Multi-objective Fuzzy Modeling Using NSGA-II

Xing Zong-Yi, Zhang Yong, Hou Yuan-Long
School of mechanical engineering
Nanjing University of Science and Technology
Jiangsu, China

Cai Guo-Qiang
School of Traffic and Transportation
Beijing Jiaotong University
Beijing, China

Abstract—An approach to construct multiple Pareto-optimal fuzzy systems based on NSGA-II 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, NSGA-II

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. Chang [11] 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. Mikut [12] 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 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. 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 time 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. A modified Gath-Geva fuzzy clustering algorithm [13] is applied in this paper to identify initial fuzzy system.

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:

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

(1)

where \( Z \) is the set of data, \( U=\left[\mu_{ik}\right] \) is the fuzzy partition matrix, \( V=\left[V_{1}, V_{2}, \cdots, V_{c}\right]^{T} \) 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, \( \mu_{ik} \) is the membership degree between the \( i \)-th cluster and \( k \)-th data, which satisfies conditions:

\[ \mu_{ik} \in [0,1] ; \sum_{i=1}^{c} \mu_{ik} = 1 \]  

(2)

The Lagrange multiplier is used to optimize the objective function (1). The minimum of \((U,V)\) is calculated as follows:

\[ \mu_{ik} = \frac{1}{\sum_{i=1}^{N} \left( D_{ik} / D_{ik} \right)^{2 \gamma(n-m)}} \]  

(3)

\[ v_{i} = \frac{1}{N} \sum_{k=1}^{N} \left( \mu_{ik} \right)^{m} v_{ik} \]  

(4)

The variance of the Gaussian function is:

\[ \sigma_{\gamma}^{2} = \frac{1}{\sum_{i=1}^{N} \sum_{k=1}^{c} \left( x_{ik} - v_{ik} \right)^{2}} \]  

(5)

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

\[ \frac{1}{D_{ik}} \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}} \exp \left( \frac{1}{2} \left( x_{ij} - v_{ij} \right)^{2} \right) \]  

(6)

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

\[ X = \left[ \begin{array}{c} X_{1}^{T} \\ X_{2}^{T} \\ \vdots \\ X_{N}^{T} \end{array} \right], y = \left[ \begin{array}{c} y_{1} \\ y_{2} \\ \vdots \\ y_{N} \end{array} \right], U_{j} = \left[ \begin{array}{cccc} u_{1j} & 0 & \cdots & 0 \\ 0 & u_{2j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{Nj} \end{array} \right] \]  

(7)

Appending a unitary column to \( X \) gives extended matrix \( X_{e} \):

\[ X_{e} = \left[ X \ 1 \right], \]  

(8)

then

\[ \theta_{i} = \left( X_{e}^{T} U_{i} \right)^{-1} X_{e}^{T} U_{i} y \]  

(9)

is the consequent parameter of the TS fuzzy system.

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

\[ M_{ij} = \frac{\sum_{k=1}^{N} \mu_{ik} f_{j}(k)}{\sum_{k=1}^{N} f_{j}(k)} \]  

(10)

where \( \mu_{ik} \) is the membership degree between the \( i \)-th cluster and the \( k \)-th data, and \( f_{j}(k) \) is defined as:

\[ f_{j}(k) = \begin{cases} 1 & \text{if } x \in g_{j} \\ 0 & \text{if } x \notin g_{j} \end{cases} \]  

(11)

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

\[ i^{*} = \arg \left( \max \left( M_{ij} \right) \right) \]  

(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 \( \epsilon > 0 \).
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 (4), (5), (9) or (12).
5) Update the partition matrix \( U \) using (3).
6) Stop if \( \left\| f^{(i)} - U^{(i-1)} \right\| \leq \epsilon \); else go to 3.

III. PREPARE YOUR PAPER BEFORE STYLING

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.

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 [14] is defined as

\[ S(A,B) = \frac{\sum_{i=1}^{N} \left[ \mu_{A}(x_{i}) \wedge \mu_{B}(x_{i}) \right]}{\sum_{i=1}^{N} \left[ \mu_{A}(x_{i}) \vee \mu_{B}(x_{i}) \right]} \]

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 \( \tau \) is a predefined threshold. It should be pointed out that threshold \( \tau \) 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.

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:

$$R : \text{If } x_1 \text{ is } \mu_1(x_1) \text{ and } x_2 \text{ is } \mu_2(x_2) \text{ and } \ldots \text{ and } x_n \text{ is } \mu_n(x_n) \text{ then the pattern } (x_1, \ldots, x_n) \text{ belongs to } g_i$$

$$R' : \text{If } x_1 \text{ is } \mu_1(x_1) \text{ and } x_2 \text{ is } \mu_2'(x_2) \text{ and } \ldots \text{ and } x_n \text{ is } \mu_n(x_n) \text{ then the pattern } (x_1, \ldots, x_n) \text{ belongs to } g_j$$

then a similarity measure of fuzzy rules is defined as $^{[15]}$

$$S_g(R, R') = \min_{k=1}^n S(\mu_{ik}, \mu_{jk})$$

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

If $S() > \lambda$, i.e., the two fuzzy rules are very similar, then only one fuzzy rule is preserved, while the other is deleted, where $\lambda$ 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 $^{[16]}$. 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, and 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:

$$H_1 = (v_{11}, \ldots, v_{cm}, \sigma_{11}, \ldots, \sigma_{cn})$$

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:

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

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 objectives, and our task is to get a set of Pareto-optimal fuzzy systems which are not dominated by each other.

Several multi-objective algorithms have been proposed, including, NSGA-II $^{[16]}$, PAES $^{[17]}$ and SPEA $^{[18]}$. 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 [16].

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:

$$P_p = f_p / \sum_{p=1}^{t} f_p$$

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

$$H'_{r} = (r_1, \ldots, r_t) \text{ and } H'_{s} = (s_1, \ldots, s_t)$$

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:

$$H'_{r}^{k+1} = (r_1, \ldots, r_k, s_{k+1}, \ldots, s_t)$$
$$H'_{s}^{k+1} = (s_1, \ldots, s_k, r_{k+1}, \ldots, r_t)$$

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

$$H'_{r}^{\lambda} = \lambda(H_{r}) + (1 - \lambda)H_{s}^{k+1}$$
$$H'_{s}^{\lambda} = \lambda(H_{s}) + (1 - \lambda)H_{r}^{k+1}$$

3) Mutation

$$H'_{r} = (r_1, \ldots, r_t) \text{ and } H'_{s} = (s_1, \ldots, s_t)$$

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. $\tilde{r}_k \in [r_k^{\min}, r_k^{\max}]$ is random number where $[r_k^{\min}, r_k^{\max}]$ is search space of $r_k$. The result offspring is:
Gaussian mutation: $x_i$ is a Gaussian distributed random number with zero mean and adaptive variance $\sigma_i$:

$$\sigma_i = ((T - t) / T)(r_{i}^{\max} - r_{i}^{\min} / 3)$$

The corresponding offspring is:

$$H^{*} = (\hat{r}_i, \cdots, \hat{r}_k)$$

where

$$\hat{r}_i = r_i + x_i$$

V. PARETO MULTI-OBJECTIVE GENETIC ALGORITHM

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>
<thead>
<tr>
<th>Parameters Setup of the Proposed Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
</tr>
<tr>
<td>Maximum generations</td>
</tr>
<tr>
<td>Initial population size</td>
</tr>
<tr>
<td>Parent population size</td>
</tr>
<tr>
<td>Child population size</td>
</tr>
<tr>
<td>Crossover probability</td>
</tr>
<tr>
<td>Mutation probability</td>
</tr>
<tr>
<td>Threshold of merging fuzzy sets</td>
</tr>
<tr>
<td>Threshold of merging fuzzy rules</td>
</tr>
</tbody>
</table>

A. Example: Mackey-Glass time series

The Mackey-Glass time series is described as follows:

$$\dot{x} = \frac{0.2x(t-17)}{1 + x^{10}(t-17)} - 0.1x(t)$$  \hspace{1cm} (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 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 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>
<thead>
<tr>
<th># Fuzzy rules</th>
<th># Fuzzy sets</th>
<th>Training RMSE</th>
<th>Testing RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paiva [19]</td>
<td>9</td>
<td>23</td>
<td>0.0228</td>
</tr>
<tr>
<td>Nauck [20]</td>
<td>129</td>
<td>35</td>
<td>0.0315</td>
</tr>
<tr>
<td>Initial</td>
<td>26</td>
<td>19</td>
<td>0.0656</td>
</tr>
<tr>
<td>Fuzzy system 1</td>
<td>5</td>
<td>8</td>
<td>7.0551e-3</td>
</tr>
<tr>
<td>Fuzzy system 2</td>
<td>4</td>
<td>7</td>
<td>7.7289e-3</td>
</tr>
<tr>
<td>Fuzzy system 3</td>
<td>3</td>
<td>6</td>
<td>1.0933e-2</td>
</tr>
<tr>
<td>Fuzzy system 4</td>
<td>2</td>
<td>5</td>
<td>2.0138e-2</td>
</tr>
</tbody>
</table>

Figure 1 and Figure 2 show the membership functions of the first Pareto-optimal fuzzy system and the comparison of system outputs and actual outputs of testing data, respectively.

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, virginica). The first class is separate from others clearly, while the second and third class are overlap slightly.
Russo [26] classified all patterns correctly; however, it is can obtain multiple accurate and interpretable fuzzy systems. With other results, which indicates that the proposed method interpretability (less number of fuzzy rules or fuzzy sets) or according to a specific situation, either the one with higher decision-maker can choose an appropriate fuzzy system Pareto-optimal fuzzy systems is described in Table IV. The initial fuzzy system. The performance of the obtained four multi-objective genetic algorithm are used to optimize the algorithm. The parameters of the second Pareto-optimal fuzzy system. Table V details the structure and parameters of the second Pareto-optimal fuzzy system.

The second pareto-optimal fuzzy system of Iris problem

<table>
<thead>
<tr>
<th># Fuzzy</th>
<th># Fuzzy</th>
<th>Classification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rules</td>
<td>sets</td>
<td></td>
</tr>
</tbody>
</table>

**R USO [26]** 3 11 97.5
Wu[22] 3 9 96.2
Shi[23] 4 12 98
Ishibuchi[24] 5 7 98
Tong[25] 3 12 98
Russo[26] 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

The second pareto-optimal fuzzy system of Iris problem

<table>
<thead>
<tr>
<th># Fuzzy</th>
<th># Fuzzy</th>
<th>Classification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rules</td>
<td>sets</td>
<td></td>
</tr>
</tbody>
</table>

**R USO [26]** 3 11 97.5
Wu[22] 3 9 96.2
Shi[23] 4 12 98
Ishibuchi[24] 5 7 98
Tong[25] 3 12 98
Russo[26] 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 IV**

**THE SECOND PARETO-OPTIMAL FUZZY SYSTEM OF IRIS**

The second pareto-optimal fuzzy system of the Iris problem

**R USO [26]** 3 11 97.5
Wu[22] 3 9 96.2
Shi[23] 4 12 98
Ishibuchi[24] 5 7 98
Tong[25] 3 12 98
Russo[26] 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

**ACKNOWLEDGMENT**

This work was supported by National Science Foundation of P. R. China (60332020), State Key Laboratory of Rail Traffic Control and Safety (SLK200810), Beijing Jiaotong University.

**REFERENCES**


**FIGURE 3. Membership functions of the second fuzzy system of Iris problem**


