A Multiobjective Genetic Algorithm for Feature Selection and Granularity Learning in Fuzzy-Rule Based Classification Systems*

O. Cordón, F. Herrera
Dept. Computer Science and A. I.
University of Granada
18071 - Granada, Spain
ocordon,herrera@decsai.ugr.es

M.J. del Jesus
Dept. Computer Science
University of Jaén
23071 - Jaén, Spain
mjjesus@ujaen.es

P.Villar
Dept. Computer Science
University of Vigo
32004 - Ourense, Spain
pvillar@uvigo.es

Abstract

In this contribution, we propose a new method to automatically learn the knowledge base of a Fuzzy Rule-Based Classification System (FRBCS) by selecting an adequate set of features and by finding an appropiate granularity for them. This process uses a multiobjective genetic algorithm and considers a simple generation method to derive the fuzzy classification rules.

1. Introduction

An FRBCS learning process must solve different problems to obtain a linguistic FRBCS with an accurate behaviour, such as:

1. Establish the granularity for the linguistic variables,
2. Obtain a fuzzy rule set with an adequate cooperation level between the rules,
3. Select the inference method, which determines the way of combining the information provided by the fuzzy rules in the classification of the examples.
4. Tune -if necessary- the fuzzy partitions for the linguistic variables, and
5. Reduce the number of input variables in high dimensional problems, in order to avoid an exponential growth in the fuzzy rule set size.

The problems 2, 3 and 4, related to the knowledge extraction process, have been solved by different learning processes based on iterative methods [4], Neural Networks [21] or Genetic Algorithms (GAs) [17, 5, 11], among others.

With respect to the first problem, the majority of methods learn the fuzzy rule set from numerical information using a predefined number of labels per variable. The usual way to proceed involves choosing a number of linguistic terms for each linguistic variable, which is normally the same for all of them. This operation mode makes the granularity and fuzzy set definitions have a significant influence on the FRBCS performance. In fact, some studies in Fuzzy Rule-Based Systems have shown that the system performance is much more sensitive to the choice of the semantics in the data base than to the composition of the rule base [7].

The fifth problem, high dimensionality with a large number of features, can be tackled from a double perspective:

- Via the compactness and reduction of the rule set, minimising the number of fuzzy rules included in it. Unnecessary rules can be eliminated with the aim of having a more co-operative rule set in order to obtain an FRBCS with better performance.
- Via a feature selection process that reduces the number of features used by the FRBCS.

Rule reduction methods have been formulated using Neural Networks, clustering techniques, orthogonal transformation methods and similarity measures [22], as well as using GA-based rule selection processes to get a co-operative set of rules from a candidate rule set [15, 5]).

Notice that, for high dimensional problems and problems where a high number of instances is available, it is difficult for the latter approaches to get small rule sets, and therefore the system comprehensibility and interpretability may not be as good as desired. For high dimensionality classification problems, a feature selection process, that determines the most relevant variables before or during the FRBCS inductive learning process, must be considered [3, 20]. It increases the efficiency and accuracy of the learning and classification stages.

Our objective is to develop a feature selection and granularity genetic learning process to obtain FRBCSs com-

*This research has been supported by CICYT PB98-1319

posed of a compact set of comprehensible fuzzy rules with high classification ability. This method uses a multiobjective GA [10] and considers a simple generation method to derive the rule base, the extension of Wang and Mendel's fuzzy rule generation method [24] for classification problems [4].

To carry out this task, this paper is organised as follows. In Section 2, the FRBCS components will be introduced. Section 3 will describe the two main problems tackled by the learning method, the feature selection and granularity learning. In Section 4 we will expose the characteristics of our proposal for the FRBCS design. The results obtained with Sonar data base will be shown in Section 5. In the last section, some conclusions will be pointed out.

2. Fuzzy Rule-Based Classification System

An FRBCS is an automatic classification system that uses fuzzy rules as knowledge representation tool. Two different components are distinguished within it:

1. The **Knowledge Base (KB)**, composed of:
    - a *Data Base* (DB), which contains the fuzzy set definitions related to the linguistic terms used in the fuzzy rules, and
    - a *Rule Base* (RB), comprised by a set of fuzzy rules that in this work are considered to have the following structure:

      R_k : If X_1 is A_1^k and ... and X_N is A_N^k then Y is C_j with r^k

      where $X_1, \ldots, X_N$ are features considered in the problem and $A_1^k, \ldots, A_N^k$ are linguistic labels employed to represent the values of the variables.

      These kinds of fuzzy rules represent, in the antecedent part, a subspace of the complete search space by means of a linguistic label for each considered variable and, in the consequent part, a class label (C_j) and a certainty degree (r^k). This numerical value indicates the degree of certainty of the classification in that class for the examples belonging to the fuzzy subspace delimited by the antecedent part.
2. The **Fuzzy Reasoning Method (FRM)**, an inference procedure which, combining the information provided by the fuzzy rules related with the example to classify, determines the class to which it belongs to.

The majority of FRBCSs (see [4, 11] among others) use the classical FRM that classifies a new example with the consequent of the fuzzy rule having the highest degree of association. Another family of FRMs that use the information provided by all the rules compatible with the example (or a subset of them) have been developed [4, 6, 16]. In this work, we use two different FRMs: maximum and normalised sum.

3. Feature Selection and Granularity Learning in an FRBCS design process

As we mentioned before, our FRBCS learning method generates the KB by selecting an adequate feature set and by finding an appropiate granularity for each selected variable. In this section, we briefly describe these problems jointly solved in our proposal.

3.1. Feature Selection Process

The main objective of any feature selection process is to reduce the dimensionality of the problem for the supervised inductive learning process. This fact implies that the feature selection algorithm must determine the best features for its design.

There are two kinds of feature selection algorithms:

- *Filter feature selection algorithms* [19], which remove the irrelevant characteristics without using a learning algorithm (e.g. by means of class separability measures). They are efficient processes but, on the other hand, the feature subsets obtained by them may not be the best ones for a specific learning process because of the exclusion of the heuristic and the bias of the learning process in the selection procedure [18].
- *Wrapper feature selection algorithms* [18, 19]. This kind of feature selection algorithms selects feature subsets by means of the evaluation of each candidate subset with the precision estimation obtained by the learning algorithm. In this form, they obtain feature subsets with the best behaviour in the classifier design. Their problem is their inefficiency since they must build the classifier for each evaluation of a candidate feature subset.

In our proposal we will use a wrapper feature selection algorithm which utilises the precision estimation provided by an efficient fuzzy rule generation process (Wang and Mendel's fuzzy rule generation process) and a GA as search algorithm. The granularity learning will provide us an additional way to select features when the number of linguistic labels assigned to a specific variable is only one (we will explain this in detail in the next section).

3.2. Granularity Learning

As previously said, the derivation of the DB highly influences in the FRBCS performance. Some approaches have been proposed to improve the FRBCS behaviour by means of a tuning process once the RB has been derived [1, 5]. However, these tuning processes only adjust the shapes of the membership functions and not the number of linguistic terms in each fuzzy partition, which remains fixed from the begining of the design process.

The methods that try to learn an appropiate granularity level per variable usually work in collaboration with an RB derivation method. A DB generation process wraps an RB learning one working as follows: each time a DB has been obtained by the DB definition process, the RB generation method is used to derive the rules, and some type of error measure is used to validate the whole KB obtained.

The works proposed in [7, 8] use Simulated Annealing and GAs to learn an appropiate fuzzy partition granularity for each variable in a Fuzzy Rule-Based System.

On the other hand, the method proposed in [14] deals with a GA to design an FRBCS. The coding scheme generates binary chromosomes of fixed length, with a segment per variable. Each segment has a pre-defined length which determines its maximum granularity. In the chromosome, an 1 indicates the center of a triangular membership function, and both extremes of the neighbour membership functions. This representation imposes several constraints in order to preserve the readability of the final FRBCS. Finally, a data covering algorithm is run to obtain the class associated to each antecedent combination.

4. Genetic Algorithm for Feature Selection and Granularity Learning

In this section, we propose a new learning approach to automatically generate the KB of a FRBCS composed of two methods with different goals:

- A genetic learning process for the DB that allows us to define:
  - The relevant variables for the classification process (feature selection).
  - The number of labels for each variable (granularity learning).

  Once the feature set and the granularity for each feature are determined, a uniform partition with triangular membership functions is considered due to its simplicity.

- A quick *ad hoc data-driven method* that derives the fuzzy classification rules considering the DB previously obtained. In this work we use the extension of Wang and Mendel's fuzzy rule generation method [24] for classification problems [4], but other efficient generation methods can be considered.

We should note that the granularity learning allows us another way of feature selection: if a variable is assigned only to one label, it has no influence in the RB, so it will not be considered as a relevant variable. A similar double-level feature selection process has been previously considered in genetic learning processes of FRBCSs such as SLAVE [11]. In [12], the authors presented a GA encoding single fuzzy rules using two binary strings with two differents information levels: the features involved in the rule (feature string) and the linguistic labels associated to each of them (value string). A feature is not considered when: i) it has a 0 bit in the first string, or ii) it has no label associated in the second. In that work, it is demonstrated that, although the feature selection process can be performed by only considering the value string (as in our case with the granularity string), it is more difficult for the learning algorithm to remove a feature working in this way and the consideration of the feature string helps it to achieve better results more quickly.

On the other hand, the main purpose of our KB design process is to obtain FRBCSs with good accuracy and high interpretability. Unfortunately, it is not easy to achieve these two objectives at the same time. Normally, FRBCSs with good performance have a high number of selected variables and also a high number of rules, thus presenting a low degree of readability. On the other hand, the KB design methods sometimes lead to a certain overfitting to the training data set used for the learning process.

To avoid these problems, our genetic process uses a multiobjective GA with two goals:

- Minimise the classification error percentage over the training data set.
- Design a compact and interpretable KB. This objective is performed by penalising FRBCSs with a large number of selected features and high granularity.

The next four subsections describe the main components of the genetic learning process.

4.1. Encoding the DB

Each chromosome will be composed of two parts to encode the relevant variables and the number of linguistic terms for variable (i.e. the granularity):

- Relevant variables (C_1): For a classification problem with N variables, the selected features are stored in a binary coded array of length N. In this array, an 1 indicates that the correspondent variable is selected for the FRBCS.
- Granularity level (C_2): The number of labels per variable is stored in an integer array of length N. In this contribution, the possible values considered are taken from the set $\{1, \ldots, 5\}$.

If v_i is the bit that represents whether the variable i is selected and l_i is the granularity of variable i, a graphical representation of the chromosome is shown next:

$$C_1 = (v_1, v_2, \ldots, v_N) \quad C_2 = (l_1, l_2, \ldots, l_N)$$
$$C = C_1 C_2$$

4.2. Initial Gene Pool

The initial population is composed of four parts. The generation of the initial gene pool is described next:

- In the first group all the chromosomes select all the features, that is, $C_1 = (1, 1, 1, \ldots, 1)$, and each one of them has the same granularity in all its variables. This group is composed of $\#val$ chromosomes, with $\#val$ being the cardinality of the significant term set, in our case $\#val = 4$, corresponding to the four possibilities for the number of labels, $2 \ldots 5$, (one label is not considered because the variable would not be selected). For each number of labels, one individual is created.
- The second part is composed of $\#val \times 4$ chromosomes and each one of them has the same granularity in all its variables. For each possible number of labels, four individuals are created, each one of them with a different percentage of randomly selected variables (75%, 50%, 25% and 10%).
- The third group has five subgroups, each one of them with the different percentages for the selected variables considered in the previous groups (100%, 75%, 50%, 25% and 10%), and all of the chromosomes with a randomly selected granularity per variable. In the experiments all of these five subgroups have 10 chromosomes.
- The fourth part is composed for the remaining chromosomes, and all of their components are randomly selected.

4.3. Evaluating the chromosome

There are three steps that must be done to evaluate each chromosome:

- Generate the DB using the information contained in the chromosome. For all the selected variables ($v_i = 1$ and $l_i > 1$), a uniform fuzzy partition is built considering the number of labels of the variable (l_i).
- Generate the RB by running a fuzzy rule learning method considering the DB obtained.
- Calculate the values of the evaluation function:
  - CPE: classification percentage error over the training set.
  - $SV \cdot AL$: with SV being the number of selected variables and AL being the granularity average of the selected variables.

4.4. Genetic operators

The following operators are considered.

4.4.1 Selection

We have used the selection mechanism of MOGA [9], which is based on the definition of Pareto-optimality. It is said that a solution dominates another when the former achieves better or equal values than the latter in all but one objective, where the former outperforms the latter. Hence, the pareto is composed of all the non-dominated solutions to the problem.

Taking this idea as a base, MOGA assigns the same selection probability to all non-dominated solutions in the current population. The method involves dividing the population into several classes depending on the number of individuals dominating the members of each class.

Therefore, the selection scheme of our multiobjective GA involves the following five steps:

1. Each individual is assigned a rank equal to the number of individuals dominating it plus one (chromosomes encoding non-dominated solutions receive rank 1).

2. The population is increasingly sorted according to that rank.

3. Each individual is assigned a selection probability which depends on its ranking in the population, with lower ranking receiving lesser probabilities.

4. The selection probability of each equivalence class (group of chromosomes with the same rank, i.e., which are non-dominated among them) is averaged.

5. The new population is created by following the Baker's stochastic universal sampling [2].

4.4.2 Crossover

The crossover works in the two parts of the chromosome at the same time. Therefore, an standard crossover operator is applied over C_1 and C_2. This operator performs as follows: a crossover point p is randomly generated in C_1 and the two parents are crossed at the p-th variable in C_1. The crossover is developed this way in the two chromosome parts, C_1 and C_2, thereby producing two meaningful descendents.

4.4.3 Mutation

Two different operators are used, each one of them acting on different chromosome parts. A brief description of them is given below:

- *Mutation on C_1*: As this part of the chromosome is binary coded, a simple binary mutation is developed, flipping the value of the gene.

- *Mutation on C_2*: The mutation operator selected for C_2 is similar to the one proposed by Thrift in [23]. A local modification is developed by changing the number of labels of the variable to the immediately upper or lower value (the decision is made at random). When the value to be changed is the lowest (1) or highest one, the only possible change is developed.

5. Experimentation

We have applied the learning method to an example base with a high feature number, Sonar data set [13], which has 208 instances of a sonar objective classification problem. Each one of these instances is described by 60 features to discriminate between a sonar output corresponding to a cylindrical metal or an approximately cylindrical rock. The training set contains 104 elements and the test set contains 104 elements, randomly selected from the whole data set.

Table 1 shows the parameter values considered for the experiments developed.

Parameter	Value
Granularity values	$\{1, \ldots, 5\}$
Population size	100
Crossover probability	0.6
Mutation probability	0.2
Number of generations	$\{100, 500\}$

Table 1. Parameter values

The best results obtained by our genetic learning process for the two FRMs considered are shown in Table 2. The best results found with the Wang and Mendel's RB generation method considering all the features selected and the same number of labels for each one of them are also shown in the top line of each FRM. The table contains the following columns:

- **FRM**: Fuzzy Reasoning Method used.
- **SV**: Number of selected variables
- **AL**: Average of the number of labels considered for the selected variables.
- **NR**: The number of rules of the FRBCS RB.
- **% tra**: Classification percentage error obtained in the training data set.
- **% tst**: Classification percentage error obtained in the test data set.

FRM	SV	AL	NR	% tra	% tst
Maximum	60	3	104	0.9	23.1
	11	3.7	101	0.0	12.5
	8	3.8	98	0.9	16.3
	7	3.7	86	2.8	18.2
	6	3.6	81	8.6	13.4
	5	3.8	75	10.5	17.3
	4	4.2	63	14.4	18.2
	3	3.6	24	25.0	18.2
	2	4.0	19	23.1	24.0
Normalised Sum	60	3	104	2.8	25.9
	6	4.5	101	0.9	15.3
	5	4.6	97	1.9	14.4
	5	4.2	67	10.5	18.2
	4	4.5	66	11.5	19.2
	3	4.3	42	19.2	23.1

Table 2. Best results obtained (% error)

As it can be observed, the proposed method achieves a significant reduction in the number of variables selected (about the 90% of the original number of features, or even more in some cases) even with an important increase of the classification rate. Besides, many solutions present also a significant decrease in the number of rules, reducing the complexity of the KB. Therefore, our multiobjective GA provides a wide set of solutions that permit an adequate choice depending on the main goal required: good performance or high degree of interpretability.

6. Conclusions

This contribution has proposed a multiobjective genetic process for jointly performing feature selection and granularity learning, which is combined with an efficient fuzzy classification rule generation method to obtain the complete KB for a descriptive FRBCS. Our method achieves an important reduction of the relevant variables selected for the final system and also adapts the granularity of each variable to the problem being solved. So, we can conclude that the proposed method allows us to significantly enhance the interpretability and accuracy of the FRBCSs generated.

References

[1] S. Abe, M.-S. Lan, and R. Thawonmas. Tuning of a fuzzy classifier derived from data. *International Journal of Approximate Reasoning*, 12:1–24, 1996.

[2] J. E. Baker. Reducing bias and inefficiency in the selection algorithms. In *Proc. of the Second International Conference on Genetic Algorithms (ICGA'87)*, pages 14–21, Hillsdale, 1987.

[3] J. Casillas, O. Cordón, M. del Jesus, and F. Herrera. Genetic feature selection in a fuzzy rule-based classification system learning process for high dimensional problems. To appear in Information Sciences, 2001.

[4] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with applications to image processing and pattern recognition.* World Scientific, 1996.

[5] O. Cordón, M. J. del Jesus, and F. Herrera. Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods. *International Journal of Intelligent Systems*, 13(10/11):1025–1053, 1998.

[6] O. Cordón, M. J. del Jesus, and F. Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 20(1):21–45, 1999.

[7] O. Cordón, F. Herrera, and P. Villar. Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing. *International Journal of Approximate Reasoning*, 25(3):187–216, 2000.

[8] O. Cordón, F. Herrera, and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. To appear in IEEE Transactions on Fuzzy Systems, 2001.

[9] C. Fonseca and P. Fleming. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In S. Forrest, editor, *Proc. of the Fifth International Conference on Genetic Algorithms (ICGA'93)*, pages 416–423. Morgan Kaufmann, 1993.

[10] C. Fonseca and P. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.

[11] A. González and R. Pérez. SLAVE: A genetic learning system based on an iterative approach. *IEEE Transactions on Fuzzy Systems*, 7(2):176–191, 1999.

[12] A. González and R. Pérez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transctions on Systems, Man and Cybernetics - Part B: Cybernetics*, 31, 2001.

[13] R. P. Gorman and T. J. Sejnowski. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1:75–89, 1988.

[14] H. Ishibuchi and T. Murata. A genetic-algorithm-based fuzzy partition method for pattern classification problems. In F. Herrera and J. Verdegay, editors, *Genetic Algorithms and Soft Computing*, pages 555–578. Physica-Verlag, 1996.

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

[16] H. Ishibuchi, T. Nakashima, and T. Morisawa. Voting in fuzzy rule-based systems for pattern classification problems. *Fuzzy Sets and Systems*, 103:223–238, 1999.

[17] H. Ishibuchi, K. Nozaki, and H. Tanaka. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Sets and Systems*, 65:237–253, 1994.

[18] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

[19] H. Liu and H. Motoda. *Feature selection for knowledge discovery and data mining.* Kluwer Academic Publishers, 1998.

[20] T. Nakashima, T. Morisawa, and H. Ishibuchi. Input selection in fuzzy rule-based classification systems. In *Proc. of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'97)*, pages 1457–1462, 1997.

[21] D. Nauck and R. Kruse. A neuro-fuzzy method to learn fuzzy classification rules from data. *Fuzzy Sets and Systems*, 89:277–288, 1997.

[22] M. Setnes, R. Babuska, U. Kaymak, and H. van Nauta-Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transctions on Systems, Man and Cybernetics - Part B: Cybernetics*, 28:376–386, 1998.

[23] P. Thrift. Fuzzy logic synthesis with genetic algorithms. In *Proc. Fourth International Conference on Genetic Algorithms (ICGA'91)*, pages 509–513, 1991.

[24] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(2):353–361, 1992.