IMPROVING TRANSPARENCY IN APPROXIMATE FUZZY MODELING USING MULTI-OBJECTIVE IMMUNE-INSPIRED OPTIMISATION[*]

JUN CHEN[&]

School of Engineering, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, United Kingdom.

MAHDI MAHFOUF

Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD, United Kingdom
E-mail: m.mahfouf@shef.ac.uk
www.shef.ac.uk/acse/staff/mm

Abstract

In this paper, an immune inspired multi-objective fuzzy modeling (IMOFM) mechanism is proposed specifically for high-dimensional regression problems. For such problems, prediction accuracy is often the paramount requirement. With such a requirement in mind, however, one should also put considerable efforts in eliciting models which are as transparent as possible, a 'tricky' exercise in itself. The proposed mechanism adopts a multi-stage modeling procedure and a variable length coding scheme to account for the enlarged search space due to simultaneous optimisation of the rule-base structure and its associated parameters. We claim here that IMOFM can account for both Singleton and Mamdani Fuzzy Rule-Based Systems (FRBS) due to the carefully chosen output membership functions, the inference scheme and the defuzzification method. The proposed modeling approach has been compared to other representatives using a benchmark problem, and was further applied to a high-dimensional problem, taken from the steel industry, which concerns the prediction of mechanical properties of hot rolled steels. Results confirm that IMOFM is capable of eliciting not only accurate but also transparent FRBSs from quantitative data.

Keywords: Interpretability, Immune-inspired multi-objective optimisation, Variable length coding scheme.

1. Introduction

Traditionally, modeling tasks involve the building of mathematical equations which can best describe the underlying process. Such a modeling practice normally requires a deep understanding of the systems under investigation, hence the reason why it is often referred to as *knowledge-Driven Modeling*. On the contrary, *Data-Driven Modeling* (DDM), inspired principally from artificial intelligence techniques, is based on limited knowledge of the modeling process and relies on data describing the input and output mapping.

DDM is able to make abstraction and generalizations of the process and plays often a complementary role to knowledge-based models. For complex systems, the linear regression may not be sufficient, which leads to the need of non-linear regression techniques. Among many of these techniques, Artificial Neural Networks (ANN), fuzzy rule-based systems (FRBS) and Neural-Fuzzy Systems (NFS) have been receiving more attention during the last two decades due to the fact of not only being able to approximate practically any given function to an arbitrary accuracy, [1] but also being able to generalise reasonably well to any previously 'unseen'

[&] Corresponding Author-
E-mail: juchen@lincoln.ac.uk; Phone: +44 (0) 1522 668901; webpage: http://www.lincoln.ac.uk/engineering/staff/jun_chen.htm.

situations. The prevalence of these nonlinear regression techniques is largely attributed to the breakthrough in the nonlinear optimisation techniques, such as the Back-Error-Propagation algorithm[2] and the bio-inspired optimisation.[3, 4]

Since the first introduction of 'fuzzy logic', FRBS have been widely used in control engineering.[2] However, the predominant approach in the traditional design of fuzzy rule-based systems highly relies on human experts, which makes the fuzzy modeling process similar to the design of expert systems except that traditional expert systems were based on the classical Boolean logic and thus were not well suited to managing the progressiveness in the underlying process phenomena.[5] More recently, FRBS has been combined with some learning components to automatically extract knowledge from resources such as data rather than from experts.[2,6] Such learning components, on the one hand, lead to a coarse FRBS, and on the other hand, can further improve the accuracy of such a coarse model to a certain degree depending on the quality of the historic data and the power of the learning mechanism. Despite of great success, eliciting FRBS through 'learning' may suffer from two serious problems, e.g. the degradation of the model in terms of interpretability (transparency) and the over-fitting to the training patterns. Taking this into account, one can find that bio-inspired optimisation, in particular Genetic Algorithms (GAs), has a long history of being incorporated into fuzzy logic[7] and demonstrate a possible route to the remedy for the previously mentioned problems.

The main aim of this paper is to present a systematic immune-inspired multi-objective fuzzy modeling approach which can simultaneously account for the interpretability of the rule-base and its predictive accuracy for regression problems. The paper is organised as follows: Section 2 discusses the formation of the multi-objective fuzzy modeling problems and the FRBSs used in this work; Section 3 shortly reviews the existing evolutionary based approaches for improving FRBS's interpretability; Section 4 introduces the proposed Immune inspired Multi-Objective Fuzzy Modeling (IMOFM) mechanism[8,9] based on the Population Adaptive based Immune Algorithm (PAIA),[10, 11] in Section 5, some implementation issues of IMOFM are discussed; experimental studies of IMOFM on a benchmark problem and on a real-world

problem taken from steel industries are given in Section 6; finally, conclusions are given in Section 7.

2. FRBS and the Formation of Multi-Objective Fuzzy Modeling Problems

2.1. *Fuzzy inference systems and FRBS*

Fuzzy inference is a process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decision can be made, or patterns discerned. The two most popular types of fuzzy inference systems are the Mamdani-type[12] and Sugeno-type[13], which vary somewhat in the way outputs are determined. The consequence part of the Mamdani-type is a fuzzy set while the consequence part of the Sugeno-type is a set of functions with the arguments that are the linguistic variables of the antecedent part. Hence, a FRBS can be formulated as follows:

$$R_i: If\ x_1\ is\ A_i^1\ and\ x_2\ is\ A_i^2, \dots, and\ x_j\ is\ A_i^j\ Then\ y_i = Z_i$$

where, A_i^j is the ith linguistic value (fuzzy set) for the jth linguistic variable x_j defined over the universe of discourse $\mho_j$; the function $\mu_{A_i^j}(x_j)$ associated with A_i^j that maps $\mho_j$ to $[0, 1]$ is the corresponding membership function; R_i represents the ith rule in the rule base, and y_i is the output of the ith rule. Typically, Z_i can be the function of the inputs or the linguistic value of the output, which differentiate FRBS into Sugeno-type (the former) and Mamdani-type (the latter).

In this paper, a special case of Sugeno-type, namely Singleton FRBS when Z_i is the zero order function of the inputs, and Mamdani-type are employed due to their abilities of expressing linguistic meanings in both of their antecedents and consequents.

2.2. *Accuracy vs. Interpretability: formation of multi-objective fuzzy modeling problems*

As Casillas *et al.* pointed out in Ref. 14, modeling is the task that simplifies a real system or complex reality with the aim of easing its understanding. Hence, the development of *reliable* and *comprehensible* models must be the main theme of any modeling tasks. By 'reliable' it is meant the model's capability of faithfully representing the real systems, in other words 'the model accuracy'. By 'comprehensible' it is meant the model's capability of expressing the behavior of the real systems

in a transparent way, in other words 'model interpretability'. However, as Zadeh conjectured in his Principle of Incompatibility,[15] it is very likely that accuracy and interpretability may well be exclusive requirements in a modeling process. Since both requirements are vital and cannot always be possessed at the same time, a good balance between them is the best outcome that one can achieve. The reflection of these in a fuzzy modeling scenario represents a dilemma of designing FRBS. As far as interpretability is concerned, attempts has indeed been made by Valente de Oliveira[16] to tackle this problem via nonlinear constraints coded within a given optimisation scheme. However, his approach was deemed to include a 'low-level interpretability' as Zhou *et al*. discussed in their comprehensive survey.[17] In Ref. 18, the authors conducted an experimental analysis of the associated algorithm and concluded that a numerical index alone is not enough for it to be widely accepted. Hence, interpretability is mainly a subjective property and normally refers to at least one or all of the following aspects in a fuzzy modeling scenario:

(i) The distribution of the fuzzy sets across each dimension should be well separated so that meaningful (distinguishable) linguistic terms can be associated with them,

(ii) The number of fuzzy sets for each dimension and the number of rules should not be excessive. This is closely related to the cognitive studies[19] which shows that the optimal number of chunks of information simultaneously held in human short-term memory should be seven, plus or minus two. This implies that redundant rules and fuzzy sets should be merged or deleted,

(iii) The number of input variables involved in each rule should be optimal, which means input variables are subject to either a global selection, in which case none of the rules in the rule base can use the deleted input variables, or a local selection, in which case the selection is done at the individual rule level.[5]

(iv) The rule base should be complete and consistent. Otherwise, the knowledge represented by the rule base is incomplete, and different conclusions given similar premises would certainly confuse its users.

Hence, 'accuracy vs. interpretability' issue in a fuzzy modeling context is a multi-objective optimization problem as shown in Fig. 1, where two competing objectives, viz. the predictive error (accuracy) and the

rule base complexity (transparency), are minimized simultaneously. The aim is to find a set of 'approximate Pareto FRBSs' as close to the true Pareto front as possible. By finding a set of solutions, human can understand the underlying problem in a much greater depth, and finally a single optimal solution to a specific scenario is selected and applied. In the case shown in Fig. 1, if one requires certain transparency of the FRBS along with its good predictive accuracy the middle circle could be the one that fulfils the user's need.
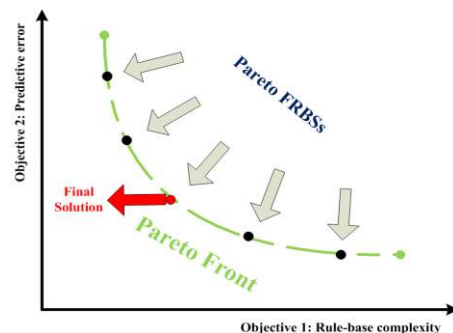

Fig. 1. Pareto front in a bi-objective fuzzy modeling case.

3. Literature Review of Previous Works

Originated from Karr's work,[7] the GA approach in fuzzy systems was initially utilized to adjust the parameters of membership functions, which leads to no significant difference when compared to other learning paradigms. The real significance of employing evolutionary algorithms (EAs) for optimizing FRBSs comes from EAs' flexibility in terms of being able to encode and evolve almost every component of the FRBS.[20] Such flexibility offers a solution so that one can take into account interpretability (structure) and predictive performance of the FRBS in a more coherent way. Broadly speaking, there currently exist two EA-based streams to tackle interpretability issues:

(i) The first stream is mainly concerned with the linguistic modeling using a Mamdani-type, in which a set of pre-specified fuzzy partitions (linguistic terms) are given *a priori* by experts or users (grid partition); the task is then to find an optimal FRBS in terms of its compactness and performance. These linguistic terms are fixed during the course of the evolution[21, 22, 23] so that their physical meanings are retained. Only fuzzy rules are subject to selection via GAs so that a compact rule-base can be evolved from a large

number of candidate rules, which should lead to a more interpretable FRBS. Since the selection process removes irrelevant and inconsistent rules, the accuracy is also improved. In Ref. 23, the GA is not only used to select the optimal combination of rules but also to learn the granularity of different fuzzy partitions for each input, which leads to a more accurate fuzzy model while the linguistic feature is not compromised. Similar effort has been reported in Ref. 24, where multi-objective evolutionary algorithms were employed to concurrently learn the partition granularities of each variable and rule bases of Mamdani fuzzy systems for regression problems. Further relevant researches include those in Ref. 25-26, apart from the rule selection, these works also tuned the linguistic terms by a modified GA. However, such tuning is only operated in a restrained space in order to maintain their original semantics.

(ii) The second stream generally uses an approximate fuzzy model as the starting point (in such a case, fuzzy partitions are extracted via some automatic learning components); the task is then to improve the model's explanatory ability, which may have been lost during the automatic learning process, through a set of similarity-driven simplification and parameter adjusting operations.[27, 28, 29, 30] Under this stream, a similarity measure is taken so that similar fuzzy sets can be merged. Consequently, similar rules are merged as well. Hence, the distinguishability of membership functions and the compactness of the rule-base are improved. In Ref. 31, the idea of rule pruning is used to remove less relevant rules within a multi-objective optimization framework. The similarity measure is not explicitly used in this work.

Comparing the two streams leads to the following: in the linguistic modeling stream, the target problems are normally associated with classifications and low-dimensional function approximation; hence, the effect of the 'curse of dimensionality' due to the grid partition and the need for the parameter tuning due to the performance requirement are not serious issues; only very recently, such a linguistic modeling framework has been adopted for high-dimensional regression problems[32]; for the second stream, high-dimensional function approximation are often the case, as a result, an approximate FRBS is a better choice to start with due to the compactness requirement and the progressive nature of the regression problems. However, to the best

of our knowledge, majority of the works within the second modeling stream were using Sugeno-type which breaches the original intention of the FRBS. It is rather 'tricky' to decide which modeling stream is more suitable. Both modeling streams have their limitations: 1) although linguistic modeling often leads to well distributed membership functions more rules are required to achieve similar predictive performances as those provided by the second modeling stream with fewer rules, this being due to the restriction imposed on the membership function search space; 2) although the second modeling stream often leads to a compact rule-base and higher predictive accuracy, the membership functions are not well distributed even after interpretability improvement; furthermore, if Sugeno-type is employed the transparency in the consequents will be lost.

In the light of the above considerations, the proposed IMOFM sits in the middle of the two modeling streams by using a compact FRBS with certain interpretability for high-dimensional regression problems. Although a Singleton/Mamdani FRBS is used in this work, unlike those in the first modeling stream, the membership functions of the proposed method can move freely within the variable intervals. Hence, it is still within the second modeling stream. However, it greatly improves the interpretability of the elicited FRBS, and can be viewed as a complement to Ref. 32 due to the fact that more compact and higher accurate FRBSs can be elicited.

4. An Immune Inspired Multi-Objective Fuzzy Modeling (IMOFM) Mechanism

IMOFM is a three-stage modeling procedure. The aim of the first two stages is to first extract an initial approximate FRBS and then to refine it in terms of its predictive accuracy. By doing so, an initial FRBS with the over-estimated number of rules can efficiently be elicited. Another reason of including the first two modeling stages, especially the second one (refinement), is that by doing so the most complex rule-base can survive under the pressure of 'Pareto' selection. Without including the refining step, the rule-base with a complex structure may be regarded inferior to the less complex rule-base in a 'Pareto' sense even if both are inaccurate in the early evolutionary stages. Hence, one may lose the chance of evolving the most accurate FRBS, which normally comes with a complex

structure. The refined initial FRBS is then used in the third stage to seed the initial population of PAIA in order to obtain a set of Pareto fuzzy models with improved interpretability. To tackle the problem of simultaneously optimizing the rule-base structure and parameters, a variable length coding scheme is adopted, and a new distance index is proposed to cope with the variable-length individuals, which should improve the efficiency of the search. For model structure optimization, a *Model Simplification* module is added in a bid to find transparent FRBSs, which will be further discussed in Section 4.3. Fig. 2 represents a schematic diagram of such a modeling framework.
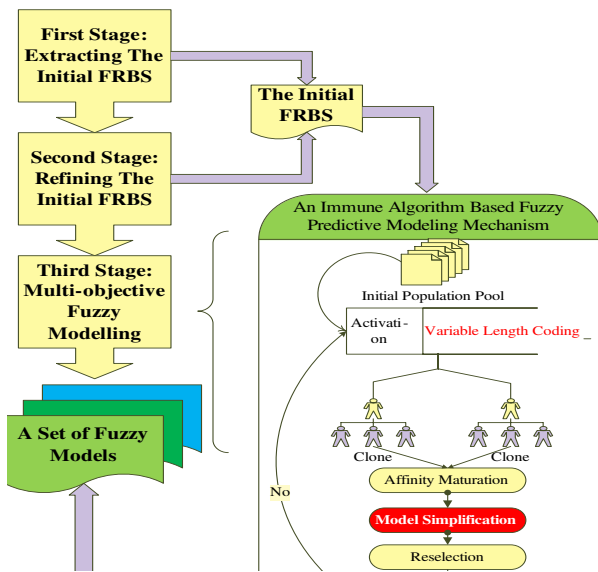

Fig. 2. The proposed IMOFM framework.

4.1. *First Stage: Elictation of Initial FRBSs*

First, an evolutionary based K-means clustering algorithm[11] is used to group the available data into a predefined number of clusters. In order to convert the obtained clusters into FRBSs, a certain mechanism has to be established so that $\mu_{A_i^j}(x_j)$ and the corresponding output Z_i can be linked with the extracted clusters. Gaussian membership functions are used for the inputs of FRBSs. In such a case, the ith identified cluster centre C_i^X in the input space corresponds directly to the centroids of the Gaussian membership functions responsible for the ith rule. The spreads of the

corresponding Gaussian membership functions are obtained by first calculating the U matrix as follows:

$$U(i,m) = \left(\sum_{l=1}^{k} \frac{\|X_m - C_i^X\|}{\|X_m - C_l^X\|} \right)^{-1} \qquad (1)$$

where, $C_1^X, C_2^X, \ldots, C_k^X$ are k cluster centers in the input space, $\|\ \|$ is the Euclidean distance, and $U(i,m)$ specifies the degree of data point m belonging to the ith cluster. Spread σ_i^j is thus deduced as follows:

$$exp\left(-\frac{1}{2} \cdot \left(\frac{x_m^j - c_i^j}{\sigma_{im}^j} \right)^2 \right) = U(i,m)$$

$$\Rightarrow \sigma_{im}^j = \sqrt{\frac{-\left(x_m^j - c_i^j\right)^2}{2 \cdot \log\left(U(i,m)\right)}} \qquad m = 1,..,N \quad (2)$$

$$\Rightarrow \sigma_i^j = \rho \cdot \max_{m \in [1,N]} \left(\sigma_{im}^j \right)$$

where, j indicates the dimension of the spread in the input space for the ith cluster, N is the total number of data points. The maximum value of σ_{im}^j is picked to ensure a certain degree of overlap between different clusters. This also ensures a smooth transition of the predictions over different regions. ρ is used to adjust the degree of overlap, and is set to 0.95 in this work without any loss of generality. Hence, the Gaussian membership function on each dimension can be specified using Eq. 3:

$$\mu_{A_i^j}\left(x_m^j\right) = \exp\left(-\frac{1}{2} \cdot \left(\frac{x_m^j - c_i^j}{\sigma_i^j} \right)^2 \right) \qquad (3)$$

For Singleton FRBS (IMOFM_S), Z_i is equal to C_i^y. If Centroid of Area (COA) defuzzification method is employed, the crisp output of the initial Singleton FRBS can be computed as below:

$$y^{crisp} = \frac{\sum_{i=1}^{k} Z_i \cdot \mu_i(X)}{\sum_{i=1}^{k} \mu_i(X)} \stackrel{\text{def}}{=} y^{crisp}(X|\theta) \qquad (4)$$

where, $\mu_i(X_m) = \mu_{A_i^1}(x_m^1) \cdot \mu_{A_i^2}(x_m^2) \cdot \ldots \cdot \mu_{A_i^n}(x_m^n) = \prod_{j=1}^{n} \exp\left(-\frac{1}{2} \cdot \left(\frac{x_m^j - c_i^j}{\sigma_i^j} \right)^2 \right)$, and $\theta = \left(c_i^y, c_i^j, \sigma_i^j | i = 1,..,k; j = 1,..,n \right)$ is the parameter vector which is subject to further tuning in the second modeling stage.

For Mamdani FRBS (IMOFM_M), the bell-shape membership functions are used for Z_i:

$$\mu_{B_i}(y_m) = \frac{1}{1 + \left(\frac{y - c_i^y}{\sigma_i^y}\right)^2} \qquad (5)$$

where, σ_i^y is obtained by using Eq. 1 and 2 but in the output space. Unlike traditional Mamdani FRBS where defuzzification is normally applied on the overall implied fuzzy set[2], IMOFM employs the center of gravity (COG) defuzzificaiton on the implied fuzzy set as below:

$$\mu_{\hat{B}_i}(y_m) = \mu_i(X_m) * \mu_{B_i}(y_m) \qquad (6)$$

Instead of using minimum and maximum for the T-norm and S-norm, IMOFM chooses 'product' and 'plus' respectively. All these modifications are to ensure computational efficiency comparing to the tranditional Mamdani implementation, and more importantly, to ensure that an analytical solution described in Eq. 7 can be deducted.

$$y^{crisp} = \frac{\sum_{i=1}^{k} c_i^y \cdot \int_y \mu_{\hat{B}_i}(y)\, dy}{\sum_{i=1}^{k} \int_y \mu_{\hat{B}_i}(y)\, dy} = \frac{\sum_{i=1}^{k} c_i^y \cdot \mu_i(X) \cdot \int_y \mu_{B_i}(y)\, dy}{\sum_{i=1}^{k} \mu_i(X) \cdot \int_y \mu_{B_i}(y)\, dy} \overset{\text{def}}{=}$$
$$y^{crisp}(X|\theta) \qquad (7)$$

where, c_i^y is the center of area of the membership function $\mu_{B_i}(y)$ and is the peak (c_i^y) if $\mu_{B_i}(y)$ is symmetric; y^{crisp} is the final defuzzified output of the FRBS. $\theta = \left(c_i^y, \sigma_i^y, c_i^j, \sigma_i^j\right)$ is the parameter vector which is subject to further fine-tuning in a bid to improve the model's predictive performance. $\int_y \mu_{\hat{B}_i}(y)\, dy$ denotes the area under $\mu_{\hat{B}_i}(y)$ over the output interval $y: [y_L, y_U]$ and $\int_y \mu_{B_i}(y)\, dy$ is calculated using Eq. (8).

$$\int_y \mu_{B_i}(y)\, dy = \sigma_i^y \left[arctan\left(\frac{y_U - c_i^y}{\sigma_i^y}\right) - arctan\left(\frac{y_L - c_i^y}{\sigma_i^y}\right) \right] \overset{\text{def}}{=}$$
$$g\left(c_i^y, \sigma_i^y\right) \qquad (8)$$

Hence, after the first stage, a Singleton/Mamdani FRBS with the pre-specified number of rules is extracted from numerical data, which is analytical and can be further refined in the second modeling stage.

4.2. *Second Stage: Refinement of Initial FRBSs*

The initial fuzzy model extracted from the first stage is not optimal from two perspectives:

(i) The structure of FRBS is not optimal as far as the interpretability is concerned. The FRBS elicited from the first stage contains redundant fuzzy sets and rules.
(ii) The membership function parameters need to be tuned further as far as the accuracy is concerned.

A constrained Back-Error-Propagation (BEP) algorithm is thus utilised to first improve the accuracy of the initial FRBS in order to seed the initial population in the third modeling stage.

❖ *BEP Algorithm for Singleton FRBS*

Recall Eq. 4 in Section 4.1, where a Singleton FRBS is defuzzified with respect to $\theta = \left(c_i^y, c_i^j, \sigma_i^j | i = 1,..,k; , j = 1,..,n\right)$. By taking the partial derivative of the predictive error ε_m with respect to each parameter in θ, a set of parameter updating laws can be obtained, where $\lambda_1 \sim \lambda_3$ and $\beta_1 \sim \beta_3$ are user-specific parameters.

Singleton Consequent Updating Law: $c_i^y(t + 1) =$
$$c_i^y(t) - \lambda_1 \cdot \varepsilon_m(t) \cdot \frac{\mu_{i(t)}(X_m)}{\sum_{i=1}^{k} \mu_{i(t)}(X_m)} + \beta_1 \cdot \Delta c_i^y(t - 1) \qquad (9)$$

Centre of the Premise Updating Law: $c_i^j(t + 1) =$
$$c_i^j(t) - \lambda_2 \cdot \varepsilon_m(t) \cdot \frac{\mu_{i(t)}(X_m) \cdot q(t)}{\sum_{i=1}^{k} \mu_{i(t)}(X_m)} + \beta_2 \cdot \Delta c_i^j(t - 1) \qquad (10)$$

Spread of the Premise Updaing Law: $\sigma_i^j(t + 1) =$
$$\sigma_i^j(t) - \lambda_3 \cdot \varepsilon_m(t) \cdot \frac{\mu_{i(t)}(X_m) \cdot r(t)}{\sum_{i=1}^{k} \mu_{i(t)}(X_m)} + \beta_3 \cdot \Delta \sigma_i^j(t - 1) \qquad (11)$$

$$\varepsilon_m(t) = y^{crisp}(X_m|\theta(t)) - y_m$$

$where:$
$$q(t) = (c_i^y(t) - y^{crisp}(X_m|\theta(t))) \cdot \left(\frac{x_m^j - c_i^j(t)}{(\sigma_i^j(t))^2}\right)$$

$$r(t) = (c_i^y(t) - y^{crisp}(X_m|\theta(t))) \cdot \frac{\left(x_m^j - c_i^j(t)\right)^2}{(\sigma_i^j)^3}$$

$$\Delta c_i^y(t - 1) = c_i^y(t) - c_i^y(t - 1)$$
$and \quad \Delta c_i^j(t - 1) = c_i^j(t) - c_i^j(t - 1)$
$$\Delta \sigma_i^j(t - 1) = \sigma_i^j(t) - \sigma_i^j(t - 1)$$

❖ *BEP Algorithm for Mamdani FRBS*

Recall Eqs. 7 and 8 already developed in Section 4.1 and using the same method except this time with respect to each parameter in $\theta = \left(c_i^y, \sigma_i^y, c_i^j, \sigma_i^j | i = 1,..,k; , j = 1,..,n\right)$, following updating formulas are derivative.

Consequent's Centre Updaing Law: $c_i^y(t + 1) =$
$c_i^y(t) - \lambda_1 \cdot \varepsilon_m(t) \cdot$

$$\frac{\mu_{i(t)}(X_m)\cdot\left[g\left(c_{i(t)}^y,\sigma_{i(t)}^y\right)+c_{i(t)}^y\cdot g'\left(c_{i(t)}^y\right)-g'\left(c_{i(t)}^y\right)\cdot y^{crisp}(X_m|\theta(t))\right]}{\sum_{i=1}^{k}\mu_{i(t)}(X_m)\cdot g(c_{i(t)}^y,\sigma_{i(t)}^y)}+\beta_1\cdot$$
$$\Delta c_i^y(t-1) \tag{12}$$

Consequent's Spread Updating Law: $\sigma_i^y(t+1)=$

$$\sigma_i^y(t)-\lambda_2\cdot\varepsilon_m(t)\cdot\frac{\mu_{i(t)}(X_m)\cdot g'\left(\sigma_{i(t)}^y\right)\cdot\left[c_{i(t)}^y-y^{crisp}(X_m|\theta(t))\right]}{\sum_{i=1}^{k}\mu_{i(t)}(X_m)\cdot g(c_{i(t)}^y,\sigma_{i(t)}^y)}+$$
$$\beta_2\cdot\Delta\sigma_i^y(t-1) \tag{13}$$

Centre of Premise Updating Law: $c_i^j(t+1)=$

$$c_i^j(t)-\lambda_3\cdot\varepsilon_m(t)\cdot\frac{g\left(c_{i(t)}^y,\sigma_{i(t)}^y\right)\cdot\left[c_{i(t)}^y-y^{crisp}(X_m|\theta(t))\right]}{\sum_{i=1}^{k}\mu_{i(t)}(X_m)\cdot g(c_{i(t)}^y,\sigma_{i(t)}^y)}\cdot$$
$$\mu_{i(t)}(X_m)\cdot\left[\frac{x_m^j-c_{i(t)}^j}{(\sigma_{i(t)}^j)^2}\right]+\beta_3\cdot\Delta c_i^j(t-1) \tag{14}$$

Spread of Premise Updating Law: $\sigma_i^j(t+1)=$

$$\sigma_i^j(t)-\lambda_4\cdot\varepsilon_m(t)\cdot\frac{g\left(c_{i(t)}^y,\sigma_{i(t)}^y\right)\cdot\left[c_{i(t)}^y-y^{crisp}(X_m|\theta(t))\right]}{\sum_{i=1}^{k}\mu_{i(t)}(X_m)\cdot g(c_{i(t)}^y,\sigma_{i(t)}^y)}\cdot$$
$$\mu_{i(t)}(X_m)\cdot\left[\frac{(x_m^j-c_{i(t)}^j)^2}{(\sigma_{i(t)}^j)^3}\right]+\beta_4\cdot\Delta\sigma_i^j(t-1) \tag{15}$$

where; $\varepsilon_m\triangleq y^{crisp}(X_m|\theta)-y_m$

$$g'\left(c_i^y\right)\triangleq g'\left(c_i^y,\sigma_i^y\right)|_{c_i^y}$$
$$=\frac{1}{1+\left(\frac{y_L-c_i^y}{\sigma_i^y}\right)^2}-\frac{1}{1+\left(\frac{y_U-c_i^y}{\sigma_i^y}\right)^2}$$

$$g'\left(\sigma_i^y\right)\triangleq g'\left(c_i^y,\sigma_i^y\right)|_{\sigma_i^y}=\frac{1}{\sigma_i^y}\cdot\left[g(c_i^y,\sigma_i^y)+\right.$$
$$\left.\frac{y_L-c_i^y}{1+\left(\frac{y_L-c_i^y}{\sigma_i^y}\right)^2}-\frac{y_U-c_i^y}{1+\left(\frac{y_U-c_i^y}{\sigma_i^y}\right)^2}\right]$$

Comparing Eqs. 9~11 with Eqs. 12~15 leads to the conclusion that the two sets of parameter updating formulas are very similar to one another. The only difference lies in the fact that the latter include extra items, such as $g\left(c_i^y,\sigma_i^y\right)$ and its partial derivatives with respect to c_i^y and σ_i^y, which allows the updating formulas to adjust the spreads of the output membership functions as well.

It is worth mentioning that in this work, a constraint handling scheme is added, which checks the boundary violation for centres during each iteration step and drives any violated centres back to the boundaries.

4.3. Third Stage: Multi-Objective Fuzzy Modeling

The previous two modeling stages can be viewed as a separate structure and parameter learning. To improve the interpretability of such an approximate FRBS,

authors in Ref. 27, 33~35 performed model simplification and fine-tuning. The learning procedure described in these research investigations can still be labeled as being a separate learning process so that model simplifications rely heavily on the pre-specified thresholds according to the designer's choice. Wang *et al.*[30] proposed a hierarchical scheme to evolve both parts. However, a rule matrix was required, which rendered the scheme vulnerable to high dimensional problems due to the exponential increase in the matrix dimension. Research work reported in Ref. 28, 31, 36 adopted a variable length coding strategy in order to cope with high dimensional problems. However, heuristic variation operators are used in these works, which did not do justice to the idea of using variable length coding. In fact, it may somehow impede the search power of the Evolutionary Algorithms (EA) as far as the real-valued optimization part is concerned. Apart from these problems, most of the aforementioned works dealt with Sugeno-type with linear functions as their consequents, which detracts from the linguistic attempts of the authors' proposed methods.

The proposed approach in this work utilizes a Population Adaptive Immune Algorithm (PAIA) within a multi-objective optimization framework and a variable length coding scheme with only effective rules being encoded. Hence, it does not suffer from 'the curse of dimensionality'. A new distance index is proposed to facilitate the use of the original variation operator in PAIA. Details of these operators and the way of formulating objective functions and the initial population pool are explained followed by the description of PAIA.

✦ *Forming Objective Functions*

Two conflicting objective functions are formulated with the first focusing on the prediction accuracy and the second on the structure simplification as described in Eq. 16; $y_{prediction\_m}$ and y_{real_m} are mth predicted and real outputs; *Nrule* is the number of fuzzy rules in FRBS; *Nset* is the total number of fuzzy sets; *RL* is the summation of the rule length of each rule ('Don't care' is not included in the rule length).

$$Obj1: RMSE=\sqrt{\frac{\sum_{m=1}^{t}\left(y_{prediction\_m}-y_{real\_m}\right)^2}{t}} \tag{16}$$
$$Obj2: Complexity=Nrule+Nset+RL$$

✦ *Formation of Initial Population Pool*

The initial population is obtained with all individuals generated around the refined FRBS obtained from the first two stages using.

$$C_{initial_i}^{\ j} = \alpha \cdot range^j \cdot randn + C_{rf_i}^{\ j}$$
$$\sigma_{initial_i}^{\ j} = \beta \cdot randn + \sigma_{rf_i}^{\ j}$$
$$C_{initial_i}^{\ y} = \alpha \cdot range^y \cdot randn + C_{rf_i}^{\ y} \qquad (17)$$
$$\sigma_{initial_i}^{\ y} = \beta \cdot randn + \sigma_{rf_i}^{\ y}$$

$$range = \min\left(\left|C_{rf} - U_{limit}\right|, \left|C_{rf} - L_{limit}\right|\right) \qquad (18)$$

where, $C_{rf_i}^{\ j}$ and $\sigma_{rf_i}^{\ j}$ are the centre and spread of the ith rule and the jth input membership function of the refined FRBS extracted from the first two modeling stages. $C_{rf_i}^{\ y}$ and $\sigma_{rf_i}^{\ y}$ are the centre and the spread of the ith rule's consequent of the refined FRBS. $randn$ is a random number within [-1, 1]. '$range$' defines the minimum interval between the centre and its corresponding upper U_{limit} and lower L_{limit} limits of the input (or the output) variable, whichever is smaller. The inclusion of '$range$' is to ensure that the newly generated centres are most likely within the inputs' (or the outputs') domains. Any violation of the domains will be dragging those centres (or consequents) back to the upper or lower limits, whichever is closest. α and β are the user specified parameters which define how much different the newly generated FRBSs are from the refined FRBS. These two parameters are set to 0.2 and 0.1 respectively and the size of the initial population in this work is set to 7. However, as will be discussed later, since PAIA is not sensitive to the size of initial population, the performance of IMOFM is not sensitive to the size of its initial population, neither is it to the aforementioned two parameters.

✧ A Variable Length Coding Scheme

Different encoding schemes have been proposed for the multi-objective fuzzy modeling and can be broadly divided into two categories: 1) encoding based on the global data-base; 2) encoding based on the effective rule parameters. The former is mainly found in the linguistic modeling stream.[23, 37, 38] The latter is mainly found in the approximate modeling stream due to the lack of global data-base. The drawback of using the first encoding scheme is that it suffers from 'the curse of dimensionality'. In such a case, the length of the chromosome grows exponentially with the increased dimensions. A typical problem associated with the first

encoding scheme and the fixed length encoding scheme[39] is illustrated in Fig. 3 (a) and (b) respectively. An ineffective real-valued optimisation may be induced because some active parameter genes may interact with the inactive ones. Conversely, if only the effective rule parameters are included in the coding, a variable length coding scheme is inevitable and is used in this work (refer to Fig. 3 (b) and (c)). In such a case, the increase in the code length is only linear to the variable's dimension.
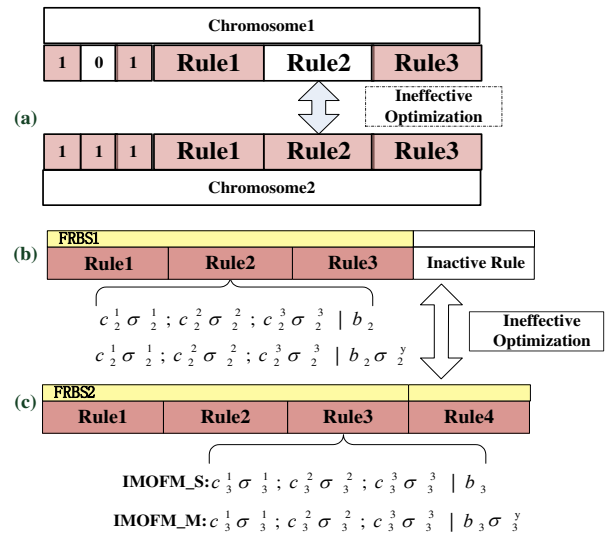

Fig. 3. (a) Ineffective optimisation caused by the interaction of inactive gene (deactivated by 0) and active gene (activated by 1); (b) and (c) variable length coding scheme for a 3-rule Singleton/Mamdani FRBS and a 4-rule Singleton/Mamdani FRBS.

✧ Variation Operators and a New Distance Index

The variation operator used in PAIA is *Affinity Maturation* which mutates copied (cloned) solutions based on their distance (affinity) to the identified good solution (see the following part for more details). Such a variation operator is used in IMOFM to optimise the encoded parameters. However, in the original PAIA, such distance is calculated between two fixed-length individuals. Given the variable length coding scheme and the unconstrained optimization used in this work, a concomitant effect of the so-called 'unordered sets of rules'[40] may occur as shown in Fig. 4, where FRBS1 and FRBS2 are exactly the same. However, because of the blind search mechanism, values encoded in 'Rule1' and 'Rule7' became different within the two FRBSs.

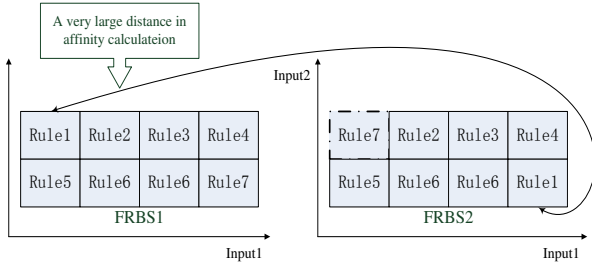Alternatively, rules may be deleted, e.g. Rule7 in FRBS2.

Fig. 4. The problems associated with the FRBS having different rule lengths and unconstrained optimisation.

Hence, a special procedure is required to align the closest rules from different FRBSs in order to have a meaningful crossover on the 'unordered sets of rules'.[40, 41] To tackle this problem, a new distance index is proposed. The basic idea is to find the distance of the closest rules in different FRBSs rather than the distance of the corresponding rules. The mathematical description of the idea is as follows:

$$dist(R_j, R_k) =$$

$$\frac{\sum_{i1=1}^{k1}\sum_{l=1}^{rl} abs\left(R_j^{i1}(l)-R_k^{C_{i1}}(l)\right)+\sum_{i2=1}^{k2}\sum_{l=1}^{rl} abs\left(R_k^{i2}(l)-R_j^{C_{i2}}(l)\right)}{rl\cdot(k1+k2)} \quad (19)$$

where, R_j and R_k are two FRBSs with $k1$ and $k2$ rules; rl is the length of the rule; $R_k^{C_{i1}}$ ($R_j^{C_{i2}}$) represents the closest rule in R_k (R_j) with respect to the $i1th$ ($i2th$) rule in R_j (R_k); $abs(*)$ is the absolute value of $*$.

❖ *Model Simplification*

As one can see from Fig. 2, a model simplification step is added to the third stage in order to remove the redundancy both in the rules and in the fuzzy sets. This is to ensure the optimization of FRBS structure along with the accuracy at the same time. Five simplification steps are involved for each mutated FRBS:

(i) **Removing Unimportant Rules:** the unimportant rules are those rules that contribute the least to any prediction error increase when not including this rule, as described by Eq. 20. This occurs because other rules may already have covered the input region under these rules.

$$Insignificant_{rule} = \min_i|RMSE_{AR} - RMSE_{\bar{\imath}}| \quad i = 1, \dots, k \quad (20)$$

where, $RMSE_{AR}$ is the root mean square error when all the rules in the rule base are used for predicting; $RMSE_{\bar{\imath}}$ is the predictive error associated with the rule base when the ith rule is temporarily excluded. Insignificant rules are deleted when the following condition is met:

$$\left(\frac{cr}{maxr}\right)\cdot rnd > p_m \quad (21)$$

where, cr is the number of rules in the current FRBS; $maxr$ is the maximum allowable number of rules, which equals the number of clusters used in the first modelling stage; rnd is a random number between [0, 1]. p_m is a design parameter which limits the fewest rules in FRBS and has been set to 0.5 in this work without any loss of generality. At each iteration step, each cloned individual has one insignificant rule removed unless the rule base reaches the fewest rules designated by Eq. 21.

(ii) **Merging Similar Rules:** Similar rules should be merged together by taking the mean values of the corresponding fuzzy sets to keep the FRBS consistent and parsimonious. To measure the similarity of rules, the so-called similarity of rule premise (SRP)[29] is used in this work. The following condition should be met for merging a pair of similar rules of each cloned individual at each iteration step:

$$SRP(i,l) \triangleq min\left\{S\left(A_i^j, A_l^j\right), \begin{matrix} j = 1, \dots, n \\ i, l = 1, \dots, k; i \neq l \end{matrix}\right\} > th\_mr = rnd \cdot (1 - mr) + mr \quad (22)$$

where, $S\left(A_i^j, A_l^j\right)$ are the similarity between two fuzzy sets and will be explained later; th_mr is the threshold which randomly changes between $[mr, 1]$ every t (specified by the user) iterations and mr is 0.95 in this work without any loss of generality

(iii) **Removing Universal Fuzzy Sets:** Fuzzy sets which meet the following condition are regarded as universal fuzzy sets and are therefore deleted:

$$S\left(A_i^j, U\right) > th\_ufs \triangleq rnd \cdot (1 - ufs) + ufs \quad (23)$$

where, U is the universal fuzzy set; th_ufs is the threshold which randomly changes between $[ufs, 1]$ every t generations and ufs is 0.85 in this work. For computation purpose, if the width of a fuzzy set is more than two times wider than the universe of discourse of the corresponding dimension, it is regarded as the universal fuzzy set.

(iv) **Merging Similar Fuzzy Sets:** Jin[42] proposed a simplified similarity measure based on the distance measure if Gaussian membership functions are

involved. Two fuzzy sets are considered to be similar if the following condition is met:

$$S(A_i^j, A_l^j) \ or \ S(A_i^y, A_l^y) > th\_sfs \triangleq rnd \cdot (1 - sfs) + sfs$$

$$S(A_i^j, A_l^j) = \frac{1}{1+\sqrt{(c_i^j-c_l^j)^2+(\sigma_i^j-\sigma_l^j)^2}} \quad (24)$$

$$S(A_i^y, A_l^y) = \frac{1}{1+\sqrt{(c_i^y-c_l^y)^2+(\sigma_i^y-\sigma_l^y)^2}}$$

where, th_sfs is the threshold which randomly changes between $[sfs, 1]$ every t generations and sfs is set to 0.95 in this work. The mean values of two similar fuzzy sets are calculated in order to substitute the original two fuzzy sets.

It is worth mentioning that all the simplification processes, except for the 'insignificant rules', have only α chance to be evoked at each iteration, where α is taken to be 20% in this work. The similarity measures mentioned in Eqs. 23 and 24 will be checked for each fuzzy set. Only the ones with the maximum similarity values will be deleted or merged during each iteration step provided the conditions are also met. For this reason and because of the elitism which records any non-dominated solution found at each iteration step, it was found (see Section 6.1) that the aforementioned thresholds are not critical parameters within certain ranges.

✧ *A Population Adaptive Immune Algorithm*

PAIA[9, 10] is an immune inspired multi-objective optimization algorithm. Fig. 5 summarizes how PAIA is adapted for IMOFM.

```
pop = random_initiation();
obj = Obj_Evaluation (pop);
[nondominated_pop, dominated_pop] = non_domination(pop, obj);
for i = 1 : Gen
    X_identified = random_pick_one(nondominated_pop);
    N = size(nondominated_pop);
    for X_j ∈ nondominated_pop
        Affinity_j = ∑_{j=1}^{N} dist(X_identified, X_j) /N;
    end
    for X_m ∈ dominated_pop
        Affinity_m = dist(X_identified, X_m);
    end
    [pop_s, pop_us] = Clonal_Selection(pop, Affinity, δ);
    [c_s, c_us]= clone(pop_s, pop_us, N_cmax);
    [c_am] = Affinity_Maturation(c_s, c_us);
    c_sim = Model_Simplification(c_am);
    obj_sim = Obj_Evaluation (c_sim);
    [pop_rs, obj_rs] = Reselection(pop, obj, c_sim, obj_sim);
    [pop, obj] = Network_suppression(pop_rs, obj_rs, σ);
    [nondominated_pop dominated_pop] = non_domination(pop, obj);
end
```

Fig. 5. Pseudo-code of PAIA for IMOFM.

As one can see from Fig. 5, after a random initial population pool is generated, one of the good solutions $x_{identified}$ from the non-dominated set will be randomly selected in order to activate the rest of the solutions. The corresponding affinity (fitness) for each solution is calculated by using $dist(*,*)$ as discussed before in Eq. 19. *Clonal_Selection()* will differentiate current solutions into selected solutions (pop_s), if their affinity values are the smallest one or smaller than a threshold δ, and otherwise into unselected solutions (pop_us). *Clone()* will assign a fraction of the maximum clone size (N_{cmax}) to each pop_s and make a copy of themselves. The higher the affinity percentage the larger the fraction is assigned. For pop_us, it will be cloned only once regardless of their affinity. *Affinity_Maturation()* will mutate clones (c) according to the following equations:

$$c\_am(i) = c(i) + \alpha \cdot N(0,1) \ i = 1, \dots, n;$$

$$a = r \times \frac{exp(Affinity)}{exp(1)} \quad (25)$$

$$r = 1 - rand^{\left(1-\frac{G}{Gen}\right)}$$

Where, *N(0,1)* is a Gaussian random variable with zero mean and standard deviation 1; *i* represents the dimension that has been chosen to mutate; *G* is the current iteration and *Gen* is the predefined total number of iterations. It is worth mentioning that, for c_s, one dimension is randomly chosen to mutate, while for c_us more than one dimensions (two, in PAIA) are randomly chosen to mutate. *Model_Simplification()* simplifies the mutated clones via Eqs. 20~24 as discussed previously. *Reselection()* selects good solutions from the combined population of c_sim and their parents based on their non-dominance and the size of the current non-dominated solutions. *Network_suppression()* is used to adaptively control the size of population at each iteration step by suppressing too closed solutions if their Euclidian distance in the objective space is less than a predefined threshold σ. In such a case, the one with larger affinity value is deleted.

The computation complexity of PAIA for the block of *non_domidation()* is $O(N^2)$, where *N* is the size of *pop* at each iteration step. For evaluating the objective functions, this complexity is governed by $O(N_{cmax})$ at each iteration step. It has been shown in Refs. 9~10 that N_{cmax} is not problem dependent and 95 is an empirically good number. Increase N_{cmax} does not lead to any improvement of the optimisation. Since *pop* and

the clone size for each *pop_s* are adaptive with respect to the search process and only relatively good solutions are selected and maintained, PAIA can largely reduce the number of evaluation times comparing to traditional EAs and is not sensitive to the initial population size.

5. IMOFM Implementation Issue

Due to the simultaneous optimisation of the rule base structure and its parameters in the third modeling stage, some issues regarding the practical implementation of the algorithm should be treated with a special caution and deserves more exploration in this Section. In the following space, two issues are discussed.

5.1. *Coding representation and the rule base*

The mechanism behind coding calls for some comments because the rule-based structure and the associated parameters are optimized for a final outcome. As a result there exist two representations: one for structure optimization and the other for parameter optimization. The interface between the two representations is of prime importance when both optimizations are performed simultaneously. Without this interface, the parameter optimisation operated over the coding representation may lose vital structural information which is constantly being modified during the structure optimisation. In order to build up such a link, two concepts, namely 'FISmap' and 'RULE', are introduced into IMOFM and are shown in Fig. 6.
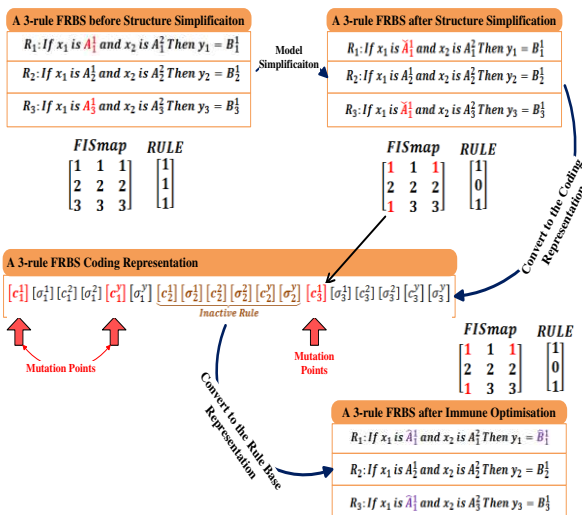

Fig. 6. If a link is set up, no missed mutation points are induced; inactive rules are not actually deleted but marked so that it will not participate in any computations afterwards.

The first concept is the so-called 'FISmap' matrix, which is a $k \times n$ matrix for IMOFM_S and $k \times (n + 1)$ for IMOFM_M. The elements of the ith row are all initialised to their row number and will be constantly updated so that it can reflect the current status of the rule base. The number stored in each element serves as the identification number of each membership function. For example, during the interpretability improvement operation at each iteration step, if the membership function of the first input in the ith rule is very similar to the membership function of the same input in the jth rule, two membership functions in the rule-base representation will be merged into a single one. In order to reflect such changes in the rule base structure, FISmap is updated, and if $i < j$, FISmap(i, 1) will remain to its initialised number 'i' and FIS(j, 1) will be updated using the smaller number 'i'. By doing so, two similar membership functions would have been combined into a single one and their corresponding identification numbers would also have been updated using the smaller value. Universal fuzzy set is marked as 'inf' ('inf' represents infinity in Matlab[®]) in FISmap. The second concept relates to a so-called vector 'RULE', which is a $k \times 1$ vector initialised with 1. This vector serves as the flag to indicate which rule in the rule-base is active and which rule is inactive. Rules are deleted or merged, which lead to the corresponding elements in 'RULE' flipping from 1 to 0 (hence, 'Rule 2' is an inactive rule as shown in Fig. 6).

5.2. *Spreads of the output membership function*

This issue only relates to IMOFM_M. Due to unconstrained optimisation, it is very likely that some spreads of the output membership functions become wide enough to be considered as the universal fuzzy set. However, it is every hard to associate any meaningful linguistic terms with the universal fuzzy set for the consequents. The solution to this problem is to impose a constraint on the spreads of the output membership functions so that they will not exceed 1 in a normalised universe of discourse.

6. Experiments

6.1. *A benchmark problem*

The benchmark example used in this section is a nonlinear static system with two inputs and one output,

which has been studied in Ref. 30. The system is defined as follows:

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, \quad 1 \le x_1, x_2 \le 5 \quad (26)$$

Although this problem is a simple low-dimensional problem, it is a very good example in terms of demonstrating how IMOFM works. The same 50 input-output data pairs as those used in Refs. 6 and 43 are collected.

The number of clusters is set to 5 in the first modeling stage. The refined 5-rule initial FRBS is used to seed the initial population in the third modeling

stage. The initial population size is set to 7. The number of iterations in the third stage is set to 1200. The network threshold (σ) of PAIA is set to 0.0008 for this example to manage the population within the solution pool. Other parameters of PAIA are kept the same as those in Refs. 9~10. In order to obtain a quantitative comparison of the proposed method with other well-known fuzzy modeling paradigms, IMOFM is compared with the methods proposed in Refs. 6, 30, 43 44 and 45. Table 1 summarizes such comparative results focusing on their predictive performances (RMSE) and the number of rules. The results in Table 1 include the average values of 30 runs.

Table 1. Comparisons of the predictive performance of different modeling methods.

Modeling Methods (Ref.)	No. of rules	No. of fuzzy sets[&]	No. of Parameters	Consequents	Performance (RMSE training)	
Ref. 44	6	12 trapzoidal[*]/Gaussian[@]	30[*]/42[@]	Singleton	0.5925[*]	0.0707[@]
Ref. 43	6	12 trapzoidal	72	Fuzzy sets	0.5639[*]	0.2811[@]
Ref. 6	5	10	25	Singleton	0.5604[*]	0.3391[@]
Ref. 30						
Initial	6	12 Gauss2mf.	66	Linear	-	0.1755[@]
Pareto FRBS1	7	6 Gauss2mf.	45	Linear		0.0298[#]
Pareto FRBS2	4	3 Gauss2mf.	24	Linear		0.0520[#]
Pareto FRBS3	3	2 Gauss2mf.	17	Linear		0.0719[#]
Ref. 45						
Pareto FRBS1	4	8 Gaussian	28	Linear		0.0656[#]
Pareto FRBS2	4	5 Gaussian	22	Linear		0.0883[#]
Pareto FRBS3	3	5 Gaussian	19	Linear		0.1382[#]
Pareto FRBS4	2	4 Gaussian	14	Linear		0.2750[#]
IMOFM_S (NB: Average results over 30 runs are presented here)						
Average execution time (Intel(R) Core(TM)2 Duo CPU, 2.27 GHz): 1st stage: 0.5460sec; 2nd stage: 120sec; 3rd stage: 213sec						
Initial FRBS	5	10 Gaussian	25	Singleton	0.5954[*]	0.0688[@]
Pareto FRBS1(30 times)	5	10 Gaussian	25	Singleton	0.0688[#]	$\sigma^2: 0$
Pareto FRBS2(30 times)	5	9 Gaussian	23	Singleton	0.0696[#]	$\sigma^2: 0$
Pareto FRBS3(29 times)	4	8 Gaussian	20	Singleton	0.0930[#]	$\sigma^2: 0.0105$
Pareto FRBS4(29 times)	3	6 Gaussian	15	Singleton	0.1417[#]	$\sigma^2: 0.0045$
Pareto FRBS5(30 times)	2	4 Guasian	10	Singleton	0.2484[#]	$\sigma^2: 0.0015$
Pareto FRBS6(25 times)	2(5[T])	3 Gaussian	6	Singleton	0.4769[#]	$\sigma^2: 0.0719$
IMOFM_M (NB: Average results over 30 runs are presented here)						
Average execution time (Intel(R) Core(TM)2 Duo CPU, 2.27 GHz): 1st stage: 0.5460sec; 2nd stage: 37sec; 3rd stage: 229sec						
Initial FRBS	5	15 Gaussian	30	Mamdani	0.6078[*]	0.0702[@]
Pareto FRBS1(14 times)	5	15 Gaussian	30	Mamdani	0.0633[#]	$\sigma^2: 0.0005$
Pareto FRBS2(25 times)	5	14 Gaussian	28	Mamdani	0.0651[#]	$\sigma^2: 0.0023$
Pareto FRBS3(22 times)	5	13 Gaussian	26	Mamdani	0.0691[#]	$\sigma^2: 0.0017$
Pareto FRBS4(26 times)	4	11 Gaussian	22	Mamdani	0.0781[#]	$\sigma^2: 0.0034$
Pareto FRBS5(25 times)	4	10 Gaussian	20	Mamdani	0.0961[#]	$\sigma^2: 0.0032$
Pareto FRBS6(28 times)	3	9 Gaussian	18	Mamdani	0.1311[#]	$\sigma^2: 0.0152$
Pareto FRBS7(28 times)	3	8 Gaussian	16	Mamdani	0.1846[#]	$\sigma^2: 0.0193$
Pareto FRBS8(25 times)	2	6 Gaussian	12	Mamdani	0.2482[#]	$\sigma^2: 0.0019$
Pareto FRBS9(28 times)	2(5[T])	5 Gaussian	10	Mamdani	0.2718[#]	$\sigma^2: 0.0617$

[&] For IMOFM_S, it is the number of fuzzy sets in its inputs; for IMOFM_M, it is the number of fuzzy sets in its inputs and output.

[*] Initial model extracted directly from data using clustering algorithms or grid partition methods.

[@] Refined model or the consequents are computed through the estimation methods.

[#] Simplified model after model simplification and parameter fine tuning.

[T] Total number of rule length.

σ^2 Stardard deviation of the results obtained from 30 runs.

One challenge associated with EAs-based multi-objective fuzzy modeling algorithms is how to include the results from different runs. This is because the algorithms of this type are stochastic in their nature. Hence, Table 1 also records the number of each FRBS' configuration found within the 30 runs using the whole three-stage modeling procedure. Most configurations are found more than 20 times within 30 runs, which suggests that the proposed modeling method is robust and consistent. It is worth mentioning at this stage that the FRBS with a short rule length was identified. This is mainly attributed to the merging of some fuzzy sets with the universal fuzzy set. The proposed method is also compared to other modeling approaches with singleton or fuzzy sets as their consequents, and it was found to represent the most accurate results with simpler rule-base structures. In contrast, Refs. 30 and 45 adopted linear Sugeno structure. As can be seen from this table, these two methods produced slightly better predictions using fewer rules, e.g. four rules, compared to five rules in the proposed work. However, due to the linear combinations in the consequents, the number of parameters involved in these two works and the proposed work is more or less the same. Due to the constraint on space, Fig. 7 only shows the Pareto fronts obtained by using IMOFM_S from one of the 30 runs.
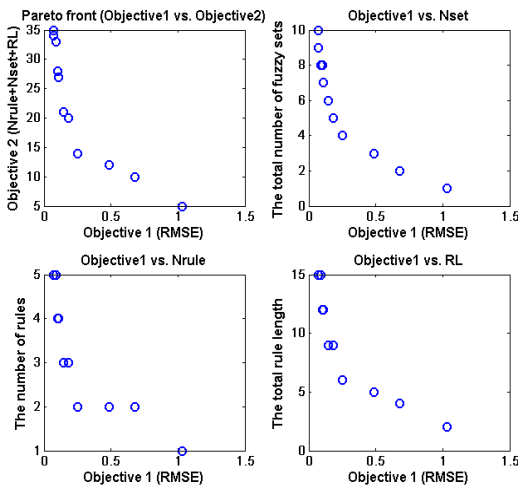

Fig. 7. The Pareto fronts obtained by using IMOFM_S.

Fig. 8 and 9 only show the results of IMOFM_M. For the third modeling stage, a 4-rule simplified FRBS with 7 fuzzy sets in its inputs and 3 fuzzy sets in its consequents is chosen for the illustration purpose. Similar results have been obtained for IMOFM_S.
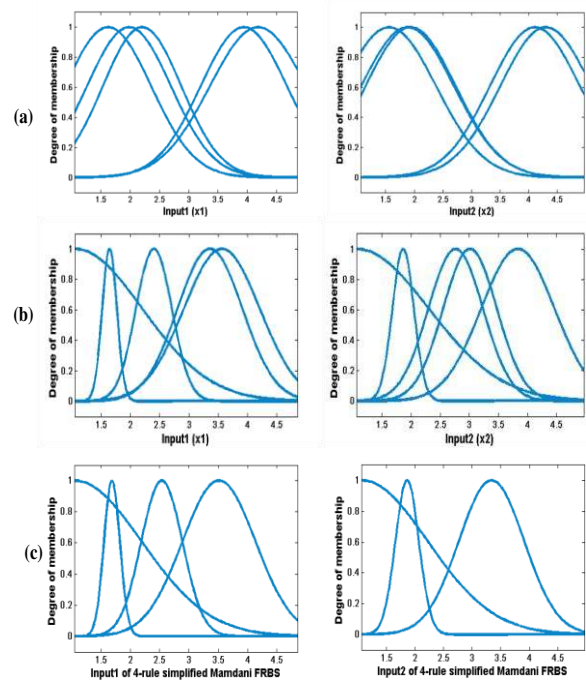

Fig. 8. The membership functions from (a) 1st stage; (b) 2nd stage; (c) 3rd stage.
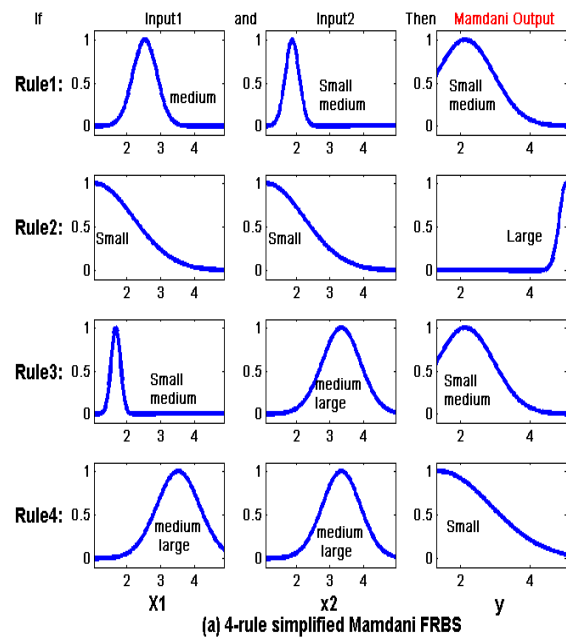

Fig. 9. 4-rule simplified Mamdani FRBS.

Fig. 10 show the predictive performances of the simplified Singleton and Mamdani FRBS. For the Singleton FRBS, a 4-rule simplified FRBS with 7 fuzzy sets is chosen for the purpose of illustration.
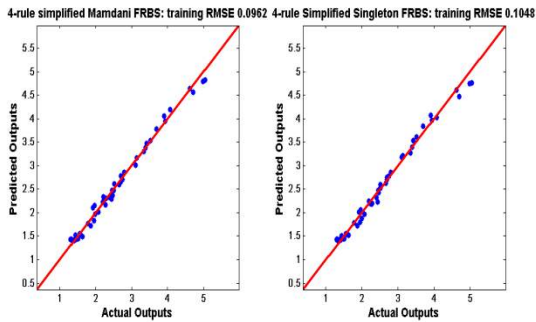
Fig. 10. Predictive performances of the simplified Mamdani and Singleton FRBSs.

Based on this benchmark problem, analysis of IMOFM_S in terms of the influences of using different modeling stages and user-specified parameters on performance is given below. The observations are always held for IMOFM_M unless otherwise stated.

✦ *Influence of the Modeling Stages on Performance*

In order to test the influences of the each modeling stages, two variants of the proposed IMOFM are investigated: 1) the combination of the first stage and the third stage; 2) only the third stage. In the first case, an initial 5-rule FRBS is generated using the clustering algorithm, which is then fed to the third stage without any refinement. While in the latter case, the initial 5-rule FRBS is randomly generated within the variable domains. Table 2 summarizes the results of the two variants.

   As shown in Table 2, more iterations are needed for the two variants to achieve a similar predictive performance as that obtained using the three-stage modeling procedure (refer to Table 1), and only a few Pareto FRBSs are obtained. The most complex structure which is supposed to evolve to the most accurate FRBS is discarded during the optimization since FRBSs with more rules may be replaced by FRBSs with fewer rules in the early iterations. In such a case, the 'Pareto' selection favours the one with a simpler structure. All these justified the inclusion of the first two stages.

✦ *Influence of the User-Specified Parameters on Performance*

IMOFM includes a set of user-defined parameters, among which some are inherited from PAIA and others are mainly associated with the third modeling stage. The investigations on how these parameters affect the

Table 2. Comparisons of the predictive performance of different modeling stages for the benchmark problem.

Modeling Methods (Ref.)	No. of rules	No. of fuzzy sets[&]	The type of FRBS	Performance (RMSE training)
IMOFM (the first stage and the third stage); numbeer of iterations: 3000				
Initial FRBS	5	10	Singleton	0.6069[*]
FRBS1	5	6	Singleton	0.1183[#]
FRBS2	4	6	Singleton	0.1268[#]
FRBS3	3	5	Singleton	0.1724[#]
FRBS4	2	4	Singleton	0.2475[#]
FRBS5	2(4[T])	2	Singleton	0.7235[#]
IMOFM (only the third stage); number of iterations: 4000				
Initial FRBS	5	10	Singleton	1.0363[*]
FRBS1	4	7	Singleton	0.1116[#]
FRBS2	4	6	Singleton	0.1223[#]
FRBS3	3	5	Singleton	0.1502[#]
FRBS4	3(8[T])	4	Singleton	0.1753[#]
FRBS5	3(7[T])	4	Singleton	0.3211[#]

[*]: Initial model extracted directly from data using clustering algorithms or grid partition methods; T: Total number of rule length;
[#]: Simplified model after model simplification and parameter fine tuning.

performance of IMOFM are carried out. The emphasis has been given to two PAIA affiliated parameters, namely the initial population size and σ, and four model-simplification parameters, namely p_m, mr, ufs and sfs.

   IMOFM is not sensitive to the size of the initial population. This parameter does have an impact on the speed of the algorithm's convergence. However, given enough evaluation times the initial population size and the accuracy of IMOFM have no causal relationship. This has been proved in Fig. 11, where the average results of 10 independent runs with the initial population size varied from 1 to 10, each of which executes 1000 iterations (which are considered as enough evaluation times) are shown. The non-dominated FRBSs with different initial population sizes produced very close Pareto fronts, which means that the initial population size is not a critical parameter.

   σ is also not a critical parameter as far as the convergence accuracy is concerned. However, it is an important factor as far as the number of the obtained Pareto solutions is concerned. Without the need for increasing the evaluation times greatly, more Pareto solutions may be obtained by simply adjusting σ. This property is regarded as one of the advantages of PAIA
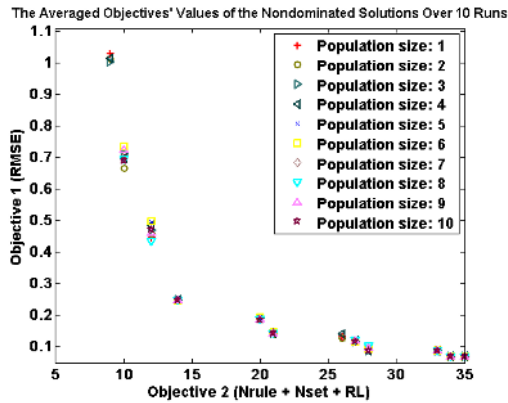
Fig. 11. The averaged Pareto fronts found in 10 independent runs with different initial population sizes.

comparing to other EAs. Fig. 12 shows the Pareto fronts obtained using 0.1 and 0.0008 as their network suppression threshold σ. The results are obtained over 10 independent runs, each of which executes 1000 iterations.
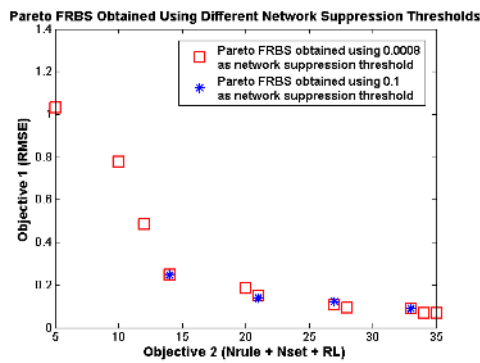

Fig. 12. Pareto FRBS obtained using different σ.

A bigger value of the threshold means more Pareto solutions will be suppressed. A Smaller value means more Pareto solutions will be allowed to enter into the memory set during each iteration step. Hence, more Pareto FRBSs are in the final population. However, as can be seen in Fig. 12, this parameter does not affect the predictive performance of the elicited model.

In order to investigate the effects of the model-simplification parameters, only one parameter is selected each time with its value varied from 0 to 1. Other parameters are kept as constant so that one can concentrate on analysing the selected parameter.

The first parameter to be investigated is p_m which is responsible for removing insignificant rules. p_m can vary from 0 to 1. In order to bypass the effects of other

model simplification parameters, mr, ufs and sfs are all set to 1. Fig. 13 shows how p_m affects the least number of rules that IMOFM can obtain. It is worth mentioning that such a number is also affected by other model simplification parameters and thus may not be strictly specified by p_m. In this work, p_m is set to 0.5.
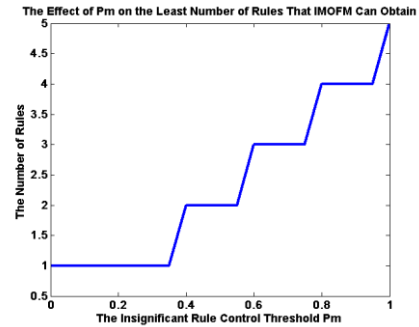

Fig. 13. The effect of p_m on the least number of rules that IMOFM can obtain.

In order to investigate the effect of the threshold mr which is responsible for merging similar rules, mr is varied from 0 to 1 with $p_m = 0.5$, $ufs = 1$ and $sfs = 1$. In such a situation, a small value of mr will result in rules being merged even if they are quite different, and this may ultimately affect the predictive performances of the elicited FRBS. Fig. 14 shows the predictive performances of four Pareto FRBSs against different values of mr. As shown in the figure, IMOFM with mr varied between 0.9 and 1 generally produces more accurate predictions. Hence, mr with a large value, e.g. between 0.9 and 1, is a preferable choice.
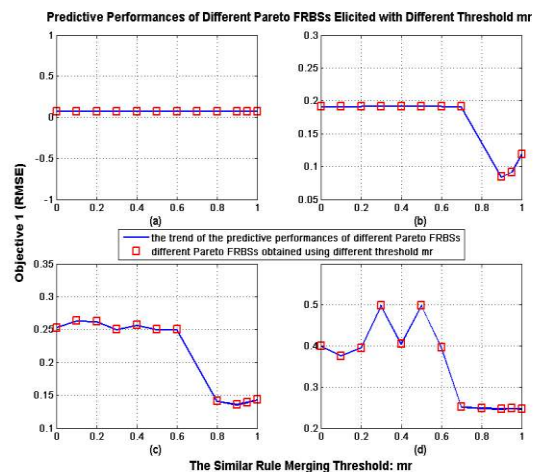

Fig. 14. (a) 5-rule FRBS; (b) 4-rule FRBS; (c) 3-rule FRBS; (d) 2-rule FRBS.

ufs is the threshold which decides when a fuzzy set can be regarded as the universal fuzzy set. A small value of this threshold means more fuzzy sets can be considered as the universal fuzzy set, and vice versa. If $p_m = 0.5$, $mr = 0.95$ and $sfs = 1$, a small value of ufs will result in more Pareto FRBS to be found since in such a case the probability of having similar rules are increased due to the deletion of some fuzzy sets. However, useful fuzzy sets may be deleted just because they are 'wide enough' to be considered as the universal fuzzy set in the case of using a small value of ufs. In such a case, the predictive performances of the elicited FRBS may be affected. Fig. 15 demonstrates how different ufs affects the predictive performances. Generally speaking, values between 0.6 and 1 are good values.
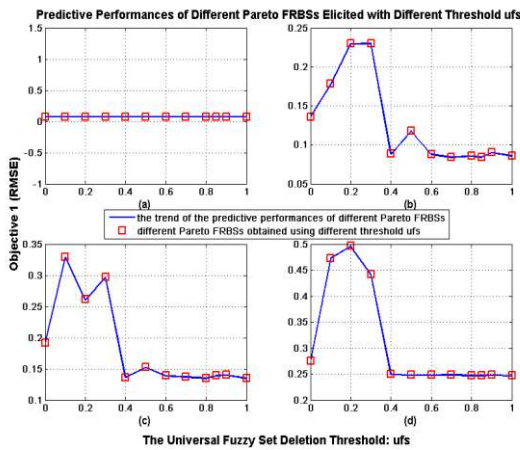

Fig. 15. (a) 5-rule FRBS; (b) 4-rule FRBS; (c) 3-rule FRBS; (d) 2-rule FRBS.

As a general guide, sfs should be set to values smaller than 1 so that not only the same fuzzy sets but also similar ones can be merged. When similar fuzzy sets are merged, rules may become similar so that they will be merged consequently, a consideration which makes the rule-base more compact. Fig. 16 confirms that when sfs takes values smaller than 1 ($p_m = 0.5$, $mr = 0.95$, $ufs = 0.85$) IMOFM can produce more Pareto FRBSs.
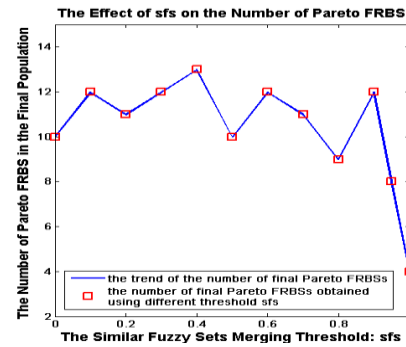

Fig. 16. The effects of *sfs* on the number of Pareto FRBSs.

6.2. *Real World Applications*

The proposed modeling method is tested further with a high-dimensional real world engineering application associated with the mechanical property prediction of hot rolled steels. Specialist heat treatments are used to develop the required mechanical properties in a range of alloy steels. The mechanical properties of the alloy steels rest with many factors of which the followings are believed to be the major ones (refer to Table 3): tempering temperature, quench type, chemical compositions of the steel, geometry of the bar, test sample location on the bar, batch distribution in the furnace, measurement tolerances and variations in the process equipment and operators.[46] Traditionally, a heat treatment metallurgist would try to balance these factors using their metallurgical knowledge and experience in a bid to obtain the desired mechanical properties. However, due to the increasing complexity of the underlying system, it becomes more difficult even for the metallurgists to tune these parameters. Given the lack of the mathematical models which can account for these complex systems and a large amount of available industrial process data associated with the systems, data-driven modeling becomes more vital for assisting the metallurgist to predict the mechanical test results without actually doing it and to understand the underlying process. Based on these models, further optimisations of the heat treatment process can also be developed, which is envisaged to be able to automate

Table 3. The inputs and output of UTS data set.

Inputs	Test Depth	Size	Site	%C	%Si	%Mn	%S	%Cr
Max.	140	381	6	0.62	0.35	1.72	0.21	3.46
Min.	4	8	1	0.12	0.11	0.35	5e-4	0.05
Inputs	%Mo	%Ni	%Al	%V	Hardening Temperature	Cooling Medium	Tempering Temperature	/
Max.	1	4.16	1.08	0.27	980	3	730	/
Min.	0.01	0.02	5e-3	1e-3	820	1	170	/
Outputs	Tensile Strength (Max.: 1842; Min.: 516.2)							

the steel design process and reduce the experimental cost.

In this part, the problem of predicting Ultimate Tensile Strength (UTS) of heat-treated steel is used as a case study, which involves knowledge acquisition from real industrial data. UTS data set consists of 3760 data samples and includes 15 inputs and one output as shown in Table 3. In order to compare with the work in Ref. 39, the UTS data set is randomly divided into two parts: 75% of the data are used for training and the remaining data are used for testing. Another 12 more recent samples are used as the unseen data set to validate the generalisation properties of the model. The maximum number of rules is set to 12. The number of iterations for the third modeling stages is set to 1200.

The results presented in Table 4 include the average values of 10 independent runs and only a few 'Pareto' FRBSs are presented due to the constraint on space. The problem of over-fitting specifically related to the second modeling stage under unseen situations is revealed in Table 4. Such over-fitting is mainly attributed to the complex structures involved in the first two modeling stages. However, the simplified fuzzy models can predict well even under unknown scenarios.

Fig. 17 shows the predictive performance of a 7-rule simplified Mamdani FRBS. Fig. 18 compares the membership functions on a few selected inputs and output from the refined FRBS and the simplified 7-rule Mamdani FRBS. A much improved transparency has been achieved. Fig. 19 shows the snapshot of the obtained approximate Pareto fronts at different iterations. The evolution starts from the most accurate FRBS and expands the Pareto front during the course of the optimization. Due to the constraints of the space, only 3 selected rules from the 7-rule simplified mamdani FRBS are shown in Appendix A.
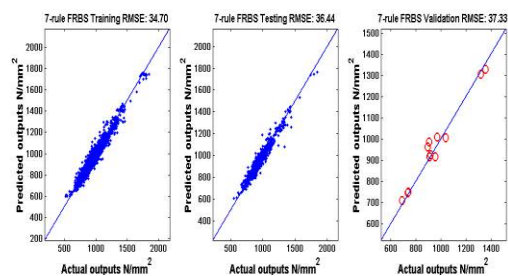

Fig. 17. The predictive performance of a 7-rule simplified FRBS (left to right: training, testing and validation).
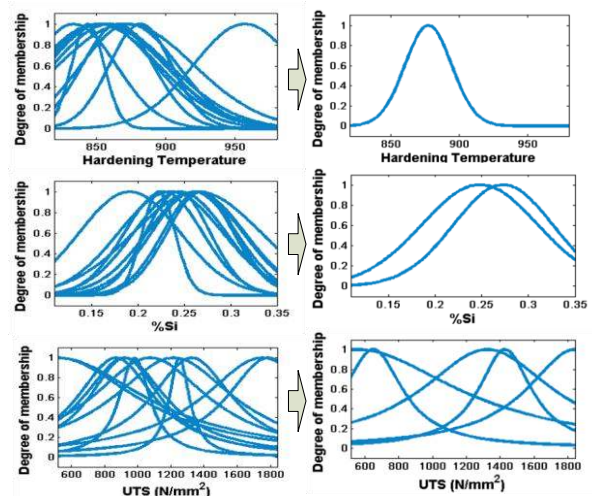

Fig. 18. The fuzzy sets of input 13 (hardening tempature), 5 (Si) and output (UTS) from the second (left) and the third modeling stage (right).

Table 4. Comparisons of the predictive performance of different modeling methods using UTS data.

Modeling Methods	First Stage (clustering algorithm)		Second Stage (single objective refining)		
	Training (RMSE)	Testing (RMSE)	Training (RMSE)	Testing (RMSE)	Validation (RMSE)
Ref. 34	100.54	108.26	37.45	43.07	-
IMOFM_S	113.54	112.32	30.93	35.65	53.61
IMOFM_M	120.43	123.44	31.21	35.49	**37.23**

Third Stage (multi-objective fuzzy modeling)				
Modeling Methods	**No. of rules**	**No. of Fuzzy sets in inputs**	**Modeling performance**	
			Training (RMSE)	Testing/ Validation
Ref. 39 Pareto FRBS1	12	Inputs: [9 11 10 12 8 10 8 9 10 10 6 11 10 10 10 10] Output: 10	37.45	43.07/-
Pareto FRBS2	9	Inputs: [9 7 8 7 5 6 4 6 8 8 2 6 7 8 7], Output: 9	42.82	43.90/-
IMOFM_S Pareto FRBS1	10	Inputs: [4 7 8 8 4 7 3 8 7 7 3 4 4 7 7], Outputs: 10	32.38	34.82/41.01
Pareto FRBS2	8	Inputs: [2 4 4 7 3 3 3 5 4 5 2 2 3 6 6], Output: 8	36.43	37.63/**31.54**
Pareto FRBS3	7	Inputs: [3 4 4 4 1 3 3 4 3 4 1 1 2 6 5], Output: 7	42.91	43.87/46.34
IMOFM_M Pareto FRBS1	10	Inputs: [8 9 10 10 6 10 6 9 9 7 4 7 6 10 9], Output: 10	31.21	35.32/**35.65**
Pareto FRBS2	7	Inputs: [5 7 7 7 2 4 3 6 6 6 2 3 1 7 7], Output: 5	34.70	36.44/**37.80**
Pareto FRBS3	6	Inputs: [2 2 2 5 2 2 1 4 3 3 0 2 1 2 4], Output: 5	45.83	44.30/49.87

Average execution time (Intel® Core(TM)2 Duo CPU, 2.27 GHz)
IMOFM_S: 1st stage: 1 min.; 2nd stage: 25 min.; 3rd stage: 3.7 hrs
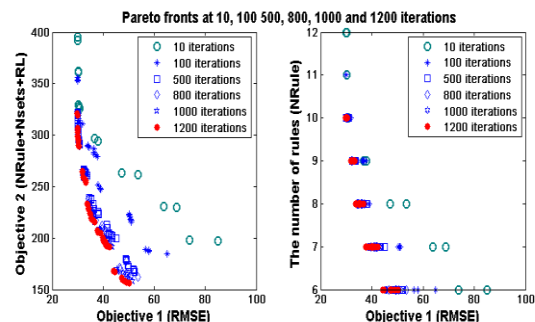IMOFM_M: 1st stage: 1 min.; 2nd stage: 30 min.; 3rd stage: 4 hrs

Fig. 19. The Pareto FRBSs at different iterations.

Table 5 summarizes the results of the UTS modeling problem using IMOFM_S with and without the variable length coding and the new distance index. Much bigger improvements have been registered for the FRBS with fewer rules since they are more prone to suffering from the problem of 'unordered set of rules'.

Table 5. The comparison of the modeling approaches with and without variable length coding scheme.

FRBS	No. of rules	IMOFM_S (without VLC) (Training RMSE)	IMOFM_S (with VLC) (Training RMSE)	Improvement (%)
Pareto FRBS1	11	29.782	29.671	0.3%
Pareto FRBS2	10	30.024	29.882	0.5%
Pareto FRBS3	8	36.762	35.740	7.0%
Pareto FRBS4	6	47.780	42.581	10.9%

7. Conclusions

In the paper, a systematic immune inspired multi-objective fuzzy modeling framework, namely IMOFM, is introduced. The main novel contributions of the proposed modeling framework are considered as follows: (1) the proposed modeling approach is not sensitive to the initial starting points due to the evolutionary based clustering algorithm used in the first stage; (2) Only the maximum allowable number of rules is required *a priori* since in the third stage a set of Pareto FRBS with different structure are elicited.; (3) due to the first two stages, the efficiency and predictive accuracy of the modeling are improved; (4) by using the variable length coding scheme and a new distance index, the problem of the so-called 'unordered set of rules' is resolved, which leads to a more efficient parameter optimization.

The similarity measures described in Eqs. 23~24 will be checked for each fuzzy set, and only the ones with the maximum similarity values will be deleted or merged during each iteration step. For this reason and because of the elitism which records any non-dominated solution found at each iteration step during the experiments these parameters were found not to be critical to the final outcome as long as they are kept within the recommended ranges.

References

1. B. Kosko, Fuzzy Systems as Universal Approximators, *IEEE Transactions on Computers*, vol. 43(11), pp. 1329-1333, 1994.

2. K. M. Passino, S. Yurkovich, *Fuzzy Control*, MA: Addison-Wesley, 1998, pp. 246-252.

3. J. H. Holland, *Adaptation in Natural and Artificial Systems*, MI: The University of Michigan Press, 1975.

4. J. D. Farmer, N. H. Packard, The Immune System, Adaptation, and Machine Learning, *Physica*, vol. 22D, pp. 187-204, 1986.

5. S. Guillaume, Designing Fuzzy Inference Systems from Data: An Interpretability-Oriented Review, *IEEE Transactions on Fuzzy Systems*, vol. 9(3), pp. 426-443, 2001.

6. M. Delgado, F. Gómez-Skarmeta Antonio, F. Martin, A Fuzzy Clustering-Based Rapid Prototyping for Fuzzy Rule-Based Modelling, *IEEE Transactions on Fuzzy Systems*, vol. 5(2), pp. 223-233, 1997.

7. C. L. Karr, Genetic Algorithms for Fuzzy Controllers, *AI Expert*, vol. 6 (2), pp. 26-33, 1991.

8. J. Chen, M. Mahfouf, Interpretable Fuzzy Modeling using Multi-Objective Immune Inspired Optimisation Algorithms, *FUZZ-IEEE 2010*, 2010.

9. J. Chen, *Biologically Inspired Optimisation Algorithms for Transparent Knowledge Extraction Allied to Engineering Materials Processing*, The University of Sheffield, *Ph.D. Thesis*, 2009.

10. J. Chen, M. Mahfouf, A Population Adaptive Based Immune Algorithm for Solving Multi-objective Optimisation Problems, in *H. Bersini & J. Carneiro (Eds.): ICARIS 2006, LNCS 4163*, pp. 280-293, 2006.

11. J. Chen, M. Mahfouf, Artificial Immune Systems as a Bio-inspired Optimisation Technique and Its Engineering Applications, in *H. W. Mo (Eds.): Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies*, pp. 22-48, 2008.

12. E. H. Mamdani, Applications of Fuzzy Algorithm for Control a Simple Dynamic Plant, *Proc. Inst. Electr. Eng.*, vol. 121 (12), pp. 1585-1588, 1974.

13. T. Takagi, M. Sugeno, Fuzzy Identification of Systems and Its Applications to Modelling and Control, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, pp. 116-132, 1985.

14. J. Casillas, O. Cordon, J. Del Jesus Mara, F. Herrera, Genetic Tuning of Fuzzy Rule Deep Structures for Linguistic Modelling, *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 13-29, 2001.

15. L. Zadeh, Outline of A New Approach to the Analysis of Complex Systems and Decision Processes, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 28-44, 1973.

16. J. V. de Oliveira, Semantic Constraints for Membership Function Optimisation, *IEEE Trans. Syst. Man. Cybern. Part A*, vol. 29 (1), pp. 128-138, 1999.

17. S. M. Zhou, J. Q. Gan, Low-level Interpretability and High-level Interpretability: A Unified View of Data-Driven Interpretable Fuzzy System Modeling, *Fuzzy Sets and Systems*, vol. 159, pp. 2091-3131, 2008.

18. J. M. Alonso, L. Magdalena, G. Gonzalez-Rodriguez, Looking for a Good Fuzzy System Interpretability Index:

An Experimental Approach, *Int. J. Approx. Reasoning*, vol. 51, pp. 115-134, 2009.

19. G. A. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *The Psychological Review*, vol. 63(2), pp. 81-97, 1956.

20. F. Herrera, Genetic Fuzzy Systems: Taxonomy, Current Research Treads and Prospects, *Evol. Intel.*, vol. 1 (1), pp. 27-46, 2008.

21. H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms, *IEEE Transactions on Fuzzy Systems*, vol. 3 (3), pp.260-270, 1995.

22. H. Ishibuchi, T. Nakashima, T. Murata, Three-objective Genetics-based Machine Learning for Linguistic Rule Extraction, *Information Sciences*, vol. 136, pp. 109-133, 2001.

23. H. Ishibuchi, T. Yamamoto, Fuzzy Rule Selection by Multi-Objective Genetic Local Search Algorithms and Rule Evaluation Measures in Data Mining, *Fuzzy Sets and Systems*, vol. 141, pp. 59-88, 2004.

24. M. Antonelli, P. Ducange, B. Zazzerini, F. Marcelloni, Learning Concurrently Partition Granularities and Rule Bases of Mamdani Fuzzy Systems in a Multi-objectie Evolutionary Framework, International Journal of Approximate Reasoning, vol. 50 (7), pp. 1066-1080, 2009.

25. M. J. Gacto, R. Alcalá, F. Herrera, Adaptation and Application of Multi-objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-based Systems, *Soft Computing*, vol. 13 (5), pp. 419-436, 2009.

26. M. J. Gacto, R. Alcalá, F. Herrera, Integration of an Index to Preserve the Semantic Interpretability in the Multi-Objective Evolutionary Rule Selection and Tuning of Linguistic Fuzzy Systems. *IEEE Transactions on Fuzzy Systems*, vol. 18(3), pp. 515-531, 2010.

27. M. Setnes, R. Babuška, U. Kaymak, H. Lemke, Similarity Measures in Fuzzy Rule Base Simplification, *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, vol. 28 (3), pp. 376-386, 1998.

28. F. Jiménez, G. Sánchez, A. F. Gómez-Skarmeta, H. Roubos, R. Babuška, Fuzzy Modeling with Multi-Objective Neuro-Evolutionary Algorithms, *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, 2002.

29. Y. Jin, W. Von Seelen, B. Sendhoff, On Generating FC^3 Fuzzy Rule Systems From Data Using Evolution Strategies, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 29 (6), pp. 829-845, 1999.

30. H. L. Wang, S. Kwong, Y. C. Jin, W. Wei, K. F. Man, Multi-objective Hierarchical Genetic Algorithm for Interpretable Fuzzy Rule-based Knowledge Extraction, *Fuzzy Sets and Systems*, Vol. 149 (1), pp. 149-186, 2005.

31. J. González, I. Rojas, H. Pomares, L. J. Herrera, A. Guill *é* n, J. M. Palomares, F. Rojas, Improving the Accuracy While Preserving the Interpretability of Fuzzy

Function Approximators by means of Multi-objective Evolutionary Algorithms, *International Journal of Approximate Reasoning*, vol. 44(1), pp. 32-44, 2007.

32. R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, A Multiobjective Evolutionary Approach to Concurrently Learn Rule and Data Bases of Linguistic Fuzzy-Rule-Based Systems, IEEE Transactions on Fuzzy Systems, vol. 17(5), pp. 1106-1122, 2009.

33. M. Setnes, H. Roubos, GA-Fuzzy Modeling and Classification: Complexity and Performance, *IEEE Transactions on Fuzzy Systems*, vol. 8 (5), pp. 509-522, 2000.

34. H. Roubos, M. Setnes, Compact and Transparent Fuzzy Models and Classifiers Through Iterative Complexity Reduction, *IEEE Transactions on Fuzzy Systems*, vol. 9 (4), pp. 516-524, 2001.

35. M. Y. Chen, D. A. Linkens, A Systematic Neuro-Fuzzy Modeling Framework With Application to Material Property Prediction, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 31 (5), pp.781-790, 2001.

36. F. Jiménez, A. F. Gómez-Skarmeta, H. Roubos, R. Babuška, Accurate, Transparent, and Compact Fuzzy Models for Function Approximation and Dynamic Modeling through Multi-Objective Evolutionary Optimisation, in *E. Zitzler et al. (Eds.): EMO 2001, LNCS* 1993, pp. 653-667, 2001.

37. M. Cococcioni, P. Ducange, B. Lazzerini, F. Marcelloni, A Pareto-based Multi-objective Evolutionary Approach to the Identification of Mamdani Fuzzy Systems, *Soft Computing*, vol. 11, pp. 1013-1031, 2007.

38. R. Acalá, M. J. Gacto, F. Herrera, A Multi-Objective Genetic Algorithm for Tuning and Rule Selection to Obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15 (5), pp. 539-557, 2007.

39. Q. Zhang, *Nature-Inspired Multi-Objective Optimisation and Transparent Knowledge Discovery via Hierarchical Fuzzy Modelling*, Ph.D. Thesis, Department of Automatic Control and Systems Engineering, The University of Sheffield, U.K, 2009.

40. L. Magdalena, Crossing Unordered Sets of Rules in Evolutionary Fuzzy Controllers, *International Journal of Intelligent Systems*, vol. 13 (10/11), pp. 993-1010, 1998.

41. M. G. Cooper, J. J. Vidal, Genetic Design of Fuzzy Controllers: The Cart and Jointed-Pole Problem, *in Proceedings of the Third IEEE Conference on Fuzzy Systems*, vol. 2, pp. 1332-1337, 1994.

42. Y. Jin, Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement, *IEEE Transactions on Fuzzy Systems*, vol. 8 (2), pp. 212-221, 2000.

43. M. Sugeno, T. Yasukawa, A Fuzzy-Logic-Based Approach to Qualitative Modeling, *IEEE Transactions on Fuzzy Systems*, vol. 1 (1), pp. 7-31, 1993.

44. Y. H. Lin, III G. A. Cunningham, S. V. Coggeshall, Using Fuzzy Partitions to Create Fuzzy Systems from

Input-Output Data and Set the Initial Weights in a Fuzzy Neural Network, *IEEE Transactions on Fuzzy Systems*, vol. 5 (4), pp. 614-621, 1997.

45. M. Y. Chen, D. A. Linkens, Rule-base Self-generation and Simplification for Data-driven Fuzzy Models, *Fuzzy Sets and Systems*, vol. 142, pp. 243-265, 2004.
46. J. Tenner, *Optimisation of the Heat Treatment of Steel using Neural Networks*, Ph.D. Thesis, Department of Automatic Control and Systems Engineering, University of Sheffield, U.K, 1999.

Appendix A
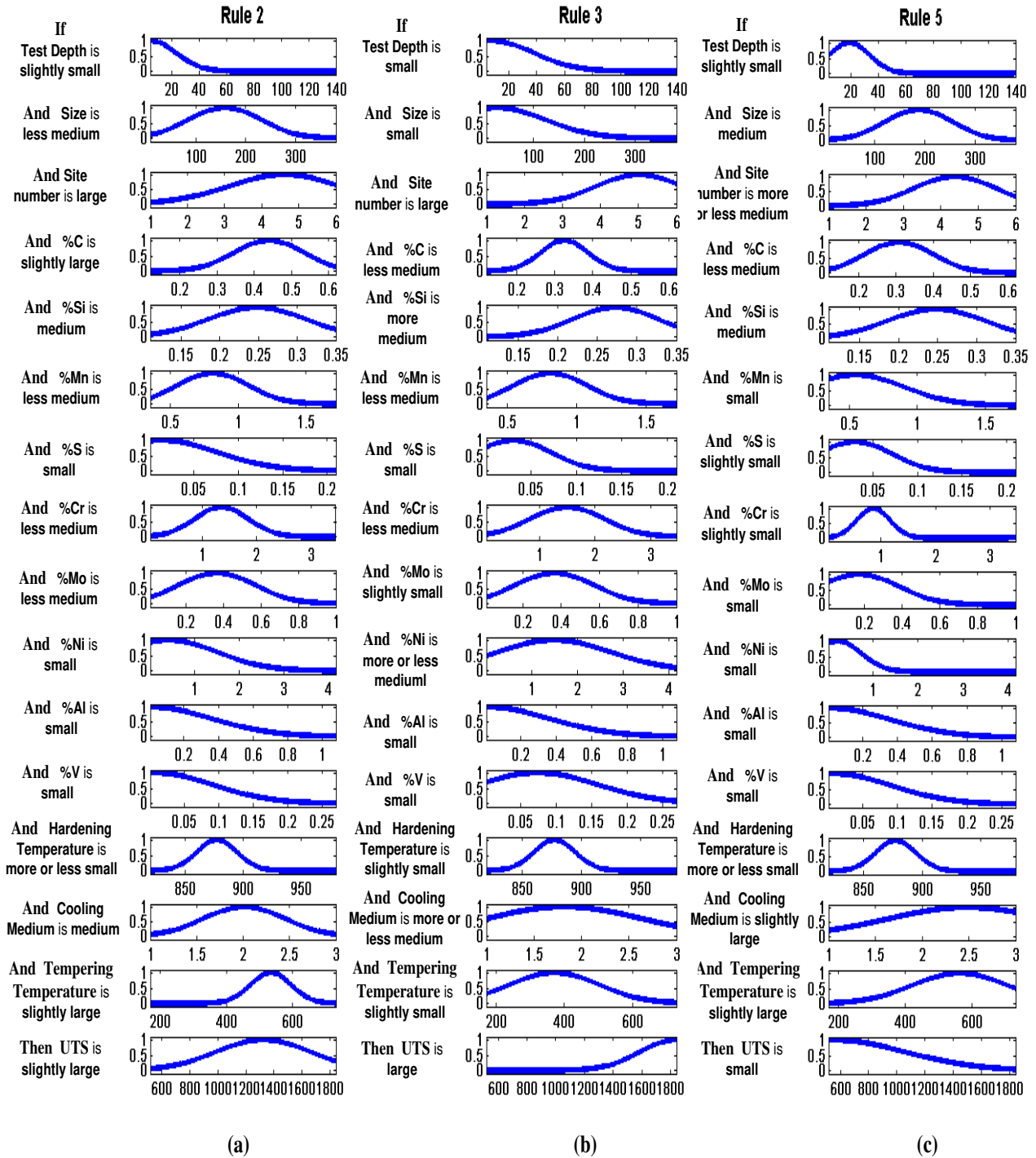
Appendix A: 3 selected rules from a 7-rule simplified Mamdani FRBS