ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

freepaper.me FREE paper

Feature extraction using maximum nonparametric margin projection

Bo Li [a,b,c,*], Jing Du [a,b], Xiao-Ping Zhang [c,d]

[a] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei 430065, China
[b] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, Hubei 430065, China
[c] Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada M5B 2K3
[d] School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form
31 October 2014
Accepted 8 November 2014

Keywords:
Dimensionality reduction
Nonparametric margin
Subspace learning
Feature extraction

ABSTRACT

Dimensionality reduction is often recommended to handle high dimensional data before performing the tasks of visualization and classification. So far, large families of dimensionality reduction methods besides the supervised or the unsupervised, the linear or the nonlinear, the global or the local have been developed. In this paper, a maximum nonparametric margin projection (MNMP) method is put forward to extract features from original high dimensional data. In the proposed method, we offer some nonparametric or local definitions to the traditional between-class scatter and within-class scatter, which contributes to remove the disadvantage that linear discriminant analysis (LDA) can not be well-performed in the cases of non-Gaussian distribution data. Based on the predefined between-class scatter and the within-class scatter, a nonparametric margin can be reasoned to avoid the small sample size (SSS) problem. Moreover, the proposed nonparametric margin will be maximized to explore a discriminant subspace. At last, we have conducted experiments on some benchmark data sets such as Palmprint database, AR face database and Yale face database. In addition, performance comparisons have also been made to some related feature extraction methods including LDA, nonparametric discriminant analysis (NDA) and local graph embedding based on maximum margin criterion (LGE/MMC). Experimental results on these data sets have validated that the proposed algorithm is effective and feasible.
© 2015 Elsevier B.V. All rights reserved.

1. Introduction

When dealing with pattern classification problems, original data always contain a great deal of information [1]. Among them, some are useful, while some are superfluous and some are even noisy. In addition, the patterns hidden in these data are always represented to high dimensional vectors. For example, a 32-by-32 appearance-based image can be viewed as a 1024-dimensional point in image space [2]. So how to extract discriminant information from original complicated data and how to represent original data with low dimensional vectors are playing more and more important role in data classification. The resulted reasons can be concluded as follows. One is that the recognition accuracy will be greatly degradated because of disturbance of noise; the other lies in that the computational cost will be very expensive if original data are directly involved in classification. It is a typical way to solve the problem using dimensionality reduction techniques. Usually, dimensionality reduction serves as automatic learning to extract features with high efficiency. Due to the property, besides pattern recognition, the topics of dimensionality reduction also appears in many other fields including data mining, computer vision, information retrieval, machine learning and bioinformatics [3,4].

Currently, researchers have developed many classical dimensionality reduction approaches, which can be categorized into two kinds based on whether the class information is considered, i.e., the supervised or the unsupervised. They are also broadly partitioned into linear methods [5–9] and nonlinear models besides artificial neural networks (ANN) [10–13]. Linear dimensionality reduction techniques try to seek a meaningful low dimensional subspace with a linear transformation, where a compact representation of input data can be provided. Among all the linear dimensionality reduction methods, principal component analysis (PCA) [5–7] and linear discriminant analysis (LDA) [6,8,9] are most well-known.

Generally speaking, PCA projects original data into a low dimensional space spanned by the eigenvectors associated to the largest eigenvalues of covariance matrix of all samples, where PCA is the optimal representation of input data in the sense of minimizing mean squared error (MSE) [14,15]. However, PCA is completely unsupervised with regard to data class information, which may result in much discriminant information missing and

* Corresponding author.
E-mail address: liberol@126.com (B. Li).

http://dx.doi.org/10.1016/j.neucom.2014.11.105
0925-2312/© 2015 Elsevier B.V. All rights reserved.

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

recognition ability weakening, especially encountering a large number of sample points [6].

Unlike PCA, LDA takes full consideration of patterns' class information. It is believed that feature extraction methods under supervised learning will be more discriminative. Thus LDA can enhance data separability more than PCA. LDA maps original data into an optimal subspace by a linear transformation. The linear transformation matrix consists of the corresponding eigenvectors which can maximize the trace ratio of the between-class scatter to the within-class scatter.

To improve the performance of the original LDA, many modified versions based on LDA have been reported, which are validated to be efficient [15–31]. Moreover, when constructing the between-class scatter and the within-class scatter, most of LDA-based methods inherit the parametric formulations in traditional LDA, which heavily depend on the fact that samples in each class should distribute as Gaussian function. However, it is not always true for all the data. In case of non-Gaussian distribution data, these LDA-based methods will lead to performance degradation. In order to overcome the problem, a nonparametric definition on the between-class scatter is presented, where boundary samples are explicitly exploited [8]. Instead of class centers, the new defined between-class scatter is formulated on the whole training set, where their contributions to discrimination are justified by the weights between inter-class sample pairs. Thus the feature extraction algorithm adapts more to those samples without taking into account data distribution model. However, the above nonparametric definition is restricted to data classification for two-class cases rather than multi-class. So Li et al. propose a nonparametric discriminant analysis (NDA) algorithm to deal with multi-class problem [32], where a new between-class scatter is advanced by extending the definition of the original nonparametric between-class scatter matrix. But the proposed NDA just pays attention to local structure information of boundary points, which can not explore all the local structure hidden in the intra-class data.

During last decade, many manifold learning approaches have been developed for nonlinear dimensionality reduction. Besides isometric mapping (ISOMAP) [33], locally linear embedding (LLE) [34], Laplacian eigenmaps (LE) [35], local tangent space alignment (LTSA) [36], maximum variance unfolding (MVU) [37] and Riemannian manifold learning (RML) [38] are their representatives. It has been shown by many examples that these methods have yielded impressive results on artificial and real world data sets. Compared to other dimensionality reduction methods, on the one hand, manifold learning can probe the essential dimensions of manifold embedded in high dimensional space, on the other hand, manifold learning expects to embed original data into a lower dimensional space by locality preserving, where the locality can be approached using k nearest neighbors (KNN) criterion. It is the characteristics of exploring manifold structure locally that manifold learning can be employed to handle data either in the non-Gaussian cases or in the Gaussian cases. Enlightened by the idea of local learning in manifold learning, both the between-class scatter and the within-class scatter can be locally modeled, based on which the proposed feature extraction method will be robustness to data with all kinds of distributions, especially non-Gaussian distribution.

In addition, the above nonparametric versions of LDA still suffer small sample size (SSS) problem, which often occurs to LDA due to case of the limited training set with high dimensionality. Under such circumstance, the within-class scatter is not positive-definite, which leads to serious instability and over-fitting. Until to now, many attempts have been tried, among which maximum margin criterion (MMC) shows its superiority [39]. On the basis of MMC, Qiu et al. present a nonparametric maximum margin criterion (NMMC) for face features extraction [40], where the between-class scatter is defined using the nearest inter-class neighbor pair. But for the within-class scatter, it is not locally constructed because the furthest intra-class points are involved. Meanwhile, the within-class scatter is formulated by the difference between the furthest intra-class data pair, which fails to explore the local structure in the intra-class data. Recently, many manifold learning based maximum margin criterion algorithms have been presented to extract discriminant features from original data, where the local scatter and the non-local scatter instead of the between-class scatter and the within-class scatter are discriminatively advanced [41–45].

In this paper, we propose a maximum nonparametric margin projection (MNMP) method for feature extraction, which tries to overcome the problems in the traditional LDA. In the proposed MNMP, different to the existing nonparametric between-class scatter and within-class scatter, we offer a novel nonparametric or local definition on them, where all the training samples rather than only the boundary points are contained, as a result, data distribution model will not be considered. At the same time, the newly defined nonparametric between-class scatter and within-class scatter are also introduced to reason a margin criterion, which will be maximized to explore a discriminant subspace.

The rest of the paper is organized as follows: Section 2 simply reviews LDA and MMC. In Section 3, the between-class scatter and the within-class scatter is parametric or locally defined, based on which a new margin criterion can be nonparametric deduced, and then the proposed MNMP is addressed in details. Compared to some related feature extraction methods such as LDA, nonparametric discriminant analysis (NDA) and local graph embedding based on maximum margin criterion (LGE/MMC), experimental results on Palmprint data, AR face data and Yale face data are offered in Section 4. At last, the paper is finished with some conclusions in Section 5.

2. Related work

2.1. Notations

The main notations used in the whole paper are summarized as follows:

- The high-dimensional input sample points will be denoted as $X_1, X_2, \ldots, X_n$. Sometimes it will be convenient to work with these sample points as a single matrix $X = [X_1, X_2, \ldots, X_n] \in \Re^{D\times n}$.
- The low-dimensional representations of $X_1, X_2, \ldots, X_n$ are represented as $Y_1, Y_2, \ldots, Y_n$. And the matrix form of these points is $Y = [Y_1, Y_2, \ldots, Y_n] \in \Re^{d\times n}$ $(d << D)$.
- n is the number of all the sample points.
- D is the dimension of input sample points.
- d is the dimension of output samples.
- k is the number of the nearest neighbors used by a particular algorithm such as k nearest neighbors.
- m_i is the i-th class mean.
- m is the mean of all the sample points.
- c is the number of the class labels.
- n_i denotes the sample number of the i-th class.
- S_B is the between-class scatter.
- S_W denotes the within-class scatter.

2.2. LDA

LDA aims to look for a linear subspace W, within which the projections of the samples from different classes are more apart, as

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

FREE paper freepaper.me دانلود کننده مقالات علمی

defined by maximizing the following ratio criterion:

$$J(W) = \max \frac{tr\left\{W^T S_B W\right\}}{tr\left\{W^T S_W W\right\}} \tag{1}$$

where the between-class scatter S_B and the within-class scatter S_W are defined as follows, respectively.

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \tag{2}$$

$$S_W = \sum_{i=1}^{c} \left(\sum_{j=1}^{n_i} (X_i^j - m_i)(X_i^j - m_i)^T \right) \tag{3}$$

Along with the orthogonal constraint of W, Eq. (1) can be solved as a generalized eigenvector and eigenvalue problem, which is stated below:

$$S_B W_i = \lambda_i S_W W_i \tag{4}$$

where W_i and λ_i are the i-th generalized eigenvector and eigenvalue of S_B with regard to S_W. The LDA solution, i.e. W, contains all the $c-1$ eigenvectors with non-zero eigenvalues because S_B has a maximal rank of $c-1$.

2.3. MMC

From the classical LDA, we know that when the trace ratio of the between-class scatter to the within-class is maximized, the samples labeled different classes will be well separated. However, LDA easily incurs SSS problem because S_W is not inversed in most cases. Thus MMC based discriminant rule is defined to be difference of the between-class scatter to the within-class scatter. The objective function of MMC can be simply described as follows:

$$J(W) = \max tr\left\{W^T (S_B - \alpha S_W) W\right\} \tag{5}$$

where α is a non-negative constant which benefits to balance the relative merits of maximizing the between-class scatter and minimizing the within-class scatter.

From Eq. (5), it can be deduced that the optimal orthogonal projection is spanned by the eigenvectors corresponding to the top d eigenvalues of the following eigen-decomposition.

$$(S_B - \alpha S_W) W_i = \lambda_i W_i \tag{6}$$

3. Maximum nonparametric margin projection

3.1. The motivation of the proposed method

As a supervised linear dimensionality reduction method, LDA is popular in feature extraction from original data due to its simply application. LDA is also called parametric discriminant analysis (PDA) since it uses the parametric form of the scatter matrix based on the assumption of Gaussian distribution [8]. However, on account of the parametric definitions on the between-class scatter and the within-class scatter and the form of trace ratio, there are three disadvantages in LDA. Firstly, LDA algorithm is based on the assumption that all classes share the Gaussian distribution with the same covariance matrix. When encountering those data with non-Gaussian distribution, LDA will not be efficient for data classification. Secondly, the number of the final LDA features must be less than $c-1$ because the rank of the between-class matrix is at most $c-1$. However, it is not enough to separate different class data with only $c-1$ features, especially in high-dimensional spaces. Thirdly, only the centers of classes are taken into account to compute both the between-class scatter and the within-class scatter and LDA is unsuccessful to capture the local structure information of all classes, which has been validated to be essential in pattern recognition [33,34]. Thus in the proposed MNMP, at first we locally define the between-class scatter and the within-class scatter, which contributes to avoid the restriction on data Gaussian distribution in traditional LDA and to make full use of local information for all samples besides both the between-class and the within-class data. Moreover, on the basis of the newly defined between-class scatter and within-class scatter, a nonparametric margin criterion can be deduced without paying attention to SSS problem. At last, the nonparametric margin criterion is justified to approach a discriminant subspace where data can be well classified.

3.2. The nonparametric between-class scatter and the nonparametric within-class scatter

In LDA, the traditional definitions on the between-class scatter and the within-class scatter just concentrate on the mean of all the samples and the mean of all the classes, where the local structure information of all the samples is missing. Thus for the between-class scatter, both data labels and local information are all taken into account. For any point X_1, we should select those labeled different class labels and ranked the first k bottom Euclidean distances to it as its k between-class nearest neighbors. Based on the k between-class nearest neighbors, its local between-class scatter can be modeled as follows:

$$S_B^i = (X_i - BLocal(X_i))(X_i - BLocal(X_i))^T \tag{7}$$

where $BLocal(X_i)$ denotes the local between-class mean, i.e.

$$BLocal(X_i) = \frac{1}{k} \sum_{i=1}^{k} X_j, c(i) \neq c(j) \tag{8}$$

where X_j is the k between-class nearest neighbors of X_i, $c(i)$ and $c(j)$ mean the labels of point X_i and point X_j, respectively.

For each point, we also seek its k between-class nearest neighbors and obtain the corresponding local between-class scatter, thus the global nonparametric between-class scatter can be rewritten to:

$$S_B^{(n)} = \sum_i (X_i - BLocal(X_i))(X_i - BLocal(X_i))^T \tag{9}$$

Similarly, because the within-class scatter reflects the clustering of the same class data, for any point X_1, those intra-class points with k shortest Euclidean distances to it can be taken as its k within-class nearest neighbors, then its local within-class scatter can be represented to:

$$S_W^i = (X_i - WLocal(X_i))(X_i - WLocal(X_i))^T \tag{10}$$

where $WLocal(X_i)$ denotes the local within-class mean, i.e.

$$WLocal(X_i) = \frac{1}{k} \sum_{i=1}^{k} X_j, c(i) = c(j) \tag{11}$$

where X_j is the k within-class nearest neighbors of X_i.

For each point, repeat Eqs. (10) and (11), and obtain the corresponding local within-class scatter, at last the global nonparametric within-class scatter is stated below:

$$S_W^{(n)} = \sum_i (X_i - WLocal(X_i))(X_i - WLocal(X_i))^T \tag{12}$$

3.3. Nonparametric margin

It is well-known that the between-class scatter characterizes the saparabity of different labeled data and the within-class scatter signifies the compactness of the same class samples. Thus

دانلود کننده مقالات علمی freepaper.me FREE paper

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

based on the predefined between-class scatter and the within-class scatter, a nonparametric margin can be reasoned with the following form:

$$M = tr\left\{S_B^{(n)} - S_W^{(n)}\right\} \tag{13}$$

where M denotes the defined nonparametric margin.

3.4. Justification

In the low dimensional space, the corresponding nonparametric between-class scatter $S_{BL}^{(n)}$ and within-class scatter $S_{WL}^{(n)}$ are expressed to:

$$S_{BL}^{(n)} = \sum_i (Y_i - BLocal(Y_i))(Y_i - BLocal(Y_i))^T \tag{14}$$

$$S_{WL}^{(n)} = \sum_i (Y_i - WLocal(Y_i))(Y_i - WLocal(Y_i))^T \tag{15}$$

According to these scatters, the nonparametric margin in the low dimensional space is formulated as follows:

$$M_L = tr\left\{S_{BL}^{(n)} - S_{WL}^{(n)}\right\} \tag{16}$$

Provided that the linear features Y can be obtained by a linear transformation, i.e. $Y_i = A^T X_i$, therefore, the nonparametric margin M_L in the low dimensional space can be represented to:

$$\begin{aligned} M_L &= tr\left\{\sum_i (Y_i - BLocal(Y_i))(Y_i - BLocal(Y_i))^T \right. \\ &\quad \left. - \sum_i (Y_i - WLocal(Y_i))(Y_i - WLocal(Y_i))^T\right\} \\ &= tr\left\{\sum_i (AX_i - A * BLocal(X_i))(AX_i - A * BLocal(X_i))^T \right. \\ &\quad \left. - \sum_i (AX_i - A * WLocal(X_i))(AX_i - A * WLocal(X_i))^T\right\} \\ &= tr\left\{A\sum_i (X_i - BLocal(X_i))(X_i - BLocal(X_i))^T A^T \right. \\ &\quad \left. - A\sum_i (X_i - WLocal(X_i))(X_i - WLocal(X_i))^T A^T\right\} \\ &= tr\left\{AS_B^{(n)}A^T - AS_W^{(n)}A^T\right\} \\ &= tr\left\{A(S_B^{(n)} - S_W^{(n)})A^T\right\} \end{aligned} \tag{17}$$

Table 1
Outline of the proposed MNMP.

Algorithm:
Step 1 Identifying *k* between-class neighbors and *k* within-class neighbors
1.1 Determine k between-class nearest neighbors and k within-class nearest neighbors;
1.2 For any point, compute the local means for the k between-class nearest neighbors and k within-class nearest neighbors, respectively;
1.3 Calculate the local between-class scatter and the local within-class scatter;
Step 2 Eigen decomposition
2.1 Obtain the nonparametric between-class scatter and the nonparametric with-class scatter;
2.2 Construct the nonparametric margin;
2.3 Achieve the linear transformation matrix A by solving the generalized eigen-decomposition $(S_B^{(n)} - S_W^{(n)})A_i = \lambda_i A_i$;
Step 3 Low dimensional embeddings
3.1 Output the corresponding low dimensional embeddings for any new coming point X_t, i.e. $Y_t = A^T X_t$.

At last, the proposed MNMP aims to approach an optimal subspace with better classification performance, where data with different labels will be more apart and samples with same class will be more clustered. In other words, the nonparametric margin should be maximized in the low dimensional space. So the objective function of the proposed MNMP can be modeled as:

$$J(A) = \max tr\left\{A(S_B^{(n)} - S_W^{(n)})A^T\right\} \tag{18}$$

Moreover, we expect to find an orthogonal subspace, i.e. $A^T A = I$.

So this optimization problem will be figured out to the following constrained objective function:

$$\begin{aligned} J(A) &= \max tr\left\{A(S_B^{(n)} - S_W^{(n)})A^T\right\} \\ s.t. \quad & A^T A = I \end{aligned} \tag{19}$$

In order to find solutions of Eq. (19), we construct the following Lagrange multiplier:

$$L(A, \lambda) = tr\left\{A(S_B^{(n)} - S_W^{(n)})A^T\right\} - \lambda(A^T A - I) \tag{20}$$

Then, the optimal orthogonal subspace A can be obtained by solving the equation below:

$$\frac{\partial L(A)}{\partial A} = 0 \tag{21}$$

At last, we have:

$$(S_B^{(n)} - S_W^{(n)})A_i = \lambda_i A_i \tag{22}$$

From Eq. (22), it can find that A is spanned by the eigenvectors associated to top d eigenvalues of the above generalized eigen-decomposition.

As a result, the outline for the proposed method can be summarized as Table 1.

4. Experiments

As described previously, an image can be represented to a point in high dimensional space. How to extract the most discriminant features and remove unwanted information is the work to be done in the following experiments. In this Section, the proposed MNMP will be applied to some benchmark databases such as Palmprint data, AR face data and Yale face data. When applying MNMP to those data sets, what should be conducted first is to construct the neighborhood graph with k nearest neighbor criterion. In the experiments, we set k to $l-1$, where l denotes the number of the

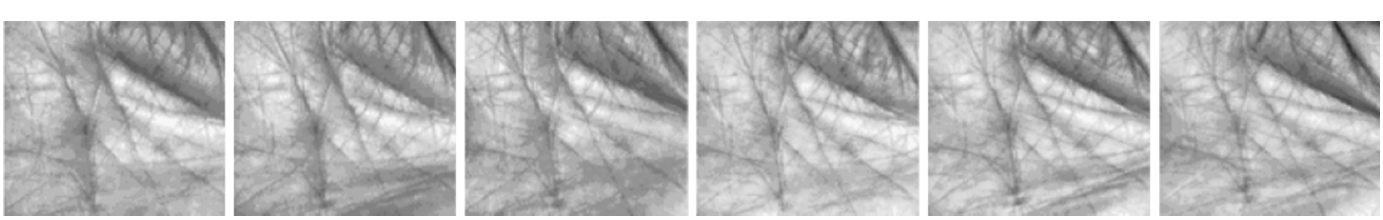

Fig. 1. The cropped sample images from PolyU Palmprint.

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

FREE paper freepaper.me دانلود کننده مقالات علمی

training samples for each class. In order to validate the efficiency of the proposed MNMP, the performances are compared to some related algorithms such as LDA, NDA and LGE/MMC. At last, the nearest neighbor classifier is adopted to classify the features extracted by LDA, NDA, LGE/MMC and the proposed MNMP, respectively.

4.1. Experiments on Palmprint data

The PolyU Palmprint database is constructed by Biometrics Research Centre, the Polytechnic University, Hongkong [46]. In this data set, there are 100 persons and each with 6 Palmprint images, so the total number of samples is 600. Moreover, 6 images for each were collected in two sessions. The first 3 images were captured in the first session, so did for the last 3 images in the second session two months later. Fig. 1 displays some cropped samples, which are of size 128 by 128. From the two session images, we randomly select 3 images as trainings and the rest as test. When constructing the neighborhood graph, k set to 2. Then we use LDA, NDA, LGE/MMC and MNMP to extract features, respectively. Moreover, in order to obtain the statistical results, we repeat the experiment ten times by randomly selecting three training samples per-person.

Shown in Table 2 are the mean accuracy and its variance by repeating the experiments ten times, where the training samples are randomly selected. From Table 2, it can find that the performance of the proposed method outperforms those of the other methods.

Fig. 2 displays the curves of the recognition accuracy at different dimensions, where the proposed MNMP gains the best recognition accuracy compared to LDA, NDA and LGE/MMC.

Table 2
Mean and the corresponding variance on Palmprint database by performing LDA, NDA, LGE/MMC and MNMP.

Methods	Mean and variance (%)	Dimensions
LDA	90.33 ± 0.79	52
NDA	91.67 ± 1.21	54
LGE/MMC	93.25 ± 1.76	70
MNMP	95.54 ± 1.48	92

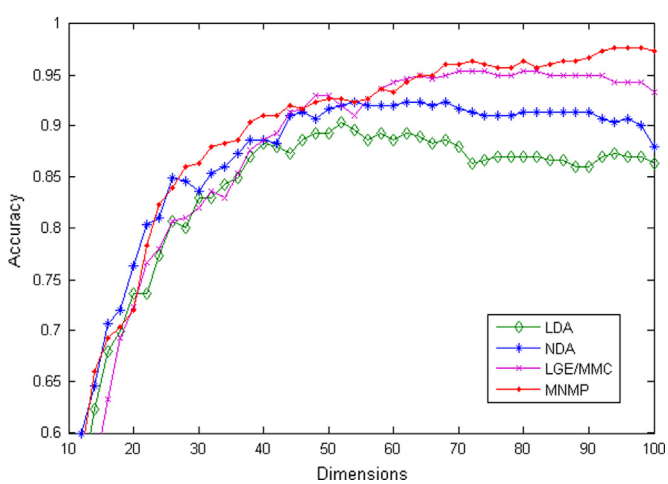

Fig. 2. The recognition accuracy curves with the varied dimensions by performing LDA, NDA, LGE/MMC and MNMP on Palmprint database.

4.2. Experiments on AR face data

AR face [47] contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each section contains 13 color images. In this experiment, 14 grayscale face images (each session containing 7) of these 120 individuals are selected. The face portion of each image is manually cropped and then normalized to be size of 40×50, which is displayed in Fig. 3.

In this experiment, we select seven images as trainings and the rest seven images as test. When constructing KNN graph, parameter k set to 6. Moreover, each experiment is repeated ten times by randomly selecting 7 images as trainings.

The statistical results on AR face data are illustrated in Table 3. Experimental results stated in Table 3 validate that the proposed MNMP is superior to the other methods besides LDA, NDA and LGE/MMC.

In order to test the dimensions' impact on the final performance, we also figure the recognition accuracy curves at different

Table 3
Mean and the corresponding variance on AR database by performing LDA, NDA, LGE/MMC and MNMP.

Methods	Mean and variance (%)	Dimensions
LDA	91.56 ± 1.15	118
NDA	93.25 ± 1.57	112
LGE/MMC	95.32 ± 1.34	118
MNMP	96.67 ± 0.95	106

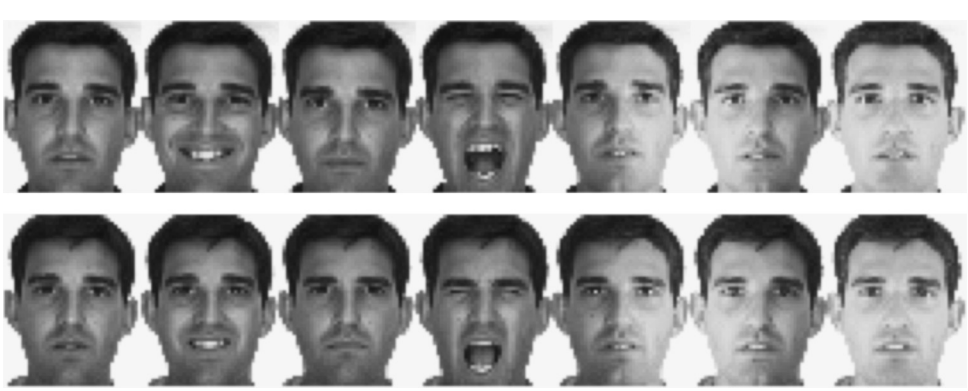

Fig. 3. The cropped images for one person in AR database.

FREE paper freepaper.me دانلود کننده مقالات علمی

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

dimensions using LDA, NDA, LGE/MMC and MNMP on AR face database, where performances are changed accompanying by different dimensions. Fig. 4 also indicates that the maximum accuracy of the proposed MNMP is larger than those of LDA, NDA and LGE/MMC.

4.3. Experiments on Yale face data

The Yale face database [48] was constructed at the Yale Center for Computation Vision and Control. There are 165 images about 15 individuals in YALE face data sets, where each person has 11 images. The images demonstrate variations in lighting condition (left-right, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised and wink), and with or without glasses. Shown in Fig. 5 is one cropped object with size of 64 by 64 from Yale database.

Firstly, we randomly select six images from Yale database as training sets and the rest five images as test sets for each class. Shown in Table 4 are the mean recognition accuracy and the corresponding variance by repeating the experiment 10 times, where the six training images are randomly selected and k set to 5 to identify the local between-class and the local within-class nearest neighbors. It is found that the proposed method is better than the other techniques.

Fig. 6 characterizes the recognition rate accuracy with varied dimensions on Yale face data. It can be found that the recognition rates for all the four methods increases greatly at low dimensions, and then, after the recognition rates climb their peak, they show the trend with great decreasing except LDA. However, the maximum accuracy of MNMP is also at the top compared to those of LDA, NDA and LGE/MMC.

5. Conclusions

In this study, we propose a maximum nonparametric margin projection (MNMP) method to extract discriminatory features from patterns. The proposed algorithm integrates both class information and local geometry, where the nonparametric between-class scatter and the nonparametric within-class scatter are locally defined. Compared to some other widely used LDA-based dimensionality reduction algorithms, the proposed method can efficiently classify those data with either Gaussian or non-Gaussian distribution. Due to the difference form of the predefined nonparametric margin, MNMP naturally avoids SSS problem, which always occurs to most of dimensionality reduction methods. Moreover, in contrast to the related method nonparametric maximum margin criterion (NMMC) [40], a significant difference between them is that the within-class scatter and the between-class scatter is locally constructed by the k nearest within-class and between-class neighbors in MNMP while NMMC just use the nearest between-class point pair and the furthest within-class point pair to formulate the within-class scatter and the between-class scatter respectively, which can not make full use of local structure information of samples. Moreover, for each point, only the nearest inter-class

Table 4
Mean and the corresponding variance on Yale face database by performing LDA, NDA, LGE/MMC and MNMP.

Methods	Mean and variance (%)	Dimensions
LDA	88.32 ± 0.96	12
NDA	93.23 ± 1.78	10
LGE/MMC	94.89 ± 1.48	12
MNMP	96.55 ± 1.04	12

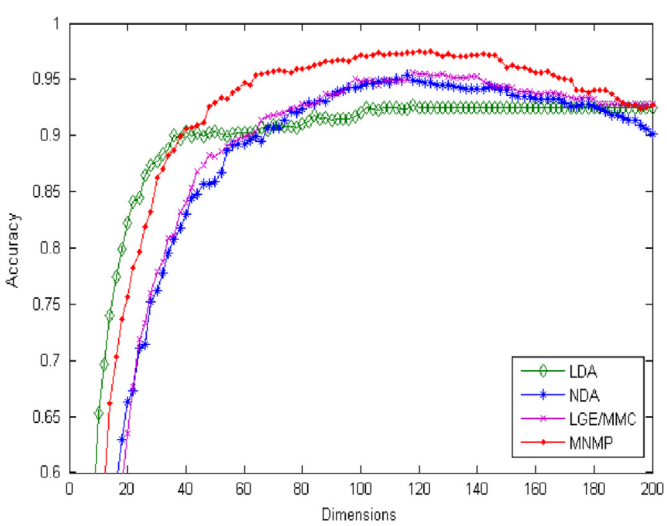

Fig. 4. The recognition accuracy curves with varied dimensions by performing LDA, NDA, LGE/MMC and MNMP on AR face database.

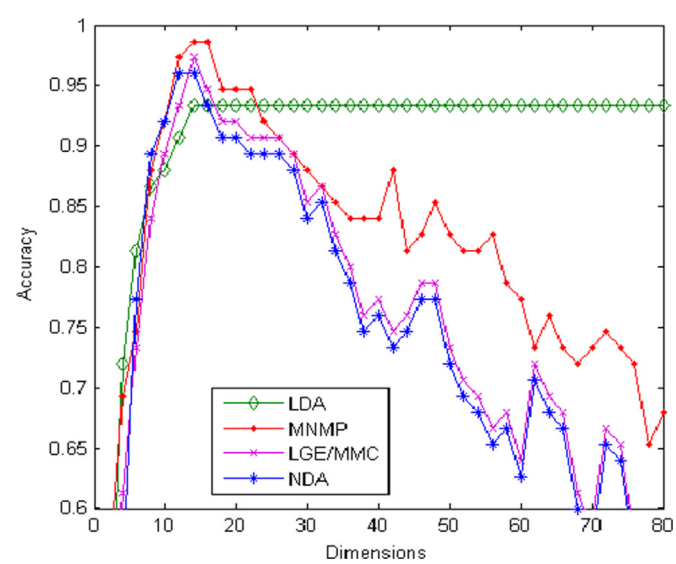

Fig. 6. The recognition rate accuracy with varied dimensions on Yale face database by using LDA, NDA, LGE/MMC and MNMP.

Fig. 5. Sample images of one person in Yale database.

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105

FREE paper freepaper.me دانلود کننده مقالات علمی

point and the furthest intra-class point are contained when constructing the nonparametric between-class scatter and the nonparametric within-class scatter respectively, which is not enough to measure the apartness among inter-class data and the compactness among intra-class data, especially outlier or noise involved. Finally, some experimental results based on benchmark databases such as AR face database, Yale face database and Palmprint database show that our proposed method is indeed effective and efficient.

Acknowledgments

This work was partly supported by the Grants of the Natural Science Foundation of China (6127303, 61273225, 61373109, 61472280, 61403287, and 61572381), China Postdoctoral Science Foundation (20100470613 and 201104173), Natural Science Foundation of Hubei Province (2010CDB03302), the Research Foundation of Education Bureau of Hubei Province (Q20121115) and the Open Project Program of the National Laboratory of Pattern Recognition (201104212).

References

[1] D.S. Huang, Systematic Theory of Neural Networks for Pattern Recognition (in Chinese), Publishing House of Electronic Industry of Beijing, 1996.
[2] Z.L. Sun, K.M. Lam, Z.Y. Dong, H. W., Q.W. Gao, C.H. Zheng, Face recognition with multi-resolution spectral feature images, PLoS One 8 (2) (2013) 1–12.
[3] D.S. Huang, H.J. Yu, Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids, IEEE/ACM Trans. Comput. Biol. Bioinform. 10 (2) (2013) 457–467.
[4] D.S. Huang, C.H. Zheng, Independent component analysis based penalized discriminant method for tumor classification using gene expression data, Bioinformatics 22 (15) (2006) 1855–1862.
[5] Joliffe, Principal Component Analysis, Springer-Verlag, Berlin, 1986.
[6] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.
[7] M. Turk, A. Pentland, Face Recognition Using Eigenfaces, in: Proceedings of IEEE Computer Society Conference on Computer Vision Pattern Recognition, 1991, pp. 586–591.
[8] K. Fukunnaga, Introduction to Statistical Pattern Recognition, second ed., Academic Press, New York, 1991.
[9] J. Ye, R. Janardan, C. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 982–994.
[10] D.S. Huang, Wen Jiang, A general CPL-AdS methodology for fixing dynamic parameters in dual environments, IEEE Trans. Syst., Man. Cybern.-Part B 42 (5) (2012) 1489–1500.
[11] D.S. Huang, Ji-Xiang Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Netw. 19 (12) (2008) 2099–2115.
[12] D.S. Huang, Radial basis probabilistic neural networks: model and application, Int. J. Pattern Recognit. Artif. Intell. 13 (7) (1999) 1083–1101.
[13] D.S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. Neural Netw. 15 (2) (2004) 477–491.
[14] Z.L. Sun, K.M. Lam, Q.W. Gao, Depth estimation of face images using the nonlinear least-squares model, IEEE Trans. Image Process. 22 (1) (2013) 17–30.
[15] H. Yang, C. Zhang, The research on two kinds of restricted biased estimators based on mean squared error matrix, Commun. Stat.: Theory Methods 37 (1) (2008) 70–80.
[16] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.
[17] P. Belhumeur, J. Hespanda, D. Kiregeman, Eigenfaces versus fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[18] H. Chen, H. Chang, T. Liu, Local discriminant embedding and its variants, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2005) 846–853.
[19] K. Etemad, R. Chellappa, Face recognition using discriminant eigenvectors, Proc. IEEE Intel. Conf. Acoust. Speech Signal Process. 4 (1996) 2148–2151.
[20] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, J. Opt. Soc. Am. A 14 (8) (1997) 1724–1733.
[21] M. Loog, R.P.W. Duin, Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004) 732–739.
[22] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 18 (8) (1996) 831–836.
[23] X. Wang, X. Tang, A unified framework for subspace face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1222–1228.
[24] X. Wang, X. Tang, Dual-space linear discriminant analysis for face recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2004) 564–569.
[25] X. Wang, X. Tang, Random sampling LDA for face recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2004) 259–265.
[26] X. Wang, X. Tang, Random sampling for subspace face recognition, Int. J. Comput. Vis. 70 (1) (2006) 91–104.
[27] M. Yang, Kernel eigenface versus kernel fisherface: face recognition using kernel methods, in: Proceedings of Fifth International Conference on Automatic Face Gesture Recognition, 2002.
[28] W. Zhao, R. Chellappa, Discriminant analysis of principal components for face recognition, in: Proceedings of Third IEEE Conference on Automatic Face Gesture Recognition, 1998, pp. 336–341.
[29] W. Zhao Discriminant component analysis for face recognition, in: Proceedings of International Conference Pattern Recognition, vol. 2, 2000, pp. 818–821.
[30] W. Zhao, R. Chellappa, N. Nandhakumar, Empirical performance analysis of linear discriminant classifiers, in: Proceedings of IEEE Conference Computer Vision Pattern Recognition, 1998, pp. 164–169.
[31] W. Zheng, C. Zou, L. Zhao, Real-time face recognition using Gram-Schmidt orthogonalization for LDA, in: Proceedings of IEEE Conference Pattern Recognition, 2004, pp. 403–406.
[32] Z. Li, D. Lin, X. Tang, Nonparametric discriminant analysis for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (4) (2009) 755–761.
[33] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[34] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[35] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[36] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, SIAM J. Sci. Comput. 26 (1) (2004) 313–338.
[37] K. Weinberger, L. Saul, Unsupervised learning of image manifolds by semidefinite programming, in: Proceedings of IEEE International Conference Computer Vision Pattern Recognition, 2004, pp. 988–995.
[38] T. Lin, H.B. Zha, Riemannian manifold learning, IEEE Trans. Pattern Anal. Mach. Intell. 30 (5) (2008) 796–809.
[39] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Netw. 17 (1) (2006) 157–165.
[40] X. Qiu, L. Wu, Nonparametric maximum margin criterion for face recognition, in: Proceedings of International Conference Image Processing, 2005, pp. 1413–1416.
[41] M. Wan, G. Shan, J. Shao, Local graph embedding based on maximum margin criterion (LGE/MMC) for face recognition, Informatica 36 (1) (2012) 103–112.
[42] P.Y. Han, A.B.J. Teoh, Maximum neighborhood margin criterion in face recognition, Opt. Eng. 48 (4) (2009) 047205.
[43] G.F. Lu, Z. Lin, Z. Jin, Face recognition using discriminant locality preserving projections based on maximum margin criterion, Pattern Recognit. 43 (10) (2010) 3572.
[44] Y. Chen, W.S. Zheng, X.H. Xu, J.H. Lai, Discriminant subspace learning constrained by locally statistical uncorrelation for face recognition, Neural Netw. 42 (1) (2013) 28–43.
[45] J. Liu, H. Huang, Local Fisher discriminant analysis with maximum margin criterion for image recognition, in: Proceedings International Conference Computer Graphics, Imaging and Visualization, 2011, pp. 92–97.
[46] ⟨http://www4.comp.polyu.edu.hk/~biometrics/⟩.
[47] A.M. Martinez, R. Benavente, The AR Face Database, ⟨http://rvl1.ecn.purdue.edu/aleix/~aleix_face_DB.html⟩, 2006.
[48] Yale University Face Database, ⟨http://cvc.yale.edu/projects/yalefaces/yalefaces.html⟩, 2002.

Bo Li received his M.Sc. and Ph.D. degree in Mechanical and Electronic Engineering from Wuhan University of Technology in 2003, Pattern Recognition and Intelligent System from University of Science and Technology of China in 2008, respectively. Now, he is an associated professor at School of Computer Science and Technology, Wuhan University of Science and Technology. His research interests include machine learning, pattern recognition, image processing and bioinformatics.

freepaper.me FREE paper

ARTICLE IN PRESS

Jing Du received her B.S. in Computer Science and Technology from Wuhan University of Science and Technology in 2013. Now, she is a master candidate in School of Computer science and technology, Wuhan University of Science and Technology. Her research interests include machine learning, bioinformatics.

Xiao-Ping Zhang (M'97, SM'02) received B.S. and Ph.D. degrees from Tsinghua University, in 1992 and 1996, respectively, both in Electronic Engineering. He holds an MBA in Finance, Economics and Entrepreneurship with Honors from the University of Chicago Booth School Of Business, Chicago, IL. He is a professor in Ryerson University. He is currently an Associate Editor for IEEE Transactions on Signal Processing, IEEE Transactions on Multimedia, IEEE Signal Processing letters and for Journal of Multimedia. His research interests include intelligent computing, bioinformatics, multimedia communication and signal processing.

freepaper.me FREE paper

Please cite this article as: B. Li, et al., Feature extraction using maximum nonparametric margin projection, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.11.105