Handwritten Chinese Character Recognition Using Modified LDA and Kernel FDA

Duanduan Yang and Lianwen Jin
School of Electronics and Information, South China University of technology, Guangzhou, 510640, P.R.China
ddyang@scut.edu.cn , eelwjin@scut.edu.cn

Abstract

The effectiveness of kernel fisher discrimination analysis (KFDA) has been demonstrated by many pattern recognition applications. However, due to the large size of Gram matrix to be trained, how to use KFDA to solve large vocabulary pattern recognition task such as Chinese Characters recognition is still a challenging problem. In this paper, a two-stage KFDA approach is presented for handwritten Chinese character recognition. In the first stage, a new modified linear discriminant analysis method is developed to get the recognition candidates. In the second stage, KFDA is used to determine the final recognition result. Experiments on 1034 categories of Chinese character from 120 sets of handwriting samples shows that a 3.37% improvement of recognition rate is obtained, which suggests the effectiveness of the proposed method.

1. Introduction

Over the last few years, kernel-based learning machines, e.g., support vector machines (SVMs) [1], kernel principal component analysis (KPCA)[2], and kernel fisher discriminant analysis (KFDA)[3], have aroused considerable interest in the fields of pattern recognition and machine learning. As nonlinear version of the well-known linear discriminant analysis (LDA), KFDA was firstly proposed by Mika [4] and Baudat [5] to overcome the limitation of the KFDA, which maps the input vectors into a high dimensional feature space and implements the KFDA in the feature space, thereby yielding a set of nonlinear discriminant vectors in the input space. Mika's work is mainly focused on two-class problems, while Baudat's algorithm is applicable for multi-class problems. Yang et al. [6] further studied kernel Fisher discriminant analysis and pointed out its essence. The KFDA turns out to be effective in many real-world applications due to its power of extracting the most discriminatory nonlinear features. However, there is hardly any Chinese Characters recognition method based on the KFDA because of the dilemma, which resides in the fact that thousands of training samples are necessary to construct a Chinese character classifier with good performance, but KFDA is hardly run with such large sample size. To solve this problem, this paper proposes a new KFDA -based method, which consists of two steps. The first step is to find the candidates for the unknown sample by the proposed *modified linear discriminant analysis(MLDA)* and a minimum distance classifier (MDC). The second step is to recognize the unknown sample in the candidates by KFDA, where KFDA is trained by a small part of training samples. As only a small part of training samples is used to train KFDA, the memory space and the computation efforts are largely reduced.

The remainder of the paper is organized as follows: In section 2, an improved LDA algorithm, MLDA, is developed. The idea of KFDA is introduced in section 3 and the two-stage KFDA -based Chinese character recognition method is presented in section 4. In section 5, the experiments are performed on 1034 categories of Chinese character from 120 sets of samples. Finally, the conclusion and the discussion are given in section 6.

2. Modified linear discriminant analysis

It has been demonstrated that Modified Quadratic Discriminant Function (MQDF) [7] can improve the classification performance of Quadratic Discriminant Function (QDF) via eigenvalue smoothing. Inspired from MQDF, we proposes a modified linear discriminat analysis (Modified LDA, MLDA) to improve the performance of LDA. The basic idea of MLDA is that the eigenvalues of minor axes of the covariance matrix of each class are set to a constant. The motivation behind this is to smooth the parameters for compensating for the estimation error on finite sample size.

2.1. Introduction of LDA

LDA searches for those vectors in the underlying space that best discriminant among classes. More formally, given a number of independent features relative to which the data is described, LDA creates a linear combination of the features that yield the largest mean differences between the desired classes. Mathematically speaking, for all the samples of all classed we define two measures, one is called within-class scatter matrix, as given by

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \qquad (1)$$

and another is called between-class scatter matrix

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \qquad (2)$$

where x_i^j is the ith sample of class j , μ_j is the mean of class j , c is the number of classes ,and N_j is the number of samples in class j ; μ represents the mean of all classes.

The goal of LDA is to maximize the between-class measure while minimizing the within-class measure. One way to do this is to maximize the ratio $\det|S_b| / \det|S_w|$. If S_w is a non-singular matrix, then the ration is maximized when the column vectors of the projection matrix, W_{lda} , are eigenvetors of $S_w^{-1} S_b$.

2.2. MLDA algorithm

MLDA tries to improve LDA performance by compensating for the estimation error of S_w. If the covariance matrix the class j is denotes as

$$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \qquad (3)$$

S_w can be denote using Σ_j:

$$S_w = \sum_{j=1}^{c} N_j \Sigma_j \qquad (4)$$

According to eigen-decomposition, Σ_j can be rewritten as:

$$\Sigma_j = \mathrm{B}_j \Lambda_j \mathrm{B}_j^T \qquad (5)$$

where $\Lambda_j = diag(\lambda_{j1}, ..., \lambda_{jd})$ with λ_{jk} , $k = 1,2,...d$, being the eigenvalues (in decreasing order) of Σ_j , $\mathrm{B}_j = [\beta_{j1}, ..., \beta_{jd}]$ and with β_{jk} , $k = 1,2,...d$, being the ordered eigenvectors. B_j is ortho-normal (unitary) such that $\mathrm{B}_j \mathrm{B}_j^T = I$.

In order to compensate for the estimation error of Σ_j on finite sample size, we use $\tilde{\Lambda}_j$ replace Λ_j to reconstruct Σ_j , $\tilde{\Sigma}_j = \mathrm{B}_j \tilde{\Lambda}_j \mathrm{B}_j^T$, where $\tilde{\Lambda}_j = diag(\lambda_{j1}, .. \lambda_{jm}, \delta_j, .. \delta_j)$, where $\delta_j = \frac{1}{d-m} \sum_{k=m+1}^{d} \lambda_{jk}$ So the reconstructed within-class scatter matrix becomes:

$$\tilde{S}_w = \sum_{j=1}^{c} N_j \tilde{\Sigma}_j \qquad (6)$$

Usually, in Chinese character recognition applications, $\tilde{S}_w$ is a non-singular matrix, then this ratio $\det \left|\tilde{S}_b\right| / \det \left|\tilde{S}_w\right|$ is maximized when the column vectors of the projection matrix, W_{mlda} , are eigenvetors of $\tilde{S}_w^{-1} S_b$.

3. Outline of KFDA

In this section, let's introduce KFDA concisely starting with two-class problems. For a given nonlinear mapping, Φ , the input data space IR^n can be mapped into the feature space H :

$$\Phi : IR^n \sim \mathrm{H}; \quad X \to \Phi(X) \qquad (7)$$

As a result, a pattern in the original input space IR^n is mapped into a potentially much higher dimensional feature vector in the feature space H . The idea of KFDA is to solve the problem of LDA in the feature space H . However, it is difficult to do so directly because it is computationally very intensive to compute the dot products in a high-dimensional feature space. Fortunately, kernel techniques can be introduced to avoid this difficulty.

Let $X_1 = \{x_1^1, ..., x_{N_1}^1\}$ and $X_2 = \{x_1^2, ..., x_{N_2}^2\}$ be samples from different classes and $X = X_1 \cup X_2 = \{x_1, ..., x_N\}$. To find the discriminant vectors in the feature space H , we need to maximize:

$$J^\Phi(w) = \frac{w^T S_b^\Phi w}{w^T S_t^\Phi w} \qquad (8)$$

where w is a discriminant vector, S_b^Φ and S_t^Φ are the between-class and within-class scatter matrices defined in feature space H.

$$S_t^\Phi = \sum_{j=1}^{2}\sum_{i=1}^{N_j}(\Phi(x_i^j) - m_j^\Phi)(\Phi(x_i^j) - m_j^\Phi)^T \qquad (9)$$
$$S_b^\Phi = (m_1^\Phi - m_2^\Phi)(m_1^\Phi - m_2^\Phi)^T$$

where x_i^j is the ith sample of class j, $m_j^\Phi = \frac{1}{N_j}\sum_{i=1}^{N_j}\Phi(x_i^j)$, $j \in \{1,2\}$.

It is easy to show that any solution $w \in H$ can be linear expanded by:

$$w = \sum_{i=1}^{N} a_i \Phi(x_i) = Q\alpha \qquad (10)$$

where $Q = (\Phi(x_1),...,\Phi(x_N))$ and $\alpha = (a_1,...,a_N)$.

To rewrite equation (8), let $Q_1 = \{\Phi(x_1^1),...,\Phi(x_{N_1}^1)\}$, $Q_2 = \{\Phi(x_1^2),...,\Phi(x_{N_2}^2)\}$. And then the Gram matrix is formed:

$$K_j = Q^T Q_j \qquad (11)$$

with the size of $N \times N_j$, where $j \in \{1,2\}$, whose elements can be determined by the virtue of kernel trick.

$$K_j(m,n) = \Phi(x_m)\Phi(x_n^j) = k(x_m, x_n^j) \qquad (12)$$

where k is a kernel function.

Substituting Eq.(10) into Eq.(8), the Fisher criterion is converted to

$$J(\alpha) = \frac{\alpha^T M \alpha}{\alpha^T N \alpha} \qquad (13)$$

where

$$M = (K_1 1_{vN_1} - K_2 1_{vN_2})(K_1 1_{vN_1} - K_2 1_{vN_2})^T$$
$$N = \sum_{j=1}^{2} K_j (I - 1_{N_j}) K_j^T \qquad (14)$$

where 1_{vN_j} is a $N_j \times 1$ column vector with elements equal to $1/N_j$, 1_{N_j} is a $N_j \times N_j$ matrix with elements equal to $1/N_j$.

Now, the problem of KFDA can be solved by finding the leading eigenvectors of $N^{-1}M$. The projection of a new pattern x onto w is given by

$$(w \bullet \Phi(x_i)) = \sum_{i=1}^{N} a_i k(x_i, x) \qquad (15)$$

More details of KFDA of two-class problem can be found in [4], and extending the KFDA in the multi-class problems can be found in [5].

4. The two-stage method for Chinese characters recognition

When applying KFDA for large vocabulary Chinese character recognition, how to reduce the memory space and computation effort is a crucial problem. For example, if 50 sets samples of 1034 categories are used to train KFDA, the size of K_j defined in equation (12) is 51700*50. The KFDA algorithm in this case is hardly run with such large size matrix. To solve the problem, a two-stage classification method is proposed (Figure 1). In our method, only a small part of samples is used to train the KFDA projection matrix for each class, which can reduce the size of K_j significantly.

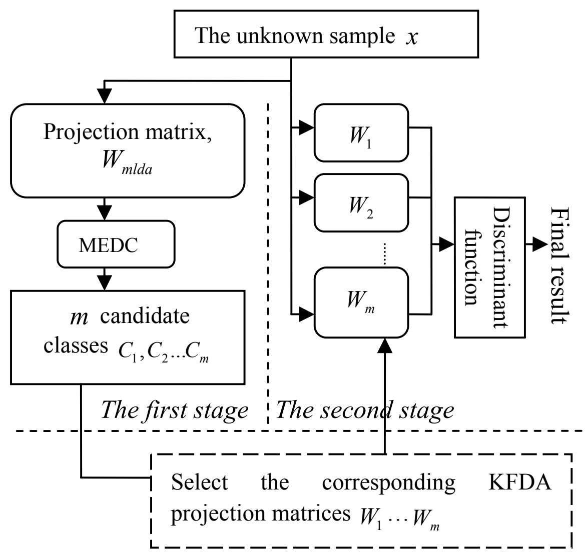

Fig.1. A two stages KFDA classification architecture

Concretely, in the first stage, when an unknown sample x is inputted, it is projected to a discriminant space by W_{mlda}, which is obtained by the proposed MLDA. Then a Minimum Euclid distance classifier (MEDC) is used to find the m candidate classes, $C_1, C_2...C_m$, of x in the discriminant space.

The second stage is to determine the final recognition result of x using KFDA in $C_1, C_2...C_m$. This can be regarded as a m two-class pattern recognition problem, which can be solved by designing m classifiers, $F^1, F^2...F^m$. F^j, that are used to distinguish whether x belongs to the jth class or not ($j = 1,...,m$). If x is mapped a high dimension by a non-linear function, Φ , F^j can be then written as following:

$$F^j(x) = \left\|W_j^T\Phi(x) - W_j Pm_j^\Phi\right\|^2 - \left\|W_j^T\Phi(x) - W_j^T Nm_j^\Phi\right\|^2 \quad (16)$$

where W_j is the projection matrix for the jth class whose columns are the discriminant vectors in feature space H. W_j can be trained by the positive and negative samples of the jth classes. Pm_j^Φ (Nm_j^Φ) is the mean of positive (negative) samples of the jth classes in H.

In order to determine the positive and negative samples for each class, let's suppose $C_1, C_2 ... C_m$ are the candidate classes of the input samples x . If $x \in C_i, i = 1,...,m$, x is treated as a positive sample for training F^i , otherwise, it is treated as a negative sample.

If the performance of F^j is good enough, there must exist $k \in \{1,2,...m\}$ satisfied $F^k(x) = \min_{j=1,...m}(F^j(x)) < 0$ and $F^j(x) > 0, \quad \forall j \neq k$. If F^j is not so good then we need to design a F^j-based rule to classify the unknown sample, which will be mentioned in the following text.

Once the training samples of W_j are determined, $W_j^T\Phi(x)$, $W_j^T Pm_j^\Phi$ and $W_j^T Nm_j^\Phi$ can be computed based on equation(15). In another word, $F^j(x)$ can be left explicit.

Next, we will show the process of recognition using KFDA. Suppose $C_1, C_2 ... C_m$ are the candidate character classes obtained by the MLDA+MDC, the corresponding distances are $d_1, d_2 ... d_m$ (ordered in increasing order) can be computed by:

$$d_i = \left\|W_{mlda}^T x - W_{mlda}^T T_i\right\|^2, \, i = 1,2,...n \quad (17)$$

where T_i is the template vector of the ith class in the original feature space and the projection matrix W_{mlda} is obtained by MLDA algorithm.

Let $F^1(x),...F^m(x)$ be the output of KFDA classifier, where $F_0 = F^k(x) = \min_{i \in \{1,...m\}}(F^i(x))$, $F_1 = \min_{i \in \{1,...m\}, i \neq k}(F^i(x))$, $k \in \{1,2,...m\}$. Based on d_i and $F^i(x), i = (1,2,...m)$, we can use the following rules to recognize the unknown sample x.

(i) If $F_0 < 0$ and $F_1 > 0$ and $abs(F_0 - F_1) > \tau$, the unknown sample belongs to C_k, where τ is a constant.

(ii) If (i) can not be satisfied, $d_1, d_2 ... d_m$ and $F^1(x),...F^m(x)$ are firstly normalized to [0,1]. Supposing $\tilde{d}_1,...\tilde{d}_m$ and $\tilde{F}^1(x),...\tilde{F}^m(x)$ are the normalization results. Then, they are combined using equation (18) to get the final recognition result.

$$f_u^i(x) = u_1 \tilde{d}_i + u_2 \tilde{F}^i(x) \quad (18)$$

where u_1 and u_2 are weighing coefficients, which are experimentally determined according to the corresponding classifying ability (in our experiments, u_1 is set to 0.6 and u_2 is set to 0.4). If $f_u^k(x) = \min_{i \in \{1,...m\}}(f_u^i(x))$, the recognition result of the unknown sample x is C_k .

5. Experiments and results

To evaluate the proposed method, 1034 categories of Chinese character from 120 sets of samples are used for our experiment. The samples from the first 100 sets, total 1034*100 samples, are used for training, the remainder samples are for testing. We use the method in [9] to extract the Gabor features of each sample. Polynomial kernel, $k(a,b) = (a \bullet b + 1)^r$, is used in our experiment (r is set to 2).

Three classifying methods, namely, Minimum Euclid distance classifier (MEDC) , LDA+MEDC， MLDA+MEDC， are compared in our experiments. The results are given in table 1.

Table1.The performances of different methods for Chinese character recognitions (1034 categories in GB2312-80)

Methods	MEDC	LDA+ MEDC	MLDA+ MEDC	Our method
Recognition rate	92.77%	94.26%	94.81%	96.14%

From table 1, we can see that the proposed MLDA improve the performance of LDA with 0.55% recognition accuracy by compensating for the estimation error of LDA. On the other hand, 3.37% and 1.88% improvements of recognition rate are obtained by the two-stage KFDA approach comparing with the MDC and MDC+LDA respectively.

6. Conclusion

The main work of this paper includes two parts. Firstly, we improve the LDA algorithm via eigenvalue smoothing to compensating for the estimation error of the parameters of LDA. Second, we propose a two-stage KFDA approach for handwritten Chinese character recognition. Experiments show the effectiveness of the proposed method.

7. Acknowledgments

This paper is sponsored by the following research foundations: New Century Excellent Talent Program of MOE of China (No.NCET-05-0736).

8. References

[1] V.Vapnik, The nature of stastical leaning Theory. New York: springer, 1995.

[2] B. Scholkopf, A. Smola, and K.R. Muller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” Neural Computation, vol. 10, no. 5, pp. 1299-1319, 1998.

[3]S. Mika, G. Ra¨tsch, B. Scho¨lkopf, A. Smola, J. Weston, and K.-R. Mu¨ ller, “Invariant Feature Extraction and Classification in Kernel Spaces,” Advances in Neural Information Processing Systems 12, Cambridge, Mass.: MIT Press, 1999.

[4] S. Mika, G. Ra¨tsch, J. Weston, B. Scho¨lkopf, and K.-R.Mu¨ ller, “Fisher Discriminant Analysis with Kernels,” Proc. IEEE Int’l Workshop Neural Networks for Signal Processing IX, pp. 41-48, Aug. 1999.

[5] G. Baudat and F. Anouar, Generalized discriminant analysis using a kernel approach, Neural Computation, Vol 12, No.10, pp. 2385–2404, 2000.

[6] J. Yang, Z. Jin, J. Yang, D. Zhang, The essence of kernel Fisher discriminant: KPCA plus LDA, Pattern Recognition, Vol 37, No.10, pp. 2097–2100, 2004.

[7] F.Kimura, K.Takashina, S.Tsuruoka, and Y.Miyake, “Modified quadratic discriminant functions and the application to Chinese character recognition ” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol9,pp.149-153,Jan1987.

[8] Aleix M. Martinez, Avinash C. Kak, "PCA versus LDA," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, Feb., 2001.

[9] Q. Huo, Y. Ge, and Z.-D. Feng, High Performance Chinese OCR Based on Gabor Features, Discriminative Feature Extraction and Model Training, Proc. ICASSP-2001, Salt Lake City, USA, May 7-11, 2001, pp.III-1517-1520.