Cooperation between the Inference System and the Rule Base by Using Multiobjective Genetic Algorithms

Antonio Márquez, Francisco Alfredo Márquez, and Antonio Peregrín

Information Technologies Department
University of Huelva, 21071 Huelva, Spain
{amarquez,alfredo.marquez,peregrin}@dti.uhu.es

Abstract. This paper[1] presents an evolutionary Multiobjective learning model achieving positive synergy between the Inference System and the Rule Base in order to obtain simpler and still accurate linguistic fuzzy models by learning fuzzy inference operators and applying rule selection. The Fuzzy Rule Based Systems obtained in this way, have a better trade-off between interpretability and accuracy in linguistic fuzzy modeling applications.

Keywords: Linguistic fuzzy modeling, interpretability-accuracy trade-off, Multiobjective genetic algorithms, adaptive inference system, adaptive defuzzification, rule selection.

1 Introduction

Interpretability and accuracy are usually contradictory requirements in the design of linguistic fuzzy models (FMs). In practice, designers must find an adequate trade-off between them for the specific application, increasing the interest of this matter in the literature [1],[2].

Two important tasks in the design of a linguistic FM for a particular application are: The derivation of the linguistic rule base (RB) and the setup of the inference system and defuzzification method. In the framework of the trade-off between *interpretability* and *accuracy* in fuzzy modeling, adaptive inference system and defuzzification method acquired greater importance [3],[4].

Recently, the use of Multiobjective Evolutionary Algorithms (MOEA) has been applied to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [5],[6],[7],[8]. Most of these works [5],[6] obtain the complete Pareto (the set of non-dominated solutions with different trade-off) by selecting or learning the set of rules better representing the example data, i.e., improving the system accuracy and decreasing the fuzzy RB system complexity. In [7],[8], authors also propose the tuning of the membership functions together with the rule selection to obtain simpler and still accurate linguistic FMs.

Following these ideas on the advantage of the use of parametric operators and the use of MOEAs to improve the trade-off between interpretability and accuracy, in this

[1] Supported by Projects TIN2005-08386-C05-01, P05-TIC-00531 and P07-TIC-03179.

E. Corchado, A. Abraham, and W. Pedrycz (Eds.): HAIS 2008, LNAI 5271, pp. 739–746, 2008.

work we present a MOEA to learn the fuzzy inference (including inference and defuzzification) and to perform rule selection for Mamdani linguistic fuzzy systems. The proposed model tries to achieve a positive synergy (this is the concept of cooperation we use) between the fuzzy operators and the RB to improve the accuracy at the same time that the RB is simplified to improve the interpretability.

In order to do this, Section 2 introduces the parametric fuzzy operators, Section 3 is devoted to describe the MOEA learning proposal, Section 4 develops an experimental study, and finally, Section 5 presents some concluding remarks.

2 Adaptive Fuzzy Operators

In this section we show the adaptive inference system as well as the adaptive defuzzification method used in our learning proposal.

2.1 Adaptive Inference System

Linguistic FRBSs for system modeling use IF - THEN rules of the following form:

$$R_i : \text{If } X_{i1} \text{ is } A_{i1} \text{ and ... and } X_{im} \text{ is } A_{im} \text{ then } Y \text{ is } B_i$$

with i = 1 to N, the number of rules of the RB, and with X_{i1} to X_{im} and Y being the input and output variables respectively, and with A_{i1} to A_{im} and B_i being the involved antecedents and consequent labels, respectively.

The expression of the Compositional Rule of Inference in fuzzy modeling with punctual fuzzification is the following one: $\mu_{B'}(y) = I(C(\mu_{A1}(x_1), \ldots, \mu_{Am}(x_m)), \mu_B(y))$, where $\mu_{B'}(\cdot)$ is the membership function of the inferred consequent, $I(\cdot)$ is the rule connective, $C(\cdot)$ is the conjunction operator, $\mu_{Ai}(x_i)$ are the values of the matching degree of each input of the system with the membership functions of the rule antecedents, and $\mu_B(\cdot)$ is the consequent of the rule.

The two components, conjunction ($C(\cdot)$) and rule connective ($I(\cdot)$) are suitable to be parameterized in order to adapt the inference system. Our previous studies in [3] show that the model based on the adaptive conjunction is a more valuable option than the one based on the adaptive rule connective. Hence, we selected the use of the adaptive conjunction in this study, in order to insert parameters in the inference system.

Taking into account the studies in [3], we have selected the Dubois adaptive t-norm with a separate connector for every rule, which expression is showed in (1).

$$T_{Dubois}(x,y,\alpha) = x \cdot y / Max(x,y,\alpha), \quad (0 \le \alpha \le 1) \tag{1}$$

2.2 Adaptive Defuzzification Interface

The most used methodology in practice, due to its fine performance, efficiency and easier implementation, is to apply the defuzzification function to every rule inferred fuzzy set (getting a characteristic value) and to compute then by a weighted average operator. This way of working is named FITA (First Infer, Then Aggregate) [9].

Attending to the studies developed in [10], in this work we consider to use a product functional term of the matching degree between the input variables and the rule antecedent fuzzy sets (h_i), $f(h_i) = h_i \cdot \beta_i$ where β_i corresponds to one parameter for each rule R_i, i=1 to N, in the RB. The adaptive defuzzification formula selected is (2).

$$y_0 = \sum_i^N h_i \cdot \beta_i \cdot V_i \Big/ \sum_i^N h_i \cdot \beta_i, \quad (2)$$

where V_i represents a characteristic value of the fuzzy set inferred from rule R_i, the Maximum Value or the Gravity Center (GC), the one selected in this paper.

3 Cooperative Evolutionary Selection of Fuzzy Rules and Learning of Adaptive Fuzzy Operators with Multiobjective Algorithms

This section describes the basics of two of the most representative second generation MOEAs, SPEA2 [12] and NSGA-II [13], as two general propose MOEAs used in this work, and later the adaptations we propose to perform the cooperative adaptation of the fuzzy operators and fuzzy rule selection.

3.1 SPEA2 and NSGA-II

The SPEA2 algorithm [11] (*Strength Pareto Evolutionary Algorithm for multiobjective optimization*) is one of the most known techniques in the Multiobjective problem solving. It is characterized by the following two aspects: a *fitness* assignment strategy, that takes into account both dominating and dominated solutions for each individual, and a density function that is estimated employing the nearest neighbourhood, which guides the search more efficiently. A deeper description of the algorithm may be found in the aforementioned paper [11].

NSGA-II algorithm [12] is also another of the most well-known and frequently-used MOEAs for general multi-objective optimization in the literature. It is a parameterless approach with many interesting principles: a binary tournament selection based on a fast non-dominated sorting, an elitist strategy and a crowding distance method to estimate the diversity of a solution. As was commented before, a deeper description may be found in [12].

3.2 Questions Related to the MOEAs

The evolutionary model uses a chromosome with threefold coding scheme (C_C+C_D+C_S) where:

- C_C encodes the α_i parameters of the conjunction connective, that is N real coded parameters (genes) for each rule, R_i ,of the linguistic RB. Each gene can take any value in the interval [0,1], that is, among minimum and algebraic product.
- C_D encodes the β_i parameters of the defuzzification. They are N real coded parameters for each rule of the linguistic RB. Each gene can take any value in the interval [0,10]. This interval has been selected according with the study developed in [10]. It allows attenuation as well as enhancement of the matching degree.

- Cs encodes the rule selection. It is a binary string of N genes, each one representing a candidate rule of the initial RB. Depending on whether a rule is selected or not, values '1' or '0' are respectively assigned to the corresponding gene.

The initial population is randomly initialized with the exception of a single chromosome with the following setup:

- Fuzzy operators part: Cc with the N genes is initiated to 0 in order to make Dubois t-norm equivalent to Minimum t-norm initially. Cd also with the N genes is initiated to 1 with the objective to begin like the standard WCOA method.
- Rule selection part, Cs, with the N rules obtained by the WM-method [13], that is, with all the genes initialized to '1'.

The crossover operator employed for the fuzzy operators part is BLX-0.5 [14] while the one used for rule selection part is HUX [15].

The fitness based on the interpretability (using the number of rules) and the accuracy (using the error measure) must be minimized.

3.3 Improvements for SPEA2 and NSGA-II

This subsection is devoted to describe the two improvements we propose for SPEA2 [11] and NSGA-II [12] respectively to perform better the searching process we pretend, that is to guide the search towards the desired Pareto zone with high accuracy with the least possible number of rules.

The *Improved Accuracy NSGA-II* (NSGA-II_{IA}) algorithm:

We propose two main changes on the NSGA-II algorithm, based on the changes proposed by [8] with a few modifications to perform better with our problem, who has a larger real part comparatively. They are the following:

- We use a restarting mechanism carried out twice at 1/3 and 2/3 of the execution of the algorithm. Each time, the most accurate individual is maintained as the sole individual in the external population. The remaining individuals are reinitiated with the same rule configuration of the best individual and fuzzy operator parameters randomly generated within the corresponding variation intervals.
- In each of the three algorithm stages (before the first restart, after the first restart and after the second restart), the number of solutions in the external population considered to form the mating pool is progressively reduced, by focusing only on those with the best accuracy. To do that, the solutions are sorted from the best to the worst (considering accuracy as sorting criterion) and the number of solutions considered for selection is reduced progressively from 100% to 50 % at the beginning 1/3, since 75% to 50% in the middle, and since 66% to 50% at the end.

The *Improved Accuracy SPEA2* ($SPEA2_{IA}$) algorithm:

In order to guide the searching process of the SPEA2, we propose to employ a method called Guided Domination Approach [16], which gives priority to the accuracy objective through a weighted function of the objectives. Focusing the searching process we can reduce the effort of the search, and a better precision in the non-dominated solutions can be obtained, because the searching effort is concentrated in a reduced zone of the Pareto, being the density of the obtained solutions higher. The weighted function of the objectives is defined in (3),

$$\Omega_i\left(f(x)\right) = f_i(x) + \sum_{j=1,\, j \neq i}^{M} a_{ij} f_j(x), \qquad i=1,2,\ldots M \tag{3}$$

where a_{ij} is the amount of gain in the *j-th* objective function for a loss of one unit in the *i-th* objective function, and M being the number of objectives. The above set of equations require fixing the matrix *a*, which has a one in its diagonal elements. This method redefines the domination concept as follows: *A solution $x^{(1)}$ dominates another solutions $x^{(2)}$, if $\Omega_i(f(x^{(1)})) \leq \Omega_i(f(x^{(2)}))$ for all $i = 1,2, \ldots, M$, and the strict inequality is satisfied at least for one objective.*

Thus, if we have two (M=2) fitness functions, the two weighted functions are showed in (4).

$$\Omega_1(f_1,f_2) = f_1 + a_{12} f_2, \qquad \Omega_2(f_1,f_2) = a_{21} f_1 + f_2 \tag{4}$$

4 Experimental Study

In order to analyze the proposed methods, we built several FMs using the learning methods showed in Table 1. WM method is considered as a reference. S and C-D are methods that perform rule selection and adaptation of fuzzy operators respectively. S + C-D means rule selection and fuzzy operators adaptation together. SPEA2, $SPEA2_{IA}$, NSGA-II and $NSGA\text{-}II_{IA}$ are the methods that learn the fuzzy operators and the rule selection together, as was previously commented.

Table 1. Methods considered for comparisson

Ref.	Method	Description
[13]	WM	Wang & Mendel algorithm
[17]	WM + S	Rule Selection
[4],[10]	WM + C-D	Adaptive Fuzzy Operators
-	WM + S + C-D	Rule Selection and Adaptive Fuzzy Operators
[11]	SPEA2	SPEA2 Algorithm
-	$SPEA2_{IA}$	Improved Accuracy SPEA2
[12]	NSGA-II	NSGA-II Algorithm
-	$NSGA\text{-}II_{IA}$	Improved Accuracy NSGA-II

4.1 Application Selected and Comparison Methodology

The fuzzy partition used for inputs and output has 5 labels. The application selected to test the evolutionary model is an electrical distribution problem [18] that has got a data set of 1059 cities with four input variables and a single output. The RB is composed of 65 linguistic rules achieved with the Wang and Mendel method [13].

We considered a 5-foder cross-validation model, i.e., 5 random partitions of the data each with 20% (4 of them with 211 examples, and one of them with 212 examples) and the combination of 4 of them (80%) as training, and the remaining one as test. We achieved a total of 30 trials for every evolutionary process, because for each one of the data partitions, the learning methods have been run 6 times. We show the averaged values of the medium square error (MSE) as a usual performance measure, computed considering the most accurate solution from each Pareto obtained for the Multiobjective algorithm. This way to work was also employed in [8] in order to compare the single objective methods with the Multiobjective ones based in to consider only the accuracy objective, letting us to see that the Pareto fronts are not only wide but also optimal, so similar solutions obtained with the WM + S + C-D should be included in the final Pareto. The MSE is computed with expression (5),

$$MSE(S)_B = \frac{1}{2}\sum_{k=1}^{P}(y_k - S(x_k))^2 \Big/ P \quad (5)$$

where S denotes the fuzzy model whose inference system uses the Dubois t-norm as conjunction operator showed in expression (2), inference operator minimum t-norm, and the adaptive defuzzificación method showed in expression (3). This measure uses a set of system evaluation data formed by P pairs of numerical data $Z_k = (x_k, y_k)$, $k=1,..,P$, with x_k being the values of the input variables, and with y_k being the corresponding values of the associated output variables.

The MOGAs population size was fixed to 200. The external population size of the SPEA2 and $SPEA2_{IA}$ was 61.

The parameters a_{12}, a_{21} used for the $SPEA2_{IA}$ have been determined after several test and fixed to 0 and 8 respectively, and give more importance to the accuracy.

4.2 Results and Analysis

The results obtained are shown in Table 2, where #R is the average number of rules, MSE_{tra} and MSE_{tst} are the average MSE for training and test respectively, σ is the standard deviation and *t-test* is the result of applying a *test t-student* (with 95 percent confidence), with the following interpretation: * represents the best average result; + means that the best result has better performance than that of the corresponding row.

Analysing the results obtained we can point out that $NSGA\text{-}II_{IA}$ and $SPEA2_{IA}$ shows a simmilar accuracy (slightly better in training) compared with the WM + S + C-D with a significative reduction in the number of rules, particularly for $NSGA\text{-}II_{IA}$. Modifications proposed in order to improve the accuracy for NSGA-II and SPEA2 let

Table 2. Results obtained

Method	#R	MSE_{tra}	σ_{tra}	t-test	MSE_{test}	σ_{test}	t-test
WM	65	56135.75	4321.42	+	56359.42	5238.57	+
WM + S	40.9	41517.01	4504.85	+	44064.67	906.64	+
WM + C-D	65	22561.77	3688.27	=	**25492.77**	830.76	*
WM + S + C-D	52.8	22640.95	2125.05	=	26444.43	854.95	+
SPEA2	38.4	24077.42	8225.82	+	29664.50	874.18	+
$SPEA2_{IA}$	47.7	22450.72	3949.57	=	25562.74	850.07	=
NSGA-II	39.1	23303.50	6295.92	+	27920.42	877.83	+
$NSGA\text{-}II_{IA}$	41.1	**22108.66**	4695.30	*	26229.72	868.95	+

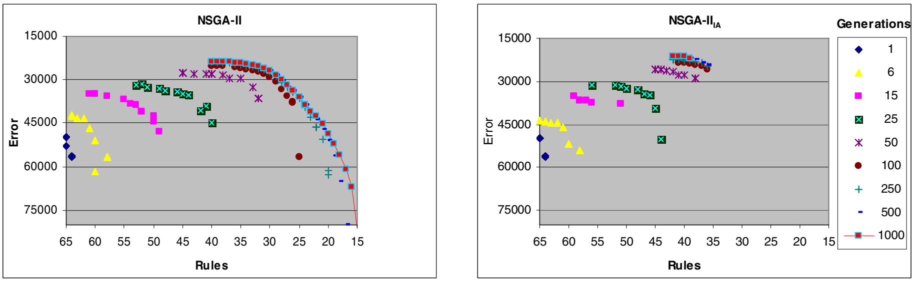

Fig. 1. Example of the Pareto front for NSGA-II and NSGA-II$_{IA}$

both models improve their accuracy compared with the standard versions of both algorithms. Due to the adaptive fuzzy parameters search space is large, it is necessary to focus the searching process on the Pareto zone with higher accuracy to reach simmilar accuracy level than the mono-objective algorithm based on the accuracy. Figure 1 shows the difference between the searching process performed by the original NSGA-II (left side) versus the improved accuracy version (right side).

5 Conclusions

In the framework of the trade-off between accuracy and interpretability, the use of Multiobjective genetic algorithm gives a set of solutions with different level of conciliation of both features. In this work we have proposed a Multiobjective evolutionary learning model where the adaptive fuzzy operator parameters are learnt together with the rule base selection. This fact allows both elements to cooperate, improving the accuracy as well as the interpretability. Methodologies to focus the searching process in a specific zone of the Pareto have also been shown useful when an objective must prevail over the other.

References

1. Casillas, J., Cordón, O., Herrera, F., Magdalena, L.: Interpretability issues in fuzzy modeling. Springer, Heidelberg (2003)
2. Casillas, J., Cordón, O., Herrera, F., Magdalena, L.: Accuracy improvements in linguistic fuzzy modeling. Springer, Heidelberg (2003)
3. Alcala-Fdez, J., Herrera, F., Márquez, F.A., Peregrín., A.: Increasing Fuzzy Rules Cooperation Based On Evolutionary Adaptive Inference Systems. Int. J. of Intelligent Systems 22(9), 1035–1064 (2007)
4. Márquez, F.A., Peregrín, A., Herrera, F.: Cooperative Evolutionary Learning of Fuzzy Rules and Parametric Aggregation Connectors for Mamdani Linguistic Fuzzy Systems. IEEE Trans. on Fuzzy Syst. 15(6), 1162–1178 (2007)
5. Ishibuchi, H., Yamamoto, T.: Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Syst. 141, 59–88 (2004)

6. Narukawata, K., Nojima, Y., Ishibuchi, H.: Modification of evolutionary Multiobjective optimization algorithms for Multiobjective design of fuzzy rule-based classification systems. In: Proc. 2005 IEEE Int. Conf. on Fuzzy Syst., pp. 809–814. Reno (2005)
7. Alcalá, R., Alcala-Fdez, J., Gacto, M.J., Herrera, F.: A Multi-Objective Evolutionary Algorithm for Rule Selection and Tuning on Fuzzy Rule-Based Systems. In: Proc. of the 16th IEEE Int. Conf. on Fuzzy Syst. FUZZ-IEEE 2007, London, pp. 1367–1372 (2007)
8. Alcalá, R., Gacto, M.J., Herrera, F., Alcala-Fdez, J.: A Multi-Objective Genetic Algorithm for tuning and Rule Selection to obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Syst. 15(5), 539–557 (2007)
9. Buckley, J.J., Hayashi, Y.: Can approximate reasoning be consistent? Fuzzy Sets and Syst. 65(1), 13–18 (1994)
10. Cordón, O., Herrera, F., Márquez, F.A., Peregrín, A.: A Study on the Evolutionary Adaptive Defuzzification Methods in Fuzzy Modelling. Int. J. of Hybrid Intelligent Syst. 1(1), 36–48 (2004)
11. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In: Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN 2001), pp. 95–100 (2001)
12. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. on Evolutionary Computation 6(2), 182–197 (2002)
13. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Trans. on Syst. Man, and Cybernetics 22(6), 1414–1427 (1992)
14. Eshelman, L.J., Schaffer, J.D.: Real-coded genetic algorithms and interval-schemata. Found. of Genetic Algorithms 2, 187–202 (1993)
15. Eshelman, L.J.: The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. Found. of Genetic Algorithms 1, 265–283 (1991)
16. Branke, J., Kaubler, T., Schmeck, H.: Guiding multi-objective evolutionary algorithms towards interesting region. Technical Report No. 399, Institute AIFB, University of Karlsruhe, Germany (2000)
17. Casillas, J., Cordón, O., del Jesus, M.J., Herrera, F.: Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. IEEE Trans. on Fuzzy Syst. 13(1), 13–29 (2005)
18. Cordón, O., Herrera, F., Sánchez, L.: Solving electrical distribution problems using hybrid evolutionary data analysis techniques. Appl. Intell. 10, 5–24 (1999)