A Kernel-Based Nonlinear Representor with Application to Eigenface Classification

ZHANG Jing\(^{1,3}\)\/, LIU Ben-yong\(^{2}\)/, TAN Hao\(^{3}\)

(1. The Information Center of Sichuan Radio and Television University Chengdu \n610073 China; 2. School of Electronic Engineering, UESTC Chengdu \n610054 China; 3. School of Computer Science and Technology, UESTC Chengdu \n610054 China)

Abstract This paper presents a classifier named kernel-based nonlinear representor (KNR) for optimal representation of pattern features. Adopting the Gaussian kernel, with the kernel width adaptively estimated by a simple technique, it is applied to eigenface classification. Experimental results on the ORL face database show that it improves performance by around 6 points, in classification rate, over the Euclidean distance classifier.

Key words kernel based nonlinear representor; face recognition; eigenfaces; Gaussian kernel; euclidean distance classifier

Classifiers play a paramount role in pattern recognition. In some of the popular methods, such as the Parzen classifier and a nonlinear support vector classifier\(^{1,2}\), the solutions are expressed as a nonlinear function \( f \) expanded by a kernel function, \( k \), of the space to which \( f \) belongs, \( f(x) = \sum_{j=1}^{M} a_j k(x, x_j) + b \), where \( x_j \) is the \( j \)-th sample point with \( j = 1, 2, \ldots, M \), \( b \) is a constant which can be set to zero in some applications. The set of coefficients, \( \{a_j\}_{j=1}^{M} \), is decided by the nature of the related problem. For example, when the error cost function for approximation is quadratic, it can be obtained by solving a linear system. When Vapnik's \( \varepsilon \)-insensitive cost function is adopted, it is obtained by the support vector machine (SVM) approximation scheme, in which a quadratic programming problem needs to be solved\(^{2}\). In our discussion, \( b \) is set to zero for simplicity and the set of coefficients is determined, in a closed form, by a measure of the classifier’s capability in optimal feature discrimination or representation. Previously, one measure of the former kind was proposed and it leaded to a classifier called a kernel-based nonlinear discriminator (KND)\(^{3}\). This paper presents another one of the latter kind.

As for dimension reduction and feature extraction, principal component analysis (PCA) and linear discriminant analysis (LDA) are two popular methods in some pattern recognition applications, such as face recognition and radar target recognition\(^{4,4}\). Since LDA aims at extracting discriminative features while PCA at representative ones, it is widely believed that LDA outperforms PCA, and thus much attention has been paid to LDA or its variations. However, some practical applications to face recognition shows that this is not always true\(^{5}\).

This paper attempts not to evaluate the performance of PCA or LDA in feature extraction, but to elucidate, by introducing a new classifier called kernel-based nonlinear representor (KNR) to eigenface classification, that how great a classifier can improve the performance of a pattern recognition system. In our study, eigenfaces are selected because we believe that it is simpler to perform PCA than LDA, especially when the small-sample-size problem and the incremental realization are taken into consideration. Adopting the Gaussian kernel, with the kernel width adaptively estimated by a simple technique, illustrative experiments are taken on the Olivetti research laboratory (ORL) face database.

1 Optimal Feature Discrimination and KND

We restrict our discussion to designing a
nonlinear classification function so that it has a certain desirable capability. It is assumed that such a function is defined on \( C^N \), a complex \( N \)-dimensional vector space, and it is denoted by \( f_0(x) \). We assume that it is an element of a reproducing kernel Hilbert space (RKHS) \( H \) which has a kernel function \( k(x, y) \) of the domain of \( C^N \). Generally \( f_0 \) is unknown but a data set \( \{(x_i, y_i)\}_{i=1}^M \) are given by sampling \( f_0 \)

\[
y = Af_0
\]  

(1)

where \( A \) is a sample operator and \( y \) is the sampled vector, an element of the \( M \)-dimensional space \( C^M \) which is consisted of \( \{y_i\}_{i=1}^M \). The study purpose is to obtain an approximation to \( f_0 \) from \( y \). In the viewpoint of an inverse problem\(^{[7]}\), a kind of inverse operator, \( X \), of \( A \), is to be found under certain criterion, so that

\[
f = XY
\]  

(2)

becomes the best approximation to \( f_0 \).

In the newly proposed discriminant measure\(^{[3]}\), the inverse operator corresponding to a given pattern class, the target class, is required to minimize the mean energy of the general outputs of all other classes. That is, for class \( c (c=1,2,\cdots,C) \) of \( C \) classes, the optimal inverse operator should satisfy

\[
X_D^{(c)} = \arg \min_{X} \left\{ \text{mean}_{i \neq c} \left\| X^{(c)} Y^{(c)} \right\|^2 \right\}
\]  

(3)

where \( Y^{(c)} \) is the sampled vector of class \( i \) \( (i \neq c, i=1,2,\cdots,C) \), \( \left\| \cdot \right\| \) denotes the norm associating with the Hilbert space \( H \). In fact, Eq.(3) is equivalent to

\[
X_D^{(c)} = \arg \min_{X} \left\{ tr(X^{(c)} Q (X^{(c)})^*) \right\}
\]  

(4)

where \( (\bullet)^* \) and \( tr(\bullet) \) respectively denote the adjoint operator and the trace of an operator, and

\[
Q = \frac{1}{C-1} \sum_{i \neq c}^C Y^{(i)} Y^{(i)^*} = Y^{(c)} \otimes Y^{(c)^*}
\]  

(5)

with \( (\bullet \otimes \bullet) \) the Neuman-Schatten product\(^{[8]} \). Eq.(4) results in a classifier called KND\(^{[3]} \).

**Proposition** The kernel-based nonlinear discriminator of class \( c \) is expressed by

\[
f_{D}^{(c)}(x) = \sum_{i=1}^M a_{D,i}^{(c)} k(x, x_i)
\]  

(6)

with the coefficients constitute the following vector

\[
a_D^{(c)} = [a_{D,1}^{(c)} \ a_{D,2}^{(c)} \cdots a_{D,M}^{(c)}]^T = (I_M - Q^+ Q) y^{(c)}
\]  

(7)

where \( T \) denotes the complex transpose of a matrix, \( Q^+ \) is the Moore-Penrose pseudoinverse of \( Q \), and \( I_M \) is the identity operator of the sampled space \( C^M \).

### 2 Optimal Feature Representation and KNR

As for the optimal approximation problem discussed above, a natural criterion is to minimize the distance between \( f \) and \( f_0 \) in the metric of the space \( H \). But it is impossible to solve this problem because both \( f_0 \) in Eq.(1) and \( X \) in Eq.(2) are unknown. Ideally, for any such an unknown function \( f_0 \), if the estimated version equals exactly to the original one, then there will be no estimation error. In this special case, Eqs.(1) and (2) yield

\[
XA = I
\]  

(8)

where \( I \) is the identity operator of the Hilbert space \( H \). It can be easily shown that a general solution to Eq.(8) is given by

\[
X = A^* + Y - A^* Y A A^*
\]  

(9)

where \( Y \) is any operator from \( C^M \) to \( H \) which satisfies

\[
(I - A^* A) Y A = I - A^* A
\]  

(10)

Eqs. (1), (2), (9), and (10) yield

\[
f = A^* f_0 + (I - A^* A) f_0 = f_1 + f_2 = f_0
\]  

(11)

with

\[
f_1 = P_{(A^*)^+} f_0 = A^* f_0
\]  

(12)

and

\[
f_2 = P_{(A^*)^+} (I - A^* A) f_0
\]  

(13)

where \( P_{(A^*)^+} \) and \( P_{(A^*)^+} \) are the orthogonal projection operators onto the range of \( A^* \) and the null space of \( A \), respectively.

Notice that the assumption led to Eq.(8) is too ideal to be practically satisfied. According to the Projection Theorem in functional analysis, if certain error is allowed to the estimated result and let

\[
f \approx f_1 = P_{(A^*)^+} f_0
\]  

(14)

then the distance between \( f \) and \( f_0 \) is minimized, in the metric of the space \( H \). It has been proved that \( \Re(A^*) \) is the largest subspace of \( H \), within which the best approximation to any desired function is obtained\(^{[9]} \).

Theorem (3.8) of Ref\.[10] and Eqs.(14), (12) and
lead us to the following theorem, proof of which is similar to that of Proposition \[^3\].

**Theorem** The kernel-based nonlinear representor (KNR) of class $c$ is expressed by

$$ f_\theta^{(c)}(x) = \sum_{j=1}^{M} a_{k,j}^{(c)} k(x, x_j) $$

with

$$ a_k^{(c)} = [a_{k,1}^{(c)}, a_{k,2}^{(c)}, \ldots, a_{k,M}^{(c)}]^T = K^c y^{(c)} $$

where

$$ (K)_{ij} = k(x_i^{(c)}, x_j^{(c)}) \quad i, j = 1, 2, \ldots, M \quad (17) $$

That is, $K$ is a matrix determined by the associated reproducing kernel of the Hilbert space $H$ and the $M$ training patterns of class $c$.

### 3 Application to Classification of Eigenfaces

This section applies KNR, together with KND, to eigenface classification. We choose the ORL face database that contains 40 distinct subjects with 10 frontal face images per subject. These images are taken at different periods of time, with different lightening conditions and facial poses, expressions, and details (with/without glasses). In this application, Gaussian kernel is adopted and a simple technique using Eq.(8) of Ref.[1] or its alteration are used to adaptively estimate the kernel width (arbitrarily used, without justification, only because the representations of a KND and a KNR are similar to that of the Parzen classifier). Before PCA is performed to obtain eigenfaces, averaging method is used to down-sample each image, from resolution of $92 \times 112$ to $23 \times 28$, to reduce the computational load. The Euclidean distance classifier is chosen for comparison because it is widely used, say in Ref.[11], to compare the effectiveness of different feature extraction methods.

We conduct two sets of experiments. In the first one, five images of each subject are randomly chosen for training, and the remained ones for test. Trained eigenfaces of all subjects are used to estimate the kernel width of common use and the coefficients of each KNR or KND. Ten different runs are performed and ten feature dimensions are selected. For each dimension, classification rates are averaged over the ten runs and the forty subjects, and the results are listed in Tab.1. Notice that the feature dimension 40 corresponds to the data compression rate 32% (about one third of the data of each down-sampled image are retained) in PCA. Near this point, the KNR classifier improves performance by around 6 points, with respect to classification rate, over the Euclidean distance classifier, which in turn performs better than the KND classifier. The more data compression is, the more relative improvement of the KNR classifier over the Euclidean distance one is obtained. Among the ten feature dimensions, the best result (Error rate of 6.6%) of the KNR is favorably comparable with those of some current methods on the same database, such as the kernel density estimation (KDE) classifier for eigenfaces (6.55%)\[^{[12]}\], the individual PCA (5.0%)\[^{[13]}\], Bayesian classifier for SVD coefficients (7.5%)\[^{[11]}\], and the direct LDA (9.2 % and about 6.0%)\[^{[14,15]}\].

**Tab.1** Classification rates (percent) of Euclidean distance (ED), KND, and KNR classifiers vs. feature dimensions (FD)

<table>
<thead>
<tr>
<th>FD</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>82.4</td>
<td>83.9</td>
<td>85.0</td>
<td>85.9</td>
<td>87.0</td>
<td>87.3</td>
<td>87.6</td>
<td>88.0</td>
</tr>
<tr>
<td>KND</td>
<td>81.2</td>
<td>81.4</td>
<td>82.2</td>
<td>81.0</td>
<td>81.3</td>
<td>82.4</td>
<td>81.9</td>
<td></td>
</tr>
<tr>
<td>KNR</td>
<td>88.5</td>
<td>90.8</td>
<td>92.6</td>
<td>92.9</td>
<td>93.0</td>
<td>93.2</td>
<td>93.4</td>
<td></td>
</tr>
</tbody>
</table>

In the second experiment, we fix the feature dimension to 40 (not the best one) to study the effect of the number of training images on the classification rate. Three to eight images per subject are randomly chosen for training, and the rest ones for test. The averaged classification rates, again over ten different runs and forty subjects, are listed in Tab.2. For the three classifiers, classification rates improve with the number of training samples. Once more, the KNR classifier performs best, it earns over 95% classification rate when more than five images per subject are chosen for training.

**Tab.2** Classification rates (percent) of Euclidean distance (ED), KND, and KNR classifiers vs. number of training images per subject (TI), with the same feature dimension of 40

<table>
<thead>
<tr>
<th>TI</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>85.2</td>
<td>86.3</td>
<td>87.3</td>
<td>90.0</td>
<td>91.2</td>
<td>91.4</td>
</tr>
<tr>
<td>KND</td>
<td>74.6</td>
<td>80.0</td>
<td>81.3</td>
<td>85.6</td>
<td>87.7</td>
<td>87.0</td>
</tr>
<tr>
<td>KNR</td>
<td>88.5</td>
<td>91.2</td>
<td>93.0</td>
<td>95.9</td>
<td>96.9</td>
<td>96.5</td>
</tr>
</tbody>
</table>
4  Conclusions

This paper presents a kernel-based nonlinear representer for optimal pattern feature representation. It has a closed form solution, and thus any quadratic programming procedure is avoided. A Gaussian kernel-based KNR is applied to classification of eigenfaces obtained by PCA from the ORL face database. The results show satisfactory performance of the KNR in face recognition. However, the previously proposed KND perform relatively poor in the experiments, and a better performance is expected if the effect of the target class is simultaneously taken into consideration.

Acknowledgement

The related work is partly supported by the Yong Scholar Foundation of UESTC (No. JX200402).

References


Brief Introduction to Author(s)

ZHANG Jing (张 晶) was born in 1972. She is now an engineer and postgraduate student. Her research interests include: computer network, information security, and signal processing. E-mail: zh.j@scrivu.net.

LIU Ben-yong (刘本永) was born in 1966. He is now an Associate Professor at UESTC. His research interests include pattern recognition, signal processing, computational intelligence, and information security. E-mail: byliu@uestc.edu.cn.

TAN Hao (谭 浩) was born in 1970. He is now an Associate Professor at UESTC. His research interests include: middleware, multimedia, and information security.