A Multi-objective GA-based Fuzzy Modeling Approach for Constructing Pareto-optimal Fuzzy systems

Xing Zong-Yi, Hou Yuan-Long, Zhang Yong, Jia Li-Min

Abstract— **An approach to construct multiple Pareto-optimal fuzzy systems based on a multi-objective genetic algorithm is proposed in this paper. First, in order to obtain a good initial fuzzy system, a modified fuzzy clustering algorithm is used to identify the antecedents of fuzzy system, while the consequents are designed separately to reduce computational burden. Second, a Pareto multi-objective genetic algorithm based on NSGA-II and the interpretability- driven simplification techniques are used to evolve the initial fuzzy system iteratively with three objectives: the precision performance, the number of fuzzy rules and the number of fuzzy sets. Resultantly, multiple Pareto- optimal fuzzy systems are obtained. The proposed approach is applied to two benchmark problems, and the results show its validity.**

Keywords— **Fuzzy modeling; fuzzy system; multi-objective genetic algorithm; Pareto-optimal; interpretability**

I. INTRODUCTION

FUZZY sets Theory, introduced by Professor Zadeh [1] thirty years ago, has been received more and more attention from researchers in a wide range of areas. Fuzzy modeling is one of the most successful disciplines that is often used in classification, data mining, simulation, prediction and control [2]. Fuzzy system can be designed based on expert knowledge; however it is difficult to acquire adequate and efficient expert knowledge for complex problems, so several approaches have been proposed to build fuzzy system from numerical data, including fuzzy clustering-based algorithms [3], neuro-fuzzy systems [4,5] and genetic fuzzy rules generation [6,7]. However all these methods only focus on fitting data with highest possible accuracy, neglecting the interpretability of the obtained fuzzy systems, which is a primary merit of fuzzy systems and the most prominent feature that distinguishes fuzzy systems from many other models.

In the recent few years, many researches have been devoted to the study of the tradeoff between interpretability and precision. Roubos [8] proposed an iterative fuzzy identification technique starting with a redundant fuzzy model obtained via fuzzy clustering in the product space of measured inputs and outputs. Successively, rule base simplification and GA-based optimization are applied iteratively to improve accuracy and reduce complexity. Papadakis [9] proposed a genetic algorithm based modeling method for building fuzzy system with scatter-type partitions. The method manages all attributes characterizing the structure of fuzzy system simultaneously, including the number of fuzzy rules, the input partition, the participating inputs of each fuzzy rule and the consequent parameters. The structure learning task is formulated as a multi-objective optimization problem which is resolved using a novel genetic-based structure learning scheme; and a genetic-based parameter learning scheme is performed for fine-tuning of the initial fuzzy system. Delgado [10] presented fuzzy modeling as a multi-objective decision-making problem, considering accuracy, interpretability and autonomy as goals. All these objectives are handled via a single-objective $\varepsilon-$constrained decision making problem, which is solved by a hierarchical evolutionary algorithm. Abonyi [11] presented a data-driven identification of fuzzy classification system for high-dimensional problems. A binary decision tree is used to select the relevant features and effective initial partition of the input domains; and the decision tree is transformed exactly to an initial fuzzy classifier. Then the obtained initial fuzzy classification system is reduced in an iterative scheme by means of similarity-driven rule-reduction. In order to improve the classification performance of the reduced fuzzy system, a genetic algorithm with a multi-objective criterion searching for both redundancy and accuracy is applied. Chang [12] addressed an automatic method to design fuzzy systems for classification via evolutionary optimization. At the beginning of the algorithm, the fuzzy system is empty with no rules in the rule base and no membership functions assigned to fuzzy variables. Then, different rules and membership functions are automatically created via VISIT algorithm by randomly assigning different initial parameters. At last, the evolutionary algorithm is used to find the optimal fuzzy system through simultaneously optimizing all the parameters of the system. Castro [13] presented an approach to build genetic fuzzy system considering both accuracy and interpretability. First, a data pre-processing step for feature selection is performed and the membership functions are generated using FCM algorithm. After that, a genetic algorithm is used for fuzzy rule base generation taking into account both comprehensibility and simplicity criteria. Finally, another genetic algorithm is executed for optimization of the previous rule base obtained, excluding unnecessary and redundant fuzzy rules. Mikut [14] presented a method for automatic and complete design of fuzzy systems from data with a user-controllable trade-off between accuracy and interpretability. The rule hypotheses are generated by

This work was supported by National Science Foundation of P. R. China (60332020), China Postdoctoral Science Foundation (2005037733), and Scientific Research Foundation of NUST (2005-AB96132).

Xing Zong-Yi is with school of mechanical engineering, Nanjing University of Science and Technology, Jiangsu, China. (corresponding author: +862584315417, e-mail: xingzongyi@tom.com)

Hou Yuan-Long is with School of Mechanical Engineering, Nanjing University of Science and Technology, Jiangsu, China.

inducing a decision tree, and are generalized by different modification of their premises, and the rule base is build by select a subset of generalized rules. Interpretability is maintained by structural choices and including interpretability criteria in the design process.

In all the above-mentioned methods, the multiple objectives are transformed into one single objective based on prior knowledge using techniques such as the weighted sum method and fuzzy expert system. However, if such prior knowledge is insufficient, or several situations should be considered, the above methods are limited, for it is difficult to determine weights of different objectives, or they can only provide one solution in a single run. In order to solve this problem, a more advanced method is needed, which could obtain multiple Pareto-optimal solutions simultaneously.

This paper presents an approach to construct multiple Pareto-optimal fuzzy systems using a multi-objective genetic algorithm considering both accuracy and interpretability. First, the modified Gath-Geva fuzzy clustering algorithm is used to construct antecedents of fuzzy system, and consequents are identified separately to reduce computational burden. Thus, a reasonably good initial fuzzy system is obtained. Second, the Pareto multi-objective genetic algorithm based on NAGS-II and the interpretability- driven simplification techniques are proposed to evolve the initial fuzzy system to optimize its structures and parameters iteratively, so both interpretability and precision of the fuzzy system are improved. As results, a set of Pareto-optimal fuzzy systems are obtained. The proposed approach is applied to the Mackey-Glass tine series and the Iris classification problem, and the results show its validity.

The paper is organized as follows. In section II, we show how to construct initial fuzzy system based on the modified Gath-Geva fuzzy clustering algorithm. Interpretability-driven simplification techniques are introduced in section III. Section IV details the Pareto multi-objective genetic algorithm based on NSGA-II. In section V, the proposed approach is demonstrated on the Mackey-Glass tine series and the Iris classification problem to show its validity. Section VI concludes the paper.

II. Construction of Initial Fuzzy system

Fuzzy clustering algorithm is a well-recognized technique to identify fuzzy systems. The fuzzy C-Means algorithm [15] and the Gustafson-Kessel algorithm [16] are the widely-used methods in fuzzy modeling. However, there are two main drawbacks to these algorithms. First, only clusters with approximately equal volumes can be properly identified, which is frequently difficult to satisfy in real systems. Second, clusters obtained are generally axes-oblique rather than axis-parallel; consequently, a decomposition error is made in their projection onto the input variables. To circumvent these problems, the modified Gath-Geva fuzzy clustering algorithm [17] is applied in this paper.

The objective function based on the minimization of the sum of weighted squared distances between the data points and cluster centers is described in the following:

$$J(Z;U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{ik}^2 \tag{1}$$

where Z is the set of data, $U = [\mu_{ik}]$ is the fuzzy partition matrix, $V = [V_1, V_2, \cdots, V_c]^T$ is the set of centers of the clusters, c is the number of clusters, N is the number of data, m is the fuzzy coefficient, μ_{ik} is the membership degree between the i-th cluster and k-th data, which satisfies conditions:

$$\mu_{ik} \in [0,1];\ \sum\nolimits_{i=1}^{C}\mu_{ik} = 1 \tag{2}$$

The Lagrange multiplier is used to optimize the objective function (6). The minimum of (U, V) is calculated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}(D_{ik}/D_{jk})^{2/(m-1)}} \tag{3}$$

$$v_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^m z_k}{\sum_{k=1}^{N}(\mu_{ik})^m} \tag{4}$$

The variance of the Gaussian function is:

$$\sigma_{ij}^2 = \frac{\sum_{k=1}^{N}\mu_{ik}(x_{jk} - v_{jk})^2}{\sum_{k=1}^{N}\mu_{ik}} \tag{5}$$

The norm of distance between i-th cluster and k-th data is

$$\frac{1}{D_{ik}^2} = \prod_{j=1}^{n}\frac{\sum_{k=1}^{N}\mu_{ik}}{N\sqrt{2\pi\sigma_{ij}^2}}\exp\left(-\frac{1}{2}\frac{(x_{jk}-v_{ij})^2}{\sigma_{ij}^2}\right)\cdot\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(y_k-\hat{y}_k)^T(y_k-\hat{y}_k)}{2\sigma_i^2}\right) \tag{6}$$

Calculation of consequents of TS fuzzy system is described as follows: given the input variable X, output y and fuzzy partition matrix U:

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_N^T \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},\ U_i = \begin{bmatrix} u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{iN} \end{bmatrix} \tag{7}$$

Appending a unitary column to X gives extended matrix X_e:

$$X_e = \begin{bmatrix} X & 1 \end{bmatrix} \tag{8}$$

then

$$\theta_i = [X_e^T U_i X_e]^{-1} X_e^T U_i y \tag{9}$$

is the consequent parameter of the TS fuzzy system.

In order to determination of consequents of fuzzy classification system, we define the function:

$$M_{ij} = \frac{\sum_{k=1}^{N}\mu_{ik} f_j(k)}{\sum_{k=1}^{N} f_j(k)} \tag{10}$$

where μ_{ik} is the membership degree between the i-th cluster and the k-th data, and $f_j(k)$ is defined as:

$$f_j(k) = \begin{cases} 1 & \text{if } x \in g_j \\ 0 & \text{if } x \notin g_j \end{cases} \tag{11}$$

For the i-th rule, the consequent can be determined:

$$i^* = \arg(\max(M_{ij}))\quad j = 1,2,\cdots,N_c \tag{12}$$

The procedure of constructing a fuzzy model based on the modified Gath-Geva fuzzy clustering algorithm is summarized as follows:

1) Choose the number of fuzzy rules, the weighting exponent, and the stop criterion $\varepsilon > 0$.

2) Generate the matrix *U* randomly. *U* must satisfy the condition (2).

3) Compute the parameters of the model using (4), (5), (9) or (12).

4) Calculate the norm of distance utilizing (6).

5) Update the partition matrix *U* using (3).

6) Stop if $\left\| U^{(l)} - U^{(l-1)} \right\| \le \varepsilon$; else go to 3).

III. Interpretability-driven Simplification Methods

A. Simplification of fuzzy sets

The initial fuzzy system obtained above by fuzzy clustering algorithm may contain redundant information in the form of similarity between fuzzy sets. The similarity of fuzzy sets makes the fuzzy system uninterpretable, for it is difficult to assign qualitatively meaningful labels to similar fuzzy sets. In order to acquire an effective and interpretable fuzzy system, elimination of redundancy and simplification of the fuzzy system are necessary.

There are three types of redundant or similar fuzzy sets in the fuzzy system: 1) a fuzzy set is similar to the universal set, 2) a fuzzy set is similar to the singleton set, and 3) the fuzzy set *A* is similar to the fuzzy set *B*.

If a fuzzy set is similar to the universal set or the singleton set, it should be removed from the corresponding fuzzy rule antecedent. As for two similar fuzzy sets, a similarity measure is utilized to determine if the fuzzy sets should be combined.

For fuzzy sets *A* and *B*, a set-theoretic operation based similarity measure [18] is defined as

$$S(A,B) = \frac{\sum_{k=1}^{N}[\mu_A(x_k) \wedge \mu_B(x_k)]}{\sum_{k=1}^{N}[\mu_A(x_k) \vee \mu_B(x_k)]} \tag{13}$$

where $\wedge$ and $\vee$ are minimum and maximum operators respectively. S is a similarity measure in [0, 1]. *S*=1 means the compared fuzzy sets are equal, while *S*=0 indicates that there is no overlap between the fuzzy sets.

If similarity measure $S > \tau$, i.e. fuzzy sets are very similar, then the two fuzzy sets *A* and *B* should be merged to create a new fuzzy set *C*, where τ is a predefined threshold. It should be pointed out that threshold τ influences the model performance significantly. A small threshold leads to a fuzzy model with low accuracy and highly interpretability. In a general way, $\tau = [0.4 - 0.7]$ is a good choice.

For the Gaussian type of fuzzy sets used in this paper, the parameters of newly merged fuzzy set *C* from *A* and *B* are defined as

$$\begin{cases} v_c = (v_A + v_B)/2 \\ \sigma_c = \sqrt{\sigma_A^2 + \sigma_B^2}/2 \end{cases} \tag{14}$$

The process of merging similar fuzzy sets is executed iteratively. For each iteration, the similarity measures between all pairs of adjacent fuzzy sets for each variable are calculated. The pair of highly similar fuzzy sets with $S > \tau$ is merged to create a new fuzzy set. The rule base of the fuzzy system is updated by substituting the new fuzzy set for the two highly similar fuzzy sets. This process continues until there are no fuzzy sets for which $S > \tau$. Then the fuzzy sets that have similarity to the universal set or the singleton set are removed.

B. Simplification of the Fuzzy Rules

During the process of simplification of similar fuzzy sets and the process of evolutionary operation, it may generate similar or same fuzzy rules, which need be reduced to improve interpretability of the fuzzy system.

Considering the following two fuzzy rules:

$$R^i: \textit{If } x_1 \textit{ is } \mu_{i1}(x_1) \textit{ and } x_2 \textit{ is } \mu_{i2}(x_2) \textit{ and } \cdots x_n \textit{ is } \mu_{in}(x_n)$$
$$\textit{then the pattern } (x_1,\cdots,x_n) \textit{ belongs to } g_i$$
$$R^j: \textit{If } x_1 \textit{ is } \mu_{j1}(x_1) \textit{ and } x_2 \textit{ is } \mu_{j2}(x_2) \textit{ and } \cdots x_n \textit{ is } \mu_{jn}(x_n)$$
$$\textit{then the pattern } (x_1,\cdots,x_n) \textit{ belongs to } g_j$$

then a similarity measure of fuzzy rules [19] is defined as

$$S_R(R^i, R^j) = \min_{k=1}^{n} S(\mu_{ik}, \mu_{jk}) \tag{15}$$

where $S(\cdot)$ is calculated with the formula (11).

If $S(\cdot) > \lambda$, *i.e.* the two fuzzy rules are very similar, then only one fuzzy rule is preserved, while the other is deleted, where λ is a predefined threshold. In a general way, $\lambda = [0.9 - 1]$ is used. As the simplification of fuzzy sets, simplification of fuzzy rules is also carried out iteratively.

IV. Pareto multi-objective Genetic Algorithm

After simplification, the initial fuzzy system is encoded to a real-coded population, which is evolved using a Pareto multi-objective genetic algorithm base on NSGA-II [20]. The process of simplification and the process of evolution are executed iteratively until multiple Pareto-optimal fuzzy systems are generated. Three elements of the algorithm, chromosome representation, multi-objective fitness function and genetic operators are detailed following.

A. Chromosome Representation

For complex system, the bit strings of binary-coded genetic algorithm becomes very long and the search space blows up, while in real-coded genetic algorithm, the variables appear directly in chromosome simply, and computation burden is relieved, so real-coded scheme is adopted in this paper.

The first chromosome is formed as a sequence of genes describing parameters in the rule antecedents of the obtained fuzzy system:

$$H_1 = (v_{11},\cdots,v_{cn},\sigma_{11},\cdots\sigma_{cn}) \tag{16}$$

Given search space

$$H^{\min} = (v_{11}^{\min},\cdots,v_{cn}^{\min},\sigma_{11}^{\min},\cdots\sigma_{cn}^{\min}) \tag{17}$$

$$H^{\max} = (v_{11}^{\max},\cdots,v_{cn}^{\max},\sigma_{11}^{\max},\cdots\sigma_{cn}^{\max}) \tag{18}$$

where $v_{ij}^{\max}$ 、 $v_{ij}^{\min}$ 、 $\sigma_{ij}^{\min}$ 、 $\sigma_{ij}^{\max}$ are maximum and minimum values of corresponding membership functions.

The other chromosomes of the initial population are created by random variation (uniform distribution) around H_1 within the search space.

B. Multi-objective Fitness Function

Fuzzy modeling requires the consideration of multiple objectives in the design process, including precision and interpretability. In this paper, precision is defined as the root-mean-square error (TS fuzzy system) or mistakenly classified patterns (fuzzy classification system), while it is difficult to quantify interpretability. According to the analysis about interpretability, we have guaranteed the features of fuzzy sets by interpretability-driven techniques, so only the number of rules and the number of fuzzy sets are included in the objective functions.

These three objectives about fuzzy modeling can be formulated as follows:

$$\text{Min } f_1(S),\ \text{Min } f_2(S),\ \text{Min } f_3(S) \tag{19}$$

where $f_1(S)$ is precision performance, $f_2(S)$ is the number of fuzzy rules, $f_3(S)$ is the number of fuzzy sets.

In general, the fuzzy system with high accuracy owns more fuzzy rules and fuzzy sets, while the fuzzy system with fewer fuzzy rules and fuzzy sets leads to low precision, so there is no single fuzzy system satisfying all the above three objective, and our task is to get a set of Pareto-optimal fuzzy systems which are not dominated by each other.

A fuzzy system S_A is said to dominate another fuzzy system S_B if the following condition holds:

$$f_1(S_A) \le f_1(S_B),\ f_2(S_A) \le f_2(S_B),\ f_3(S_A) \le f_3(S_B) \tag{20}$$

and at least one of the following inequalities holds:

$$f_1(S_A) < f_1(S_B) \tag{21}$$

$$f_2(S_A) < f_2(S_B) \tag{22}$$

$$f_3(S_A) < f_3(S_B) \tag{23}$$

The condition (20) shows that no objective of S_B is worse than S_A. Any inequality of (21)-(23) means that at least one objective of S_B is better than S_A. When a fuzzy system S is not dominated by any other fuzzy systems, S is regard as a Pareto-optimal fuzzy system.

Several multi-objective algorithms have been proposed, including, NSGA-II [20], PAES [21] and SPEA [22]. In this paper, we use the NSGA-II algorithm due to its high searching ability and easy implementation. For more details about the NSGA-II algorithm, please see the reference [20].

C. Genetic Operators

There are three genetic operators in multi-objective genetic algorithm: selection, crossover and mutation. In order to hold variety of chromosomes, several randomly selected methods for each operator are adopted in this paper.

1) *Selection*:

The *roulette wheel* selection method is used to select individuals to operate. For chromosome H_p with fitness value f_p, the selected probability is:

$$p_p = f_p \Big/ \sum_{p=1}^{L} f_p \tag{24}$$

In order to prevent optimal chromosomes are ignored, elitist selection are used at the same time, i.e., the best chromosome is always preserved in population.

2) *Crossover*

$H_r^t = (r_1, \cdots, r_l)$ and $H_s^t = (s_1, \cdots, s_l)$ are selected chromosome for crossover in *t*-generation. The following two crossover operators are adopted randomly.

Simple arithmetic crossover: k is randomly selected position of chromosome. The result offspring are:

$$H_r^{t+1} = (r_1, \cdots, r_k, s_{k+1}, \cdots, s_l) \tag{25}$$

$$H_s^{t+1} = (s_1, \cdots, s_k, r_{k+1}, \cdots, r_l) \tag{26}$$

Whole arithmetic crossover: $\lambda \in [0,1]$ is a uniform distributed random number. The result offspring are:

$$H_r^{t+1} = \lambda(H_r^t) + (1-\lambda)H_s^t \tag{27}$$

$$H_s^{t+1} = \lambda(H_s^t) + (1-\lambda)H_r^t \tag{28}$$

3) *Mutation*

$H_r^t = (r_1, \cdots, r_l)$ and $H_s^t = (s_1, \cdots, s_l)$ are selected chromosome for crossover in t generation. T is total number of generations. The following mutation operators are adopted randomly.

Uniform mutation: r_k is randomly selected element of chromosome. $\hat{r}_k \in [r_k^{\min}, r_k^{\max}]$ is random number where $[r_k^{\min}, r_k^{\max}]$ is search space of r_k. The result offspring is:

$$H_r^{t+1} = (r_1, \cdots, \hat{r}_k, \cdots, r_l) \tag{29}$$

Gaussian mutation: x_k is a Gaussian distributed random number with zero mean and adaptive variance σ_k

$$\sigma_k = ((T-t)/T)((r_k^{\max} - r_k^{\min})/3) \tag{30}$$

The corresponding offspring is:

$$H_r^{t+1} = (\hat{r}_1, \cdots, \hat{r}_k, \cdots, \hat{r}_l) \tag{31}$$

where

$$\hat{r}_k = r_k + x_k \tag{32}$$

D. Pseudo-code of the Algorithm

The pseudo-code of the proposed algorithm is present in the following

```
Begin
  g:=0
    Initialize fuzzy system Fs
    P(0)=encode(Fs)
  While not done do
    g:=g+1
    A^1 (g) = decode (P (g))
    A^2 (g) = simplify (A^1 (g))
    C^2 (g) = calculate (A^2 (g))
    Fs^2(g) = A^2 (g) + C^2 (g)
    Fs (g) = NSGA-II (Fs (g) + Fs (g))
    P^1 (g) = encode (Fs (g))
    P (g) = operate (P^1 (g))
  End while
End
```

where encode() converts antecedents of fuzzy system into population, decode() converts population into antecedents, simplify() reduces the obtained antecedents of fuzzy system, calculate() identify consequents of fuzzy system, NSGA-II() generates multiple fuzzy systems based on non-domination level and crowding distance, operate() executes three genetic operators.

V. EXPERIMENTS ANS RESULTS

In order to examine the performance of the proposed approach, two benchmark problems, the Mackey-Glass time series and the Iris classification problem, are demonstrated in this section. Table I gives the parameter setups of the algorithm. All simulation programs are realized under Matlab 7.0 environment.

TABLE I
PARAMETER SETUPS OF THE PROPOSED APPROACH

Parameters	Values
Maximum generations	100
Initial population size	40
Parent population size	40
Child population size	40
Crossover probability	0.5
Mutation probability	0.1
Threshold of merging fuzzy sets	0.4
Threshold of merging fuzzy rules	1

A. Example: Mackey-Glass time series

The Mackey-Glass time series is described as follows:

$$\dot{x} = \frac{0.2x(t-17)}{1+x^{10}(t-17)} - 0.1x(t) \tag{33}$$

The goal is to predict $x(t+6)$ from $x(t)$, $x(t-12)$and $x(t-18)$.1000 data points are generated using the fourth order Runge-Kutta method with a step length of 0.1 and the initial condition $x(0)=1.2$, where 500 pair of data are used for training and the others for test. The sampling data of $x(t+6)$ is showed in Figure 1.

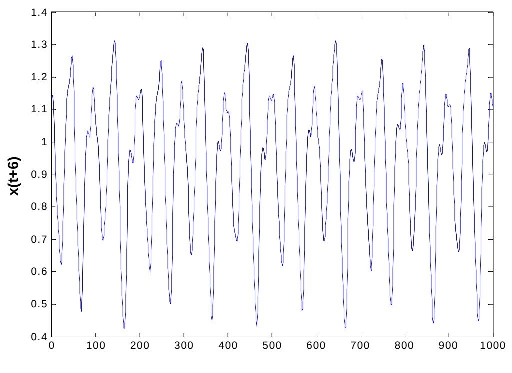

Figure 1. The Mackey-Glass time series

The initial fuzzy system is obtained by the fuzzy clustering and the least square method. The *RMSE* (Root Mean Square Error) of training data is 0.0657, and the *RMSE* of test data is 0.0646. The number of fuzzy rules is 5, and the number of fuzzy sets is 20.

The interpretability- driven simplification techniques and the multi-objective genetic algorithm are used to optimize the initial fuzzy system. The performance of the obtained four Pareto-optimal fuzzy systems is described in Table II. The decision-marker can choose an appropriate fuzzy system according to a specific situation, either the one with higher interpretability (less number of fuzzy rules or/and fuzzy sets) or the one with less error.

Table II also shows a comparison between the proposed method and other published systems, which indicates that the proposed approach is able to find multiple fuzzy systems than any other algorithm with higher precision performance and less number of fuzzy rules and fuzzy sets. In conclusion, the proposed method can obtain multiple interpretable and accurate fuzzy systems.

TABLE II
COMPARISON OF RESULTS OF THE MACKEY-GLASS TIME SERIES

	# Fuzzy rules	# Fuzzy sets	Training RMSE	Testing RMSE
Paiva [23]	9	23	0.0228	0.0239
Nauck[24]	129	35	0.0315	0.0332
	26	19	0.0656	0.0671
This paper				
Initial	5	20	0.0657	0.0646
Fuzzy system 1	5	8	7.0551e-3	6.8453e-3
Fuzzy system 2	4	7	7.7289e-3	7.6689e-3
Fuzzy system 3	3	6	1.0933e-2	1.0860e-2
Fuzzy system 4	2	5	2.0138e-2	1.9991e-2

Figure 2 and Figure 3 show the membership functions of the first Pareto-optimal fuzzy system and the comparison of system outputs and actual outputs of testing data, respectively. Table III details the structure and parameters of the first Pareto-optimal fuzzy system.

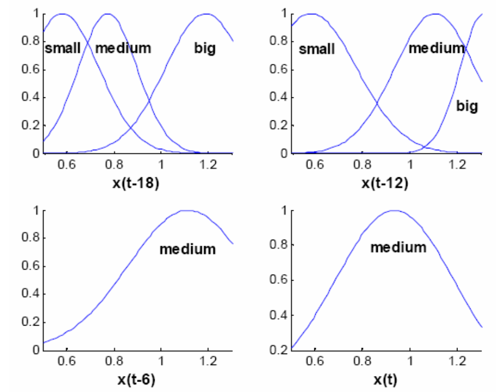

Figure 2. Membership functions of the first fuzzy system of time series

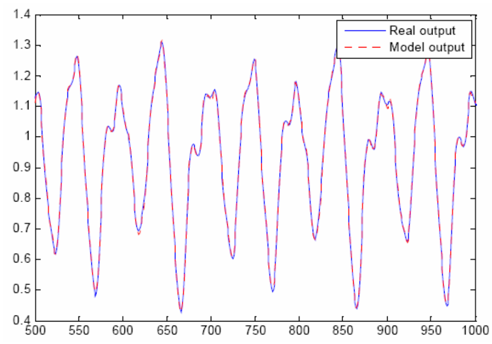

Figure 3. Comparison of system outputs and real outputs

TABLE III

THE FIRST PARETO-OPTIMAL FUZZY SYSTEM OF TIME SERIES

The first Pareto-optimal fuzzy system of Mackey-glass time series
R1: **If** $x(t)$ is medium, $x(t\ \text{-}6)$ is medium , $x(t\text{-}12)$ is small, $x(t\text{-}18)$ is small, **then** x(t+6) = 0.51 $x(t)$+0.11 $x(t\ \text{-}6)$-0.47 $x(t\text{-}12)$-0.15 $x(t\text{-}18)$+0.062
R2: **If** $x(t)$ is medium, $x(t\ \text{-}6)$ is medium, $x(t\text{-}12)$ is medium, $x(t\text{-}18)$ is medium, **then** $x(t\ +6)$ = -0.037 $x(t)$-0.47 $x(t\ \text{-}6)$-0.76x3-0.034 $x(t\text{-}18)$+1.42
R3: **If** x(t) is medium, x(t-6) is medium, $x(t\text{-}12)$ is big,$x(t\text{-}18)$ is big, **then** $x(t\ +6)$ = 0.74 $x(t)$-0.22 $x(t\ \text{-}6)$-0.11 $x(t\text{-}12)$-0.16 $x(t\text{-}18)$+0.579
R4: **If** $x(t)$ is medium, $x(t\ \text{-}6)$ is medium, $x(t\text{-}12)$ is small, $x(t\text{-}18)$ is big, **then** $x(t\ +6)$ = 1.45 $x(t)$-0.57$x(t\text{-}6)$+0.59$x(t\text{-}12)$+0.28 $x(t\text{-}18)$-0.459
R5: **If** $x(t)$ is medium, $x(t\ \text{-}6)$ is medium, $x(t\text{-}12)$ is medium, $x(t\text{-}18)$ is big, **then** $x(t\ +6)$ = -0.017 $x(t)$+0.53 $x(t\ \text{-}6)$-1.27 $x(t\text{-}12)$-0.656 $x(t\text{-}18)$+2.316
Parameters of antecedents of the fuzzy system:
$x(t)$: medium=[0.936,0.061] $x(t\ \text{-}6)$: medium=[1.11,0.064]
$x(t\ \text{-}12)$: small=[0.58,0.038] medium=[1.1,0.029] big=[1.31,0.01]
$x(t\ \text{-}18)$: small=[0.58,0.025] medium=[0.77,0.016] big=[1.19,0.029]

B. Example: Iris classification System

The Iris classification system is a benchmark problem in classification and pattern recognition studies. It contains 50 measurements of four features (*sepal length*, *sepal width*, *pental length*, *pental width*) from each of three species (*setosa*, *versicolor*, *virinica*). The first class is separate from others clearly, while the second and third class are overlap slightly. Figure 4 shows the two-dimension (*sepal length*, *sepal width*) measurement, where "*" denotes data of *setosa* class, "o" denotes data of *versicolor* class, "+" denotes data of *virinica* class.

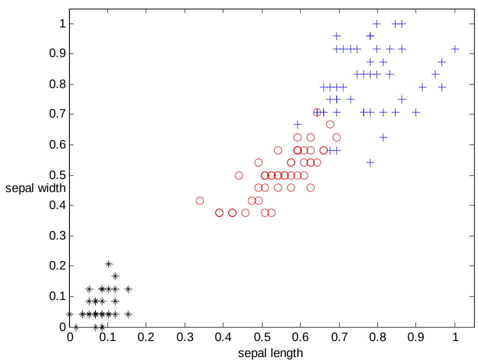

Figure 4. Iris data: *setosa* (*), *versicolor* (o), *virinica* (+)

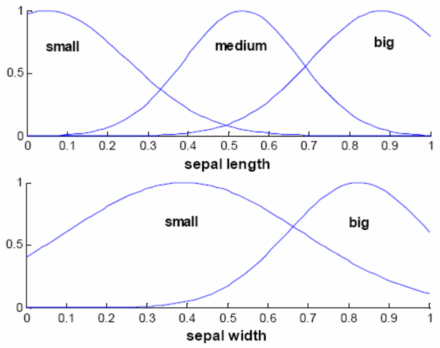

Figure 5. Membership functions of the second fuzzy system of Iris problem

TABLE IV

COMPARISON OF RESULTS OF THE MACKEY-GLASS TIME SERIES

	# Fuzzy rules	# Fuzzy sets	Classification rate (%)
Wang [25]	3	11	97.5
Wu[26]	3	9	96.2
Shi[27]	4	12	98
Ishibuchi[28]	5	7	98
Tong[29]	3	12	98
Russo[30]	5	18	100
This paper			
Initial	9	36	95.3
Solution 1	7	10	98.7
Solution 2	4	5	98
Solution 3	3	5	96
Solution 4	3	4	94.7

TABLE V

THE SECOND PARETO-OPTIMAL FUZZY SYSTEM OF IRIS

The second pareto-optimal fuzzy system of the Iris problem
R1: **If** *sepal length* is big , *sepal width* is big **then** output is *virinica*
R2: **If** *sepal length* is small , *sepal width* is small **then** output is *setosa*
R3: **If** *sepal length* is medium, *sepal width* is big **then** output is *virinica*
R4: **If** *sepal length* is medium , *sepal width* is small **then** output is *versicolor*
Parameters of antecedents of the fuzzy system:
sepal length: small=[0.048457,0.040494] medium=[0.53457,0.020748] big=[0.88115,0.030973]
sepal width: small=[0.39215,0.084652] big=[0.82529,0.030372]

The initial fuzzy system is obtained by the fuzzy clustering algorithm. The precision performance is 95.3%, and the number of fuzzy rules is 9, and the number of fuzzy sets is 36.

The interpretability- driven simplification techniques and the multi-objective genetic algorithm are used to optimize the initial fuzzy system. The performance of the obtained four Pareto-optimal fuzzy systems is described in Table IV. The decision-marker can choose an appropriate fuzzy system according to a specific situation, either the one with higher interpretability (less number of fuzzy rules or fuzzy sets) or the one with less error. Table IV also shows the comparisons with other results, which indicates that the proposed method can obtain multiple accurate and interpretable fuzzy systems. Russo [30] classified all patterns correctly; however it is difficult to interpret the system for containing too many fuzzy rules and fuzzy sets

Figure 5 shows the membership functions of the second Pareto-optimal fuzzy system. Table V details the structure and parameters of the second Pareto-optimal fuzzy system.

VI. CONCLUSIONS

In this paper, we presented an approach based on a multi-objective genetic algorithm to construct accurate and interpretable fuzzy system. First, a modified fuzzy clustering algorithm is used to construct antecedents of fuzzy system, and consequents are identified separately to reduce computational burden. Second, the multi-objective genetic algorithm based on NSGA-II and simplification techniques are proposed to evolve the initial fuzzy system to optimize its structures and parameters iteratively, so both interpretability and precision of the system are improved. The proposed approach is applied to the Mackey-Glass tine series and the Iris classification problem, and the results show its validity.

REFERENCES

[1] L. A. Zadeh. Fuzzy sets [J]. *Information and Control*, 1965, 8: 338-353.

[2] R. Babuska. *Fuzzy Modeling for Control* [M]. Boston: Kluwer Academic Publisher, 1998

[3] A. F. Gomez-Skarmeta, M. Delgado, M. A. Vila. About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy Sets and Systems*, 1999, 106(2): 179-188.

[4] H. T. Lefteri, E. Robert. *Fuzzy and Neural Approaches in Engineering*. Wiley, 1997.

[5] H. T. Lefteri, E. U. Robert. *Fuzzy and Neural Approaches in Engineering*. New York: Wiley, 1997

[6] O. Cordon, F. Herrera. *Genetic Fuzzy models: Evolutionary Tuning and Learning of Fuzzy Rule Bases*. Singapore: World Scientific, 2000

[7] O. Cordon, F. Gomide, F. Herrera. Ten Years of Genetic Fuzzy models: Current Framework and New Trends. *Fuzzy sets and systems*, 2004, 141(1): 5-31.

[8] J. A. Roubos, M. Setnes. Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Trans. on Fuzzy Systems*, 2001, 9(4):516-524.

[9] S. E. Papadakis, J. B. Theocharis. GA-based fuzzy modeling approach for generating. TSK models. *Fuzzy Sets and Systems*, 2002, 131(2): 121-152.

[10] M. R. Delgado, F. Zuben, F. Gomide. Multi-Objective Decision Making: Towards Improvement of Accuracy, Interpretability and Design Autonomy in Hierarchical Genetic Fuzzy models. *Proc. IEEE Int. Conf. on fuzzy systems*, Hawai, 2002: 1222-1227.

[11] J. Abonyi, J. A. Roubos, and F. Szeifert. Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization. *Int. J. of Approximate Reasoning*, 2003, 32(1): 1-21.

[12] X. Chang, J. H. Lilly. Evolutionary Design of a Fuzzy Classifier From Data. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 2004, 34(4): 1894- 1906.

[13] P. A. D. Castro, H. A. Camargo. Focusing on Interpretability and Accuracy of a Genetic Fuzzy System. *Proc. IEEE Int. Conf. on fuzzy systems*, Reno Nevada.2005, 696-701.

[14] R. Mikut, J. Jakel, L. Groll. Interpretability issues in data-based learning of fuzzy systems. *Fuzzy Sets and Systems*, 2005, 150(2): 179-197.

[15] J. C. Bezdek. *Pattern Recognition with fuzzy objective algorithm*. New York: Plenum Press, 1981

[16] D. Gustafson, W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. *Proc. IEEE conf. on decision and control*. San Diego, USA, 1979: 761-766.

[17] J. Abonyi, B. Babuska, F. Szeifert. Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 2002, 32(5): 612-621.

[18] M. Setnes, R. Babuska, U. Kaymak, H. R. N. Lemke. Similarity Measures in Fuzzy Rule Base Simplification. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 1998, 28(3): 376-386.

[19] Y. Jin. Advanced *Fuzzy Systems Design and Applications*. NewYork：Physical-verl, 2003.

[20] K. Deb, A. Pratap, S. Agarwal. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation*, 2002, 6(2): 182-197.

[21] J, D. Knowles, D. Corne. Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation*, 2000, 8(2): 149-172.

[22] E. Zitzler, L. Thiele. Multi-Objective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Trans. on Evolutionary Computation*, 1999, 3(4):257-271.

[23] R. P Paiva, A. Dourado. Interpretability and learning in neuro-fuzzy models. *Fuzzy Sets and Systems*. 2004, 147(2004): 17-38.

[24] D. D. Nauck. Fuzzy data analysis with NEFCLASS. *Approximate reasoning*, 2003, 32:103-130.

[25] J. S. Wang, G. C. S. Lee. Self-adaptive neuro-fuzzy inference system for classification application. *IEEE Trans. Fuzzy System*. 2002, 10(6): 790–802.

[26] T.P. Wu, S.M. Chen. A new method for constructing membership functions and fuzzy rules from training examples. *IEEE Trans. on Systems, Man and Cybernetics*, *Part B*, 1999, 29(1):25-40.

[27] Y. Shi, R. Eberhart and Y. chen. Implementation of evolutionary fuzzy. Systems. *IEEE Trans. Fuzzy System*, 1999, 7(2): 109-119.

[28] H. Ishibuchi, T. Nakashima, T. Murata. Three-objective genetics-based machine learning for linguistic rule extraction. *Information Science*, 2001, 136(1-4): 109-133.

[29] S. Tong, Y. Shen, Z. Liu. Approach to construct fuzzy classification system with clustering. *Control and Decision*. 2001, 16(SUPP1):737-740. (In Chinese).

[30] M. Russo. Genetic fuzzy learning. *IEEE Trans. Evolutionary Computation*. 2000, 4(3): 259–273.