A Methodology to Generate Compact and Accurate Fuzzy Knowledge Bases based on Fuzzy Clustering and Evolutionary Selection and Tuning

Ruth Mª Toscano, Javier Aroba, Antonio Peregrín

Abstract— **A new methodology to learn descriptive linguistic Fuzzy Rule-based System Knowledge Bases from examples based on the combination of fuzzy clustering and evolutionary simultaneous rule selection and membership functions tuning is presented in this work. Fuzzy clustering is used to achieve a preliminary description of the data, in other words to obtain information on the definition of the linguistic terms and rules instead of predefined linguistic terms and rules that use them. The evolutionary algorithm obtains the final compact and accurate knowledge base selecting a subset of rules with high level of cooperation and fine-tuning the linguistic terms involved. The results obtained with this proposal improves accuracy as well as complexity through the number of rules compared with a classic algorithm and a reference algorithm both well known in the literature, as the experimental study developed shows, using several different data sets.**

I. Introduction

In the framework of the design of fuzzy linguistic rule – based systems (FRBSs) for fuzzy modelling, the main element related with the problem to solve is the Knowledge Base (KB) that basically contains the fuzzy Rule Base (RB) as well as the fuzzy membership functions definition of the related variables. For this reason, the RB learning [1, 2, 3] and membership functions tuning [4, 5, 6] are problems that have traditionally been of great interest and widely studied in the specialist literature [7].

Recent research on the design of fuzzy linguistic models [6] has focused on methods aimed at generating FRBSs with an appropriate trade-off-between two usually contradictory features, accuracy and interpretability, in the sense of system complexity, so as to obtain reliable and understandable models.

Learning the KB or its elements from examples automatically allows fuzzy models to be created easily. The balance between interpretability, in the sense of complexity, and accuracy directly affects the KB design. When the main aim is to obtain a compact set of rules, i.e. when the interpretability is the main target, the accuracy decreases; in contrast, when higher accuracy is intended, the number of rules usually rises. Learning methodologies for interpretable and still accurate FRBSs is therefore an issue of interest nowadays.

This work was supported by Projects TIN2008-06681-C06-06 and P07-TIC-03179. R M. Toscano, J. Aroba and A. Peregrín are with the Department of Information Technologies, University of Huelva, 21819 Palos de la Fra. Huelva, Spain (e-mail: ruth.toscano@diesia.uhu.es; aroba@dti.uhu.es; peregrin@dti.uhu.es).

In line with these ideas, in this paper we present a methodology for the design of linguistic KBs from examples aimed at a more readily achievable accuracy and interpretability by using fuzzy clustering and evolutionary selection of the rules and tuning of the membership functions. Fuzzy clustering, also well known in fuzzy modelling [8], lets us create rules that individually describe the training data set while the evolutionary selection and fine tuning [4,6] favour a good level of cooperation between rules (inside the RB) and between the RB and the membership functions of the related variables, in order to have compact and accurate KBs. The novelty of this proposal is the way of combining fuzzy clustering and genetic fuzzy systems selection and tuning in a procedure to obtain Mamdani descriptive fuzzy systems knowledge bases for fuzzy modelling with higher levels of accuracy and compactness than other methodologies.

To explain how this is achieved, Section II describes the proposed methodology in depth, Section III shows the experimental study carried out with four applications and comparing the proposed methodology with two other learning methodologies, and finally Section IV presents some concluding remarks.

II. CEST Methodology

This paper proposes a new methodology for generating KBs for descriptive fuzzy linguistic systems from examples. It is based on the idea of placing rules and membership functions where they are needed to describe the training data set, instead of equilaterally partitioning the universe of the variables and then looking for the rules to describe the examples to cover. This is accomplished by using fuzzy clustering. Additionally, the proposed methodology benefits from the use of evolutionary algorithms to select the subset of rules with the best cooperation together and to tune the membership functions. We designate it CEST methodology (Clustering and Evolutionary Selection and Tuning).

This section describes in detail the elements of this methodology, which has two sequential steps: The first step is devoted to generating a set of candidate rules to depict the training data set and is carried out using a fuzzy clustering algorithm. This candidate rule set can be considered as an approximative fuzzy linguistic system.

978-1-4244-8126-2/10/$26.00

The first step concludes transforming the approximative fuzzy system to descriptive. The rules generated in this way are good rules to describe the data set individually, but they cannot perform together in a fuzzy system where there is interaction between rules. The second step employs an evolutionary algorithm to select the rules and tune the membership functions, both simultaneously to achieve the best collaboration between the two elements of the KB: RB (set of rules) and the data base (membership functions). Each stage is described in the following separate subsections.

A. Generation of Candidate Rules by Means of a Fuzzy Clustering Algorithm

Classical clustering algorithms generate a partition of a data set so that each item is assigned to a cluster. These algorithms use the so called "rigid partition" derived from classical sets theory: the elements of the partition matrix obtained from the data matrix (with n elements) can only contain values 0 or 1; with zero indicating null membership and one indicating whole membership to each of the c partitions (clusters). That is, the elements must fulfil:

$$
\begin{aligned}
&a)\quad \mu_{ik} \in \{0,1\},\ 1\le i \le c,\ 1\le k \le n \\
&b)\ \sum_{i=1}^{c} \mu_{ik} = 1,\ 1\le k \le n \\
&c)\ 0 \le \sum_{k=1}^{n} \mu_{ik} \le n,\ 1\le i \le c
\end{aligned} \quad (1)
$$

where μ_{ik} represents the membership degree of the k^{th} element to the i^{th} cluster.

Fuzzy partition is a generalization of the previous one, so that it holds the same conditions and constraints for its elements, except that in this case real values between zero and one are allowed (partial membership grade). Therefore, samples may belong to more than one group, so the selecting and clustering capacity of the samples increases. From this we may deduce that the elements of a fuzzy partition fulfil the conditions given in (1), except that now condition (a) will be written as:

$$\mu_{ik} \in [0,1],\ 1\le i \le c,\ 1\le k \le n \quad (2)$$

The Isodata algorithm [9] was modified by [10] and generalized in [11] and [12] until conversion into the well-known general-purpose fuzzy-clustering algorithm Fuzzy C-Means (FCM) [13]. This algorithm is based on the minimization of distances between two data points and the prototypes of cluster centres (c-means). Basically, this algorithm attempts to classify n elements $x_k \in X$ with $1 \le k \le n$, with p characteristics each, that is, $X \subset \Re_p$, into c fuzzy clusters, assigning a membership degree μ_{ik} (2). To this end, the algorithm to try minimize the following cost function (3)

$$J_m\ (U,P: X) = \sum_{k=1}^{n}\sum_{i=1}^{c} (\mu_{ik})^m D_{ik}^2 \quad (3)$$

where $U=(\mu_{ik})$ is the membership matrix of X, $P=[v_1, v_2, ..., v_c]$ is a vector of cluster centre prototypes which must be determined, and $m \in [1,\infty]$ is a weighting exponent which determines the degree of fuzziness of the resulting clusters (in this paper m=2 was considered) and is the norm used for measuring distances.

$$D_{ik}^2 = \| x_k - v_i \|^2 = (x_k - v_i)^T (x_k - v_i) \quad (4)$$

Finally, the cost function J is minimized to obtain the components of U and P, that is, the membership matrix and the vector of cluster centre prototypes. The necessary conditions to minimize J are:

$$\mu_{ik} = \left[\sum_{j=1}^{c} \left[\frac{\| x_k - v_i \|}{\| x_k - v_j \|} \right]^2 \right]^{-1} \forall i,k \quad (5)$$

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^2 x_k}{\sum_{k=1}^{n} (\mu_{ik})^2} \forall i \quad (6)$$

The final goal of this phase of the proposed methodology is to obtain a fuzzy rules system from a multi-parametric quantitative dataset. To do so, a fuzzy clustering algorithm [14] based on the methodology proposed by Sugeno and Yasukawa in [15] was applied, with the aim of building a fuzzy model based on fuzzy rules of the form:

$$R_i\text{: IF } x_i \in A_i \text{ THEN } y \in B_i \quad (7)$$

Where $X=[x_1, x_2,..., x_n] \in \Re_n$ are input variables (antecedents), $A=[A_1, A_2,..., A_n]$ are n fuzzy sets, $y \in \Re$ is an output variable (consequent), and $B=[B_1, B_2,..., B_m]$ are m fuzzy sets.

Therefore, the methodology used to obtain the fuzzy model basically consists of applying the fuzzy partition FCM to the output parameter y. As a result of this process, the membership grade is obtained of each output element of the dataset to each fuzzy set B_i. Once a partition of the output space in fuzzy clusters B_i is obtained, a projection of these clusters on the input space is carried out, obtaining a fuzzy set in $\Re_n$ as result, which projected on each axis assigns to each input parameter x_i a fuzzy set A_i (7), as illustrated in Figure 1.

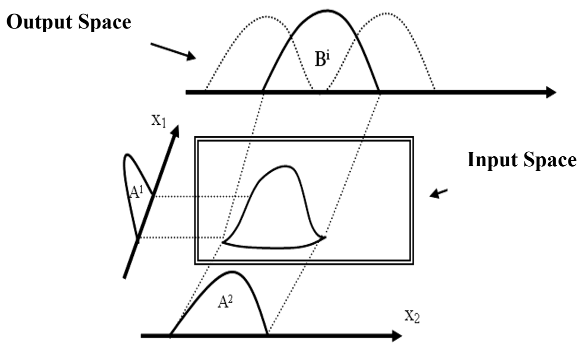

Fig. 1: Projection of fuzzy clusters

The number of candidate fuzzy rules that are generated by means of this methodology is based on the approach proposed by Fukuyama in [16], which basically consists of measuring the variance of the elements in a cluster and the variance between clusters. In this way, the optimum number of clusters is the value that minimizes the variance in each cluster and maximizes the variance between them. In this paper, the number of rules generated were empirically determined and we determined to use two or four times the optimum calculated with these criteria depending on the problem, with the aim of starting the next evolutionary selection step with a number of rules large enough to let the algorithm perform the search process better, which is based on selecting a subset with high cooperation, as we describe in the next subsection.

It is important to point that this methodology proposes the partitions of the variables in the universe of discourse quite differently from most methods for generating fuzzy KBs that use a uniform partition independently of the dataset to be described. It also creates rules together with the definition of the membership functions, instead of producing rules using predefined membership functions.

As commented at the beginning of Section II, the result of the rule generation process in this point is an approximative fuzzy system because each individual rule has its own labels for each variable. Moreover, membership functions are of trapezoidal shape. Figure 2 shows the approximate KB obtained at this point of the procedure.

However, we want to achieve a descriptive, more interpretable fuzzy system. To do so, we transform the output of the clustering process in the way described here. In this paper, we have opted to assign triangular fuzzy labels (antecedent and consequent). Therefore, we convert the trapezoidal membership functions into triangular labels, where the amplitude of the base of the triangle corresponds to the largest base of the trapezoid.

The next step is to obtain a descriptive fuzzy system from the approximative one, that is, to reduce the number of membership functions we have for each variable to a fixed number, previously defined by the user, number of membership functions. To do so, an algorithm based on the K-nearest neighbour is employed. Hence, we obtain a prefixed number of different groups of triangular membership functions for each variable, and each of these groups will be represented by a single triangle whose position and base are calculated as the average of their represented triangular fuzzy labels. In this way, the approximative fuzzy KB has been transformed into a descriptive one derived from it. Figure 3 shows an example of a RB obtained following this method from the RB of Figure 2. It can be seen that the first rule in the first variable uses the same label as the third rule, and likewise the second with the seventh and the fourth with the fifth, and so on.

The transformation of the approximate system rules into a descriptive one involves an approximation, so each individual rule loses accuracy in this process. This is partially offset by the membership functions tuning performed in the next step of the methodology.

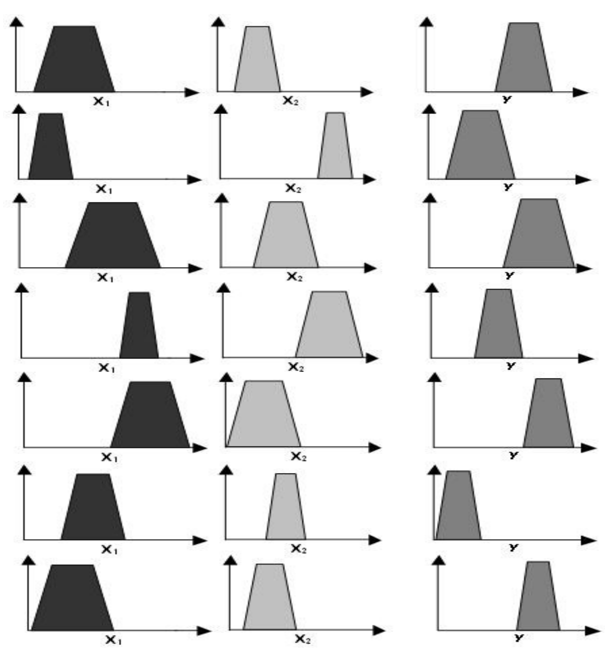
Fig. 2: Example of candidate rules by means of an approximative system

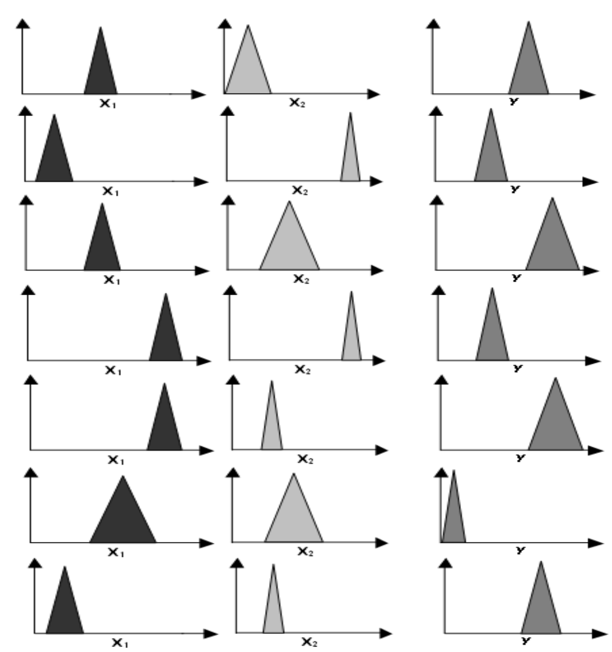
Fig. 3: Example of candidate rules by means of a descriptive system

B. Genetic Selection and Tuning of the Candidate Rules Set

The set of rules and membership functions of each variable involved that were generated in the previous step are used in this second step to find the final set of rules, and to tune the membership functions.

The method used in this paper is similar to the one presented in [6]. It uses an evolutionary algorithm in which both elements, candidate rules and membership functions associated with each variable, are encoded in each chromosome. Thus, the scheme selected is a dual coding (CS_S+CS_T), shown in Figure 4, where:

1) CS_S encodes the rules to be selected: This is carried out using a binary string of N genes, each one representing a candidate rule from the clustering process performed in the first step of the proposed methodology. The initial set of candidate rules are good rules individually, but can be redundant rules, conflicting, or in general rules with low level of cooperation. The selected rules will have the value "1" on their corresponding gene in the chain, whereas "0" means the opposite. This way of selecting a rule set allows the evolutionary algorithm to choose a subset of rules with a higher level of cooperation between them [3], that is, rules that operate well together in the fuzzy rule-based system with defuzzifier where interaction between rules is one of the main points.

2) CS_T encodes the membership functions to be tuned: Using the so-called lateral tuning [4], which uses only the displacement and the amplitude of the membership functions. This method for tuning membership functions has some advantages over classical tuning [5]: It enables the position and amplitude of membership functions to be found more easily, thanks to the use of a smaller number of parameters, two against three for triangular shape membership functions, and more interestingly, lateral tuning keeps the symmetry of the labels, benefiting the interpretability of the resulting system. The lower number of degrees of freedom may seem to be a drawback, but a lower number of parameters actually lets us find better solutions easily [4].

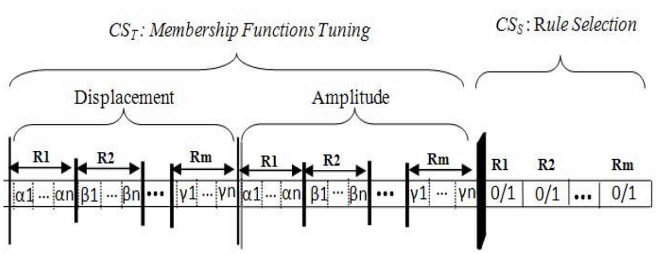

Fig. 4: Scheme Dual Coding (CS_S+CS_T)

The evolutionary model employed in this work is derived from CHC model [17]. This model has a good balance between exploration and exploitation, making it a good choice for problems with complex search spaces. The evolutionary model is characterized in that each generation uses a parent population of size M to generate an intermediate population of M individuals. These individuals will be randomly paired and used to generate M' descendants, giving rise to a competition for survival, where the best M chromosomes from parents and offspring populations are selected to form the new generation.

No mutation is applied during the recombination phase. Instead, when the population converges or the search stops making progress (i.e., the difference threshold has dropped to zero and none of the newly generated offspring are better than any member of the parent population), the population is reinitialized. The restarted population completely consists of random individuals except for one of them, which must be the best individual found so far.

Although the CHC algorithm was designed for binary-encoded chromosomes, there are versions for use with real-encoded chromosomes. This is the one used in this work for the chromosome part that contains the lateral tuning of membership functions. In these cases, the BLX-α crossover (α=0.5) is used in order to recombine the parent's genes. This produces two descendants for each pair of parents, so that the offspring generated by this crossover operator is the same size as the initial population. The Hamming distance is computed by translating the real-coded genes into strings and by taking into account whether each character is different or not. Only those string pairs which differ from each other by a number of bits (mating threshold) are mated. The initial threshold is set to L/4 where L is the length of the string. When no offspring is inserted into the new population, the threshold is reduced by 1.

The learning of both elements simultaneously, rule selection and lateral tuning of membership functions, lets us obtain not only a set of rules with a good level of cooperation, but at the same time their corresponding best combination of membership function tuning to achieve good precision. Figure 5 illustrates an example of the final RB after the last step considering that they have been tuned only (none of them were unselected).

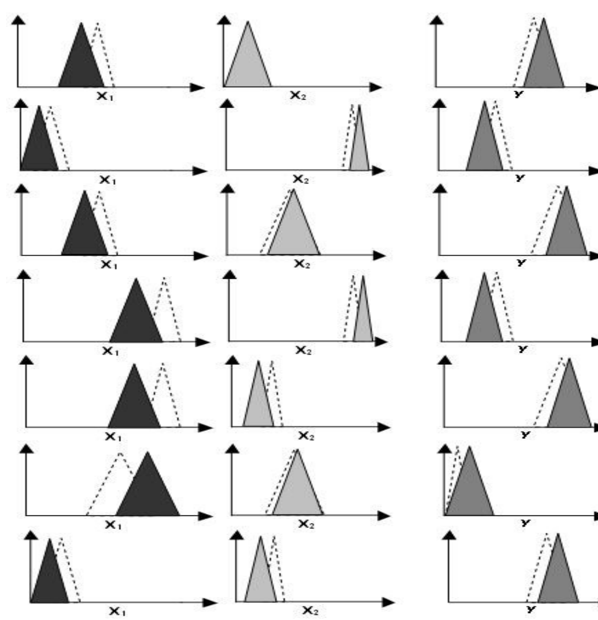

Fig. 5: Example of final rules after genetic selection and tuning

The objective to be minimized by the evolutionary algorithm is the classic Mean Square Error (MSE), which is a standard average accuracy performance whose expression is (8):

$$ECM\ (S)_B = \frac{\frac{1}{2}\sum_{k=1}^{P} (\, y_k - S\,(x_k)\,)^2}{P}, \qquad (8)$$

Where S is the fuzzy model, P is the number of pairs in the dataset $Z_k = (x_k, y_k)$, k=1,..,P, with x_k the values of input variables and y_k their associated values corresponding to the output variable.

III. Experimental Study Carried Out

We shall describe the experimental study developed to show the usefulness of the proposed methodology, CEST. To do so, we selected four data sets and compared the results obtained by CEST with those obtained with a well known reference method, WM [1] and also with COR [3], which is also a well known methodology that offers a set of rules with a good level of cooperation between them instead of the best individual rules.

Taking into account that CEST methodology includes a second step that reduces the rule set and tunes the membership functions, and in order to carry out the comparative study in similar conditions, we decided to use WM and COR, adding to them the second step of CEST, which we name "EST" (Evolutionary Selection and Tuning), so the resulting methods compared with CEST will be WM+EST and COR+EST. The EST significantly increases the compactness and accuracy of the WM and COR conventional RBs.

A. Description of the Problems and Comparison Methodology

We selected four problems with different number of variables and amount of data for the experimental study (these data sets are available at http://www.keel.es/ [18]). They are as follows:

1) An electrical distribution problem (SE), [19]: Estimates the maintenance cost of medium voltage lines in a city. There are 1059 data items, 4 continuous input variables and one output. *2) The data set Weather in Ankara (WA)*, [20]: Contains the weather information for Ankara from 01/01/1994 to 05/28/1998. From a given set of features, the aim is to predict the mean temperature. There are 1609 data items, 9 continuous input variables and one output.

3) The data set Quake (Q): A regression data set where the task is to approximate the strength of an earthquake given the depth of its focal point, its latitude and longitude. There are 2178 data items, 3 continuous input variables and one output.

4) The data set Treasury (TR): Contains the Economic data information of the USA from 04/01/1980 to 04/02/2000 on a weekly basis. From a set of given features, the goal is to predict 1 Month CD Rate. There are 1049 data items, 15 continuous input variables and one output.

We used the average values of the MSE to measure the average accuracy for each model. We also used 5-cross validation, i.e. 5 random data partitions, each with 20% (4 with 211 examples and one with 212 for the SE problem, 4 with 322 examples and one with 321 for the WA problem, 4 with 435 examples and one with 438 for Problem Q and 4 with 210 examples and one with 209 for Problem TR). The combination of 4 of them (80%) is used as training and the fifth as test. We ran a total of 30 experiments for each evolutionary learning process, that is, we use 5 seeds for the random number generator and 6 partitions.

The corresponding RBs were generated for each partition. using the WM [1] and COR [3] methods. These methods generate about 65 rules in the case of problem SE, 156 in the problem WA, 54 in the problem Q and 75 in the problem TR. The aforementioned number of rules represents the values before the process of posting their RBs with EST.

In the case of the CEST methodology, empirically we chose to generate twice the optimum number of rules according to the criteria described in [19] for problem SE, i.e. an average of about 20 rules, an average of 66 rules for the WA problem, which is four times the optimum, and finally for problems Q and TR we use the optimum, which is 15 and 90 rules, respectively.

The population size of the CHC evolutionary algorithm is 50 individuals randomly initialized within their ranges of variation except one, initialized with the following configuration:

1) CS_S part of the chromosome (rule selection): All genes with "1", that is, all the rules provided by the clustering step are active at the beginning.

2) CS_T part, (membership functions tuning): With the initial position of the labels resulting from the previous stage, that is, the position computed by the K-nearest neighbour algorithm.

For all problems, the number of evaluations of the evolutionary algorithm used is 200.000, determined empirically by several previous tests.

The objective function of the evolutionary algorithm is the aforementioned of expression (7).

B. Results and Analysis

This section shows and analyses the results obtained in the experimental study.

The different tables from I to IV show the results corresponding to problems SE, WA, Q and TR respectively. The columns in these tables show, from left to right, the KB learning method, the average number of rules obtained by them, and the MSE in training and test. Likewise, we have included along with the MSE columns the result of applying the Student t-test (column t-test) with 95% confidence to the best average result of the corresponding column, comparing

1 to 1 with the rest of average results. The interpretation of this column is as follows:

* Indicates the result with the best average.

+ Indicates a significantly worse performance than the best.

TABLE I
RESULTS OBTAINED WITH THE SE PROBLEM.

Method	#R	MSE_{TRAIN}	t-test	MSE_{TEST}	t-test
WM	65	56135.75	+	56359.42	+
COR	65	50711.82	+	54584.76	+
WM+EST	35.5	19084.121	+	23256.302	+
COR+EST	41.8	15793.401	+	19781.520	+
CEST	10.8	12274.421	*	13749.437	*

TABLE II
RESULTS OBTAINED WITH THE WA PROBLEM.

Method	#R	MSE_{TRAIN}	t-test	MSE_{TEST}	t-test
WM	156	16.063	+	16.403	+
COR	156	7.150	+	7.357	+
WM+EST	63.8	5.5969318	+	5.8172814	+
COR+EST	57.3	1.8875130	+	2.9646900	+
CEST	41.3	0.0066863	*	0.0072634	*

TABLE III
RESULTS OBTAINED WITH THE Q PROBLEM.

Method	#R	MSE_{TRAIN}	t-test	MSE_{TEST}	t-test
WM	54	0.0458	+	0.0467	+
COR	54	0.0288	+	0.0289	+
WM+EST	34.2	0.022898968	+	0.0241609	+
COR+EST	41.5	0.017190032	+	0.0176492	+
CEST	12.7	0.015080329	*	0.0157390	*

TABLE IV
RESULTS OBTAINED WITH THE TR PROBLEM.

Method	#R	MSE_{TRAIN}	t-test	MSE_{TEST}	t-test
WM	75	1.636	+	1.632	+
COR	75	1.519	+	1.524	+
WM+EST	51.6	0.0269343	+	0.0444616	+
COR+EST	42.1	0.0170835	+	0.03616134	+
CEST	23.1	0.0040693	*	0.00521205	*

Viewing the results, we can analyse the following:

- Results obtained by methods WM and COR are comparatively worse than if we use the CEST methodology or add EST to WM and COR.
- The improvement of the proposed method CEST is clearly visible, improving the accuracy (reducing MSE) in training as well in test, and complexity (with few rules), and also independently of the problem.
- In the case of problem SE (Table I) and problem Q (Table III), the reduction of rules is very important, while the improvement in the MSE is only slightly better. Nevertheless, viewing problems WA (Table II) and TR (Table IV), the most important improvement is shown by the MSE, while the improvement in the number of rules is also important, but lower. Problems WA and TR have more variables, so we think that this may be the reason: when the problem has several variables, the improvements are greater in accuracy, whereas if the problem has few variables, the best improvements are obtained in the compactness of the RB. All in all, we consider that we must continue studying this effect in order to characterize it.

Finally, we wish to point out that the advantage of CEST, is not the evolutionary selection and tuning, since WM+EST and COR+EST are also using this step. Nor is the main point of CEST the fuzzy clustering, because the candidate rules generated by the first step are very good rules individually, but cannot be used directly in a fuzzy rule-based system with defuzzifier because they have a very high error rating, as a consequence of a lot of redundant rules and in general, due to a low cooperation between rules. CEST achieves good results because it joins a good rule candidate generation process with an especially competent method to select a good subset of rules and obtain a very good tuning of the membership functions associated with the set of rules selected.

IV. CONCLUSION

This work proposed a methodology to obtain fuzzy linguistic RBs from examples taking the number of labels desired for the granularity of the variables. The methodology is based on the combination of two strategies: the generation of good individual candidate rules based on fuzzy clustering, without prior definition of the membership functions, and the evolutionary selection of the rules together with tuning of the membership functions, in order to obtain what is possibly the best subset of cooperating rules and their associated membership tuning. KBs obtained in this way are compact and accurate, and the results obtained in the experimental study show that the complete design of KBs can improve accuracy and interpretability in the sense of complexity significantly more than post-processing methods such as selection and tuning alone.

In future works, we would like to find a relation between the number of rules to be generated in the clustering step and the data set features, and of course continue with the experimental study, using more and different data sets to validate the good results presented in this work. It could also be interesting to continue working in terms of interpretability: in this work we have simply considered the system complexity using the number of rules, but it is possible to define it using many other aspects and indexes [6, 7, 21, 22, 23, 24, 25] that can be considered and added.

REFERENCES

[1] L.X. Wang, J.M. Mendel. Generating fuzzy rules by learning from examples. IEEE Trans. on Systems, Man, and Cybernetics 22:6, Pág.1414-1427, 1992.

[2] P. Thrift. Fuzzy logic synthesis with genetic algorithms. Proceedings of the 4th International Conference on Genetic Algorithms Morgan Kaufmann, Pág. 509-513, 1991.

[3] J. Casillas, O. Cordón, F. Herrera. COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing

cooperation among rules. IEEE Trans.on Fuzzy Systems, 32:4, Pág. 526–537, 2002.
[4] R. Alcalá, J. Alcalá-Fdez, F. Herrera. A Proposal for the Genetic Lateral Tuning of Linguistic Fuzzy Systems and its Interaction with Rule Selection. IEEE Trans. on Fuzzy Systems 15:4, Pág. 616-635, 2007.
[5] C. Karr. Genetic algorithms for fuzzy controllers. AI Expert 6:2, Pág. 26-33, 1991.
[6] J. Casillas, O. Cordón, M.J. Del Jesus, F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability for linguistic modeling. IEEE Trans. on Fuzzy Systems 13:1, Pág. 13-29, 2005.
[7] F. Herrera. Genetic Fuzzy Systems: Taxonomy, current research trends and prospects. Evolutionary Intelligence, 1, Pág. 27-46, 2008.
[8] M. Setnes, R. Babuˇska, H. B. Verbruggen. Rule-Based Modeling: Precision and Transparency. IEEE Trans. on Systems, Part C: Applications and Reviews 28:1, Pág. 165-169, 1998.
[9] G.H. Ball, D.J. Hall. ISODATA, a novel method of data analysis and pattern classification. Technical report, Stanford Research Institute, 1965.
[10] J.C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3, Pág. 32-57, 1973.
[11] J.C. Bezdek. Pattern Recognition with fuzzy objective function algorithm. New York: Plenum Press, 1981.
[12] J.C. Bezdek, R. Ehrlich, W. Full. FCM: The Fuzzy c-Means Clustering Algorithm. Computers and Geosciences 10:2-3, Pág. 191-203, 1984.
[13] Hoppner, F., Klawonn, F. A contribution to convergence theory of fuzzy c means and derivatives. IEEE Trans. on Fuzzy Systems 11 (5). Pág. 682–694, 2003.
[14] Keller, A., Klawonn, F. Fuzzy clustering with weighting of data variables. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (6). Pág. 735–746, 2000.
[15] M. Sugeno, A. Yasukawa. A Fuzzy-Logic Based approach to qualitative Modeling. IEEE Trans. on Fuzzy Systems 1, Pág. 7-31, 1993.
[16] Y. Fukuyama, M. Sugeno. A new method of choosing the number of clusters for fuzzy c-means method. Proceedings of the 5th Fuzzy Systems Symposium, Pág. 247-250, 1989.
[17] L.J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. Foundations of genetic Algorithms 1, Pág. 265-283, 1991.
[18] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera. Keel: A software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13, Pag. 307-318, 2009.
[19] O. Cordón, F. Herrera, L. Sánchez. Solving electrical distribution problems using hybrid evolutionary data analysis techniques. Applied Intelligence 10, Pág.5-24, 1999.
[20] M. Erdem Kurul, E. Tuzun. Avaible:www.wunderground.com; http://www.wunderground.com/global/stations/17128.
[21] J. Valente de Oliveira. Semantic constraints for membership functions optimization. IEEE Trans. Systems Man and Cybernetics Part A. 29:1. Pág.128-138, 1999.
[22] S. Guillaume. Designing fuzzy inference systems from data: an interpretability-oriented review. IEEE Trans. on Fuzzy Systems. 9:3. Pág. 426-443, 2001.
[23] J.M. Alonso and L. Magdalena, "A Conceptual Framework for Understanding Fuzzy Systems" in Proc. of IFSA World Congress (IFSA-EUSFLAT'09) , Lisbon, Portugal, 2009, pp. 119-124.
[24] M. Zhou and J. Q. Gan, "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling", *Fuzzy Sets and Systems*, vol. 159, no. 23, pp. 3091–3131, 2008.
[25] J. Casillas, O. Cordón, F. Herrera and L. Magdalena. *Interpretability issues in fuzzy modeling*. Springer-Verlag, 2003.