Interpretable Fuzzy Modeling using Multi-Objective Immune-Inspired Optimization Algorithms

Jun Chen, Mahdi Mahfouf

***Abstract*—In this paper, an immune inspired multi-objective fuzzy modeling (IMOFM) mechanism is proposed specifically for high-dimensional regression problems. For such problems, high predictive accuracy is often the paramount requirement. With such a requirement in mind, however, one should also put considerable efforts in making the elicited model as interpretable as possible, which leads to a difficult optimization problem. The proposed modeling approach adopts a multi-stage modeling procedure and a variable length coding scheme to account for the enlarged search space due to the simultaneous optimization of the rule-base structure and its associated parameters. IMOFM can account for both Singleton and Mamdani Fuzzy Rule-Based Systems (FRBS) due to the carefully chosen output membership functions, the inference and the defuzzification methods. The proposed algorithm has been compared with other representatives using a simple benchmark problem, and has also been applied to a high-dimensional problem which models mechanical properties of hot rolled steels. Results confirm that IMOFM can elicit accurate and yet transparent FRBSs from quantitative data.**

I. 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 the 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 the 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 facts of not only being able to approximate practically any given function to an arbitrary accuracy [1], but also being able to generalize reasonably well for any previously 'unseen' situations. The prevalence of these nonlinear regression techniques is largely attributed to the breakthrough in the nonlinear optimization techniques, such as the Back-Error Propagation (BEP) algorithm [2, p. 246-252] and the bio-inspired optimization [3]-[4].

Since the first introduction of 'fuzzy logic', FRBSs have been widely used in systems and control engineering [2]. However, the predominant approach in the traditional design of FRBS highly relies on human experts. Although learning components can be further incorporated into the procedures of coarse model inducement [5] and its further refinement [2, p. 246-252], it may suffer from two serious problems, viz. the deterioration of the model's interpretability and the over-fitting to the training patterns. Taking this into account, one can find that bio-inspired optimization, in particular Generic Algorithms (GAs), has a long history of being incorporated into fuzzy logic [6] and demonstrate a possible route to the remedy for the mentioned two 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 organized as follows: Section II discusses the formation of the multi-objective fuzzy modeling problems and the FRBSs used in this work; Section III shortly reviews the existing evolutionary based approaches for improving FRBS's interpretability; IMOFM which represents an alternative tactic to solve interpretability issues will be introduced in Section IV; in Section V, in order to evaluate the performances of IMOFM, the algorithm is tested using a simple benchmark problem and is applied to the prediction of the mechanical properties of alloy steels; finally, conclusions are given in Section VI.

II. Fuzzy Rule Based Systems and the Formation of the Multi-Objective Fuzzy Modeling Problem

A. Fuzzy Rule Based Systems (FRBS)

Two fuzzy modeling paradigms, viz. Singleton [7] and Mamdani [8] FRBSs, are employed in this work due to their abilities of expressing linguistic meanings in both of their antecedents and consequents. Generally speaking, 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 *i*th linguistic value (fuzzy set) for the *j*th

Manuscript received January 31, 2010. This work was supported in part by the EPSRC under Grant EP/F023464/1.

Jun Chen is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK (phone: +44 (0)114 222 5613; e-mail: jun.chen@shef.ac.uk).

Mahdi Mahfouf is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK (phone: +44 (0)114 222 5607; fax: +44 (0)114 222 5624; e-mail: m.mahfouf@shef.ac.uk).

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 *i*th rule in the rule base, and y_i is the output of the *i*th rule. Typically, Z_i can be the function of the inputs or the linguistic value of the output, which differentiate FRBS into Singleton (the former) and Mamdani (the latter) FRBS. It is worth noting that Singleton FRBS is a special case of TSK FRBS [7] when Z_i is the zero order function of the inputs. In some sense, Singleton FRBS shares the basic feature of Mamdani FRBS if one considers singleton consequents as a special type of fuzzy sets.

B. Formation of the Multi-objective Fuzzy Modeling Problem

As Casillas *et al.* pointed out in [9], 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 comprehensible way, in other words 'the model interpretability'. However, as Zadeh conjectured in his Principle of Incompatibility [10], it is very likely that accuracy and interpretability may well be exclusive requirements in a modeling process. The reflection of these in a fuzzy modeling scenario represents a dilemma of designing FRBS.

The 'accuracy vs. interpretability' issue in a fuzzy modeling context can also be formulated as a multi-objective optimization problem. Fig. 1 shows the Pareto front in a bi-objective fuzzy modeling scenario where two competing objectives, viz. the predictive error (accuracy) and the rule base complexity (interpretability), are minimized simultaneously. The aim is to find a set of 'approximate Pareto FRBSs' as close to the true Pareto front as possible.

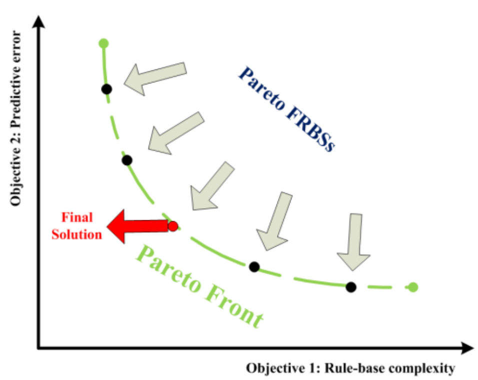

Fig. 1. Pareto front in a bi-objective fuzzy modeling case.

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 finally selected and applied. In the case shown in Fig. 1, if one requires certain interpretability (transparency) of the FRBS along with its good predictive accuracy the middle circle could be the one that fulfils the user's need. As already stated in [11], this should result in a 'minimal' human intervention during the modeling process.

III. Literature Review of Previous Works

Originated from Karr's work [6], 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 [12]. Such a flexibility offers a solution so that one can take into account the interpretability (structure) and the predictive performance of the FRBS in a more coherent way. Broadly speaking, there currently exist two different EA-based streams to tackle the interpretability issues:

1) The first stream is mainly concerned with the linguistic modeling using a Mamdani type of FRBSs, in which a set of pre-specified fuzzy partitions (linguistic terms) are given a priori by experts or users (grid partition). These linguistic terms are fixed during the course of the evolution [13]-[15] so that their physical meanings are retained. Only the fuzzy rules are subject to the 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. Further relevant researches include those in [16]-[17], 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.

2) 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 procedures; hence, there is not a global data-base a priori); 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 [18]-[21]. 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.

Comparing the two streams leads to the following: 1) in the linguistic modeling stream, the target problems are normally associated with classifications and low-dimensional regression problems; this is because that, for such problems, the effect of the 'curse of dimensionality' due to the grid partition and the need for the parameter tuning due to the predictive accuracy requirement are not serious issues; only very recently, such a linguistic modeling framework has been adopted for high-dimensional regression problems [17]; 2) for high-dimensional regression problems, an approximate FRBS may represent a better choice to start with due to the accuracy and compactness requirement. However, to the best of our knowledge, majority of the works within the second modeling stream were using TSK FRBS, which breach the original intention of eliciting an interpretable 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 TSK FRBS is employed the transparency in the consequents will be lost.

In the light of the above considerations, the proposed algorithm 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 which use similar types of FRBS within 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 [17] due to the fact that more compact and higher accurate FRBSs can be elicited.

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

An immune inspired multi-objective fuzzy modeling (IMOFM) procedure consists of three stages which follow principles of vaccination and the secondary response of the immune systems. Furthermore, the third modeling stage utilizes a Population Adaptive Immune Algorithm (PAIA) [22], [23] as the search engine in search of the optimal structure and parameters. In the following space, Artificial Immune Systems (AIS), in particular PAIA, are first introduced followed by the description of IMOFM.

A. *AIS for Multi-objective Optimization*

The basic idea of using AIS for optimization is emanated from the Clonal Selection Principle and Network Hypothesis [4]. The Clonal Selection Principle describes the basic features of an immune response to an antigenic stimulus, and establishes the idea that only those antibodies that recognize the antigen are selected to proliferate. The analogies of this in AIS are composed of the followings: 1) *Activation* calculates the affinity (fitness) for each antibody (solution) so that an adaptive number of clones can be selected and produced; 2) *Affinity Maturation* mutates the selected good clones so that more search space can be explored; 3) *Reselection* selects good candidate solutions from the combined parents and clones to provide selection pressure to effectively drive the candidate solutions towards the Pareto front. The Network Hypothesis states that antibodies can be stimulated by recognizing other antibodies, and for the same reason can be suppressed by being recognized. Such a suppression mechanism allows the regulation of the over-stimulated antibodies to maintain a stable memory. The reflection of this in AIS is the so-called *Network Suppression* which is used to regulate the population so that it is adaptive to the search process. By synergizing all the above components, PAIA is proposed in [22], [23]. In [24], the authors further proposed a multi-stage optimization procedure by incorporating the concept of vaccination. The idea is to first use a single objective optimization stage, acting as the vaccination process, to efficiently find any one of the solutions resting on the Pareto front. Then, PAIA can act as a post-processing procedure to expand such a solution along the Pareto front. For fuzzy modeling, small modifications of the *Activation* step are needed and will be discussed accordingly in Part B. Details regarding PAIA and multi-stage optimization are referred to [22] and [24, ch. 3].

B. *IMOFM for Obtaining a Set of FRBSs*

IMOFM is a three-stage modeling procedure. The first two stages are equivalent to the vaccination process in the first stage of the multi-stage optimization procedure. The aim is to first extract an initial approximate FRBS and then to refine it in terms of its predictive accuracy. By doing so, an initial 'vaccine model' 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 them 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 'vaccine model' 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 after the *Affinity Maturation* in a bid to find transparent FRBSs. Fig. 2 represents a schematic diagram of such a modeling framework.

1) *Elicitation of the Initial Singleton/Mamdani FRBSs*

Firstly, an evolutionary based *K*-means clustering algorithm [25] 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

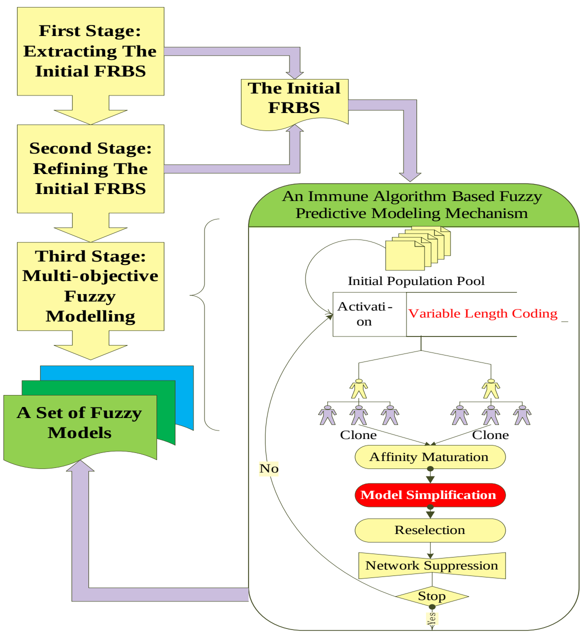

Fig. 2. The proposed IMOFM framework.

established so that $\mu_{A_i^j}(x_j)$ and the corresponding output Z_i (refer to Section II) 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 functions are obtained by first calculating the U matrix as follows:

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

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:

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

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

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

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 (3):

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

For Singleton FRBS, 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:

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

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 $\theta = \left(c_i^y, c_i^j, \sigma_i^j \middle| i = 1,..,k;, j = 1,..,n\right)$ is the parameter vector which is subject to further tuning in the second modeling stage. For Mamdani FRBS, the bell-shape membership functions are used for Z_i:

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

where, σ_i^y is obtained by using (1) and (2) but in the output space. Unlike traditional Mamdani FRBS where defuzzification is normally applied on the overall implied fuzzy set [2, p. 64], IMOFM employs the center of gravity (COG) defuzzificaiton on the implied fuzzy set as below:

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

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

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

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 (8).

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

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

2) *Refinement of Initial FRBSs*

In the second modeling stage, a BEP with momentum terms algorithm [2, p. 246-252] is developed to first improve the accuracy of the initial FRBS by adjusting the parameters in θ so that the rule-base with the over-estimated number of rules can also survive under the pressure of 'Pareto' selection. By taking the partial derivatives of (4) and (7) with respect to each parameter included in θ, one can obtain a set of parameter updating laws (due to the space, interested readers are referred to [24]). One problem associated with the BEP updating formulas is that they include no constraints with respect to the update mechanism of these parameters. Hence, during the course of the optimization, the centers of the membership functions are likely to be placed outside the boundaries. Hence, in this work a simple constraint handling scheme is added, which checks the boundary violation for centers during each iteration step and drives any violated centers back to the boundaries.

3) Multi-objective Fuzzy Modeling

a) 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 (9); $y_{prediction\_m}$ 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).

$$Objective1: RMSE = \sqrt{\frac{\sum_{m=1}^{t}\left(y_{prediction\_m} - y_{real\_m}\right)^2}{t}} \qquad (9)$$
$$Objective2: Complexity = Nrule + Nset + RL$$

b) Variable Length Coding Scheme and Initial Population

As far as the multi-objective fuzzy modeling is concerned, different encoding schemes have been proposed 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 [13]-[15], in which a string or a rule matrix is formed as the chromosome in order to select the effective rules and linguistic terms from the candidate set. [17] represents the variant of the first encoding scheme, in which the encoding comprised the structure coding and the parameter (data-base) coding. Key to the first encoding scheme is that the global data-base is usually kept unchanged or only varied in a constrained search space. The latter is mainly found in the approximate modeling stream due to the lack of global data-base. Since only the effective rule parameters are included in the coding, a variable length coding scheme is inevitable and is used in this work. Fig. 3 gives an example of how to encode FRBSs. As shown in Fig. 3, the increase of the code length is only linear to the variable's dimension, which effectively tackles the efficiency of the search and the curse of dimensionality.

The initial population is obtained with all individuals generated around the 'vaccine model' using (10) and (11). Where, ${C_{vaccine}}_i^j$ and ${\sigma_{vaccine}}_i^j$ are the centre and spread of the *i*th rule and the *j*th input membership function in the

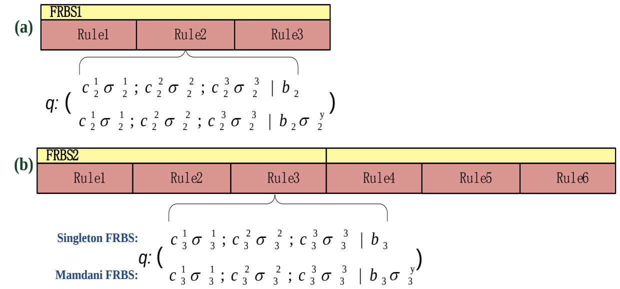

Fig. 3. Variable length coding scheme for: (a) a three-rule Singleton/Mamdani FRBS; (b) a six-rule Singleton/Mamdani FRBS.

original vaccine FRBS extracted from the first two modelling stages. ${C_{vaccine}}_i^y$ and ${\sigma_{vaccine}}_i^y$ are the centre and the spread of the *i*th rule's consequent. $randn$ is a random number within [0, 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. α and β are the user specified parameters which define how much different the newly generated FRBSs are from the original vaccine one, and are set to 0.2 and 0.1 respectively.

$$\begin{gathered} {C_{initial}}_i^j = \alpha \cdot range^j \cdot randn + {C_{vaccine}}_i^j \\ {\sigma_{initial}}_i^j = \beta \cdot randn + {\sigma_{vaccine}}_i^j \\ {C_{initial}}_i^y = \alpha \cdot range^y \cdot randn + {C_{vaccine}}_i^y \\ {\sigma_{initial}}_i^y = \beta \cdot randn + {\sigma_{vaccine}}_i^y \end{gathered} \qquad (10)$$

$$range = \min(|C_{vaccine} - U_{limit}|, |C_{vaccine} - L_{limit}|) \qquad (11)$$

Such a 'forming' approach only acquires the knowledge about the maximum allowable number of rules (i.e. the pre-specified number of clusters) and the data so that emphasis of the third modeling stage is placed on the automatic elicitation of a set of FRBSs in the 'Pareto' sense. The size of the initial population in this work is set to 7.

c) Variation Operators and a New Distance Index

The variation operator used in PAIA is *Affinity Maturation* which mutates the selected good solutions according to their affinity values (aff_val) as described in (12). Such a variation operator is used in IMOFM to optimize the associated parameters encoded in *q* (see Fig. 3).

$$q_{new}(k) = q_{old}(k) + \alpha \cdot N(0,1)\ k = 1, \ldots, t; \qquad \alpha = \frac{exp(aff\_val)}{exp(1)} \qquad (12)$$

where *N(0, 1)* is a Gaussian random variable with zero mean and standard deviation 1. k is a dimension index within the length of q that has been chosen to mutate.

One problem in using (12) is associated with aff_val which is originally calculated based on the distance 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' [26] may occur. As pointed in [26], variations over 'unordered' individuals are equivalent to combing the mother's gen for good vision and father's gen for curly hair, which does not make much sense. To tackle this problem, a new distance index is proposed for the calculation of aff_val. The basic idea is to align the closest rules from different individuals in order to have a meaningful variation.

$$dist(R_j, R_k) = \frac{\sum_{i1=1}^{k1}\sum_{l=1}^{rl} abs\left(R_j^{i1}(l) - R_k^{C_{i1}}(l)\right) + \sum_{i2=1}^{k2}\sum_{l=1}^{rl} abs\left(R_k^{i2}(l) - R_j^{C_{i2}}(l)\right)}{rl\cdot(k1+k2)} \tag{13}$$

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 $*$.

d) Model Simplification

A model simplification step is added as shown in Fig. 2. The aim is to remove the redundancy both in the rules and in the fuzzy sets. On the rule level: 1) one of the insignificant rules (rules that contribute the least to any prediction error increase when not include these rules) is removed unless the rule base reaches the fewest rules designated by the user; 2) one of singleton rules [21] (rules whose comprising fuzzy sets are similar to singleton set) is removed; 3) the most similar pair of rules based on the Similarity of Rule Premise (SRP) [20] are merged. On the fuzzy sets level: 1) one fuzzy set that is the most similar to the universal fuzzy set is labeled as 'Don't care'; 2) two most similar fuzzy sets from the inputs and output dimensions are merged to form a single fuzzy set based on the similarity measure $S(A_i^j, A_l^j)$ [20]. It is worth mentioning that the above operations are executed for each cloned FRBS at each iteration step. A set of thresholds which control the various similarity measures are specified by the users. Interested readers are referred to [24, ch.5] for more details. During the experiments, it is found that the thresholds are not the critical parameters due to the following two reasons: 1) only one fuzzy rule or two fuzzy sets are removed or merged at each iteration step; 2) elitism is adopted to record any non-dominated solution found at each iteration step.

V. Experiments

A. 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 [27]. The system is defined as follows:

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, \qquad 1 \le x_1, x_2 \le 5 \tag{14}$$

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 [5] and [27] 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 using (10) and (11). The initial population size is set to 7. The number of iterations in the third stage is set to 1200. Table I summarizes such comparative results focusing on their predictive performances (RMSE) and the number of rules. The results in Table I include the average values of 30 runs. Fig. 4 shows the membership function distribution of input2 (x_2) through the three modeling stages.

TABLE I
Comparisons Of The Predictive Performance Of The Different Modeling Methods For The Benchmark Problem

Modeling Methods (Ref.)	No. of rules	No. of fuzzy sets&	The type of FRBS	Performance (RMSE training)	
[5]	6	12	Mamdani	0.5639*	0.2811@
[27]	5	10	Singleton	0.5604*	0.3391@
IMOFM_S (IMOFM for Singleton FRBS):					
Initial FRBS	5	10	Singleton	0.5954*	0.0688@
FRBS1(30 times)	5	9	Singleton	0.0696#	σ^2: 0
FRBS2(30 times)	5	8	Singleton	0.0875#	σ^2: 0.004
FRBS3(29 times)	4	8	Singleton	0.0930#	σ^2: 0.010
FRBS4(29 times)	3	6	Singleton	0.1417#	σ^2: 0.005
FRBS5(25 times)	2(5T)	4	Singleton	0.4769#	σ^2: 0.072
IMOFM_M (IMOFM for Mamdani FRBS):					
Initial FRBS	5	15	Mamdani	0.6078*	0.0702@
FRBS1(25 times)	5	14	Mamdani	0.0651#	σ^2: 0.002
FRBS2(22 times)	5	13	Mamdani	0.0691#	σ^2: 0.002
FRBS3(26 times)	4	11	Mamdani	0.0781#	σ^2: 0.003
FRBS4(28 times)	3	9	Mamdani	0.1311#	σ^2: 0.015
FRBS5(28 times)	2(5T)	5	Mamdani	0.2718#	σ^2: 0.062

&: 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.

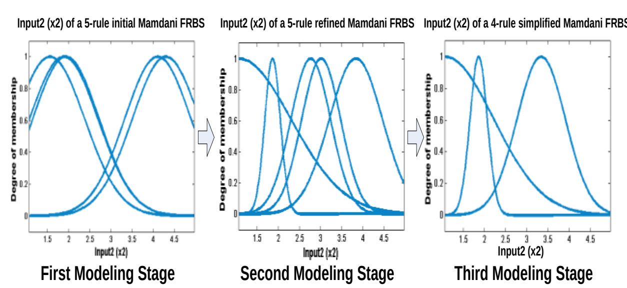

Fig. 4. The membership function distribution of input2.

Since different runs will lead to slightly different FRBS configurations, Table I also records the number of each FRBS' configuration found within the 30 runs. Most configurations are found more than 20 times within 30 runs, which suggests that IMOFM is robust and consistent. The Comparing to other Singleton and Mamdani modeling approaches, IMOFM was found to represent the most accurate results with simpler rule-base structure.

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 latter case, the initial 5-rule FRBS is randomly generated within the variable domains. Table II summarizes the results of the two variants. It is worth mentioning that for the two variants, the thresholds for the model simplification are set at higher values so that the merging operation only happens when the fuzzy sets or rules are 'very' similar. This is to ensure that the FRBSs with more rules are given a better chance of surviving in the early stages of the evolution. In [19], a 'niche' concept is used to maintain a set of FRBSs with the same number of rules. A substitution only happens within each niche so that one can evolve a set of FRBSs with the different number of rules without the worry of losing individuals with more rules. The proposed three-stage procedure does not need the aforementioned 'niche' concept if all the stages work as a unified procedure. In such a case, the most accurate FRBS is always the one with the number of rules close to the maximum value. More importantly, this accurate FRBS will direct the search from the most complex structure (the more accurate one) to the simplest ones (the less accurate ones). This ensures the coexistence of FRBSs with various complexities during the 'Pareto' selection.

TABLE II
COMPARISONS OF THE PREDICTIVE PERFORMANCE OF THE 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.

As shown in Table II, 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 I), 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 for the reasons already described. All these justified the inclusion of the first two stages.

B. *Real World Applications*

In this section, the problem of predicting the Ultimate Tensile Strength (UTS) of heat-treated steel is studied, which features a high dimensional, nonlinear and sparse data space. The UTS data set consists of 3760 data samples and includes 15 inputs and 1 output which is the UTS with the values between $516.2N/mm^2$ and $1842N/mm^2$. In order to compare with the work in [28], the 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 unseen examples are also included. All the parameter settings related to IMOFM are the same as those used for the benchmark problem except that the initial number of rules is set to 12. The results presented in Table III include the average values of 10 independent runs and only a few 'Pareto' FRBS are presented due to the constraints on space.

TABLE III
COMPARISONS OF THE PREDICTIVE PERFORMANCE FOR THE DIFFERENT MODELING METHODS USING THE UTS DATA

Modeling Methods	First Stage (clustering algorithm)		Second Stage (single objective refining)		
	Training (RMSE)	Testing (RMSE)	Training (RMSE)	Testing (RMSE)	Validation (RMSE)
[28]	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)	**Modeling performance** Testing/ Validation
[28]				
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

As shown in Table III, the problem of over-fitting specifically related to the second modeling stage (vaccine FRBS) under unseen situations is revealed in Table III. 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. 5 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. Table IV summarizes the results of the UTS modeling problem using IMOFM 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’.

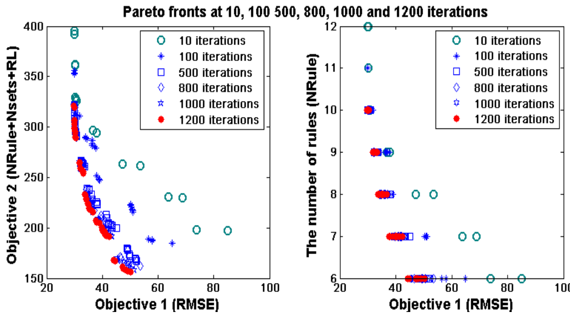

Fig. 5. The Pareto FRBSs at different iterations.

TABLE IV
THE COMPARISON OF THE MODELING APPROACHES WITH AND WITHOUT VARIABLE LENGTH CODING SCHEME

FRBS Configurations	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%

VI. CONCLUSION

The main novelty of IMOFM is considered to be as follows: 1) the initial number of rules in the rule-base is not an important factor anymore since in the third stage a set of Pareto FRBS with different structure are elicited; only the maximum allowable number of rules is required *a priori*; 2) due to the vaccination process, the efficiency and predictive accuracy of the modelling are improved; 3) by using the variable length coding scheme, the problem of the ‘unordered set of rules’ is resolved, which leads to a more efficient parameter optimisation; 4)the proposed method represents one of the first attempts which uses an approximate Mamdani FRBS for high-dimensional regression problems, and can be viewed as a complement to the linguistic Mamdani fuzzy modeling approach.

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] 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.

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

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] J. Chen, *Biologically Inspired Optimisation Algorithms for Transparent Knowledge Extraction Allied to Engineering Materials Processing*, The University of Sheffield, *Ph.D. Thesis*, 2009, ch. 3.

[25] J. Chen, M. Mahfouf, “An Immune Algorithm Based Fuzzy Predictive Modeling Mechanism using Variable Length Coding and Multi-objective Optimisation Allied to Engineering Materials Processing”, in *proceedings of the 2008 IEEE International Conference on Granule Computation (GrC2008)*, pp.26-28, 2008.

[26] 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.

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

[28] Q. Zhang, *Nature-Inspired Multi-Objective Optimisation and Transparent Knowledge Discovery via Hierarchical Fuzzy Modelling*, *Ph.D. Thesis*, The University of Sheffield, U.K, 2009, ch.6.