

Locality Preserving Fisher Discriminant Analysis with Clustering

Lishan Zou

College of Information and Automobile Engineering,
Guangzhou City Polytechnic,
Guangzhou, 510405, China

Yuechao Wang

College of Information and Automobile Engineering,
Guangzhou City Polytechnic
Guangzhou, 510405, China

Zhenzhou Chen

Computer School South China Normal University
Guangzhou, 510631, China

Xiaorong Wu

School of CISCO Informatics China, Guangdong
University of Foreign Studies
Guangzhou, 510420, China

Abstract-Fisher discriminant analysis (FDA) is an important feature extraction method for many classifiers. However, it tends to give undesired results if samples in some classes form several separate clusters, i.e., multimodal. This paper proposed a new feature extraction method called locality preserving Fisher discriminant analysis with clustering (LPFDA) for multimodal data. First new classes are formed by clustering data according to labels, then the between-subclass scatter matrix and within-subclass scatter matrix are computed by new classes, finally the vectors are choose which will maximize the Fisher criterion function as the discriminant vector . When our method is applied to the recognition problems of digits and images, and the experimental results show the better performance than the original one.

Keywords-Fisher discriminant analysis, Locality preserving, Feature extraction, Clustering

I. INTRODUCTION

The goal of feature extraction is to reduce dimensionality by projection of D-dimensional vector onto d-dimensional vector ($d < D$) while most of "information" contained in the data is preserved [1]. In this paper, we consider the supervised feature extraction problem where samples are accompanied with class labels.

Fisher discriminant analysis (FDA) is a popular method for linear feature extraction, which maximizes between-class scatter and minimizes within-class scatter [2]. This traditional FDA is known to work well and is practically useful even now. However, it tends to give undesired results if samples in some class form several separate clusters (i.e., multimodal) [3]. Multimodality is often observed in many practical applications such as disease diagnosis, hand-written digit recognition, etc.

The local structure of the data is preserved for features extracted by Locality-preserving projection (LPP) [4]. However, LPP is an unsupervised feature extraction method which does not take the label information into account.

Clustering is also based on locality. Cluster analysis is the organization of a collection of patterns into clusters based on similarity [5]. But clustering only keeps nearby data pairs in the original space.

In this paper, we propose a new feature extraction method called locality preserving Fisher discriminant analysis with clustering (LPFDA). LPFDA combines the ideas of FDA and Clustering: first subclasses are formed by clustering algorithms according to classes, then between-subclass separability is maximized while within-subclass local structure is preserved.

The remainder of the paper is organized as follows: Section 2 gives a brief review of FDA. Section 3 shows how to combine FDA method and clustering. Section 4 provides some experiments of LPFDA. Finally, section 5 gives a brief summary of the present method

II. FISHER DISCRIMINANT ANALYSIS AND CLUSTERING

In this section, we review Fisher discriminant analysis for feature extraction and clustering.

A. Fisher discriminant analysis

Here we briefly review the definition of Fisher criterion for feature extraction [1,2]. With some abuse, we refer to the feature extraction method based on the Fisher criterion as Fisher discriminant analysis (FDA), where the original FDA embeds the data samples only in one-dimensional space.

Let $x_i \in \mathbb{R}^d$ ($i = 1, 2, \dots, n$) be d-dimensional samples and $y_i \in \{1, 2, \dots, l\}$ be associated class labels, where n is the number of samples and l is the number of classes. Let n_i be

$$n = \sum_{i=1}^l n_i$$

the number of samples in the class i:

Let SW and SB be the within-class scatter matrix and the between-class scatter matrix defined by

$$S_W = \sum_{j=1}^l \sum_{i: y_i=j} (x_i - m_j)(x_i - m_j)^T$$

$$S_B = \sum_{i=1}^l n_i (m_i - m_0)(m_i - m_0)^T$$

where m_i is the mean vector of class i and m_0 is the global mean vector:

$$m_i = \frac{1}{n_i} \sum_{j: y_i=i} x_j \quad m_0 = \frac{1}{n} \sum_{i=1}^n x_i$$

The FDA transformation matrix is defined as follows:

$$J_{FDA}(W) = \frac{\text{trace}\{W^T S_B W\}}{\text{trace}\{W^T S_W W\}}$$

It is known that W is given by $W = \{w_1, w_2, \dots, w_r\}$, where w_i ($i=1, \dots, r$) are the generalized eigenvectors associated to the generalized eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ of the following generalized eigenvalue problem

$$S_B w = \lambda S_W w$$

Then the linear transform $y = W^T x$ is called FDA.

To study the behavior of FDA, let's see an example. In Figure.1, from two-dimensional two-class data (Figure.1 (a)), a one-dimensional feature (Figure.1 (b)) is obtained by FDA. We can see that FDA gives an undesired result.

B. Clustering

Clustering [6] can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. We can show this with a simple graphical example (Figure.2)

III. LOCALITY PRESERVING FISHER DISCRIMINANT ANALYSIS (LPFDA) WITH CLUSTERING

As illustrated in Figure.1, FDA can perform poorly if samples in some class form several separate clusters (i.e., multimodal).

To overcome this problem, we propose combining the idea of FDA and Clustering: first subclasses are formed by clustering algorithms according to classes, then between-subclass separability is maximized while within-subclass local structure is preserved. We call our new method locality preserving Fisher discriminant analysis (LPFDA) with clustering.

A. Clustering

Let $x_i \in \mathbb{R}^d$ ($i = 1, 2, \dots, n$) be d -dimensional samples and $y_i \in \{1, 2, \dots, l\}$ be associated class labels, where n is the number of samples and l is the number of classes.

For the following feature extraction by FDA, one can not perform clustering algorithms on the whole samples directly. First the samples are divided into subsamples according to class labels. Let S_i be the set contains the

samples whose class label is y_i . Then the clustering algorithms are performed on each S_i . Suppose the clustering number of subsamples S_i is C_i . The clustering number of the whole samples is

$$k = \sum_{i=1}^l C_i, \quad (k > l)$$

Let $z_j \in \{1, 2, \dots, k\}$ be the j th cluster of the whole samples, then we get new samples with difference subclass labels.

B. LPFDA with Clustering

Let $x_i \in \mathbb{R}^d$ ($i = 1, 2, \dots, n$) be d -dimensional samples and $z_j \in \{1, 2, \dots, k\}$ be associated subclass (clusters) labels, where n is the number of samples and k is the number of subclasses (clusters). Let n_j be the number of samples in the

$$n = \sum_{j=1}^k n_j$$

subclass j :

Let S_{Wsub} and S_{Bsub} be the within-subclass scatter matrix and the between-subclass scatter matrix defined by

$$S_{Wsub} = \sum_{j=1}^k \sum_{i: z_i=j} (x_i - m_j)(x_i - m_j)^T$$

$$S_{Bsub} = \sum_{j=1}^k n_j (m_j - m_0)(m_j - m_0)^T$$

where m_j is the mean vector of subclass j and m_0 is the global mean vector:

$$m_j = \frac{1}{n_j} \sum_{i: z_i=j} x_i \quad m_0 = \frac{1}{n} \sum_{i=1}^n x_i$$

The LPFDA transformation matrix is defined as follows:

$$J_{LPFDA}(W) = \frac{\text{trace}\{W^T S_{Bsub} W\}}{\text{trace}\{W^T S_{Wsub} W\}}$$

It is known that W is given by $W = \{w_1, w_2, \dots, w_r\}$, where w_i ($i=1, \dots, r$) are the generalized eigenvectors associated to the generalized eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ of the following generalized eigenvalue problem

$$S_{Bsub} w = \lambda S_{Wsub} w$$

Then the linear transform $y = W^T x$ is called LPFDA.

IV. NUMERICAL RESULTS

In this section, we apply feature extraction algorithms LPP, FDA and LPFDA to benchmark data sets for visualization and classification.

A. The Datasets and Algorithms

The following datasets are used in our experiments:

Dataset A: An artificial dataset obey the two-dimensional normal distribution. There are three normal distributions (clusters) of two classes with different mathematical expectations and variances.

Dataset B: The “Pendigits” database from the UCI repository. Pendigits is a Pen-based recognition problem of handwritten digits (0~9). We produce a series subsets of Pendigits $B_i (i = 3, \dots, 10)$ which is a classification problem of i classes.

Dataset C: The “Iris” database from the UCI repository.

To compare the methods above, we use linear support vector machines (SVM) [6] and K-nearest neighbors (KNN) [7] algorithm as classifiers.

B. Results and Analysis

Figure.3, Figure. 4 and Figure.5 show the results of data visualization on datasets above. For dataset A and C, FDA tends to mix samples of different classes, while LPP and LPPFDA separates them very well. For For dataset B3, FDA and LPPFDA separates them equally well, but LPP tends to mix samples of di@erent classes.

Table. I describe the classification accuracy of KNN and SVM on features extracted by FDA, LPP and LPPFDA. For FDA, there are only $n-1$ features can be extracted form n -classes problem. But LPPFDA can extract more than $n-1$ features (at most cluster number-1) form n -classes problem. So the default features number of the methods above is $n-1$.

For KNN algorithm, We can see that FDA performs well on most dataset But on dataset A, FDA is very poor because A is apparently multimodal. LPP is the most unstable methods of all for LPP is an unsupervised feature extraction method. LPPFDA is the best method of all for feature extraction. Though LPPFDA performs little worse than FDA when $n-1$ features are extracted (such as B3 and B4), LPPFDA performs better than FDA when n features are extracted.

For SVM with RBF kernel ($\gamma=1$), we also can see that LPPFDA is not so good as FDA (on dataset B). Because LPPFDA has a locality preserving nature which is not necessary for SVM.

Overall, LPPFDA is found to be useful in data visualization and classification tasks with some kind of classification methods.

V. CONCLUSIONS

Feature extraction based on the Fisher criterion (FDA) works well given data samples of each class form a single cluster (i.e., unimodal). On the other hand, samples in some class can be multimodal.

In this paper, we first showed that clustering methods can be used to form new class labels for samples. Based the samples with new class labels, then we proposed a localized variant of FDA called locality preserving Fisher discriminant analysis (LPPFDA) with clustering. Lastly, our method is applied to digits and images recognition problems, and the experimental results show that the performance of proposed method is superior to the original method.

Our feature work includes extending LPPFDA to nonlinear feature extraction by kernel trick and studying how clustering algorithms impact on LPPFDA.

ACKNOWLEDGMENT

The research was Supported by Foundation for Distinguished Young Talents in Higher Education of Guangdong (No.396-GK100018)

REFERENCES

- [1] Øivind Due Trier; Anil K Jain; Torfinn Taxt. Feature Extraction Methods For Character Recognition - A Survey, Pattern Recognition [J], 1996, PP:641-662.
- [2] Fisher, R. A. The use of multiple measurements in taxonomic problems, Annals of Eugenics [B], 1936, PP: 179-188.
- [3] Fukunaga, K. Introduction to statistical pattern recognition., Boston: Academic Press [B], Inc. Second edition, 1990.
- [4] HE Xiao-fei; Niyogi Partha. Locality preserving projections, In S. Thrun, L. Saul and B. Schölkopf (Eds.), Advances in neural information processing systems 16, Cambridge, MA: MIT Press [J]. 2004.
- [5] Gan Guo-jun. Data clustering: theory, algorithms, and applications, Society for Industrial and Applied Mathematics [B], 2007.
- [6] Zhang Shao-wu; Pan Quan; Zhang Hong-cai; Zhang Yunlong; Wang Haiyu. Classification of protein quaternary structure with support vector machine, Bioinformatics [J], 2003, PP: 2390-2396.
- [7] Zhang Min-Ling; Zhou Zhi-Hua. L-KNN: A lazy learning approach to multi-label learning. Pattern Recognition [J], 2007, PP: 2038-2048.

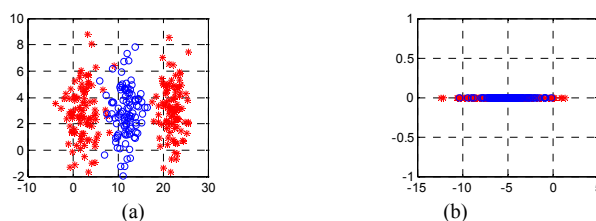


Figure.1 Example of feature extraction by FDA

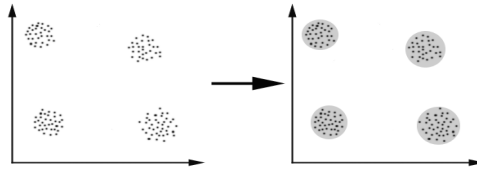


Figure.2 example of clustering

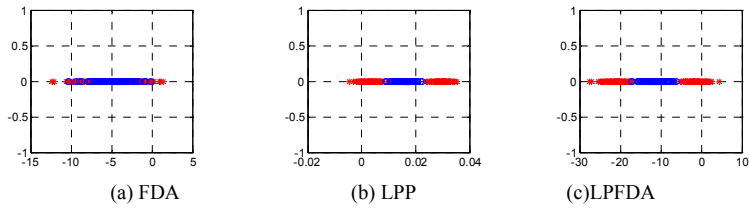


Figure. 3 Results of data visualization on dataset A

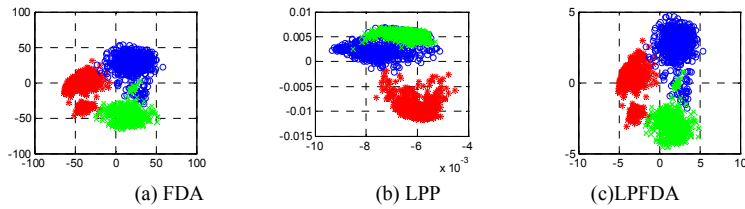


Figure. 4 Results of data visualization on dataset B_3

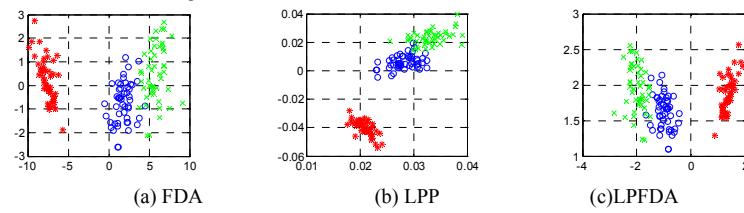


Figure. 5 Results of data visualization on dataset C

TABLE I. THE CLASSIFICATION ACCURACY ON DATASETS

Method	FDA		LPP		LPPFDA		LPPFDA with n features	
	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
A	57.93	66.67	33.77	66.67	97.43	97.10	99.20	99.10
B_3	97.20	89.38	78.67	35.14	94.98	74.20	98.26	71.33
B_4	96.57	77.99	88.16	48.56	96.49	73.23	96.84	76.84
B	95.20	95.60	97.37	97.67	95.34	95.67	95.54	96.24
C	96.67	98.67	97.33	97.33	98.00	97.33	97.33	98.67