

Mixture of Experts for Persian Handwritten Word Recognition

R. Ebrahimpour*, S. Sarhangi** and F. Sharifzadeh***

Abstract: This paper presents the results of Persian handwritten word recognition based on Mixture of Experts technique. In the basic form of ME the problem space is automatically divided into several subspaces for the experts, and the outputs of experts are combined by a gating network. In our proposed model, we used Mixture of Experts Multi Layered Perceptrons with Momentum term in the classification Phase. Applying this term makes three effects in our system: a) increase convergence rate, b) obtain the optimum performance in our system, c) and escape from the local minima on the error surface. We produce three different Mixture of Experts structure. Experimental result for proposed method show an error rate reduction of 6.42 % compare to the mixture of MLPs experts. Comparison with some of the most related methods indicates that the proposed model yields excellent recognition rate in handwritten word recognition.

Keywords: Mixture of Experts, Persian Handwritten Word Recognition, Neural Network Ensembles, OCR.

1 Introduction

Work on neural networks has been motivated right from its inception by the recognition that human brain computes in a different way from the conventional digital computers. Neural networks represent a multi-disciplinary subject with roots in the neurosciences, mathematics, statistics, physics, computer science, and engineering. They are powerful data modeling tools, which can capture and represent complex input/output relationships [1-3].

Developing robust optical character recognition (OCR) techniques would be very rewarding in today technology. Some of the successful applications are: mail sorting, form data entry, bank checking processing, etc. A lot of researches in this area are attended as an interesting field and it has tangible progress in the recognition systems. But these advances are limited to English, Chinese and Indian languages [4-5, 9].

There has been very limited reported research on the scripts of Persian languages, although it is usual language in Middle East. Recently different feature extraction and classification methods applied in the Persian handwritten recognition systems [6-8].

The process of handwritten word recognition of any script can be broadly broken down into four stages:

1. Pre-Processing
2. Segmentation
3. Feature extraction
4. Classification

Skewness and skeletonizing of the image in preprocessing stage simplify the processing of other stages. The segmentation stage takes in a page image and separates the different logical parts, like text from graphics, lines of a paragraph, and characters of a word. The feature extraction stage analyzes a text segment and selects a set of features that can be used to uniquely identify the text segment. The selection of a stable and representative set of features is the heart of pattern recognition system design. Among the different design issues involved in building an OCR system, perhaps the most consequential one is the selection of the type and set of features. The classification stage is the main decision making stage of an OCR system and uses the features extracted in the previous stage to identify the word according to preset rules.

However handwritten word recognition problems are often too complicated for a single classifier to solve it. Thus, the committee machine created, that it is

Iranian Journal of Electrical & Electronic Engineering, 2011.

Paper first received 20 Oct. 2010 and in revised form 29 Aug. 2011.

* The Author is with Brain and Intelligent Systems Lab., Department of Electrical and Computer Engineering, Shahid Rajaei Teacher Training University Lavizan, Tehran, Iran.

E-mail: ebrahimpour@ipm.ir

** The Author is with the Department of Mechatronics, Islamic Azad University south Tehran Branch, Tehran, Iran.

E-mail: somavehsarhangi@yahoo.com

*** The Author is with the Department of Computer Science, Shahid Bahonar University of Kerman, Iran.

Email: fSharifzade@mail.uk.ac.ir

proved which it has more accurate results than a single predictor. In other words, to improve the performance in classification particularly for complex problems such as those involving limited number of patterns, high-dimensional feature sets, and highly overlapped classes it is needed to use the ensemble methods [16].

S. Günter and H. Bunke in [10] the influence of the vocabulary size and the number of training samples on the performance of three ensemble methods in the context of handwritten word recognition is examined. C. C. Tappert et al. had a survey which describes the state of the art of online handwriting recognition during a period of renewed activity in the field [11]. S. Madhvanath and V. Govindaraju attempted to take a fresh look at the potential role of the holistic paradigm in handwritten word recognition [12]. In [13] the authors address some issues relevant to the design of serial classifier combinations.

Now research on Farsi (Persian) scripts is receiving increasing attention due to the increasing interest in automatic processing of handwritten data. R. Safabakhsh and P. Adibi introduces a complete system for recognition of Farsi Nastaaligh handwritten words using a continuous-density variable-duration hidden Markov model [14]. H. Soltanzadeh, M. Rahmati applied the support vector machines in the classification stage of their work [15].

Generally there are two main strategies in combining classifiers: fusion and selection. In classifier fusion, it is supposed that each ensemble member is trained on the whole feature space, whereas in classifier selection, each member is assigned to learn a part of the feature space. This way, in the former strategy, the final decision is made by considering the decisions of all members, while in the latter strategy, the final decision is made by aggregating the decisions of one or a few of experts [17, 18]. Combining classifiers based on the fusion of outputs of a set of different classifiers have been developed as a method for improving the recognition performance [19, 20]. Classifier fusion is categorized into two classes, trainable and non trainable. Trainable fusion methods such as decision templates, stack generalization are better than Non trainable fusion method like as product, average. R. Ebrahimpour and F. Sharifzadeh, used static methods to Persian handwritten digit recognition [21]. Classifier selection has not attracted as much attention as classifier fusion. However, classifier selection is probably the better of the two strategies, if trained well [17].

In this paper, we propose a new neural computational model for Persian handwritten word recognition which is essentially based on the so-called mixture of experts (ME) architecture, that originally proposed by Jacobs et al. [22] and falls under the category of classifier selection.

Essentially, the idea behind combining classifiers in ME is based on the so-called divide-and-conquer principle, according to which a complex computational

task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks [16]. In ME, the problem space is divided into several subspaces for the experts, and then the outputs of experts are combined by a gating network, that is gating network implements competitive selection between a numbers of simple homogeneous modules (experts).

Since the original paper on ME [22], a huge number of variants of this paradigm have been developed. In the conventional ME [16], the expert and gating networks were linear classifiers, however, for more complex classification tasks, the expert and gating networks could be of more complicated types. For instance, Ebrahimpour et al. [23] proposes a face detection model, in which they use MLPs in forming the gating and expert networks to improve the face detection accuracy. In [24], authors use MLPs with one hidden layer as experts and an RBF network as the gating network, for designing compensators for intensity modulated radiation therapy. In [25] the idea was extended with a hierarchical mixture models.

Expectation–Maximization (EM) [26] algorithm has been introduced to the ME architecture in order that the learning process is separated in a way that fits well with the modular structure [25- 28]. Since the EM algorithm learns only the cluster centroids and not intermediary points, it will not work well on non linear examples.

In this study we use a momentum term in training of MLP experts in ME, which with using momentum term in updating weights process we can obtain the optimum performance with more speed. Another benefit of using this term is escaping of local minimum on the error surface.

Hereafter we referred this method as a Mixture of Multi Layer Perceptrons Experts with Momentum constant. This paper is organized as follows. In Section 2 we describe proposed feature extraction method. In Section 3 we provide a brief description of the Mixture of Experts structures with linear and Multilayer Perceptrons, MLP, as experts. Section 4, is devoted to experimental results. Finally Section 5, concludes and summarizes the paper.

2 Feature Extraction Method

In this section we apply a scale invariant gradient based method for Persian handwritten word recognition feature extraction. It should be noted that for this method first, thinning must be applied on the word images. Thinning is the process of reducing thickness of samples to just a single pixel. By the thinning method shape information of patterns preserved and data size reduced [8]. Thinning method removes pixels so that a pattern without holes shrinks to a minimally connected stroke, and a pattern with holes shrinks to a connected ring halfway between each hole and the outer boundary. After applying the thinning method, the word images are decomposed into a number of separate images

corresponding to four 3×3 masks. Indeed with this mask lines, 0° , 45° , 90° and 135° degrees, word images are separated. Fig. 1 shows these masks. For example the first mask used for separating the lines with 0° degree from the word image. These lines can be extracted as explained below. Instead of input word image considered as a zero matrix with the size of original image. This mask moved from left to right over the image and specifies the new values of elements of this matrix. The assigned value of this matrix depends on the value of the pixels in original word image. This means that if the value of the pixel (i,j) and two neighbors in right and left, $(i,j-1)$ and $(i,j+1)$, were 1 in the original image, then in the defined matrix we assign the value 1 in the position of (i,j) . This procedure repeat for other three masks and each word image decompose into four separate images corresponding to these masks.

In the next stage, each of separated images uniformly partition into 8 sectors around the center of image. The number of black pixels in each sector is calculated. The value of each sector is normalized by dividing it upon the total number of black pixels in word image and it is used for feature value of that sector. Hence for each of these separated images we have eight feature values leading to a feature vector of 32 elements. Thus the needed feature vector is extracted for each input image. Fig. 2 displays the stages of this method.

3 Mixture of Experts

From a computational aspect, according to the principle of divide and conquer, a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces. The combination of experts is lead to make up a combination of classifiers.

Mixture of experts is one of the famous methods in the category of dynamic structures of classifier combining, in which the input signal is directly involved in actuating the mechanism that integrates the outputs of the individual experts into an overall output [17]. Consider a modular neural network (Fig. 3) in which the learning process proceeds by fusing self organized and supervised forms of learning. The experts are technically performing supervised learning in that their individual outputs are combined to model the desired response. There is, however, a sense in which the experts are also performing self-organized learning; that is they self-organize to find a good partitioning of the input space so that each expert does well at modeling its own subspace, and as a whole group they model the input space well. The learning algorithm of the mixture structure is described in [22].

However, in our models, in order to improve the performance of the expert networks, and consequently

the handwritten word recognition accuracy, we use our revised version of ME in which MLPs instead of linear networks or experts are used [23, 29].

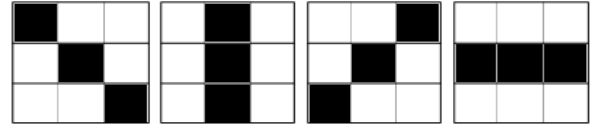


Fig. 1 The masks that used for decomposing word images into a number of separate images.

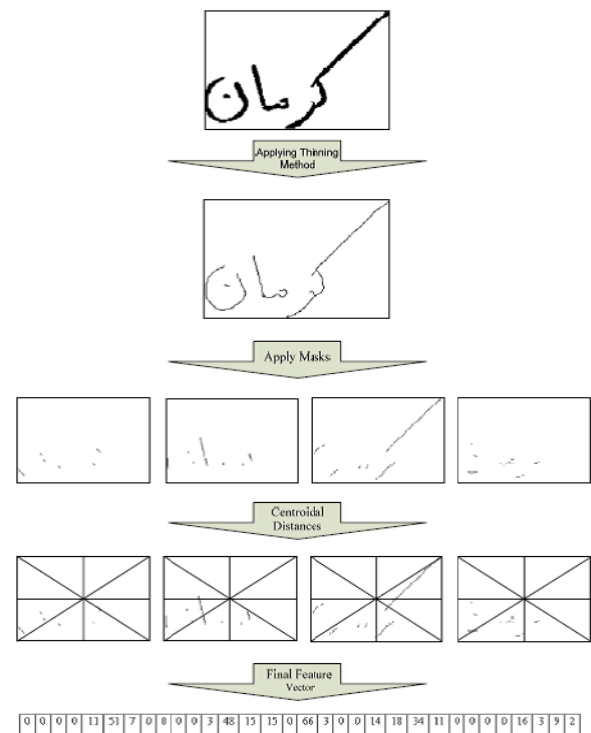


Fig. 2 The stages of the applied feature extraction method for Persian handwritten word recognition.

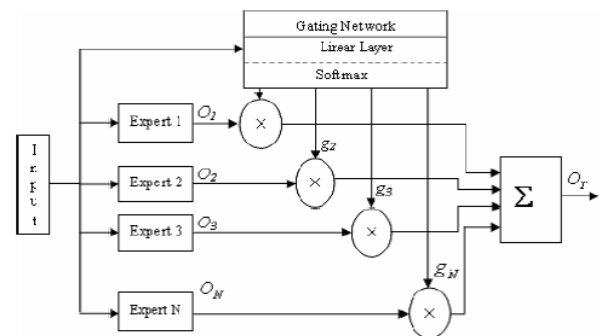


Fig. 3 The mixture of experts is composed of expert networks and a gating network. The experts compete to learn the training patterns and the gating network mediates the competition. The gating network is simultaneously trained to combine the experts' outputs.

4 Proposed Classification Method

Our proposed model is calculated to achieve robust handwritten recognition with the mentioned feature extraction, Section 2, and a mixture of MLP experts with Momentum term in the classification stage.

Each expert is a one-hidden-layer MLP, with Momentum term, that computes an output vector O_i as a function of the input vector x and a sigmoid function as the activation function and a set of parameters such as weights of hidden and output layer. It is considered that each expert specializes in a different area of the input space. The g_i calculated by gating network to each of the experts' outputs, O_i . The gating network determines the g_i as a function of the input vector x and a set of parameters such as weights of the hidden layer, the output layer and a sigmoid function as the activation function. The g_i can be used in estimating of the prior probability that expert i can generate the desired output y . The gating network is composed of two layers: the first layer is an MLP network, and the second layer is a softmax nonlinear operator as the gating network's output. The gating network computes O_g , which is the output of the MLP layer of the gating network, then applies softmax function to get:

$$g_i = \frac{\exp(O_{gi})}{\sum_{j=1}^N \exp(O_{gj})}, \quad i = 1, 2, 3 \quad (1)$$

So the g_i is nonnegative and sum to 1. The final mixed output of the entire network is:

$$O_T = \sum_i O_i g_i, \quad i = 1, 2, 3 \quad (2)$$

The "normalized" exponential transformation of Eq. (1) may be viewed as a multi-input generalization of the logistic function. It keeps the rank order of its input values, and is a differentiable generalization of the "winner-takes-all" operation of picking the maximum value, so referred to as softmax. The weights of MLPs are learned using the back-propagation, BP, algorithm, in order to maximize the log likelihood of the training data given the parameters. Assuming that the probability density associated with each expert is Gaussian with identity covariance matrix, MLPs obtain the following online learning rules:

$$\Delta W_y = \eta_e h_i (y - O_i) (O_i (1 - O_i)) O h_i^T \quad (3)$$

Our method of improving the learning rate, and also avoiding the peril of instability, is to improve the delta rule of Eq. (3) by including a Momentum term:

$$\Delta W(n)_y = \alpha \Delta W(n-1)_y + \eta_e h(n)_i (y - O(n)_i) (O(n)_i (1 - O(n)_i)) O h(n)_i^T \quad (4)$$

$$\Delta W(n)_y = \alpha \Delta W(n-1)_h + \eta_e h(n)_i W(n)_y^T (y - O(n)_i) (O(n)_i (1 - O(n)_i)) o h(n)_i \quad (5)$$

$$\Delta W(n)_{yg} = \alpha \Delta W(n-1)_{yg} + \eta_g (h(n) - g(n)) (O(n)_g (1 - O(n)_g)) O h(n)_g^T \quad (6)$$

$$\Delta W(n)_{hg} = \alpha \Delta W(n-1)_{yg} + \eta_g W(n)_{yg}^T (h(n) - g(n)) O(n)_g (1 - O(n)_g) o h(n)_g (1 - o h(n)_g) x_i \quad (7)$$

The inclusion of Momentum term in the back-propagation algorithm tends to increase speed of descent in steady downhill directions and has a stabilizing effect in directions that oscillate in sign. The incorporation of Momentum term in the back-propagation algorithm represents a minor modification to the weight update process, yet it may have some useful effects on the learning behavior of the algorithm. The Momentum term may also have the benefit of preventing the learning process from terminating in a shallow local minimum on the error surface. Where η_e and η_g are learning rates for the experts and the gating network, respectively, $O h_i$ is the output of expert network's hidden layer, and h_i is an estimate of the posterior probability that expert i can generate the desired output y :

$$h_i = \frac{g_i \exp(-\frac{1}{2}(y - O_i)^T (y - O_i))}{\sum_j g_j \exp(-\frac{1}{2}(y - O_j)^T (y - O_j))} \quad (8)$$

This can be reason of as a softmax function computed on the inverse of the sum squared error of each expert's output, smoothed by the gating network's current estimate of the prior probability that the input pattern was drawn from expert i 's area of specialization. As the network's learning process progresses, the expert networks "compete" for each input pattern, while the gating network rewards the winner of each competition with stronger error feedback signals. Thus, over time, the gate partitions the handwritten word space in response to the expert's performance.

5 Experimental Results

To implement the proposed model, the Iranshahr dataset is used. Iranshahr consists of 780 samples of 30 city names of Iran. There are 26 handwritten samples for each city name, which have written by 26 different persons. In these experiments, we have selected 20 samples for training and 6 samples for testing. So, used training set consists of 600 samples and used testing set includes 180 samples. All used samples have scanned at 96 dpi resolution in the gray scale format, before using in feature extraction stage. Next, all of them converted to the binary format with a constant threshold value. Afterward, images centralized in a 184×320 pixels frame. Some sample images are shown in Fig. 4.

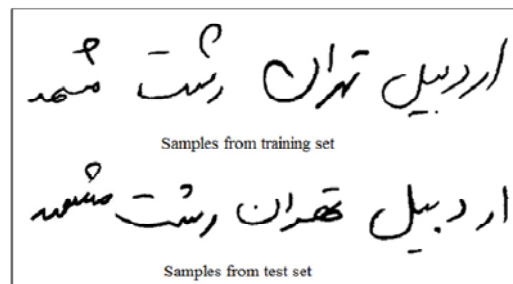


Fig. 4 Samples of city names of Iran from training and test set.

Selection is one of the main stages in any type of feature extraction technique. Selection is probably the most important stage to transform input space into the feature space, which must have a good discriminating power in classification. In Section 2 we proposed a feature extraction method for our experiment. As described in the next sections, image we extract a 32 dimensional feature vector from each image sample. To evaluate the performance of the ME method and also to exhibit the advantages of using it in recognition of handwritten Persian words, we performed several experiments. First, we compared it with single MLP and Linear ME method. Second, we implemented an experiment to use MLP as an experts in ME structure, and then, we specified the best structure of ME. Finally, we remarked the role of momentum in the learning algorithm to increasing the training speed. In the proposed approach, the connective weights of single MLP classifier are estimated using the back-propagation algorithm. In addition, to set the learning parameters of MLP, four fold cross validation are used. Also, the learning rate of MLP was 0.1. In MLP structure; we utilized different numbers of neurons in the hidden layer. As it observed, with more than 45 hidden neurons the results has a few variance in recognition rate.

In attention to the results of using single MLP, it must be attended that, because of the complexity of Persian handwritten word recognition system, used combining methods can demonstrated more accurate results with regards to single MLP. We know that, the ME is the one of the most famous combination algorithms. So, we conducted our experiments based on ME method. In the conventional ME used linear networks as experts and gating network, but as mentioned above, Persian handwritten recognition, is a fully complicated problem. Therefore, we must use stronger experts in ME structure. Thus, in this work MLPs are used in the architecture of ME.

We performed our experiments using MLP, Mixture of Linear Experts, Mixture of MLP Experts, and Mixture of Momentum MLP Experts respectively. Each experiment was repeated 10 times, using randomly picked training and testing sets form Iranshahr dataset. By setting the number of proposed method components into 32 components, we evaluated the performance of Mixture of Momentum MLP Experts, and compared it with other networks. To perform a reasonable evaluation, we compared our method with a single MLPs (Single MLPs has one hidden layer, which is trained using the back-propagation algorithm), with the mixture model, which consisted of three simple MLPs. Experimental results support our claim that three simple MLPs in the mixture architecture performs the recognition task much better than a single MLPs. To conduct a fair comparison, all networks have trained and tested on the same training and testing sets. By considering the results, it is evident that the performance of the Mixture of MLPs Experts with Momentum network is superior to that of others. Table 1 lists the training parameters in detail. The best structure, in the terms of recognition rate, for the experts and gating network is shown in this table, too.

To reach the optimal number of neurons in hidden layer of experts, we evaluated this task with different number of neurons. As shown in Table 2, experts with 17 neurons in their hidden layer reveal the best performance.

To gain the optimal number of epochs to achieve the best recognition rate in both Mixture of MLP Experts with Momentum and the conventional Mixture of Experts, we repeated the same experiment with different epochs in training networks and considered their performance. As shown in Fig. 5, Mixture of MLP Experts with Momentum needs less epochs to gain its best result.

Table 1 The details of the training parameters in as well as the recognition rates of the MLP, Mixture of Linear Experts, Mixture of MLP Experts and Mixture of MLP Experts with Momentum on the test set.

	Multi Layer Perceptrons	Mixture of Linear Experts	Mixture of MLP Experts	Mixture of MLP Experts with Momentum
Topology	32:45:30	3 linear Experts	Experts: 32:17:30 Gating: 32:9:30	Experts: 32:17:30 Gating: 32:9:30
Learning rate	0.1	Experts:0.19	Experts:0.19 Gating:0.09	Experts: 0.19 Gating: 0.09 Momentum: 0.6
Max Percentage	86.34	88.33	90.5	91.11

Table 2 Recognition rates of different topologies of the Mixture of Experts.

No. of hidden neurons of gating network	710	814	917	1219	1621
No. of hidden neurons of experts					
Recognition Rate (%) Mixture of linear Experts	82.11	86.23	88.33	89.1	87.23
Mixture of MLP Experts	82.95	87.58	90.5	88.51	87.5
Mixture of MLP Experts with Momentum	83.32	87.95	91.11	89.3	88.33

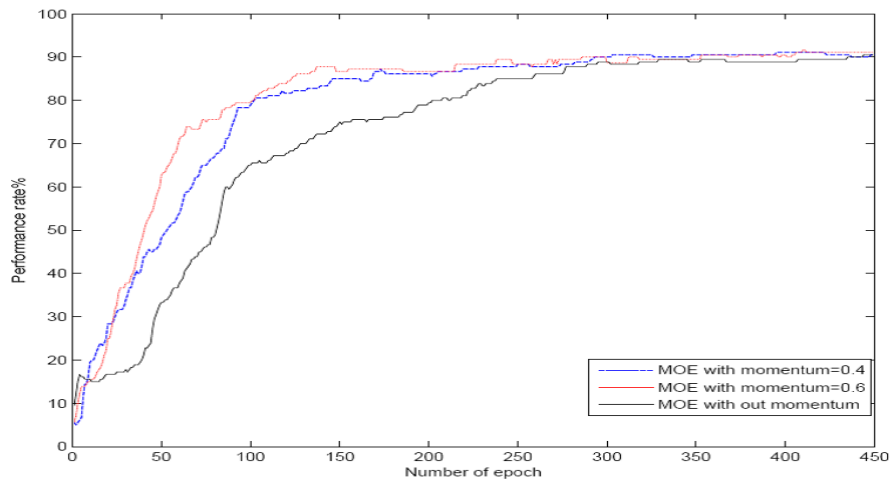


Fig. 5 Performance rate versus number of epoch for Mixture of MLP Experts with and without Momentum.

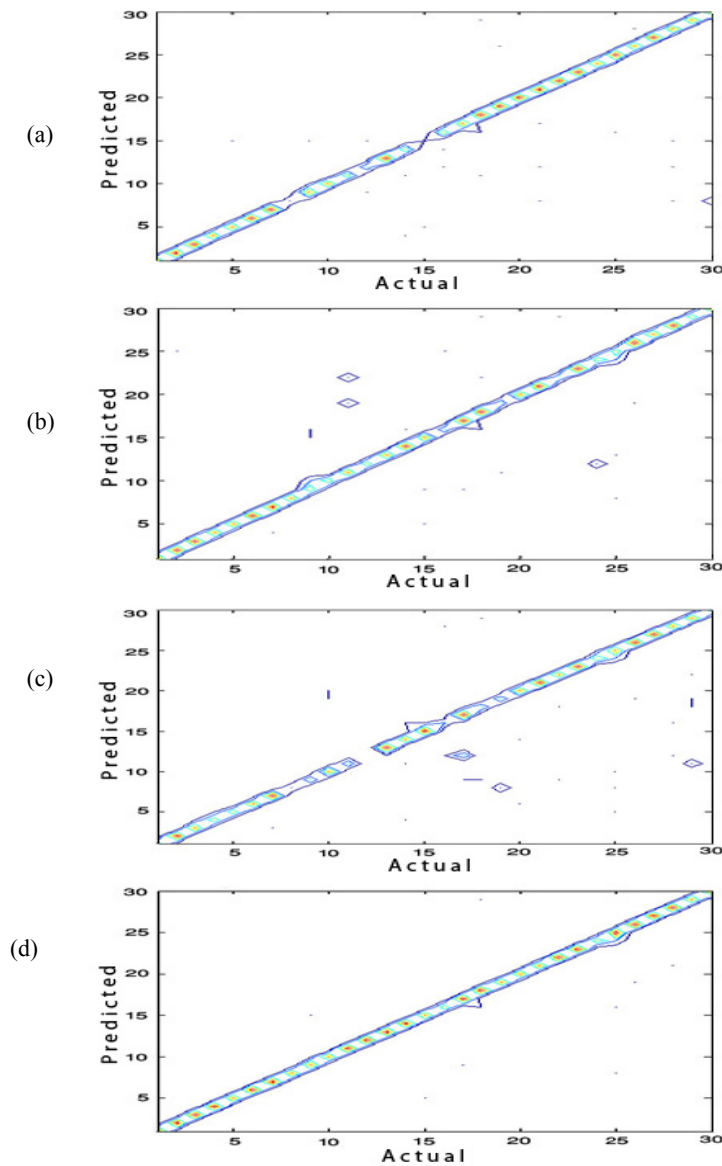


Fig. 6 (a-c) Confusion matrix of each expert in ME, performing handwritten word recognition on the test set. (d) Confusion matrix of the ME model. It is clearly shown that combining the output of experts considerably improves the recognition rate.

In other words, it is faster in comparison with the same network which is trained without Momentum term. An important issue in combining classifiers is the diversity of classifiers in learning the input data. When base classifiers of an ensemble are diverse, i.e. they learn different areas of the input space; they obtain specialization on a specific area of the input space and consequently have fewer errors on those areas.

This way, combining the outputs of classifiers that are perfectly divers, improves the overall performance of the ensemble. To have diverse classifiers, the training procedure can be affected by input representation of patterns, training samples, learning procedure and supervision strategy, on which correlation reduction techniques is based. To see how the errors are distributed over classes, which shows the level of classifiers diversity, we form a confusion matrix using the testing data set. A confusion matrix [30] contains information about actual and predicted classifications done by a classification system.

Performance of such systems is commonly evaluated using the data in the matrix. Fig. 6 illustrates a graphic representation of the confusion matrices of each expert of a ME, along with the overall performance of the whole ME network. Note that off-diagonal marks on each figure show misclassified images; therefore a denser left diagonal represents less mistakes. Considering Fig. 6(d), the role of combining experts in a ME framework is demonstrated by improved recognition rate, which is the result of combining the correct recognition of each expert.

6 Conclusion

This paper presented the use of a modified ME structure to improve handwritten word recognition. Our ME is composed of three modified local experts and gating network, which were MLPs with a Momentum term used in the training process. Our proposed ME was trained, cross validated and tested on the Iranshahr dataset. The recognition rates achieved by the modified ME turned out to be higher than that of a single MLPs and conventional ME models with linear networks as their experts and gating networks, and also those with MLPs experts trained without the momentum term.

References

- [1] Ghods L. and Kalantar M., "Long-Term Peak Demand Forecasting by Using Radial Basis Function Neural Networks", *Iranian Journal of Electrical and Electronic Engineering(IJEEE)*, Vol. 6, No. 3, pp. 175-182, 2010.
- [2] Hariri M., Shokouhi S. B. and Mozayani N., "An Improved Fuzzy Neural Network for Solving Uncertainty in Pattern Classification and Identification", *Iranian Journal of Electrical and Electronic Engineering(IJEEE)*, Vol. 4, No. 3, pp. 71-78, 2008.
- [3] Mosavi M. R., "Infrared Counter-Countermeasure Efficient Techniques using Neural Network, Fuzzy System and Kalman Filter", *Iranian Journal of Electrical and Electronic Engineering*, Vol. 5, No. 4, pp. 215-222, 2009.
- [4] Suen C. Y., Nadal C., Mai T. A., Legault R. and Lam L., "Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts", In: *Proc.IWFHR*, pp. 131-143, 1990.
- [5] Kato N., Suzuki M., Omachi S., Aso H. and Nemoto Y., "A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance", *Pattern Analysis and Machine Intelligence*, Vol. 21, No. 3, pp. 258-262, 1999.
- [6] Borji A., Hamidi M. and Mahmoudi F., "Robust Handwritten Character Recognition with Features Inspired by Visual Ventral Stream", *Neural Processing Letters*, Vol. 28, No. 2, pp. 97-111, 2008.
- [7] Suen C. Y., Izadinia S., Sadri J. and Solimanpour F., "Farsi script recognition: a survey", in: *Proceedings of the Summit on Arabic and Chinese Handwriting Recognition*, pp. 101-110, 2006.
- [8] Amin A., "Off-line Arabic character recognition: the state of the art", *Pattern Recognition*, Vol. 31, No. 5, pp. 517-530, 1998.
- [9] Roy V., Madhvanath S., Anand S. and Sharma Raghunath R., "A Framework for Adaptation of the Active-DTW Classifier for Online Handwritten Character Recognition", *International Conference on Document Analysis and Recognition*, Barcelona, August 2009.
- [10] Gunter S. and Bunke H., "Multiple Classifier Systems in Off line Handwritten Word Recognition— on the Influence of Training Set and Vocabulary Size", *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) word scientific journal*, Vol. 18, No. 7, pp. 1303-1320, 2004.
- [11] Tappert C. C., Suen C. Y. and Wakahara T., "The State of the Art in Online Handwriting Recognition", *pattern analysis and machine intelligence*, Vol. 12, No. 8, pp. 787-808, 1990.
- [12] Madhvanath S. and Govindaraju V., "The Role of Holistic Paradigms in Handwritten Word Recognition", *pattern analysis and machine intelligence*, Vol. 23, No. 2, pp. 149-164, 2001.
- [13] Madhvanath S. and Govindaraju V., "Serial classifier combination for handwritten word recognition", *Third International Conference on Document Analysis and Recognition (ICDAR'95)*, Vol. 2, pp. 911, 1995.
- [14] Safabakhsh R. and Adibi P., "Nastaalighhandwritten word recognition using a continuous-density variable-duration HMM", *The*

- Arabian Journal for Science and Engineering*, Vol. 30, No. 1B, pp. 95-118, 2005.
- [15] Soltanzadeh H. and Rahmati M., "Recognition of Persian handwritten digits using image profiles of multiple orientations", *Pattern Recognition Letters*, Vol. 25, No. 14, pp. 1569-1576, 2004.
- [16] Nadal C., Legault R. and Suen C. Y., "Complementary Algorithms for Recognition of totally Unconstrained Handwritten Numerals", *Proc. 10th Int. Conf. Pattern Recognition*, Vol. A, pp. 434-449, 1990.
- [17] Haykin S., *Neural Networks—A Comprehensive Foundation*, second ed., Prentice-Hall, 1998.
- [18] Kuncheva L. I., *Combining Pattern Classifiers: Methods and algorithms*, John Wiley & Sons. Inc. 2004.
- [19] Tax David M. J., van Breukelen M., Duin Robert P. W. and Kittler J., "Combining multiple classifiers by averaging or by multiplying", *Pattern Recognition*, Vol. 33, No. 9, pp. 1475-1485, 2000.
- [20] Liu C. L., "Classifier combination based on confidence transformation", *Pattern Recognition*, Vol. 38, No. 1, pp. 11-28, 2005.
- [21] Ebrahimpour R. and Sharifzadeh F., "Persian Handwritten Digit Recognition with Classifier Fusion: Class Conscious versus Class Indifferent Approaches", *World Academy of Science, Engineering and Technology*, Vol. 57, No. 33, pp. 552-559, 2009.
- [22] Jacobs R. A., Jordan M. I., Nowlan S. J. and Hinton G. E., "Adaptive mixtures of local experts", *Neural Comput.*, Vol. 3, No. 1, pp. 79-87, 1991.
- [23] Ebrahimpour R., Kabir E. and Yousefi M. R., "Face Detection Using Mixture of MLP Experts", *Neural Processing Letters*, Vol. 26, No. 1, pp. 69-82, 2007.
- [24] Goodband J. H., Haasa O. C. L. and Millsb J. A., "A mixture of experts committee machine to design compensators for intensity modulated radiation therapy", *Pattern Recognition*, Vol. 39, No. 9, pp. 1704-1714, 2006.
- [25] Jordan M. I. and Jacobs R. A., "Hierarchical mixture of experts and the EM algorithm", *Neural Comput.*, Vol. 6, No. 2, pp. 181-214, 1994.
- [26] Dempster A. P., Laird N. M. and Rubin D. B., "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society. Series B(Methodological)*, Vol. 39, No. 1, pp.1-38, 1977.
- [27] Hong X. and Harris C. J., "A mixture of experts network structure construction algorithm for modelling and control", *Appl. Intell.*, Vol. 16, No. 1, pp. 59-69, 2002.
- [28] Guler I. and Derya Ubeyli E., "A mixture of experts network structure for modeling Doppler ultrasound blood flow signals", *Computers in Biology and Medicine*, Vol. 37, No. 7, pp. 565-582, 2005.
- [29] Ebrahimpour R. and Kabir E., Esteky H. and Yousefi M. R., "A Mixture of Multilayer Perceptron Experts Network For Modeling Face/Nonface Recognition in Cortical Face Processing Regions", *Intelligent Automation and Soft Computing*, Vol. 14, No. 2, pp. 145-156, 2008.
- [30] Shipp C. A. and Kuncheva L. I., "Relationship between Combination methods and measures of diversity in combining classifiers", *Information Fusion*, Vol. 3, No. 2, pp. 135-148, 2002.



Reza Ebrahimpour is an Assistant Professor at the Department of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran. He is also a research scientist at School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran. He received the B.S. degree in Electronics

Engineering from Mazandaran University, Mazandaran, Iran and the M.S. degree in Biomedical Engineering from Tarbiat Modares University, Tehran, Iran, in 1999 and 2001, respectively. He obtained Ph.D. degree in the field of Computational Neuroscience from the School of Cognitive Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran in July 2007. Dr. Ebrahimpour is the author or co-author of more than 80 international journal and conference publications in his research areas, which include Human and Machine Vision, Computational Neuroscience, Neural Networks, Pattern Recognition and Multiple Classifier Systems.



Somayeh Sarhangi was born in Tehran, Iran in 1983. She achieved the B.Sc. degree in computer engineering from Institute of higher education Sadjad. She has an M. Sc. in Mechatronics from the faculty of engineering, Azad University south Tehran Branch, Iran. Her research interests are machine vision, neural network, Combining Classifiers.



Fatemeh Sharifzadeh was born in Tehran, Iran. She received B.Sc. degrees in Computer Science from the University of Tehran, in 2009. She is now a student of Computer Science in M. Sc. degree. Her interest areas are artificial intelligence, Machine Learning, Intelligent Systems, Pattern Recognition, and Computational Neuroscience.