A New Approach to the Design of Combining Classifier Based on Immune Algorithm

Moon Hwan Kim *, Keun Ho Jeong **, Young Hoon Joo ** and Jin Bae Park *

*Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
**School of Electronic and Information Engineering, Kunsan National University, Kunsan, Korea

Abstract: This paper presents a method for combining classifier which is constructed by fuzzy and neural network classifiers and uses classifier fusion algorithms and selection algorithms. The input space of combing classifier is divided by the extended hyperbox region proposed in this paper to guarantee non-overlapped data property. To fuse the fuzzy classifier and the neural network classifier, we propose the fusion parameter for the overlapped data. In addition, the adaptive learning algorithm also proposed to maximize classifier performance. Finally, simulation examples are given to illustrate the effectiveness of the method.

Keywords: Fuzzy classifier, neural network classifier, immune algorithm, fuzzy hyperbox

1. Introduction

Over the last few years, there has been an ever-increasing interest in the area of classification system. Many problems in business, science, industry, and medicine can be treated as classification problems. Examples include bankruptcy prediction, credit scoring, medicine diagnosis, quality control, handwritten character recognition and speech recognition.

The final goal of designing classification system is to maximize the accuracy of designed classifier. This objective led to the development of different classification scheme for any classification problem to be solved. It had been observed in such design studies that the combination of classifier can overcome the limitation of each classifier [1]. These observations motivated the relatively interest in combining classifier. Generally, there are two basic approaches a combination of classifier algorithms may take: classifier fusion and dynamic classifier selection. In classifier fusion algorithms, individual classifiers are applied in parallel and their outputs are combined [1]. The dynamic classifier selection attempts to predict which single classifier is most likely to new correct for a given data sample. Only one output of the selected classifier is considered in the final decision. Especially, there are many researches to development classifier fusion algorithms including the majority vote [2], [3], the Borda count [4], unanimous consensus [3], [4], thresholded voting [3] and polling methods which utilize heuristic decision [5], [6].

In this paper, we propose the design of new combining classifier system that combines the fuzzy classifier and the neural network classifier using combination algorithms. To combine different classifiers, we use two aspects of combination algorithms: the input space partition in classifier selection algorithm and the parallel form of discriminant function of classifier fusion algorithm. The proposed classifier divides sample data by overlapping property based on the extended fuzzy min-max hyperbox. If data belongs to overlapping data class, the classifier determines the class of data using a discriminant function. The discriminant function calculated by fuzzy rule uses fuzzy TSK classifier model [7]. The consequent of each rule is composed the combination of the fuzzy membership function and the neural network discriminant function with the fusion parameter. The fusion parameter that represents fusion ratio is determined by immune algorithm. To optimize combing classifier, we propose adaptive learning algorithm including immune algorithm.

This work has been supported by EESRI(R-2003-B-078), which is funded by MOCIE(Ministry of commerce, industry and energy).

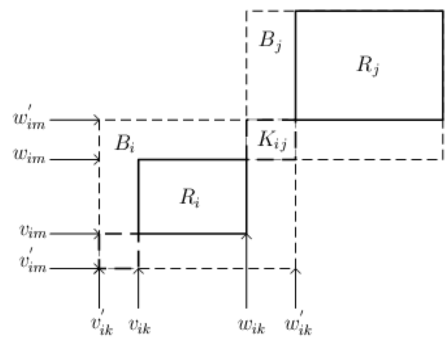

Fig. 1. Extended fuzzy min-max hyperbox region

2. Designing the combining classifier

2.1. Extended fuzzy min-max hyperbox region

In this paper, we propose the new hyperbox called extended fuzzy min-max hyperbox to distinguish the overlapped region and the non-overlapped region. Hyperboxes, defined by pair of min-max points, and their corresponding membership functions are used to create fuzzy subsets of the n-dimensional pattern space [8]. Extended fuzzy min-max hyperboxes are also defined by pair of min-max points, and their corresponding membership function. The general hyperbox that used in other research needs amount of hyperboxes to represent data's distribution. However, the proposed hyperbox that has two sub hyperboxes can describe the each class's data distribution using one hyperbox. The extended hyperbox has a inner sub hyperbox and a outer sub hyperbox. The inner sub hyperbox describes the reliable region that is distributed by non-overlapped data. The outer sub hyperbox has the region that describes entire data distribution. Figure 1 represents extended fuzzy min-max hyperbox region. A region $\mathcal{R}_i$ represents the inner sub hyperbox of the class labeled as i. A region $\mathcal{B}_i$ represents the boundary region which is a part of outer sub hyperbox region excepted the inner sub hyperbox region. The inner sub

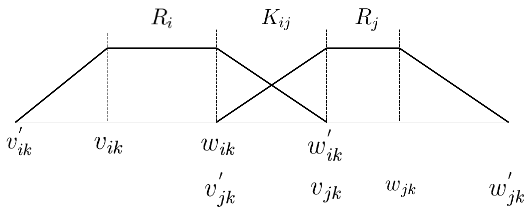

Fig. 2. The membership function of the extended fuzzy min-max region

hyperbox is described by a pair of min-max points v_{ik}, w_{ik}. Similarly, the outer sub hyperbox is described by a pair of min-max points v'_{ik}, w'_{ik}. Especially, A region $\mathcal{K}_{ij}$ has the complex region that is distributed by overlapped data. In this paper, we separate data by overlapping property and apply proper classification method.
Figure 2 expresses the membership function corresponding to the extended fuzzy min-max hyperbox region. This membership function $\mu_{A_{ik}}(x)$ denotes the form of the trapezoidal membership function.

$$\mu_{A_{ik}}(x) = \begin{cases} \frac{x_i - v_{ik}}{v_{ik} - v'_{ik}}, & v_{ik} < x < v'_{ik} \\ 1, & v_{ik} < x < w_{ik} \\ \frac{w'_{ik} - x_i}{w'_{ik} - w_{ik}}, & w_{ik} < x < w'_{ik} \\ 0, & others \end{cases} \tag{1}$$

2.2. Design of combination of classifiers

The fundamental architecture of combining classifier that is proposed in this paper is described as following fuzzy rule:

R_k: IF x_1 is A_{1k} and ... and x_n is A_{nk} then $g_{k1}= z_{k1}$ and $g_{kn} = z_{kn}$

where the subscript n is the number of feature, A_{ik} is the membership function of the ith feature of the kth rule and g_{ki} is the discriminant function of $k, (1 \le k \le c)$ class of the ith feature.
The proposed classifier determines the classification result C using the maximum discriminant function value.

$$C = \arg\{\max g_i(x)\} \tag{2}$$

To combine the fuzzy classifier and the neural network classifier, we propose the value of discriminant function z_{im} as

$$z_{im}(x) = \beta \frac{\tau_i}{c}(x) + (1-\beta) n_{mi}(x) \tag{3}$$

where β is the fusion parameter, c is the number of class, $\tau_i(x)$ is the firing strength function of class i [7] and n_{im} is the output of ith network classifier of the feature m.
Especially, to guarantee the data belonged to region $\mathcal{R}$, the output of each classifier has following values,

$$\begin{aligned} \mu_{A_{mi}}(x) = c, &\quad x_m \in \mathcal{R}_i \\ n_{mi}(x) = c, &\quad x_m \in \mathcal{R}_i \end{aligned} \tag{4}$$

In addition, when the kth feature data x_k belongs to class i, the classification error $\Psi(x_k)$ can be defined as

$$\Psi(x) = \sum_{j=1, j\neq i, g_i(x) < g_j(x)}^{c} \frac{1}{\{g_j(x) - g_i(x)\}^2} \tag{5}$$

Theorem 1: Let the discriminant function of the fuzzy rule i be a $g_i(x)$, the difference between the firing strength functions of ith and jth fuzzy rule be a $\Delta T_{ij} = \{\tau_i(x) - \tau_j(x)\}$, the difference between neural network output value of ith and jth of kth feature be a $\Delta N_{ij} = \sum_{k=1}^{M} n_{ki}(x) - n_{kj}(x)$. The discriminant function $g_i(x)$ satisfies the followings:

$$g_i(x) > g_j(x), 1 < j < c, j \neq i, \tag{6}$$

if the fusion parameter β satisfy the followings:

$$\max_{j=1, j\neq i, \Delta N_{ij} < \Delta T_{ij}}^{c} \{\frac{\Delta N_{ij}}{\Delta N_{ij} - \Delta T_{ij}}\} < \beta < \min_{j=1, j\neq i, \Delta N_{ij} > \Delta T_{ij}}^{c} \{\frac{\Delta N_{ij}}{\Delta N_{ij} - \Delta T_{ij}}\}. \tag{7}$$

Proof: The discriminant function $g_i(x)$ can be calculated as

$$g_i(x) = \frac{\sum_{k=1}^{M} z_{ki} \cdot \tau_k(x)}{\sum_{k=1}^{M} \tau_k(x)}. \tag{8}$$

where $M, (M = c)$ is the number of the rule. Next, classification condition (6) be represented using (8)

$$g_i(x) - g_j(x) > 0 \tag{9}$$

$$\frac{\sum_{k=1}^{M} z_{ki}(x) \cdot \tau_k(x)}{\sum_{k=1}^{M} \tau_k(x)} - \frac{\sum_{k=1}^{M} z_{kj}(x) \cdot \tau_k(x)}{\sum_{k=1}^{M} \tau_k(x)(x)} > 0 \tag{10}$$

$$\frac{\sum_{k=1}^{M} z_{ki}(x) \cdot \tau_k(x) - \sum_{k=1}^{M} z_{kj}(x) \cdot \tau_k(x)}{\sum_{k=1}^{M} \tau_k(x)} > 0 \tag{11}$$

$$\frac{\sum_{k=1}^{M} (z_{ki}(x) - z_{kj}(x)) \cdot \tau_k(x)}{\sum_{k=1}^{M} \tau_k(x)} > 0 \tag{12}$$

$$\sum_{k=1}^{M} \{z_{ki}(x) - z_{kj}(x)\} > 0, \; \textit{because all } \tau_k(x) > 0 \tag{13}$$

By exchange z_{ki} with (3), we obtain simplified classification condition,

$$\sum_{k=1}^{M} \{z_{ki}(x) - z_{kj}(x)\} > 0 \tag{14}$$

$$\sum_{k=1}^{M} \beta\{\frac{\tau_i}{c}(x) - \frac{\tau_i}{c}(x)\} + (1-\beta)(n_{ik}(x) - n_{jk}(x)) > 0 \tag{15}$$

$$\beta\{\tau_i(x) - \tau_j(x)\} + \sum_{k=1}^{M} \{(1-\beta)(n_{ik}(x) - n_{jk}(x))\} > 0 \tag{16}$$

$$\beta \Delta T_{ij} + (1-\beta)\Delta N_{ij} > 0 \tag{17}$$

As a result, the fusion parameter β has upper and lower bound depending on the value of ΔT_{ij} and ΔN_{ij}. Finally, (7) guarantee the bound of fusion parameter(6). ■
On top of this, some data which misclassified can't have boundary that satisfy (7). To find the optimal fusion parameter, we need to consider the data that can't satisfy (7).
Corollary 1: Let boundary of fusion parameter is denoted as followings:

$$B_s(N_s) < \beta < B_l(N_l) \tag{18}$$

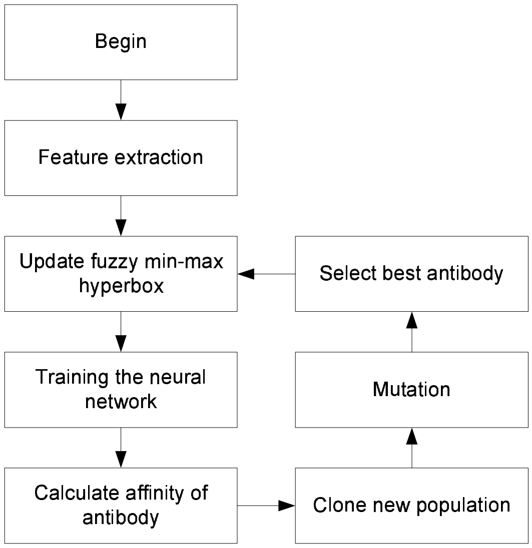

Fig. 3. Adaptive learning algorithm for combining classifier

where $B_s(N_s)$ is the N_sth largest value of fusion parameter and $B_l(N_l)$ is the N_lth smallest value of it.

If sample data is classified incorrectly, minimums N_s and N_l that satisfying (18) guarantee minimum classification error $\Psi(x_k)$.

Proof: Differences $N_l - 1$ and $N_s - 1$ mean the number of class that does not satisfy (6) by Theorem 1. We can easily know that the increase of differences mean increase of classification error by definition of classification error (5). Therefore, if we guarantee minimum of differences, we can get the minimum classification error. ■

3. Adaptive learning algorithm for combining classifier

3.1. On-line adaptation using immune system

The basic concept of on-line adaptation using immune system is tuning the classification system by data-driven method. In general, The data driven-method provide more flexible classification system and more extensible than another tuning methods [8]. The combination classifier proposed in this paper presents tuning methods for the fuzzy hyper box, the neural network and the fusion parameter. Figure 3 represents the steps of the whole algorithm.

3.2. Feature extraction

The extended fuzzy min-max hyperbox region has some drawbacks which caused by using just one hyperbox to one class data. However, using feature extraction method, we can reduce the fault of the extended fuzzy min-max hyperbox. In this paper, we use a new feature extraction based on PCA (principal component analysis) method [12]. Let $\mathbf{X}= [x_1\ x_2\ ..\ x_n]^T$ is the original feature, and $\mathbf{Y}= [y_1\ y_2\ ...\ y_n]^T$ is the extracted feature. Then, we can get first principal component $\mathbf{u}_1$ from $\mathbf{X}$ using PCA. The first new feature J that has the largest region $\mathcal{R}$ is denoted as

$$J = \mathbf{u}_i{}^T\mathbf{X} \tag{19}$$

To reduce the complex region $\mathcal{K}$, we need the feature that has long distance between another classes. Therefore, the new feature is determined by using the first new feature which the smallest complex regions.

$$y_{kd} = x_{kd} + j_d \tag{20}$$

where y_{kd} is the the kth new feature data of the dth data, x_{kd} is the the kth previous feature data of the dth data and j_d is the first new feature of the dth data.

3.3. Construction extended fuzzy min-max hyperbox region

Let data x_k be a new kth feature data. The outer min-max points v'_{ik}, w'_{ik} and inner min-max points v_{ik}, w_{ik} are updated by x_k.

If x_k belongs to $\mathcal{R}_j, (j = 1, .., c, j \neq i)$, the next inner min-max points have

$$\begin{aligned} v_{ik}^{next} &= v_{ik} \\ w_{ik}^{next} &= w_{ik}. \end{aligned} \tag{21}$$

If x_k does not belong to $\mathcal{B}_j, (j = 1, .., c, j \neq i)$, the next inner min-max points have

$$\begin{aligned} v_{ik}^{next} &= \min\{v_{ik}, x\} \\ w_{ik}^{next} &= \min\{w_{ik}, x\}. \end{aligned} \tag{22}$$

In addition, the next outer min-max points always have

$$\begin{aligned} v_{ik}^{'next} &= \min\{v'_{ik}, x\} \\ w_{ik}^{'next} &= \max\{w'_{ik}, x\}. \end{aligned} \tag{23}$$

3.4. Tuning the neural network discriminant function

The basic tuning goal of the neural network discriminant function can be described as

$$\begin{aligned} \alpha < n_{ik}(x_k) < 1, &\quad k = i \\ 0 < n_{ik}(x_k) < \alpha, &\quad k \neq i \end{aligned} \tag{24}$$

where α is a tolerance that has very small value.

Because of (3), the neural network discriminant function is evaluated by feature. Therefore, n_{ki} is represented as SISO(single input single output) multi-layer neural network. Because the four layer neural network guarantee general function approximator [11], generally, four-layer neural networks are used for n_{ki}.

If data x belongs to K_{ij}, the neural network classifier discriminant function is trained with a input vector T_i and a output vector T_o,

$$T_i = [x_k] \tag{25}$$

$$T_o = [e_{ij}] \tag{26}$$

$$\tag{27}$$

where e_{ij} is the delta function which is denoted as,

$$e_{ij} = \begin{cases} 1, & i = j \\ 1, & i \neq j \end{cases} \tag{28}$$

3.5. Immune algorithm for the fusion parameter tuning

The immune algorithm is a kind of evolutionary optimization method. The immune algorithm is designed based on immune system that plays a important role that protects body of human or animal. An immune system detects germs and produces proper antibody (Ab) to eliminates germs. Basically, the immune system remembers germs infect body once for the fast response. Similar to the immune system, the immune algorithm has detection ability and reproducing ability for objects. Because of the immune system reinforces its system by new germs that come from out of body, the immune algorithm is trained by data-driven method [9].

The fusion parameter has dynamic boundary (7) that changed by input data x. Therefore, we need to remember all boundary condition of data and find the optimal condition that makes every possible effort satisfying (7). The satisfaction of (7) is evaluated to use affinity evaluation function that represents fitness value of antibody.

The immune algorithm has following steps:

Step 1 Randomly choose an antigen Ag_k and present it ro all Ab's in the repertoire $Ab_{ij} = Ab_{ij\{r\}} \cup Ab_{ij\{m\}}$, $(r+m = N)$. To optimize the fusion parameter, the antibody Ab_{ij} is constructed by fusion parameter.

Step 2 Determine the vector $\mathbf{f}_j$ that contains affinity of Ag_j to all the N Ab's in Ab. The affinity is evaluated using object function $f_{obj}(x)$. The object function $f_{obj}(x)$ can be denoted as,

$$f_{obj}(x) = \begin{cases} \frac{x - B_s(N_s^{'})}{M - B_s(N_s^{'})}, & B_s(N_s^{'}) \le x < M \\ \frac{B_l(N_l^{'}) - x}{B_l(N_l^{'}) - M}, & M \ne< x < B_l(N_l^{'}) \end{cases} \tag{29}$$

where M is the mean between $B_s(N_s^{'})$ and $B_l(N_l^{'})$, $N_s^{'}$ and $N_l^{'}$ are the minimum values satisfying (18) of new sample data.

Step 3 Select the n highest affinity Ab's from Ab_{ij} to compose a new set of $Ab_{ij\{n\}}^k$ of high-affinity Ab's related to A_{g_k}.

Step 4 The n selected Ab's will be cloned independently and proportionally to their affinities, generating a repertories C^k of clones.

Step 5 The repertories C^k is submitted to an affinity maturation process, generating a population C_k* of matured clones.

Step 6 In mutation process, Ab's are mutated until affinity of them is greater than $\mathbf{f}_j$ or loop counter is over maximum loop

Step 7 Determine the affinity $\mathbf{f}_j^*$ of matured clones C^k in relation to antigen Ag_j.

Step 8 From this set of clones C_k*, reselect the one with highest affinity (A_j^*) in relation to Ag_j to be a candidate to enter the set of memory antibodies $Ab_{\{m\}}$

Step 9 Finally, replace the m lowest $\mathbf{f}_j^*$ Ab's from Ab with the m highest $\mathbf{f}_j$ candidate Ab's from C_k*

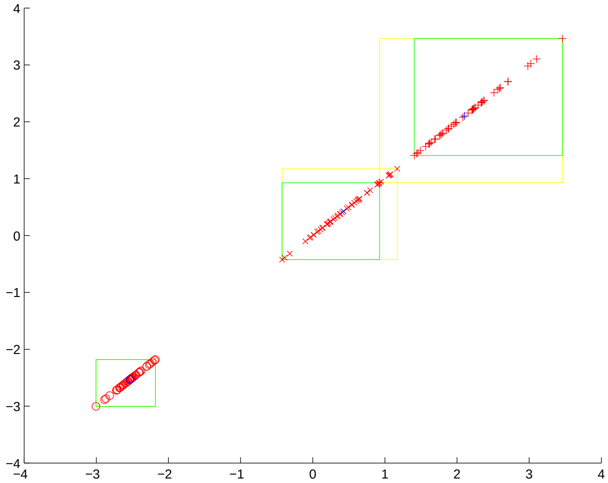

Fig. 4. The extended fuzzy min-max hyperbox region of sepal length

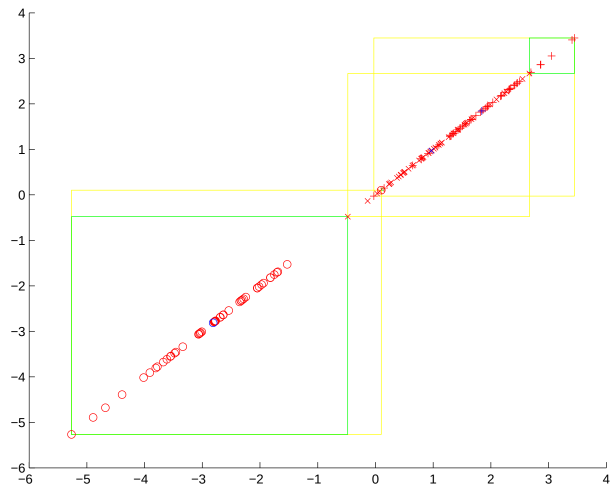

Fig. 5. The extended fuzzy min-max hyperbox region of sepal width

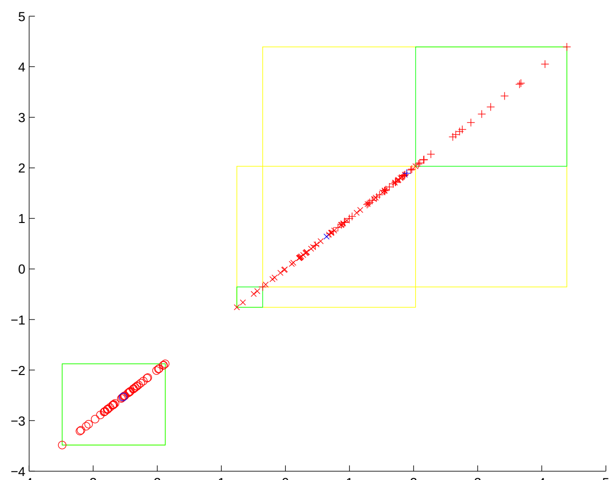

Fig. 6. The extended fuzzy min-max hyperbox region of petal length

4. Simulation Results

We apply the proposed methods to the iris data to verify the effectiveness of our methods. The classification problem of

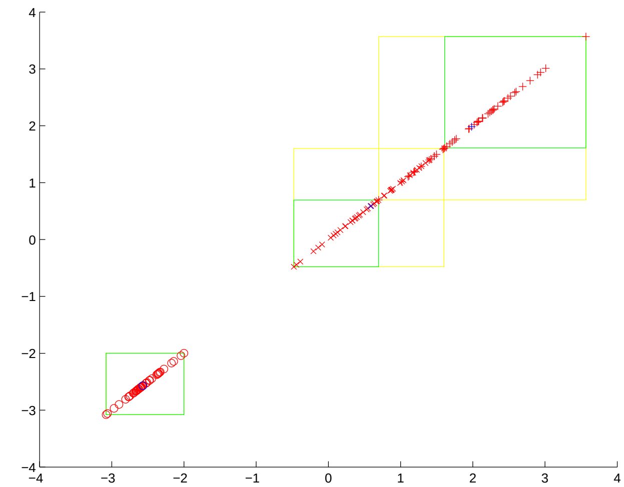

Fig. 7. The extended fuzzy min-max hyperbox region of petal width

the iris data is to classify three species of iris by the four-dimensional pattern vectors. There are 50 samples of each class in this classification problem. To examine the learning ability for training pattern, we use all the 150 samples as training pattern and performed computer simulations.

Table 1 shows the number of data belongs to region $\mathcal{R}$, $N(\mathcal{R})$, and the classification accuracy with feature extraction. When classifier has data from feature extraction that proposed in this paper, we can confirm the increase of $N(\mathcal{R})$ and the classification accuracy. Figures 4–7 show the extended fuzzy min-max hyperbox of each feature. In addition, the fusion parameter is determined using the immune algorithm. Figure 8 shows the classification accuracy (C.A.) by the fusion parameter. The fusion parameter of classifier with feature differs from classifier without feature extraction. This difference shows that effectiveness of neural network in complex region. The data is distributed in complex region has no common rule to classify. Therefore, we use neural network to classify data in complex region because the neural network has distribution free property [10].

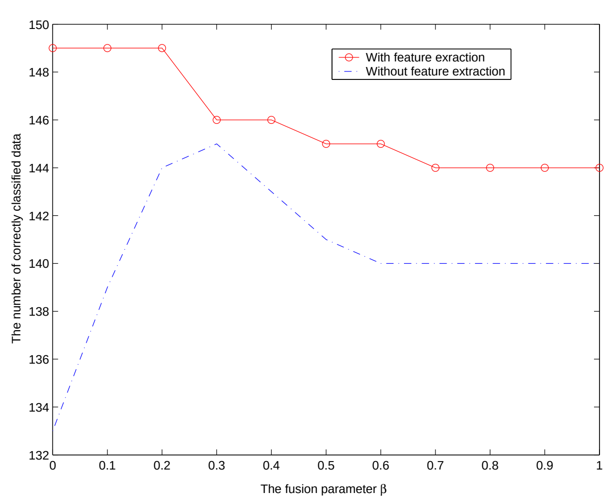

Fig. 8. The changes of fusion parameter with feature extraction

Table 1. The effect of feature extraction

$N(\mathcal{R})$	f_1	f_2	f_3	f_4	sum	C.A.
X	26	5	106	90	227(37.83%)	95.33%
Y	138	53	68	110	369(61.5%)	99.33%

Table 2. Comparing classification accuracy

Ref.	the number of rule	C.A.
[13]	5	96.67%
[14]	8	96.3%
[15]	4	97.33%
Ours($\beta = 1.98$)	3	99.33%

Finally, We can confirm the superior of combining classifier proposed in this paper from Table 2. The number of fuzzy rule is considered to compare the classification ability.

5. Conclusion

This paper proposes a new combing classifier that combines the fuzzy classifier and the neural network classifier. To combine two classifiers, we propose the extended fuzzy hyperbox for classifier selection and the fusion parameter for classifier fusion. The extended hyperbox distinguishes the data by overlapping property and guarantees correct classification of the non-overlapped data. Also, the fusion parameter combines the output of the fuzzy classifier and the neural network classifier to guarantees correct classification of the non-overlapped data.

The classifier presented in this paper is optimized by using the adaptive learning algorithm. The adaptive learning algorithm is composed by four algorithm: feature extraction, extended fuzzy hyperbox tuning, neural network learning, and immune algorithm for the fusion parameter. Each algorithm is driven by sample data except feature extraction algorithm. Finally, simulation results show the proposed tuning algorithms correctly worked and show the superior of proposed classifier.

References

[1] K. Wood, W.P. Kegenlmeyer Jr., K. Boewyer, "Combination of Muliple Classifier Using Local Accuracy Estimates," IEEE Transactions on Patten Analysis and Machine Intellegence, vol. 19, no. 4, 1997.

[2] L. Lam and C.Y. Suen, "A Theoritical Analysis of the Application of Majority Voting to Patten Recognition," Proc. 12th International Conference on Pattern Recognition and Computer Vision, pp. 418-420, Jeusalem, 1994.

[3] L. Xu, A. Krzyzak and C.Y. Suen, "Methods of Combining Multiple classifier and Thier Applcation to Handwriting Recognition," IEEE Transactions on Systems, Man and Cybernetics, vol. 22, no. 3, 1992.

[4] T.K. Ho, J.J. Hull and S.N. Srihari, "Decision Combina-

tion in Multiple Classifier Systems," IEEE Transactions on Patten Analysis and Machine Intellegence, vol. 16, pp. 66-75, 1994.

[5] F. Kimura and M. shidar, "Handwriting Numerical Recognition Based on Multiple Algorithms," Pattern recognition, vol. 24, No. 10, pp. 969-983, 1991.

[6] C. Nadal, R. Legault and C.Y. Suen, "Complementary Algorithms for Recognition of Totally Unconstrained Handwritten Numerals," Proc. 10th International Conference on Petten Recognition, pp. 443-449, Atalntic City, N.J., 1990.

[7] L. Kuncheva, "How Good Are Fuzzy If-Then Classifiers?," IEEE Transactions on Systems, Man and Cybernetics, vol. 30, no. 4, 2000.

[8] P.K. Simpson, "Fuzzy Min-Max Neural Networks? Part 1: Classification", IEEE Transactions on Neural Networks, vol. 3, pp. 776-786, 1992.

[9] L.N. De Castro, F.J. Von Zuben, "Learning and Optimization Using the Clonal Selection Principle," IEEE Transactions on Evolutionary Computation, vol. 6, pp. 239 -251, 2002.

[10] G.P. Zhang, "Neural Networks for Classification: a Survey," IEEE Transactions on Systems, Man and Cybernetics, Vol. 30, pp. 451 -462, 2000.

[11] S. Abe, "Pattern Classification: Neuro-fuzzy Mehtods and Their Comprison," Springer, pp.23-24, 2001.

[12] J. E. Jackson, "A User's Guide to Principal Components," John Wiley and Sons, pp. 1-25, 1991.

[13] Y. Shi, R. Eberhat and Y. Chen,"Implementation of Evolutionary Fuzzy Systems," IEEE Transaction on Fuzzy Systems, vol. 7, pp. 109-119, 1999.

[14] R. Li, M. Mukaidono and I.B. Turksen, "A Fuzzy neural Network for Pattern Classification and Feature Selection", Fuzzy Sets Systems, Vol. 130, pp. 101-108, 2002.

[15] T.P. Hong and J.B. Chen, "Processing Indivisual Fuzzy Attributes for Fuzzy Rule Induction," Fuzzy Sets Systems, vol. 112, pp. 127-140, 2000.