Learning Hierarchical Representations for Face Recognition using Deep Belief Network Embedded with Softmax Regress and Multiple Neural Networks

Hai-jun Zhang*[1, a], Nan-feng Xiao[1, b]

[1]School of Computer Science and Engineering, South China University of Technology, China

*,[a]nihaoba_456@163.com, [b]xiaonf@scut.edu.cn

Abstract   In face recognition and classification, feature extraction and classification based on insufficient labeled data is a well-known challenging problem. In this paper, a novel semi-supervised learning algorithm named deep belief network embedded with Softmax regress (DBNESR) is proposed to address this problem. DBNESR first learns hierarchical representations of feature by deep learning and then makes more efficient classification with Softmax regress. At the same time we design many kinds of classifiers based on supervised learning: BP, HBPNNs, RBF, HRBFNNs, SVM and multiple classification decision fusion classifier——hybrid HBPNNs-HRBFNNs-SVM classifier. The conducted experiments validate: Firstly, the proposed semi-supervised deep learning algorithm DBNESR is optimal for face recognition with the highest and most stable recognition rates; Second, the semi-supervised learning algorithm has better effect than all supervised learning algorithms; Third, hybrid neural networks has better effect than single neural network.

1 Introduction

Face recognition (FR) is one of the main areas of investigation in biometrics and computer vision. It has a wide range of applications, including access control, information security, law enforcement and surveillance systems. FR has caught the great attention from large numbers of research groups and has also achieved a great development in the past few decades [1-2]. However, FR suffers from some difficulties because of varying illumination conditions, different poses, disguise and facial expressions and so on [3-4]. A plenty of FR algorithms have been designed to alleviate these difficulties [5-6]. FR includes three key steps: image preprocessing, feature extraction and classification. Image preprocessing is essential process before feature extraction and also is the important step in the process of FR. Feature extraction is mainly to give an effective representation of each image, which can reduce the computational complexity of the classification algorithm and enhance the separability of the images to get a higher recognition rate. While classification is to distinguish those extracted features with a good classifier. Therefore, an effective face recognition system greatly depends on the appropriate representation of human face features and the good design of classifier [7].

In this paper, we first make image preprocessing to eliminate the interference of noise and redundant information, reduce the effects of environmental factors on images and highlight the important information of images. At the same time, in order to compensate the deficiency of geometric features, it is well known that the original face images often need to be well represented instead of being input into the classifier directly because of the huge computational cost. So PCA and 2D-PCA are used to extract geometric features from preprocessed images, reduce their dimensionality for computation and attain a higher level of separability. At last, we propose a novel semi-supervised learning algorithm named deep belief network embedded with Softmax regress (DBNESR) as classifier for FR, design many kinds of classifiers based on supervised learning and make experiments to validate the effectiveness of the algorithm.

2 Images preprocessing

Images often appear the phenomenon such as low contrast, being not clear and so on in the process of generation, acquisition, input, etc. of images due to the influence of environmental factors such as the imaging system, noise and light conditions so on. Therefore it needs to make images preprocessing. The process of images preprocessing is as following.

2.1 Face images filtering

We use median filtering to make smoothing denoising for images. Median filter is a kind of nonlinear operation, as shown in Eq.1.

$$f'(i,j) = \underset{s}{Med}\{f(i,j)\} \tag{1}$$

Where, s is the filter window.

2.2 Histogram equalization and dimensionality-reduced

The purpose of histogram equalization is to make images enhancement, improve the visual effect of images, make redundant information of images after preprocessing less and highlight some important information of images.

It is well known that the original face images often need to be well represented instead of being input into the classifier directly because of the huge computational cost. As one of the popular representations, geometric features are often extracted to attain a higher level of separability. Here we employ multi-scale two-dimensional wavelet transform to generate the initial geometric features for representing face images.

2.3 Feature extraction

2.3.1 Extract features with PCA

Suppose that there are N facial images $\{X_i\}_{i=1}^N$, X_i is column vector of M dimension. Calculate the average face of all sample images as following:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^N X_i \tag{2}$$

Calculate the difference of faces, namely the difference of each face with the average face as following:

$$d_i = X_i - \overline{X}, i = 1, 2, \cdots, N \tag{3}$$

Therefore, the images covariance matrix C can be represented as following:

$$C = \frac{1}{N}\sum_{i=1}^N d_i d_i^T = \frac{1}{N}AA^T$$
$$A = (d_1, d_2, \cdots, d_N) \tag{4}$$

Through Eq.5 the eigenvalues of covariance matrix C can be calculated.

$$u_i = \frac{1}{\sqrt{\lambda_i}} Av_i, (i = 1, 2, \cdots, N) \tag{5}$$

v_i is the orthogonal normalization eigenvector of $A^T A$. Through the formula as following:

$$t = \min_k \left\{ \frac{\sum_{j=1}^k u_j}{\sum_{j=1}^N u_j} > \alpha, k \le t \right\} \tag{6}$$

Where, usually set $a = 90\%$, can get the eigenvalues face subspace $U = (u_1, u_2, \cdots, u_t)$.

2.3.2 Extract features with 2D-PCA

As opposed to PCA, 2D-PCA is based on 2D image matrices rather than 1D vectors so the image

matrix does not need to be transformed into a vector prior to feature extraction. Instead, an image covariance matrix is constructed directly using the original image matrices and its eigenvectors are derived for image feature extraction.

3 Designing the classifiers of supervised learning

Usually the classifiers based on supervised learning are often used for FR, in the paper we design two types of classifiers. One is the type of supervised learning classifiers and the other is semi-supervised learning classifiers [8].

3.1 Single BP neural network and hybrid BP neural networks (HBPNNs)

The BP neural network is a kind of multilayer feed-forward network according to the back-propagation algorithm for errors, is currently one of the most widely used neural network models [9]. The recognition and classification of the face images is an important application for the BP neural network in the field of pattern recognition and classification.

When the number scale of human face images isn't big, generalization ability and operation time of the single BP neural network are ideal, and with the increase of numbers of identification species, the structure of BP network will become more complicated, which causes the time of network training to become longer, slower convergence rate, easy to fall into local minimum and poorer generalization ability and so on. In order to eliminate these problems we design the hybrid BP neural networks (HBPNNs) composed of multiple simple BP networks to replace the complex single BP network for FR.

3.2 Single RBF neural network and hybrid RBF neural networks (HRBFNNs)

Radial Basis Function (RBF) simulates the structure of neural network of the adjustment and covering each other of receiving domain of human brain, can approximate any continuous function with arbitrary precision. With the characteristics of fast learning, won't get into local minimum.

Similar to the HBPNNs we design the hybrid RBF neural networks (HRBFNNs) composed of multiple simple RBF networks to replace the complex single RBF network for FR.

3.3 Support Vector Machine (SVM)

SVM is a novel machine learning technique based on a statistical learning theory that aims at finding optimal hyper-planes among different classes of input data or training data in high dimensional feature space. Therefore, the original problem becomes linearly separable.

3.4 Multiple classification decision fusion classifier (MCDFC)——hybrid HBPNNs-HRBFNNs- SVM classifier

The different classifiers have different performance. Fusion of multiple classifiers integrating their respective characteristics can make the classification effect and robustness further improvement. We use the weighted voting for decision fusion of each classifier.

4 Designing the classifier of semi-supervised learning

Recently, semi-supervised learning, which uses a large amount of unlabeled data together with labeled data to build better learners, has attracted more and more attention in pattern recognition and classification [10]. In the paper we design a novel classifier of semi-supervised learning——deep belief network embedded with Softmax regress (DBNESR) for FR. DBNESR first learns hierarchical representations of feature by deep learning [11] and then makes more efficient classification with Softmax regress.

4.1 Softmax regression

Softmax regression is a generalization of the logistic regression in many classification problems [12]. The hypothesis function is as following:

$$h_\phi(x^{(i)}) = \begin{bmatrix} p(Y^{(i)}=1 \mid x^{(i)}; \phi) \\ p(Y^{(i)}=2 \mid x^{(i)}; \phi) \\ \vdots \\ p(Y^{(i)}=k \mid x^{(i)}; \phi) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\phi_j^T X^{(i)}}} \begin{bmatrix} e^{\phi_1^T X^{(i)}} \\ e^{\phi_2^T X^{(i)}} \\ \vdots \\ e^{\phi_k^T X^{(i)}} \end{bmatrix} \tag{7}$$

Where, $\phi_1, \phi_2, \cdots, \phi_k \in \Box^{D+1}$ denote model parameters vector, the cost function is as following:

$$J(\phi) = -\frac{1}{L}\left[\sum_{i=1}^{L}\sum_{j=1}^{k} 1\{Y^{(i)} = j\}\log\frac{e^{\phi_j^T X^{(i)}}}{\sum_{l=1}^{k} e^{\phi_l^T X^{(i)}}}\right] \tag{8}$$

4.2 Deep belief network embedded with Softmax regress (DBNESR)

DBN uses Restricted Boltzmann Machine (RBM) [13-14] of unsupervised learning networks as building blocks for the multi-layer learning systems and uses a supervised learning algorithm named BP (back-propagation) for fine-tuning after pre-training. Its architecture is shown in Fig.1.
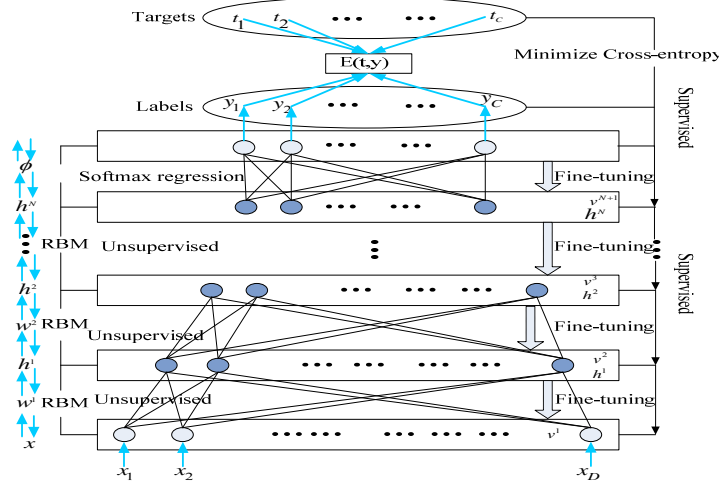

Fig.1 Architecture of deep belief network embedded with Softmax regress (DBNESR)

For unsupervised learning, we define the energy of the joint configuration (h^{k-1}, h^k) as [15]:

$$E(h^{k-1}, h^k; \theta)$$
$$= -\sum_{i=1}^{D_{k-1}}\sum_{j=1}^{D_k} w_{ij}^k h_i^{k-1} h_j^k - \sum_{i=1}^{D_{k-1}} b_i^{k-1} h_i^{k-1} - \sum_{j=1}^{D_k} c_j^k h_j^k \tag{9}$$

For supervised learning, the DBM architecture is trained by C labeled data. The optimization problem is formulized as:

$$\arg\min err = -\sum_k p_k \log \hat{p}_k - \sum_k (1 - p_k)\log(1 - \hat{p}_k) \tag{10}$$

5 Experiment results and analysis

5.1 Experimental environment

All the experiments are carried out in MATLAB R2010b environment running on a desktop with Intel® Core™2 Duo CPU T6670 @2.20GHz and 4.00 GB RAM. Face Recognition Databases are ORL Face Database.

5.2 Image processing

In the experiment we first make images preprocessing by median filtering、histogram equalization and multi-scale two-dimensional wavelet transform so on. After images preprocessing we extract features respectively with PCA and 2D-PCA and compare their effects as following:

(a) Original image

t=10 t=20 t=30 t=40 t=50 t=10 t=20 t=30 t=40 t=50

(b) PCA principal component reconstruction images　　　(c) 2D-PCA principal component reconstruction images
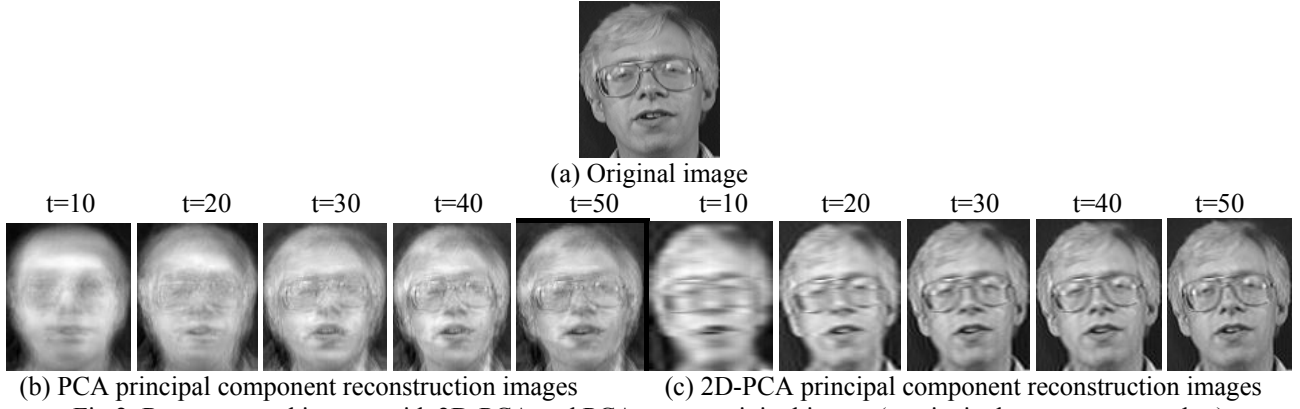
Fig.2　Reconstructed images with 2D-PCA and PCA versus original image (t: principal component number)

We can see that the images of reconstruction by 2D-PCA are clearer than the images of reconstruction by PCA when extracting same number of principal components.

5.3 Recognition effects based on different hidden layers and different neurons of DBNESR

In the experiment we set up different hidden layers and each hidden layer with different neurons for DBNESR. DBNESR structures and learning epochs used are separately shown in Table 1.

TABLE 1 Different hidden layers of DBNESR and learning epochs used in this experiment

Serial number	DBNESR structures	Unsupervised learning epochs	Supervised learning epochs
1	400-200-100-50-20-40	10	1000
2	400-200-300-100-50-40	100	20
3	400-200-300-100-40	50	50
4	400-200-300-40	200	100
5	500-1000-40	200	200
6	1000-500-40	200	200

Almost all the recognition rates of these DBNESR structures are more than 90%, in particular the effects of 500-1000-40 and 1000-500-40 are best and most stable. Therefore, the DBNESR structures used in this experiment are 1000-500-40.

5.4 Recognition rates of different recognition methods

In this experiment, in order to validate the performance of our proposed algorithm——DBNESR is optimal for FR, we compare our proposed algorithm with some other methods such as BP, HBPNNs, RBF, HRBFNNs, SVM and MCDFC. The experiment is repeated for 20 times and the experimental results are shown in Fig. 3.
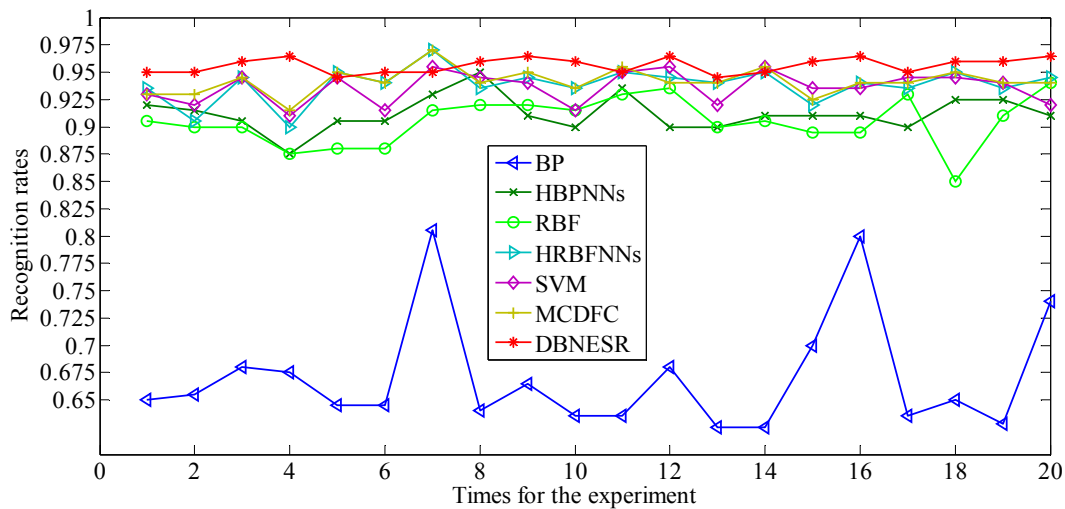

Fig.3 The curve figures of recognition rates of different recognition methods

6 Conclusions

In this paper, we present a novel semi-supervised learning algorithm named deep belief network embedded with Softmax regress (DBNESR). DBNESR first learns hierarchical representations of feature by deep learning and then makes more efficient classification with Softmax regress. And we

also design many kinds of classifiers based on supervised learning: BP, HBPNNs, RBF, HRBFNNs, SVM and multiple classification decision fusion classifier (MCDFC) －－ hybrid HBPNNs-HRBFNNs-SVM classifier. The conducted experiments validate: Firstly, the proposed semi-supervised deep learning algorithm DBNESR is optimal for face recognition with the highest and most stable recognition rates; Second, the semi-supervised learning algorithm has better effect than all supervised learning algorithms; Third, hybrid neural networks has better effect than single neural network.

Acknowledgements

References

[1] J. Wright, Y. Ma, J. Mairal, et al., Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.

[2] S.J. Wang, J. Yang, M.F. Sun, et al., Sparse tensor discriminant color space for face verification, IEEE Trans. Neural Netw. Learn. Syst. 23 (6) (2012) 876–888.

[3] Z. Fan, Y. Xu, W. Zuo, D. Zhang, et al., Modified principal component analysis: an integration of multiple similarity subspace models, IEEE Trans. Neural Netw. Learn. Syst. (2014), http://dx.doi.org/10.1109/TNNLS.2013.2294492.

[4] Y. Xu, X. Li, J. Yang, et al., Integrating conventional and inverse representation for face recognition, IEEE Trans. Cybern. (2013), http://dx.doi.org/10.1109/TCYB.2013.2293391.

[5] H. Zhang, Z. Zhang, Z. Li, et al., Improving representation based classification for robust face recognition, J. Mod. Opt. (2014), http://dx.doi.org/10.1080/09500340.2014.915064.

[6] S.J. Wang, H.L. Chen, et al., Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine, Neural Process. Lett. 39 (1) (2014) 25–43.

[7] Wanggen Wan, Zhenghua Zhou, Jianwei Zhao, Feilong Cao, A novel face recognition method: Using random weight networks and quasi-singular value decomposition, Neurocomputing 151 (2015) 1180–1186.

[8] Martin Längkvist, Lars Karlsson, Amy Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recognition Letters, Volume 42 (1 June 2014) 11-24.

[9] Yingjie Xu, Tao You, Chenglie Du, An integrated micromechanical model and BP neural network for predicting elastic modulus of 3-D multi-phase and multi-layer braided composite, Composite Structures, Volume 122 (April 2015) 308-315.

[10] X. Zhu, Semi-supervised Learning Literature Survey, Technical Report, University of Wisconsin Madison, Madison, WI, USA, (2007).

[11] Jürgen Schmidhuber, Deep Learning in neural networks: An overview Neural Networks, 61 (2015) 85–117.

[12] Wenping Hu, Yao Qian, Frank K. Soong, Yong Wang, Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers, Speech Communication, Volume 67 (March 2015) 154-166.

[13] Asja Fischera, Christian Igelb, Training restricted Boltzmann machines: An introduction, Pattern Recognition, Volume 47, Issue 1 (January 2014) 25–39.

[14] Noel Lopes, Bernardete Ribeiro, Towards adaptive learning with improved convergence of deep belief wetworks on graphics processing units, Pattern Recognition, Volume 47, Issue 1 (January 2014) 114-127.

[15] G.E. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.