#### 2319

# LETTER Design of Hierarchical Fuzzy Classification System Based on Statistical Characteristics of Data\*

Chang Sik SON<sup>†</sup>, Member, Yoon-Nyun KIM<sup>††</sup>, Kyung-Ri PARK<sup>†††</sup>, and Hee-Joon PARK<sup>††††a)</sup>, Nonmembers

**SUMMARY** A scheme for designing a hierarchical fuzzy classification system with a different number of fuzzy partitions based on statistical characteristics of the data is proposed. To minimize the number of misclassified patterns in intermediate layers, a method of fuzzy partitioning from the defuzzified outputs of previous layers is also presented. The effectiveness of the proposed scheme is demonstrated by comparing the results from five datasets in the UCI Machine Learning Repository.

key words: hierarchical fuzzy classification system, fuzzy partitioning, rule generation

#### 1. Introduction

The design of a fuzzy rule-based system is commonly a time-consuming activity involving the acquisition of knowledge, reduction of the total number of rules, and tuning of the involved parameters, all of which affect the accuracy and/or interpretability of the fuzzy systems. In recent years, various approaches have been proposed to generate fuzzy rules from training data for classification problems: heuristic approach [1], neuro-fuzzy approach [2], genetic approach [3], [4], and hierarchical approach [5] to overcome "curse of dimensionality" problems. Almost all of these studies were aimed at developing methods that have classification ability comparable to those of the conventional methods with less computational burden while dealing with highdimensional problems. However, the total number of rules in standard fuzzy systems (SFSs) increases exponentially with the number of input variables. To overcome this explosion problem, a hierarchical fuzzy system (HFS) has been devised [6], the theoretical validity for which is established in [7]. Since this pioneering work, various other related studies have also been performed, including the approximation capabilities of a nonlinear function in HFSs [8], [9],

<sup>†</sup>The author is with the Dept. of Medical Informatics, School of Medicine, Keimyung University, Republic of Korea.

<sup>††</sup>The author is with the Dept. of Internal Medicine, School of Medicine, Keimyung University, Republic of Korea.

<sup>†††</sup>The author is with CheongShim International Academy, Republic of Korea.

<sup>††††</sup>The author is with the Dept. of Biomedical Engineering, School of Medicine, Keimyung University, Republic of Korea. Corresponding author.

\*This work was supported by the grant No. RTI 04-01-01 from the Regional Technology Innovation Program of the Ministry of Knowledge Economy (MKE) and the attached research institute Research Grant of Keimyung University in 2007.

a) E-mail: hjpark@kmu.ac.kr

DOI: 10.1587/transinf.E93.D.2319

and the evolutionary approach to optimize the parameters of nonlinear consequents [10] where the results have confirmed that HFSs are useful for overcoming the explosion problem in SFSs. However, the main problem with HFSs is that the outputs in the intermediate layers do not possess a physical meaning when the outputs of the previous layer and inputs of the subsequent layers are defined as mapping variables for the next layers. Thus, to reduce this problem, [11] describes a mapping rule-based scheme to reduce the number of rules involved in the middle layers of the HFSs. This letter proposes a scheme for designing a hierarchical fuzzy classification system (HFCS) based on statistical characteristics of the data. Also, we present a method that minimizes the number of misclassified patterns in intermediate layers to improve the classification capability of HFCS. We then apply the proposed method to classification problems and discuss its effectiveness.

## 2. Design of Hierarchical Fuzzy Classification System

Figure 1 shows a structure of the proposed HFCS, comprised of n-1 two input fuzzy systems with n input variables.

In Fig. 1 (a), two input variable  $a_1$  and  $a_2$  are fed into the first layer, whose output  $y_1^*$  is then combined with another input variable  $a_3$  into the second layer, and this procedure continuous until all the input variables have been used. Each substructure has two main components, fuzzy partitioning and rule generation as shown in Fig. 1 (b), where  $y_j^*(j = 1, ..., n - 1)$  denotes the defuzzified outputs of *j*-th fuzzy system in the HFCS.



**Fig. 1** Block diagram of HFCS. (a) Structure of n input HFCS and (b) Substructure of *j*-th fuzzy system in the HFCS.

Manuscript received December 24, 2009.

Manuscript revised April 23, 2010.

#### 2.1 Fuzzy Partitioning

Most of high-dimensional data is overlapped in feature space relative to some classes and its distribution is heterogeneous for each attribute. This section describes how to partition the input space into fuzzy regions by the statistical characteristics of the labeled data. Let  $X = \{x_i | x_i = (x_{i1}, x_{i2}, ..., x_{in}), i = 1, 2, ..., s\}$  be a set of *s* training patterns with *n* attributes,  $a_j = \{x_{1j}, x_{2j}, ..., x_{sj}\}$ , (j = 1, 2, ..., n) be a set of the *j*-th input attributes, and  $C = \{c_k | k = 1, 2, ..., m\}$  be a set of labels. Fuzzy regions for each input attribute are determined by the following procedures: **Step 1:** Find the whole domain of the input attribute  $a_j$  and individual domain  $I_{jk}$ ,  $I_{jk} \in c_k$  for each label, as shown in Fig. 2.

**Step 2:** Find an overlapped region  $O_j = [o_j^{L_j}, o_j^{U_j}]$  of  $I_{jk}$  and  $I_{jk+1}$ , where  $o_j^{L_j}$  and  $o_j^{U_j}$  denote the minimum and maximum of every overlapped region, respectively.

**Step 3:** Calculate the number of fuzzy partitions  $P_j$  given by

$$P_{j} = round\left(\left\{\max\left(a_{j}\right) - \min\left(a_{j}\right)\right\} / \alpha \cdot \sigma_{j}\right)$$
(1)

where *round*() denotes a function for rounding-off to the nearest integer, and  $\sigma_j$  is a standard deviation value of the input attribute  $a_j$ . In Eq. (1), the denominator  $\alpha \cdot \sigma_j$  is used to select the different number of fuzzy partitions because the data distribution for each attribute is heterogeneous in the set of labeled patterns. Also the adjustable value of a free parameter  $\alpha$  ( $\alpha > 0$ ) is in the range satisfying the condition  $P_j \ge 2$ . The whole domain of each input attribute is modified as  $D_j^* = [\min(a_j) - W_j, \max(a_j) + W]$  to ensure a variation space, where  $W_j$  is the half-width of fuzzy membership functions as shown in Fig. 3.

$$W_j = \left(o_j^{U_j} - o_j^{L_j}\right) / \left(P_j - 1\right) \tag{2}$$



**Fig.2** An overlapped region  $O_i$  of the input attribute  $a_i$ .



**Fig.3** Fuzzy membership functions defined on overlapping regions when  $P_j = 3$ .

#### 2.2 Rule Generation

Next, we consider the following type of fuzzy rules in the proposed HFCS.

1) First layer (fuzzy system 1)

$$R_k^1$$
: If  $a_1$  is  $A_{k,1}$  and  $a_2$  is  $A_{k,2}$  Then  $y_1$  is  $c_k$ 

where  $R_k^1$  denotes the *k*-th fuzzy rule in first layer (i.e., fuzzy system 1) of HFCS,  $A_{k,1}$  and  $A_{k,2}$  are the linguistic value for the input variable  $a_1$  and  $a_2$ ,  $y_1$  is a dummy output variable, and  $c_k(k = 1, ..., m)$  is the label with a trapezoid-type membership function defined in the domain [0,1], as shown in Fig. 4. In Fig. 4, the parameters  $z_1$ ,  $z_2$ , and  $z_3$  denote the points to define fuzzy membership functions with two linguistic labels.

The compatibility grade of an input pattern  $x_i$  is defined with the antecedent part  $A_k = A_{k,1} \times A_{k,2}$  of the fuzzy rule  $R_k^1$ using the product operator as

$$\mu_k^1(\mathbf{x}_i) = \mu_{k,1}(x_{i1}) \times \mu_{k,2}(x_{i2})$$
(3)

where  $\mu_{k,j}(\cdot)$  is the membership function of the antecedent fuzzy sets  $A_{k,1}$  and  $A_{k,2}$ . The input pattern  $x_i$  is classified by the candidate rule  $R_k^{1*}$  defined as

$$\mu_k^{1*}(\mathbf{x}_i) = \max_k \left\{ \mu_k^1(\mathbf{x}_i) : k = 1, \dots, m \right\}$$
(4)

However, there exist some conflicting rules among all possible candidate rules generated by Eq. (4) when different labels involve the same rules. To overcome this limitation, [12] discussed the method that accepts only the rule with the maximum degree in a conflict group among the generated candidate rules. Although [12] reduces a conflict problem in the generated rules, it is dependent on the positions of fuzzy membership functions defined on each domain interval. Thus, the proposed method uses two criteria, the maximum degree of the candidate rules and its frequency for each label, to minimize the conflicting rules without tuning of parameters to adjust rule weights. Table 1 shows a procedure, the selection of the consequent label, to resolve the conflict problem. In Table 1,  $\alpha = fre(R_k^{1*} \Rightarrow c_p)$  and  $\beta = fre(R_k^{1*} \Rightarrow c_q)$  are the total number of fuzzy rules belonging to p- and q-th label among the overall candidate rules generated by Eq. (4), where  $fre(\cdot)$  denotes a frequency function. Also,  $\gamma = max(R_k^{1*} \Rightarrow c_p)$  and  $\delta = max(R_k^{1*} \Rightarrow c_q)$ denote the maximum degrees among fuzzy rules for the two labels, respectively. 'NA' represents a "not allowable rule"



**Fig.4** Fuzzy membership functions defined on dummy variable *y*<sub>1</sub>.

in case the two criteria are the same.

2) Subsequent layer (fuzzy system  $2 \dots n - 1$ )

$$R_k^{j+1}$$
: If  $y_j$  is  $c_k^*$  and  $a_{j+2}$  is  $A_{k,j+2}$  Then  $y_{j+1}$  is  $c_k$ 

where  $c_k^*$  is the linguistic label with a trapezoid fuzzy membership function, which is newly constructed from the defuzzified outputs  $y_i^*$  of previous layer (see Fig. 6).

$$y_{j}^{*} = \sum_{y_{j}} y_{j} \cdot \mu_{\tilde{k}}(y_{j}) \left| \sum_{y_{j}} \mu_{\tilde{k}}(y_{j}), j = 1, \dots, n-1 \right|$$
  
where  $\mu_{\tilde{k}}(y_{j}) = \max_{k} \left[ \mu_{k}^{j}(\mathbf{x}_{i}) \times \mu_{k}(y_{j}) \right]$  (5)

 $\mu_k^j(\mathbf{x}_i) \times \mu_k(y_j)$  represents the fuzzy implication operation between the antecedent and consequent parts in the fuzzy rule, which is newly defined by Eqs. (3), (4) and Table 1, fired by the input pattern  $\mathbf{x}_i$ .  $\mu_{\bar{k}}(y_j)$  is the fuzzy outputs, i.e., degree of fulfillment, of the rule obtained from the max operation after carrying out the fuzzy implication. The fuzzy partitioning of subsequent layer in HFCS is adjusted by the following procedures:

**Step 1:** Find a boundary point for fuzzy partition from the defuzzified outputs of the previous layer, as shown in Fig. 5.

$$z_{2}^{*} = (T_{c_{1}} + T_{c_{2}}) / 2, \ z_{1}^{*} = z_{2}^{*} - \tau_{j}, \ z_{3}^{*} = z_{2}^{*} + \tau_{j},$$
  
where  $T_{c_{1}} = \max_{y_{j}^{*} \in c_{1}} y_{j}^{*}, \ T_{c_{2}} = \min_{y_{j}^{*} \in c_{2}} y_{j}^{*}$  (6)

Table 1	Two criteria	to resolve th	e conflict	problem.
---------	--------------	---------------	------------	----------

Maximum degree Frequency	$\gamma > \delta$	$\gamma = \delta$	$\gamma < \delta$
$\alpha > \beta$	$c_p$	$c_p$	$c_p$
$\alpha = \beta$	$c_p$	NA	$c_q$
$\alpha < \beta$	$c_q$	$c_q$	$c_q$



Fig. 5 An example of boundary point.



**Fig.6** Fuzzy membership function constructed from the defuzzified outputs of previous layer.

 $\tau_j$  is a standard deviation value for the defuzzifed outputs  $y_j^*, y_j^* \in c_1$  and  $y_j^* \in c_2$  denote the defuzzified output intervals for each label, i.e., its belonging region, in the overall inference result obtained from Eq. (5).

**Step 2:** Construct a new fuzzy membership function based on the boundary point found in Step 1, as shown in Fig. 6.

In this study, we adjusted the parameters of fuzzy membership functions from the defuzzified outputs of previous layer, and this procedure continuous until all the input variables in the proposed HFCS have been used.

# 3. Experimental Results

To show the effectiveness of the proposed HFCS, experiments for the three-type classification methods (statistical classifiers, SVMs with three kernel functions, and fuzzy rule learning methods with rules weights) and the proposed method were performed on five well-known datasets [13], Haberman's survival, Blood transfusion service center, New thyroid disease, Pima Indians diabetes, and Wisconsin breast cancer original. Main characteristics of these datasets are given in Table 2. Each dataset was spilt into 10 groups randomly, and all samples of each class are uniformly assigned to these groups. A group was used as the testing dataset, the other nine groups as training dataset.

During 10-fold cross validation (CV), the combination of inputs in the proposed HFCS was determined according to the degree of redundancy  $N_j \in [0, 1]$  (i.e., ascending order from lower to higher redundancy) calculated from Eq. (7). Then the exhaustive search is used to find the value of pa-

	Table 2	Used datasets.	
Dataset	No. of sample	No. of input feature	No. of class
Hab.	306	3	2
Blood	748	4	2
NT	215	5	3
PID	768	8	2
WBCO	683	9	2

 Table 3
 Comparison results for statistical classifiers and SVMs.

Deterat	Statistical classifiers* (%)		SVMs* (%)			
Dataset	LDA	QDA	kNN	Poly.	RBF	Sig.
Hab.	74.82	75.13	66.97	26.47	70.61	73.53
Blood	76.74	38.10	64.45	23.79	74.60	76.21
NT	91.26	96.75	97.21	13.96	70.74	69.81
PID	77.10	74.38	70.33	65.11	65.11	65.11
WBCO	96.05	95.02	96.05	93.71	89.46	65.01
Avg.	83.19	75.88	79.00	44.61	74.10	69.93
±Sd	±9.74	±23.62	±16.23	±33.66	±9.23	±5.00

\* Experimental results were obtained using KEEL Ver.1.2 software [14].

Datasat	Fuzzy rule learning methods* (%)				
Dataset	Chi_RW Ishibuchi99		Fuzzy slave	Shi et al.	HFCS (%)
(P=2) 73.53 (P=4) 71.23	(P=2) 73.53, (P=3) 73.19,	(P=2) 73.53, (P=3) 73.20,	(P=2) 73.20, (P=3) 74.49,	(P=2) 71.23, (P=3) 71.56,	$\alpha = 1.27, 77.74$
	(P=4) 71.23	(P=4) 73.20	(P=4) 71.55	(P=4) 74.12	(L <sub>1</sub> :17.4, L <sub>2</sub> :5.9)
Dlood	(P=2) 76.07, (P=3) 76.61,	(P=2) 76.21, (P=3) 76.07,	(P=2) 76.47, (P=3) 75.81,	(P=2) 76.74, (P=3) 75.54,	$\alpha = 1.62, 76.20$
Біооц	(P=4) 76.87	(P=4) 76.21	(P=4) 76.47	(P=4) 75.94	$(L_1:22.5, L_2:8.2, L_3:4.0)$
NT	(P=2) 84.24, (P=3) 78.16,	(P=2) 78.61, (P=3) 75.84,	(P=2) 94.46, (P=3) 88.94,	(P=2) 84.26, (P=3) 86.97,	$\alpha = 0.9, 93.03,$
N1 (P=4) 89.85		(P=4) 69.81	(P=4) 79.57	(P=4) 85.11	(L1:35.4, L2:17.2, L3:19.9, L4:18.2)
PID (P=2) 65.76, (P= (P=4) 73.44	(P-2) 65 76 (P-3) 72 40	(D-2) 65 11 $(D-2)$ 64.08	( <b>D</b> -2) 76 71 ( <b>D</b> -2) 74 26	( <b>D</b> -2) 72 78 ( <b>D</b> -2) 71 87	$\alpha = 1.3, 75.01,$
	(P-2) 05.70, $(P-3)$ 72.40, (P-4) 72.44	(P-2) 05.11, $(P-3)$ 04.98, (P-4) 68.24	(P-2) 70.71, $(P-3)$ 74.30,	(P-2) 72.78, $(P-3)$ 71.87,	$(L_1:17.5, L_2:8,7, L_3:10.5, L_4:12.4,$
	(r-4) 75.44	(r-4) 08.24	(r-4)/2.14	(r-4) 72.50	L <sub>5</sub> :11.1, L <sub>6</sub> :9.4,L <sub>7</sub> :8.6)
					$\alpha = 1.2, 95.76,$
WBCO	(P=2) 94.29, (P=3) 91.21,	(P=2) 95.17, (P=3) 96.34,	(P=2) 96.50, (P=3) 95.03,	(P=2) 94.58, (P=3) 95.47,	(L1:8.2, L2:8.4, L3:6.0, L4:4.9, L5:6.8,
	(P=4) 73.21	(P=4) 95.76	(P=4) 96.63	(P=4) 93.13	L <sub>6</sub> :6.0,
					L <sub>7</sub> :4.6, L <sub>8</sub> :5.2)
Avg.±Sd	(P=2) 78.78 ± 10.89	(P=2) 77.73 ± 11.00	(P=2) 83.47 ± 11.08	(P=3) 80.28 ± 10.54	83.55 ± 10.00

Table 4 Comparison results for HFCS and fuzzy rule learning methods.

rameter  $\alpha$  as described in Eq. (1).

$$N_j = \sum_{\mathbf{x}_i \in c_k} h\left(x_{ij}\right) / s, \ x_{ij} \in a_j, \ j = 1, \dots, n$$
(7)

where h(.) represents the number of occurrence frequency when the attribute values  $x_{ij}(i = 1, ..., s)$  correspond to more than two labels of *m* class labels. *s* is the total number of data points, i.e., attribute values. Tables 3 and 4 show the average classification accuracy rates of testing during 10fold CV. In Tables 3 and 4, the parameter values of SVMs (i.e., kernel types: Polynomial, RBF, and Sigmoid functions; c: 1000; eps: 0.001; degree(d): 10; gamma(g): 1.0; coef0(r): 1.0; nu(n): 0.5; p: 1.0; and shrinking(h): 0.0) are used as the default values provided by KEEL Ver.1.2 [14]. In each Table, 'Avg±Sd' denotes the average performance of each classification method for five datasets. In Table 4, 'P' denotes the number of fuzzy partitions and 'L<sub>i</sub>' denotes the average number of rules in each layer of the HFCS.

For five datasets, we observed that the classification performance of the proposed HFCS was maximized when the values of parameter  $\alpha$  were updated as 1.27, 1.62, 0.9, 1.3, and 1.2. Moreover, we confirmed that the proposed HCFS with the different number of fuzzy partitions produces better average testing performance than those of the three-type classification methods for five datasets, even though it has slightly lower classification accuracy rate than that of Fuzzy slave (P=2) with the highest classification performance among the fuzzy rule learning methods. The results showed that the performance in HFCS is strongly dependent on the number of fuzzy partitions defined in overlapping regions.

# 4. Conclusions

In this letter, we have proposed a scheme for designing a hierarchical fuzzy classification system based on statistical characteristics of the labeled data. In order to show the influence of the different number of fuzzy partitions in overlapping regions, we have presented the classification performances according to the parameter  $\alpha$  (i.e.,  $\alpha \cdot \sigma_j$ ) in Eq. (1) on five datasets. The effectiveness of the proposed method was demonstrated by comparing the experimental results of the five datasets using the proposed scheme and four methods. The problem on the determination of parameter  $\alpha$  without using the exhaustive search remains to be solved in further studies.

## Acknowledgment

The authors wish to thank both the reviewer and the editor for their invaluable advice in revising this paper.

#### References

- S.-M. Chen and F.-M. Tsai, "Generating fuzzy rules from training instances for fuzzy classification systems," Expert Systems with Applications, vol.35, no.3, pp.611–621, 2008.
- [2] A. Quteishat and C.P. Lim, "A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification," Applied Soft Computing, vol.8, no.2, pp.985–995, 2008.
- [3] H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule learning," Inf. Sci., vol.136, no.1-4, pp.109–133, 2001.
- [4] J. Casillas, O. Cordon, M.J. Del Jesus, and F. Herrera, "Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems," Inf. Sci., vol.136, no.1-4, pp.135–157, 2001.

- [5] P. Salgado, "Rule generation for hierarchical collaborative fuzzy systems," Appl. Math. Modell., vol.32, no.7, pp.1159–1178, 2008.
- [6] G.V.S. Raju, J. Zhou, and R.A. Kisner, "Hierarchical fuzzy control," Int. J. Control, vol.54, no.5, pp.1201–1216, 1991.
- [7] L.X. Wang, "Universal approximation by hierarchical fuzzy systems," Fuzzy Sets Syst., vol.93, no.2, pp.223–230, 1998.
- [8] C. Wei and L.X. Wang, "A note on universal approximation by hierarchical fuzzy systems," Inf. Sci., vol.123, no.3-4, pp.241–248, 2000.
- [9] X.-J. Zeng and J.A. Keane, "Approximation capabilities of hierarchical fuzzy systems," IEEE Trans. Fuzzy Syst., vol.13, no.5, pp.659–

672, 2005.

- [10] M.R. Delgado, F.V. Zuben, and F. Gomide, "Hierarchical genetic fuzzy systems," Inf. Sci., vol.136, no.1, pp.29–52, 2001.
- [11] M.L. Lee, H.Y. Chung, and F.M. Yu, "Modeling of hierarchical fuzzy systems," Fuzzy Sets Syst., vol.138, no.2, pp.343–361, 2003.
- [12] L.-X. Wang and J.M. Mendel, "Generating fuzzy rules by learning from examples," IEEE. Trans. Syst. Man Cybern., vol.22, no.6, pp.1414–1427, 1992.
- [13] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/
- [14] KEEL software, http://sci2s.ugr.es/keel/