Handwritten Chinese Character Recognition Using Local Discriminant Projection with Prior Information

Honggang Zhang[1,2], Jie Yang[1], Weihong Deng[2], Jun Guo[2]

[1] *Carnegie Mellon University*
[2] *Beijing University of Posts and Telecommunications*
E-mail:{*zhhg, jie.yang*}@*cs.cmu.edu*, {*whdeng, guojun*}@*bupt.edu.cn*

Abstract

In this paper, we propose a new method to model the manifold of handwritten Chinese characters using the local discriminant projection. We utilize a cascade framework that combines global similarity with local discriminative cues to recognize Chinese characters. We find the similarity of different characters using a nearest-neighbor (NN) classifier, and followed by the Local Discriminant Projection with Prior Information (LDPPI) to map similar characters within a cluster to a low-dimensional space. We evaluate the proposed method on two large public datasets, ETL9B which contains 607,200 handwritten characters from 200 people, and HCL2000 which contains 3,755,000 characters written by 1,000 people. The experimental results demonstrate that the proposed method achieves 0.74% error rate on ETL9B database and 1.88% on HCL2000 database.

1. Introduction

One of the important issues in handwritten character recognition is to improve the recognition performance for similar characters, in particular, in the recognition of Chinese characters which include a large number of similar characters. Feature dimensionality reduction plays a very crucial role in similar characters recognition. Many dimensionality reduction methods have been proposed to the discrimination of similar characters for improving the overall recognition, such as linear methods based on Linear Discriminant Analysis (LDA) [2] [6], and nonlinear methods such as kernel SVM and kernel LDA [8]. LDA implicitly assumes all pattern classes have an equal covariance, when the practical data is heteroscedastic, LDA cannot perform in an optimal way. It is difficult to use kernel based methods in handwritten Chinese character recognition because of the fact that thousands of training samples are needed to construct the classifier, and difficulties in choosing the kernel function.

Recently, manifold learning plays an important role in many pattern applications. Several nonlinear techniques, such as Isomap, LLE, have been introduced to learn the intrinsic manifold embedded in the ambient space of high dimensionality. However, these nonlinear techniques yield maps that are defined only on the training data points and it remains unclear how to evaluate the maps on novel testing points. To address this problem, many linear methods are proposed [1] [3] [7] [9]. Locality preserving projection (LPP) is proposed in [3]. Specifically, LPP finds an embedding that preserves local information and may be simply extended to any new sample. LPP first projects the data to a subspace by Principal Component Analysis(PCA), which may lose some discriminative information for classification. Marginal Fisher Analysis(MFA) is developed using the graph embedding framework based on a Marginal Fisher Criterion [7]. MFA tends to compress the points of an individual together even though they are really far away in original space. However, this makes against extracting efficient discriminative information.

In this paper, we consider the similar Chinese characters reside on a nonlinear manifold structure and propose a cascade framework that combines global similarity with local discriminative cues to classify handwritten Chinese characters. First, we find the similarity of different characters using a nearest-neighbor(NN) classifier, and then map similar characters within a cluster to a low-dimensional space by the Local Discriminant Projection with Prior Information(LDPPI). The NN classifier can easily handle a large scale problem of character recognition, and the LDPPI provides a robust method to find the local features which discriminate the similar characters. The experimental results demonstrate that the proposed method achieves 0.74%

error rate on ETL9B database and 1.88% on HCL2000 database.

The rest of this paper is organized as follows. Section 2 models global similarity; Section 3 details our LDPPI algorithm; The experimental results on large set databases are then followed in Section 4. Finally we give concluding remarks and future work.

2. Modeling Global Similarity by NN

Our goal is to use manifold learning methods to solve the problem of handwritten Chinese character recognition. We propose to use a cascade of NN and LDPPI approach to handle this problem. The proposed approach includes two phases, training and recognition. In the training phase, we first cluster similar characters by NN, and train the LDPPI transform matrix. In the recognition phase, each input image is first classified into one of similar characters group by NN, and then analyzed by the LDPPI models.

The purpose of modeling global similarity is to cluster characters with the similar global structure, e.g., , , , into a group. Unlike most existing methods of handwritten character recognition which dynamically use top n candidates as the similar characters in the recognition process, we model the similar characters statically in the training process by an NN classifier. First, we establish the center of each class in the feature space using the clustering algorithm, and then project every training sample into its class by the NN classifier. In order to recognize a large set of characters in fast speed, we extract 196-dimensional Directional Element Feature(DEF) and adopt the City Block Distance with Deviation (CBDD) to measure the distance between the prototype and the input sample instead of Euclidean distance in both training and testing process [4].

After using NN classifier to model the global similarity on training data, we get each character's similar character group and the priori error rate of each pairwise similar characters. If given input character is c^j and the NN classifier output is c^i, the definition of error rate is:

$$P(c^i|c^j) = \frac{\sum_{k=1}^{C_j}(NN(c_k^j) = c^i)}{\sum_{l=1}^{C}\sum_{k=1}^{C_l}(NN(c_k^l) = c^i)}, \quad (1)$$

where $NN(c_k^j)$ is the classification result when the input character is c_k^j, c_k^j is the kth character sample of c^j, C_j is the number of training samples of c^j, C is the number of classes. We sort the error rates in a descent order and set the maximum number of similar characters in a group is $M = 50$. Then we build directed graph models for each Chinese character, as shown in Fig.1.
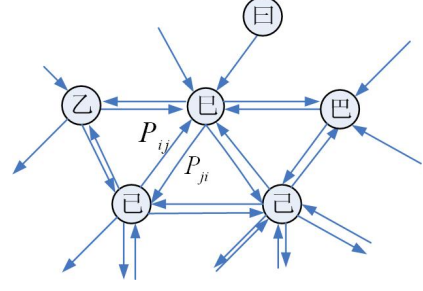

Figure 1. The similar character graph of character . The weight is the priori error rate.
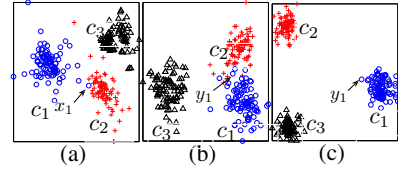

Figure 2. An illustration of LDPPI and LPP, (a) the original space, (b) the projection by LPP, (c) the projection by LDPPI

The vertex is the character in a group, and the directed edge indicates the error rate from a father node to a son node. The directed graph will be further considered in the following local discrimination.

3. Modeling Discriminative Cues Using LDPPI

For each similar character group, the character set is represented as a matrix $X = [x_1, x_2, \cdots, x_N], x_i \in \mathbb{R}^m$, where N is the number of samples and m is the feature dimension. The class label of the sample x_i is assumed to be $c_i \in \{1, 2, \cdots, M\}$, where M is the number of similar characters. We want to find projection matrix W which maps an m-dimensional data space to a d-dimensional subspace ($d \ll m$): $W^T : \mathbb{R}^m \to \mathbb{R}^d$.

In contrast with LPP, LDPPI aims to find the discriminating projection for classification, which not only preserves the locality but also discriminates samples by separating the global class centers apart with prior information. Fig.2 illustrates an example of three classes, $c_1(+)$, $c_2(\circ)$ and $c_3(\triangle)$. x_1 is a sample of c_1 but overlapped and near c_2 as shown in Fig.2(a). After supervised LPP projection, locality is preserved in each class, but y_1 (projection of x_1) is still overlapped

and has higher probability of misclassification because y_1 is still near c_2 shown in Fig.2(b). After projection by LDPPI, as shown in Fig.2(c), the centers of c_1 and c_2 are separated far away and the samples in c_1 and c_2 are compact, y_1 will be classified correctly.

Two graphs are introduced to construct class-specific manifold,the k-nearest native neighbor graph and the similar characters graph. The native neighbor graph is defined as a weighted graph $G^n = \{V, S^n\}$, with $V = \{x_i\}_{i=1}^N$,

$$S_{ij}^n = \begin{cases} e^{\frac{-\|x_i - x_j\|^2}{t}}, & x_i \in \mathcal{N}_k(x_j), x_j \in \mathcal{N}_k(x_i) \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where S^n is a sparse symmetric $N \times N$ matrix with S_{ij}^n weighting the edge connecting vertices x_i and x_j , $\mathcal{N}_k(x_i)$, called k-nearest native neighbors of x_i, is the set of k nearest neighbors having the same label with x_i.

The other graph is the similar character graph as showed in Fig.1, which is denoted by $G^s = \{V, S^s\}$, with $V = \{m_i\}_{i=1}^M$,

$$S_{ij}^s = \begin{cases} 1 - P_{ij}, & i \neq j \\ 1, & i = j, \end{cases} \quad (3)$$

Note that the higher S_{ij}^s is, the more similar the two characters are, and the larger distance we will push them apart after projection. Therefore, in order to discriminate similar characters, we push the class centers apart according to the prior information on the training samples.

Thus the objective functions of LDPPI are as follows:

$$\min \sum_{i=1}^N \sum_{j=1}^N (y_i - y_j)^2 S_{ij}^n, \quad (4)$$

and

$$\max \sum_{i=1}^M \sum_{j=1}^M (m_i - m_j)^2 S_{ij}^s. \quad (5)$$

Suppose W is a transformation vector, that is, $Y = W^T X$, The objective function can be formulated as:

$$\begin{aligned} w^* &= \arg\min_w \frac{\sum_{ij}^N (w^T x_i - w^T x_j)^2 S_{ij}^n}{\sum_{ij}^M (w^T f_i - w^T f_j)^2 S_{ij}^s} \\ &= \arg\min_w \frac{w^T X L^n X^T w}{w^T F L^s F^T w}, \end{aligned} \quad (6)$$

where $L^n = D^n - S^n$ is the Laplacian of the neighbor graph G^n, D^n is a diagonal matrix with $D_{ii}^n = \sum_j S_{ij}^n$, and $L^s = D^s - S^s$ is the Laplacian of the similar characters graph G^s, D^s is a diagonal matrix with

$D_{ii}^s = \sum_j S_{ij}^s$, $f_i = \frac{1}{C^i} \sum_{j=1}^{C^i} x_j$. Minimizing the numerator of objective function is to ensure that if two samples x_i^c and x_j^c are close then y_i^c and y_j^c are close as well. While maximizing the denominator is an attempt to ensure that if f_i and f_j are easily misclassified then m_i and m_j are far, and it is easy to discriminate the similar characters.

The vectors w^* that minimize the objective function are given by minimum eigenvalue solutions to the generalized eigenvalue problem. From the above description, the Local Discriminant Projection with Prior Information algorithm can be described as follows.

1. Feature extraction. For local discriminant, we extract 512-dimensional gradient features [5] and project the data set $\{x_i\}_{i=1}^N$ onto the low-dimensional subspace. Let $\Phi_{GF}(m, k)$ denote the feature extraction.

2. Constructing the k-nearest native neighbor graph and similar characters graph according to equations (2) and (3).

3. Solving locality discriminant criterion in equation (6). Computing the eigenvector and eigenvalue of the following generalized eigenvalue problem:

$$X L^n X^T w = \lambda F L^s F^T w. \quad (7)$$

4. Outputting the final linear transformation matrix. Let $w_0, w_1, \cdots, w_{k-1}$ be the solutions of (7), eigenvalues sorted in an ascending order $0 \leq \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$, and $d(d \leq k)$be the output dimension. We have the transformation matrix:

$$W_{LDPPI} = [w_0, w_1, \cdots, w_{d-1}]. \quad (8)$$

4. Experimental Results

We evaluate the proposed method using two large databases of handwritten Chinese characters, ETL9B and HCL2000. ETL9B was collected by the Electro Technical Laboratory of Japan, and contains binary images of 3036 characters, and each of them has 200 binary images. Funded by Chinese 863 High-tech project, we built up HCL2000, which currently contains 3,755,000 simplified Chinese characters written by 1,000 different people. For ETL9B, we use the first 20 and last 20 images of each character for testing, and the rest 160 images for training. For HCL2000, we use 700 sets labeled from xx001 to xx700 for training, and the rest sets labeled from hh001 to hh300 for testing.

The recognition process has two steps. First, an input image is preprocessed and classified by the NN classifier to be assigned into a character group [4] [5]. Then we use the gradient feature for further discrimination. Gradient feature vectors are subsequently transformed to the LDA, LPP, MFA and LDPPI representations with

class label information respectively. For fair comparisons, in both LPP and LDPPI experiments, we set $k = 20$ as the parameter for constructing the graph and use the simple-mind weight scheme for all the following experiments. For MFA, its parameter $k_1 = 3$ and $k_2 = 20$. The maximum number of similar characters in a group is 50. The results on ETL9B and HcL2000 are shown in Fig.3 and Fig.4 respectively.
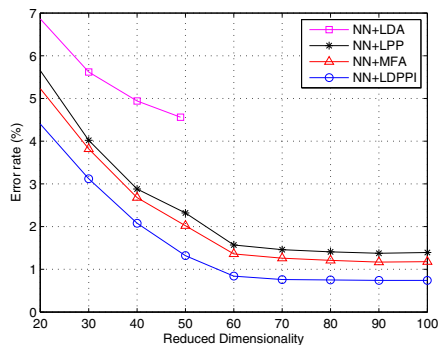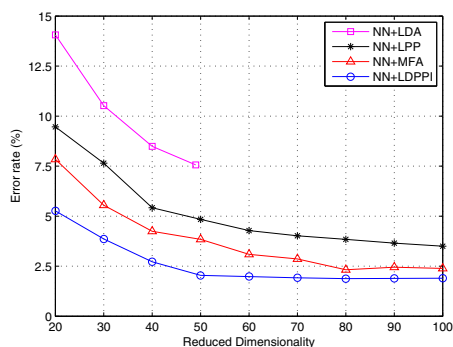

Figure 3. Experimental results on ETL9B

Figure 4. Experimental results on HCL2000

Based on the experimental results, LDPPI outperforms other methods, and MFA, LPP are better than LDA. LDA yields some meaningful projections since handwritten characters of the same class are mapped close to each other. However, LDA discovers only the Euclidean structure, and can not reveal the underlying nonlinear manifolds that handwritten characters lie on. Therefore its discriminating power is limited. LPP and MFA are supervised methods and preserve local neighborhood information. Their discriminating powers are better than LDA, since similar handwritten characters contain variations and significant overlapping among different classes. LDPPI provides the best projection since it discriminates the overlapped classes with the prior information, especially for the most easily misclassified characters.

5. Conclusions

The main contribution of this paper is to have successfully demonstrated the application of state-of-the-art manifold learning techniques in a classic pattern recognition problem. We have employed LDPPI to map similar characters within a cluster to a low-dimensional space which has more discriminative power. The experimental results indicate that the proposed method achieves 0.74% error rate on ETL9B database and 1.88% on HCL2000 database. LDPPI is linear, but the performance of LDPPI in the kernel space needs to be further examined.

6 Acknowledgments

References

[1] W. deng, J. Hu, J. Guo, and H.-G. Zhang. Comments on 'globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics'. *IEEE trans. PAMI*, 30(8), 2008.

[2] T.-F. Gao and C.-L. Liu. Lda-based compound distance for handwritten chinese character recognition. *Proceeding of ICDAR'07*, 2007.

[3] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face recognition using laplacian faces. *IEEE trans. PAMI*, 27(3):328–340, 2005.

[4] N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto. A handwritten character recognition system using directional element feature and asymmetric mahalanobis distance. *IEEE trans. PAMI*, 21(3):258–262, 1999.

[5] C.-L. Liu. Normalization-cooperated gradient feature extraction for handwritten character recognition. *IEEE trans. PAMI*, 29(8):1465–1469, 2007.

[6] H.-L. Liu and X.-Q. Ding. Improve handwritten character recognition performance by heteroscedastic linear discriminant analysis. *Proceeding of ICPR'06*, 2006.

[7] S.-C. Yan, D. Xu, and B. Zhang. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. PAMI*, 29(1):40–51, 2007.

[8] D. Yang and L.-W. Jin. Handwritten chinese character recognition using modified LDA and kernel FDA. *Proceeding of ICDAR'07*, 2007.

[9] W.-W. Yu, X. Teng, and C.-Q. Liu. Face recognition using discriminant locality preserving projections. *Image and Vision Computing*, 24:239–248, 2006.