A Local Tangent Space Alignment Based Transductive Classification Algorithm

Jianwei Yin[1], Xiaoming Liu[1], Zhilin Feng[1, 2], and Jinxiang Dong[1]

[1] Department of Computer Science and Technology, Zhejiang University, China
[2] Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, China
zjuyjw@zju.edu.cn, liuxiaoming@cs.zju.edu.cn, djx@cs.zju.edu.cn

Abstract. LTSA (local tangent space alignment) is a recently proposed method for manifold learning, which can efficiently learn nonlinear embedding low-dimensional coordinates of high-dimensional data, and can also reconstruct high dimensional coordinates from embedding coordinates. But it ignores the label information conveyed by data samples, and can not be used for classification directly. In this paper, a transductive manifold classification method, called QLAT (LDA/**Q**R and **L**TS**A** based **T**ransductive classifier) is presented, which is based on LTSA and TCM-KNN (transduction confidence machine-k nearest neighbor). In the algorithm, local low-dimensional coordinates is constructed using 2-stage LDA/QR method, which not only utilize the label information of sample data, but also conquer the singularity problem of traditional LDA, then the global low-dimensional embedding manifold is obtained by local affine transforms, finally TCM-KNN method is used for classification on the low-dimensional manifold. Experiments on labeled and unlabeled mixed data set illustrate the effectiveness of the method.

Keywords: manifold learning; local tangent space alignment; transductive inference; LDA/QR.

1 Introduction

Dimension reduction has long been an important problem in the fields of pattern classification, data mining and machine learning. With the development of information technology, especially the development of internet, more and more high-dimensional data, such as gene data, images and video emerges, the requirement of dimension reduction becomes more urgent.

Many high-dimensional data in real-world applications can be modeled as sets of points or vectors lying close to a low-dimensional nonlinear manifold. Discovering the structure of the manifold from such a sample of data points is a very challenging problem. Many dimension reduction algorithms have been proposed, and can be classified to two classes roughly: linear methods and nonlinear methods. PCA (principal component analysis) and LDA (linear discriminant analysis) are the most popular linear dimension reduction methods. While they have the advantages of easy understandable, simple to implement and can catch the linear structure of data, they can

F. Schwenker and S. Marinai (Eds.): ANNPR 2006, LNAI 4087, pp. 93 – 106, 2006.

not discover the nonlinear structure of data. In reality, many higher dimension data is embedded in a low nonlinear manifold, and there have some cues to show the low-dimensional embedding is consistent with human perception[1]. To address the shortcomings of the linear methods, kernel PCA method and kernel LDA method have been proposed by many researchers. Recently, there has been considerable interest in developing efficient algorithms, the so called manifold learning methods, to construct nonlinear low-dimensional manifolds from sample data points in high-dimensional spaces, and these methods have been regarded as effective approaches for nonlinear dimension reduction. In Isomap algorithm [2], pairwise *geodesic* distances of the data points instead of the Euclid distance are used with MDS (multidimensional scaling). The LLE (locally linear embedding) method [3] constructs a local geometric structure that is invariant to translations and orthogonal transformations in a neighborhood of each data point, and seeks to project the data points into a low-dimensional space that best preserves those local geometries. (A related method using Hessian matrices is presented in [4]). LTSA (local tangent space alignment) [5] methods constructs a local tangent space for each data point, and obtains the global low-dimensional embedding through affine transformation of the local tangent spaces.

While the LTSA algorithm can learn the low-dimensional nonlinear embedding coordinates of the higher-dimensional data, and can reconstruct the higher-dimensional coordinates from the low-dimensional coordinates. But as pointed out in [6], the best representative features are not always the best discriminant features for general classification task. In LTSA, class label information of data is ignored and so it can not be applied for classification directly. In the paper, we try to use the class label information and extend the LTSA algorithm from dimension reduction to classification problem. Traditional classification algorithms try to make the trained classifier optimal for all possible future data samples, but in practical, it is not needed and the classifier is usually only required to be optimal for specific unseen data sets. Transductive inference[7,8] learns the classification for unseen data directly from known data, and is more economic than traditional algorithms. Integrating LTSA and the idea of transductive inference, we proposed a TCM-KNN (transduction confidence machine-k nearest neighbor) [7,8] based manifold classification algorithm, called QLAT (LDA/**Q**R and **L**TS**A** based **T**ransductive classifier). The algorithm uses improved 2-stage LDA/QR algorithm [9] to construct local low-dimensional coordinate, then use LTSA method to retrieve the global embedding map for dimension reduction, finally, uses TCM-KNN on the low-dimensional embedding space for classification.

The rest of the paper is organized as follows: in Section 2, preliminary backgrounds are introduced, including LDA/QR, LTSA and TCM-KNN. In section 3, we describe in detail our proposed QLAT algorithm. Experiments result on synthetic and real data sets are presented in section 4. In section 5, we conclude and predict future work.

2 Preliminaries

QLAT algorithm is based on LTSA and TCM-KNN, which use the idea of LTSA to construct global embedding coordinates through affine transformation of the local

space. TCM-KNN is the transductive version of KNN algorithm, LTSA and TCM-KNN are introduced in 2.2 and 2.3 sections respectively. In the first stage of the original LTSA, PCA is used to construct the local coordinate, LDA is required to utilize the class label information. An intrinsic limitation of classical LDA is the so-called singularity problem, to deal with the singularity problem and improve the performance of the algorithm, we use 2-stage LDA/QR algorithm instead of classical LDA to construct the local embedding space. In section 2.1, LDA/QR algorithm is introduced briefly.

2.1 LDA/QR Algorithm

LDA/QR is a 2-stage dimension reduction algorithm proposed by Ye etc [9]. In the first stage of the algorithm, the separation of different classes is maximized via QR decomposition on the small matrix composed of class centers. This stage can be used independently as a dimension reduction, and the distinct property of this stage is the low time/space complexity. The second stage of LDA/QR refines the first stage by addressing the issue of within-class distance, and can be solved using the similar method for classical LDA, that is, by applying eigen-decomposition method.

2.2 LTSA Algorithm

LTSA is a nonlinear dimension reduction algorithm operated on tangent space. Data are assumed to lie on noised nonlinear low-dimensional manifold in the algorithm. Local tangent spaces are constructed for every data point with their k nearest neighborhoods. The final global coordinates are obtained through transfer, scaling, rotation and alignment of the local tangent spaces. During the alignment process, the local coordinates of a data point in the neighborhood with respect to the tangent space are to be preserved by all means. Min etc in [10] have proved that the local tangent space can be constructed with the eigen-vectors of the local covariance matrix, so the local tangent space projection problem can be converted into local PCA problem. Finally, the problem of obtaining global embedding coordinates can be converted into eigen-value problem of matrix.

2.3 TCM-KNN

TCM-KNN is a transductive algorithm. Transductive inference is a type of local inference that moves from particular to particular. In contrast to inductive inference where one uses given empirical data to find the approximation of a functional dependency and then uses the obtained approximation to evaluate the values of a function at the points of interest, one estimates the values of a function only at the points of interest in one step. The transductive inference approach uses the whole training set to infer a rule for each new exemplar. Transductive inference has a strong connection with Kolmogorov complexity, and is related with the notion of randomness deficiency, which is a measure of randomness. TCM-KNN is a transductive version of KNN method.

3 LDA/QR and LTSA Based Transductive Classifier (QLAT)

QLAT is based on LTSA and TCM-KNN. Firstly, local tangent space is constructed for each sample data using its nearest neighborhoods, discriminant analysis is performed on local tangent space, and low-dimensional local coordinates are obtained for nearest neighborhoods. Then, global low-dimensional coordinates are achieved through affine transforming of local spaces. Finally, TCM-KNN algorithm is performed on the low-dimensional manifold space. The procedure of construct low-dimensional manifold is similar with the LTSA, but the mathematic induce process have some differences for class label information is ignored in LTSA. In the following, we give the detail induce process.

Notation. We use I to denote the identity matrix, e to denote the column vector with all the element 1, $\|*\|_2$ denotes the 2-norm of a vector or matrix, $\|*\|_F$ denotes the Frobenius norm, A^T denotes the transpose of A, and A^+ denotes the Moore-Penrose generalized inversion.

Sample data set X, containing L labeled data and U unlabeled data, are assumed to evenly sampled from a noised low-dimensional manifold. That is, $X = X_L \cup X_U$, where $X_L = \{(x_1, y_1), ..., (x_L, y_L)\}$, $X_U = \{x_{L+1}, ..., x_{L+U}\}$, $y_i \in C = \{c_1, ..., c_{|C|}\}$, C is the class label set, let N=L+U, then $x_i = f(\tau_i) + \varepsilon_i$, $\tau_i \in R^d$, $x_i \in R^D$, i=1,…,N, and $D \geq d$. The classification problem is : given the labeled data set X_L and unlabeled data set X_U, label the sample data x_j in X_U with y_j, $y_j \in C$, $j = L+1, ..., N$.

To obtain the local coordinate of a data point p, LTSA uses the k nearest neighborhoods of p, and the local coordinates can be obtained with local PCA. As mentioned above, this local coordinate is not optimal for the problem of classification. Discriminant analysis is needed to utilize the class label information of data points, furthermore, data samples of each class are required to perform discriminant analysis on the local tangent space. However, in practice, labeled data is usually small and most data are unlabeled, so the direct applying of LDA is not appropriate. So, we use the method of PCA+LDA to obtain the local coordinates using nearest neighborhoods, and obtain the global embedding coordinate through affine transformations of the local coordinates. For a data point p, k nearest neighborhoods are used in LTSA, but the k nearest neighborhood can not guarantee the needed sample number of each class, that is, LDA may not be applicable, CK-NN construction method is proposed in paper [11]

to deal this problem for Isomap algorithm, and we apply a simple extension of KNN local space construction here. For each point p, find k_l the nearest neighborhoods p_j^i for each class C_i, i=1,…|C|, j=1,…,k_l, and $p_j^i \in X_L$, then, LDA is performed with these data points, while PCA is performed with the k'_l unlabeled nearest neighborhoods p_j (j=1,…,k'_l) of p, $p_j \in X_U$, usually $k'_l=k_l$. The calculation of optimal d dimensional approximation of data point p in the affine space is equal to the optimizing problem:

$$\min_{x,\theta,Q} \sum_{j=1}^{k_l'} \| p_j - (x + Q\theta_j) \|_2^2 + \lambda \frac{|Q^T S_W^p Q|}{|Q^T S_B^p Q|} = \min_{x,\Theta,Q} \| X^{p'} - (xe^T + Q\Theta) \|_F^2 + \lambda \frac{|Q^T S_W^p Q|}{|Q^T S_B^p Q|} \quad (1)$$

where Q is a D×d dimensional orthogonal matrix, $\Theta = [\theta_1, ..., \theta_{k'_l}]$, $X^{p'}$ is the k'_l unlabeled nearest neighborhoods of p, S_W^p is the within-class scatter matrix of p and S_B^p is the between-class scatter matrix. The definitions of S_W^p and S_B^p are: $S_W^p = \frac{1}{N^p} \sum_{i=1}^{|C|} \sum_{x \in X_i^p} (x - m_i^p)(x - m_i^p)^T$, $S_B^p = \frac{1}{N^p} \sum_{i=1}^{|C|} N_i^p (m_i^p - m^p)(m_i^p - m^p)^T$, where N^p is the labeled data number near to p, $N^p=|C|\times k_l$, N_i^p is the sample number near to p and belonging to class C_i, m_i^p is the mean of X_i^p, m^p is the overall mean of labeled data near to p.

The direct solving of above optimizing problem is difficult, and the problem can not be solved when S_B is singular, so we change the target function. There are many improvement of classical LDA, for example Pseudoinverse LDA, PCA+LDA, LDA/GSVD, LDA/QR. Among them, LDA/QR is a recently proposed 2-stage LDA algorithm by Ye etc, between-class distance is maximized during the first stage, and the optimization problem is $G = \arg\max_{G^T G = I_d} trace(G^T S_B G)$, which can be solved using QR decomposition. Within-class distance is minimized during the second stage, and the

target optimization problem is $W = \arg\min_{W} \text{trace}((W^T \widehat{S_B} W)^{-1}(W^T \widehat{S_W} W))$, where $\widehat{S_B}$ and $\widehat{S_W}$ are the reduced between-class and within-class scatter matrices respectively, and can be solved using eigen-decomposition of $\widehat{S_b^{-1}} \widehat{S_w}$. During the first stage of the original LDA/QR algorithm, only information of labeled data is utilized, that is, S_B is only related with labeled samples and information of large unlabeled sample is not utilized. We utilize the information of unlabeled data with LDA/QR algorithm, that is, change the optimization problem into

$$G = \arg\max_{G^T G = I_d} trace(G^T S_B G + \lambda G^T S_T G) \tag{2}$$

Where S_T is the total scatter matrix of unlabeled data set near to p, $S_T = \sum_{x_j^p \in X^{p'}} (x_j^p - m^p)(x_j^p - m^p)^T$, λ is a parameter used to adjust the weight of labeled samples and unlabeled samples in the construction of local coordinates, which in fact, is also the adjust of weight between LDA and PCA. The second stage of LDA/QR is the same. When the LDA/QR finished, the low-dimensional representation of x_i is $z_i = G^T x_i = G^T(x_i - \bar{x} + \bar{x}) = G^T \bar{x} + G^T(x_i - \bar{x})$, where G is a D×d matrix. Comparing it with LTSA, the local coordinates in the low-dimensional space of x_j^i near to x_i is $\theta_j^{(i)} = G^T(x_j^i - \overline{x^i})$, so $x_j^i = \overline{x_i} + G\theta_j^i + \zeta_j^i$, where $\zeta_j^i = (I - GG^T)(x_j^i - \overline{x^i})$ is the reconstruction error, and $\theta_j^{(i)}$ is the local coordinates of x_j^i in the low-dimensional space near x_i.

Now consider constructing the global coordinates τ_i, i=1,…,N, in the low-dimensional embedding space based on the local coordinates $\theta_j^{(i)}$ which represents the local geometry. Assuming the global coordinates can be obtained with affine transform of the local coordinates. Let τ_{ij} is the global embedding coordinate of x_{ij}, then $\tau_{ij} = \overline{\tau_i} + L_i \theta_j^{(i)} + \varepsilon_j^{(i)}$, j=1,…,$N_i$, i=1,…,N, N_i is the number of nearest

neighborhoods used during constructing the local coordinates of x_i, N_i is not related with data point x_i, and $N_i=|C|*k_l+k'_l=(|C|+1)*k_l$ if $k'_l=k_l$, denote $M=N_i$, $\overline{\tau_i}$ is the mean of τ_{ij}, j=1,…,M, $L_i \in R^{d\times d}$ is a local affine transformation matrix that needs to be determined, $\varepsilon_j^{(i)} \in R^d$ is the local reconstruction error. Denoting $T_i=[\tau_{i1},...,\tau_{iM}]\in R^{d\times M}$, $E_i=[\varepsilon_1^{(i)},...,\varepsilon_M^{(i)}]\in R^{d\times M}$ and $\Theta_i=[\theta_1^{(i)},...,\theta_M^{(i)}]\in R^{d\times M}$, we have $T_i=\frac{1}{M}T_i ee^T+L_i\Theta_i+E_i$, and the local reconstruction error matrix E_i has the form:

$$E_i=T_i(I_M-\frac{1}{M}ee^T)-L_i\Theta_i\in R^{d\times M} \tag{3}$$

To preserve as much as possible the local geometry in the global low-dimensional space, we seek to find τ_{ij} and L_i to minimize the reconstruction errors $\varepsilon_j^{(i)}$, i.e.,

$$\sum_i \|E_i\|_2^2 \triangleq \sum_i \left\|T_i(I-\frac{1}{M}ee^T)-L_i\Theta_i\right\|_2^2 = \min \tag{4}$$

Obviously, the optimal alignment matrix L_i that minimizes the local reconstruction error $\|E_i\|_F$ for a fixed T_i, is given by $L_i=T_i(I-\frac{1}{M}ee^T)\Theta_i^+$, and therefore, $E_i=T_i(I-\frac{1}{M}ee^T)(I-\Theta_i^+\Theta_i)$.

Let $T=[\tau_1,...,\tau_N]\in R^{d\times N}$ and $S_i \in R^{N\times N}$ be the 0-1 selection matrix such that $TS_i=T_i$, where τ_i is the global low-dimensional embedding coordinates of x_i, i=1,…,N. We then need to find T to minimize the overall reconstruction error $\sum_{i=1}^N \|E_i\|_F^2=\|TSW\|_F^2$, where $S=[S_1,\cdots,S_N]\in R^{N\times N^2}$, and $W=diag(W_1,\cdots,W_N)\in R^{N^2\times N^2}$ with

$$W_i=(I_N-\frac{1}{N}ee^T)(I_N-\Theta_i^+\Theta_i)\in R^{N\times N} \tag{5}$$

To uniquely determine T, we will impose the constraints $TT^T=I_d$. It turns out that the vector e of all ones is an eigen-vector of

$$B\triangleq SWW^TS^T \tag{6}$$

corresponding to the zero eigen-value. Therefore the optimal T is given by the d eigenvectors of B corresponding to the 2nd to d+1st smallest eigen-values, i.e., $T = [\tau_1, ..., \tau_N] = [u_2^T, ..., u_{d+1}^T]^T$, where τ_i is a d-dimensional column vector and u_j is the corresponding eigenvector of the jth smallest eigen-value of matrix B. There, the low-dimensional embedding coordinate of x_i is τ_i, i=1,…,N.

After obtaining the global embedding coordinates, classification can be applied on the low-dimensional manifold space. We adopt the TCM-KNN[7,8] for the classification task. As mentioned before, TCM-KNN is a transductive inference method, and it seeks to find, from all possible labelings L(W) on the working set W, the one that yields the largest randomness deficiency, i.e., the most probable labeling. Randomness deficiency is, however, not computable. One has to approximate it instead, using a slightly modified Martin-Lof test for randomness and the values taken by such randomness tests are called p-values. Given a sequence of distances from exemplar i to other exemplars, the strangeness of i with putative label y is defined as:

$$\alpha_y(i) = (\sum_{j=1}^{k} d_{ij}^{y})(\sum_{j=1}^{k} d_{ij}^{-y})^{-1} \tag{7}$$

The strangeness measure $\alpha_y(i)$ is the ratio of the sum of the k nearest distance d from the same class (y) divided by the sum of the k nearest distances from all the other classes (-y). The strangeness of an exemplar increases when the distance from the exemplars of the same class becomes larger and when the distance from the other classes becomes smaller. The smaller the strangeness, the larger its randomness deficiency is. The p-value for a working exemplar j (with putative label y) can be computed as:

$$p_y(j) = \frac{f(\alpha_1) + \cdots + f(\alpha_l) + f(\alpha_{new}^{y})}{(l+1) f(\alpha_{new}^{y})} \tag{8}$$

where l is the cardinality of the training set T, α_{new}^{y} is the strangeness measure of classifying a new sample into putative class y, f is monotonic nondecreasing function with f(0)=0, which can be defined as f(α)=α. TCM-KNN classify a sample j into class y, if

$$p_y(j) = \arg\max_{y}(p_y(j)) \tag{9}$$

With the above explanation, the procedure of QLAT can be presented as follows:

Table 1. QLAT Algorithm

Input: Data set X (including labeled data set X_L and unlabeled data set X_U), target embedding low-dimension d, number of nearest neighborhoods k_l and k'_l (usually equals to k_l) of LDA/QR, weight parameter λ, and k in TCM-KNN. Output: class label y'_i of data in X_U, low embedding coordinates T of samples in X.
1. dimension reduction with QR decomposition for each data point i in high-dimensional data space, maximizing the local between-class distances; 2. processing with the second stage of LDA/QR for each data point, minimizing the within-class distance; 3. affine transforming the local coordinates, the global embedding coordinate τ_i of data point i is given by the eigenvectors corresponding to 2~d+1 minimal eigenvalues of matrix B; 4. with the global low embedding coordinates, calculating the $p_y(x_i)$ of samples in X_U with TCM-KNN algorithm, and classifying it with the class label $y_i = \arg\max_y p_y(x_i)$.

During the first stage of LDA/QR, QLAT utilize the information of labeled data set and unlabeled data set simultaneously, not only maximizing the betweem-class distances, but also utilizing the geometry information of data distribution. Compared with original LDA/QR algorithm, it utilizes the geometry information of sample data more effectively, and compared with original LTSA algorithm, it utilizes the class label information of sample data more effectively.

4 Experimental Setup and Results

In order to evaluate QLAT method, we have conducted several experiments on synthetic data sets and real datasets. Experiment results on Swiss-roll 3D data set, 2D synthetic data set, MNIST data set, ORL data set and Yale B data set are presented in this section.

4.1 Synthetic Data

Swiss-roll data set [3] was sampled evenly from noiseless 3D Swiss-roll surface, the data set does not have class label information, we use the data set to test the low-dimensional embedding capability of QLAT. The generating function is as follows:

$$t = (3 * pi / 2) * (1 + 2 * rand(1, N));$$
$$s = 21 * rand(1, N);$$
$$X = [t. * \cos(t); s; t. * \sin(t)];$$

LDA/QR+LTSA is used, TCM-KNN is not used, and only the first stage of LDA/QR is used, i.e., LDA is not used. Experiment results with different k_l values are shown in Fig.1, N=4000, d=2 in the experiments. Similar results of LTSA with N=2000, d=2 and different k values are presented in paper [5]. From the result, it can be seen that QLAT algorithm can effectively discover the low-dimensional embedding structure of high-dimensional data.

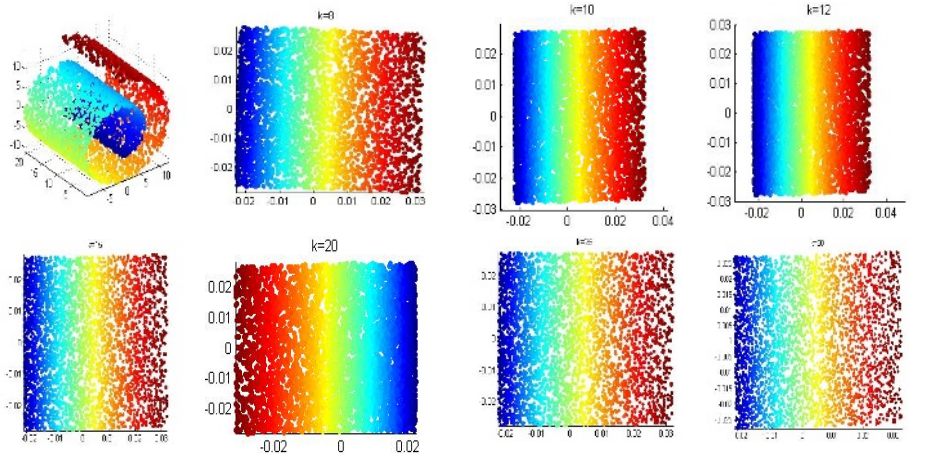
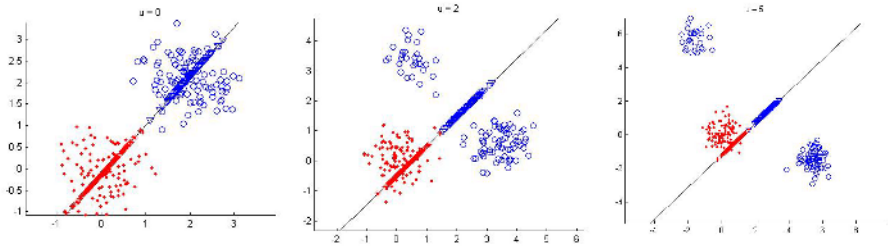


Fig. 1. 3D swiss roll data. Generating coordinates and computed coordinates by QLAT with different k_l values, k_l=8,10,12,15,20,25,30 respectively.

The 2D synthetic data set [9] contains 200 data points from two classes (each has 100 points) in the 2D space. Data in the first class is generated from a Gaussian whose mean is [0,0], and data in the second class is generated from a mixture of two Gaussians: The first one has 30 points with the mean $[2,2]-[\sqrt{2}/2, -\sqrt{2}/2]\mu$, and the second one has 70 points with the mean $[2,2]+[\sqrt{2}/2, -\sqrt{2}/2]\mu$ (for some μ). All these Gaussians have covariance $0.5I_2$. The low-dimensional embedding results with different μ are presented in Fig.2, TCM-KNN is not used, and $k_l=k'_l=5$. It can be seen that QLAT can reduce the dimension of the data, meanwhile, keeping the separability of data of different classes.



(a) (b) (c)

Fig. 2. Visualization of 2D synthetic data with different μ and their projections via QLAT, μ=0, 2, 5 corresponding to (a),(b) and (c)

4.2 MNIST Data Set, ORL Data Set and Yale B Data Set

The MNIST database of handwritten digits has a training set of 60,00 examples, and a test set of 10,000 examples. The sample numbers of each class in training set varies from 5842 to 6742, and the sample number of each class in test set is 1,000. It is a subset of a large set available from NIST. The digits have been size-normalized and centered in a fixed-size image. The ORL face data set contains 400 face images of 40 individuals. The image size is 92×112. The face images are perfectly centralized. The major challenge on this data set is the variation of the face post. There is no lighting variation with minimal facial expression variations and no occlusion. We use the whole image as an instance, that is, the dimension of an instance is 92×112=10,304. The Yale B data set contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions). The difference of viewing conditions dramatically increases the within-class variations of the data set. In this study, we use a subset of Yale B data set, which contains 1,280 face images, that is, each person with 4 poses and 32 illumination conditions. Its image size is 640 × 480. We crop the image from the row 80 to 480 and the column 150 to 450, and then subsample the cropped images with sample step 4×4. The dimension of each instance is 101×76=7,676.

In these experiments, we explore the performance of QLAT and compare it with other methods, including PCA-KNN, PCA+LDA-KNN, TCM-KNN, LDA/QR-TCM. Some parameters in the experiments are set as follows: the component number in PCA and the PCA stage in PCA+LDA are set as p=100, the output dimension for LDA is k-1, where k is the number of class labels here. Furthermore, in QLAT method, the k_l and λ need to be decided, the performance of QLAT with different k_l and λ on MNIST are presented in Tab.2, k is set to be 1 during the TCM-KNN step. To explore the utilizing of unlabeled samples in QLAT, label information of half labeled training samples are discarded, i.e., half of the training samples have class labels and the other half do not have. It can be seen that the classification performance is best when k_l is between 8 and 12, and λ is about 1.0. If the k_l is too small, the estimation error of between-class scatter matrix will become large when analyzing the local structure. While if the k_l is too big, the influence of remote data points is improperly enlarged, which can not represent the locality of the analysis,

besides, which increases the amount of computation. As to λ, which balance PCA and LDA during the local analysis. If λ is taken too small, QLAT is similar to LDA/QR and can not utilize the information of unlabeled samples. While if λ is taken too large, which can not take advantage of labeled samples. Of course, the most optimal values of k_l and λ are related with the specific data set to be classified. In the next experiment, we take k_l=10 and λ=1.0.

Table 2. Classification error of QLAT on MNIST with different k_l and λ (K=1)

k \ λ	0	0.2	0.6	1.0	3.0	5.0	10.0
5	3.28	3.24	2.89	2.74	3.15	3.84	4.78
8	3.18	3.41	2.38	2.24	2.93	4.12	4.32
10	3.82	3.17	2.53	2.32	3.18	5.02	4.22
12	3.46	3.25	3.51	2.23	2.84	5.41	6.74
20	9.25	7.54	4.83	2.34	3.52	8.64	10.32
30	10.36	8.44	5.81	4.32	6.23	10.21	12.4

Figure 3 shows the classification error results of different methods on MNIST, ORL and Yale B data sets. For the MNIST and Yale B data sets, class label information of 1/3 training sample is discarded, while for the ORL data set, all the class label information of training data set is used. The most interesting result lies in the classification accuracy results on Yale B data set. We observe that PCA+LDA-KNN, PCA+LDA-TCM, LDA/QR-TCM and QLAT distinctly outperform the PCA-KNN method. Recall that the images in the Yale B data set contains large variations of poses and illumination conditions, whose direct consequence is the large within-class variation of each individual. The effort of minimizing the within-class variation achieves distinct success in this situation. While PCA does not have the effort in minimizing the within-class variation, which predicts its poor performance in this situation.

Besides the major observation mentioned above, it can also be seen that TCM-KNN outperforms traditional KNN. In all the methods above, QLAT can achieve the best

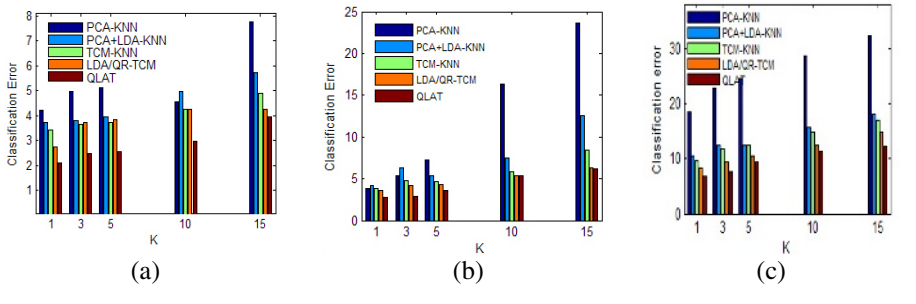


Fig. 3. Performance comparison of different methods on 3 data sets, (a) for MNIST, (b) for ORL and (c) for Yale B

performances on all the three data sets, especially on the Yale B data set. As to the ORL data set, the performance improvement of QLAT compared to LDA/QR-TCM is not significant. On ORL data set, the performances of most methods can achieve above 90%. This is mainly due to the relatively small within-class variations in these data. Recall that ORL face images contains small pose variations and have no obstruction. Finally, it can be seen that KNN with k=1 usually performs the best by all algorithms on all three image data sets.

5 Conclusions and Future Work

We have described QLAT, a classification algorithm based on LTSA and TCM-KNN, which extends the usage field of LTSA algorithm from dimension reduction to classification problem. Compared with LTSA, it not only utilize the geometry information of unlabeled data set, also utilize the class label information of labeled data, and utilizes 2-stage LDA/QR instead of traditional LDA during constructing the local embedding coordinates. Compared with traditional KNN, QLAT uses TCM-KNN algorithm for classification on low-dimensional manifold and can effectively utilize the distribution information of testing samples. Experiment results show that QLAT is an effective manifold classification method.

In future, we plan to investigate improvement in QLAT algorithm. Such as the parameter values of k_l and λ during constructing local embedding coordinates need to be decided in QLAT, how to obtain the optimal values for a specific data set need further investigation. Furthermore, the integration of LTSA and other transductive inference algorithms also needs investigation.

Acknowledgement

The work has been supported by the National High-Tech. R&D Program for CIMS, China (No.2003AA411021) and the highlight R&D Program of Zhejiang Province (No.2004C11053, No. 2005C21078).

References

1. Seung, S., Daniel, D.L., The manifold ways of perception. Science, 2000, 290(5500): 2268-2269.
2. Tenenbaum, J.B., de Silva, V., and Langford, J.C.,, A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, 2000, 290(5500): 2319-2323.
3. Roweis, S.T. and Saul, L.K., Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500): 2323-2326.
4. David, L.D., Caroe, G., Hessian eigenmaps Locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Sciences of the United States of America, 2003. 100(10): 5591-5596.
5. Zhang, Z.Y., Zha, H.Y., Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM Journal of Scientific Computing, 2004. 26(1): 313-338.

6. Yan, S.C., Zhang, H.J., Hu, Y.X., etc, Discriminant Analysis on Embedded Manifold. Proceeding of the 8th Europe Conference on Computer Vision-ECCV 2004, LNCS, 3021:121-132.
7. Proedrou, K., Nouretdinov, I., Vovk, V,, etc, Transductive Confidence Machines for Pattern Recognition. Proceedings of the 13th European Conference on Machine Learning, 2002, LNAI, 2430:381-390
8. Li, F.Y., Wechsler, H., Open Set Face Recognition Using Transduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005. 27(11): 1686-1697.
9. Ye, J.P., Li, Q., A two-stage linear discriminant analysis via QR-decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 929-941.
10. Min, W.L., Lu K., HE X.F., Locality pursuit embedding, Pattern Recognition, 2004. 37(4): 781-788.
11. Wu, Y.M., Chan, K.L., Wang, L., Face recognition based on discriminative manifold learning. in Proceedings of the 17th International Conference on Pattern Recognition. 2004, 4:171-174.