2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

A Study on Performance Improvement of Fuzzy Min-Max Neural Network Using Gating Network

Byoung-Dong Kwak, Kwang-Hyun Park, Z. Zenn Bien
Dept. of EECS, KAIST
E-mail : goodkwak@ctrsys.kaist.ac.kr, akaii@robotian.net , zbien@ee.kaist.ac.kr

Abstract **- Fuzzy Min-Max Neural Network (FMMNN) is a powerful classifier. It has, however, some problems. Learning result depends on the presentation order of input data and the training parameter that limits the size of hyperbox. The latter problem affects the result seriously. In this paper, the new approach to alleviate that without loss of on-line learning ability is proposed. The committee machine is used to achieve the multi-resolution FMMNN. Each expert is a FMMNN with fixed training parameter. The advantages of small and large training parameters are used at the same time. The parameters are selected by performance and independence measures. The Decision of each expert is guided by the gating network. Therefore the regional and parametric divide and conquer scheme are used. Simulation shows that the proposed method has better classification performance.**

I. INTRODUCTION

Recently, human friendly man-machine interface has been an important issue. The development of technology gave the machine the ability to communicate with human by human friendly ways, those are, voice, gesture, facial expression, and so on. The classifier of the machine is important to achieve a powerful human friendly man-machine interface. This paper deals with the classifier.

Since Fuzzy Min-Max Neural Network (FMMNN) has developed by Simpson [1], many researchers have taken great interest in its usefulness. FMMNN has a very simple structure, and a fast learning speed. Because of those advantages, it can be simply implemented as hardware.

It has, however, some problems in spite of its superiority. Learning result depends on the presentation order of input data and the training parameter that limits the size of hyperbox. The generalization ability is also affected by the training parameter, because it imposes the same constraint on the whole feature space. Especially the latter problem, how to decide the training parameter that limits the size of hyper-box, is very critical.

Many researchers have proposed some modifications to alleviate the problem. Most of them had to store the training data and use it many times and didn't use the contraction process. [2][3][4][5] Most different aspect of proposed method of this paper from previous works is that both single pass on-line training and multi-resolution property are achieved by the use of the advantages that are considered disadvantages of small and large training parameters.

This paper organized as follow. Next section explains FMMNN briefly and the advantages of small and large training parameters. In section 3, the proposed method is explained. In section 4, the simulation result is given. And section 5 gives some comments and conclusion.

II. FMMNN and Characteristics of Training Parameter

FMMNN uses a hyperbox. A hyperbox is a fuzzy set that has full membership value inside the region defined by its min point and max point in the feature space and smaller membership value outside the region. Each hyperbox is associated with a unique class. FMMNN is comprised of three layers. First layer supplies input features to hidden layer. Second layer, hidden layer, is hyperbox layer. Each neuron, hyperbox, of second layer computes the membership value of input pattern. Third layer, output layer, has neurons as much as the number of the output classes. Each neuron of third layer determines membership value of input pattern with respect to the class. Final decision can be crisp as well as fuzzy.

FMMNN learning process consists of expansion and contraction. The learning process begins whenever a training pattern is presented. First, the closest hyperbox of the same class as input pattern is found and expand to include the input. If the hyperbox is not found or does not meet the expansion criteria, a new hyperbox is created and added to the neural network. The expansion criteria determines whether the hyperbox is expanded or created. To do this, the training parameter that limits the maximum size of hyperbox is used. If the distance between min and max point with respect to any dimension of expanded hyperbox to include input is smaller than the training parameter, then the expansion criteria is met [6].

It is the most difficult part in FMMNN learning to determine the training parameter. Small training parameter produces many hyperboxes so that It can cause overfitting, but on the other hand, It can learn training data in detail. Large training parameter produces relatively small hyperboxes so that it can cause underfitting, but on the other hand, it can reject to learn some data.

III. Proposed Committee Machine Scheme

Most of previous works about how to decide the training parameter aimed at the perfect training data learning with the smallest number of hyperboxes. To do this, the training data was used iteratively so that the advantage of original FMMNN, fast single-pass on-line learning ability, was lost. And when the training data is not perfect, for example, the data is acquired in an uncontrolled environment, the perfect learning may not be the best answer for the generalization performance.

To cope with those problems, simultaneous use of the error rejection of the underfitting and the fine learning of the overfitting is proposed. Committee machine with experts that have different training parameter can use those at the same time. Furthermore, This scheme still has not only advantages of original FMMNN, that is, fast and single-pass on-line learning ability that was lost in the previous works, but also multi-resolution training parameter.

The committee machine is comprised of three layers. First layer just supplies input to second layer. Second layer consists of experts which have different training parameter. Third layer is output layer that decides final result. The weights between first and second layers are '1'. The weights between second and third layers are controlled by gating network with respect to the input. the described structure is shown in figure 1.

The learning process consists of expert pool training, expert selection, and gating network training. All experts and gating network are a FMMNN with a single training parameter, so the training of those is same as Simpson's algorithm. Two training data sets are needed. One is for experts training and the other is for expert selection and gating network.

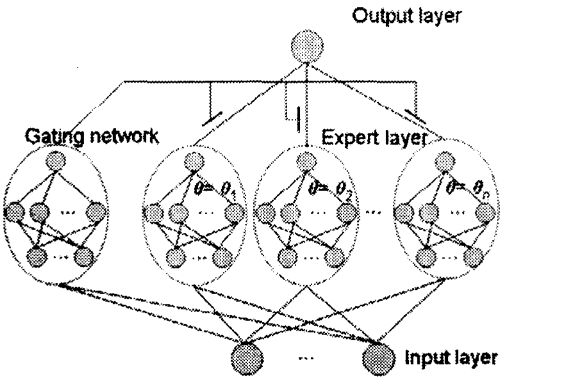

figure 1. The structure of the committee machine

Expert pool is comprised of many FMMNNs. Each FMMNN has a unique training parameter. By training data set I, all experts in the pool are trained. By training data set II, the best performance expert is selected. Then the other experts are selected by the performance of itself and the independence from the best expert. Then the committee machine is constructed with selected experts. The gating network is also trained by training data set II. The output class of the gating network is the ID of expert. Therefore the input of the gating network is features of input data that is labeled as the ID of expert that output the right classification with respect to original class of input pattern.

The committee machine works as follows. When test input is presented to the network, the input layer supplies input to the input layer of each experts and the gating network. Each experts computes membership values of all classes corresponding to input pattern. The gating network computes membership values of all experts. The output of each expert are multiplied by the membership value of the gating network and supplied to the output layer. The output layer of the committee machine sums all input from the second layer as formula (1).

$$\mu_{sum,j}(\mathbf{x}_k) = \sum_{i=1}^{M} \mu_{i,j}(\mathbf{x}_k)\mu_{gate,i}(\mathbf{x}_k) \qquad (1)$$

where i is the ID of the expert, j is output class, subscript *gate* means the membership function of the gating network and M is the number of selected experts Then final decision is the class of maximum final membership value as formula (2).

$$\text{final decision} = \arg\max_{j} \mu_{sum,j}(\mathbf{x}_k) \qquad (2)$$

Figure 2 shows described decision making process.

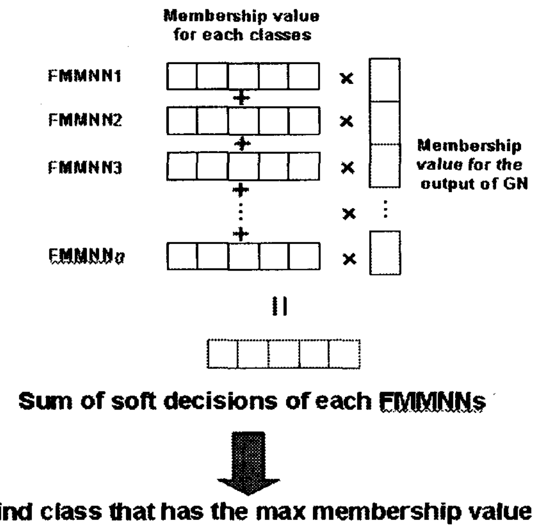

figure 2. Decision making process

IV. Simulation

Initial setting is as follow. Expert pool size is 20. Training parameters of each expert is from 0.025 to 0.5 by step size 0.025. The number of the selected experts is 5. Weights which are used to sum performance and independence measures are 3 and 2 respectively.

Facial expression data was used. [7] Data were randomly divided into 10 groups. 5 out of 10 groups were used for the training and the rest for the test. 3groups out of 5 training data groups were used for expert training and the rest for expert selection and gating network training. Total $252({}_{10}C_5)$ times simulation was done to alleviate the input data presentation order dependency problem. The results were averaged.

Performance of the proposed method is classification rate 95.7978%. the classification rates of selected training parameters are in the table (1)

Table(1) simulation result of original FMMNN with several Θ

θ	0.1	0.2	0.3	0.4	0.5
Mean (rate %)	95.5	89.68	80.7	7.83	71.2
Standard deviation	1.96	2.17	1.94	2.29	2.52

V. Conclusion

Committee machine with experts that have different training parameter was proposed to achieve the multi-resolution FMMNN that still has the fast single-pass on-line learning ability. Each expert is a original FMMNN so that expert has a fast, single-pass, and on-line learning ability. Gating network divides problem regionally and experts divides problem parametrically. Therefore regional and parametric divide and conquer are used.

For the further work, parameters, for example, the size of expert pool, the number of selected experts, and weights for performance and independence measures were set in the heuristic way. The systematic method to set them is needed.

References

[1] Patrick K. Simpson "Fuzzy Min-Max Neural Networks—Part 1: Classification," IEEE Transaction on Neural Networks, Vol. 3, NO. 5, September 1992.

[2] Meneganti M., Saviello F.S., Tagliaferri R. "Fuzzy neural networks for classification and detection of anomalies," IEEE Transactions on Neural Networks, Vol. 9 Issue: 5, Page(s): 848 -861Sep 1998

[3] Chen Xi, Jin Dongming, Li Zhijian, "Recursive training for multi-resolution fuzzy min-max neural network classifier," Proceedings of 6th International Conference on Solid-State and Integrated-Circuit Technology, Vol. 1 , pp. 131-134, 2001

[4] Rizzi, A.; Panella, M.; Frattale Mascioli, F.M., "Adaptive resolution min-max classifiers," Neural Networks, IEEE Transactions on , Volume: 13 Issue: 2 , Page(s): 402 -414 March 2002

[5] Gabrys, B., "Agglomerative learning for general fuzzy min-max neural network," Proceedings of the 2000 IEEE Signal Processing Society Workshop, Volume: 2, 11-13 Page(s): 692 -701 Dec. 2000

[6] Bogdan Gabrys, Andrzej Bargiela, "General Fuzzy Min-Max Neural Networks for Clustering and Classification," IEEE Transaction on Neural Networks, Vol. 11, NO. 3, May 2000.

[7] Dae-Jin Kim, Zeungnam Bien, "Design of a Personalized Classifier using Soft Computing Techniques and Its Application to Facial Expression Recognition," ISIS 2003