Fuzzy Integral-based Neural Network Ensemble for Facial Expression Recognition

Z.Y. Wang
School of Computer Science and Engineering
South China University of Technology
Guangzhou, China

N.F. Xiao
School of Computer Science and Engineering
South China University of Technology
Guangzhou, China

***Abstract*-Neural network is widely used in pattern classification. But an unavoidable fact is that many complex problems cannot get ideal correct rate of classification with a single neural network. Neural network ensembles provides a serial of theories and methods to simply ensemble outputs of some low performance neural networks, and the ensemble output can get higher correct rate. This paper tries to use fuzzy integral to complete the information integration of member classifier. A more effective fuzzy density function using in fuzzy integral was also proposed in this paper. Facial Expression Recognition (FER) was used to validate the method. Many simulation experiments were done to analyze the validity of neural network and different parameters. The Experiment data also proved the new fuzzy density function is more effective.**

Keywords-neural network ensembles; ensemble learning; information fusion; fuzzy integral; facial expression recognition

I. INTRODUCTION

Hansen and Salamon[1] introduced the Neural Network Ensemble(NNE) methods in 1990. They proved that ensemble a number of neural networks can significantly improve the generalization ability of neural network system.

Facial Expression Recognition (FER) is the extraction, analysis and classification of Facial expression features to help computer understand emotions of human beings. In 1971, Ekman and Friesen have identified six primary expressions: happiness, sadness, fear, disgust, surprise and anger[2]. Facial Expression classification is to "identify" the face expression. Neural network is a big family of the classifiers [3] [4].

In this paper, firstly the effectiveness of NNE and common approach of ensemble is introduced. Then an NNE method based on fuzzy integral according to the definition and characteristics of the fuzzy integral and fuzzy density is designed. Some discussion is done on the selection methods. Experiment using the FER as experimental subjects. The comparisons of effective between fuzzy integral method and other NNE methods are done. The influence of some parameters and functions in fuzzy integral is discussed and verified by the experimental result. Finally, future research goals and direction is proposed.

II. NEURAL NETWORK ENSEMBLE

A *Effectiveness of Neural Network Ensemble*

NNE makes use of multiple neural networks to learn the same problem. The output of NNE is work out from all outputs of its member networks. This method can significantly improve the generalization ability of neural network system, is regarded as a very effective way of learning. Krogh[5] proposed the formula of neural network ensemble generalization error by theoretical analysis. It indicates the performance of NNE neural network meets or exceeds the average performance of each member network in any case. Therefore, to enhance the generalization ability of NNE, on the one hand should improve the generalization ability of individual networks, on the other hand should be increased the differences between each member network as much as possible.

B *Integration Methods for NNE*

Many researches have been done and various integration methods have been proposed on ensemble learning , such as Bagging[6], Adaboosting[7], a simple vote, Bayesian voting, DS evidence theory[8], fuzzy integral[9] and the like. There are also studies[10][11] pointed out that the condition of effective ensemble learning is error rate of every single network's should be less than 0.5, otherwise it will improve the error rate of the results of ensemble. Neural network is widely used in classification. But single network has its limitation that a single vast scale NN with many hidden layers and nodes is difficult to train and control because of fitting excessive and local minimum problem. NNE based classifier can improve the result of classification.

III. FUZZY INTEGRAL BASED NEURAL NETWORK INTEGRATION

A *Fuzzy Measure and Fuzzy Integral*

1) Fuzzy measures: Definition 1: Let(X, Ω) denotes a measurable space, g:Ω→ [0, 1] is a set of functions, has the following properties: (1) boundary conditions: $g(\Phi)=0, g(x)=1$; (2) monotonic: $g(A)\le g(B)$, if $A\subset B\subset \Omega$; (3) $\{A_j\}_{i=1}^{\infty}$ is a collection of measurable increasing sequence, then $\lim_{i\to\infty} g(A_i)=g(\lim_{i\to\infty} A_i)$. The function g is called fuzzy measure, defined monotonic determine the g can be added not necessarily, that means the union set of two uncorrelated sets are not related to the collection cannot get the measures by adding measure of each part directly. Fuzzy measure is a monotonic and normalized set of functions. It can be used as an extension of probability measure. Sugeno[12] proposed fuzzy measure, meet the

following additional features. For all $A, B \subset X$, and $A \cap B = \Phi$ satisfies:

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \quad (1)$$

where, λ>-1.

2) gλ- fuzzy measure fuzzy density properties: Make X={x1,x2…xn} a finite set, defines gi = gλ{xi} , gi is called a fuzzy density function. According to Eq.(1):

$$g_\lambda(X) = \frac{\prod_{i=1}^{n}(1+\lambda g_i)}{\lambda} \quad (2)$$

According to the definition of fuzzy measure, gλ(X)=1, so get these high-order equation:

$$1+\lambda = \prod_{i=1}^{n}(1+\lambda g_i) \quad (3)$$

3) Sugeno fuzzy integral and Choquet fuzzy integral: Let X={x1,x2…xn} a finite set with a function h:X→[0,1]. Assumptions: h(x1)≥ h(x2)… ≥h(xn) (if disorderly, it is sorted in descending order). In the set of X, fuzzy measures gλ based Sugeno fuzzy integral as follows:

$$E = \max_{i=1}^{n}(\min(h_{x_i}, g_\lambda(A_i))) \quad (4)$$

In the Eq.(4). Let gi be a fuzzy measure for gλ, according to Eq.(1):

$$g_\lambda(A_i) = g^1 + g_\lambda(A_{i-1}) + \lambda g^1 g_\lambda(A_{i-1}) \quad (5)$$

where, λ can be calculated by the Eq. (3).

We also introduced Choquet fuzzy integral:

$$E = \sum_{i=1}^{n}(h(x_i) - h(x_{i-1}) g_\lambda(A_i)) = \sum_{i=1}^{n} h(x_i)(g_\lambda(A_{i+1}) - g_\lambda(A_i)) \quad (6)$$

In the Eq.(6).Choquet fuzzy integral can be seen as weighted sum of h(x1), h(x2)…h(xn), and the weight depends on the sort of {xi}, and sort of {xi} depends on the corresponding function h(xi) values.

B Fuzzy Integral and Neural Network

Fuzzy integral based integration method is an extension of the weighted average method, which can be applied to the neural network ensemble. h(xi) ∈[0,1] represents the results for classification of member network xi, fuzzy density gi indicates the importance of the member network, fuzzy integral is the nonlinear integration of the importance of each member network result. The results of fuzzy integral indicates the overall result from NNE for the target classification

Let X denotes the input to be classified. Assuming n member neural networks are used, and m categories can be divided. h(i, j) represents the i-th neural network considers the value of X belongs to the class j.e(i) for the fuzzy integral, which integrates multiple member neural networks output on the i class. Finally, we compare e(i), which corresponds to the maximum value of i is the output of NNE output. Literature[13] used Sugeno integral to handle recognition and classification of remote sensor images, and got a better result than a single neural network classification.

C The choice of Fuzzy Density

According to the fuzzy integral Eq.(4) and Eq.(6) ,the results of integration mainly related to two factors: First, the output of each member networks h, this can increase the classify ability of the member networks. Second is fuzzy density gi . The choice of fuzzy density function has a very important influence to the result of fuzzy integral. The choice the fuzzy density function has different methods in many literatures. Literature[14] considering the misclassification in fuzzy density factors, but it has possibility to let gi =0 or gi =1. In extreme cases, it will make a destruction of the validation of density. Literature[15] proposed method of determining fuzzy density from the correct rate, describing quality and distinguishable rate, but calculation is slightly complicated. Next, a method for determining the fuzzy density is proposed. Assume multiple neural network classifiers NN={nn1,nn2,…nnk} for NNE. X={x1, x2…xn}denotes the data to be classified and X can be divided into m categories, the classification results are expressed as Y={y1,y2…yn}. For the m-dimensional vector yi={yi1,yi2,..yim}, if xi belongs to the class j, then yij = 1, the remaining components are 0. The output for X of the k-th classifier is $O_k = \{o_k^1, o_k^2, ...o_k^n\}$. Define the confusion matrix of classifier k as below

$$P^k = \begin{bmatrix} p_{11}^k & p_{12}^k & \cdots & p_{1m}^k \\ p_{21}^k & p_{22}^k & \cdots & p_{2m}^k \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1}^k & p_{m2}^k & \cdots & p_{mm}^k \end{bmatrix} \quad (7)$$

where, p_{ij}^k denotes the amount of the class i input data was classified to class j by the k-th classifier. When i = j, it denotes the correct work while other cases present the classifier is not correct. Define the following parameters:

The correct classification rate of class i:

$$R_k^i = \frac{p_{ii}^k}{\sum_{j=1}^{m} p_{ij}^k}, R_k^i \in [0,1] \quad (8)$$

The error distance of class i:

$$D_k^i = \sqrt{\sum_{j=1}^{n} \frac{(o_k^i - y_j)^2}{n}}, D_k^i \in (0,1) \quad (9)$$

The failure level of class i:

$$M_k^i = \sum_{j=1,}^{m, j\neq 1} \left(\frac{p_{ji}^k}{\sum_{j=1}^{m} p_{ji}^k} \right), M_k^i \in [0,1] \quad (10)$$

therefore, we can define the fuzzy density of classifier k to class i :

$$g_k^i = \alpha R_k^i + \beta(1 - D_k^i) + \gamma(1 - M_k^i) \quad (11)$$

where, α、β、γ is the weight for the three factors for the classifiers ability, α+β+γ=1. For example, α=0.7, β=0.2, γ=0.1. Noted, $g_k^i \in [0,1)$ does not satisfy the fuzzy density (0,1) values condition, so when $g_k^i = 0$, let $g_k^i = \delta$, where δ is a positive number Infinite close to zero.

Fuzzy density according to the above method, including the correct rate of classifier, error distance between the standard output and actual output, and the degree of misclassification contains information of misclassification samples. It is a more comprehensive reflection of the importance of the member classifier in NNE.

IV. EXPERIMENTS

A Experimental Environment and Parameter Settings

JAFFE database[16] which contains 213 images of Japanese female facial expressions, was used in this study. Every image corresponds to one of the 7 categories of expression. All face images with a resolution of 150×100(pixel) and have handled with facial recognition, grey level normalization and resizing. Matlab's neural network toolbox was used as a simulation tool.

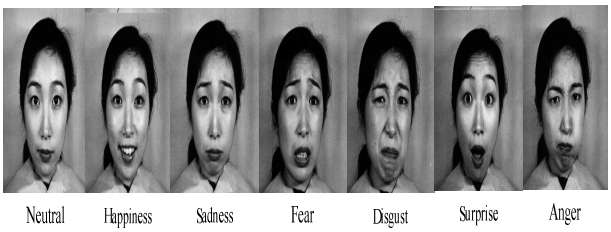

FIGURE I. THE SEVEN CLASSES OF HUMAN FACE EXPRESSION.

B Comparison of Single Neural Network and Neural Network Ensemble

Five neural networks were trained and calculated the correct rate for each neural network separately. Then use the integrated approach of Choquet fuzzy integral to fuse the five neural networks output, calculate the correct classification rate of integration. The experiment was repeated five times. The 10 neural networks are ensemble with the simple mean value algorithm and the result was shown in Table 1. The data of Table 1 shown the accuracy rates of single neural network were not high. But the result of NNE using the fuzzy integral is significantly higher than that of single neural network.

TABLE I. THE COMPARION OF SINGLE NN AND NNE.

	Single neural network					NNE
	NN1	NN 2	NN 3	NN 4	NN5	
1	0.84	0.72	0.81	0.68	0.74	0.86
2	0.85	0.72	0.81	0.77	0.73	0.89
3	0.82	0.79	0.86	0.76	0.75	0.91
4	0.81	0.69	0.73	0.69	0.68	0.84
5	0.88	0.78	0.79	0.79	0.65	0.91
Avg	0.84	0.74	0.8	0.738	0.71	0.882

C Influence of Parameters in Member Networks Setting

First, considered the Influence of the number of members to the NNE result. Different number of member network were used to ensemble the output. Experiment was repeated five times. If could be seen from the data in Fig.2, when the k value was relatively small, k increased, the accuracy rate was also increasing corresponding. But if k continue increased, the number of members of the member network had been unable to significantly increase the effect of NNE. This indicated that the effect of NNE cannot increase unlimited with the increasing of number of member networks, when the number of members reached a certain degree, the newly added members was no longer continue to improve the generalization ability to NNE.

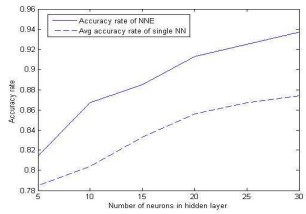

FIGURE II. QUANTITY OF MEMBER NN AFFECTS THE RESULT OF NNE.

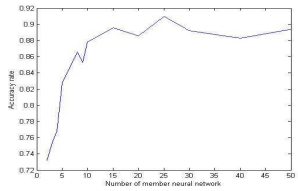

FIGURE III. PERFORMANCE OF MEMBER NN AFFECTS THE RESULT OF NNE.

According NNE theory, improve the individual members of the generalization ability of neural networks can improve the overall generalization ability to NNE. In the experiment, we used a method to increase the number of neurons in hidden layer of the member network to improve generalization ability of NNE. The results shown in Fig.3, along with members of the hidden layer gradually increased, NNE classification accuracy was also improved.

D Comparison of Different Fuzzy Density Function

Experiments used three kinds of fuzzy density function for the same set of data to test NNE classifier. Conducted a total of five experiments to obtain the data shown in Table 2. The "density #3" is the fuzzy density function proposed in this paper, "density #1" and "density #2", respectively, random fuzzy density and accuracy fuzzy density. As can be seen from the data, the accuracy rate of "density #1" was the lowest, it

was means that it is fuzzy density have a significant impact on the integration of the NNE results, the method of setting the importance or weight (ie fuzzy density)of member network is not an effective and reliable approach. Besides the accuracy rate, the density #3 also considered the error distance and classification effectiveness factors, so the its accuracy rate was slightly higher than the density #2 .

TABLE II. THE COMPARION OF RESULT OF NNE WITH DIFFERRNT FUZZY MESSURE.

	1	2	3	4	5	Avg
Density #1	0.86	0.83	0.82	0.87	0.78	0.832
Density #2	0.85	0.84	0.87	0.91	0.88	0.87
Density #3	0.83	0.86	0.89	0.91	0.92	0.882

V. CONCLUSIONS

This paper established a neural network ensemble classifier model, made use of fuzzy integral method to fuse the member networks' output. A fuzzy density is designed which measure the importance of member networks of NNE from three aspects: accuracy rate, error distance and failure extent. Through experiments, the NNE model was applied to FER. The result data verified the effectiveness of the NNE, proved simply increase the generalization ability of member networks can continuously improve the quality of NNE and achieved the FER problem correctly classified. Finally experiment compared the effect of using different fuzzy density function to the NNE result, the data shown that the proposed fuzzy density function in this paper has certain advantages in the accuracy rate.

NNE can use a simple method to enhance the ability of weak classifiers, but the problem is its long computation time, a large-scale NNE also require heavy computation for training and fusion. The next step of research is to further improve the NNE algorithms and model to establish a parallel NNE, and enable it to run in parallel computing environments, such as grid or cloud computing platform. The next tasks is to apply the identify and classify capability of NNE to the field of machine vision applications such as face recognition or facial expression classification.

ACKNOWLEDGEMENTS

We would like to express out gratitude for provide financial support from National Natural Science Foundation of China(61171141) and the special cooperative research funds of Guangdong Province and China Education Ministry(2012B091100448).

REFERENCES

[1] Hansen L K, Salamon P. Neural network ensembles. IEEE Trans Pattern Analysis and Machine Intelligence. 12(10), pp.993-1001, 1990.

[2] P. Ekman and W. V. Friesen, The Facial Action Coding System.San Francisco: Consulting Psychologist Press, 1978

[3] Kobayashi and F. Hara. Facial Interaction between Animated 3D Face Robot and Human Beings, Proc. Int'l Conf. Systems, Man, Cybernetics, pp.3732-3337;1997.

[4] Tian Y., Takeo K. and Cohn J. Recognizing Facial Actions by Combining Geometric Features and Regional Appearance Patterns ,CMU-R I-TR-01-01,2001, ,pp.1-25

[5] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning. Neural information processing systems. pp.231-238,1995.

[6] Breiman L. ,Bagging Predictors. Machine Learning, 24(2), pp.123-140, 2002,.

[7] Freund Y, R. E. Schapire. Experiments with a New Boosting Algorithm. Thirteenth International Conference on Machine Learning.pp.148-156,1996.

[8] Xu, L. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, IEEE Transactions on Systems, Man and Cybernetics. 22(3),pp.418- 435,1992.

[9] Roli E, Kittler J. Fusion of Multiple Classifiers. Information Fusion. 3(4), pp.243 - 244,2002.

[10] G. Valentini, F. Masulli. Ensembles of learning Machines. Neural Nets WIRN Vietri-02.M.Marinaro and R. Tagliaferri, Eds.: Springer-Verlag, Heidelberg, 2002, Invited Review.

[11] Kittler, J, Hatef, M, Duin, R. P, et al. J. On Combining Classifiers. IEEE Trans. Pattern Anal. Mach. 226-239,1998.

[12] M.Sugeno. Theory of Fuzzy Integrals and Its Application. Tokyo: Tokyo Institute of Technology, 1974.

[13] Reddy G., Mohan B., Pal, Sankar. Classification of Remotely Sensed Images Using Neural-Network Ensemble and Fuzzy Integration.Pattern Recognition and Machine Intelligence. Heidelberg： Springer Berlin, pp.350-355,2005.

[14] Li Yurong,Qiao Bin,Jiang Jingping. An Algorithm for Combining Based on Fuzzy Integral and Genetic Algorithms. Computer Engineering and Applications. 12,pp.119-121,2002.

[15] Wang Xiaojun. A Study on Multi-classifiers Fusion Based on Fuzzy Integrals. Tianjin: Hebei University, 2005.

[16] JAFFE. http://www.kasrl.org/jaffe.html