

An Immune Algorithm Based Fuzzy Predictive Modeling Mechanism using Variable Length Coding and Multi-objective Optimization Allied to Engineering Materials Processing

Jun Chen and Mahdi Mahfouf

*The University of Sheffield, UK, Department of Automatic Control and Systems Engineering
Institute for Microstructural and Mechanical Process Engineering:*

The University of Sheffield (IMMPETUS)

{jun.chen, m.mahfouf}@sheffield.ac.uk

Abstract

In this paper, a systematic multi-objective fuzzy modeling approach is proposed, which can be regarded as a three-stage modeling procedure. In the first stage, an evolutionary based clustering algorithm is developed to extract an initial fuzzy rule base from the data. Based on this model, a back-propagation algorithm with momentum terms is used to refine the initial fuzzy model. The refined model is then used to seed the initial population of an immune inspired multi-objective optimization algorithm in the third stage to obtain a set of fuzzy models with improved transparency. To tackle the problem of simultaneously optimizing the structure and parameters, a variable length coding scheme is adopted to improve the efficiency of the search. The proposed modeling approach is applied to a real data set from the steel industry. Results show that the proposed approach is capable of eliciting not only accurate but also transparent fuzzy models.

1. Introduction

Fuzzy Rule-Based Systems (FRBS) have the additional ability to integrate human expertise in the form of vague or imprecise statements rather than crisp mathematics, for many real-world systems' knowledge can only be described by experts with nature language. Depending on what degree to which such expertise is involved, fuzzy modeling can belong to white box, black box or grey box modeling. Previous researches on fuzzy modeling are mainly concerned with how to synthesis a rule-base with domain-dependent knowledge from human experts, such as operators, and hence render the task of tuning the parameters associated with the antecedent and consequent parts as an optimization problem, e.g. recursive least-squares or gradient-based method. Without the tuning

process following the synthesis step, the above approach is indeed equivalent to white-box modeling and the built model can be regarded as descriptive FRBS [1], which gives rise to four limitations: 1) often, expert knowledge is not available or is limited ; 2) it is very hard to handle problems with a significant amount of data to be processed and analysed; 3) it suffers from the curse of dimensionality; 4) the way to design such a fuzzy system is not domain-independent and thus no systematic design procedure can be followed. In all these cases, pure knowledge extraction from experts fails to provide a satisfactory solution. However, discovering knowledge from data can help in overcoming the aforementioned limitations by augmenting FRBS with an additional learning ability.

In the past two decades, many successes in the hybridization of FRBS and learning methods have been registered. The most representative one must be the so-called neuro-fuzzy system, which incorporates learning methods usually used in neural networks into FRBS [2]. Almost at the same time, attempts of hybridizing clustering methods with fuzzy systems were carried out and led very promising results [3]. The aim of these types of hybridization is to automatically induce rules from large collections of learning data. Despite the great success using the aforementioned paradigms, the following limitations have been identified: 1) the designer still needs to set the granulation level or the number of clusters; 2) the need to set start points for clustering and neural networks; 3) most importantly, the hence elicited FRBS can only be described as approximate FRBS [1] rather than the descriptive one. The main drawback compared to the descriptive one is the degradation in terms of interpretability of the rule base due to the automatic learning process, which yields overlapped fuzzy sets. Although such approximate FRBS retains some interpretability, it may become more black-box oriented although often its performance is very much improved compared to the descriptive one.

The shift between the descriptive and approximate FRBS represents two extremes of designing fuzzy models. The last two decades have witnessed the popularity of the latter by compromising the interpretability with accuracy, which deviates from the original intention of FRBS. In the light of the above discussion, the main thrust of this paper is to present a systematic way of building FRBS, which preserves some degree of interpretability during the automatic learning process, while at the same time retaining a high level of accuracy. To this aim, section 2 describes a global evolutionary based clustering method, which is developed to create an initial model with the maximum allowable number of rules. Section 3 presents a constrained back-propagation (BP) algorithm with momentum terms, which serves to improve the accuracy of the initial FRBS. An immune inspired multi-objective optimization algorithm is incorporated in the last stage by realizing that the increased interpretability is often a contradictory goal against the objective of the accuracy, which can be found in section 4. To account for the need of simultaneously optimizing rule base structure and membership parameters, a variable length coding scheme is devised and presented in section 4. Finally, the experimental results for predicting the Tensile Strength (TS) of alloy steels are presented in section 5.

2. First stage: an evolutionary based clustering algorithm-G3Kmeans

Clustering is incorporated into fuzzy modeling especially when the numerical data has a very high dimensionality. The purpose is to extract the relationship between independent system variables so that the initial fuzzy structure with only a few rules can be obtained. Among many methods, attempts have been made to obtain clusters by minimizing the within-cluster-distance and maximizing the inter-cluster-distance, which results in points within the identified clusters being as close to each other as possible. K-means clustering [4] and fuzzy C-means [5] are two instances of this family. By understanding the nature of fuzzy modeling and fuzzy partition, it is reckoned that the method based on this idea is more close to the modeling requirement since one wishes to isolate data points into sub-regions so that each one can be described by a localized fuzzy rule.

Three most often used clustering algorithms for fuzzy partition are K-means, fuzzy C-means and subtractive clustering [6]. The first two are found to be very sensitive to the initial points and often get stuck at some local optima. The third one is found to only produce near optimal partitions and is usually used as a method to estimate the number of clusters as *a priori* for other clustering algorithms. Hence, an evolutionary based clustering algorithm-G3Kmeans, representing the hybridization of G3PCX [7] and K-means, is put forward

in this paper to overcome such drawbacks. The purpose of the hybridization is to utilize the global search capability of GA to find a set of cluster centers so that a within-cluster-distance criterion is minimized. By doing so, the sensitivity to the initial settings is avoided in the first place, and more importantly a good global fuzzy partition is extracted, which can ease the optimization in the second stage. Since G3PCX is a real coded GA, it is very efficient in terms of solving this real parameter optimization problem. The detailed steps included in G3Kmeans are as follows.

(1) **Initialization:** The randomly generated k cluster centers are encoded in each chromosome in a concatenated form. P chromosomes are generated in the initial population.

(2) **Assigning data points:** Each data point is assigned to one cluster with the center of C_j using equation (1):

$$X_m \in C_i : \text{if} \left\{ \begin{array}{l} \|X_m - C_i\| < \|X_m - C_l\| \\ m=1,2,\dots,t; i,l=1,2,\dots,k, l \neq i \end{array} \right. \quad (1)$$

Where, $\|$ is the Euclidean norm and t is the number of data samples. After the assignment, cluster centers encoded in the chromosome are updated by calculating the mean value of each cluster.

(3) **Fitness computation:** the fitness value of each individual is calculated using equation (2):

$$\varpi(C_1, C_2, \dots, C_k) = \sum_{l=1}^k \sum_{x_m \in C_l} \|X_m - C_l\|^2 \quad (2)$$

Where, ϖ is a within-cluster-distance metric to be optimized (minimized), and C_1, C_2, \dots, C_k are k cluster centers.

(4) **Parent-Centric Crossover (PCX):** Generate λ offspring from the parents using the Parent-Centric Crossover [9].

(5) **Fitness computation:** the cluster centres and fitness values of the offspring are updated and calculated again as what have been done in the step 2 and 3 accordingly.

(6) **Parents to be replaced:** choose two parents at random from the population P .

(7) **Replacement:** From the combined subpopulation of two chosen parents and λ created offspring, choose the best two solutions and replace the chosen two parents (in step 6) with these solutions.

(8) **Iteration:** the aforementioned steps from step 2 are repeated for a specified generations, and the final solution is the one with the smallest objective value at the end of the execution.

As will be seen in Section 3, the Gaussian membership function is used. In such a case, the identified cluster centres C correspond directly to the centers of Gaussian membership functions. The spread of the Gaussian membership function is obtained by first calculating the U matrix as follows:

$$U(i, m) = \left(\sum_{l=1}^k \frac{\|X_m - C_l\|}{\|X_m - C_l\|} \right)^{-1} \quad (3)$$

Where, U specifies the degrees of data points belonging to each cluster center. Spread σ_i^j is then calculated as follows:

$$\sigma_i^j = \max \left(\sqrt{\frac{-(X_m^j - C_i^j)^2}{2 \cdot \log(U(i, m))}} \right) \quad m = 1, 2, \dots, t \quad (4)$$

Where, j indicates the dimension of the spread for the i th cluster. With centres and spreads obtained from the clustering algorithm, the Gaussian membership function can be specified as follows:

$$A_i^j(x_m^j) = \exp \left(-\frac{1}{2} \cdot \left(\frac{x_m^j - c_i^j}{\sigma_i^j} \right)^2 \right) \quad (5)$$

3. Second stage: refining the initial model with a back-propagation algorithm

After the first stage, a FRBS with the pre-specified number of rules is extracted from the numerical data. This initial fuzzy model is not optimal from two perspectives: 1) the structure of FRBS is not optimal as far as the interpretability is concerned; 2) the membership function parameters need to be tuned further. A constrained back-propagation (BP) algorithm is thus utilized to first improve the accuracy of the initial FRBS so that a 'vaccine model' [8] can be obtained for the further operation in the multi-objective optimization stage.

The well-known problem associated with BP lies in its local search limitation. However, this problem is somehow relaxed by using a global clustering algorithm in the first stage, which enhances its ability of locating good global FRBS in the sense that a within-cluster-distance is globally minimized. If the centre of average defuzzification, production inference rule, singleton fuzzification and Gaussian membership function are employed, a fuzzy system can be represented as follows:

$$y_q^{crisp} = \frac{\sum_{i=1}^k b_i^q \cdot \mu_i}{\sum_{i=1}^k \mu_i} = y_q^{crisp}(X_m | \theta) \quad (6)$$

Where, q is the number of outputs and is set as 1 in this work; k is the number of rules in the rule base; b_i is the centre of the output membership functions; μ_i is the matching degree of the i th rule as described as follows:

$$\mu_i(X_m) = A_1^1(x_m^1) \cdot A_2^2(x_m^2) \cdot \dots \cdot A_n^n(x_m^n) \quad (7)$$

θ is the parameter vector subject to the minimization of the mean square error. Since the convergence properties of BP can sometimes be improved via the additions of 'momentum terms', the following parameter update laws are derived:

$$b_i(t+1) = b_i(t) - \lambda_1 \cdot \varepsilon_m(t) \cdot \frac{i(t)}{\sum_{k=1}^k i(t)} + \beta_1 \cdot \Delta b_i(t-1) \quad (8)$$

$$c_i^j(t+1) = c_i^j(t) - \lambda_2 \cdot \varepsilon_m(t) \cdot \frac{i(t) \cdot q(t)}{\sum_{k=1}^k i(t)} + \beta_2 \cdot \Delta c_i^j(t-1) \quad (9)$$

$$\sigma_i^j(t+1) = \sigma_i^j(t) - \lambda_3 \cdot \varepsilon_m(t) \cdot \frac{i(t) \cdot r(t)}{\sum_{k=1}^k i(t)} + \beta_3 \cdot \Delta \sigma_i^j(t-1) \quad (10)$$

$$\begin{aligned} \varepsilon_m &= y^{crisp}(x_m | \theta) - y_m \\ \text{where, } q &= (b_i - y^{crisp}(X_m | \theta)) \cdot \left(\frac{x_m^j - c_i^j}{(\sigma_i^j)^2} \right) \\ r &= (b_i - y^{crisp}(X_m | \theta)) \cdot \left(\frac{x_m^j - c_i^j}{(\sigma_i^j)^3} \right) \\ \Delta b_i(t-1) &= b_i(t) - b_i(t-1) \end{aligned} \quad (11)$$

As one can see from Equations 8 to 10, there are no constraints for updating these parameters. Hence, during the course of the optimization, centres are likely to be placed outside the boundaries. Although this does not affect the ultimate accuracy of FRBS, it does cause confusion for the users when assigning linguistic labels. Hence, in this work, a constraint handling scheme is added, which checks the boundary violation for centres during each iteration and drive any violated centres back to the boundaries. The step size λ and the gain of momentum term β are all set to 0.035 in this work.

4. Third stage: multi-objective fuzzy modeling

An optimal FRBS could be obtained by optimizing the rule-base structure and membership function parameters either simultaneously or separately. The previous two stages can be viewed as the instances of a separate structure and parameter learning. The drawbacks of the separate learning are as follows: 1) only a sub-optimal result may be obtained since the structure and parameters need to cooperate to provide a satisfactory FRBS; 2) the separate structure learning relies strongly on the human preference. Hence, only problem 2, namely the need to set the start points, as mentioned in the Section 1 would have been solved by using G3Kmeans. One still has to set the initial granulation level and only an approximate FRBS can be elicited.

To improve the interpretability of such an approximate FRBS, authors in [9] performed model simplifications and a fine-tuning operation is required. It is still a separate learning process so that model simplifications rely heavily on the pre-specified thresholds according to the designer's preference. In [10], authors proposed a hierarchical scheme to evolve both parts. However, a rule matrix is required, which would be vulnerable to high dimensional problems due to the exponential increase in the matrix dimension.

The proposed approach in this work utilizes a multi-objective optimization framework and a variable length coding scheme, which does not suffer from the curse of dimensionality. A set of FRBS representing the trade-offs of interpretability and accuracy are obtained through a single run, and only the maximum number of rules is required as *a priori*, which reduces any user intervention during the whole design process to a minimum level.

4.1. A framework of the immune inspired multi-objective fuzzy modeling

Fig. 1. shows the framework of the proposed modeling approach which is based on [8], in which the authors have proved that a multi-stage optimization procedure can reduce the computational load of the whole search process greatly. G3Kmeans and BP in this paper function exactly the same as the first step of that procedure to extract the so-called ‘vaccine model’.

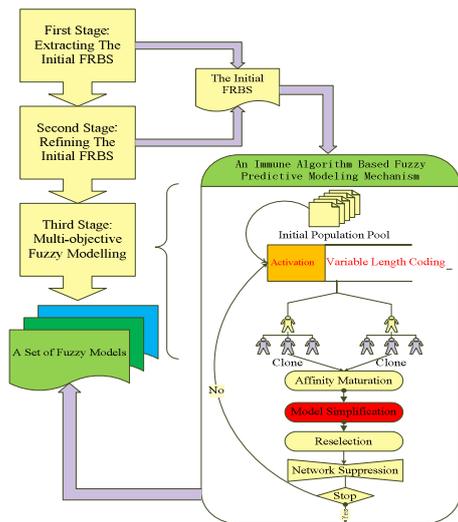


Fig. 1. The framework of the proposed immune based fuzzy predictive modeling methodology

A Population Adaptive based Immune Algorithm (PAIA) [8] is utilized in the third stage to optimize both the structure and parameters simultaneously. Activation calculates the affinity (fitness) for each Antibody (solution) so that an adaptive number of clones can be produced. Affinity maturation mutates the clones so that more search space can be explored. Reselection selects good candidate solutions from the combined parents and clones to provide a selection pressure to effectively drive the candidate solutions towards the Pareto front over many iteration steps; Network suppression is used to regulate the dynamics of the population so that it can adapt to the problem. Although immune algorithms accidentally resemble some characteristics of Genetic Algorithms, a more efficient search could be induced

since the population is adaptive. Variable length coding and model simplifications are added to the original PAIA to account for the structure optimization as will be explained in the next two sub-sections. Further details on PAIA can be found in [8].

4.2. Variable length coding scheme

A variable length coding, which only encodes effective rules, is employed in this work to account for the efficiency of the search and the curse of dimensionality. Since only the parameters of effective rules are encoded, the increase of the code length is only linear to the dimension, which is not the case for the hierarchical coding. Some previous works [1] fixed the length of coding according to the maximum allowable number of rules, and filled the empty places with random values if FRBS has fewer rules as shown in Fig. 2.

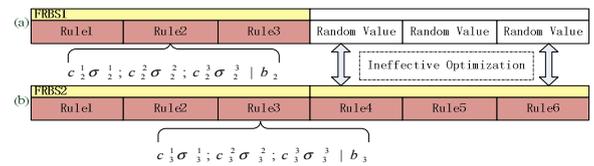


Fig. 2. Example of the variable length coding for a two-input system: (a) three-rule FRBS; (b) six-rule FRBS

Since most heuristic search methods rely on a distance calculation (interaction between individuals) in the phenotypic space, which is the major force to direct the search, an ineffective optimization may be induced because some parts of the long FRBS may interact with the random part of the one with fewer rules. So a method is needed to address the problem of individuals with different lengths. Unconstraint optimization causes another problem as shown in Fig. 3, where FRBS1 and FRBS2 are exactly the same. However, because of the blind search mechanism, values encoded in the Rules 1 and 7 are different within the two FRBSs. In a different instance, rules may be deleted, e.g. Rule7 in FRBS2. Thus, a very large distance is produced as the affinity measure is used in such a case. In PAIA this would lead to a very large mutation, however, only a small or non jump is needed.

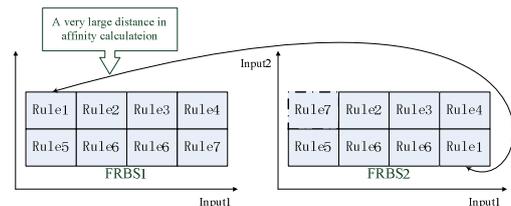


Fig. 3. The problems associated with FRBS having different rule lengths and unconstrained optimization

To tackle the aforementioned problems, a new distance index is proposed to calculate the affinity for PAIA in the activation step. The basic idea is to find the distance of the closest rules in different FRBSs rather than the distance of corresponding rules. Hence, Rule1 in FRBS1 will be paired with Rule7 in FRBS2. The mathematical description of the idea is as follows:

$$dist(R_j, R_k) = \frac{\sum_{i=1}^{k1} \sum_{l=1}^{r_l} (R_j^{i1}(l) - R_k^{C1}(l)) + \sum_{i=1}^{k2} \sum_{l=1}^{r_l} (R_k^{i2}(l) - R_j^{C2}(l))}{rl \cdot (k1 + k2)} \quad (12)$$

Where, R_j, R_k are two FRBS with $k1$ and $k2$ rules; rl is the length of the rule; R_k^{C1} (R_j^{C2}) represent the closest rule in R_k (R_j) with respect to the i lth ($i2$ th) rule in R_j (R_k). The above distance index is used to replace the one in the original PAIA for calculating affinity (refer to [8] for more details about affinity calculation).

4.3. Model simplification

A model simplification step is added to PAIA. The aim is to remove the redundancy both in the rules and in the fuzzy sets so that one can pursue the FRBS structure optimization along with the accuracy at the same time.

(I) Rule pruning: three scenarios can be regarded as having redundant rules and thus need to be pruned.

i. *Insignificant rules:* inspired by the idea behind neural network pruning, insignificant rules are the ones that contribute the least to any prediction error increase when not including this rule. This occurs because other rules may have already covered the area under these rules. Insignificant rules are deleted when the following condition is met:

$$(cr / \max r) \cdot rnd > p_{in} \quad (13)$$

Where, cr is the number of rules in the current FRBS; $\max r$ is the maximum allowable number of rules; rnd is a random number between $[0, 1]$. P_{in} is a design parameter which limits the fewest rules in FRBS and is 0.5 in this work.

ii. *Redundant rules:* rules that only cover a small amount of data are regarded as redundant rules since they may not be fired in most cases. They are deleted subject to the following condition:

$$\sum_{j=1}^n \sigma_i^j / n < th_rdr = rnd \cdot rdr \quad (14)$$

Where, n is the input dimension; th_rdr is a design parameter which randomly changes between $[0, rdr]$ every t generations and rdr is 0.01 in this work.

iii. *Merging similar rules:* during the optimization rules may have similar fuzzy sets in the antecedent. Those rules should be merged together by taking the mean values of those fuzzy sets to keep FRBS consistent and parsimonious. The following condition should be met:

$$\min \left\{ S(A_i^j, A_l^j), \quad j=1, \dots, n \right. \\ \left. i, l=1, \dots, k; i \neq l \right\} > th_mr = rnd \cdot (1 - mr) + mr \quad (15)$$

Where, $S(A_i^j, A_l^j)$ 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 generations and mr is 0.9 in this work.

(II) Redundant fuzzy sets: one can encounter two situations.

i. *Universal fuzzy sets:* fuzzy sets which meet the following condition are regarded as universal fuzzy sets and thus are deleted:

$$S(A_i^j, U) > th_ufs = rnd \cdot (1 - ufs) + ufs \quad (16)$$

Where, U is the universal fuzzy set; th_ufs is the threshold which randomly changes between $[ufs, 1]$ every t generations and ufs is 0.65 in this work.

ii. *Merging similar fuzzy sets:* two fuzzy sets are considered to be similar if following condition is met:

$$S(A_i^j, A_l^j) > th_sfs = rnd \cdot (1 - sfs) + sfs \\ S(A_i^j, A_l^j) = \frac{1}{1 + \sqrt{(c_i^j - c_l^j)^2 + (\sigma_i^j - \sigma_l^j)^2}} \quad (17)$$

Where, th_sfs is the threshold which randomly changes between $[sfs, 1]$ every t generations and sfs is 0.9 in this work. Means values of two similar fuzzy sets are calculated to substitute the original two fuzzy sets.

It is worth noting here that all simplification processes, expect for the ‘insignificant rules’, only have α chance to be evoked at each generation, where α is 20% in this work.

5. Experimental studies

To validate the proposed modeling framework, it is applied to the modeling of Tensile Strength (TS) of alloy steels. In this work, 3760 TS data are used. 75% of the data are used for training and the remaining data are used for checking. Another 12 more recent samples are used as the unseen data set to validate the generalisation properties of the model. The TS data includes 15 inputs and one output. The inputs consists of the weight percentages for the chemical compositions, the test depth, the size of the specimen and the site where it has been produced, the cooling medium, the quenching and tempering temperatures. The output is the tensile strength itself.

Two objectives are formulated with the first focusing on the prediction accuracy and the second on the structure simplification as follows:

$$\text{Objective 1: } RMSE = \sqrt{\frac{\sum_{k=1}^m (y_{prediction} - y_{real})^2}{m}} \quad (18) \\ \text{Objectiv 2: } Nrule + Nset + RL$$

where, $y_{prediction}$ and y_{real} are predicted and real outputs respectively; N_{rule} is the number of fuzzy rules in FRBS; N_{set} is the total number of fuzzy sets; RL is the summation of the rule length of each rule.

Fig. 4 shows the Pareto fronts obtained through the third stage. Table 1 includes the detailed parameters of a 9-rule fuzzy model selected from a set of Pareto models and its validation performance upon 12 unseen data points.

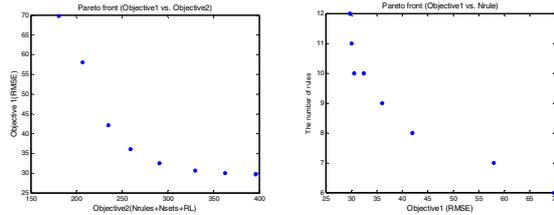


Fig. 4. The Pareto fronts obtained from the third stage

Table 1. The parameters of a 9-rule fuzzy model

9-rule fuzzy model	Parameters
The number of fuzzy rules	9
The number of fuzzy sets in each inputs	Inputs: [7; 6; 6; 6; 5; 7; 6; 8; 7; 6; 2; 6; 6; 9; 9]
RMSE for training	36.013
RMSE for checking	39.767
RMSE for validation	48.619

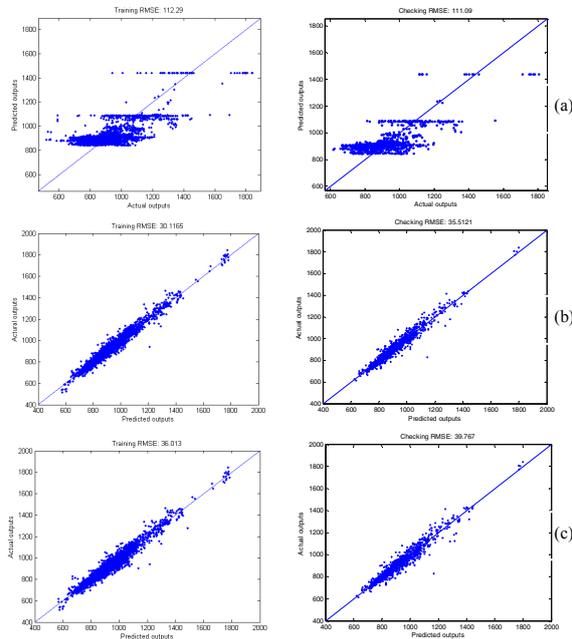


Fig. 5. The prediction performances of the three stages: (a) the initial FRBS; (b) the refined FRBS; (c) a 9-rule FRBS

Fig. 5 shows the results obtained from the three stages. The first two stages give FRBS consisting of a maximum of 12 rules. The third stage produce a set of Pareto FRBS and only a 9-rule model is presented here. The training

RMSEs are 112.29, 30.12 and 36.01 for the three stages respectively. The checking RMSE is 111.09, 35.51 and 39.77. Fig. 6 shows the simplified fuzzy sets in input 11 compared to the refined 12-rule model.

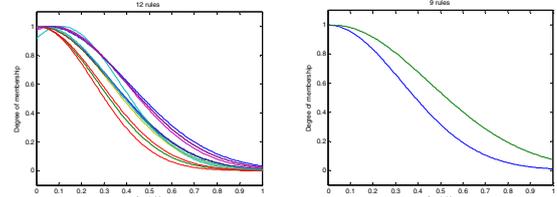


Fig. 6. The fuzzy sets of input 11: (right) the refined 12-rule model; (left) the optimized 9-rule model

6. Conclusion

The proposed modeling approach is not sensitive to the initial settings. The initial granulation level 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 as a prior. This simple framework provides the user with more options for a set of eliciting optimal models and leaves the designer's intervention to a minimum level.

References

- [1] Cordon, O., Herrera, F., Hoffann, F., Magdalena, L., *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, World Scientific, Singapore, 2001.
- [2] Jang, J.-S.R., "ANFIS: Adaptive-Network-Based Fuzzy Inference System", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 23(3), 1993, pp. 665-685.
- [3] Gomez-Skarmeta, A. F., Delgado, M., Vila, M. A., "About the Use of Fuzzy Clustering Techniques for Fuzzy Model", *Fuzzy Sets and Systems*, Vol. 106, 1999, pp. 179-188.
- [4] Hartigan, J. A., Wong, M. A., "A K-Means Clustering Algorithm", *Applied Statistics*, Vol. 28 (1), 1979, pp. 100-108.
- [5] Bezdek, J. C., Ehrlich, R., Full, W., "FCM: The Fuzzy C-means Clustering Algorithm", *COMP. GEOSCI.*, Vol. 10 (2-3), 1984, pp. 191-203.
- [6] Chiu, S., "Fuzzy Model Identification Based on Cluster Estimation", *J. of Intelligent & Fuzzy Systems*, Vol. 2 (3), 1994
- [7] Deb, K., Anand, A., Joshi, D., "A Computationally Efficient Evolutionary Algorithm for Real-Parameter Optimization", *Evolutionary Computation*, Vol. 10 (4), MIT Press, 2002, pp. 371-395.
- [8] Chen, J., Mahfouf, M., "Artificial Immune Systems as a Bio-inspired Optimization Technique and Its Engineering Applications", *Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies*, 2007 (to be appear).
- [9] Chen, M. Y., Linkens, D. A., "A Systematic Neuro-Fuzzy Modeling Framework With Application to Material Property Prediction", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 31 (5), 2001, pp.781-790
- [10] Wang, H. L., Kwong, S., Jin, Y. C., Wei, W., Man, K. F., "Multi-objective Hierarchical Genetic Algorithm for Interpretable Fuzzy Rule-based Knowledge Extraction", *Fuzzy Sets and Systems*, Vol. 149 (1), 2005, pp. 149-186.