Improving interpretability in approximative fuzzy models via multi-objective evolutionary algorithms.

Gómez-Skarmeta, A.F.
University of Murcia
skarmeta@dif.um.es

Jiménez, F.
University of Murcia
fernan@dif.um.es

Sánchez, G.
University of Murcia
gracia@dif.um.es

Abstract

Current research lines in fuzzy modeling mostly tackle with improving the accuracy in descriptive models, and the improving of the interpretability in approximative models. This paper deals with the second issue approaching the problem by means of multi-objective optimization in which accurate and interpretability criteria are simultaneously considered. Evolutionary Algorithms are specially appropriated for multi-objective optimization because they can capture multiple Pareto solutions in a single run of the algorithm. We propose a multi-objective evolutionary algorithm to find multiple Pareto solutions (fuzzy models) showing a trade-off between accuracy and interpretability. Additionally neural network based techniques in combination with ad hoc techniques for interpretability improving, are incorporated into the multi-objective evolutionary algorithm in order to improve the efficacy of the algorithm.

Keywords: Takagi-Sugeno Fuzzy models, Multi-objective Evolutionary Algorithms, Pareto optimality, Radial Basis Function (RBF) Neural Networks.

1 Introduction

Evolutionary Algorithms (EA) [2] have been recognized as appropriated techniques for multi-objective optimization because they perform a search for multiple solutions in parallel [1]. EAs have been applied to learn both the antecedent and consequent part of fuzzy rules, and models with both fixed and varying number of rules have been considered [5]. Also, EAs have been combined with other techniques like fuzzy clustering [3] and neural networks [6]. This has resulted in many complex algorithms and, as recognized in [11] and [7], often the transparency and compactness of the resulting rule base is not considered to be of importance. In such cases, the fuzzy model becomes a black-box, and one can question the rationale for applying fuzzy modeling instead of other techniques.

Most evolutionary approaches to multi-objective fuzzy modeling consist of multiple EAs, usually designed to achieve a single task each, which are applied sequentially to obtain a final solution. In these cases each EA optimizes the problem attending to one criterion separately which is an impediment for the global search. Simultaneous optimization of all criteria is more appropriate. Other approaches are based on classical multi-objective techniques in which multiple objectives are aggregated into a single function to be optimized. In this way a single EA obtains a single compromise solution. Current evolutionary approaches for multi-objective optimization consist of a single multi-objective EA, based on the Pareto optimality notion, in which all objective are optimized simultaneously to find multiple non-dominated solutions in a single run of the EA. These approaches can also be considered from the fuzzy modeling perspective [4]. The advantage of the classical approach is that no further interaction with the decision maker is required, however it is often difficult to define a good aggregation function. If the final solution cannot be accepted, new runs of the EA may be required until a satisfying solution is found. The advantages of the Pareto approach are that no aggregation function has to be defined, and the decision maker can choose the most appropriate solution according to the current decision environment at the end of the EA run. Moreover, if the decision environment changes, it is not always necessary to run the EA again. Another solution may be chosen out of the family of non-dominated solutions that has already been obtained.

In this paper, we propose a multi-objective neuro-EA (MONEA) to obtain a parameter estimation for fuzzy models. In section 2, the fuzzy model is defined and in section 3 and 4, techniques to improve transparency and compactness of rule set and training are approached respectively. The criteria taken into account are discussed in section 5. The main components of the multi-objective neuro-EA are described in section 6. The optimization model and decision making is described in section 7. Section 8 shows the results obtained for a test problem. Section 9 concludes the paper and indicates lines for future research.

2 Fuzzy Model Identification

We consider Takagi-Sugeno (TS) rule-based models [10] which are especially suitable for the approximation of dynamic systems. The rule consequents are often taken to be linear functions of the inputs:

$$\begin{array}{l} R_i : \textbf{If } x_1 \textit{ is } A_{i1} \textbf{ and} \ldots x_n \textit{ is } A_{in} \textbf{ then} \\ \hat{y}_i = \zeta_{i1}x_1 + \ldots + \zeta_{in}x_n + \zeta_{i(n+1)},\ i = 1, \ldots, M \end{array}$$

Here $\mathbf{x} = (x_1, \ldots, x_n)$ is the input vector, $\hat{y}_i$ is the output of the ith rule, A_{ij} are fuzzy sets defined in the antecedent space by membership functions $\mu_i : \Re \rightarrow [0,1]$, $\zeta_{ij} \in \Re$ are the consequent parameters, and M is the number of rules.

$$\hat{y} = \sum_{i=1}^{M} p_i(\mathbf{x})\hat{y}_i$$

where $p_i(\mathbf{x})$ is the normalized firing strength of the ith rule: Each fuzzy set A_{ij} is described by an asymmetric gaussian membership function:

$$\mu_{A_{ij}}(x_j) = \begin{cases} \exp\left(-\frac{\left(c_{ij}-x_j\right)^2}{2\sigma_{lij}^2}\right) & if\ x_j < c_{ij} \\ \exp\left(-\frac{\left(x_j-c_{ij}\right)^2}{2\sigma_{rij}^2}\right) & if\ x_j \geq c_{ij} \end{cases}$$

This fuzzy model is defined by a radial basis function neural network. The number of neurons in the hidden layer of an RBF neural network is equal to the number of rules in the fuzzy model. The firing strength of the ith neuron in the hidden layer matches the firing strength of the ith rule in the fuzzy model. We apply an asymmetric gaussian membership function defined by three parameters, the center c, left variance σ_l and right variance σ_r. Therefore, each neuron in the hidden layer has these three parameters that define its firing strength value.

The neurons in the output layer perform the computations for the first order linear function described in the consequents of the fuzzy model, therefore, the ith neuron of the output layer has the parameters ζ_i that correspond to the linear function defined in the ith rule of the fuzzy model.

3 Rule Set Simplification Technique

Automated approached to fuzzy modeling often introduce redundancy in terms of several similar fuzzy sets that describe almost the same region in the domain of some variable. According to some similarity measure, two similar fuzzy sets can be merged or separated. The merging-separation process is repeated until fuzzy sets for each model variable are not similar. This simplification may results in several identical rules, which are removed from the rule set.

We consider the following similarity measure between two fuzzy sets A and B:

$$S(A,B) = \max\left\{\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|}\right\}$$

If $S(A,B) > \theta$ (we use $\theta = 0.6$), fuzzy sets A and B are merged in a new fuzzy set C as follows:

$$\begin{array}{l} c_C = \alpha c_A + (1-\alpha)c_B \\ c_C^l = \min\left\{c_A^l, c_B^l\right\} \\ c_C^r = \max\left\{c_A^r, c_B^r\right\} \end{array} \qquad (1)$$

where $\alpha \in [0,1]$ determines the influence of A and B in the new fuzzy set C:

$$\alpha = \frac{c_A^r - c_A^l}{c_A^r - c_A^l + c_B^r - c_B^l}$$

If $\eta_2 < S(A,B) < \eta_1$ (we use $\eta_2 = 0.6$), fuzzy sets A y B are splitted as follows:

$$\begin{array}{lll} If\ c_A < c_B\ then & \sigma_A^r \leftarrow \sigma_A^r(1-\beta), & \sigma_B^l \leftarrow \sigma_C^l(1-\beta) \\ in\ other\ case & \sigma_A^l \leftarrow \sigma_A^l(1-\beta), & \sigma_B^r \leftarrow \sigma_B^r(1-\beta) \end{array}$$

where $\beta \in [0,1]$ indicates the amount of separation between A y B (we use $\beta = 0.1$).

4 Training of the RBF neural networks

The RBF neural networks associated with the fuzzy models can be trained with a gradient method to obtain more accuracy. However, in order to maintain the transparency and compactness of the fuzzy sets, only the consequent parameters are trained. The training algorithm incrementally updates the parameters applying the gradient descent method.

5 Criteria for Fuzzy Modeling

We consider three main criteria: (i) accuracy, (ii) transparency, and (iii) compactness. It is necessary to define quantitative measures for these criteria by means of appropriate objective functions which define the complete fuzzy model identification.

The accuracy of a model can be measured with the *mean squared error*:

$$MSE = \frac{1}{K}\sum_{k=1}^{K}(y_k - \hat{y}_k)^2$$

where y_k is the true output and $\hat{y}_k$ is the model output for the kth input vector, respectively, and K is the number of data samples.

For the second criterion, transparency, we consider the *similarity* S among distinct fuzzy sets in each variable:

$$S = \max_{\substack{i,j,k \\ A_{ij} \neq A_{kj}}} S(A_{ij}, A_{kj})$$

This is an aggregated similarity measure for the fuzzy rule-based model with the objective to minimize the maximum similarity between the fuzzy sets.

Finally, measures for the third criterion, the compactness, are the number of rules, M and the number of different fuzzy sets L of the fuzzy model.

6 Multi-objective Neuro-Evolutionary Algorithm (MONEA)

The main characteristics of MONEA are the following:

- The proposed algorithm is a Pareto-based multi-objective EA for fuzzy modeling, i.e., it has been designed to find, in a single run, multiple non-dominated solutions according to the Pareto decision strategy. There is no dependence between the objective functions and the design of the EA, thus, any objective function can easily be incorporated. Without loss of generality, the EA minimizes all objective functions.

- Constrains with respect to the fuzzy model structure are satisfied by incorporating specific knowledge about the problem. The initialization procedure and variation operators always generate individuals that satisfy these constrains.

- The EA has a variable-length, real-coded representation. Each individual of a population contains a variable number of rules between 1 and *max*, where *max* is defined by a decision maker. Fuzzy numbers in the antecedents and the parameters in the consequent are coded by floating-point numbers.

- The initial population is generated randomly with a uniform distribution within the boundaries of the search space, defined by the learning data and model constrains.

  • The EA searches for among rule sets treated with the technique described in section 3 and trained as defined in section 4, which is an added ad hoc technique for transparency, compactness and accuracy.

- Chromosome selection and replacement are achieved by means of a variant of the preselection scheme. This technique is, implicitly, a niche formation technique and an elitist strategy. Moreover, an explicit niche formation technique has been added to maintain diversity respect to the number of rules of the individuals. Survival of individuals is always based on the Pareto concept.

- The EAs variation operators affect at the individuals at different levels: (i) the rule set level, (ii) the rule level, and (iii) the parameter level.

6.1 Representation of Solutions

An individual I for this problem is a rule set of M rules defined by the weights of the RBF neural network. With n input variables, we have for each individual the following parameters:

- centers c_{ij}, left variances σ_{lij} and right variances σ_{rij}, $i = 1, \ldots, M$, $j = 1, \ldots, n$

- coefficients for the linear function of the consequent ζ_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, n+1$

6.2 Variation Operators

As already said, an individual is a set of M rules. A rule is a collection of n fuzzy numbers (antecedent) plus $n+1$ real parameters (consequent), and a fuzzy number is composed of three real numbers. In order to achieve an appropriate exploitation and exploration of the potential solutions in the search space, variation operators working in the different levels of the individuals are necessary. In this way, we consider three levels of variation operators: rule set level, rule level and parameter level.

Rule Set Level Variation Operators: Rule Set Crossover (interchanges rules), Rule Set Increase Crossover (increases the number of each child rules adding rules of the other), Rule Set Mutation (deletes or adds a new rule), and Rule Set Train (trains the RBF neural network).

Rule Level Variation Operators: Rule Arithmetic Crossover (arithmetic crossover of two random rules) and Rule Uniform Crossover (uniform crossover of two random rules).

Parameter Level Variation Operators: Arithmetic crossover, uniform crossover, not uniform mutation, uniform mutation and small mutation. These operators excluding the last one have been studied and described by other authors. The small mutation produced an small change in the individual and it is suitable for fine tuning of the real parameters.

6.3 Selection and Generational Replacement

In each iteration, MONEA executes the following steps:

1. Two individuals are picked at random from the population.

2. These individuals are crossed, mutated and trained to produce two offspring.

3. Performs the simplification in the offspring.

4. The first offspring replaces the first parent and the second offspring replaces the second parent only if:
   - the offspring is better than the parent and
   - the number of rules of the offspring is equal to the number of rules of the parent, or the niche count of the parent is greater than *minNS* and the niche count of the offspring is smaller than *maxNS*.

An individual I is better than another individual J if I dominates J. The niche count of an individual I is the number of individuals in the population with the same number of rules as I. *minNS* and *maxNS*, with $0 \leq minNS \leq \frac{PS}{max} \leq maxNS \leq PS$ (PS is the population size), are the minimum and maximum niche size respectively.

The preselection scheme is an implicit niche formation technique to maintain diversity in the population because an offspring replaces an individual similar to itself (one of their parents). Implicit niche formation techniques are more appropriate for fuzzy modeling than explicit techniques, such as sharing function, which can provoke an excessive computational time. However, we need and additional mechanism for diversity with respect to the number of rules of the individuals in the population. The added explicit niche formation technique ensures that the number of individuals with M rules, for all $M \in [1, max]$, is greater or equal to *minNS* and smaller or equal to *maxNS*. Moreover, the preselection scheme is also an elitist strategy because the best individual in the population is replaced only by a better one.

7 Optimization Model and Decision Making

After preliminary experiments in which we have checked different optimization models, the following remarks can be maded:

- Instead of minimizing of the number of rules M we have decided to search for rules sets with a number of rules within an interval $[1, max]$ where a decision maker can feel comfortable. The explicit niche formation technique ensures the EA always contains a minimum of representative rule sets for each number of rules in the populations. Then, we do not minimize the number of rules during the optimization, but we will take it into account at the end of the run, in a posteriori decision process applied to the last population.

- It is very important to note that a very transparent model will be not accepted by a decision maker if the model is not accurate. In most fuzzy modeling problems, excessively low values for similarity hamper accuracy, for which these models are normally rejected. Alternative decision strategies, as *goal programming*, enable us to reduce the domain of the objective functions according to the preferences of a decision maker. Then, we can impose a goal g_S for similarity, which stop minimization of the similarity in solutions for which goal g_S has been reached.

- The measure L (number of different fuzzy sets) is considerably reduced by the rule set simplification technique. So, we do not define an explicit objective function to minimize L.

According to the previous remarks, we finally consider the following optimization model:

$$\begin{array}{ll} Minimize & f_1 = MSE \\ Minimize & f_2 = max\{g_S, S\} \end{array}$$

At the end of the run, we consider the following a posteriori decision strategy applied to the last population to obtain the final compromise solution:

1. Identify the set $X^* = \{x_1^*, \ldots, x_p^*\}$ of non-dominated solutions according to:

$$\begin{array}{ll} Minimize & f_1 = MSE \\ Minimize & f_2 = S \\ Minimize & f_3 = M \end{array}$$

2. Choose from X^* the most accurate solution x_i^*; remove x_i^* from X^*;

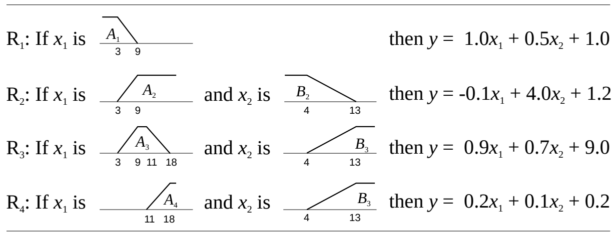
3. If solution x_i^* is not accurate enough or there is no solution in the set X^* then STOP (no solution satisfies);

4. If solution x_i^* is not transparent or compact enough then go to step 2;

5. Show the solution x_i^* as output.

8 Experiments and Results

We consider the modeling of the rule base given by Sugeno in [9]:



The corresponding surface is shown in Figure 1. In [8] a model with four rules was identified from sampled data ($N = 546$) by the supervised clustering algorithm, which was initialized with 12 clusters. This model was optimized using a Genetic Algorithm to result in a MSE of 1.6. Table 1 shows results obtained with MONEA using the following values for the parameters: $PS = 100$, $minNS = 5$, $maxNS = 30$, cross probability 0.9, mutation probability 0.9, train probability 0.2, $g_S = 0.35$, and $max = 5$.

Table 1: Solutions obtained with MONEA.

M	**L**	**MSE**	**S**
1	2	125.291	0.0
2	4	25.606	0.348
3	5	2.187	0.349
4	5	1.017	0.349
5	5	0.910	0.350

We finally choose a compromise solution (4-rules fuzzy model) according to the described decision process. Figure 2 shows the local model, the surface generated by the model, fuzzy sets for each variable and the prediction error.

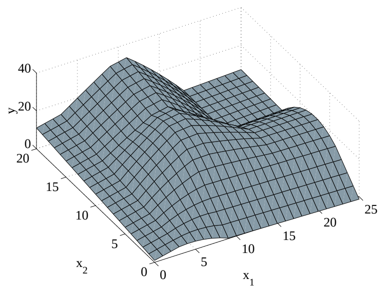
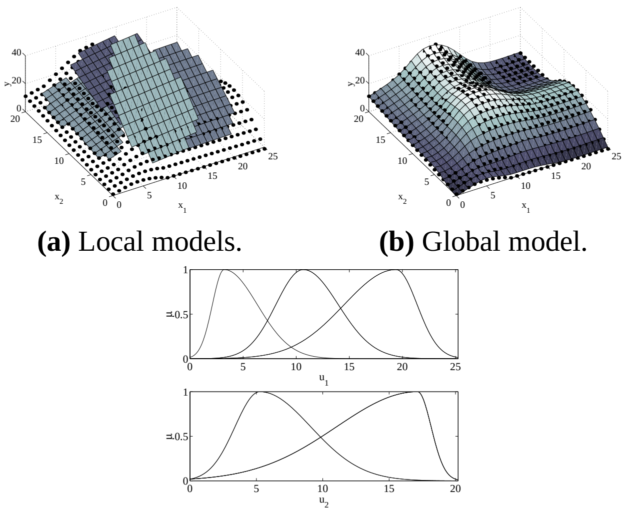


Figure 1: Real surface for the example in [9].



(a) Local models. **(b)** Global model.

(c) Fuzzy sets.

Figure 2: Accurate, transparent and compact fuzzy model for the example in [9].

9 Conclusions

This paper remarks some results in the combination of Pareto-based multi-objective evolutionary algorithms, neural networks and fuzzy modeling. Criteria such as accuracy, transparency and compactness have been taken into account in the optimization process. An implicit niche formation technique (preselection) in combination with other explicit techniques with low computational costs have been used to maintain diversity. These niche formation techniques are appropriated in fuzzy modeling if excessive amount of data are required. Elitism is also implemented by means of the preselection technique. A goal based approach has been proposed to help to obtain more accurate fuzzy models. Results obtained are good in comparison with other more complex techniques reported in literature, with the advantage that the proposed technique identifies a set of alternative solutions.

Acknowledgements

The authors thank the Ministerio de Ciencia y Tecnología (MCyT) and the Fondo Europeo de Desarrollo Regional (FEDER) for the support given to this work under the project TIC2001-0245-C02-01.

References

[1] C.A. Coello (2000). List of references on multi-objective evolutionary optimization. http://www.lania.mx/∼ccoello/EMOO/EMOObib.html

[2] D.E. Goldberg (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

[3] A.F. Gómez-Skarmeta and F. Jiménez (1999). Fuzzy modeling with hibrid systems. *Fuzzy Sets and Systems*, 104:199-208.

[4] H. Ishibuchi, T. Murata, and I. Türksen (1997). Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets and Systems*, 89:135-150.

[5] H. Ishibuchi, T. Nakashima, and T. Murata (1999). Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Transactions on Systems, Man, and Cubernetics - Part B: Cybernetics*, 29/5):601-618.

[6] M. Russo. FuGeNeSys - a fuzzy genetic neural system for fuzzy modeling (1998). *IEEE Transactions on Fuzzy Systems*, 6(3):373-388.

[7] M. Setnes, R. Babuska, and H. B. Verbruggen (1998). Rule-based modeling: Precision and transparency. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications & Reviews*, 28:165-169.

[8] M. Setnes and J.A. Roubos (1999). Transparent Fuzzy Modeling using Fuzzy Clustering and GA's, *18th International Conference of the North American Fuzzy Information Processing Society*, 198-202.

[9] M. Sugeno and G.T. Kang (1998). Structure identification of fuzzy model *Fuzzy Sets and Systems*, 28:15-23.

[10] Takagi, T., Sugeno, M. (1985). *Fuzzy identification of systems and its application to modeling and control*. IEEE Transactions on Systems, Man and Cybernetics, 15:116-132.

[11] J. Valente de Oliveira (1999). Semantic constraints for membership function optimization. *IEEE Transactions on Fuzzy Systems*, 19(1):128-138.