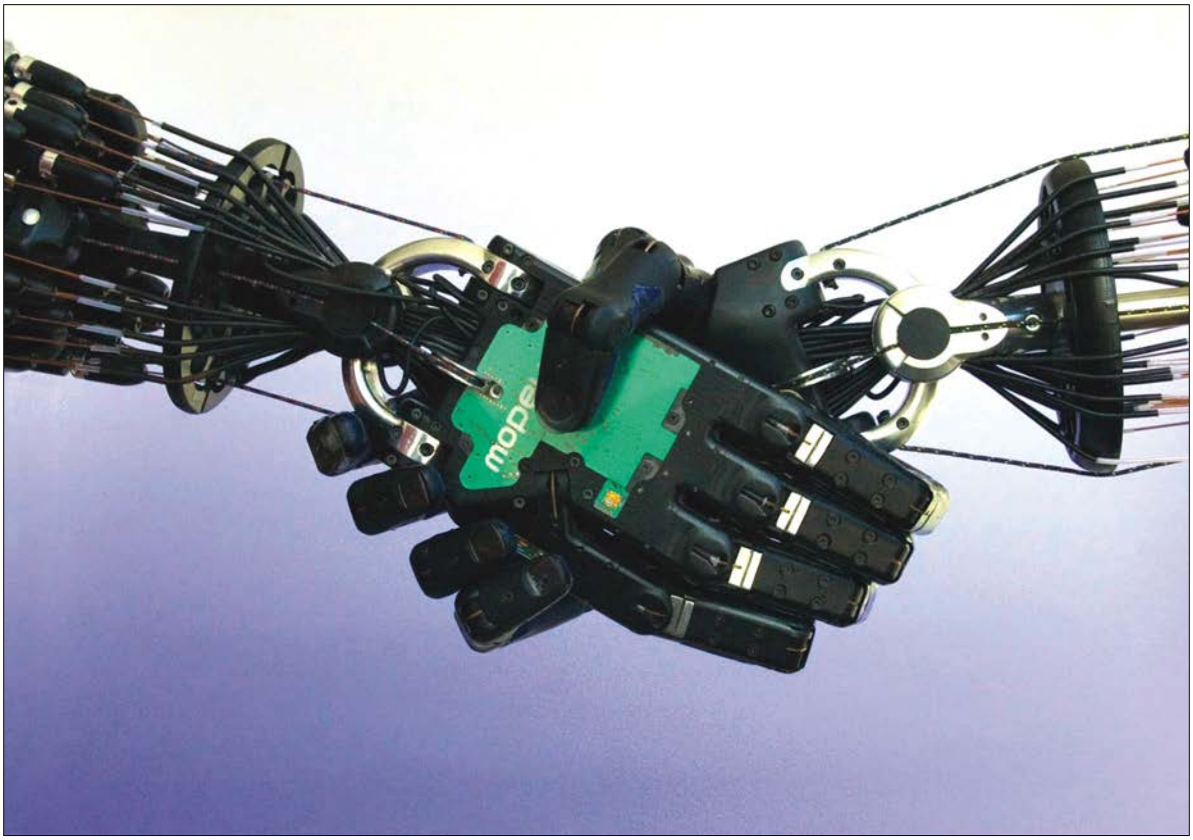

Journal of Artificial Intelligence

ISSN 1994-5450

Journal of Artificial Intelligence 3 (3): 168-173, 2010
ISSN 1994-5450 / DOI: 10.3923/jai.2008.28.33

Facial Gender Classification with Eigenfaces and Least Squares Support Vector Machine*

[1]Cheng Quanhua, [2]Liu Zunxiong and [1]Di Guoqiang
[1]School of Information Technology,
Jiangxi University of Finance and Economics, Nanchang 330013, China
[2]School of Information Engineering, East China Jiaotong University,
Nanchang 330013, China

Abstract: The techniques of eigenfaces and Least Squares Support Vector Machine (LS-SVM) classifier are combined to categorize gender from facial knowledge in this study, both of which explored. We will firstly establish the eigenfaces from the training images and obtain the projection coefficients for training and testing images in the space spanned by the eigenfaces. The LS-SVM classifiers are built with training coefficients, which are used for classifying training and testing images and classification accuracy percentage values are calculated. The experiments are implemented with our self-made facial images and the results demonstrate that LS-SVM classification has better performance than the other classification algorithms. We use cross validation to determine the number of selected principal components and kernel function parameter. In comparison, Fisher's Linear Discriminant (FLD) analysis is also implemented.

Key words: Facial gender classification, eigenfaces, LS-SVM classifier, cross validation

INTRODUCTION

Gender Identification is a basic task of face recognition, deciding the gender according to the face image. It is all known that many social interactions and services critically depend on the correct gender perception of the parties involved. It is visual information from human face that provides one of the more important sources of information for gender classification. We can tell very easily the gender of a person based on only a frontal-view image with the person's hair, nose, eyes, mouth and skin. That is to say, gender classification is very straightforward for us. But the situation is different for computer. So there are many investigations and researches on gender classification within the area of artificial intelligence (Moghaddam and Yang, 2002; Anthony *et al*., 2000), which are mainly about human image representation and classification algorithms.

In applications about face images, Principal Component Analysis (PCA) is a well established dimension reduction technique (Valentin *et al*., 1997; Abdi *et al*., 1995), which generates an eigenspace composed of eigenfaces based on a certain ensemble of images. The idea is to project images into this eigenspace and use the first few projected components for classification. The traditional classification algorithms include linear discriminant classifier (Valentin *et al*., 1997), Nearest-Neighbor Classifiers (Duda *et al*., 2001), Neural Networks (Golomb *et al*., 1991; Cottrell and Metcalfe, 1991). With the significant progress in human face recognition techniques, some new powerful methods are introduced in gender classification, such as Support Vector Machine (SVM) (Moghaddam and Yang, 2002). It is a widely used method of classification based on statistic learning theory (Burges, 1998) and the optimization problem is with Structural Risk Minimization and not the Empirical Risk Minimization induction. In this way does SVM possess better generalization ability. Moreover the kernel functions

Corresponding Author: Liu, Zunxiong, School of Information Engineering, East China Jiaotong University, Nanchang 330013, China

**Originally Published in Journal of Artificial Intelligence, 2008*

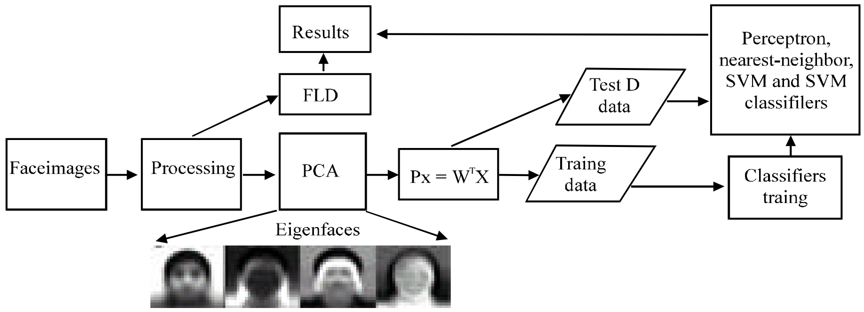

Fig. 1: The flow chat for the simulation experiments

which meet Mercers theory are incorporated into SVM, so the nonlinear problems in low dimensional space can be solved in linear scheme in the corresponding high dimensional space. Least Squares Support Vector Machines (Suykens and Vandewalle, 1999), as a modified Version for Support Vector Machine (SVM), takes the equality constraints in place of the inequality counterparts with SVM, the solution follows from solving a set of linear equations, instead of quadratic programming for classical SVMs. Recently, Kim H. Chul *et al*. (2006) took Gaussian Process Classifiers (GPCs) for gender classification and got good results.

Fishers Linear Discriminant (FLD) is another popular form of dimensionality reduction, which is also a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space (Yambor, 2000).

In this study, Least Squares Support Vector Machine (LS-WSVM) classifier is proposed for the task of facial gender classification. The face images are first reduced in dimensionality using PCA and the subset of the coefficients are presented to train the corresponding classifiers which are tested on the unknown faces. The gender classification performance with the LS-SVM classifier will be compared with the other techniques. This study flowchart is shown as (Fig. 1).

EIGENFACE APPROACH

Here every image is treated as a vector in a very high dimensional space and Eigenfaces are the eigenvectors of the covariance matrix of the training image vectors, which incorporate the variation amongst the face images. Now each image in the training set would have its contribution to the eigenvectors (variations). Eigenfaces represent its contribution in the variation between the images, which dimension is the same as the face image vector. Moreover they can be displayed. The face image can be represented as a linear combination of the Eigenfaces, the combination coefficients are the projection of the original face image into Eigenfaces space. The amount of overall variation that one eigenface counts for, is actually known by the eigenvalue associated with the corresponding eigenvector. If the eigenface with small eigenvalues are neglected, then an image can be a linear combination of the selected eigenfaces. The training image coefficients are used to train the classifier.

The face gray image which size is N pixels in width and N pixels in height, can be considered as a vector of dimension N^2. Given N = 100, the corresponding image becomes a vector of dimension 10,000 or equivalently a point in a 10,000-dimension space. Let the training images be I_1, I_2, ...I_M, then the average face is calculated as

$$\psi = \left(\frac{1}{M}\right)\sum_{i=1}^{M} I_i \ .$$

All the training images are mean-adjusted to obtain $\varphi = I_1 - \psi$. We can get the covariance matrix with the training image vectors:

$$C = \left(\frac{1}{M}\right)\sum_{i=1}^{M} \varphi_i \varphi_i^T = AA^T \tag{1}$$

Where, $A = [\varphi_1, \varphi_2, ... \varphi_M]$, the dimension of matrix C is $N^2 \times N^2$, with which there exist N^2 eigenvectors and eigenvalues. However, their calculations would be impractical to implement. Generally, the number of the training images is less than the number of pixels in an image. We can obtain the related eigenvectors through solving an $M \times M$-dimensional matrix $A^T A$ instead of a $N^2 \times N^2$ matrix $A A^T$. Supposing that the eigenvalue problem with $A^T A$ meets as follows:

$$A^T Av_i = \lambda_i v_i \quad (2)$$

Where, λ_i is the eigenvalue related to the eigenvector v_i with the $M \times M$ covariance matrix and there are M eigenvectors. Premultiplied by A Eq. 2 is transformed into

$$A A^T Av_i = \lambda_i Av_i \quad (3)$$

The right hand side gives us the M eigenfaces of the order N^2 by 1. All such vectors would make the imagespace of dimensionality M. By selecting the M' eigenfaces which have the largest associated eigenvalues (M<M), the eigenfaces now span aM'-dimensional subspace instead of N^2 without losing much more information.

LS-SVM FOR FACIAL GENDER CLASSIFICATION

LS-SVM is a powerful technique used for pattern recognition including face recognition. Given a training vector set of N images $\{x_k, y_k\}_{k=1}^{N}$ with the input data $x_k \in R^{M\,p}$ which is the combination coefficients vector of the selected eigenfaces for k-th image and $y_k \in \{1,-1\}$. Kernel mapping can map the training examples in eigenfaces space into a high-dimension feature space in which linear classification can be implemented with the mapped eigenfaces space data. In LS-SVM, the model is constructed by solving the following optimal problem:

$$\min_{w,b,e} J_p(w,e) = \frac{1}{2} w^T w + C \frac{1}{2} \sum_{i=1}^{N} e_i^{\,2} \quad (4)$$

subject to

$$y_i[w^T \varphi(x_i) + b] = 1 - e_i, i = 1, \cdots, N$$

The discrimination function takes the form:

$$y = \text{sign}\,[w^T \varphi\,(x) + b] \quad (5)$$

with $\varphi(\cdot)$: $R^n \rightarrow R^{nh}$ a function which maps the input space into a so-called high dimensional (possibly infinite dimensional) feature space, C the regularization parameter, weight vector $w \in R^{nh}$ in primal weight space, error variables $e_k \in R$ and bias term b. Note that the cost function J_p consists of a sum of squared fitting error (SSE) and a regularization term.

The model should be computed in the dual space, the Lagrangian function can be defined:

$$L(w, b, e, a) = J_p(w, e) - \sum_{i=1}^{N} a_i \{y_i w^T j \varphi(x_i) + y_i b + e_i - 1\} \quad (6)$$

Where, $\alpha_i \in R$ are Lagrange multipliers. For this nonlinear classifier formulation, the Lagrangian is solved, which results in the following dual problem to be solved in α, b:

After elimination of w and e, one obtains the solution

$$\begin{bmatrix} 0 & y^T \\ y & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 1_N \end{bmatrix} \tag{7}$$

with $y = [y_1, y_2, \dots y_N]^T$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$, $1 = [1, \dots 1]^T$, $k_{ij} = k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)^T$ is sapplied with the Ω matrix, $\Omega_{ij} = y_i\, y_j\, \varphi(x_i)^T \varphi(x_j) = y_i y_j K(x_i, x_j)$, I, j =1, ..., N.

The classifier in the dual space takes the form:

$$y(x) = \sum\nolimits_{k=1}^{N} a_k y_k K(x, x_k) + b \tag{8}$$

Where, a and b are the solution to Eq. 7. The chosen kernel function should be positive definite and satisfy the Mercer condition.

SIMULATIONS AND RESULTS

The comparison experiments will be implemented in the section. The used face dataset consists of 200 male and 200 female images, whose subjects is from our school taking the National Computer Rank Examination (NCRE). These images were taken using a frontal-view, almost in the same size and of the format JPEG. Then, they were converted to HSV images and keep only the value plane. After that, All faces were rescaled to 100 by 100 pixels in size and are normalized to the range (0,1). The simulations were done 10 separate times, for every experiment the training and testing data were set as following: Firstly, all the male and female images were random-sorted separately. Twenty males and 20 females were selected in sequence as testing set for one simulations, leaving the rest of the images to train the algorithms including least squares support vector machine, thus allowing all of the images to be in a test set exactly once; Then the training and testing were carried out for every data partition, the trained classifiers were all tested in the training set and testing set and the algorithm results were averaged across all tests to increase their statistical significance.

Before training the classifiers, the corresponding eigenfaces should be achieved with the dimension reduction technique PCA with the training data. All images were projected into the space spanned by the eigenfaces to obtain the reduced data, which were used for training and testing the classifiers. As an example, the first 10 eigenfaces were plotted in Fig. 2 for the first-time simulation.

With the LS-SVM classifiers, the kernel function is taken as RBF function because RBF kernel function,

$$\varphi(X_i)^T(X_j) = e^{\|x_i - x_j\|^2 / 2*\theta^2}$$

has the property of Symmetry. For the width (θ, p) for RBF function and the chosen primary component number p, we perform cross-validation with (θ, p) combination. In the simulations, the value of θ is selected as 16.

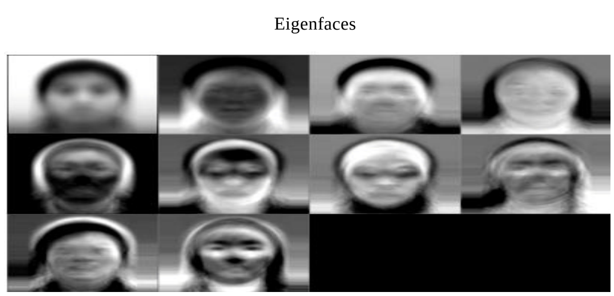

Fig. 2: Eigenfaces with 360 images for the first simulation

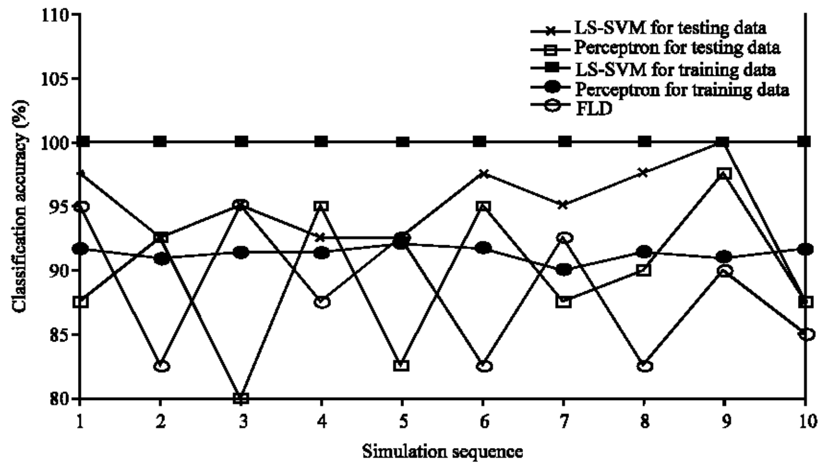

Fig. 3: Classification accuracy with LS-SVM, perceptron and FLD for 10 separate simulations

Table 1: Average and SD. of accuracy for 10 simulations on testing data with different number of eigenfaces

p	2	3	4	5	6	7	8	9	10	11	12	13
Avg. of Acc.	79.50	82.30	93.50	94.00	94.30	94.30	94.00	94.80	94.80	94.80	94.80	94.30
SD. of Acc.	06.82	06.92	04.28	04.59	04.09	04.09	03.37	02.99	03.62	03.22	04.32	04.09
p	14.00	15.00	16.00	17.00	18.00	19.00	20.00	21.00	22.00	23.00	24.00	25.00
Avg. of Acc.	94.30	94.80	95.50	94.50	94.50	94.00	95.00	95.50	95.80	95.80	95.80	95.80
SD. of Acc.	04.09	03.99	03.07	04.05	04.05	04.89	03.91	03.07	03.13	03.13	03.13	03.74

Varying the number of eigenfaces will have the effect on performance of LS-SVM classifiers, we implemented experiments with different number of eigenfaces from 2 to 25 for the 10 separate data partitions, the averages and deviations of classification accuracy for 10 simulations are calculated with respect to the number of eigenfaces. The LS-SVM classifiers match the training data perfectly well, which classification accuracy values are all 100% and the performance of the LS-SVM verses the number of eigenfaces used during preprocessing for testing data are shown in Table 1.

The performance of the algorithm for the testing data was nearly stable when using eight or more eigenfaces (Table 1). Because the performance of the algorithm did not increase when using more than eight eigenfaces and the calculation complexity can be reduced.

In present experiments, perceptron classification was implemented with the same face partitions, the classification accuracy values on the training and testing data were calculated respectively as comparison. Figure 3 shows the classification accuracy values on 10 training data and 10 testing data with the two models, in which the lines with asterisks are for LS-SVM classifiers and the lines with squares are for Percetron classifiers. From Fig. 3 it is known that the classification accuracy values on training data are better than on testing data, the reason for it is that the classification models should match training data. The training accuracy average value remains 100% with LS-SVM, but it is 91.278% for Perceptron, the performances in the training data have better stability for both LS-SVM and Perceptron. We also obtain the test accuracy average value of 94.75 with LS-SVM, which is more five percent points than the classification accuracy average value for Perceptron. Among the four curves, the testing accuracy values with Perceptron swing a lot for the 10 separate simulations, averages in 89.5%.

However, Fisher's Linear Discriminant analyses are implemented on the face images for 10 separate times, the test data taking 10% of all the face images, the average accuracy is 88.5%, not satisfying. Because FLD combines dimensionality reduction to 1D and classification, it was applied just in the case that final data dimension was one. In comparison, the performance data for FLD, PCA with Perceptron and LS-SVM while reducing the dimensionality to 1D, are shown in Table 2.

Table 2: Average accuracy for 10 simulations with FLD and PCA

	FLD	PCA(1D)+Perceptron	PCA(1D)+LS-SVM
Avg. of Acc.	88.5%	40.5%	50.9%

Table 2 shows that when the dimension number is selected as one the performance with PCA is much worse, but the accuracy data for FLD is better. That reason is that in the case of one dimension there exists more information loss with PCA. So the threshold criterion should be used to choose the principal components on the work with PCA.

CONCLUSION AND FUTURE WORKS

Since LS-SVM theory built, it has been widely used for pattern recognition, data-mining, system identification and etc. LS-SVM classifier is of high quality and efficiency, which settles linear problems in high dimension mapping space and avoids quadratic optimization problem. We take LS-SVM technique for facial gender classification and implement experiments with NCRE face images from our university. The experiment results on training and testing images demonstrate that LS-SVM has better classification performance.

The technique of Primary Component Analysis is utilized to reduce face images dimension, which transforms the experimental images onto lower-dimensional eigenface space to finish data preprocessing. Since the number of eigenfaces, that is, the number of coefficients to represent the image, can be chosen arbitrary, we also use cross validation to determine and settle that ten coefficients would adequately represent the data and would not affect algorithm performance significantly. However, dimension reduction is an important research aspect with face recognition and there are many methods and their modified version available, we will devote ourselves to data preprocessing and different classification algorithm comparison.

REFERENCES

Abdi, H., D. Valentin, B.E. Edelman and A.J. O'Toole, 1995. More about the difference between men and women: Evidence from linear neural networks and the principal component approach. Perception, 24: 539-562.

Anthony, D., D. Jeffrey and S. Juan, 2000. Gender Recognition-Final Project for EE368. http://white.stanford.edu/~dicarlo/ee368/.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2: 2.

Chul, K.H., K. Daijin, Z. Ghahramani and S.Y. Bang, 2006. Appearance-based gender classification with Gaussian processes. Pattern Recognition Lett., 27: 618-626.

Cottrell, G.W. and J. Metcalfe, 1991. EMPATH: Face, emotion and gender recognition using holons. In: Advances in Neural Information Processing Systems, pp: 564-571.

Duda, R.O., P.E. Hart and D.G. Storkm, 2001. Pattern Classification. John Wiley and Sons.

Golomb, A., D.T. Lawrence and T.J. Sejnowski, 1991. SEXNET: A neural network identifies sex from human faces. In: Advances in Neural Information Processing Systems, pp: 572-577.

Moghaddam, B. and M.H. Yang, 2002. Gender classification with support vector machines. IEEE. Transact. Pattern Anal. Mach. Intelligence, 24: 707-711.

Suykens, J.A.K. and J. Vandewalle, 1999. Least squares support vector machine classifiers. Neural Processing Lett., 9: 293-300.

Valentin, D., H. Abdi, B. Edelman and A.J. O'Toole, 1997. Principal component and neural network analyses of face images: What Can Be Generalized in Gender Classification? J. Math. Psychol., 41: 398-412.

Yambor, W.S., 2000. Analysis of PCA-based and fisher discriminant-based image recognition algorithms. Maintained by Colorado State University, USA.