A Novel Method for Face Recognition based on Genetic Algorithm Optimized Kernel Extreme Learning Machine

Jiajun ZHANG
School of Mathematics and Statistics, Zaozhuang University, Zaozhuang, China
Tel.: +86-632-3786721, fax: +86-632-3786721
E-mail: zhangJiajun1975@yeah.net

Received: 20 September 2013 /Accepted: 22 November 2013 /Published: 30 December 2013

Abstract: The artificial intelligent classifiers have been proven to be efficient in face recognition; however, to meet the demand of online recognition, they need enhance the recognition accuracy and speed. In order to resolve this issue, the kernel extreme learning machine (KELM) has been proposed to provide quick and accurate pattern recognition ability. The only parameter need be determined in KELM is the neuron number of hidden layer. Suitable neuron number will accelerate the training procedure. However, little work has been done to select proper neuron number in the application of face recognition. To address this issue, this paper presents a new method that uses the genetic algorithm (GA) to optimize the KELM parameter for face recognition. After the determination of proper hidden layer neuron number, the face recognition accuracy and speed of KELM could meet the online application requirements. Experiments have been carried out to evaluate the proposed method. The performance of the GA-KELM was compared with KELM, ELM, and LS-SVM. The analysis results indicate that the proposed GA-KELM outperforms its rivals in terms of both recognition accuracy and training speed. Copyright © 2013 IFSA.

Keywords: Biometric identification, Face recognition, GA, KELM.

1. Introduction

Face recognition has been widely used in areas of human-computer interaction, security systems, criminal identification, teleconference, image and film processing, etc [1]. Human face expressions could be regarded as low dimensional manifold resided in the original images [2]. If recognizing the face images in the low dimensional space, the recognition precision and speed will be improved significantly [3]. Especially for practical online applications, the face recognition precision and speed are strictly demanded. Hence, how to develop a practicable tool for the face recognition attracts extensive attentions. To meet the industry requirements, a possible solution for high performance of face recognition is to reduce the dimensionality of the image matrix into a low dimensional space. By doing so, distinct features could be easily extracted by signal processing techniques. To realize this, the principal component analysis (PCA) has been introduced into the face recognition [4]. Although the PCA [4] is very efficient for image pre-processing by extract some distinct features of the original image data into a low dimensional space, their main limitations lie in the extraction of nonlinear properties of the original data [5]. In contrast to PCA, the Kernel PCA (KPCA) can extract nonlinear properties from the original data [5]. Then useful face expression features can be
obtained in the low dimensional space. The remaining question is how to efficiently utilize these features.

Up to date, the artificial intelligent methodologies have been widely used in the face image recognition. Two common techniques are the artificial neural network (ANN) and support vector machine (SVM). For some type of ANNs, such as BP NN and RBF NN, they are able to adaptively learn the inherent patterns hidden in the given data; however, they often suffer from local minima and slow convergence speed [6]. For SVM, it needs to set the kernel function, error control parameters, and penalty coefficient. It is difficult to select these parameters. Hence, although ANN and SVM have contributed a lot in machine learning and data analysis [6], they face some challenging issues (i.e. slow learning speed and poor learning scalability). These disadvantages limit their applications in practice.

In order to overcome this problem, the kernel extreme learning machine (KELM) has been proposed as an integration of ANN and SVM to provide quick and accurate pattern recognition ability [7]. The KELM is a variant of the conventional LS-SVM but adopts the structure of single-hidden layer feed-forward networks (SLFNs). The classical ANN and SVM learning algorithms require setting several defined parameters and may produce the local minimum. However, the KELM only needs to set up the number of hidden layer nodes of the network [7, 8]. Siddartha et al [3] employed the extreme learning machine (ELM) to extract features in face recognition. Zong and Huang [9] introduced the ELM into the face recognition.

Experimental tests have been carried out in the study to show high performance of the ELM in multi-label face recognition applications. Further, they [7] have presented the KELM in the face recognition and found that the KELM outperforms LS-SVM in terms of both recognition prediction accuracy and training speed. However, a parameter optimization mechanism of the KELM or ELM has not developed in existing work. Proper setting of the neuron number of the KELM can enhance the training speed and accuracy [10]. It is therefore imperative to develop an optimization mechanism for the KELM.

To address the mentioned issue, a new solution based on the GA-KELM has been proposed for face recognition in this work. The GA is used to tune the neuron number of the KELM to make the KELM achieve high generalization performance. A series of experimental tests have been implemented to evaluate and verify the effectiveness of the proposed method in face recognition. The recognition performance of the proposed GA-KELM has been compared with KELM, ELM, and SVM in terms of both recognition prediction accuracy and training speed.

2. The Proposed Face Recognition Method

As mentioned above, this work will introduce the GA-KELM method for the face recognition. The face recognition is carried out in a low dimensional space of the original images through the transform of KPCA. The theories about KPCA and GA-KELM are briefly described as follows.

2.1. Kernel Principal Component Analysis

Given $X = [x_1, x_2, ..., x_n] \in R^{d \times n}$, where $d$ is the data dimension and $n$ is sample number, the covariance matrix $C$ can be expressed as

$$C = \frac{1}{n}XX^T$$

Conduct the eigenvalue decomposition we can obtain

$$C = U \Lambda U^T = \sum \lambda_i u_i u_i^T,$$

where $\Lambda$ is the eigenvalues and $U$ is the eigenmatrix. Then, the original matrix $X$ can be projected along the first several eigenvalues to form a new reduced matrix $X_k \in R^{k \times n}$, where $k$ is the new dimension. This is the PCA algorithm. To extend PCA to higher dimensional space, i.e. Reproducing Kernel Hilbert Space (RKHS), a nonlinear map has been used, i.e. $\phi: R \rightarrow Q$ [5]. Then, the kernel functions can be adopted in the RKHS to evaluate the inner product of two points:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Assuming that the data are centered in $Q$, the covariance matrix for KPCA can be obtained by

$$\bar{C} = \frac{1}{n} \phi(X)\phi(X)^T$$

Using the eigenvalue decomposition we can get the kernel principal components.

2.2. GA optimized KELM

Given $\{(x_i, t_i) : i = 1, 2, ..., N; x_i \in R^d, t_i \in R^q\}$, where $x$ is the feature vector and $t$ is the class label vector, the below SLFN is used to identify the sample [8]

$$\sum_{i=1}^m \beta_i g(\alpha_i^T x - b_i) = o_i, j = 1, 2, ..., N,.$$
where \( m \) is the number of hidden neuron; \( o_i \) is the output of \( j^{th} \) sample; \( g(\cdot) \) is the activation function; \( b_i \) is the threshold of the \( i^{th} \) hidden neuron; \( \alpha_j \) and \( \beta_i \) are the input and output weight vectors, respectively. If the output \( o \) can approximate \( t \), we derive

\[
\sum_{i=1}^{m} \beta_i g(\alpha_i^T x_j - b_i) = o_j = t_j, j=1, 2, ..., N. \tag{6}
\]

(6) can be written compactly as

\[
G\beta = T,
\]

where

\[
G = \begin{bmatrix}
g(\alpha_1^T x_1 - b_1) & \cdots & g(\alpha_m^T x_1 - b_m) \\
\vdots & \ddots & \vdots \\
g(\alpha_1^T x_N - b_1) & \cdots & g(\alpha_m^T x_N - b_m)
\end{bmatrix},
\]

\[
\beta = [\beta_1, \cdots, \beta_m]^T \quad \text{and} \quad T = [t_1, \cdots, t_N]^T.
\]

To solve (7), the ELM adopts a least squares error to get solution \( \hat{\beta} \):

\[
\hat{\beta} = G^+ T, \tag{1}
\]

where \( G^+ \) is the Moore-Penrose generalized inverse of \( G \). Function \( g(\cdot) \) is usually unknown, we can incorporate kernel functions in \( g(\cdot) \). This is the so called KEML. The kernel matrix

\[
K = \begin{bmatrix}
K(x_1; x_1) & \cdots & K(x_1; x_N) \\
\vdots & \ddots & \vdots \\
K(x_N; x_1) & \cdots & K(x_N; x_N)
\end{bmatrix} (K(\cdot) \text{ is the kernel function})
\]

is introduced into (7) and (8) to estimate the output of the KELM:

\[
o = KT \tag{9}
\]

Herein, the Gaussian kernel function (RBF) is adopted

\[
K(x_1; x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma}\right), \tag{10}
\]

where \( \sigma \) is the width of RBF.

The number of hidden neuron \( m \) needs to be specified. In order to obtain a proper \( m \), the GA has been used to tune it in the training processing of the KELM. An initial value of \( m \) is firstly randomly selected; then GA uses the inheritance, mutation, selection, and crossover [11] to search the optimized value of \( m \).

A diagram block of the proposed face recognition method is shown in Fig. 1.

3. Experiments and Results

The Yale [12] database has been used to evaluate and verify the proposed method in this paper. The Yale contains frontal grayscale face images of 15 people, with 11 face images of each subject. Facial expression variations include normal, sleepy, surprised, happy, sad, and wink. Out of these 11 face images of each individual, 5 images are taken as training data and remaining 6 images are used for testing.

In the present work, the KPCA was employed to extract useful features of the face images in a low dimensional space. Then the GA-KELM was applied to the recognition of the face images. To evaluate the performance of the proposed method, the KPCA was compared with some popular extraction algorithms (i.e. PCA, independent component analysis (ICA), and linear discriminant analysis (LDA)) with the same classifier. Moreover, the GA-KELM was compared with KELM, ELM, and SVM with the same feature extraction algorithm. Fig. 2 and Fig. 3 show the comparison results.

It can be seen in Fig. 2 that the KPCA based feature extraction can generate the best face recognition rate against the PCA, ICA and LDA. The best recognition rate of KPCA is 93.7 % when the input feature dimension of the GA-KELM classifier is 38. In contrast, the best recognition rates of PCA, ICA and LDA are 82.7 %, 83.7 % and 88.7 %, respectively. Hence, the KPCA extraction could improve the face recognition rate by 5.1 % or better among the 4 feature extraction algorithms. This is because the kernel trick can enhance the ability of KPCA to deal with the nonlinear components hidden in the original data. Distinct features can be extracted by KPCA to improve the face recognition rate.
Fig. 2. The comparison of the face recognition performance between KPCA, PCA, ICA, and LDA using the same GA-KELM classifier.

Fig. 3. The comparison of the face recognition performance between GA-KELM, KELM, ELM, and SVM using the same KPCA feature extraction algorithm.

In Fig. 3, one can note that the GA-KELM outperforms the rest and obtains the best face recognition rate. It can be also seen in Fig. 3 that the KELM can get better face recognition rate than that of ELM and SVM. This is because KELM adopts the kernel function to avoid the parameter of output bias in ELM [7]; hence there only one parameter needs to be determined. In the optimization processing, the less parameters to be tuned, the better optimization performance could be obtained. Thus, the KELM can provide better face recognition rate than that of ELM and SVM. However, in the KELM, if the neuron number could be optimized, the recognition performance will be improved. This explains why the GA-KELM gets better face recognition rate than that of KELM.

To highlight the effectiveness of the proposed face recognition method, we present the comparison between different feature extraction algorithms and different face recognition classifiers. Table 1 lists the comparison results. The hybrid of the KPCA, PCA, ICA, and LDA extraction algorithms and the GA-KELM, KELM, ELM, and SVM classifiers has been investigated in terms of both recognition accuracy and training speed. It can be seen in the table that both the recognition accuracy and training speed of the proposed method is among the best; the GA-KELM based face recognition with the same feature extraction algorithm can attain better performance than that of the KELM, ELM and SVM; the training speeds of the GA-KELM and KELM are equal but better than that of the ELM and SVM. Hence, the comparison results listed in Table 1 verify that the KELM has fast training speed owning to only one parameter of the structure, and the GA-KELM can enhance the face recognition rate.

Table 1. The comparison results of the face recognition.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition rate</th>
<th>Training time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA-SVM</td>
<td>78.7 %</td>
<td>0.027 s</td>
</tr>
<tr>
<td>PCA-ELM</td>
<td>78.3 %</td>
<td>0.021 s</td>
</tr>
<tr>
<td>PCA-KELM</td>
<td>80.3 %</td>
<td>0.018 s</td>
</tr>
<tr>
<td>PCA-GA-KELM</td>
<td>82.7 %</td>
<td>0.017 s</td>
</tr>
<tr>
<td>ICA-SVM</td>
<td>79.7 %</td>
<td>0.027 s</td>
</tr>
<tr>
<td>ICA-ELM</td>
<td>79.3 %</td>
<td>0.020 s</td>
</tr>
<tr>
<td>ICA-KELM</td>
<td>81.3 %</td>
<td>0.017 s</td>
</tr>
<tr>
<td>ICA-GA-KELM</td>
<td>83.7 %</td>
<td>0.017 s</td>
</tr>
<tr>
<td>LDA-SVM</td>
<td>86.3 %</td>
<td>0.026 s</td>
</tr>
<tr>
<td>LDA-ELM</td>
<td>85.7 %</td>
<td>0.021 s</td>
</tr>
<tr>
<td>LDA-KELM</td>
<td>87.7 %</td>
<td>0.018 s</td>
</tr>
<tr>
<td>LDA-GA-KELM</td>
<td>88.7 %</td>
<td>0.017 s</td>
</tr>
<tr>
<td>KPCA-SVM</td>
<td>87.7 %</td>
<td>0.027 s</td>
</tr>
<tr>
<td>KPCA-ELM</td>
<td>86.3 %</td>
<td>0.021 s</td>
</tr>
<tr>
<td>KPCA-KELM</td>
<td>91.3 %</td>
<td>0.017 s</td>
</tr>
<tr>
<td>KPCA-GA-KELM</td>
<td>93.7 %</td>
<td>0.017 s</td>
</tr>
</tbody>
</table>

4. Conclusions

In order to develop a practicable approach for the face recognition, a new method based on the integration of KPCA, GA and KELM has been proposed to meet the industrial requirements on the recognition accuracy and computation speed. The KELM only has one parameter to be determined and hence it is reasonable and reliable to employ the GA to optimize this parameter. By doing so, satisfactory face recognition performance could be obtained. The innovation of this work lies in the development and implementation of the KPCA and GA-KELM in the face recognition for the first time. A series of experimental tests have been carried out to verify the performance of the new method. The experimental analysis has showed satisfactory and effective face image identification performance of the proposed method. In addition, through comparison between different feature extraction algorithms (i.e. KPCA, PCA, ICA, and LDA) and different face recognition classifiers (i.e. GA-KELM, KELM, ELM, and SVM) it has shown that the performance of the proposed KPCA-GA-KELM method is superior to its rivals in...
terms of both recognition accuracy and training speed. Thus, the composition of KPCA feature extraction and GA-KELM classification show promising applications in the domain of face recognition.

Future research will focus on the industrial practice of the newly proposed method and it would be interesting to explore further the interplay between nonlinear feature extraction algorithms and kernel extreme learning machines.

References


[12]. The Yale database (http://cvc.yale.edu/).

2013 Copyright ©, International Frequency Sensor Association (IFSA). All rights reserved. (http://www.sensorsportal.com)