

An Adaptive Neighborhood Choosing of the Local Sensitive Discriminant Analysis Algorithm

Gao Weijun^{*}, Zhang Hong and Ke Ming

School of Computer and Communication, Lanzhou University of Technology, Gansu, Lanzhou, 730050, P.R. China

Abstract: The curse of dimensionality is a problem of machine learning algorithm which is often encountered on study of high-dimensional data, while LSDA (Locality Sensitive Discriminant Analysis) can solve the problem of curse of dimensionality. However, LSDA can not fully reflect the requirements that the manifold learning for neighborhood, by using the adaptive neighborhood selection method to measure the neighborhood, it proposes an adaptive neighborhood choosing of the local sensitive discriminant analysis algorithm. Experimental results verify the effectiveness of the algorithm from the ORL and YALE face database.

Keywords: Locality Sensitive Discriminant Analysis, Manifold learning, Neighborhood Choosing, Face Recognition.

1. INTRODUCTION

Face recognition has a great prospect in the identification, file management, credit card verification, human-computer interaction and other aspects of human life. It has become the research focus of pattern recognition and computer vision and other areas. Face data has feature of "high-dimensional, small sample", so having a reasonable dimensionality reduction for the data is particularly important. Subspace analysis is an important method of face recognition as it has a small computational cost but strong descriptive and divisibility characteristics. Among these methods, principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are the most widely used linear dimension reduction algorithms.

In recently years, many scholars proposed some methods which is aimed at discovering the nonlinear manifold structure of data, such as ISOMAP (ISOMetric MAPPING) [3], LLE (Locally Linear Embedding) [4], LE (Laplacian Eigenmap) [5] and so on. However, these methods can not be effectively for testing sample points, that is, Out-of-sample problem [6]. The linearization of manifold learning algorithm makes the neighbor of each sample's local neighborhood as a linear space, it can effectively solve the above problem. The representative algorithms include LPP (Locality Preserving Projection) [7], LSDA (Locality Sensitive Discriminant Analysis) [8], LPP is a linear approximation of Laplacian Eigenmaps which can remain neighbors relationship under the process of dimensionality reduction. LPP do not use the sample label information, the divisibility of unsupervised dimensionality reduction methods have limitations, so for face recognition, supervised

dimensionality reduction methods can achieve more satisfactory results. LSDA is designed to overcome the shortcomings of LDA which is unable to found the local geometry, the aim of this method is to discover the local geometrical characteristics of the data sets, and find the optimal projection to maximize the interval of sample points from different classes, while similar sample points are closer. However, the existing way of built neighborhood of LSDA is based on prior experience to setting the parameters of neighbors or neighbors. Manifold learning algorithms are more sensitive to its parameters, the parameter is too large or too small will affect the performance of the algorithm.

So the paper proposes an adaptive neighborhood selection of local sensitivity discriminant analysis algorithm for the problem of neighborhood choosing. This algorithm uses adaptive neighborhood selection algorithm under LSDA to choose the size of each sample's neighborhood. The algorithm can strike a neighbor sample data more accurately through the adaptive algorithm and further improve the robustness of the classification.

2. LOCALITY SENSITIVE DISCRIMINANT ANALYSIS

The goal of LSDA is that the similar data points of original space need more closely in low-dimensional space, and after the dimensionality reduction, the distance of different types of data points in the low-dimensional space should be as large as possible. Suppose $\{(x_1, l_1), (x_2, l_2), \dots, (x_N, l_N)\}$ are the training sample data sets, where $x_i \in R^D$ represents a D-dimensional vector, is category label of sample x_i . For every data point x_i , LSDA finds the k nearest point, denoted by $N(x_i)$. Construct the every data point's within-class map G_w and inter-class map G_b .

Through the projection matrix $A_{D \times d}$, the data point x_i can get the low-dimensional mapping $y_i \in R^d, d < D$, so $y_i = A^T x_i$.

Use $W_{w,ij}$ and $W_{b,ij}$ to represent weight matrix of map G_w and G_b which are defined as follows:

$$W_{b,ij} = \begin{cases} 1, x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, \text{other} \end{cases}$$

$$W_{w,ij} = \begin{cases} 1, x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, \text{other} \end{cases}$$

Where means nearest neighbor which have different marks with x_i , $N_w(x_i)$ means k nearest neighbor which have same marks with x_i .

To solve the projection direction after dimensionality reduction for classification [10], put the data points which have edge connection in with-class map as close as possible after projection, while put the data points which have edge connection in inter-class map as far as possible. Therefore, the aim function of LSDA mapping can be expressed by the following equation:

$$\min \sum_{ij} (y_i - y_j)^2 W_{w,ij} \quad (1)$$

$$\max \sum_{ij} (y_i - y_j)^2 W_{b,ij} \quad (2)$$

In minimize equation (1), the same class nearest neighbor points in high-dimensional space should be close in low-dimensional space after dimensionality reduction, and in maximize equation (2), the different class nearest neighbor points in high-dimensional space should be far in low-dimensional space after dimensionality reduction.

From the objective function (1):

$$\begin{aligned} & \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{w,ij} \\ &= \frac{1}{2} \sum_{ij} (a^T x_i - a^T x_j)^2 W_{w,ij} \\ &= \sum_i a^T x_i D_{w,ii} x_i^T a - \sum_{ij} a^T x_i D_{w,ij} x_j^T a \\ &= a^T X D_w X^T a - a^T X W_w X^T a \end{aligned} \quad (3)$$

From the objective function (2):

$$\begin{aligned} & \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{b,ij} \\ &= \frac{1}{2} \sum_{ij} (a^T x_i - a^T x_j)^2 W_{b,ij} \\ &= a^T X (D_b - W_b) X^T a \\ &= a^T X L_b X^T a \end{aligned} \quad (4)$$

Under the constraints $a^T X D_w X^T a = 1$, eigenvector set which corresponds to the largest eigenvalue in generalized

eigenvalue problem can be used as the final LSDA projection matrix:

$$X(\mu L_b + (1-\mu)W_w)X^T a = \lambda X D_w X^T a \quad (5)$$

Where is a constant coefficient within $[0,1]$, and it is determined by experimental verification. L_b is Laplace operator in inter-class map, W_w is weight matrix of within-class, D_w is diagonal matrix.

3. ADAPTIVE NEIGHBORHOOD CHOOSING OF THE LOCAL SENSITIVE DISCRIMINANT ANALYSIS

3.1. Adaptive Neighborhood Choosing Algorithm

The existing algorithm almost uses the global K -neighborhood in constructing the nearest neighbor graph which is more difficult for treatment of non-uniform manifold. Therefore, this paper follows the adaptive neighborhood selection method that proposed in reference [9] to determine the size of the neighborhood of each point which could solve the problem. If the proportion of Euclidean distance and corresponding geodesic distance is smaller between two points, the more bent of the local manifold between two points, the smaller of neighborhood parameter, and vice versa. The neighborhood parameter determination method needs to calculate geodesic distance. If we directly calculate all input data between geodesic distances, time complexity is very high which is approach to $O(|N|^3)$ and it is more time-consuming. In face, we only use the Euclidean distance to select an appropriate neighborhood which uses this neighborhood to calculate geodesic distance of sample points. The method is first to identify the local area size K , and then make as the center, select the nearest neighbor points constitute the local region and estimate geodesic distance between to other K points. Finally, use the relationship between geodesic distance and Euclidean distance to dynamically determine size of each point's neighborhood. The algorithm is as follows:

Algorithm 1: Identify neighborhood parameters of each point

Input: is high-dimensional observed data, and are the initial size of neighborhood

Output: the neighborhood parameter K_i of each point $X_i \in X$.

Step 1: Use the Euclidean distance to calculate the K -neighborhood of point in set X and constitute the local data set $N_K(X_i)$.

Step 2: Use ISOMAP method to calculate the local geodesic distance for any two points of local data set $N_K(X_i)$, it mainly have two steps:



Fig. (1). Example of ORL face database images.

① Construct weight graph according to neighborhood of and K . is corresponding to the data in $N_K(X_i)$, is edge set which is connecting two points in V , $(X_i, X_j) \in E$, if is K -neighborhood of X_j , The distance between and is Euclidean distance $d_e(X_i, X_j)$.

② Estimate all point geodesic distance $d_g(X_i, X_j)$ of local manifolds from through getting shortest distance between any two points in G . First, for any $(X_i, X_j) \in E$, set $d_g(X_i, X_j) = d_e(X_i, X_j)$; Otherwise, set $d_g(X_i, X_j) = \infty$. Then use all t and have a iterative calculation of all $d_g(X_i, X_j) = \min\{d_g(X_i, X_j), d_g(X_i, X_t) + d_g(X_t, X_j)\}$.

Step3: Calculate the proportion between all points in of the Euclidean distance and geodesic distance and take it as a bending measurement of $x_i \in X$ in local data manifold.

Step4: calculate the neighborhood parameter of X_i . Its basic idea is that those which have mean of data should put as neighborhood parameter, other data points have a regulation as a center of K . That is:

3.2. The step of Facial Feature Extraction Based on LSDA

Step 1. for any $x_i \in X_N$, select nearest neighbor K and heterogeneous point neighborhood K of each point's same class point by algorithm 1, then build the within and between class weight matrix W_W, W_B .

Step 2. Solve the eigenvalue and eigenvector for (5). Get eigenvector of D nonzero eigenvalues, the feature vector can be written as $a_1, a_2, \dots, a_d, a_{d+1}, \dots, a_D$, that are the eigenvector which are correspond to the d largest eigenvalues.

Step 3. Set $A = [a_1, a_2, \dots, a_d]$, it is the projection matrix of $D \times d$. Samples in D -dimensional space can transform through and project to D dimensional space.

4. THE EXPERIMENT OF FACE RECOGNITION

We have an experiment on the ORL and YALE face database to verify the performance of the algorithm, and compare with PCA, LDA, LPP, LSDA Respectively. The

initial neighborhood $K = 4$, experiments use the nearest neighbor classifier to classify [11], and recognition rate is the maximum average of 10 times experiment. All face images are grayscale images and will be scaled to 32×32 .

4.1. The Experiment on ORL Face Databases

ORL face database have 40 people and everyone has 10 images, these face images are got at different times. It includes the different perspectives, different facial expressions (eyes open, eyes closed, smiling, angry, happy, surprised, etc.) and different facial details (different hair, wearing or not wearing glasses, with or without beard), At the same time, it contains no more than 20° tilt and rotation. The different images of ORL face database are shown in Fig. (1).

For ORL face database, we extract $l(l=2,3,\dots,6)$ images randomly for training samples, The rest $10-l$ pictures are test samples.

From Table 1 and Fig. (2), our algorithm achieves the highest recognition rate on ORL face database. Because it uses adaptive neighborhood to selected sample point and makes it more accurate to maintain the locality of manifold space, so recognition rate is the highest. At the same time we can see the recognition rate of LSDA algorithm is higher than the LDA and PCA algorithm which can fully reflects the advantages for local characteristics of manifold learning.

4.2. The Experiment on YALE Face Database

To further verify the effectiveness of the algorithm, experiment uses Yale database which includes 165 images, it has an expression samples of 15 person and everyone has 11 images. These images have quite big changes in lighting conditions and facial expressions (normal, sad, happy, blinking and surprised, etc.) and details (to wear or not wear glasses) are also have an obvious changes. Fig. (3) shows 10 different images of YALE face database.

We extract $l(l=2,3,\dots,6)$ images randomly for training samples from YALE face database, The rest $11-l$ pictures are test samples.

From Table 2 and Fig. (4), we can see that the images on YALE library get interference by light, gesture, facial expressions is high, the recognition rate from all the methods listed in the text are relatively low. PCA and LDA can

Table 1. Comparing of recognition results on ORL face database.

Algorithm	Number of Samples and the Correct Recognition Rate%				
	2	3	4	5	6
PCA	66.31	77.58	81.34	85.76	88.12
LDA	72.15	82.34	89.06	91.26	92.45
LPP	74.82	84.53	89.74	92.38	93.72
LSDA	77.41	85.68	91.23	93.27	95.40
Our algorithm	78.63	87.69	93.57	95.16	97.06

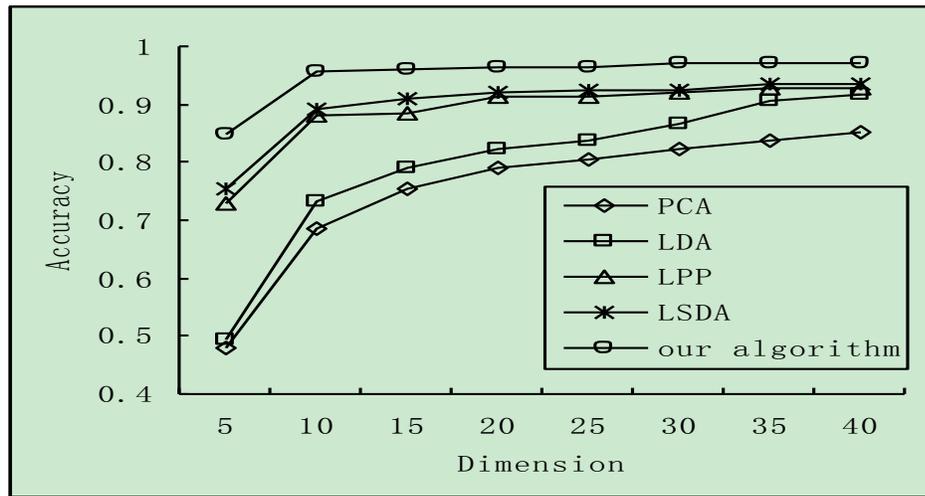


Fig. (2). 5Train recognition rates on ORL face database.



Fig. (3). Examples of YALE face atabase images.

Table 2. The comparing of recognition results on YALE face database.

Algorithm	Number of Samples and the Correct Recognition Rate%				
	2	3	4	5	6
PCA	42.47	49.34	54.38	56.43	59.25
LDA	47.41	59.67	70.08	74.16	78.72
LPP	56.27	68.28	76.24	77.55	80.92
LSDA	56.51	69.14	77.35	80.73	83.69
Our algorithm	57.90	71.52	80.76	83.15	85.56

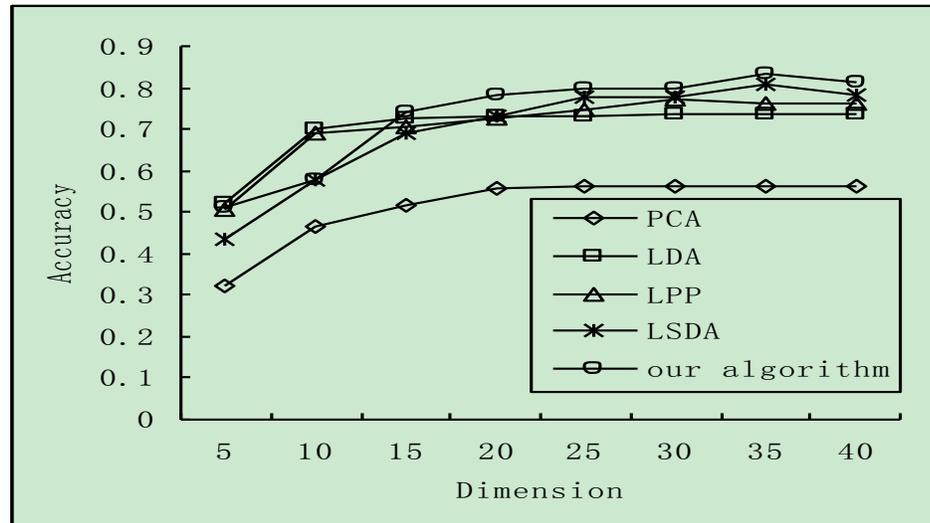


Fig. (4). 5-Train recognition rates on YALE face database.

maintain global linear structure to have dimensionality reduction [12], while it can not describe the nonlinear manifold very accurately, recognition rate is lower than the LSDA which can maintain the local linear and our algorithm, the recognition rate of our algorithm is higher than the LSDA.

4.3. Experimental Analysis

We use adaptive neighborhood selection algorithm to calculate the neighbor points which is based on LSDA algorithm, the classified projection matrix that we obtained is conducive to effective. Experiments show that the algorithm reflects the very strong robustness for posture, facial expressions, illumination and other changes, and the effect is best when compared with several other algorithms. For identification, too much features will be carry the information which is not conducive to classify, when the dimension reaches a certain level, the recognition rate is no longer increasing with the increasing of the number of dimensions, and it may even decline. Because the selection of neighbors point for LPP and the LSDA use pre-set parameters k , the recognition rate is lower than our algorithm. The recognition rate of PCA is lower than LDA, because the PCA is unsupervised. The recognition rate of LDA is lower than the LSDA, and the recognition rate of LSDA is lower than our algorithm, it shows that the adaptive neighborhood selection algorithm can achieve more satisfactory performance.

CONCLUSION

For neighborhood selection of this manifold learning algorithm, we propose an adaptive neighborhood selection of local sensitive neighborhood discriminant analysis algorithm, the algorithm build neighborhood adaptively by the relationship of geodesic distance and Euclidean distance and it can reflect the local linear structure of the manifold. However, when the degree of nonlinearity is very high, the neighborhood will be too small and can not form the connected neighborhood graph. Through kernel techniques

to expand non-linear algorithm is the focus on future research.

ABOUT THE AUTHORS

First Author: Gao Weijun, associate professor, M.S. Graduated from the Graduate University of Chinese Academy of Sciences, research in field of pattern recognition, high performance computing and software engineering.

Second Author: Zhang Hong, Ph.D, is learning at School of Computer and Communication, Lanzhou University of Technology. The research field is large data processing.

Third Author: Ke Ming, Ph.D. Graduated from the Lanzhou University of Technology. The research field is the theory of complex network, biological information processing, pattern recognition.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the national science foundation of china (61263047).

REFERENCES

- [1] M. Turk and A. Pentland, "Eigenface for recognition", *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 72-86, 1997.
- [2] P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [3] S. Dhanya and M. Pankaj, "Wilscy Comparison of PCA, LDA and Gabor Features for Face Recognition Using Fuzzy Neural Network", *Advances in Intelligent Systems and Computing*, vol. 177, pp. 413-422, 2013.

- [4] S.L. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.
- [5] M. Belkin and P. Niyogi "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural Computation*, vol.15, no. 6, pp. 1373-1396, 2003.
- [6] Y. Bengio, J. Palement, P. Vincent, O. Delalleau, N. L. Roux and M. Ouimet, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering", *Neural Computation*, vol. 16, no. 10, pp. 2179-2219, 2004.
- [7] P. Hu, "Application research of gabor filter and LPP algorithms in face recognition", *Lecture Notes in Electrical Engineering*, vol. 144, pp. 483-489, 2012.
- [8] C. Deng and H. Xiaofei, "Locality sensitive discriminant analysis," *International Joint Conference on Artificial Intelligence*, 2007.
- [9] G. Wen, L. Jiang and J. Wen, "Dynamically determining neighborhood parameter for locally linear embedding", *Journal of Software*, vol. 19, no. 7, pp. 1666-1673, 2008.
- [10] S. Karamizadeh, S.M. Abdullah, M. Zamani, A. Kherikhah, "Pattern recognition techniques Studies on Appropriate Classifications", *Lecture Notes in Electrical Engineering*, vol. 315, pp. 791-799, 2015.
- [11] R. Marek, O. Tomasz and H. Image, "Processing, pattern recognition, and semantic understanding techniques", *Advances in Computer Vision and Pattern Recognition*, pp. 1-70, 2015.
- [12] C. Hou, F. Nie, H. Wang, D. Yi, C. Zhang, "Learning high-dimensional correspondence via manifold learning and local approximation," *Neural Computing and Applications*", vol. 24, no. 7-8, pp. 1555-1568, June 2014.

Received: December 15, 2014

Revised: January 04, 2015

Accepted: February 25, 2015

© Weijun et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.