad - 2 \text{trace}(A^T X (D' - Q) X^T A) \\ &= 2 \text{trace}(A^T X (\tilde{D} - \tilde{L}) X^T A), \end{aligned}$$

where $P = [p_{ij}]_{n \times n}$, $Q = [q_{ij}]_{n \times n}$, $\tilde{D} = D - D'$, $\tilde{L} = P - Q$, $D = \text{diag}\{\sum_j p_{1j}, \dots, \sum_j p_{nj}\}$ and $D' = \text{diag}\{\sum_j q_{1j}, \dots, \sum_j q_{nj}\}$, the problem (10) is equivalent to

$$\min_A \text{trace}(A^T X (\tilde{D} - \tilde{L}) X^T A). \quad (12)$$

By imposing a constraint $\text{trace}(A^T X \tilde{D} X^T A) = 1$, the problem (12) can be rewritten as

$$\begin{aligned} & \min_A \text{trace}(A^T X (\tilde{D} - \tilde{L}) X^T A) \\ & \text{s.t. } \text{trace}(A^T X \tilde{D} X^T A) = 1, \end{aligned}$$

which is equivalent to

$$\max_A \frac{\text{trace}(A^T X \tilde{L} X^T A)}{\text{trace}(A^T X \tilde{D} X^T A)}, \quad (13)$$

Let a is a projection vector, then the problem (13) is the following optimization problem:

$$\max_a \frac{a^T X \tilde{L} X^T a}{a^T X \tilde{D} X^T a},$$

which shows that DLPP is a special case of GF.

3.4 PCA Underlying GF (GPCA)

PCA is one of the most popular algorithms for feature combination first proposed in [10]. It seeks an optimal transformation to map data points to a low-dimension subspace and preserves the construction of data's total convince as possible. The criterion of PCA is as follows:

$$\begin{aligned} \max_A \sum_{i=1}^n \|y_i - \bar{y}\|^2 \\ \text{s.t. } A^T A = I, \end{aligned} \quad (14)$$

where $I \in R^{d \times d}$ is an identical matrix and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is the mean of all samples in low-dimension space. Since

$$\begin{aligned} & \sum_{i=1}^n \|y_i - \bar{y}\|^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^T (y_i - \bar{y}) \\ &= \sum_{i=1}^n (x_i - \mu)^T A A^T (x_i - \mu) \\ &= \text{trace}(A^T S_t A) \end{aligned}$$

and we have shown that $S_t = X C X^T$ in Section 3.2, the problem (14) is equivalent to

$$\begin{aligned} \max_A \text{trace}(A^T X C X^T A) \\ \text{s.t. } A^T A = I. \end{aligned} \quad (15)$$

For the sake of convenience, we substitute the constraint $A^T A = I$ by $\text{trace}(A^T X X^T A) = 1$, and then the problem (15) can be rewritten as

$$\begin{aligned} \max_A \text{trace}(A^T X C X^T A) \\ \text{s.t. } \text{trace}(A^T X X^T A) = 1, \end{aligned}$$

which is equivalent to

$$\max_A \frac{\text{trace}(A^T X C X^T A)}{\text{trace}(A^T X X^T A)}. \quad (16)$$

Let a is a projection vector, then the problem (16) is the following optimization problem:

$$\max_a \frac{a^T X C X^T a}{a^T X X^T a},$$

which shows that PCA is a special case of GF.

3.5 MFA Underlying GF (GMFA)

MFA utilizes the graph embedding to construct an intrinsic graph and a penalty graph. The former is used to characterize the intraclass compactness, and the latter is designed to characterize the interclass separability (for details, see [14-15]). The intraclass compactness and interclass separability are defined by the

terms S_c and S_p , respectively,

$$\begin{aligned} S_c &= \sum_{i,j} \|A^T x_i - A^T x_j\|^2 W_{ij}^c, \\ S_p &= \sum_{i,j} \|A^T x_i - A^T x_j\|^2 W_{ij}^p, \end{aligned}$$

where

$$\begin{aligned} W_{ij}^c &= \begin{cases} 1, & i \in N_{k_1}^+(j) \text{ or } j \in N_{k_1}^+(i), \\ 0, & \text{otherwise,} \end{cases} \\ W_{ij}^p &= \begin{cases} 1, & i \in N_{k_2}^-(j) \text{ or } j \in N_{k_2}^-(i), \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

$N_{k_1}^+(i)$ is the index set of the k_1 nearest neighbors of the data point x_i in the same class and $N_{k_2}^-(i)$ is the index set of the k_2 nearest neighbors of the data point x_i in which the class labels are different from the class label of x_i . Let $D^c = \text{diag}\{\sum_j W_{1j}^c, \dots, \sum_j W_{nj}^c\}$ and $D^p = \text{diag}\{\sum_j W_{1j}^p, \dots, \sum_j W_{nj}^p\}$, then

$$\begin{aligned} S_c &= \sum_{i,j} \|A^T x_i - A^T x_j\|^2 W_{ij}^c \\ &= 2\text{trace}(A^T X (D^c - W^c) X^T A) \end{aligned}$$

and

$$\begin{aligned} S_p &= \sum_{i,j} \|A^T x_i - A^T x_j\|^2 W_{ij}^p \\ &= 2\text{trace}(A^T X (D^p - W^p) X^T A). \end{aligned}$$

Consequently, the criterion of MFA is as follows:

$$\max_A \frac{\text{trace}(A^T X (D^p - W^p) X^T A)}{\text{trace}(A^T X (D^c - W^c) X^T A)}. \quad (17)$$

Let a is a projection vector, then the problem (17) is the following optimization problem:

$$\max_a \frac{a^T X (D^p - W^p) X^T a}{a^T X (D^c - W^c) X^T a},$$

which shows that MFA is a special case of GF.

3.6 MMC Underlying GF (GMMC)

MMC aims to get a low-dimension feature subspace, in which the margin between classes is maximized (for details, see [13,16]). the margin between classes is defined by

$$d(i, j) = d(\mu^{(i)}, \mu^{(j)}) - \text{trace}(S_i) - \text{trace}(S_j),$$

where $d(\mu^{(i)}, \mu^{(j)}) = \|\mu^{(i)} - \mu^{(j)}\|^2$, $\mu^{(i)}$ and S_i denote the mean vector and covariance matrix of the class i in the low-dimension space, respectively. The criterion of MMC is as follows:

$$\max J = \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j d(i, j), \quad (18)$$

where p_i and p_j are the prior probabilities of class i and class j , respectively. It can be derived that $J = \text{trace}A^T(S_b - S_w)A$ (see [17]), and then the criterion (18) can be rewritten as

$$\max_A \text{trace}(A^T(S_b - S_w)A). \quad (19)$$

By imposing a constraint $\text{trace}(A^T X X^T A) = 1$, the problem (19) can be rewritten as

$$\begin{aligned} &\max_A \text{trace}(A^T(S_b - S_w)A) \\ &s.t. \text{trace}(A^T X X^T A) = 1, \end{aligned}$$

which is equivalent to

$$\max_A \frac{\text{trace}(A^T(2S_b - S_t)A)}{\text{trace}(A^T X X^T A)} \quad (20)$$

since $S_t = S_b + S_w$.

Let a is a projection vector. By substituting $S_b = X B X^T$ and $S_t = X C X^T$ derived in section 3.1 into (20), the problem (20) can be changed into the following optimization problem:

$$\max_a \frac{a^T X (2B - C) X^T a}{a^T X X^T a},$$

which shows that MMC is a special case of GF.

4 New Algorithms for GF

We propose a framework based on graphs (2) in Section 2. Let $S_1 = X B X^T$ and $S_2 = X C X^T$, then the GF (2) can be written as

$$\max_a \frac{a^T S_1 a}{a^T S_2 a}. \quad (21)$$

We often meet the question that S_2 is singular during the way to solve the problem (21). In order to avoid computing the singularity of S_2 , by means of the study of FDA improved algorithms (for details, see [2-6]), we propose three efficient and effective algorithms in this section to solve the problem (21), which are called regularized GF (RGF), GF based on null space (NGF) and GF based on singular value decomposition (GF/SVD), respectively.

4.1 RGF

Let I denote the identical matrix and $\alpha > 0$. We know that $S_2 + \alpha I$ is a nonsingular matrix because that S_2 is a symmetric positive semidefinite matrix. Consequently, we can substitute $a^T S_2 a$ by $a^T (S_2 + \alpha I) a$ in problem (21) and then get the following optimization problem:

$$\max_a \frac{a^T S_1 a}{a^T (S_2 + \alpha I) a}. \quad (22)$$

We can show that the problem (22) can be transformed into the following generalized eigen-equation:

$$S_1 a = \lambda (S_2 + \alpha I) a. \quad (23)$$

By solving the equation (23), we can get the optimal projection a as the eigenvector corresponding to the maximum eigenvalue. The specific algorithm is as follows.

Algorithm 1. RGF

1. Construct the affinity matrix B and constraint matrix C .
2. Compute $S_1 = X B X^T$, and $S_2 = X C X^T$.
3. Solve the generalized eigen-equation (23) with a given $\alpha > 0$ and get the eigenvectors a_1, \dots, a_d corresponding to eigenvalues $\lambda_1 \geq \dots \geq \lambda_d$, respectively.
4. Let $A = [a_1, \dots, a_d]$.

4.2 NGF

In this subsection, in order to avoid the singularity of S_2 in problem (21), we first consider the null space $\text{null}(S_2)$ of S_2 and then consider the range space of S_1 on the $\text{null}(S_2)$.

Let $\text{rank}(S_2) = r_2$. By the singular value decomposition (SVD) of S_2 :

$$S_2 = [U_{21} \ U_{22}] \begin{bmatrix} \Sigma_{21} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_{21}^T \\ U_{22}^T \end{bmatrix}, \quad (24)$$

where $U_{21} \in R^{m \times r_2}$ and $U_{22} \in R^{m \times (m-r_2)}$ are column orthogonal matrices and $\Sigma_{21} \in R^{r_2 \times r_2}$ is a diagonal matrix with non-increasing positive diagonal components, we can get that $\widetilde{S}_2 = U_{22} U_{22}^T S_2 U_{22} U_{22}^T = 0$. It is easy to show that $\text{null}(S_2) = \text{span}(U_{22})$. Let

$$\widetilde{S}_1 = U_{22} U_{22}^T S_1 U_{22} U_{22}^T.$$

Next, we consider the range space $\text{range}(\widetilde{S}_1)$ of \widetilde{S}_1 . Let $\text{rank}(S_1) = r_1$, then $\text{rank}(\widetilde{S}_1) = r_1$. By the SVD of the matrix \widetilde{S}_1 :

$$\widetilde{S}_1 = [\widetilde{U}_{11} \ \widetilde{U}_{12}] \begin{bmatrix} \widetilde{\Sigma}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{U}_{11}^T \\ \widetilde{U}_{12}^T \end{bmatrix}, \quad (25)$$

where $\widetilde{U}_{11} \in R^{m \times r_1}$ and $\widetilde{U}_{12} \in R^{m \times (m-r_1)}$ are column orthogonal matrices and $\widetilde{\Sigma}_{11} \in R^{r_1 \times r_1}$ is a diagonal matrix with non-increasing positive diagonal components, we can get that

$$\widetilde{\widetilde{S}}_1 = \widetilde{U}_{11}^T U_{22} U_{22}^T S_1 U_{22} U_{22}^T \widetilde{U}_{11} = \widetilde{\Sigma}_{11}.$$

It is easy to show that $\text{range}(\widetilde{S}_1) = \text{span}(\widetilde{U}_{11})$ and

$$\widetilde{S}_2 = \widetilde{U}_{11}^T U_{22} U_{22}^T S_2 U_{22} U_{22}^T \widetilde{U}_{11} = 0.$$

Consequently, we can get an optimal solution $U_{22} U_{22}^T \widetilde{U}_{11}$ of the problem (21), which can be assigned as an optimal dimensionality reduction matrix A . The specific algorithm is as follows.

Algorithm 2. NGF

1. Construct the affinity matrix B and constraint matrix C .
2. Compute $S_1 = X B X^T$ and $S_2 = X C X^T$.
3. Compute the SVD (24) of S_2 and let $\widetilde{S}_1 = U_{22} U_{22}^T S_1 U_{22} U_{22}^T$.
4. Compute the SVD (25) of \widetilde{S}_1 and get \widetilde{U}_{11} .
5. Let $A = U_{22} U_{22}^T \widetilde{U}_{11}$.

4.3 GF/SVD

Let $\text{rank}(S_1) = r_1$ and consider the SVD of the matrix S_1 :

$$S_1 = [U_{11} \ U_{12}] \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_{11}^T \\ U_{12}^T \end{bmatrix}, \quad (26)$$

where $U_{11} \in R^{m \times r_1}$ and $U_{12} \in R^{m \times (m-r_1)}$ are column orthogonal matrices and $\Sigma_{11} \in R^{r_1 \times r_1}$ is a diagonal matrix with non-increasing positive diagonal components. It is obvious that $\widetilde{S}_1 = U_{11}^T S_1 U_{11} = \Sigma_{11}$ is a nonsingular matrix.

Let $\widetilde{S}_2 = U_{11}^T S_2 U_{11}$. In the most applications, we have $\text{rank}(S_2) \geq \text{rank}(S_1)$ and then \widetilde{S}_2 is also a nonsingular matrix (for details, see [19]). Consequently, we can get an optimal dimensionality reduction matrix A . The specific algorithm is as follows.

Algorithm 3. GF/SVD

1. Construct the affinity matrix B and constraint matrix C .
2. Let $S_1 = X B X^T$ and $S_2 = X C X^T$.
3. Compute the SVD (26) of S_1 and get U_{11} .
4. Let $\widetilde{S}_1 = U_{11}^T S_1 U_{11}$ and $\widetilde{S}_2 = U_{11}^T S_2 U_{11}$.
5. Compute the eigenvectors g_1, \dots, g_d of the matrix $(\widetilde{S}_2)^{-1} \widetilde{S}_1$ corresponding to the first d largest eigenvalues.
6. Let $A = U_{11} [g_1, \dots, g_d]$.

5 Experiments and Analysis

In this section, in order to evaluate the performance of algorithms RGF, NGF and GF/SVD, we make a series of experiments on KSC data recorded in March 23, 1996 with six kinds of feature combination

methods, regularized GFDA (RGFDA), GFDA based on null space (NGFDA), GFDA based on SVD (GFDA/SVD), regularized GLPP (RGLPP), GLPP based on null space (NGLPP) and GLPP based on SVD (GLPP/SVD). The KSC data belongs to NASA's AVIRIS remote sensing image and can be downloaded from <http://www.csr.utexas.edu/hyperspectral/data/KSC>. We choose three different sets of data:

- T_1 : KSC_TRAIN_expt8_Rate5.mat,
- T_2 : KSC_TRAIN_expt9_Rate5.mat,
- T_3 : KSC_TRAIN_expt10_Rate5.mat.

In each set of data, we take respectively the first 20 data as samples in classes 1-3. Each sample has 176 features. All experiments are performed on a Pentium 2.52GH with 2G RAM and programmed in the MATLAB 7.11.0.

The first group of experiments are made on T_3 and the results can be found in Figures 1-6. From Figures 1-3, we can see that RGFDA and NGFDA are more effective than GFDA/SVD for the classification results of low dimension samples, and RGFDA is more effective than NGFDA for the classification results of low dimension samples. RGFDA, NGFDA and GFDA/SVD can separate completely class 1, but GFDA/SVD can not separate class 2 and class 3. From Figures 4-6, we can see that GLPP/SVD is more effective than RGLPP and NGLPP, and NGLPP is more effective than RGLPP for the classification results of low dimension samples. RGLPP, NGLPP and GLPP/SVD can separate completely class 1, but RGLPP can not separate class 2 and class 3.

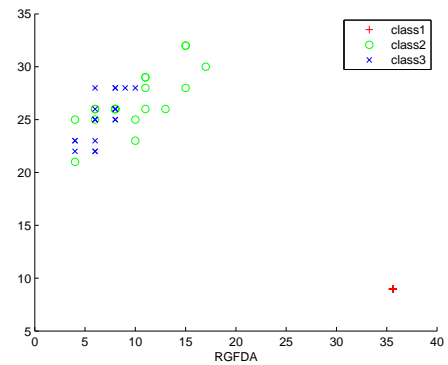


Figure 1: RGFDA

The second group of experiments are made on T_1 and the results can be found in Figures 7-12. From Figures 7-9, we can see that RGFDA and NGFDA are more effective than GFDA/SVD for the classification results of low dimension samples, and RGFDA is more effective than NGFDA. RGFDA, NGFDA and GFDA/SVD can separate class 1 completely,

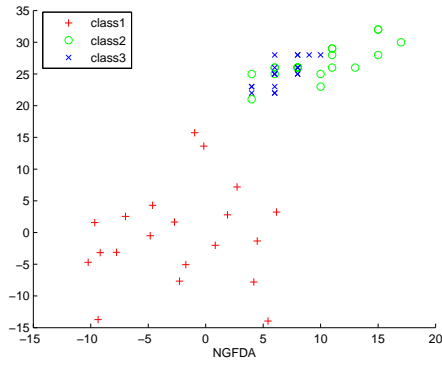


Figure 2: NGFDA

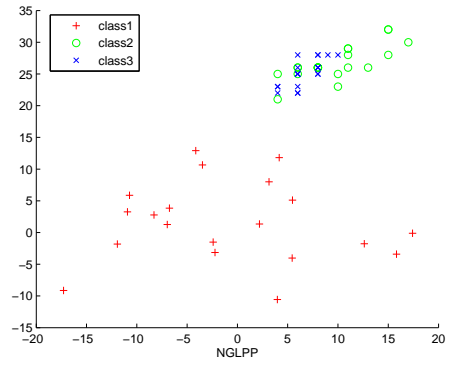


Figure 5: NGLPP

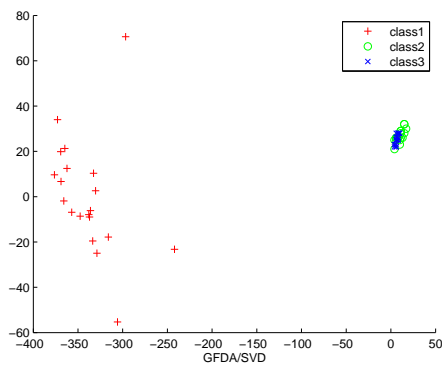


Figure 3: GFDA/SVD

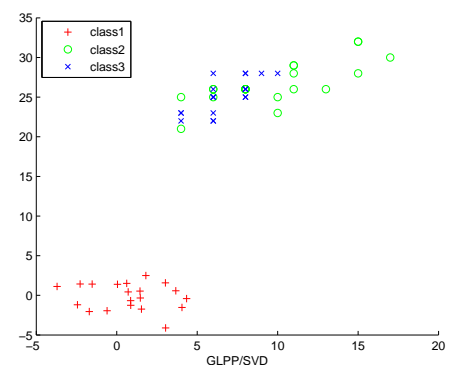


Figure 6: GLPP/SVD

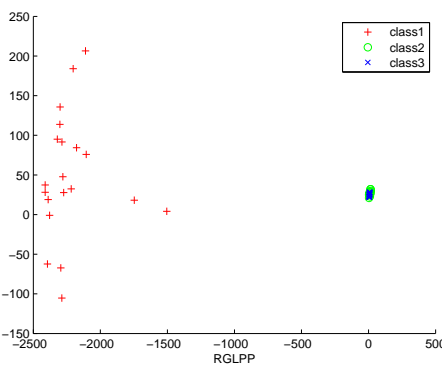


Figure 4: RGLPP

and class 3.

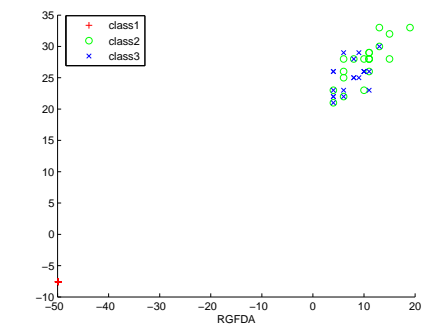


Figure 7: RGFDA

but GFDA/SVD can not separate class 2 and class 3. From Figures 10-12, we can see that NGLPP and GLPP/SVD are more effective than RGLPP, and GLPP/SVD is more effective than NGLPP for the classification results of low dimension samples. RGLPP, NGLPP and GLPP/SVD can separate completely class 1, but RGLPP can not separate class 2

The third group of experiments are made on T_2 and the results can be found in Figures 13-18. From Figures 13-15, we can see that RGFDA and NGFDA are more effective than GFDA/SVD for the classification results of low dimension samples, and RGFDA is more effective than NGFDA. RGFDA, NGFDA

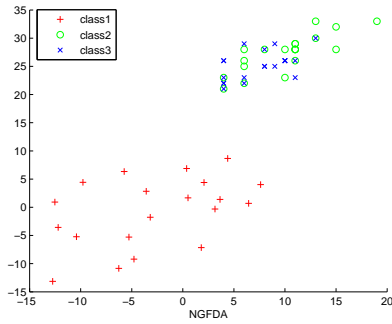


Figure 8: NGFDA

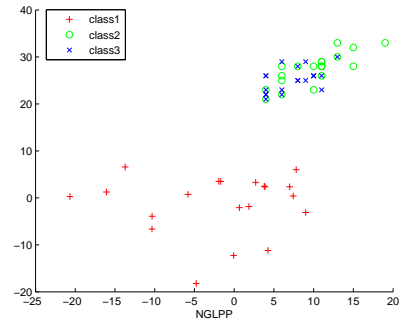


Figure 11: NGLPP

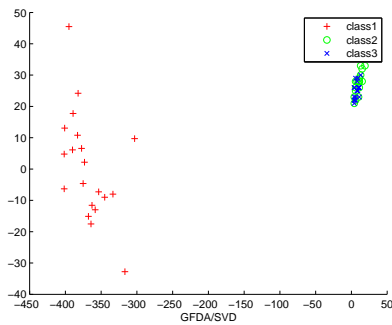


Figure 9: GFDA/SVD

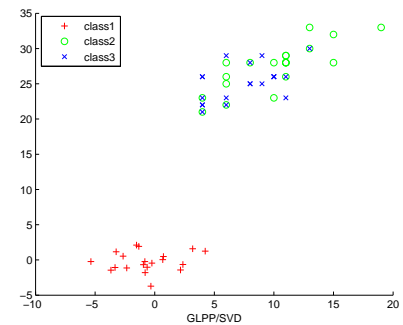


Figure 12: GLPP/SVD

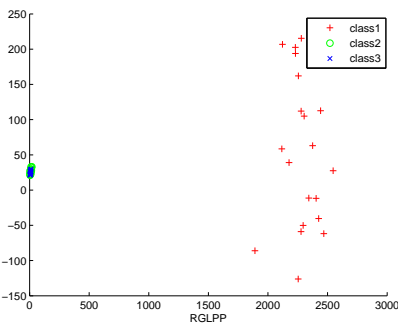


Figure 10: RGLPP

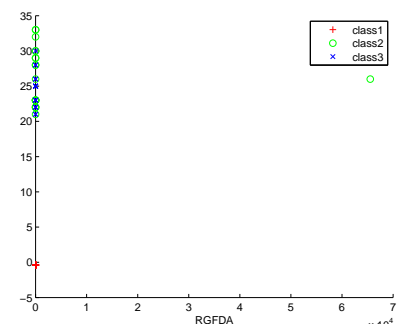


Figure 13: RGFDA

and GFDA/SVD can separate completely class 1, but GFDA/SVD can not separate class 2 and class 3. From Figures 16-18, we can see that GLPP/SVD is more effective than RGLPP and NGLPP, and NGLPP is more effective than RGLPP for the classification results of low dimension samples. RGLPP, NGLPP and GLPP/SVD can separate completely class 1, but RGLPP can not separate class 2 and class 3.

In short, we can see from three groups of experiments that RGFDA is more effective than NGFDA and GFDA/SVD for GFDA, and GLPP/SVD is more

effective than RGLPP and NGLPP for GLPP.

6 Conclusion

In this paper, we firstly present a unified framework based on graphs and show that the framework contains the global methods FDA, PCA and local methods LPP, DLPP, MFA, MMC as special cases. Secondly, in order to avoid the singularity problem of matrix S_2 , we propose three new effective algorithms

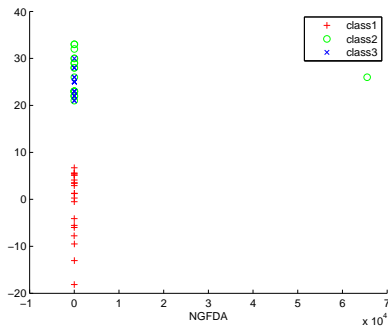


Figure 14: NGFDA

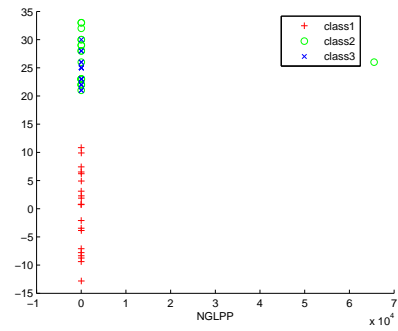


Figure 17: NGLPP

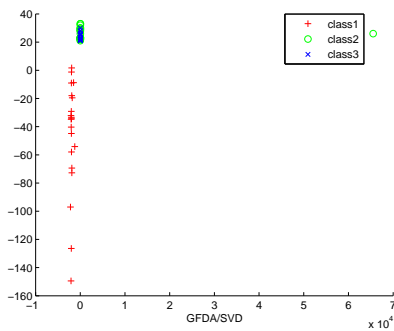


Figure 15: GFDA/SVD

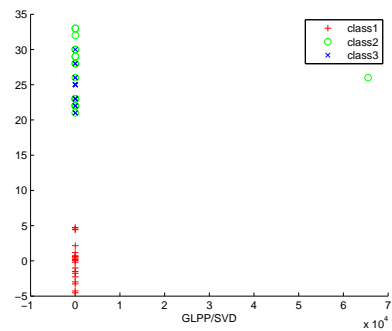


Figure 18: GLPP/SVD

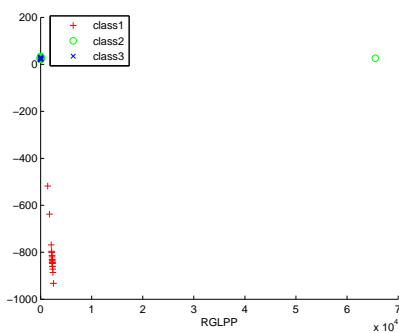


Figure 16: RGLPP

RGF, NGF, GF/SVD for GF. In addition, In order to illustrate the efficient and effective of the proposed algorithms, we make 18 experiments on three different sets of data taken from AVIRIS database with global method GFDA and local method GLPP. Experiment results show that the proposed algorithms are workable, and RGFDA is more effective than NGFDA and GFDA/SVD for GFDA and GLPP/SVD is more effective than RGLPP and NGLPP for GLPP.

In this paper, we consider the framework based on graphs (GF) for supervised problems. However,

we often encounter semi-supervised problems in real world. So, we can extend the supervised GF to semi-supervised case.

Acknowledgements This work is supported by Natural Science Foundation of Shandong Province (ZR2009AL006), P. R. China.

References:

- [1] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Second Edition), Academic Press, New York, 1990.
- [2] J. H. Friedman, Regularized discriminant analysis, *J. Am. Stat. Assoc.* 84 (405), 1989, pp. 165–175.
- [3] L. Chen, H. M. Liao, M. Ko, J. Lin, G. Yu, A new LDA-based face recognition system which can solve the small sample size problem, *Pattern Recognition* 33, 2000, pp. 1713–1726.
- [4] P. Howland, M. Jeon, H. Park, Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition, *SIAM J. MATRIX Anal. Appl.* 25 (1), 2003, pp. 165–179.

- [5] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, *Pattern Recognition* 34, 2001, pp. 2067–2070.
- [6] W. Zheng, L. Zhao, C. Zou, An efficient algorithm to solve the small sample size problem for LDA, *Pattern Recognition* 37, 2004, pp. 1077–1079.
- [7] J. Gao, L. Fan, L. Xu, Solving the face recognition problem using QR factorization, *WSEAS Transactions on Mathematics* 11 (8), 2012, pp. 728-737.
- [8] J. Yang, L. Fan, Weighted Generalized Kernel Discriminant Analysis Using Fuzzy Memberships, *WSEAS Transactions on Mathematics* 10 (10), 2011, pp. 346-357.
- [9] J. Gao, L. Fan, Kernel-based Weighted Discriminant Analysis with QR Decomposition and Its Application to Face Recognition, *WSEAS Transactions on Mathematics* 10 (10), 2011, pp. 358-367.
- [10] I. T. Jolliffe, Principal component analysis, 2nd ed, *Springer*, New York, 2002.
- [11] X. F. He, P. Niyogi, Locality preserving projections, *Proceedings of the Conference on Advances in Neural Information Processing System* 15 (6), 2003, pp. 135–160.
- [12] X. He, S. Yan, Y. H. P. N. Zhang, H. J. Face, Recognition using laplacianfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (3), 2005, pp. 328–340.
- [13] X. Ma, Y. Tan, T. Zhao, H. Tian, Face Recognition Based on Maximizing Margin and Discriminant Locality Preserving Projection, *Journal of Information and Computational Science* 7 (7), 2010, pp. 1551–1559.
- [14] S. Yang, D. Xu, B. Zhang, H. Zhang, Graph embedding and extension: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (1), 2007, pp. 40–50.
- [15] C. Chen, J. Shi, Local Correlation Discriminant Analysis and Its Semi-supervised Extension, *Journal of Electronics (China)* 28 (3), May 2011, pp. 289–296.
- [16] H. F. Li, T. Jiang, K. S. Zhang, Efficient and robust feature extraction by maximum margin criterion, *IEEE Transaction on Neural Networks* 17 (1), 2006, pp. 157–165.
- [17] Y. Cui, L. Fan, Feature extraction using fuzzy maximum margin criterion, *Neurocomputing* 86, 2012, pp. 52–58.
- [18] W. Yang, S. Zhang, W. Liang, A Graph Based Subspace Semi-supervised Learning Framework for Dimensionality Reduction, *J D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV, LNCS 5303*, 2008, pp. 644–677.
- [19] C. H. Park, H. Park, A comparison of generalized linear discriminant analysis algorithms, *Pattern Recognition* 41, 2008, pp. 1083–1097.