

Combining Interpretable Fuzzy Rule-Based Classifiers via Multi-Objective Hierarchical Evolutionary Algorithm

Jingjing Cao

Department of Computer Science
City University of Hong Kong
Kowloon, Hong Kong
jingcao3@student.cityu.edu.hk

Hanli Wang

College of Electronic and Information Engineering
Tongji University
Shanghai, China
hanli.wang@gmail.com

Sam Kwong

Department of Computer Science
City University of Hong Kong
Kowloon, Hong Kong
cssamk@cityu.edu.hk

Ke Li

Department of Computer Science
City University of Hong Kong
Kowloon, Hong Kong
keli.genius@gmail.com

Abstract—The contributions of this paper are two-fold: firstly, it employs a multi-objective evolutionary hierarchical algorithm to obtain a non-dominated fuzzy rule classifier set with interpretability and diversity preservation. Secondly, a reduce-error based ensemble pruning method is utilized to decrease the size and enhance the accuracy of the combined fuzzy rule classifiers. In this algorithm, each chromosome represents a fuzzy rule classifier and compose of three different types of genes: control, parameter and rule genes. In each evolution iteration, each pair of classifiers in non-dominated solution set with the same multi-objective qualities are examined in terms of Q statistic diversity values. Then, similar classifiers are removed to preserve the diversity of the fuzzy system. Finally, experimental results on the ten UCI benchmark datasets indicate that our approach can maintain a good trade-off among accuracy, interpretability and diversity of fuzzy classifiers.

Index Terms—Multi-objective evolutionary algorithm (MOEAs), Fuzzy rule-based systems (FRBCs), Interpretability, Ensemble diversity, Ensemble pruning.

I. INTRODUCTION

The fuzzy rule-based systems (FRBCs) are promising due to they relate to the logic of human beings and the rules they produced are easy for human to understand. To generate an understandable fuzzy systems, the concept of interpretability is taken into account to preserve semantic meaning and complexity of FRBCs. For decades, most of researchers devote to find a good trade-off between interpretability and accuracy associated with fuzzy model ([13], [23]). With regard to these two criteria, multi-objective evolutionary algorithms (MOEAs) are utilized as effective global search approaches to optimize these objectives simultaneously ([6], [22]). However, by means of MOEAs, a set of non-dominated solutions with respect to different objectives, instead of a single fuzzy rule classifier are obtained. In this case, the problem of selecting the suitable and reliable solutions as final classifier has given rise to a growing interest by many researchers.

Ensemble approaches are widely employed to design reliable classifiers by fusing or selecting multiple classifiers. The two most popular ensemble methods are bagging [1] and boosting [16], where adaptive boosting, i.e., AdaBoost is the most well known boosting algorithm [5] in the field. The ensemble methods have been proved to be effective approaches to improve the performance of classification prediction. However, they suffer the problems on how to guarantee each classifier training different aspects of data and how to select useful classifiers among all component classifiers. Thus, ensemble diversity and pruning become hot issues to investigate the classifier ability in ensemble field. Though no general theoretical framework is proposed, diversity issue is still considered as an important part to construct successful ensemble classifiers. And pruning is presented mainly based on the assumption that only appropriate subensemble learner set can give rise to the better classification accuracy than complete ensemble classifiers.

Currently, several ensemble methods have been introduced in the design of MOEAs based FRBCs because a large number of classifiers could be achieved after each run of MOEAs. Yusuke et al. [14] designed a multi-classifier coding scheme on chromosome and worked on an entropy-based diversity objective in genetic searching process. Further, both non-pair-wise and pair-wise diversity measures have been utilized to examined the classification performance. However, their coding scheme limits the number of classifiers and no attention is paid to including ensemble pruning step in algorithm. Hisao et al. [9] presented a fuzzy rule-based ensemble classifier by utilizing evolutionary multi-objective optimization algorithm and received a good trade-off between interpretability and accuracy of the model. Whereas the method only combined the non-dominated fuzzy rule-based classifiers without considering ensemble diversity and pruning. In our approach, the proposed chromosome encoding structure relaxes the restriction of clas-

sifier number, and the diversity between any two classifiers is examined in non-dominated solution set which not only ensures the complete diversity among non-dominated fuzzy rule-based set, but also edits the similar classifiers to make final hypothesis more compact.

In this paper, we use the multi-objective hierarchical evolutionary algorithm (MOHEA) to generate a non-dominated fuzzy rule-based classifier set. In the optimization process, low-level and high-level interpretability are both taken into account to enhance the interpretability of FRBCs. Additional, ensemble diversity and pruning methods are utilized to handle the non-dominated rule sets. The Q statistic diversity scheme maps the metric distance between two individuals with equal objectives to solution space, which reveals intrinsic quality of their classification performance. The reduce-error based pruning method decreases the size of non-dominated rule set and improves the classification performance simultaneously. Therefore, when a final classifier is acquired, a good trade-off among accuracy, interpretability and diversity of the classifier is preserved.

This paper is organized as follows. The preliminary knowledge on interpretability and diversity of FRBCs is presented in the next section. Section 3 briefly elaborates on the description of a multi-objective hierarchical evolutionary algorithm (MOHEA). Interpretability and ensemble diversity preservation methods for MOHEA is presented in Section 4. In Section 5, the proposed method is applied to 10 UCI data for quantifying the performance of the classifier. Finally, the last section conclude the work of the paper.

II. PRELIMINARY KNOWLEDGE ON INTERPRETABILITY AND DIVERSITY OF FRBCS

Reviews in [13] and [23] categorized the interpretability measures from different perspectives. Mencar et al. [13] discussed interpretability constraints in terms of fuzzy sets, frames of cognition, fuzzy information granules, fuzzy rules, fuzzy models and fuzzy model adaptation. While Zhou et al. [23] divided the interpretability into two groups: low-level and high-level interpretability. Low-level interpretability criteria are obtained on the fuzzy set level and high-level criteria are achieved on the fuzzy rule level. Further, the pair-wise diversity measure based on classifier level is also introduced.

A. Low-Level Interpretability

Completeness and Distinguishability: The concepts of completeness and distinguishability of fuzzy systems are usually expressed through a fuzzy similarity measure in ([2], [8], [17]). This similarity measure can be interpreted in many different ways depending on the application context. However, an important definition is given in [8]: Similarity between fuzzy sets is defined as the degree to which fuzzy sets are equal. In fact, if the similarity of two neighboring fuzzy sets is zero or too small, it means that the fuzzy partitioning in this fuzzy variable is incomplete or the two fuzzy sets do not overlap enough. On the other hand, if the similarity is too big,

then it indicates that the two fuzzy sets overlap too much, and the distinguishability between them is poor.

In the following, let A and B be two fuzzy sets of fuzzy variables x (on the universe U) with the membership functions $u_A(x)$ and $u_B(x)$, respectively. The symbol s represents the similarity value of these two fuzzy sets $s = S(A, B)$, $s \in [0, 1]$. We use the following similarity measure between fuzzy sets [17]:

$$S(A, B) = \frac{M(A \cap B)}{M(A \cup B)} = \frac{M(A \cap B)}{M(A) + M(B) - M(A \cap B)} \quad (1)$$

where $M(A)$ denotes the cardinality of the fuzzy set A , and the operators \cap and \cup represent the intersection and union, respectively. There are several methods to calculate the similarity. In our approach, we use the following form to calculate the similarity of fuzzy sets because it is computationally simple and effective ([15], [18]):

$$S(A, B) = \frac{\sum_{j=1}^m [u_A(x_j) \wedge u_B(x_j)]}{\sum_{j=1}^m [u_A(x_j) \vee u_B(x_j)]} \quad (2)$$

on a discrete universe $U = x_j | j = 1, 2, \dots, m$. \wedge and \vee in (2) are the minimum and maximum operators.

B. High-Level Interpretability

1) *Compactness*: A compact fuzzy system means that it has the minimal number of fuzzy antecedents and fuzzy rules. In addition, the number of fuzzy variables is also worth being considered. A compact fuzzy system is always desirable when the number of input variables increases.

2) *Completeness*: Considering the fuzzy systems with rule consequent defined as class labels, every class label should be included in at least one rule of the fuzzy system. This constraint is seldom noticed as an interpretability aspect due to it is often satisfied in fuzzy genetic systems. But in our work, it is an essential interpretability constraint. Suppose we have a three class problem with two input variables, then take the following fuzzy system with three rules for example:

R_1 : If x_1 is small and x_2 is large, Then Class 1;

R_2 : If x_1 is large, Then Class 3;

R_3 : If x_1 is middle and x_2 is middle, Then Class 1;

The above fuzzy system is called not completeness because none rule would be activated for class 2.

C. Diversity among Ensemble Classifiers

As for fuzzy ensemble classifiers, the diversity issue has been interested by researchers ([11]). For a typical classification problem on data set $X = \{X_1, X_2, \dots, X_n\}$, suppose D_i denotes the i th classifier with vector $y_i = [y_{1,i}, y_{2,i}, \dots, y_{n,i}]^T$ such that, $y_{j,i} = 1$ if classifier D_i correctly classified pattern X_j , otherwise, $y_{j,i} = 0$. In the following, N^{pq} is the number of X_i which satisfied that $y_{j,i} = p$ and $y_{j,i} = q$.

$$Q_{i,k} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (3)$$

Q Statistic is a common utilized pair-wise diversity measure. Through formula (3), it implies that the lower the value, the more diverse the classifiers. Only two statistically independent classifiers could obtain $Q = 0$. The formula is proposed based upon the principle that classifiers that tend to recognize the same objects correctly will have positive values of Q , and those which commit errors on different objects will render Q negative.

III. PROPOSED MULTI-OBJECTIVE HIERARCHICAL EVOLUTIONARY ALGORITHM

This section will provide a specific description of a multi-objective hierarchical evolutionary algorithm (MOHEA) based on our previous works ([21], [20], [19]). Firstly a hierarchical chromosome formulation with three types of genes are used to represented a individual classifier. In each evolution iteration, not only interpretability preservation methods are applied to produce valid offsprings and reduce redundancy of the fuzzy system, but also the ensemble diversity measure is employed to construct diverse and compact non-dominated solution set.

A. Hierarchical chromosome coding and rule structure

A hierarchical chromosome consists of three types of genes: control genes, parameter genes and rule genes. The activation of the parameter genes is governed by the value of the control genes. When a control gene is “1”, then the corresponding parametric gene is activated. Otherwise, it is deactivated.

For each fuzzy variable x_i , we determine the possible maximal number of fuzzy sets M_i so that it can sufficiently represent this fuzzy variable. For N dimensional problems, there are totally possible fuzzy sets. So there are control genes coded as bits 0 or 1, where 1 is assigned to represent that the corresponding parameter gene, which is dominated by this control gene, is selected for involvement in an evolutionary process; otherwise, 0 is for turning off. We apply the Gaussian combinational membership functions (abbreviated as Gauss2mf) to depict the antecedent fuzzy sets (i.e., a combination of two Gaussian functions). The Gauss2mf function depends on four parameters σ_1 , c_1 , σ_2 and c_2 as given by

$$\mu(x; \sigma_1, c_1, \sigma_2, c_2) = \begin{cases} \exp\left[\frac{-(x-c_1)^2}{2\sigma_1^2}\right] : x < c_1 \\ 1 : c_1 \leq x \leq c_2 \\ \exp\left[\frac{-(x-c_2)^2}{2\sigma_2^2}\right] : c_2 < x \end{cases} \quad (4)$$

where σ_1 and c_1 determine the shape of the leftmost curve. The shape of the rightmost curve is specified by σ_2 and c_2 . So we use the parameter list $[\sigma_1, c_1, \sigma_2, c_2]$ to represent one parameter gene(i.e., a fuzzy set expressed in the form of a Gauss2mf membership function).

Rule structure is typically shown as a relational matrix, a decision table or a list of rules ([3]). Unlike rule structure coding in ([21]), this paper adopted rule structure ([20], [19]) and added it into chromosome as a cell called rule genes. That is, each individual of the fuzzy rule set population, it is represented as a concatenated string of the length $N \times N_{rule}$, where N_{rule} is a predefined integer to describe the size of

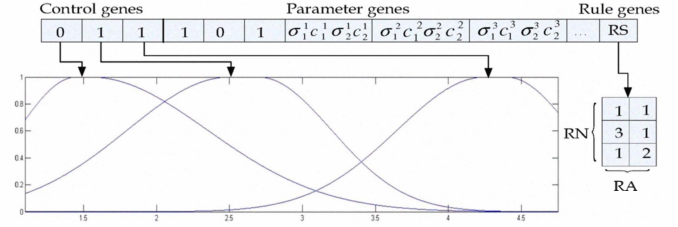


Fig. 1. Proposed chromosome structure with three types of genes

the initial fuzzy rule base. The fuzzy rule sets are randomly initialized so that the cell value of the concatenated string represents one of the fuzzy sets of the corresponding fuzzy variable or is equal to zero indicating “don’t care” conditions. In this work the range of N_{rule} is set as $[3, 10]$ in evolutionary process.

Fig.1 shows an example of our hierarchical formulation. The control genes of chromosome represent two input variables x_1 and x_2 with three fuzzy sets each (A_{ij} denotes the j th fuzzy set of the i th input variable). The first three control genes 0 1 1 represent the corresponding fuzzy sets are activated. For the rule genes RS, it contains a rule matrix, which each row represents a rule, and each column denotes an input variable. Giving the cell values, three rule antecedents are extracted as (1) if x_1 is A_{11} and x_2 is A_{21} ; (2) if x_2 is A_{21} ; (3) if x_1 is A_{11} and x_2 is A_{22} . Rules (2) only has x_2 because for input variable x_1 , only two fuzzy sets are activated, so 3 represents “don’t care” case.

B. Calculation of Rule Consequents

For the classification problems, the heuristic procedure ([10]) is applied to generate rule consequents from the training patterns. Suppose the classification problem has c class in the n -dimensional pattern space. The rule could be expressed in the following form:

R_i : If x_1 is A_{i1} and, ..., and x_n is A_{in} ,

Then Class C_i with $CF = CF_i$,

where R_i represents the i th fuzzy rule, A_{i1}, \dots, A_{in} are the antecedent fuzzy sets, C_i is the consequent class(i.e., one of the given c classes), and CF_i is the grade of certainty of R_i . For each class, calculate the sum of the compatibility grades of the training patterns $X^i = [x_1^i, x_2^i, \dots, x_n^i]$ for the rule R_j as:

$$\beta_{Class\ h}(R_j) = \sum_{X^i \in Class\ h} u_j(X^i), \quad h = 1, 2, \dots, c, \quad (5)$$

where $\beta_{Class\ h}(R_j)$ is the sum of the compatibility grades of the training patterns in Class h for R_j . Find Class \hat{h}_j that has the maximum value of $\beta_{Class\ h}(R_j)$ as the consequent class for R_j :

$$\beta_{Class\ \hat{h}_j}(R_j) = \max\{\beta_{Class\ 1}(R_j), \dots, \beta_{Class\ c}(R_j)\}, \quad (6)$$

If the maximum value of \hat{h}_j can not be uniquely specified, that is, there are some classes that have the same maximum value, the fuzzy rule R_j is removed from the rule base. Then we calculate the certainty factor CF_j as follows:

$$CF_j = \frac{\beta_{Class \hat{h}_j(R_j)} - \bar{\beta}}{\sum_{h=1}^c \beta_{Class h(R_j)}}, \quad (7)$$

$$\text{where } \bar{\beta} = \frac{\sum_{h \neq \hat{h}_j} \beta_{Class h(R_j)}}{c-1}$$

C. Crossover and Mutation

From the structure of proposed hierarchical encoding, the crossover and mutation operators are different for each type of genes in the chromosome. With regards to the different functions of each type of genes, crossover are not operate on hierarchical (control, parameter) genes and rule genes simultaneously. They share equal chance to do crossover for all individuals in current population. For control genes, multi-point crossover is applied considering the reason of simplicity. As for the parameter genes which are represented in real numbers, BLX- α crossover [4] is applied because BLX- α (in particularly $\alpha = 0.5$) crossover has turned out to be the best crossover operators for the real-coded GA based on the experiment results reported in [7]. One-point crossover is used on rule genes ([20]). The crossover operation randomly selects a different cut-off point for each parent to generate offspring rule sets.

Similarly, multi-point mutation and non-uniform mutation are applied to the control and parameter genes respectively. For rule genes, a mutation operator randomly replaces each element of the rule sets string with other linguistic value in a certain probability is satisfied. Elimination of existing rules and addition of new rules can also be used as mutation operations. Such mutation operators change the number of rules in the rule sets string. Note that the crossover and mutation operations may generate equal rules. In this case, the consistency of fuzzy systems is guaranteed by deleting equal rules.

D. Combining Interpretable Fuzzy Classifiers in Non-dominated Solution Set

When the proposed algorithm satisfies the stopping condition, a non-dominated solution set containing many classifiers are obtained. To combine these classifiers effectively, the reduce-error pruning method [12] is presented to prune the obtained classifier set. The reduce-error pruning method is utilized in ensemble field which sequently added the classifier with lowest classification error. Finally, a subensemble with lowest training error can be constructed. The relationship between subensemble S_k and S_{k-1} is defined as follows:

$$S_k = S_{k-1} \cup \arg \max_{\theta} \sum_{(\mathbf{x}, y) \in Z_{train}} I(H_{S_{k-1} \cup D_{\theta}}(\mathbf{x}) = y), \quad (8)$$

where $\theta \in X \setminus S_{k-1}$. $I(\cdot)$ is a indicator function, the result is 1 if $H_{S_{k-1} \cup D_{\theta}}(\mathbf{x}) = y$, else it returns 0.

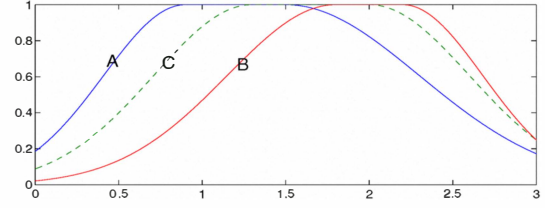


Fig. 2. Merging A and B to create C

IV. INTERPRETABILITY AND DIVERSITY PRESERVATION METHODS FOR MOEA

A. Interpretability preservation at fuzzy sets base level

1) *Merging Similar Fuzzy Sets:* An example of the similarity measure between two fuzzy sets is given as in (3). If the similarity value is greater than a given threshold, then we merge these two fuzzy sets to generate a new one. Considering two fuzzy sets A and B with the membership functions $u_A(x; a_1, a_2, a_3, a_4)$ and $u_B(x; b_1, b_2, b_3, b_4)$, the resulting fuzzy set C with the membership function $u_C(x; c_1, c_2, c_3, c_4)$ is defined from merging A and B by

$$\begin{aligned} c_1 &= \min(a_1, b_1), c_2 = \lambda_2 a_2 + (1 - \lambda_2) b_2 \\ c_3 &= \lambda_3 a_3 + (1 - \lambda_3) b_3, c_4 = \min(a_4, b_4). \end{aligned} \quad (9)$$

The parameters $\lambda_2, \lambda_3 \in [0, 1]$ determine the relative importance about the influence of the fuzzy sets A and B have on C . The threshold for merging similar fuzzy sets plays an important role in the improvement of interpretability. According to our experience, values in the range $[0.4, 0.7]$ may be a good choice. In our approach, we set the threshold equal to 0.55. Fig. 2 illustrates the case for merging A and B to create C .

2) *Removing Fuzzy Sets Similar to the Universal Set or Similar to a Singleton Set:* If the similarity value of a fuzzy set to the universal set $U(u_U(X) = 1)$ is greater than an upper threshold or smaller than a lower threshold, then we can remove it from the rule base. In the first case, the fuzzy set is very similar to the universal set and in the latter case, similar to a singleton set. Neither case is desirable for interpretable rule base generation. We set $\theta_U = 0.9, \theta_s = 0.05$ in this work.

B. Interpretability preservation at fuzzy rules base level

1) *Generating Rule Set with Completeness:* For each new generated chromosome, it should involve at least one rule for each class. However, the rule generation process of evolutionary algorithm may not always satisfies this constraint and changes to invalid classifier. To deal with this problem, a regeneration operator is adopted by producing valid rule set by repeating generating chromosome R times if classifier does not contain the completeness rule for each class. If the chromosome still does not satisfy the constraint, regeneration operator stops. This operator is applied after creating initial population, crossover and mutation operators.

2) *Keeping the Compactness by Multi-objective Evaluation Criteria*: The proposed MOGHA uses the following three criteria to evaluate fuzzy rule set candidates and preserve the compactness of fuzzy rules simultaneously: (1) Accuracy: the accuracy is measured in terms of mean-squared error (MSE) for function approximation problems and classification error rates for classification problems; (2) the number of fuzzy rules; (3) the average length of fuzzy rules: the total number of the rule antecedents displayed in the rule base divide by the number of fuzzy rules.

C. Diversity preservation at fuzzy classifiers base level

In each iteration, a non-dominated solution set is updated by utilizing the selection mechanism of the NSGA-II algorithm. In order to compare the fuzzy rule set candidates, we predefine the preference for the three criteria. The accuracy is predefined the first priority and the other two criteria about the interpretability are predefined the same second priority. In other words, we first compare two fuzzy rule set candidates according to the accuracy only. If these two candidates have the same accuracy level based on our preference, then we compare the other two criteria to determine which rule set candidate is better.

However, for the two candidates have the same values on all three criteria, the selection mechanism of the NSGA-II algorithm does not consider this case due to it is sparsely happened in continuous space. While in fact, It is common happened for the fuzzy genetic system. Actually, even two classifiers have the same value on these criteria, they may represent a complete different fuzzy systems and misclassified different patterns. Thus, to handle this situation and enhance the diversity for the final non-dominated solution set, Q statistic diversity measure is employed. That is, if S_1 and S_2 have the same objective values, we maps them to the class label decision space and calculate their Q statistic value δ . If δ is higher than a threshold δ_0 , the two classifiers are considered similar and one of them is deleted randomly. It should be noted that if S_1 and S_2 are the same classifier, we remove the same classifier to preserve compactness as well. In our work $\delta_0 = 0.9$.

V. EXPERIMENT SETUP AND COMPARISON RESULT

In order to estimate the classification performance of the presented algorithm, our experiments are tested on ten UCI machine learning repository benchmarks (<http://www.ics.uci.edu/mllearn/>), where TABLE I provide the basic data information. Also, a 5x2 cross-validation is conducted in this paper by repeatedly running two-fold cross-validation five independent iterations. It should be noted that 5x2cv is normally employed to evaluate the effect of ensemble approaches. Matlab 7.9 is applied to implement the experiments.

TABLE II and III collect the average and standard derivation of testing MSE in simulation experiment by comparing KNN(K-Nearest Neighbour algorithm), Bagging KNN and AdaBoost KNN with reduce-error pruning method employed, the proposed MOHEA selects only one solution with the lowest training MSE and proposed MOHEA with ensemble

TABLE I
DATASET INFORMATION

Data	Pattern #	Feature #	Class #
Australian	690	14	2
Breast	699	9	2
Cmc	1473	9	3
German	1000	24	2
Glass	214	9	6
Heart	270	13	2
Iris	150	4	3
Pima	768	8	2
Wdbc	569	30	2
Wdbc	198	30	2

TABLE II
AVERAGE OF TESTING MSE BY COMPARING KNN, BAGGING KNN, ADABOOST KNN AND THE PROPOSED MOEHA AND ENSEMBLE MOEHA ALGORITHM

Data	KNN	BagKNN	AdaKNN	MOEHA	EnsMOEHA
Iris	0.0520	0.0507	0.0587	0.0507	0.0440
Beast	0.0444	0.0441	0.0484	0.0506	0.0372
German	0.3190	0.3184	0.3272	0.2984	0.2994
Heart	0.2259	0.2222	0.2326	0.2519	0.2222
Pima	0.3003	0.3003	0.3102	0.2563	0.2471
Wdbc	0.0450	0.0436	0.0503	0.0801	0.0742
Wdpc	0.3212	0.3111	0.3374	0.2364	0.2323
Aus.	0.1980	0.1893	0.2064	0.1530	0.1449
Glass	0.5720	0.5709	0.5724	0.5192	0.5157
Bupa	0.4133	0.4133	0.4116	0.4023	0.4116

TABLE III
STANDARD DEVIATION OF TESTING MSE BY COMPARING KNN, BAGGING KNN, ADABOOST KNN AND THE PROPOSED MOEHA AND ENSEMBLE MOEHA ALGORITHM

Data	KNN	BagKNN	AdaKNN	MOGHA	EnsMOGHA
Iris	0.0087	0.0087	0.0060	0.0128	0.0037
Breast	0.0042	0.0036	0.0037	0.0071	0.0054
German	0.0123	0.0088	0.0099	0.0155	0.0032
Heart	0.0275	0.0258	0.0269	0.0284	0.0168
Pima	0.0129	0.0141	0.0142	0.0091	0.0063
Wdbc	0.0044	0.0038	0.0058	0.0190	0.0210
Wdbc	0.0201	0.0146	0.0176	0.0097	0.0071
Aus.	0.0107	0.0102	0.0087	0.0090	0.0057
Glass	0.0096	0.0078	0.0110	0.0057	0.0087
Bupa	0.0240	0.0238	0.0191	0.0285	0.0394

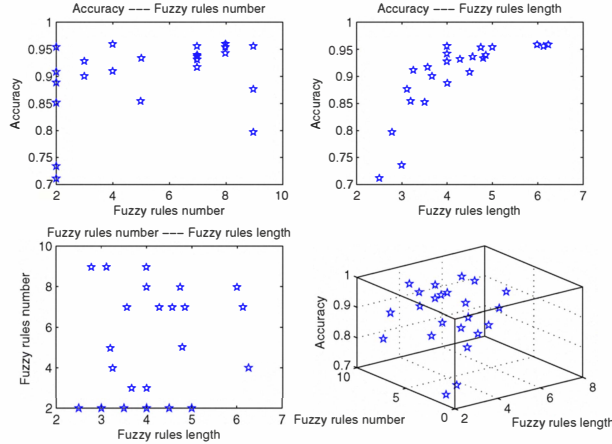


Fig. 3. Non-dominated Fuzzy Rule-Based Set of Breast Dataset

TABLE IV
NON-DOMINATE FUZZY RULE-BASED CLASSIFIERS WITH ENSEMBLE PRUNING FOR BREAST DATASET

No.	Accuracy	Input	No.Rules	Rule.leng.
1	0.9600	7{0,3,1,2,3,3,0,3,2}	8	6.00
2	0.9600	8{2,1,2,2,0,2,1,2,3}	4	6.25
3	0.9571	8{2,1,1,1,0,1,3,1,1}	9	4.00
4	0.9571	8{3,0,1,2,2,3,1,1,2}	7	6.14

scheme. From TABLE II, we can see that the proposed ensemble MOHEA has lowest error rate on test for most data sets(7/10). The MOHEA without ensemble outperforms other methods on two datasets while Bagging KNN with pruning also outperforms others in two classifier. KNN and AdaBoost KNN with pruning show the worst result on all datasets. Table III represents similar results as TABLE II, which illustrates that the proposed ensemble MOHEA is more reliable than others.

Fig.3 shows the trade-off among multi-objectives of the non-dominated solutions on breast dataset. The upper left figure illustrates the trade-off between the accuracy and fuzzy rules number, the upper right figure shows the trade-off between the accuracy and fuzzy rules length, the lower left figure is for the trade-off between the accuracy and fuzzy rules length, and the lower right figure represents the trade-off among the three objectives: accuracy, fuzzy rules number, and fuzzy rules average length. The TABLE IV shows the corresponding non-dominate fuzzy rule-based classifiers selected through ensemble pruning.

VI. CONCLUSION

This paper proposes a multi-objective evolutionary hierarchical algorithm with ensemble measures. Firstly, three different types of genes chromosome structure is employed. Then, ensemble diversity and pruning methods are utilized to handle the non-dominated rule sets. Besides, several genetic operations are proposed to preserved low-level and high-level

interpretable of the FRBCs.

REFERENCES

- [1] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [2] G. Castellano, A.M. Fanelli, and T. Roselli. A GA-based approach to optimize fuzzy models learned from data. 2002.
- [3] O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena, O. Cordon, F. Herrera, and F. Hoffmann. *Genetic fuzzy systems*. World Scientific, 2001.
- [4] L.J. Eshelman. Real-coded genetic algorithms and interval-schemata. *Foundations of genetic algorithms*, 2:187–202, 1993.
- [5] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.
- [6] M.J. Gacto, R. Alcalá, and F. Herrera. Integration of an index to preserve the semantic interpretability in the multiobjective evolutionary rule selection and tuning of linguistic fuzzy systems. *Fuzzy Systems, IEEE Transactions on*, 18(3):515–531, 2010.
- [7] F. Herrera, M. Lozano, and J.L. Verdegay. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, 12:265–319, 1998.
- [8] H. Ishibuchi, T. Nakashima, and T. Kuroda. A hybrid fuzzy GBML algorithm for designing compact fuzzy rule-based classification systems. In *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on*, volume 2, pages 706–711. IEEE, 2000.
- [9] H. Ishibuchi and Y. Nojima. Evolutionary multiobjective optimization for the design of fuzzy rule-based ensemble classifiers. *International Journal of Hybrid Intelligent Systems*, 3(3):129–145, 2006.
- [10] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *Fuzzy Systems, IEEE Transactions on*, 3(3):260–270, 1995.
- [11] L.I. Kuncheva and C.J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2):181–207, 2003.
- [12] Gonzalo Martínez-Muñoz, Daniel Hernández-Lobato, and Alberto Suárez. An analysis of ensemble pruning techniques based on ordered aggregation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):245–259, 2009.
- [13] Corrado Mencar and Anna Maria Fanelli. Interpretability constraints for fuzzy information granulation. *Inf. Sci.*, 178(24):4585–4618, 2008.
- [14] Y. Nojima and H. Ishibuchi. Effects of diversity measures on the design of ensemble classifiers by multiobjective genetic fuzzy rule selection with a multi-classifier coding scheme. *Hybrid Artificial Intelligence Systems*, pages 755–763, 2008.
- [15] H. Roubos and M. Setnes. Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *Fuzzy Systems, IEEE Transactions on*, 9(4):516–524, 2001.
- [16] R.E. Schapire. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.
- [17] M. Setnes, R. Babuska, U. Kaymak, and H.R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 28(3):376–386, 1998.
- [18] M. Setnes and H. Roubos. Ga-fuzzy modeling and classification: complexity and performance. *Fuzzy Systems, IEEE Transactions on*, 8(5):509–522, 2000.
- [19] H. Wang, S. Kwong, Y. Jin, and C.H. Tsang. Agent based multi-objective approach to generating interpretable fuzzy systems. *Multi-Objective Machine Learning*, pages 339–364, 2006.
- [20] H. Wang, S. Kwong, Y. Jin, W. Wei, and K.F. Man. Agent-based evolutionary approach for interpretable rule-based knowledge extraction. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(2):143–155, 2005.
- [21] H. Wang, S. Kwong, Y. Jin, W. Wei, and KF Man. Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. *Fuzzy sets and systems*, 149(1):149–186, 2005.
- [22] Y. Zhang, X. Wu, Z. Xing, and W.L. Hu. On generating interpretable and precise fuzzy systems based on pareto multi-objective cooperative co-evolutionary algorithm. *Applied Soft Computing*, 11(1):1284–1294, 2011.
- [23] Shang-Ming Zhou and John Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159(23):3091–3131, 2008.