SECURING INTERPRETABILITY OF FUZZY MODELS FOR MODELING NONLINEAR MIMO SYSTEMS USING A HYBRID OF EVOLUTIONARY ALGORITHMS

M. EFTEKHARI, M. EFTEKHARI, M. MAJIDI AND H. NEZAMABADI POUR

Abstract. In this study, a Multi-Objective Genetic Algorithm (MOGA) is utilized to extract interpretable and compact fuzzy rule bases for modeling nonlinear Multi-input Multi-output (MIMO) systems. In the process of nonlinear system identification, structure selection, parameter estimation, model performance and model validation are important objectives. Furthermore, securing low-level and high-level interpretability requirements of fuzzy models is especially a complicated task in case of modeling nonlinear MIMO systems. Due to these multiple and conflicting objectives, MOGA is applied to yield a set of candidates as compact, transparent and valid fuzzy models. Also, MOGA is combined with a powerful search algorithm namely Differential Evolution (DE). In the proposed algorithm, MOGA performs the task of membership function tuning as well as rule base identification simultaneously while DE is utilized only for linear parameter identification. Practical applicability of the proposed algorithm is examined by two nonlinear system modeling problems used in the literature. The results obtained show the effectiveness of the proposed method.

1. Introduction

Identification of a nonlinear dynamical process is the art of finding a mathematical model which specifies that process accurately. Structure selection, parameter estimation, model performance and model validation are important objectives in the process of non-linear system identification. Therefore, some conflicting objectives must particularly be considered in the identification task when the process is MIMO and nonlinear [20, 14, 24]. In the case of nonlinear system identification, many mathematical models have been proposed in the literature among which Takagi-Sugeno and Kang (TSK) fuzzy inference systems have attracted a great attention [18, 1, 16]. This popularity is due to the transparency, simplicity and no need for defuzzification in these models [18, 1]. Recently, some studies have been performed in order to develop usages for TSK models in MIMO nonlinear system identification [5, 7, 6]. The main goal of these studies is to achieve objectives of system identification as well as securing the interpretability of fuzzy model. In a recent research, a twofold taxonomy of interpretability has been presented in the

Received: October 2010; Revised: February 2011; Accepted: March 2011

Key words and phrases: Multi-objective, Evolutionary, Fuzzy identification, Compact, Interpretability.

literature based on the previous studies in this field [26]. According to this classification, low-level and high-level interpretability are introduced. The low-level interpretability defines some constraints for the sake of securing interpretability at fuzzy set level, while the high-level interpretability is defined at the fuzzy rule level. As mentioned in [26], distinguishability is the most important semantic constraint for achieving low-level interpretability. According to Occam's razor parsimony principle in machine learning, from all models that can describe a process accurately the simplest one is the best. So, parsimony is a basic criterion for achieving high-level interpretability. Therefore, distinguishability and parsimony are the first concerns toward achieving both low and high-level interpretability separately. Obtaining a suitable interpretability is a complex and computationally prohibitive problem especially when the goal is to design approximate fuzzy models [5, 6]. Recently, a post-processing multi-objective evolutionary algorithm has been proposed which performs rule selection and tuning of fuzzy rule based systems with three objectives: accuracy, low-level interpretability and parsimony [10]. As mentioned in [10], the well-known Wang and Mendel method has been firstly applied to obtain an initial Mamdani rule base. Secondly, they have tried to keep the partitions and the meanings of membership functions to their original values by defining suitable low-level interpretability measures namely semantic based indices. Furthermore, one classical complexity (i.e. parsimony) measure has been considered for reducing number of rules. Several researches related to interpretability of fuzzy models have been reported in [10] as well. Moreover, there has been a great interest in developing methodologies for securing the interpretability of TSK fuzzy models in both low and high levels [5, 6, 27, 19]. In one of these recently published researches, a linguistic modifier has been proposed to generate distinguishable fuzzy sets and to obtain good local model interpretability [27]. The linguistic modifier has been utilized to characterize membership functions whose centers and shapes can be updated automatically [27]. Besides, securing local model interpretability could have a side effect on global model accuracy [27]. Finally, a hybrid objective function has been proposed by which a tradeoff between global approximation ability and local model interpretation can be achieved [27]. Another recently reported study has been performed in order to address one specific aspects of complexity reduction, i.e. reduction of the number of output MFs in 0-th order TSK models [19]. In all of these studies accuracy and transparency are two main objectives that are simultaneously taken into account. Owing to many objectives in nonlinear system identification task, developing transparent fuzzy models for this task induces a lot of complexities [5, 6]. Efforts have been devoted by the authors in order to extract interpretable fuzzy models for nonlinear MIMO processes by means of single objective EAs [5, 6]. In these studies both accuracy and complexity indices are formulated into a single weighted objective called information criterion [5, 6].

This paper uses a hybrid of two powerful evolutionary search techniques, MOGA and DE, for the purpose of nonlinear system identification. The most important advantage of MOGA is its ability to handle multiple conflicting objectives by the concept of Pareto-optimal set of the solutions [2]. In addition, objectives in MOGA can be considered subject to several constraints. Moreover, constraints can be formulated as distinct objectives beside the main objectives of problems. Also, DE is a greedy, fast and easy to implement evolutionary algorithm which does not require parameter control techniques; Since DE has two parameters namely crossover probability and a scaling factor called F. The MOGA-fuzzy algorithm employed in this work is similar to the Pittsburgh approach [15]. In the proposed algorithm, each chromosome in the population represents a complete knowledge base (KB) except the consequence parts of the rules. In other words, the MOGA algorithm is only allowed to determine the number of rules and the antecedent parts. Then, for each chromosome in GA, the DE is employed to compute the rule consequents instead of GA. Thus, it can greatly improve the search efficiency of GA and exploits the training data in a more effective way. The proposed algorithm then calculates the performance and compactness criteria to evaluate the fitness of each chromosome. In order to produce valid and interpretable models in solution space, model validity and transparency criteria are imposed as some constraints into the performance objective.

2. **Fuzzy Modeling**

Fuzzy modeling and identification from input-output process data has proved to be effective for uncertain nonlinear dynamic systems [1, 5, 6]. The most frequently applied model which is Takagi-Sugeno Fuzzy Inference System (TS-FIS), tries to decompose the input space of the nonlinear model into fuzzy subspaces and then approximate the system in each subspace by a simple linear regression model [1]. Hence, TS models are often used to represent nonlinear dynamic systems by interpolating between local Linear Time-Invariant (LTI) and Auto-Regressive with eXogenous (ARX) input models. So far, most of the attention have been devoted to single input, single output (SISO) or multi-input, single-output (MISO) systems. Recently, some methods for multi-input, multi-output systems (MIMO) have been investigated in [5, 7, 6].

The aim of constructing an FIS is to obtain a set of fuzzy rules that describe the system behavior as accurate as possible, given a set of operating data and if available an initial set of linguistic rules collected from experts. In this paper, a general fuzzy ARX structure is assumed to describe the system. In such a structure, the system is represented by the following nonlinear vector function:

$$\hat{y} = h(y(k-1), ..., y(k-n_p), u(k-n_d), ..., u(k-n_q-n_d+1)) \qquad (1)$$

where,h represents the nonlinear model,$y = [y_1, y_2, ..., y_{ny}]^T$ is an n_y dimensional output vector, $\hat{y}$ is the one-step-ahead predicted output vector,$u = [u_1, u_2, ..., u_{nu}]^T$ is an n_u dimensional input vector, n_p and n_q are maximum lags considered for the outputs and inputs, respectively, and n_d is the minimum discrete dead time.

3. **Evolutionary Search in Fuzzy System Identification**

3.1. **Multi-objective Approach.** Due to the ability of Evolutionary Algorithms (EAs) in exploring huge search spaces and exploiting global optimums in reasonable time, these algorithms have been employed widely in recent researches. On the other hand, in the case of modeling, obtaining a tradeoff between the accuracy and the parsimony of a model is an important problem in the field of machine learning research which severely needs optimization techniques. In recent years, EAs have been extensively used in order to develop fuzzy models regarding Occam's razor principle [20, 18]. The problem of nonlinear system identification (i.e. modeling dynamical nonlinear processes) is an optimization task that inherently involves optimizing multiple conflict objective functions. Thus, nonlinear system identification by means of simple and transparent fuzzy models is a complicated optimization problem with multiple conflicting objectives [20, 14, 5, 6, 10, 27]. A comprehensive literature review regarding multi-objective EAs for designing fuzzy systems has been provided in [14, 10, 27].

The MOGAs, proposed by Fonseca and Fleming [8, 9] are especially appropriate to solve these problems because a comparison is made in these algorithms to determine whether one solution dominates the other or not, and they can return a set of Pareto solutions in a single run of the algorithm. MOGA algorithm uses the concept of dominated solutions to find the Pareto-optimal solutions. Definition of the dominated solution described in [2] is as follows: For two solutions S^1 and S^2, a solution S^2 is said to dominated by the other solution S^1 (i.e. S^1 is a non-dominated solution), if the following conditions are satisfied:

1) The solution S^1 is no worse than S^2 in all objectives.
2) The solution S^1 is strictly better than S^2 in all objectives.

3.2. **Differential Evolution.** A recent and extremely simple EA is the so-called Differential Evolution (DE) suggested by Storn and Price [5, 21, 22]. Recent studies in the field of nonlinear system identification by fuzzy models confirm the applicability of DE in this area of research [5]. Additionally, DE is much simpler to implement than the other EAs and it appears to require less parameter tuning. DE algorithms create new individuals by adding the vector difference between two randomly chosen individuals to a third one. DE uses both crossover and mutation. However, both operations need to be redefined in the DE context. DE attempts to create p^m, a mutated form of any individual p^1, using the vector difference of randomly picked individuals. There are several variants of the original DE, the particular one used in this work is the DE/rand/1 algorithm proposed by [21, 22] in which the vector difference of two randomly selected individual p^2 and p^3 are used such that:

$$p^m = p^1 + F(p^2 - p^3) \tag{2}$$

Where F is the scaling factor, usually between 0 and 2. In DE this operation is known as mutating with vector differentials. Next, the crossover is applied between any individual member of the population p^4 and the mutated vector p^m, which is done essentially by swapping the vector elements in the corresponding location. This is done probabilistically and the decision of doing (or not doing) the crossover is considered by a crossover rate $0 < \eta \leq 1$. The new vector p^5, produced in this way, is known as the trial vector. It has to be assured that the trial vector inherits at least one variable from the mutated vector p^m, so that it does not become an exact replica of the original parent vector. In DE the trial vector is allowed to pass on to the next generation if and only if, its fitness is higher than that of its parent vector p^4. Otherwise, the parent vector proceeds to the next generation. Consequently, DE is a greedy scheme, because one of the parents always competes against its own offspring for appearing in the next generation. Therefore, DE converges very fast.

4. Proposed MOGA-Fuzzy Design Procedure

Past researches in the field of designing fuzzy models show that the simultaneous design of Membership Functions (MFs) and fuzzy rules can enhance the performance of fuzzy systems [13]. Therefore, the MOGA-fuzzy algorithm employed in this work is similar to the Pittsburgh approach. In the algorithm, each chromosome in the population represents a complete knowledge base (KB) except the consequence parts of the rules. In this section, after introducing the employed MFs, chromosome representation and fitness function design are described in detail.

4.1. **Membership Function Design.** Results of past studies approve that membership functions must be flexible enough to develop an accurate fuzzy model [17]. However, flexible membership functions need additional tuning parameters to adjust their shapes. Recent studies confirm that inflexible MF may lead to redundant fuzzy rules [24, 25]. Thus, as described in [24], Gaussian combinational MFs (abbreviated as Gauss2mf) is utilized to depict the antecedent fuzzy sets (i.e. a combination of two Gaussian functions) in this work. A Gauss2mf can be represented by the parameter list $[\sigma_1, c_1, \sigma_2, c_2]$. Where σ_1 and c_1 determine the shape of the leftmost curve. The shape of the rightmost curve is specified by σ_2 and c_2. For the sake of simplicity in GA encoding, every input is considered to be in interval $[-b, b]$, and three parameters C,L and σ are encoded in a chromosome for each MF. Variables $[\sigma_1, c_1, \sigma_2, c_2]$ can be derived as:

$$\begin{aligned} \sigma_1 &= \sigma_2 = \sigma \\ c_1 &= c \\ c_2 &= c + L \end{aligned} \tag{3}$$

Where $c \in [-b-0.5, b+0.5]$,$L \in [0, 3b]$ and $\sigma \in [0, 0.5]$. The advantages of this encoding are described in [13] and the most important ones are as listed below:

1) Confining all genetic searches within feasible regions. This method is more reliable than penalty approaches for handling constraints.

2) if $c_1 \leq -b + 0.3$ and $c_2 \geq b - 0.3$ or $c_1, c_2 \leq -b$,$c_1, c_2 \geq b$, the condition is viewed as a "don't care" condition. Since don't care conditions can be omitted, short rules with a less number of antecedent conditions can be obtained.

These advantages help to secure low-level interpretability. It can lead to high-level interpretability provided that all MFs in antecedent part of a rule to be achieved as don't cares thereby that rule must be omitted from rule base.

4.2. **Fitness Design and Chromosome Representation.** The objectives of designing an efficient fuzzy model to identify a MIMO system are as follows:

1) Minimizing the Mean Square Error (MSE) between the actual and model outputs.

2) Minimizing the number N_r of fuzzy rules.

3) Maximizing the number of don't cares in antecedents of all selected rules.

This can be achieved by minimizing the following measure. $M_d = (1 - \frac{NA_d}{NA_r})$ Where NA_d is the number of don't cares in antecedents of all the selected rules. NA_r is the number of antecedents of all the selected rules.

The second objective results in securing the high-level interpretability and the third one leads to secure interpretability in both low and high levels as mentioned in the previous section. Therefore, only good models in terms of complexity and performance are identified. Validation criteria are utilized as imposed constraints to these objectives which cause to produce valid models. In the case of NARX model validation, the noise model is not specifically estimated and consequently, the residuals may be colored. Specific tests are required and the estimated nonlinear model will be unbiased if and only if [20]:

$$XC_{u^2\varepsilon^2}(\tau) = 0 \ \forall\tau \tag{4}$$

$$XC_{u^2\varepsilon}(\tau) = 0 \ \forall\tau \tag{5}$$

$$XC_{u\varepsilon}(\tau) = 0 \ \forall\tau \tag{6}$$

Where XC denotes the cross correlation, ε is the residual vector (containing all residuals), and u is the input vector (containing all inputs). The target value for these constraints is set up to be at the 95 percent confidence limit. The correlation based validation objectives are contained in a $(2\tau + 1)$ element vector. In order to define these functions as scalar, the infinity norm is utilized as follows:

$$N^{inf}_{u^2\varepsilon^2} = \|XC_{u^2\varepsilon^2}\|_\infty \tag{7}$$

$$N^{inf}_{u^2\varepsilon} = \|XC_{u^2\varepsilon}\|_\infty \tag{8}$$

$$N^{inf}_{u\varepsilon} = \|XC_{u\varepsilon}\|_\infty \tag{9}$$

The identification process then evolves through regions of the search space where valid fuzzy models are located (i.e. the above scalars should be less than or equal to the predefined values).

Similar fuzzy sets represent almost the same region in the universe of discourse of a fuzzy variable; i.e., they describe the same concept. To improve the low-level interpretability of fuzzy models (i.e. distinguishability), several methods have been proposed [26, 10, 27, 4]. Some of them focused on the tradeoff between numerical accuracy and linguistic interpretability [26, 10, 27, 19, 4]. In these methods,

...	c_j^i	L_j^i	σ_j^i	...	$c_j^{N_u}$	$L_j^{N_u}$	$\sigma_j^{N_u}$	r_j	...

Table 1. Chromosome Representation

a formulation of some constraints is imposed in the optimization of the MFs to guarantee semantic integrity [10, 4]. In this work, the peak value of intersection MF of fuzzy MFs on each input is forced to be in $0.1 - 0.7$ range. This constraint results in obtaining dissimilar MFs for each input variable which is concisely called securing distinguishability.

One of the most used fuzzy rule reduction techniques is the rule selection that leads to high level interpretability (i.e. parsimony). Rule selection is the task of finding an optimal subset of rules via a search mechanism. In this study, rule selection is performed by encoding some control genes in GA's chromosome as shown in Table 1. A chromosome consists of control genes for selecting significant fuzzy rules and parametric genes for encoding the antecedent MFs. In this design, rule selection and antecedent MF tuning are simultaneously determined to obtain an accurate and compact fuzzy model.

Where $N_u = n_u(n_q + n_d) + n_y n_p$ and $r_j, j = 1, ..., N_R$ is the control gene represented by one bit for eliminating unnecessary fuzzy rule. If $r_j = 0$, the j^{th} fuzzy rule is excluded from the rule base, otherwise it is included. For evaluating each chromosome N_r rules out of N_R are included. Parameters c_j^i ,L_j^i and $\sigma_j^i (i = 1, ..N_U, j = 1, ...N_R)$ are employed for determining the antecedent fuzzy set for i^{th} input in j^{th} rule.

The fitness function is designed as follows:

Procedure Fitness (Chromosome Ch_i)

• Construct the antecedent parts of a TS-FIS model according to Ch_i. Inputs which have don't care conditions are eliminated. Some of rules are excluded from the rule base.

• For each output Tune the consequent parameters of FIS rules by DE (given the training data).

• If the infinity norms defined in equations (6-8) are less than or equal to predefined values and low-level interpretability constraints are satisfied.

• Return the costs regarding to performance of model (objective No.1), high-level interpretability index (i.e., No. of rules, objective No.2) and low-level interpretability measure (objective No. 3).

End Procedure

For each chromosome DE is utilized to determine the consequent parameters of included rules. Finally, the obtained fuzzy model is evaluated and fitness assignment is done based on three defined objectives. As described in this section, parsimony (high-level interpretability) is achieved by rule selection through control genes of chromosomes and minimizing the number of rules as one of the objectives in MOGA (objective No.2). Low-level interpretability is also secured by maximizing the number of don't care MFs (i.e. minimizing M_d , objective No.3) as well as imposing distinguishability constraints. Therefore, both low and high levels of interpretability, accuracy and validity of fuzzy models are considered as objectives in the proposed optimization procedure. Consequently, all objectives of fuzzy based nonlinear MIMO system identification task are taken into account for optimization simultaneously.

5. Simulation Setup

MOGA setup: The population size and generation numbers are set to $p_s = 300$ and $G = 50$ respectively. The 1-point crossover with the probability of 0.7 is employed. Classical mutation with probability 0.05 is utilized. The tournament selection method by size 2 is used.

DE setup: The population size and generation number are set to $p_s = 30$ and $G = 20$, respectively. Due to the fact that consequent parameters are linear parameters and the estimation of them is not a complicated task, p_s and G are considered relatively small values. A convergence criterion in terms of error is also employed in order to terminate DE immediately (i.e. before reaching the generation No. 20). The DE parameters are defined to be $\eta = 0.8$ and $F = 0.2$. These values of parameters are usually used for most of problems.

Fuzzy models: Two case studies are considered in this paper. In the first one the input and output ranges, i.e. n_p, n_q and n_d parameters are considered the same as the literature [24]. In the second case study, data values are normalized in the range $(-1\ 1)$ and the model is obtained by the proposed procedure. Then the model output is denormalized into its actual scale. Maximum lags considered for the outputs and inputs and minimum discrete dead time are $n_p = n_q = 5$,$n_d = 1$ respectively. The initial number of rules ($N_R = 5$) is considered for the first case study and 6 for the second.

5.1. Simulation Results.

To show the effectiveness of the proposed algorithm two case studies adopted from the literature are described in the following.

Case 1: The non-linear system studied in [24] is taken as the first example:

$$y(k) = g(y(k-1), y(k-2)) + u(k) \tag{10}$$

where

$$g(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)(y(k-1)-0.5)}{1+y^2(k-1)+y^2(k-2)} \tag{11}$$

The system output depends on both its previous values and the current input. The goal is to approximate the non-linear component $g(y(k-1), y(k-2))$ of the system with a fuzzy model. As in [24], 400 simulated data points are generated from the system model (10). Starting from the equilibrium state $(0, 0)$, 200 samples of training data are obtained with a random input signal $u(k)$ uniformly distributed in $(-1.5, 1.5)$, followed by 200 samples of testing data obtained using a sinusoidal input signal $u(k) = sin(2\pi k/25)$. These provided input and output data are depicted in Figure 1.

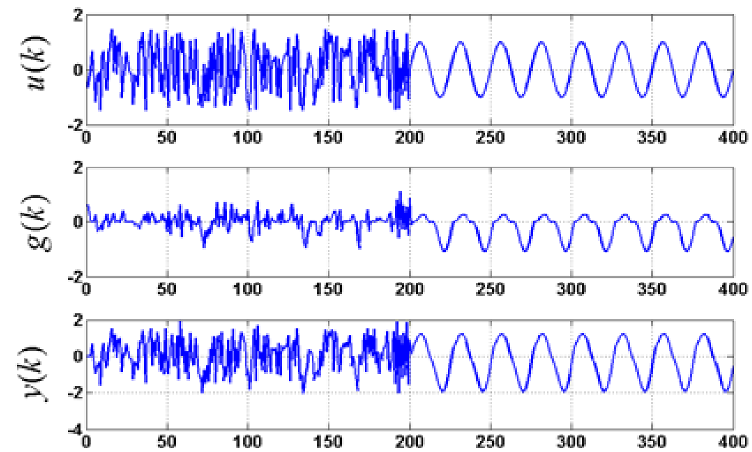

Figure 1. Produced Train and Test Data Samples

If $y(k-1)$ is Middle and $y(k-2)$ is Low then
$g(k) = 0.7967y(k-1) + 0.0456y(k-2) + 0.0534$
If $y(k-2)$ is Middle then $g(k) = 0.0155y(k-2) - 0.0044$
If $y(k-1)$ is Middle and $y(k-2)$ is High then
$g(k) = 0.5890y(k-1) - 0.0016y(k-2) - 0.0144$

Table 2. Obtained TS-FIS Model

The obtained TS-FIS model with three rules is given in Table 2 . The related input MFs distributions for FIS are shown in Figure 2.
The actual (solid line) and model output (dotted line) for the trained and test data are depicted in Figures 3 and 4.

The performance of the proposed fuzzy model is compared with those in [24]. The comparative results are listed in Table 2. It can be seen that compared with the fuzzy models in the literatures, the proposed model keeps good balance between numerical accuracy and model simplicity (i.e. compactness and transparency).

Case 2: **Continuous Stirring Tank Reactor.** The process is a Continuous Stirring Tank Reactor (CSTR), where the reaction is exothermic and the concentration is controlled by regulating the coolant flow. Schematic diagram of Continuous stirring tank reactor is shown in Figure 5.

In Figure 5, F_0, C_{A0} and T_0 are feed flow rate of mass, inlet feed concentration and inlet feed temperature respectively and F_j, T_{j0} are inlet coolant mass flow and temperature, V, T, C_A, F and T_j are reactor volume, temperature, concentration, output feed flow rate and output coolant temperature. In this case, the input is coolant flow to jacket and the outputs are product concentration and temperature. The numbers of data randomly taken from [3] as public domain bench mark is 1000 and the sample time is 6 seconds. Seventy percent of the data pairs are selected

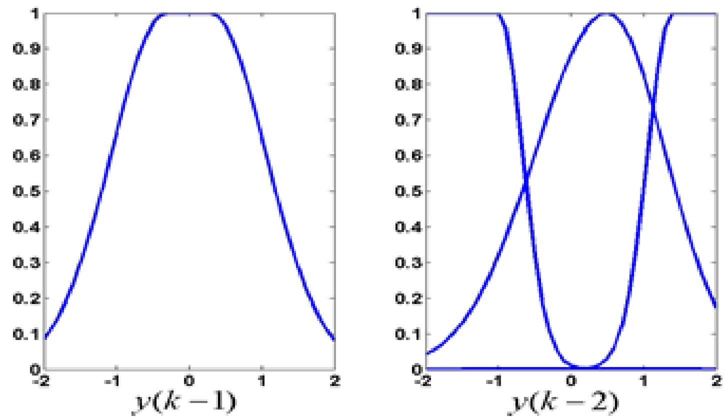

FIGURE 2. Inputs MFs Distribution

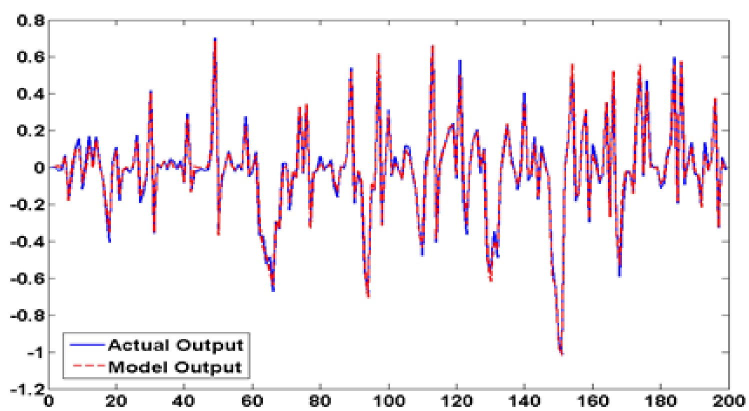

FIGURE 3. Actual and Model Outputs for Training Data

Ref.	No. of rules	No. of MFs	MSE train	MSE test
[24]	5 rules (initial)	10 Gauss2mf.	1.40×10^{-3}	2.63×10^{-3}
	5 rules (optimized)	3 Gauss2mf.	2.38×10^{-4}	3.01×10^{-4}
	4 rules (optimized)	3 Gauss2mf.	5.46×10^{-4}	5.44×10^{-4}
	4 rules (optimized)	3 Gauss2mf.	5.61×10^{-4}	2.49×10^{-4}
This paper	5 rules (initial)	MOGA initialization		
	3 rules (optimized)	4 Gauss2mf.	4.57×10^{-4}	1.56×10^{-4}

TABLE 3. Comparing the Best Recently Reported Fuzzy Model and That of This Paper. In All Cases the Consequent Parts are Linear

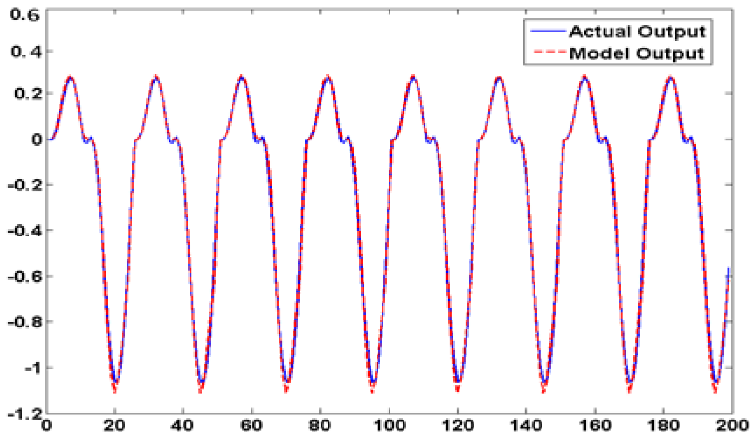

Figure 4. Actual and Model Outputs for Test Data

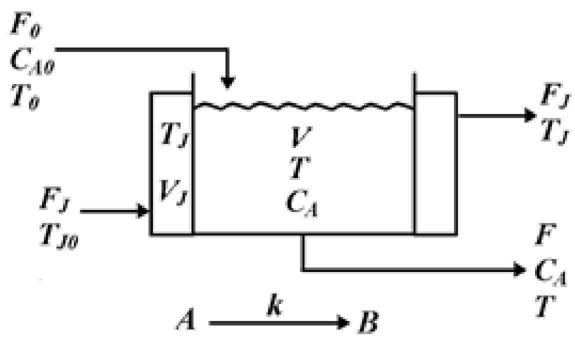

Figure 5. Continuous Stirring Tank Reactor

randomly for training and the rest are used for the test. For the first 200 data of training and test samples, actual and model outputs are depicted in Figures 6a, 6b and Figures 7a, 7b respectively.

Table 4 shows the only extracted rule for TS-FIS model. As can be seen in Table 4, some of the inputs have been omitted (i.e. 9 inputs are selected out of 15) and number of rules has been optimized (i.e. one rule). When the procedure is terminated, the Pareto-optimal set of solutions is given. Designer can select various solutions in terms of the defined objectives values. Therefore, different solutions in terms of accuracy, interpretability, compactness and correctness are available.

Figures 9,10 depict the Pareto-fronts of two objectives namely MSE and objective related to the number of don't care for the first and second outputs respectively. These Pareto-fronts are in feasible space and are illustrated by dash lines. The coordinates of points located on Pareto-fronts are objective values of non-dominated solutions.

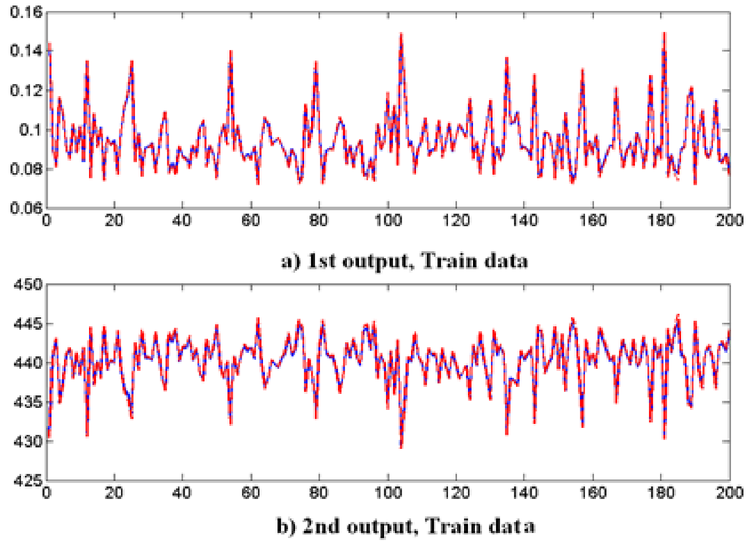

Figure 6. Actual and Model Output for the First and Second Outputs on Train Data

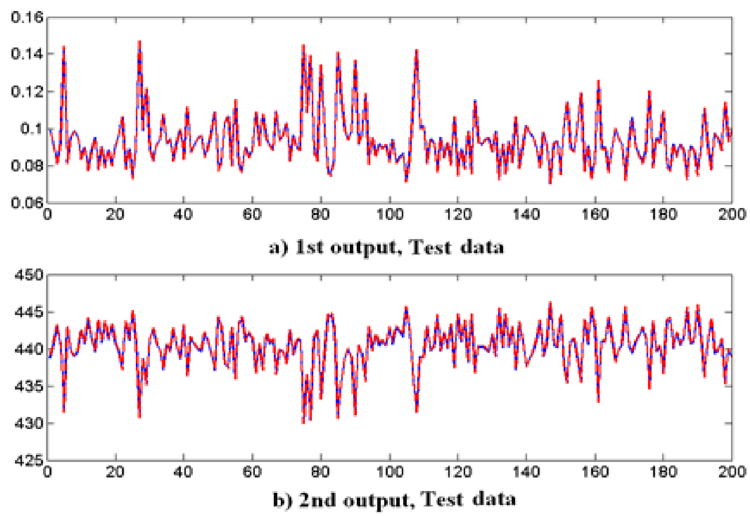

Figure 7. Actual and Model Output for the First and Second Outputs on Test Data

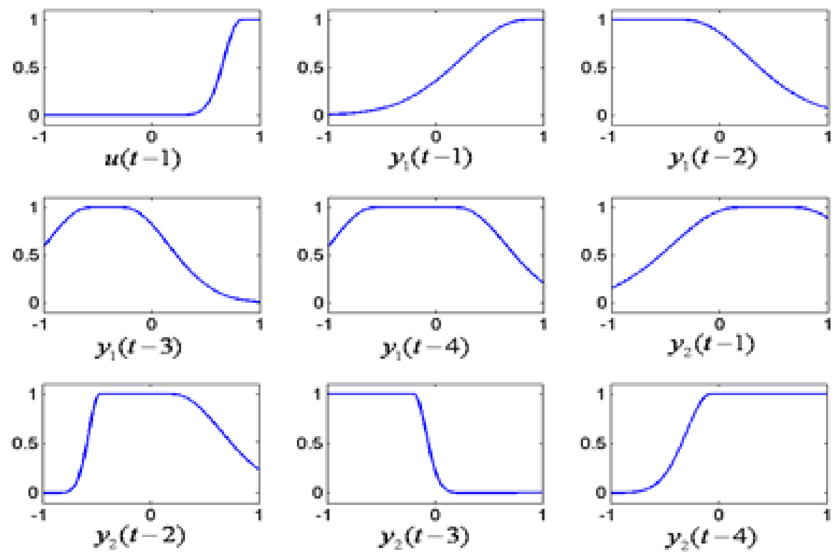

FIGURE 8. Plot of Antecedent MFs Distributions

If $u(t-1)$ is High and $y_1(t-1)$ is High and $y_1(t-2)$ is Low and $y_1(t-3)$ is Middle and $y_1(t-4)$ is Middle and $y_2(t-1)$ is Middle and $y_2(t-2)$ is Middle and $y_2(t-3)$ is Low and $y_2(t-4)$ is High then
$y_1(t) = 0.0429\ u(t-1) + 1.3649\ y_1(t-1) - 0.6939 y_1(t-2) -0.0395\ y_1(t-3) + 0.5999\ y_1(t-4) - 0.4421\ y_2(t-1) +0.5235\ y_2(t-2) + 0.0653\ y_2(t-3) - 0.4882\ y_2(t-4) - 0.0014$
and
$y_2(t) = -0.1160\ u(t-1) - 0.5031\ y_1(t-1) + 1.1602\ y_1(t-2) +0.4123\ y_1(t-3) + 0.19\ y_1(t-4) + 0.4770\ y_2(t-1) -0.5176\ y_2(t-2) - 0.3727\ y_2(t-3) + 0.1312\ y_2(t-4) + 0.0019$

TABLE 4. The Extracted Rule

Many dominated solutions in terms of their objectives values are indicated by stars in both figures. An expert can select one solution from Pareto-optimal set of solutions where the fitness values are desirable. The selected points of Pareto-fronts in this paper are circled in Figures 9,10. These points are to extent in the middle of Pareto-fronts and yield a good tradeoff between two objectives values. The fuzzy model which its fitness values are indicated by squares on Pareto-fronts of Figures 9,10 is compared with those in [5, 6].

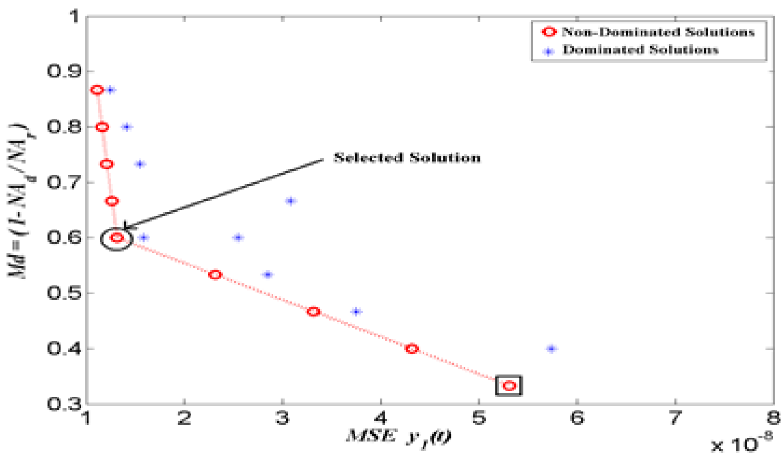

Figure 9. Feasible Pareto-front for Two Objectives, x-axis: is the MSE of the First Output, y-axis: Objective Value Related to Don't Cares, when $N_R = 1$. Selected Solution has been Highlighted

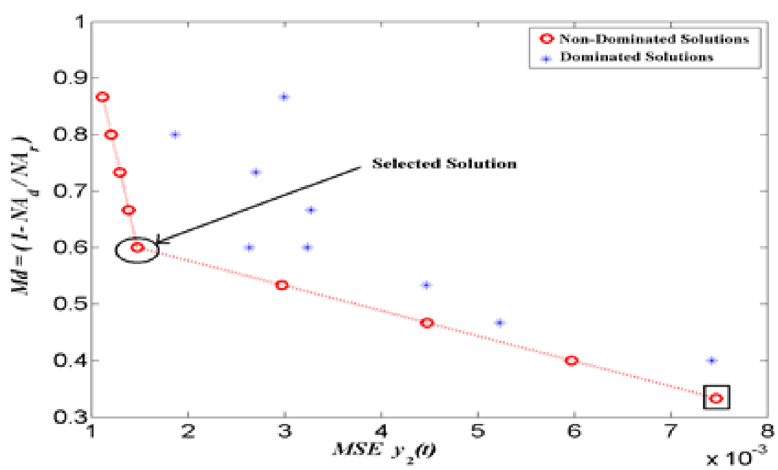

Figure 10. Feasible Pareto-front for Two Objectives, x-axis: is the MSE of the Second Output, y-axis: Objective Value Related to Don't Cares, when $N_R = 1$. Selected Solution has been Highlighted

The comparative results are listed in Table 5. It can be seen that compared with the fuzzy models in the literatures, the proposed model keeps good balance between numerical accuracy and model simplicity (i.e. compactness and transparency).

For completeness, in the following we are going to discuss the robust optimization problem. Robust design is an engineering methodology for optimization in

Ref.	No. of rules	No. of MFs	MSE train	MSE test
[5]	6 rules (initial)	36 Gauss mf.		
	1 rule (optimized)	6 Gauss mf.	1.07×10^{-5} (1st)	6.5×10^{-8} (1st)
			0.30 (2nd)	3.3×10^{-3} (2nd)
[6]	1 rule (optimized)	5 Gauss mf.	6.62×10^{-8} (1st)	9.02×10^{-8} (1st)
			2.0×10^{-3} (2nd)	1.5×10^{-3} (2nd)
This paper	6 rules (initial)	MOGA initialization		
	1 rule (optimized)	5 Gauss2mf.	5.21×10^{-8} (1st)	4.99×10^{-8} (1st)
			7.51×10^{-3} (2nd)	1.56×10^{-3} (2nd)

Table 5. Comparing the Best Resulted Fuzzy Model of This Paper with the Best Recently Reported Ones. 1st and 2nd Indicate the First and Second Outputs Respectively

which the sensitivity to the various causes of variations is minimal [12, 23]. Evolutionary algorithms have some parameters and the optimization performance depends greatly on changing their values [12, 23]. Therefore, in recent researches the Taguchi's parameter design method has been employed in EAs to achieve robustness in optimization [12, 23]. Our proposed algorithm can achieve the population distribution closest to the target namely the robustness of optimization. Thirty runs of the proposed procedure are carried out for each problem. Accordingly, there are two main reasons for assessing the robustness ability. Firstly, the proposed procedure utilizes the abilities of two algorithms: MOGA and DE. Secondly, the standard deviations of objective functions values are relatively small on different regions of Pareto-front over 30 runs of the proposed procedure. MOGA performs the task of MFs tuning as well as rule base structure identification simultaneously. The resulting Pareto-front of objectives confirms the diversity of the obtained solutions over the objective space. This diversity is a result of using fitness sharing in our MOGA. Besides, DE is utilized only for linear parameter identification (i.e. the consequence parameters of the first order TSK model). Duo to this division of optimization tasks among MOGA and DE, the overall optimization task is performed in a robust manner.

6. **Concluding Remarks**

In this paper, a novel approach is introduced for construction of TS-fuzzy models which take the accuracy, compactness, transparency and correctness into account for nonlinear system identification. In order to consider different non-commensurate objectives, the use of MOGA for fuzzy non-linear system identification provides an identification technique by simultaneous evaluation of the model structure, the performance, transparency and validation characteristics of the candidate models. EAs are powerful in finding solutions in a complicated search space, yet they are computationally expensive. The most difficult aspect in fuzzy modeling involves identification and finding the effective fuzzy rules and determining the antecedent

MFs parameters simultaneously. For the sake of computational efficiency, DE was utilized to determine the consequent parameters instead of encoding them in the MOGA chromosome. Small values for number of generations and population size were utilized in DE while a convergence criterion was employed to terminate the algorithm immediately (i.e. before reaching the last generation). The proposed approach has been successfully applied to two benchmark problems taken from the literature: 1) a nonlinear plant with two inputs and one output and 2) a CSTR process. Simulation results demonstrate the capability of the proposed method in finding accurate, compact and transparent models. The obtained models are well-interpretable, i.e. three rules with distinct MFs on each input for the first case study and one rule for the second one. The resulting Pareto-set solution from MOGA provides user with the opportunity of selecting various solutions. Also, the user can interact with the algorithm to define softer constraints and goals when the algorithm is being run. Thirty runs of the proposed procedure for each problem were performed. In different regions of obtained Pareto-fronts, the resulted relatively small standard deviations of fitness functions values show that the method is robust to estimate the fuzzy model parameters.

References

[1] J. Abonyi, *Fuzzy Model Identification for Control*, Birkhauser, Boston, 2003.

[2] K. Debs, *Multio-bjective optimization Using Evolutionary Algorithms*, John Wiley and Son Ltd, 2001.

[3] B. L. R. De Moor and ed., *DaISy: Database for the Identification of Systems*, Department of Electrical Engineering, ESAT/SISTA, K. U. Leuven, Belgium, URL: http://www.esat.kuleuven.ac.be/sista/daisy/.

[4] V. De Oliveira, *Semantic constraints for membership function optimization*, IEEE Trans. SMC-A, **29(1)** (1999), 128-138.

[5] M. Eftekhari, S. D. Katebi, M. Karimi and A. H. Jahanmiri, *Eliciting transparent fuzzy model using differential evolution*, Applied Soft Computing, **8** (2008), 466-476.

[6] M. Eftekhari and S. D. Katebi, *Extracting compact fuzzy rules for nonlinear system modeling using subtractive clustering, GA and unscented filter*, Applied Mathematical Modelling, **32** (2008), 2634-2651.

[7] B. Feil, J. Abonyi, J. Madar, S. Nemeth and P. A rva, *Identification and analysis of MIMO systems based on clustering algorithm*, Acta Agraria Kaposvariensis, **8(3)** (2004), 191-203.

[8] C. M. Fonseca and P. J. Fleming, *Multi-objective optimization and multiple constraint handling with evolutionary algorithms-part I: application example*, IEEE Trans. Syst. Man and Cybernetics, **28(1)** (1998), 26-37.

[9] C. M. Fonseca and P. J. Fleming, *Multi-objective optimization and multiple constraint handling with evolutionary algorithms-part II: a unified formulation*, IEEE Trans. Syst. Man and Cybernetics, **28(1)** (1998), 38-47.

[10] M. J. Gacto, R. Alcala and F. Herrera, *Integration of an Index to Preserve the Semantic Interpretability in the Multiobjective Evolutionary Rule Selection and Tuning of Linguistic Fuzzy Systems*, IEEE Transactions on Fuzzy Systems, **8(3)** (2010), 515-531.

[11] S. Y. Ho, H. M. Chen, S. J. Ho and T. K. Cehn, *Design of accurate classifiers with a compact fuzzy rule base using an evolutionary scatter partition of feature space*, IEEE Tans. Systems, Man and Cybernetics, Part B: Cybernetics, **34(2)** (2004), 1031-1044.

[12] W. H. Ho, J. H. Chou and C. Y. Guo, *Parameter identification of chaotic systems using improved differential evolution algorithm*, Nonlinear Dynamics, **61** (2010), 29-41.

[13] A. Homaifar and E. McCormick, *Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms*, IEEE Trans. Fuzzy Syst., **3** (1995), 129-139.

[14] H. Ishibuchi, *Multiobjective genetic fuzzy systems: Review and future research directions,* Proc. of IEEE InternationalConference on Fuzzy Systems, London, UK, July **23-26**, (2007) 913-918.

[15] C. Z. Janikow, *A knowledge intensive genetic algorithm for supervised learning*, Machine Learning, **13** (1993), 198-228.

[16] L. Ljung, *System identification toolbox: user's guide*, The MathWorks, 2004.

[17] S. Medasani, J. Kim and R. Krishnapuram, *An overview of membership function generation techniques for pattern recognition*, Int. J. Approx. Reasoning, **19(3-4)** (1998), 391-417.

[18] O. Nelles, *Nonlinear System Identification*, Springer-Verlag, Berlin Heidelberg, 2001.

[19] A. Riid and E. Rustern, *Interpretability improvement of fuzzy systems: reducing the number of unique singletons in zeroth order Takagi-Sugeno systems*, Proceedings of (2010) IEEE International Conference on Fuzzy Systems, Barcelona, Spain, (2010), 2013-2018.

[20] K. Rodriguez-vazquez, *Multiobjective evolutionary algorithms in non-linear system identification. PhD thesis,* Department of Automatic Control and Systems Engineering, The University of Sheffield, 1999.

[21] R. Storn and K. Price, *Differential evolution-a simple and efficient adaptive scheme or global optimization over continuous spaces, technical report TR-95-012*, International Computer Science Institute, Berkley, 1995.

[22] R. Storn and K. Price, *Differential evolution a simple and efficient heuristic for global optimization over continuous spaces*, J. Global Optim. **11** (1997), 341-359.

[23] J. T. Tsai, J. H. Chou and T. K. Liu, *Tuning the structure and parameters of a neural network by using hybrid Taguchi-genetic algorithm*, IEEE Trans. on Neural Networks, **17** (2006), 69-80.

[24] 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 Systems, **149** (2005), 149-186.

[25] H. Wang, S. Kwong, Y. Jin, W. Wei and K. F. Man, *Agent-based evolutionary approach for interpretable rule-based knowledge extraction*, IEEE Trans. on Systems, Man and Cybernetics-Part C, **35** (2005), 143-155.

[26] S. M. Zhou and J. Q. Gan, *Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling*, Fuzzy Sets and Systems, **159** (2008), 3091-3131.

[27] S. M. Zhou and J. Q. Gan, *Extracting Takagi-Sugeno fuzzy rules with interpretable submodels via regularization of linguistic modifiers*, IEEE Transactions on Knowledge and Data Engineering, **21(8)** (2009), 1191-1204.

Mojtaba Eftekhari, Faculty of Islamic Azad University, Sirjan branch, ,Sirjan, Kerman, Iran
E-mail address: m.eftekhari59@gmail.com

Mahdi Eftekhari*, Department of Computer Engineering, School of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
E-mail address: m.eftekhari@uk.ac.ir

Maryam Majidi, Department of Computer Engineering, School of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
E-mail address: majena67@yahoo.com

Hossein Nezamabadi pour, Department of Electrical Engineering, School of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
E-mail address: nezam_h@yahoo.com

*Corresponding author