On Generating Fuzzy Systems based on Pareto Multi-objective Cooperative Coevolutionary Algorithm

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

Abstract: An approach to construct multiple interpretable and precise fuzzy systems based on the Pareto Multi-objective Cooperative Coevolutionary Algorithm (PMOCCA) is proposed in this paper. First, a modified fuzzy clustering algorithm is used to construct antecedents of fuzzy system, and consequents are identified separately to reduce computational burden. Then, the PMOCCA and the interpretability-driven simplification techniques are executed to optimize the initial fuzzy system with three objectives: the precision performance, the number of fuzzy rules and the number of fuzzy sets; thus both the precision and the interpretability of the fuzzy systems are improved. In order to select the best individuals from each species, we generalize the NSGA-II algorithm from one species to multi-species, and propose a new non-dominated sorting technique and collaboration mechanism for cooperative coevolutionary algorithm. Finally, the proposed approach is applied to two benchmark problems, and the results show its validity.

Keywords: Coevolutionary algorithm, fuzzy modeling, fuzzy system, multi-objectives.

1. INTRODUCTION

Fuzzy systems have been successfully applied to various areas such as classification, simulation, data mining, pattern recognition, prediction and control. When they are constructed traditionally based on expert knowledge, the obtained fuzzy systems are usually well understandable, while when they are built from numerical data, the generated fuzzy systems are not necessarily interpretable.

In the recent few years, many researches have been devoted to the study of the tradeoff between interpretability and precision [1,2,3-13]. Particularly, evolutionary computation, or genetic algorithm (GA), has received a lot of attention owing to its robustness and the ability to obtain global optimum solutions. GA is used widely to generate fuzzy rules and adjust the parameters of fuzzy systems [14]. However, when the simultaneous optimization of the antecedents and the number of rules and other factors concerning fuzzy modeling is needed, it is difficult to achieve a satisfactory fuzzy system, because the chromosome represented the complete solution is too long and the learning performance is deteriorated. The hierarchical genetic algorithm [4,15] and the cooperative coevolutionary algorithm [13] are two promising approaches to solve this problem.

This paper proposes a new Pareto Multi-Objective Cooperative Coevolutionary Algorithm (PMOCCA) to construct multiple Pareto-optimal fuzzy systems from numerical data, considering both interpretability and precision. 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. Then, the PMOCCA and 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. In this step, different from traditional weighted sum method in which multiple objectives are combined to one objective and only a single fuzzy system can be obtained in one run, we propose a novel non-dominated sorting technique that can find a set of Pareto-optimal fuzzy systems in a single run based on NSGA-II. The advantages of the paper are: (1) the interpretability-driven simplification techniques are used to simplify the fuzzy sets and fuzzy rules, thus the interpretability of the fuzzy system is improved. (2) The number of rules, the antecedents of the fuzzy rules and the parameters of the antecedents are

Manuscript received March 22, 2006; revised May 13, 2007; accepted June 19, 2007. Recommended by Editorial Board member Eun Tai Kim under the direction of Editor Jin Young Choi. This work was supported by National Science Foundation of P. R. China (60332020), China Postdoctoral Science Foundation.

Zong-Yi Xing, Yong Zhang, and Yuan-Long Hou are with the School of Mechanical Engineering/Automation, Nanjing University of Science and Technology, Jiangsu, China (e-mails: xingzongyi@gmail.com, zy69813@gmail.com, yuanlongh@sohu.com).

Jia li-Min is with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China (e-mail: jialm@vip.sina.com).

optimized simultaneously by the PMOCCA. (3) A set of Pareto-optimal solutions can be acquired in a single run so that the decision-maker can choose the most appropriate solution. (4) The obtained fuzzy systems are interpretable and accurate.

This paper is organized as follows. In Section 2, we review the TS fuzzy system and interpretability issues in fuzzy system. Section 3 constructs an initial fuzzy system based on fuzzy clustering algorithm. The interpretability-driven simplification techniques are described in Section 4. The complete PMOCCA is detailed in Section 5. Section 6 provides some experiments and results before concluding in Section 7.

2. PRELIMINARIES

2.1. TS fuzzy system

The typical fuzzy rule of the Takagi-Sugeno (TS) Fuzzy model [16] has the form:

$$R_i : \text{if } x_1 \text{ is } A_{i1}, x_2 \text{ is } A_{i2}, \cdots, x_n \text{ is } A_{in}$$
$$\text{then} \quad \hat{y}_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + a_{i0},$$

where x_j is the j-th input variable, A_{ij} is the fuzzy set of the j-th input variable in i-th rule, y_i are output of the i-th fuzzy rule.

The output of the TS fuzzy model is computed using the normalized fuzzy mean formula:

$$y(k) = \sum_{i=1}^{c} p_i(x)\hat{y}_i,$$

where c is the number of rules, and P_i is the normalized firing strength of the ith rule:

$$p_i(x) = \frac{\prod_{j=1}^{n} A_{ij}(x_j)}{\sum_{i=1}^{c} \prod_{j=1}^{n} A_{ij}(x_j)}.$$

In this paper, the Gaussian membership function is used to represent the fuzzy set A_{ij}

$$A_{ij}(x_j) = \exp(-\frac{(x_j - v_{ij})^2}{2\sigma_{ij}^2}),$$

where v_{ij} and σ_{ij} represent the center and the variance of the Gaussian function respectively.

2.2. Interpretability issues in fuzzy systems

Different from the objective property of precision, interpretability is a subjective property of fuzzy systems, which depends on several factors. Although there is no formal definition for interpretability, several characteristics are believed to be essential. These are described as follows [1-8]:

1) The number of variables and rules: a high-dimensional fuzzy system is difficult to interpret. The fuzzy system should use as few variables as possible. A fuzzy system with a large rule base is less interpretable than a fuzzy system containing only few rules. Experientially, the number of fuzzy rules of an interpretable fuzzy system is no more than ten, which is determined by the limit of human intelligence.
2) Characteristics of fuzzy rules: each rule should employ the fewest possible membership functions (fuzzy sets), i.e., variables. For all fuzzy rules, they should be consistent with one another. For any effective input, at least one fuzzy rule should be fired.
3) Characteristics of membership functions: each membership function should be convex. The adjacent membership functions should be moderately overlapped, and the overlapped value cannot be too large or small. Generally, 0.5 is a good choice. Membership functions of any variable should cover the whole universe. The number of membership functions should be compatible with the number of conceptual entities which a human being can handle.

3. CONSTRUCTON OF AN INITIAL FUZZY SYSTEM

In order to guarantee the effective of the PMOCCA, a good initial fuzzy system is preferred. Fuzzy clustering algorithm is a well-recognized technique to identify fuzzy systems. The fuzzy C-Means algorithm [17] and the Gustafson-Kessel algorithm [18] 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 [19] 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:

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

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

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

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

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

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

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 [20]:

$$X_e = [X \quad 1] \tag{8}$$

then

$$[a_{i1}, \cdots a_{ip}, a_{i0}] = [X_e^T U_i X_e]^{-1} X_e^T U_i y \tag{9}$$

is the consequent parameter of the TS fuzzy system.

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 clusters (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).
4) Calculate the norm of distance utilizing (6).
5) Update the partition matrix U using (3).
6) Stop if $\left\| U^{(l)} - U^{(l-1)} \right\| \leq \varepsilon$, else go to 3).

4. INTERPRETABILITY-DRIVEN SIMPLIFICATION TECHNIQUES

4.1. Simplification of fuzzy sets

Fuzzy systems obtained by the fuzzy clustering and the PMOCCA may contain redundant information in terms of similarity between fuzzy sets (membership functions). The redundancy makes the fuzzy systems uninterruptible, for it is difficult to assign meaningful terms 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 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 [21] is defined as

$$S(A,B) = \frac{\sum_{k=1}^{N} [\mu_A(x_k) \wedge \mu_B(x_k)]}{\sum_{k=1}^{N} [\mu_A(x_k) \vee \mu_B(x_k)]}, \tag{10}$$

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

$$\begin{cases} v_c = (v_A + v_B)/2 \\ \sigma_c = \sqrt{\sigma_A^2 + \sigma_B^2} / 2. \end{cases} \tag{11}$$

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.

4.2. Simplification of 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 : if x_1 is $\mu_{i1}(x_1)$, x_2 is $\mu_{i2}(x_2)$, $\cdots$, x_n is $\mu_{in}(x_n)$

then $\hat{y}_i =$

R_j : if x_1 is $\mu_{j1}(x_1)$, x_2 is $\mu_{j2}(x_2)$, $\cdots$, x_n is $\mu_{jn}(x_n)$

then $\hat{y}_j = \cdots$

then a similarity measure of rules [8] is defined as

$$S_R(R_i, R_j) = \min_{k=1}^{n} S(\mu_{ik}, \mu_{jk}), \quad (12)$$

where $S(\cdot)$ is calculated with the formula (10).

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.

5. PARETO MULTI-OBJECTIVE COOPERATIVE COEVOLUTIONARY ALGORITHM (PMOCCA)

Coevolutionary algorithm refers to the simultaneous evolution of two or more species with coupled fitness. In coevolutionary algorithm, the fitness of one individual depends on the fitness of individuals of other species and on its interaction with them. Coevolutionary algorithm can be classified into competitive coevolutionary algorithm and cooperative coevolutionary algorithm. In competitive coevolutionary algorithm, the species compete with each other; while in cooperative coevolutionary algorithm, the species interact with other species to improve their survival. This paper studies on how to build interpretable and accurate fuzzy systems using the Pareto multi-objective cooperative coevolutionary algorithm (PMOCCA).

There are five issues in the PMOCCA which are elaborated in this section: (1) problem decomposition; (2) collaboration formation of multi-species; (3) multi-objective function of the PMOCCA; (4) non-dominated sorting method and collaboration mechanism; (5) evolutionary operators.

5.1. Problem decomposition: species and encoding

Coevolutionary algorithm is used to deal with the simultaneous optimization of antecedents of fuzzy rules and parameters of fuzzy sets. So the fuzzy system is decomposed into two species: antecedents of fuzzy rules (*Species A*) *and* parameters of fuzzy sets (*Species B*).

5.1.1 *Species A*: coding and initial population

Species A represents the index of fuzzy sets that appear in the antecedent of the fuzzy rule with binary code. The phenotype of a chromosome is demonstrated as following:

β_{11}	$\cdots$	β_{1n}	$\cdots$	β_{ij}	$\cdots$	β_{c1}	$\cdots$	β_{cn}

where $\beta_{ij} = \{0, 1, 2, \cdots, c\}$. The number of 0 is utilized to represent the "*don't care*" [22] condition; while other values represent that the j-th fuzzy set of i-th rule is accepted.

The initial population is generated by introducing a chromosome that represents the antecedents of the obtained initial fuzzy system. The remaining chromosomes are generated randomly.

5.1.2 *Species B*: coding and initial population

Species B contains all parameters of the membership functions (fuzzy sets) defined in the fuzzy system with real code. The first chromosome is formed as a sequence of genes describing parameters in the rule antecedent:

$$H_1 = (v_{11}, \cdots, v_{cn}, \sigma_{11}, \cdots \sigma_{cn}). \quad (13)$$

Given search space $[H^{\min}, H^{\max}]$:

$$H^{\min} = (v_{11}^{\min}, \cdots, v_{cn}^{\min}, \sigma_{11}^{\min}, \cdots \sigma_{cn}^{\min}), \quad (14)$$

$$H^{\max} = (v_{11}^{\max}, \cdots, v_{cn}^{\max}, \sigma_{11}^{\max}, \cdots \sigma_{cn}^{\max}), \quad (15)$$

where $v_{ij}^{\max}$, $v_{ij}^{\min}$, $\sigma_{ij}^{\min}$, $\sigma_{ij}^{\max}$ are maximum and minimum values of corresponding membership functions. The remaining chromosomes are created by random variation (uniform distribution) around H_1 within the search space.

5.2. Collaboration formation of multi-species

In the PMOCCA, individuals of multi-species collaborate with each other to jointly provide antecedents of the fuzzy rules. In order to explain how the species form collaborations explicitly, we give an example with five fuzzy rules and two variable, i.e., c=5 and n=2. Two illustrational individuals are selected from each species as:

0,0		3,0		1,4		1,5		1,4	
v_{11}		$\cdots$		v_{52}	σ_{11}		$\cdots$		σ_{52}

where the first and second line represent individuals from *Species A* and *Species B* respectively. Then antecedents of the five fuzzy rules are:

$$
\begin{aligned}
&R_1 : \\
&R_2 : \text{if } x_1 \text{ } is \text{ } A_{21}(v_{31}, \sigma_{31}) \\
&R_3 : \text{if } x_1 \text{ is } A_{31}(v_{11}, \sigma_{11}),\, x_2 \text{ is } A_{32}(v_{42}, \sigma_{42}) \\
&R_4 : \text{if } x_1 \text{ is } A_{41}(v_{11}, \sigma_{11}),\, x_2 \text{ is } A_{42}(v_{52}, \sigma_{52}) \\
&R_5 : \text{if } x_1 \text{ is } A_{51}(v_{11}, \sigma_{11}),\, x_2 \text{ is } A_{52}(v_{42}, \sigma_{42}),
\end{aligned}
$$

where $R_1,\ldots,R_5$ represent the collaborated fuzzy rules, x_1, x_2 are two variables.

The first and second genes of the first individual (upper line) are zero, which indicates that none of any fuzzy set is fired, and the first fuzzy rule is empty, and is excluded from rule base. The fourth gene of the first individual is zero, which shows that x_2 in R_2 is a *don't care* term. All parameters between R_3 and R_5 are identical, so R_3 and R_5 should merge to one rule in the process of simplification. Finally, the optimized fuzzy system contains only two fuzzy rules.

The jointly obtained antecedents above, combined with consequents computed by formulas (9), integrate the complete fuzzy system.

5.3. Multi-objective function

In a multi-objective problem, these objectives are often conflict with each other, so there is no single optimal solution, and the goal is to find a set of Pareto-optimal solutions.

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, while it is difficult to quantify interpretability. According to the analysis about interpretability in Section 2, we have guaranteed characteristics of fuzzy sets (membership functions) by interpretability-driven methods, so only the number of rules and the number of fuzzy sets are included in the objective function.

These mentioned objectives are detailed following.

Precision: the mean square error (*MSE*) is used as precision criterion:

$$f_1(S) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{16}$$

where N is the number of data, $\hat{y}_i$ is output of fuzzy system, y_i is the measured output.

Interpretability: the number of fuzzy rules $f_2(S)$ and the number of fuzzy sets $f_3(S)$.

The three objectives about fuzzy system can be formulated as follows

$$Min\ f_1(S),\ \ \min\ f_2(S),\ \ \min\ f_3(S), \tag{17}$$

where $f_1(S)$ is precision, $f_2(S)$ 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:

$$
\begin{aligned}
&f_1(S_A) \le f_1(S_B), f_2(S_A) \\
&\le f_2(S_B), f_3(S_A) \le f_3(S_B)
\end{aligned}
\tag{18}
$$

and at least one of the following inequalities holds:

$$f_1(S_A) < f_1(S_B), \tag{19}$$

$$f_2(S_A) < f_2(S_B), \tag{20}$$

$$f_3(S_A) < f_3(S_B). \tag{21}$$

The condition (18) shows that no objective of S_A is worse than S_B. Any inequality of (19)-(21) means that at least one objective of S_A is better than S_B. 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 [23], PAES [24] and SPEA [25]. 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 [23].

5.4. A new non-dominated sorting method and collaboration mechanism

A classical collaboration strategy of multi-objective cooperative coevolutionary algorithm has been proposed by Potter [26]. The multiple objectives are combined into a single objective using weighted sum method. The best individual and randomly selected individuals are selected as representatives. In order to determine fitness value of individual, the individual collaborates with representatives of other species to construct a set of antecedents of fuzzy systems. These obtained antecedents and consequents combine to form complete fuzzy systems. Then the objective values of these fuzzy systems are calculated and the minimum value is assigned to the individual.

For the sake of clarity, we give an example of how two species are collaborated with each other, where

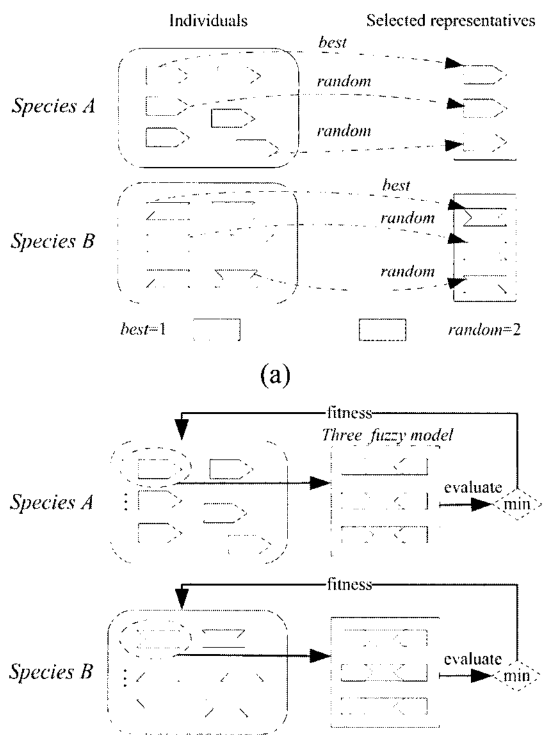


Fig. 1. Collaboration mechanism of single species.

the population size is 40 and there are three representatives. In Fig. 1(a), we gather the best individual and two randomly selected individuals acting as representatives for each species. In Fig. 1(b), one individual of *Species A* collaborate with representatives from *Species B*, thus forming three solutions (antecedents of fuzzy systems). The obtained antecedents and the consequents calculated by formula (9) combine to form complete fuzzy systems. These fuzzy systems are evaluated based on the weighted objective function and the minimum value is assigned as the final fitness to the individual being evaluated. In a similar way, fitness evaluation of other individuals in *Species A* and individuals of *Species B* are accomplished.

The algorithm introduced above can only obtain a single optimal solution, however there is always a set of Pareto-optimal solutions for a multi-objective problem. In order to obtain multiple fuzzy systems, we generalize the NSGA-II algorithm from single species to multi-species, and propose a new non-dominated sorting method and collaboration mechanism described in Fig. 2.

For the sake of clarity, we also use an example with two species to illustrate the PMOCCA, where the number of individuals of each species is 40, and the number of representatives is two.

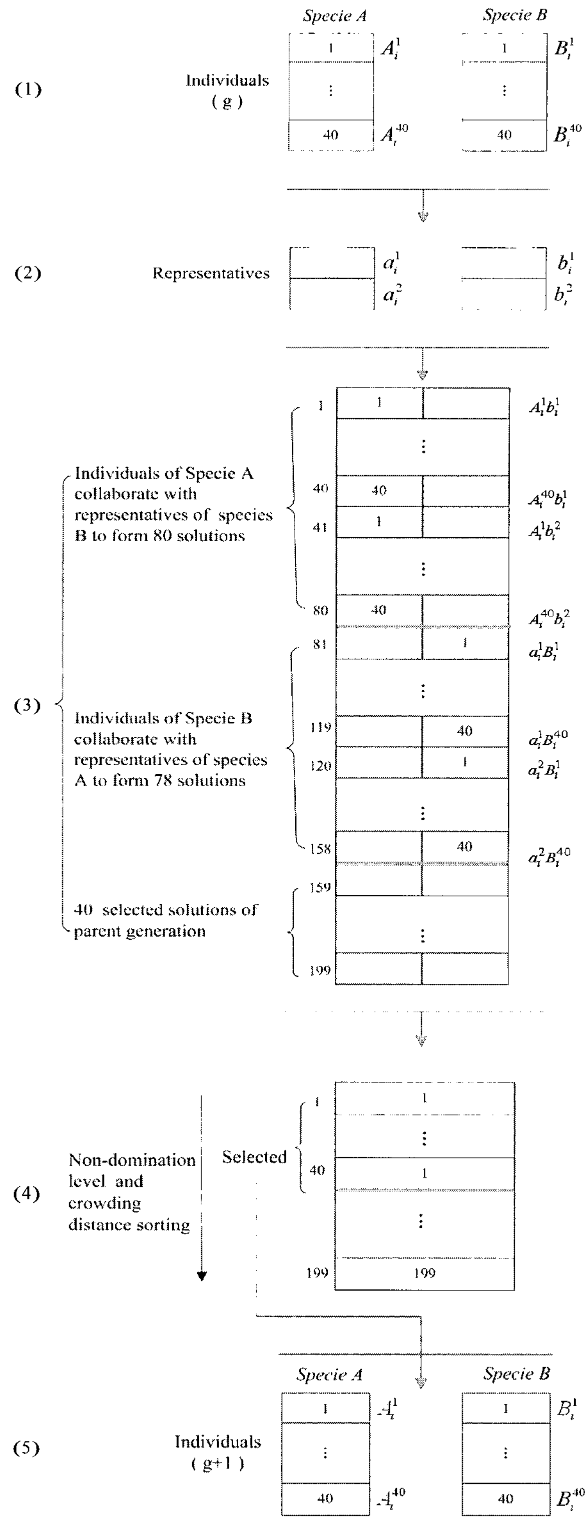


Fig. 2. Non-dominated sorting collaboration mechanism of the proposed PMCCA.

(1) The antecedents of the initial fuzzy system are decomposed into the first individual of the two species. Other individuals of *Specie A* are generated randomly, and other individuals of *Specie B* are created by uniform distribution within the search space.
(2) The first (best) individual and a random individual are selected as representatives of each species.

(3) The individuals in *Specie A* cooperate with representatives of *Specie B* to form antecedents of fuzzy system. Individuals of *Specie B* also construct antecedents in a similar way. The number of antecedent solutions is 158.
(4) After being simplified, the obtained antecedents and the consequents calculated by formula (9) are combined to form complete fuzzy systems. Based on NSGA-II, the non-dominated sorting and the crowding distance sorting are performed on the obtained fuzzy systems with the defined objective functions, i.e., precision, the number of fuzzy rules and the number of fuzzy sets, in descending order.
(5) The first 40 fuzzy systems are selected and decomposed into individuals of two species. The tournament selection, crossover and mutation are performed respectively on the two species to reproduce offspring subpopulations.
(6) If the stop criterion is not meet, go to step (2)~(5), otherwise the algorithm is stopped. Different from the first generation in step (3), the 40 solutions of parent generation are added to the sorting procedure, thus the total solutions is 199.

5.5. Evolutionary operators

Selection is the first operator applied in the proposed PMOCCA. For all these two species, a tournament technique is used to select individuals for the next generation. In order to avoid divergence of the algorithm, the tournament is combined with an elitist strategy to ensure that the best chromosome will be chosen.

The second operator is crossover. For *Species A*, the one-point crossover is adopted due to their binary-encoding. The simple arithmetic crossover, the whole arithmetic crossover and the heuristic crossover are selected randomly in *Species B*.

The role of third operator, i.e., mutation, is to introduce diversity into a population. For *Species A*, the bit inversion mutation is used due to their binary-encoding. The uniform mutation, the multiple uniform mutation and the Gaussian mutation are selected randomly in *Species B*.

5.6. Pseudo-code of the algorithm

The pseudo-code of PMOCCA is present in the following

Begin
  g:=0
    Initialize fuzzy system $F(0)$
    $P(1)$=decompose($F(0)$)
  While not done **do**
    g:=g+1
    For each species S, $S = A, B$
      $R_S(g)$ = Represent ($P(g)$)
      $A(g)$ = Collaborate($R_S(g)$, $P(g)$)
      $A'(g)$ = Simplify($A(g)$)
      $C'(g)$ = Calculate($A'(g)$)
      $F_S'(g)$ = Compose($A'(g)$, $C'(g)$)
    End for
    $F(g)$ = NSGA-II ($F_A'(g)$, $F_B'(g)$, $F(g-1)$);
      % $F(0) = \Phi$
    $P'(g)$ = decompose($F(g)$)
    $P(g+1)$ = Operate ($P'(g)$);
  End while
End

where decompose() converts antecedents of fuzzy system into species, decompose() converts population into antecedents of species, Represent() selects representatives of species, Collaborate() means that representatives collaborate with individuals to form antecedents of fuzzy systems, simplify() reduces the obtained antecedents of fuzzy systems, calculate() identify consequents of fuzzy systems, NSGA-II() generates multiple fuzzy systems based on non-domination level and crowding distance, operate() executes three evolutionary operators.

6. EXPERIMENTS AND RESULTS

In order to examine the performance of the proposed PMOCCA, two benchmark problems, the second-order nonlinear plant and the Mackey-Glass time series, are demonstrated in this section. Table 1 gives the parameter setups of the algorithm. All simulation programs are realized under Matlab 7.0 environment.

6.1. Experiment: the second-order nonlinear plant

We consider the second-order nonlinear plant studied in [4,27-29]:

$$y(k) = g(y(k-1), y(k-2)) + u(k), \tag{35}$$

where

Table 1. Parameter setups of the PMOCCA.

Parameters	Values
Maximum generations	100
Population size	40
Crossover probability of *Species A*	1
Mutation probability of *Species A*	0.2
Elitism rate of *Species A*	0.1
Crossover probability of *Species B*	0.85
Mutation probability of *Species B*	0.1
Elitism rate of *Species B*	0.025
Best representatives N_{cf}	1
Random representatives N_{cr}	1
Threshold of merging fuzzy sets	0.4
Threshold of merging fuzzy rules	1

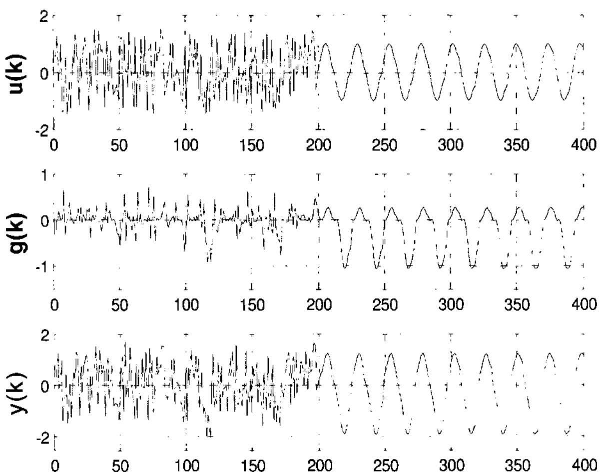


Fig. 3. Input, unforced system and output of the second-order nonlinear plant.

$$g(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)(y(k-1)-0.5)}{1+y^2(k-1)+y^2(k-2)}. \tag{36}$$

The goal is to approximate the nonlinear component $g(y(k-1), y(k-2))$ of the plant with a fuzzy system. For this purpose, 400 simulated data are generated from the plant (35). The first two hundreds of training data are obtained with a random input signal $u(k)$ uniformly distributed in the interval [-1.5 1.5], while the other two hundreds of validation data are generated by using a sinusoid input signal $u(k) = \sin(2\pi k/25)$. The total simulated data are shown in Fig. 3.

The initial fuzzy system is obtained by the fuzzy clustering and the least square method. In order to compare performance of the algorithm with other published results, the number of fuzzy clusters is five.

The simulation results are summarized in Table 2 where four Pareto-optimal fuzzy systems are generated. The results indicate that the proposed method can obtain multiple accurate and interpretable fuzzy systems.

Wang [4] proposed a new scheme based on multi-objective hierarchical genetic algorithm to extract interpretable fuzzy systems. The initial fuzzy system is constructed using fuzzy clustering, and then is optimized by the multi-objective genetic algorithm. The obtained fuzzy systems are almost equivalent to our results, but the 1# fuzzy system we constructed has the highest precision performance.

Yen [27] proposed several orthogonal transformation methods to select fuzzy rules. The initial fuzzy system contained 25 rules, among which only 20 rules were remained after simplification. This approach can construct accurate fuzzy systems, but the obtained fuzzy systems have too many fuzzy rules and fuzzy sets.

Table 2. Fuzzy systems of nonlinear plant.

Ref.	No. rules	No. sets	MSE (train)	MSE (validation)
[4]	5	10	1.4032e-3	2.6267e-3
	5	3	2.3773e-4	3.0116e-4
	4	3	5.4611e-4	5.4360e-4
	4	3	5.6086e-4	2.4885e-4
[27]	25	25	2.3092e-4	4.0717e-4
	20	20	6.8341e-4	2.3836e-4
[28]	5	10	4.9e-3	2.9e-3
	5	10	1.4e-3	5.9e-4
	5	5	8.3e-4	3.5e-4
[29]	4	2	1.40e-4	1.53e-4
	9	3	1.26e-5	1.2e-5
	16	4	1.50e-6	3.4e-6
This paper				
Initial	5	10	6.0e-3	6.6e-3
1#	4	4	5.9202e-5	3.9307e-5
2#	5	3	2.3013e-4	2.2895e-3
3#	4	3	3.3833e-4	6.0245e-4
4#	3	3	1.6233e-3	1.4781e-2

Table 3. 1# fuzzy system of nonlinear plant.

Fuzzy rules	R^1: **If** y(k-1) is big, y(k-2) is small, **then** g(k)=-0.37453y(k-1)+0.10912y(k-2) +0.28266 R^2: **If** y(k-1) is small, y(k-2) is big, **then** g(k)=-0.47701y(k-1)+0.09224y(k-2) -0.03940 R^3: **If** y(k-1) is big, y(k-2) is big, **then** g(k)=0.35558y(k-1)+0.099222y(k-2) -0.25662 R^4: **If** y(k-1) is small, y(k-2) is small, **then** g(k)=0.47591y(k-1)+0.1043y(k-2) +0.032723
Antecedent parameters	
y(k-1): small=[-0.83053, 0.56478], big=[0.92494, 0.43772] y(k-2): small=[-0.58531, 0.35528], big=[0.66867, 0.42768]	

Roubos [28] obtained an initial redundant fuzzy system by fuzzy clustering, and adopted the method of merging similar fuzzy sets and genetic algorithms to reduce the initial fuzzy model iteratively and finally used genetic algorithms to optimize all parameters of the fuzzy model. The two optimized fuzzy systems can reach high precision with few rules and sets, however, all these two fuzzy systems are dominated by our 1# fuzzy system.

Kim [29] identified antecedents of fuzzy systems using a unique evolutionary algorithm, and calculated consequents by least square method. The occurrence

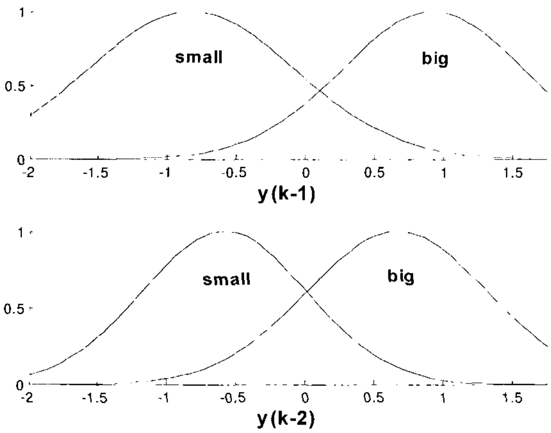


Fig. 4. Membership functions of the 1# fuzzy system of the nonlinear plant.

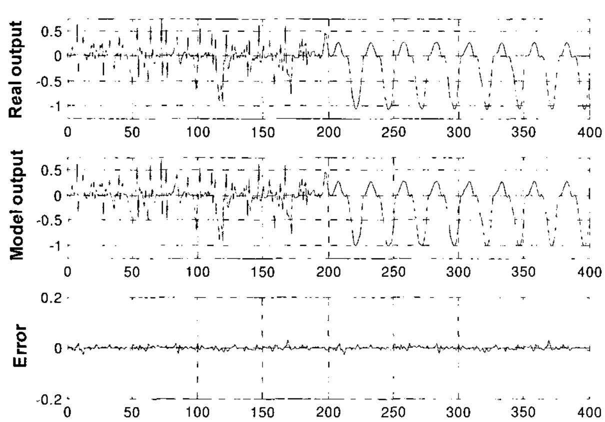


Fig. 5. Comparison of 1# fuzzy system outputs and real outputs of the nonlinear plant.

of the multiple overlapping membership functions was resolved with the help of the fitness function. The obtained fuzzy systems can be assigned meaningful labels to membership functions, and the highest precision performance is obtained by the fuzzy system with 16 rules.

In the obtained 1# fuzzy system, the MSE errors of training data and validation data are 5.9202e-5 and 3.9307e-5 respectively, which indicate that 1# fuzzy system has high precision and high generalization capacity. Compared to other published approaches in [4,27-29], the 1# fuzzy system also has high interpretability for its only contains four fuzzy rules and four fuzzy sets. Table 3 details the obtained 1# fuzzy system. Fig. 4 depicts the membership functions of 1# fuzzy system, and Fig. 5 compares the 1# fuzzy system outputs and real outputs.

6.2. Experiment: Mackey-Glass time series

The Mackey-Glass time series is described as:

$$\dot{x} = \frac{ax(t-\tau)}{1+x^{10}(t-\tau)} - bx(t), \tag{33}$$

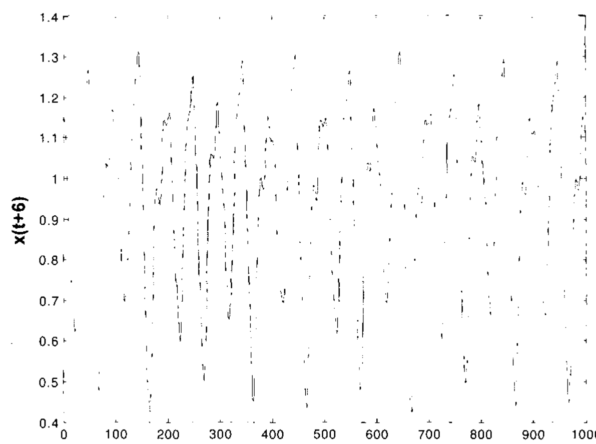


Fig. 6. Sampling data of $x(t+6)$ in the time series.

where a=0.2, b=0.1, and τ=17 as in [30]. 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 Fig. 6.

The initial fuzzy system is obtained by the fuzzy clustering and the least square method. The *MSE* error of training data is 0.0657, and the *MSE* error 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 methods 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 4. 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 4 also shows the comparison with other published results, which indicates that the proposed

Table 4. Fuzzy systems of nonlinear plant.

Ref.	No. rules	No. sets	MSE (train)	MSE (validation)
[4]	9	23	0.0228	0.0239
[29]	16	8	0.0014	0.0013
[30]	129	35	0.0315	0.0332
	26	19	0.0656	0.0671
This paper				
Initial	5	20	0.0657	0.0646
1#	5	7	0.0052175	0.0051334
2#	5	6	0.005299	0.0053031
3#	4	6	0.0064568	0.0063298
4#	5	5	0.0085456	0.0084214
5#	3	6	0.0085849	0.0085999
6#	4	5	0.0096844	0.009595

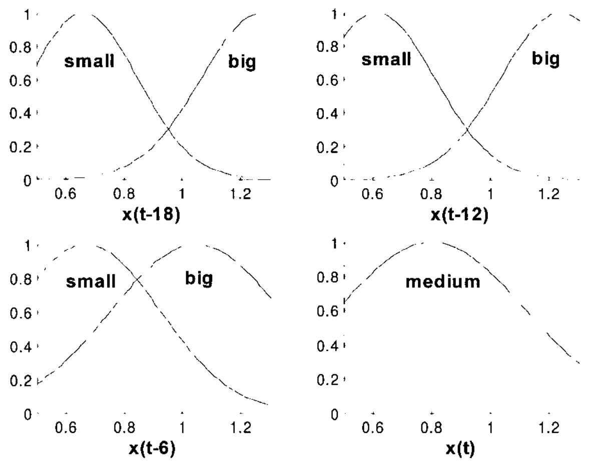


Fig. 7. Membership functions of the 1# fuzzy system of the time series.

Table 5. 1# fuzzy system of the time series.

Fuzzy rules	R1: **If** x(t-6) is small, x(t-12) is small, **then** x(t+6)=0.9831x(t)-43678x(t-6)+0.8414 x(t-12)-0.01937x(t-18)-0.02515 R2: **If** x(t) is medium, (t-12) is big, x(t-18) is small, **then** x(t+6)=-0.50397x(t)+ 0.99034x(t-6)-1.125x(t-12)+0.68576 x(t-18)+1.1839 R3: **If** x(t) is medium, x(t-6) is big, x(t-12) is big, x(t-18) is big, **then** x(t+6)= 0.31641x(t)+ 0.15879x(t-6)-0.46131x (t-12)+0.07292x(t-18)+0.6029 R4: **If** x(t) is medium, x(t-6) is small, x(t-18) is big, **then** x(t+6)=-0.3874x(t) +0.8729x(t-6)-1.1872x(t-12)-1.2476 x(t-18)+2.8725 R5: **If** x(t) is medium , x(t-12) is small, x(t-18) is small, **then** x(t+6)=0.4755x(t) +0.17824x(t-6)+0.4111x(t-12)+0.35525 x(t-18)-0.05043
Antecedent parameters	
x(t): medium =[0.79914,0.10275] x(t-6): small=[0.67052, 0.063765] big=[1.0471,0.085645] x(t-12): small=[0.61178,0.03998] big=[1.2375,0.04141] x(t-18): small=[0.66109, 0.03526] big=[1.2703, 0.04251]	

method can obtain multiple accurate and interpretable fuzzy systems. The fuzzy systems obtained by Kim [29] have highest precision performance; however these fuzzy systems use up to 16 fuzzy rules.

Table 5 details the obtained 1# fuzzy system of time series. Fig. 7 depicts the membership functions of the 1# fuzzy system, and Fig. 8 compares the 1# fuzzy system outputs and real outputs of time series.

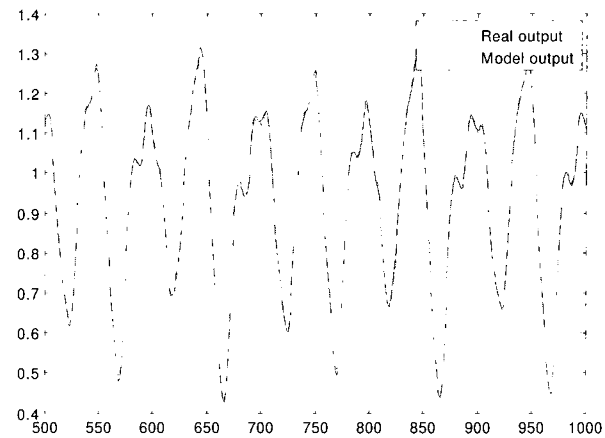


Fig. 8. Comparison of 1# fuzzy system outputs and real outputs of the time series.

7. CONCLUSIONS

In this paper, we presented an approach, named PMOCCA, to construct accurate and interpretable fuzzy systems. First, preliminaries, including the TS fuzzy system and the interpretability issues, are stated. Then, a modified fuzzy clustering algorithm is used to construct antecedents of fuzzy system, and consequents are identified separately to reduce computational burden. Finally, the PMOCCA and the interpretability-driven simplification are employed to evolve the initial fuzzy system iteratively; resultantly, multiple Pareto-optimal fuzzy systems with high precision and interpretability are obtained. The simulation results on two benchmark problems illustrate validity of the method.

REFERENCES

[1] J. Casillas, O. Cordón, and F. Herrera, *Interpretability Issues in Fuzzy Modeling*, Springer, 2003.

[2] J. Casillas, O. Cordón, and F. Herrera, *Accuracy Improvements in Linguistic Fuzzy Modeling,* Springer, 2003.

[3] R. Mikut, J. Jakel, and L. Groll, "Interpretability issues in data-based learning of fuzzy systems," *Fuzzy Sets and Systems*, vol. 150, no. 2, pp. 179-197, 2005.

[4] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, "Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction," *Fuzzy Sets and System*, vol. 149, no. 1, pp. 149-186, 2005.

[5] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through

iterative complexity reduction," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 4, pp. 516-524, 2001.

[6] D. D. Nauck, "Fuzzy data analysis with NEFCLASS," *Approximate Reasoning*, vol. 32, no. 2-3, pp. 103-130, 2003.

[7] Y. Jin, "Fuzzy modeling of high-dimensional systems complexity reduction and interpretability improvement," *Fuzzy Sets and Systems*, vol. 8, no. 2, pp. 212-221, 2000.

[8] Y. Jin, Advanced *Fuzzy Systems Design and Applications*, Physical-verl, New York, 2003.

[9] O. Kaynak, K. Jezernik, and A. Szeghegyi, "Complexity reduction of rule based models: a survey," *Proc of IEEE Int. Conf. on Fuzzy Systems*, pp. 1216-1222, 2002.

[10] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans on Fuzzy Systems*, vol. 9, no. 3, pp. 426-443, 2001.

[11] T. Sudkamp, A. Knapp, and J. Knapp, "Model generation by domain refinement and rule reduction," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 33, no. 1, pp. 45-55, 2003.

[12] M. Setnes and H. Hellendoorn, "Orthogonal transforms for ordering and reduction of fuzzy rules," *Proc of IEEE Int. Conf. on Fuzzy Systems*, pp. 700-705, 2000.

[13] C. A. Pena and M. Sipper, "FuzzyCoCo: A cooperative-coevolutionary approach to fuzzy modeling," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 5, pp. 727-737, 2001.

[14] O. Cordon and F. Herrer, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Rule Bases*, World Scientific, Singapore, 2000.

[15] M. R. Delgado and J. Fernando, "Hierarchical genetic fuzzy systems," *Information Sciences*, vol. 136, no. 1-4, pp. 29-52, 2001.

[16] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. on Systems Man and Cybernetics*, vol. 15, no. 1, pp. 116-132, 1985.

[17] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Algorithm*, Plenum, New York, 1981.

[18] D. Gustafson and W. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *Proc. of IEEE Conf. on Decision and Control*, pp. 761-766, 1979.

[19] J. Abonyi, B. Babuska, and F. Szeifert, "Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models," *IEEE Trans. on Systems, Man and Cybernetics, Part B*, vol. 32, no. 5, pp. 612-621, 2002.

[20] R. Babuska, *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Boston, 1998.

[21] M. Setnes and R. Babuska, "Similarity measures in fuzzy rule base simplification," *IEEE Trans. on Systems Man and Cybernetics*, vol. 28, no. 3, pp. 376-386, 1998.

[22] H. Ishibushi, T. Nakashima, and T. Murata, "Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 5, pp. 601-618, 1999.

[23] K. Deb, A. Pratap, and S. Agarwal, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002.

[24] J. D. Knowles and D. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," *Evolutionary Computation*, vol. 8, no. 2, pp. 149-172, 2000.

[25] E. Zitzler and L. Thiele, "Multi-objective evolutionary algorithms: A comparative case study and the strength Pareto approach," *IEEE Trans. on Evolutionary Computation*, vol. 3, no. 4, pp. 257-271, 1999.

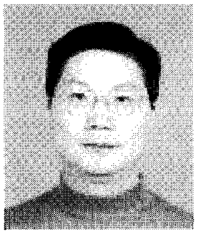
[26] M. Potter and A. D. Kenneth, "Cooperative coevolution: Architecture for evolving coadapted subcomponents," *Evolutionary Computation*, vol. 8, no. 1, pp. 1-29, 2000.

[27] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetic*, vol. 29, no. 1, pp. 13-24, 1999.

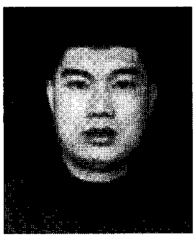
[28] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 4, pp. 516-524, 2001.

[29] M. S. Kim, C. H. Kim, and J. J. Lee, "Building a fuzzy model with transparent membership functions through constrained evolutionary optimization," *International Journal of Control, Automation, and Systems*, vol. 2, no. 3, pp. 298-309, 2004.

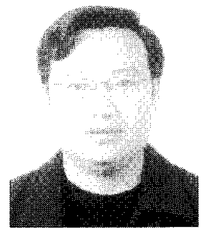
[30] R. P. Paiva and A. Dourado, "Interpretability and learning in neuro-fuzzy systems," *Fuzzy Sets and Systems*, vol. 147, no. 1, pp. 17-38, 2004.



Zong-Yi Xing received the doctor degree in China Academy of Railway Science and Technology in 2003. Currently, he is a Teacher in Nanjing University of Science and Technology. His research interests include fuzzy modeling, industry process control, and servo system.



Yong Zhang received the doctor degree in Nanjing University of Science and Technology in 2006. Now, he is a Teacher in Nanjing University of Science and Technology. His research interests include fuzzy modeling and intelligent control.



Yuan-Long Hou received the master degree in Nanjing University of Science and Technology. Now, he is a Professor in Nanjing University of Science and Technology. His research interests include intelligent control, servo system.

Li-Min Jia received the doctor degree in China Academy of Railway Science and Technology in 2003. Currently, he is a Professor in Beijing Jiaotong University. His research interests include computational intelligence, Security Technology transportation intelligent technology and system.