Locally Linear Representation Fisher Criterion

Bo Li, Jin Liu, Zhong-Qiu Zhao, and Wen-Sheng Zhang

Abstract—In this paper, a novel supervised dimensionality reduction method based on LLE is put forward, which is titled locally linear representation Fisher criterion (LLRFC). In the proposed LLRFC, the class information of the original data has been fully considered, according to which an inter-class graph and an intra-class graph can be well modeled respectively. Meanwhile, the neighborhoods in the inter-class graph consist of samples with various labels and the neighborhoods in the intra-graph are just composed of points sampled from the same class. Then the locally linear representation technique is introduced to optimize the reconstruction weights in both graphs. At last, the Fisher criterion with maximum inter-class scatter and minimum intra-class scatter is reasoned. Experiments on some benchmark face data sets have been conducted and the results validate the proposed method’s performance.

I. INTRODUCTION

During last decades, manifold learning methods have been attracting many concentrations on machine learning and pattern recognition. Among them, besides isometric mapping (ISOMAP) [1], Hessian locally linear embedding (HLLE) [2], Laplacian eigenmaps (LE) [3-4], local tangent space alignment (LTS) [5], maximum variance unfolding (MVU) [6] and Riemannian manifold learning (RML) [7], locally linear embedding (LLE) [8-9] is computationally simpler and can offer useful results on a broader range of manifolds. But LLE also has some intrinsic drawbacks such as sensitiveness to noise, selection of neighbors, ill-conditioned eigenproblems, out-of-sample problem and ignorance of class information [10]. When applying LLE to classification, attentions are mainly paid to its supervised learning because it is crucial to improve recognition performance.

Recently, more and more supervised versions of LLE have been presented to deal with multiple manifolds corresponding to various classes. In the original LLE, the manifold local geometry is usually explored using k nearest neighbors (KNN) graph, where Euclidean distance is involved. In most cases, some points with different labels may also have a shorter Euclidean distance than those with the same class when constructing the KNN graph, which leads to wrong neighborhoods for classification that contain neighbors from different classes. To address the problem, a method to adjust neighborhood weights is advanced by taking the class information into account, where the distance between any two points belonging to different classes is defined to be relatively larger than their Euclidean distance while those distances between two points with the same label are preserved. The work was first presented by de Ridder et al. [11], where only the Euclidean distances between points belonging to different classes are simply enlarged by adding a constant. Instead of enlarging the between-class distances, Wen et al. utilized a nonlinear function to shrink the within-class distances [12], which shows similar impacts on recognition performance. These methods just either enlarge between-class distances or shrink within-class distances. Thus Zhang brought forward an enhanced supervised LLE model by reducing within-class distance and expanding between-class distance simultaneously [13]. Combining to the class information, these methods endeavor to increase the accuracy of LLE by adjusting the distances between neighborhood points rather than by selecting the neighborhoods points. So Hui et al. [14] and Zhao et al. [15] imposed a strict constraint on construction of KNN graph that only points with the same class can be considered to be neighbors. Nevertheless, when points are not sampled densely, the neighborhoods points determined by the method mentioned above will be not enough to explore the manifold geometry structure. Thus, Han et al. proposed a method to make a supplement [16]. According to the ascending Euclidean distances, the first same class samples are predefined as neighborhood points, and then the remaining neighbors are searched from points with different classes. Later, Zhang and Zhao introduced the probability-based distance that can enlarge the Euclidean distance for labeled and unlabelled points [17-18]. However, this modified version of LLE mainly just takes advantage of class information to adjust the distances between points or to select the neighborhood points in KNN graph, where more parameters are introduced with the augment of the application difficulty.

In addition, some other supervised LLE algorithms combined with LDA have also been boomed. Based on the projection distances of the preprocessed points in LDA subspace, Pang et al. selected the k minimum-distance points...
as the neighbor set of each data point and then used LLE [19],
which can be viewed as the mode of LDA+LLE. Zhang et al.
presented a unified framework of LLE and LDA [20-21].
This framework essentially equals to LLE+LDA. Pang et al.
also brought forward a model which is linearly constructed by
the objective function of LLE and LDA with some constraints
[22]. The model can be either LLE or LDA when the
coefficient is one or zero, respectively. Furthermore, a local
Fisher embedding (LFE) was put forward by de Ridder et al.
in 2004 [23], where local geometry and global class
information were combined to a Fisher formulation. Li et al.
also proposed a supervised LLE algorithm named local linear
discriminant embedding (LLDE) based on the fact that the
embedding cost function is invariant to translation under
sum-to-one constraint in LLE. Furthermore, the translations
can be optimized by a modified LDA [24]. However, those
methods employ class information to obtain discriminant
features globally instead of locally.

In this paper, different to previous works, a novel
supervised LLE method, named locally linear representation
Fisher Criterion (LLRFC) is brought forward, where label
information is introduced to locally construct different graphs
including an intra-class graph and an inter-class graph. In the
intra-class graph, the neighbor for any point is composed of
points with same label. On the contrary, in the inter-class
graph, any point and its neighbors should be labeled with
different classes. Moreover, the neighborhood points are
chosen from the sorted Euclidean distances to the point. After
construction of both graphs, the least locally linear
reconstruction can be modeled to obtain the optimal weights
between nodes in the corresponding neighborhoods. Because
each neighborhood in the intra-class graph consists of points
from the same class, for classification, it attempts to project
them with more compactness. But when projected into a low
dimensional space, the inter-class graph will show more
separability since its nodes represents various classes’ data
respectively. Thus a Fisher criterion can be reasoned to detect
a subspace where samples in the inter-class graph can be
partitioned and points in the intra-class graph can be
compressed as soon as possible.

The rest of this paper is organized as follows. Section 2
reviews related work including LDA and LLE. The proposed
LLRFC algorithm is described and justified in Section 3.
Section 4 presents some experiments. Finally, Section 5
draws some conclusions.

II. RELATED WORK

There are many dimensionality reduction methods related
to the proposed LLRFC. In the following, some famous
works such as LDA and LLE will be described.

A. LDA

Let \( S_B \) and \( S_W \) denote the between-class scatter and the
within-class scatter, which are defined as follows, respectively.

\[
S_B = \sum_{i=1}^{n} n_i (m_i - m)(m_i - m)^T
\]

\[
S_W = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (X_{ij} - m_i)(X_{ij} - m_i)^T
\]

where \( c \) is the number of class labels, \( n_i \) denotes the sample
number of the \( i \)th class, \( m_i \) and \( m \) stand for the \( i \)th class mean
and the mean of all sample points, respectively.

LDA looks for a linear subspace \( A \), within which the
projections of the samples from different classes are best
separated, as defined by maximizing the following
discriminant criterion.

\[
J(A) = \max \frac{tr[A^T S_B A]}{tr[A^T S_W A]}
\]

Along with the orthogonal constraint of \( A \), it can be solved
as a generalized eigenequation, which is stated below.

\[
S_B A = \lambda S_W A
\]

where \( A \) and \( \lambda \) are the \( i \)th generalized eigenvector and
eigenvalue of \( S_B \) with regard to \( S_W \). The LDA solution
contains the corresponding eigenvectors with non-zero
eigenvalues.

B. LLE

The goal of LLE is to map the high dimensional data into a
low dimensional manifold space, where the neighborhoods
including neighbors and their reconstruction weights will be
kept unchanged. The outline of LLE can be summarized as
follows.

Step1: For each data point \( x_i \), identify its \( k \) nearest
neighbors by KNN criterion or super-ball criterion;

Step2: Compute the reconstruction weights minimizing the
error of linearly reconstructing \( x_i \) by its \( k \) nearest neighbors;

Step3: Compute the low-dimensional embedding \( y \) for \( x \) that best preserves the local geometry represented by
the reconstruction weights.

Step 1 is typically done using Euclidean distance to define
neighborhood, although more sophisticated criterion may
also be involved in, such as Euclidean distance in kernel
space or cosine distance.

After identifying the \( k \) nearest neighbors of point \( x_i \), Step2
seeks the best reconstruction weights. Then repeating Step 1
and Step 2 to all the data points, at last all the reconstruction
weights consist of a weight matrix \( W \).

Step 3 in the LLE algorithm computes the optimal low
dimensional embedding \( y \) based on the weight matrix \( W \).
After deducing, it can be found that \( y \) is composed of the
eigenvectors associated with the first \( d \) smallest eigenvalues
except zero of the sparse matrix.

III. LOCALLY LINEAR REPRESENTATION FISHER CRITERION

A. Motivation

The original LLE is often applied for data visualization
because it can probe the intrinsic low dimensional manifold
structure embedded in high dimensional data space. However,
LLE can not efficiently extract the features for classification. The reason probably lies in that LLE non-linearly mines the high dimensional data by preserving the local manifold structure without considering the class information, i.e. LLE explores manifold locality by KNN graph, which is composed of data with the same label as well as data with different labels according to the sorted Euclidean distances. Due to the prosperity of locality preserving, the neighborhood for any point will be mapped into a low dimensional space without any alteration. Firstly, for any point, its k nearest neighbors will not change. Secondly, the reconstruction weights among a point and its k nearest neighbors will be preserved. At last, the same conclusion can be drawn to the labels of any point and its k nearest neighbors. However, in most cases, any points and its k nearest neighbors are labeled with various classes, which results in great difficulties in data classifying.

Therefore, both manifold locality preserving and class information should be combined to improve the recognition performance of LLE. On the one hand, the locality should be preserved, which helps to mine manifold distributed data; on the other hand, the class information can be used to supervise the construction of neighborhood to improve the classification performance. Based on the above motivations, a novel supervised method is proposed, where an inter-class graph and an intra-class graph characterizing the corresponding data are constructed on the basis of data labels. In the inter-class graph, the nearest neighbors of any point are sampled from points with different labels. While in the intra-class graph, any neighborhood contains points with the same class. Moreover, either the inter-class graph or the intra-class graph is constructed using KNN criterion, which can approximately approach the locality of manifold. Thus, both class information and manifold locality preserving can be well integrated for dimensionality reduction.

Fig 1 (b) can be determined by KNN criterion in the original LLE algorithm, the corresponding intra-class graph neighborhood and the inter-class graph neighborhood are illustrated in Fig 1 (c) and Fig 1 (d), respectively.

B. The intra-class graph and the inter-class graph

On the basis of the data labels, an intra-class graph can be constructed. For any sample, its k nearest neighbors should be of the same label as the sample; meanwhile, these neighbors are selected with the first k bottom Euclidean distances to the sample. In the intra-class graph, due to its local linearity, the sample can be well reconstructed by its k nearest neighbors with the optimal weights, which is stated as follows.

$\varepsilon(W_{\text{intra}}) = \min \left\{ \sum_{i,j} W_{ij}^2 \right\}$

Suffered the sum-to-one constraint for the reconstruction weights, Eq (5) can also be rewritten as follows.

$\varepsilon(W_{\text{intra}}) = \min \left\{ \sum_{i,j} W_{ij}^2 \right\}$

Then, it can be represented to the following formulation.

$\varepsilon(W_{\text{intra}}) = \min \left\{ \sum_{i,j} W_{ij}^2 \right\}$

where $W_{ij}^2 = [W_{i1}^2, W_{i2}^2, ..., W_{ik}^2], \ X_j = [X_{j1}, X_{j2}, ..., X_{jk}], \ G_i = X_i \cdot X_j$

However, in some cases, the local gram matrix is not positive definite, which can be usually avoided by adding a small multiple of identity matrix to the local gram matrix.

Noted the sum-to-one constraint in the objective function mentioned above, using Lagrange function, the weights can be obtained as follows.

$W_{\text{intra}} = \frac{\sum G_i^2}{\sum G_i^2}$

Based on the above objective function, the weights between any two points in the intra-class graph can be defined as follows.

$W_{\text{intra}} = \left\{ \begin{array}{ll} \frac{\sum G_i^2}{\sum G_i^2} & \text{if} \ X_j \in \text{Intra}(X_i) \\ 0 & \text{otherwise} \end{array} \right.$

where $\text{Intra}(X_i)$ denotes the intra-class neighborhood of point $X_i$.

Meanwhile, the inter-class graph can also be established as follows. Firstly, for any point $X_i$, super-ball or KNN criterion is introduced to determine its pre-defined neighborhood; secondly, the points located in the pre-defined neighborhood is sorted according to their ascending Euclidean distances to point $X_i$ and then the bottom k points with different labels to that of point $X_i$ are selected as its inter-class neighbors, which comprise the inter-class neighborhood of point $X_i$. The same process as in the intra-class graph is repeated to achieve the
optimal reconstruction weights in the inter-class graph.

\[
W_{\text{inter}} = \begin{cases} 
\sum_{i} G_{ij} X_{j} \in \text{Inter}(X_{i}) \\
0 \quad \text{otherwise}
\end{cases}
\] (10)

where \(\text{Inter}(X_{i})\) stands for the inter-class neighborhood of point \(X_{i}\).

C. Justification

As mentioned above, both the intra-class graph and the inter-class graph can be constructed with the corresponding optimal weights between nodes. Moreover, the proposed LLRFC aims to explore a low dimensional space with the best classification accuracy, where the data with the same label should be more clustered and the points with different labels should be projected farther. Note that the neighborhood for any node in the intra-class graph is of the same classes, thus it will contribute to data classification in two aspects to minimize reconstruction error in the neighborhoods of the intraclass graph. One is locality preserving, which the manifold learning methods just pursue. Moreover, it will assure that those nodes are located in a super-plane as far as possible, in other words, the smaller reconstruction error in the neighborhood, the more compactness of neighborhood points in the low dimensional space. The other is data classification. The neighborhood points in the intra-class graph are of the same class and it will help to search an optimal subspace with more compressed neighbors for any node, which meets data classification well. The minimized reconstruction error in the low dimensional space can be rewritten as follows.

\[
\varepsilon(Y) = \min \left\| Y - \sum_{j=1}^{k} W_{\text{inter}} Y \right\| 
\] (11)

Eq. (11) offers the minimized reconstruction error by the fixed weights \(W_{\text{inter}}\) between a point \(Y\) and its intra-class neighbors in the low dimensional space. As for all the points \(Y\), an intra-class scatter can be defined to minimize all the neighborhoods reconstruction errors in the corresponding matrix form as follows.

\[
S_{\text{inter}} = \min tr \{ Y M_{\text{inter}} Y^T \} 
\] (12)

where \(M_{\text{inter}} = (I - W_{\text{inter}}) Y (I - W_{\text{inter}})\).

However, for the task of classification, a low dimensional subspace will be found where the points with different labels will be more separable. Note that the point and its inter-class neighbors belong to different classes in any neighborhood, so in the proposed LLRFC, the reconstruction error in the neighborhoods of the inter-class graph should be maximized to scatter data points with different labels as far as possible, which will contribute more to data classification. Moreover, the reconstruction weights between neighborhood points are still preserved in the low dimensional space, so we can construct the following objective function with the fixed inter-class reconstruction weights.

\[
\varepsilon(Y) = \max \left\| Y - \sum_{j=1}^{k} W_{\text{inter}} Y \right\| 
\] (13)

Similarly, the inter-class scatter with the corresponding matrix form can be defined below.

\[
S_{\text{inter}} = \max tr \{ Y M_{\text{inter}} Y^T \} 
\] (14)

where \(M_{\text{inter}} = (I - W_{\text{inter}}) Y (I - W_{\text{inter}})\).

Both the inter-class scatter and the intra-class scatter are defined to find a low dimensional subspace with minimum compactness and maximum separability. Consequently, the optimal embeddings can be approached by solving the following multiple objective functions, simultaneously.

\[
S(Y) = \max \left\{ \frac{tr \{ Y M_{\text{inter}} Y^T \} }{tr \{ Y M_{\text{inter}} Y^T \} } \right\} 
\] (15)

Obviously, the above two objective functions can be transferred into a single one, where the inter-class scatter can be maximized and the intra-class scatter can be minimized in a low dimensional space, simultaneously.

\[
S(Y) = \max \left\{ \frac{tr \{ Y M_{\text{inter}} Y^T \} }{tr \{ Y M_{\text{inter}} Y^T \} } \right\} 
\] (16)

It must be noted that out-of-sample problem often occurs to manifold learning methods, thus a linear transformation \(Y = A'X\) is introduced to overcome the problem. Considering the constraint that \(Y_{\text{out}} Y^T = n I_{p\times d}\), which can be rewritten to \(A'X X^T A = n I\). So Eq. (16) can also be represented to the following constrained Fisher criterion.

\[
S(A) = \max \left\{ \frac{tr \{ A'X M_{\text{inter}} X^T A \} }{tr \{ A'X M_{\text{inter}} X^T A \} } \right\} 
\] (17)

s.t. \(A'X X^T A = n I\)

For Eq. (17), Lagrange multiplier method can also be adopted to find the linear transformation \(A\) as expected.

\[
\frac{\partial S}{\partial A} = 2X M_{\text{inter}} X^T A - 2XX^T A = 0 
\] (18)

Then, the following generalized eigen-equation can be deduced.

\[
X M_{\text{inter}} X^T A = \lambda X M_{\text{inter}} X^T A 
\] (19)

From Eq. (19), it can be concluded that \(A\) is spanned by the eigenvectors corresponding to top \(d\) eigenvalues of the above generalized eigen-equation.

IV. EXPERIMENTS

In this Section, the proposed LLRFC is compared with several related dimensionality reduction methods including linear discriminant analysis (LDA), neighborhood preserving embedding (NPE) [29], LLDE, which have been applied to AR and CMU PIE face data. Note that LLRFC has a Fisher form which may result in the small sample size (SSS) problem under the condition that the sample number is smaller than the feature dimensions of the original data [25-26]. So PCA is introduced in advance, where the dimensionality of the original data is reduced to some
dimensions to avoid the SSS problem and 100 percent energy is preserved. And then, LDA, NPE, LLDE and LLRFC are exploited. At last, a classifier with the nearest neighbor criterion is adopted to identify the features extracted by LDA, NPE, LLDE and the proposed LLRFC, respectively.

A. Experiments on AR face data

The AR face [27] contains over 4,000 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 and involved. The face portion of each image is manually cropped and then normalized to be size of $40 \times 50$ pixels.

Shown in Fig 2 are the classification accuracy curves with varied dimensions using LDA, LLRFC, LLDE and NPE with 5, 6, 7, 8 samples per-class selected as training sets, respectively, where the dimensions of LDA can only be reduced to 119. From Fig 2, it can be easily found that the proposed LLRFC can achieve the better accuracy than the others. Particularly, when 8 samples per-class are selected for training, LLRFC steadily outperforms the others.

Fig 3 exhibits the mean accuracies and their standard deviations by LDA, LLRFC, LLDE and NPE with 5, 6, 7, 8 training samples each class, where the experiments are repeated 20 times. From Fig 3, it can be found that the proposed LLRFC acquires the best mean accuracy no matter how many training samples are chosen.

When applying LDA, LLRFC, LLDE and NPE to extract features from the original AR face data, the maximum accuracies and the corresponding dimensions are shown in Table 1, where the proposed LLRFC outperforms the other methods.

B. Experiments on CMU PIE face data

The CMU PIE face image database [28] is becoming a benchmark database to test and evaluate feature extraction methods for face recognition, where 68 subjects with 41,368 face images as a whole are contained. The face images were captured by 13 synchronized cameras and 21 flashes under varying pose, illumination, and expression. In this experiment,
Firstly, the relations between classification performance and different feature dimensions are explored, where 60, 70 and 90 training samples can be randomly selected from the CPU PIE subset and the rest 110, 100 and 80 samples are employed for test. Shown in Fig 4 are the accuracy curves with varied dimensions and training samples. The recognition accuracies show the increasing trend with the increasing of dimensions. However, when the recognition accuracies achieve its maxima, they almost keep unchanged.

From Fig 4 it can be found that LLRFC superiors to LDA, LLDE and NPE after the dimension climb to 40.

Displayed in Fig 5 are the mean accuracies and the corresponding standard deviations with various training samples in CMU PIE subset, where each experiment has been conducted 10 times. The conclusion can also be drawn from Fig 8 that no matter how many training samples are chosen, the mean accuracies of LLRFC are always at the top level compared to those of LDA, LLDE and NPE.

Table 2 presents the maximum accuracies and the corresponding dimensions for various training samples.
Compared to LDA, LLDE and NPE, LLRFC achieves the best classification performance at 80, 84, 76, 84 dimensions for 60 trainings, 70trainings, 80trainings, 90trainings, respectively.

V. CONCLUSIONS

In this paper, we reviewed some extensions to LLE to solve the supervised learning problem, which either use class information to adjust distances between two nodes in KNN graph or combine LDA to LLE for features extraction discriminatingly. Different to those supervised versions, the LLRFC is presented for face recognition, where a novel supervised way is introduced. In LLRFC, labels are introduced to supervise the construction of the inter-class graph and the intra-class graph respectively. In the intra-class graph, a point and its KNN neighborhood points belong to the same class. Oppositely, the label for any point must different to those of its k nearest neighbors in the inter-class graph. Thus the least reconstruction error trick can be exploited in both the inter-class graph and the intra-class graph to achieve the optimal reconstruction weights. At last, preserving the weights, a Fisher criterion can be modeled to explore a low dimensional space for face recognition. Compared to other state-of-the-art feature extraction methods, which are relative to the proposed LLRFC, experiments results on AR and CMU PIE face data show that the our LLRFC is efficient and feasible. However, in the proposed LLRFC, the neighborhood parameter k in the inter-class graph and the intra-class graph are set to be the same. In the future, the works will be performed to explore the classification performance under circumstance that k is different when constructing both graphs.

ACKNOWLEDGMENTS

This work was partly supported by the grants of the National Natural Science Foundation of China (61005076, 61273303, 61273225 &U1135005), China Postdoctoral Science Foundation (20100470613 & 2011041473) · Natural Science Foundation of Hubei Province (2010CDB03302 &2011CDC076), the Research Foundation of Education Bureau of Hubei Province (Q20121115), the Program of Wuhan Subject Chief Scientist (201150530152), Hong Kong Scholars Program (XJ2012012) and the Open Project Program of the National Laboratory of Pattern Recognition (201104212).

REFERENCES