Niching genetic feature selection algorithms applied to the design of fuzzy rule-based classification systems

José Joaquín Aguilera, Manuel Chica, María José del Jesus and Francisco Herrera

***Abstract*— In the design of Fuzzy Rule-Based Classification Systems (FRBCSs) a feature selection process which determines the most relevant features is a crucial component in the majority of the classification problems. This simplification process increases the efficiency of the design process, improves the interpretability of the FRBCS obtained and its generalization capacity. Most of the feature selection algorithms provide a set of variables which are adequate for the induction process according to different quality measures. Nevertheless it can be useful for the induction process to determine not only a set of variables but also different set of variables. These sets of variables can be used for the design of a set of FRBCSs which can be combined in a multiclassifier system, improving the prediction capacity increasing its description capacity.**

In this work, different proposals of niching genetic algorithms for the feature selection process are analyzed. The different sets of features provided by them are used in a multiclassifier system designed by means of a genetic proposal. The experimentation shows the adaptation of this type of genetic algorithms to the FRBCS design.

I. INTRODUCTION

The Fuzzy Rule-Based Classification Systems (FRBCSs) are composed of a set of fuzzy rules and a fuzzy reasoning method that generalizes the extracted knowledge from the data in order to classify new data. Traditionally, the main objective in the design of this type of systems has been maximizing precision, although nowadays interpreting the generated fuzzy rules set is becoming more important [6]. Some of the determining factors to interpret a set of rules are the type and number of fuzzy rules, the definition of the fuzzy sets and the number of variables implied in each rule.

When the classification problem has a high number of variables, the design of an interpretable FRBCS is more difficult because of the dimension problem, since an increment in the number of variables implies an exponential increase in the fuzzy rules search space.

In the specialized literature, two solutions have been considered for the problem of high dimensionality in the learning of FRBCSs:

- A one-step design of a post-processing phase to compact and reduce a group of rules previously obtained ([1], [10]).
- To carry a feature selection process out to determine the most relevant variables before or during the inductive learning process of the system. We are going to face up to this problem from this point of view.

José Joaquín Aguilera, Manuel Chica and María José del Jesus are with the Department of Computer Science, University of Jaén, Paraje Las Lagunillas s/n, Jaén, Spain (phones: +34 953 212 879 and +34 953 212 444; emails: {jjaguile, mchica, mjjesus}@ujaen.es).

Francisco Herrera is with the Department of Computer Science and Artificial Intelligence, University of Granada, Periodista Daniel Saucedo Aranda, s/n, Granada, Spain (phone: +34 958 240 598; email: herrera@decsai.ugr.es).

A feature selection process can be defined as a search process of a subset of the complete set of variables with the objective of removing irrelevant and/or redundant features, and to obtain a simpler classification system with greater precision in the classification. This problem has been solved by means of Genetic Algorithms (GAs) [15] in different proposals ([3], [7], [26], [30], [34]).

The feature selection problem has a strong multimodal character, since multiple optimums (local or global) can exist, and a standard GA has some difficulties to obtain adequate results without being trapped in a partial good solution. Besides, in some problems, it is convenient to obtain not only a set of adequate characteristics but also the different optimum subsets. In this work, a study on the use of different proposals of niching GAs [27] is presented for feature selection processes prior to the FRBCS design. Then, with the sets of features obtained a genetic design of a multiclassifier system is analyzed.

To do so, this article is organized in the following way: in section II some preliminary concepts are introduced. The different proposals of niching GAs for feature selection are presented in Section III. In section IV the design of our multiclassifier system is described and the experimentation carried out with an analysis of the results obtained is shown in Section V. Finally, in section VI the conclusions are exposed and future works are presented.

II. PRELIMINARY

In this section preliminary concepts related with the four topics of this proposal are described: FRBCSs, feature selection processes, niching genetic algorithms and multiclassifier systems.

A. *Fuzzy rule-based classification systems*

An FRBCS is an automatic classification system that use fuzzy rules as a knowledge representation tool. Two main components can be distinguished in it:

- The knowledge base, formed by the database (where the semantic of the fuzzy sets is defined) and the fuzzy rule set.
- A fuzzy reasoning method: it combines the rules with the example to classify and determines the class which it belongs to.

In the specialized bibliography, most of the FRBCSs use, as fuzzy reasoning method, either the maximum or the sum approaches [11]. In the first case, the class of the new

1-4244-1210-2/07/$25.00 ©2007 IEEE.

example is that what provides the rule that has a greater association degree with the example. In the second one, all the matching rules with the example are taken into account and it will be classified by the most voted class.

B. Feature selection

The feature selection problem can be defined as a search process of P features from an initial set S of N variables, with $P <= N$. It aims to eliminate irrelevant and/or redundant features and to obtain a simpler classification system. Also it can get a better accuracy during the classification process the one designed by using all the features [20], [23].

Besides the search algorithm, in a feature selection algorithm an important component is the evaluation function that provides a measure of the quality for the feature subsets. Depending on this function, the feature selection algorithms can be classified in: *filter* models that use evaluation measures based on separability of classes; and *wrapper* models that use an estimation of the precision in the classification process (designing a classification system from the selected variables).

Genetic algorithms have been developed for feature selection by using both filter and wrapper approaches ([3], [7], [16], [26], [30], [34]).

The niching genetic proposals which are presented and analyzed in this work are based on the wrapper approach, because they use as a part of the evaluation function the estimated precision obtained by an FRBCS produced by means of a simple fuzzy rule generation method: the extension of the *Wang* and *Mendel* method [31] for classification problems [9]. An alternative way is to use the performance of a Case-Based Reasoning as assessment of the reliability of a feature subset [32] or the accuracy provided by the k-NN rule [7].

C. Niching genetic algorithms

The feature selection problem has a multimodal character because multiple optimum solutions could been found in the search space. Therefore, in this type of problems, a standard evolutionary process can cause the premature convergence leaving the exploration of the rest of the search space [15], [24].

Niching GAs preserves the population diversity by using niches techniques [24], [29] which divide the population in different niches. In this way the solutions in each different area or niche can survive during the evolutionary process independently of their global quality. This approach helps to maintain the necessary diversity and to get the different optimum solutions of a multimodal problem.

Different proposals of niching GAs are presented in the specialized bibliography ([4], [14], [18], [22], [27], [28]). The most used ones are the spatial niching GAs which promotes the creation of population niches in the same GA execution. In this category the following algorithms must be highlighted:

1) *Sharing* algorithms, based on organizing the population individuals in niches penalizing the quality of the individuals according to their proximity to other solutions. Examples of this GAs are *fitness sharing* [15] and *continuously updated sharing* [25], among others.
2) *Crowding* algorithms, based on replacement schemes by proximity. Examples of this type of algorithms are the *deterministic crowding* [24] and the *multi-niches crowding* [8].
3) *Clearing* algorithms. They modify the quality of the individuals to create niches in which survive a certain number of solutions, the rest disappear. *Clearing* [28] is the most representative example of this type of GAs.

In the specialized literature, niching GAs have been used for feature selection. Particularly, methods have been developed based on *deterministic crowding* and *sharing* models. In [5] a *deterministic crowding* binary coded GA was applied for the selection of features in a Chilean wines classification problem. A GA with *fitness sharing*, *clustering*, and random immigrant applied to the feature selection problem is described in [19].

In this work, three evolutionary proposals of feature selection have been selected to be developed and analyzed: *fitness sharing*, *clearing* and *deterministic crowding*, described in Section III.

D. Multiclassifiers

In a multiclassifier system a set of classifiers $D = D_1, D_2, ..., D_L$ is combined in order to increase the classification accuracy. In this kind of systems the final classification will not be obtained by only one classifier, it will be taken by a group of them.

Different proposals on multiclassifier systems are presented in the specialized bibliography [21]. These systems differ on the aggregation procedure, the type of classifier output or the type of each single classifier. Apart from the combination paradigms, the individual classifiers which could be used in a multiclassifier can be built using different subsets of features [33], different subsets of data [2] or different types of classifier models.

Two types of combination can be considered in the multiclassifier design: classifier selection [12] and classifier fusion [33]. The presumption in classifier selection is that what each classifier is an expert in some local area of the feature space. On the other hand, using classifier fusion assumes that the classifiers are trained over the whole data set, and they are considered as *competitive* rather than *complementary*. These two ideas can be merged. Instead of nominating one classifier, we can nominate a small group of them. Then we can give them a weight and take the output of the classifier which has the highest rate or use for the classification the majority vote [33], among other techniques.

In this paper, the multiclassifier is designed with FRBCSs learnt using different sets of features (obtained with the niching genetic feature selection proposals described in Section III). The FRBCSs are combined by means of a fusion in which the individual FRBCSs are weighted (and also selected) with a genetic algorithm described in Section IV.

III. Niching genetic proposals for feature selection

In this section, three niching genetic proposals for feature selection based on *fitness sharing*, *clearing* and *deterministic crowding* are described.

A. Common elements of the proposals

The proposals have the following common elements:

- Binary code. The length of the chromosome is equal to the maximum number of features for each problem.
- A wrapper approach. The estimation of the accuracy has been used as evaluation measure. It is given by an FR-BCS obtained from the subset of variables represented in the chromosome and using a simple fuzzy rules induction method: the extension of Wang and Mendel method [31] for classification problems [9]. In order to get a balance between accuracy and dimensionality reduction the fitness function is the following:

$$fitness(z) = \lambda \cdot acc(z) - (1-\lambda) \cdot \frac{feats(z)}{total\_feat} \quad (1)$$

  where *acc(z)* represents the accuracy for the FRBCS learnt by means the subset of features coded in the individual z, *feats(z)* is the number of features represented in the chromosome z and *total_feat* is the maximum number of features for the problem, i.e. N.
- Niching GAs use a distance measure to know if two individuals, i and j, belong or not belong to the same niche ($d(i,j)$). According to the codification used in these proposals, the Hamming distance is considered.
- The feature selection GA that uses sharing and clearing is a generational GA and uses the binary tournament as a selection operator.
- All the proposals include the one-point crossover operator and a simple mutation (it changes one gene at random).

In the following subsections the different elements of each proposal are described.

B. Sharing-GA for feature selection

This algorithm was introduced by Holland [17] and improved by Goldberg and Richardson [14]. The classical sharing method is based on modifying the search space of the GA and penalizing solutions belonging to zones of the population with many individuals to favor the exploration of other less populated zones. The population level of a zone is determined by the number of solutions which are within an specified area (given by the niche radius σ).

A modified fitness function is used to promote the evolution of these independent niches. This function is equal to the original fitness function (1) divided by the value of the niche to which belongs to:

$$f'_i = \frac{f_i}{m_i} \quad (2)$$

where m_i is calculated by adding the function sharing of all the members of the population:

$$m_i = \sum_{j=1}^{L} sh(d_{i,j}) \quad (3)$$

L represents the size of the population, and $d(i,j)$ the distance between the individuals i and j. The function $sh()$ gets a value that represents the similarity degree between two individuals:

$$sh(d_{i,j}) = \begin{cases} 1-(d_{i,j}/\sigma)^{\alpha} & \text{if } d_{i,j} < \sigma \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

σ denotes a threshold that shows the difference between two individuals to know if they belong, or not, to the same niche, and α is the slope of the *sharing* function whose most used value is 1.

This *sharing* method has been included in the evolutionary process, after all individuals are evaluated and before applying the selection operator.

C. Clearing-GA for feature selection

The *clearing* algorithm was proposed by Pétrovsky in [28] and is based on the observation that in natural life the available resources are limited and are different among species. This allows to create biodiversity and reduces the competition among the individuals of different species permitting the cohabitation in the same area. According to this, a clearing-GA limits the selective pressure modifying the population after the evaluation phase but before doing the selection. The goal is to maintain the dominant individuals (with the best fitness values) in each niche.

The basic *clearing* algorithm maintains the fitness of the dominant individuals and eliminates the others. However, the algorithm can be generalized to permit the inclusion of various individuals in the same niche (the capacity of a niche κ).

The algorithm works sorting decreasingly according to their fitness. The first individual is the dominant one, because there are no better individuals than it. Only the κ-better individuals of each niche survive. The fitness of the remaining individuals is set to 0. Then, the process will be repeated but only with the individuals whose fitness is greater than 0. The rest of the evolutionary process continues with this population.

D. Crowding-GA for feature selection

The *crowding* GA uses a replacement scheme based on the similarity to promote the formation of niches. The *deterministic crowding* GA, proposed by Mahfoud [24], randomly split up the population in reproductive couples. When the descendants are obtained, each one of them competes against one of its parents in a tournament to determine which individual will survive in the following generation. The competition is carried out in the following way: each father competes with its most equal descendant. The survival is obtained according to its quality (eq. 1).

IV. Fuzzy Multi-classifiers via the Genetic Learning of Weights

With the niching GAs designed we can obtain not only the best solution of the solution space, but a group of the best solutions: one or more solutions per niche. Applying the multiclassifier paradigm to the FRBCSs built with these best feature subsets we can obtain a multiclassifier with higher performance.

To do so, the classifier fusion with the majority vote [33] is used: each classifier assigns a single class label to the instance x, i.e., the classifier votes for the class. The final class label of x is the most voted.

In order to ponderate the output of each FRBCS we have associated a weight to each one of them by means a GA. Therefore, our multiclassifier is composed of a group of FRBCSs with their corresponding weights:

$$multiclassifier = \omega_1 \times clasif_1 + \omega_2 \times clasif_2 + ... + \omega_n \times clasif_n$$

where $\omega_1, ..., \omega_n \in [0,1] \wedge (\omega_1 + ... + \omega_n) = 1$ are the different weights learnt by the GA.

The main characteristics of the GA are:

- A CHC approach has been used [13].
- Real representation. The length of the chromosome is set up to the maximum number of FRBCSs considered.
- Fitness function: the accuracy classifying with the multi-classifier. The classification process is based on classify with each FBRCS, ponderating the FRBCS output with respect to its weight and its confidence and finally obtaining the most voted class.
- BLX-α crossover.
- Avoiding incest. To measures the distance between two parents' genes (real values), a modified version of the *hamming* distance is considered: The distance between two genes are 0 when their real values are very close (ponderated by a similarity threshold, in our case, 0.05) and 1 when are very different.

If the GA sets the weight of an FRBCS to 0, this FRBCS will be removed from the multiclassifier because it has no relevance in the classification process.

V. Experimentation and Analysis of Results

For the experimentation carried out, three databases [1] have been used: *Ionosphere* (351 instances, 34 features, 2 classes); *Wisconsin* (570 instances, 30 features, 2 classes); and *Vehicle* (846 instances, 18 variables, 4 classes).

The results obtained by our proposals (*GA-sharing, GA-clearing and GA-crowding*) are compared with those obtained by a standard generational GA, by the CHC algorithm and by the learning algorithm utilized, *Wang* and *Mendel*, without applying any feature selection algorithm.

The parameters corresponding to the GAs used in the experimentation are shown in Table I. Also, the following parameters are common for all the experimentation:

- Size of the population: 100
- Number of evaluations: 5,000
- Probability of crossing: 0.6
- Probability of mutation: 0.01
- λ: 0.7 and 1.0
- For the error estimation 10-fold cross validation error method is used.
- The FRBCSs used for the error estimation are built using an uniform fuzzy partition with 3 triangular fuzzy sets crossed at level 0.5, computing the compatibility degree with the t-norm minimum and using as fuzzy reasoning method the maximum and the normalized sum.
- All the algorithms have been executed three times with different seeds.

TABLE I

Algorithms and parameters used

Algorithm	Parameters	
CHC	divergence rate $= 0.35$	
Clearing	$\sigma = 3$	$\kappa = 0, 1, 3$
Sharing	$\sigma = 3$	$\alpha = 1$

Tables II, III and IV shown the accuracy and number of features used (average and standard deviation for all the partitions and executions) with: two values of the λ parameter, two fuzzy reasoning methods, with the typical GA, with the CHC algorithm, with the niching GAs and with the FRBCS built with the Wang and Mendel algorithm without feature selection process.

TABLE II

Classification Results for the IONOSPHERE database

	$\lambda = 0.7$		$\lambda = 1$	
	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$
Fuzzy reasoning method: sum				
GA	83,67 (0,06)	3,33 (0,88)	67,05 (0,03)	18,67 (2,81)
CHC	84,66 (0,05)	3,60 (0,89)	66,18 (0,03)	18,17 (3,13)
Cl. 0	86,23 (0,04)	**3,07** (0,69)	66,76 (0,04)	16,70 (3,47)
Cl. 1	**86,24** (0,04)	3,10 (0,61)	67,03 (0,05)	**16,30** (3,83)
Cl. 3	85,49 (0,04)	3,20 (0,71)	66,94 (0,04)	17,40 (2,61)
Sh.	80,04 (0,07)	4,47 (1,09)	**67,72** (0,03)	18,13 (2,52)
Cr.	82,25 (0,05)	3,47 (0,86)	67,12 (0,04)	17,83 (2,57)
W&M	62,97	34	-	-
Fuzzy reasoning method: maximum				
GA	80,71 (0,06)	4,30 (1,06)	67,44 (0,03)	19,50 (2,90)
CHC	82,05 (0,06)	3,73 (1,08)	67,15 (0,03)	20,67 (2,31)
Cl. 0	83,73 (0,06)	**3,67** (0,71)	67,54 (0,03)	19,87 (3,35)
Cl. 1	83,01 (0,05)	3,80 (0,61)	67,44 (0,03)	**18,10** (2,99)
Cl. 3	**84,22** (0,06)	3,80 (0,66)	67,34 (0,03)	19,03 (2,65)
Sh.	77,99 (0,06)	4,67 (1,21)	**68,02** (0,04)	19,30 (3,47)
Cr.	80,38 (0,06)	3,77 (1,19)	67,74 (0,03)	19,70 (3,19)
W&M	65,54	34	-	-

The results show us that for all the problems the precision obtained with an FRBCS design with all characteristics is surpassed with the ones obtained by using as a previous step the multimodal feature selection algorithms. The improve is greater when the problem dimension increases.

[1] Obtained from the UCI Repository http://www.ics.uci.edu/~mlearn/MLRepository.html

TABLE III

CLASSIFICATION RESULTS FOR THE VEHICLE DATABASE

	$\lambda = 0.7$		$\lambda = 1$	
	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$
Fuzzy reasoning method: sum				
GA	40,46 (0,03)	1 (0)	43,06 (0,03)	7,77 (1,48)
CHC	**40,62**(0,03)	1(0)	44,09(0,02)	**7,03**(1,25)
Cl. 0	40,50 (0,03)	1 (0)	44,13 (0,02)	7,10 (1,16)
Cl. 1	40,58 (0,03)	1 (0)	**44,28** (0,03)	7,23 (1,17)
Cl. 3	40,55 (0,03)	1 (0)	43,93 (0,03)	7,47 (1,25)
Sh.	40,58 (0,03)	1 (0)	42,98 (0,03)	7,50 (1,33)
Cr.	**40,62** (0,03)	1 (0)	42,99 (0,03)	7,43 (1,41)
W&M	39,60	18	-	-
Fuzzy reasoning method: maximum				
GA	56,34 (0,05)	3,70 (0,60)	57,71 (0,05)	13,03 (0,93)
CHC	57,05 (0,04)	3,87 (0,73)	58,62 (0,05)	13,23 (0,97)
Cl. 0	**59,10** (0,05)	4,00 (0,00)	57,98 (0,05)	13,03 (0,91)
Cl. 1	58,78 (0,05)	3,97 (0,18)	**58,85** (0,05)	13,20 (1,00)
Cl. 3	58,78 (0,05)	3,97 (0,18)	57,92 (0,05)	**13,00** (0,82)
Sh.	56,02 (0,05)	**3,63** (0,67)	58,50 (0,05)	13,33 (0,88)
Cr.	57,88 (0,05)	3,83 (0,38)	58,34 (0,05)	13,17 (1,02)
W&M	57,45	18	-	-

TABLE IV

CLASSIFICATION RESULTS FOR THE WISCONSIN DATABASE

	$\lambda = 0.7$		$\lambda = 1$	
	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$	$\overline{acc}(\sigma)$	$\overline{feat}(\sigma)$
Fuzzy reasoning method: sum				
GA	**92,96** (0,03)	2,70 (0,70)	95,08 (0,03)	12,23 (0,90)
CHC	92,50(0,02)	2,60(0,67)	94,26(0,03)	12,10(1,06)
Cl. 0	92,21 (0,03)	2,53 (0,86)	**95,25** (0,03)	**11,63** (1,13)
Cl. 1	92,33 (0,02)	2,40 (0,77)	94,90 (0,03)	12,03 (1,22)
Cl. 3	92,33 (0,03)	2,47 (0,78)	94,55 (0,03)	11,93 (1,23)
Sh.	92,74 (0,03)	2,70 (0,65)	94,32 (0,03)	12,30 (1,42)
Cr.	92,62 (0,02)	**2,20** (0,55)	94,20 (0,03)	12,27 (0,83)
W&M	89,09	30	-	-
Fuzzy reasoning method: maximum				
GA	91,39 (0,04)	1,73 (0,91)	**92,85** (0,03)	**9,13** (5,02)
CHC	90,80 (0,05)	1,27 (0,45)	89,69 (0,04)	13,57 (4,95)
Cl. 0	91,39 (0,04)	1,67 (0,71)	90,10 (0,04)	11,67 (4,31)
Cl. 1	91,51 (0,04)	1,63 (0,72)	89,22 (0,04)	11,10 (5,40)
Cl. 3	**91,62** (0,04)	1,67 (0,80)	90,21 (0,04)	11,33 (3,63)
Sh.	91,04 (0,04)	**1,13** (0,35)	89,28 (0,04)	14,60 (3,89)
Cr.	90,57 (0,05)	1,67 (0,48)	89,34 (0,04)	14,13 (4,21)
W&M	88,57	30	-	-

If niching GAs are compared with generational GA, it can be observed that normally all multimodal results improve, but with more differences in the problem with greater dimension, *Ionosphere*.

Examining only the results of the niching feature selection algorithms it can be observed that the *clearing* algorithm with value $\kappa = 0$ is the one with best results. But, in some problems, it is surpassed by the clearing AG with $\kappa = 1$ and that are represented as a second algorithm for general performance.

The experimentation shows that for the feature selection problem the method *clearing* has better results than *sharing*, *deterministic crowding*, and the standard GA.

If we observe the differences among the use of one or another inference method, in this work normalized sum or maximum, we cannot state that one is better than the other. According to the problem we examine, one method works better than another. So, for *Ionosphere* the method of the sum obtains better precision, but for *Vehicle* more precision is obtained if we use the maximum inference method. With *Wisconsin*, the sum method obtains more precision but reduces more features than the maximum one.

It must be highlighted the high reduction in the feature subset used for the FRBCSs in all the problems. So much that *Vehicle* always considers only one feature (we started with 18 initial features). This dimensionality reduction is more evident when the value *0.7* is used as ponderation error. This combination of weights for fitness function boosts the dimension reduction with values of adequate precision. Even in problems with high number of variables, as *Ionosphere*, precision is increased.

TABLE V

RESULTS FOR THE MULTICLASSIFIER OF EACH DATABASE

	Type of Classifier	$\overline{Tst}$	$\overline{Tra}$
Ionosphere	Individual FRBCS	86,24	89,88
	Multiclassifier (proportional weights)	88,82	90,09
	Multiclassifier (evolved weights)	**89,11**	**92,09**
Wisconsin	Individual FRBCS	95,25	94,76
	Multiclassifier (proportional weights)	**97,83**	96,66
	Multiclassifier (evolved weights)	97,66	**97,31**
Vehicle	Individual FRBCS	59,10	75,42
	Multiclassifier (proportional weights)	65,59	83,58
	Multiclassifier (evolved weights)	**65,95**	**85,61**

In Table V the accuracy in testing and training sets for individual FRBCS, for a multiclassifier using proportional weights and for a multiclassifier with evolved weights are shown. As we can see, the multiclassifier using fusion and a GA for the learning of the weights has the best generalization capacity, increasing its description capacity. Moreover, the multiclassifier system lets the classification of examples described by means of different features subsets due to it integrates different FRBCSs built with different features subsets. However, it must be noted that the interpretability of the FRBCSs could be decreased due to the increase of the number of fuzzy rules considered. A postprocessing stage which selects rules could be considered.

VI. CONCLUSIONS AND FUTURE WORKS

In this work different niching genetic feature selection algorithms have been described and analyzed. Moreover, the different optimum solutions provided by them (and therefore, the different FRBCSs learnt using these feature subsets) are combined in a multiclassifier designed by using an evolutionary approach. The proposals has been applied to three different problems: *Ionosphere, Wisconsin* and *Vehicle*.

The three niching genetic algorithms implemented, *clearing*, *sharing*, and deterministic *crowding*, have shown to be better than the simple GA and the CHC approach for this problem. The niching genetic algorithms has obtained sets with very few variables, with cardinality equal to 3 or 4, for problems that initially are described by 34 features, and even so, the precision has improved. The GA with *clearing* is the niching genetic algorithm that presents better performance in precision, reduction and convergence.

The improvement in accuracy and simplicity is not the main advantage of the niching GAs for dimensionality reduction in FBRCS design. They provide not only a solution but also a set of optimum solutions. These sets of variables can be used for the design of a set of FRBCSs which can be combined in a multiclassifier system. In this paper an evolutionary proposal for the design a multiclassifier improving the prediction capacity increasing the description capacity is described. Moreover, the multiclassifier system design in this way lets the classification of examples described by means different features subsets due to it integrates different FRBCSs built with different features subsets.

As future works, we consider to do an study on the best solutions provided by the niching GAs for feature selection in order to extract more knowledge analyzing relations between variables and to develope of new proposals based on advanced niching GAs.

ACKNOWLEDGMENT

This work has been supported by TIN 2005-08386-C05-03 and TIN 2005-08386-C05-01 projects.

REFERENCES

[1] J. Abonyi, J.A. Roubos, F. Szeifert, *Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization*, International Journal of Approximate Reasoning 32 (1): 1-21, 2003.

[2] E. Alpaydin, *Voting over multiple condensed nearest neighbours*, Artificial Intelligence Review 11: 115-132, 1997.

[3] J. Bala, K. De Jong, J. Huang, H. Vafaie, H. Wechsler, *Using learning to facilitate the evolution of features for recognizing visual concepts*, Evolutionary Computation 4(3): 297-311, 1997.

[4] D. Beasley, D.R. Bull, R.R. Martin *A sequential niche technique for multimodal function optimization* Evolutionary Computation 1(2): 101-125, 1993.

[5] N.H. Beltrán, M.A. Duarte-Mermoud, S.A. Salah, M.A. Bustos, A.I. Pe-Neira, E.A. Loyola, J.W. Jalocha *Feature selection algorithms using Chilean wine chromatograms as examples*, Journal of food engineering, 67: 483-490, 2005.

[6] J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), *Interpretability issues in fuzzy modeling*, Springer-Verlag Series Studies in Fuzziness and Soft Computing 128, 2003.

[7] J. Casillas, O. Cordón, M.J. Del Jesus, F. Herrera, *Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems*, Information Sciences 136: 135-157, 2001.

[8] W. Cede, V.R. Vemuri, *Analysis of speciation and niching in the multiniche crowding GA*, Theoretical Computers Science 229: 177-197, Elsevier, 1999.

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

[10] O. Cordón, M.J. Del Jesus, F. Herrera, M. Lozano, *MOGUL: a methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach*, International Journal of Intelligent Systems 14 (11): 1123-1143, 1999.

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

[12] B.V. Dasarathy and B.V. Sheela, *A composite classifier system design: concepts and methodology*, Proceedings of IEEE, 67: 708-713, 1978.

[13] L.J. Eshelman., *The CHC Adaptative Search Algorithm: how to have safe search when engaging in Nontraditional Genetic Recombination.* Foundations of Genetic Algorithms, 265-283, Morgan Kaufmann Publishers, 1991.

[14] D.E. Goldberg and J. Richardson, *Genetic Algorithms with sharing for multimodal function optimization.*, In proceeding of the Second International Conference on Genetic Algorithms: 41-49. J.J. Grefensette Ed. Hillsdale, NJ: Lawrence Erlbaum. 1987.

[15] D.E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning.*, Addison-Wesley 1989.

[16] A. González, R. Pérez, *Selection of relevant features in a fuzzy genetic learning algorithm*, IEEE Transactions on Systems, Man and Cybernetics Part B 31(3): 417-425, 2001.

[17] J.H. Holland, *Adaptation in Natural and Artificial Systems.*, Ann Arbor, MI: University of Michigan Press, 1975.

[18] J.J. Hu, E.D. Goodman, *Robust and efficient genetic algorithms with hierarchical niching and a sustainable evolutionary computation model*, Proc. GECCO: 1220-1232, 2004.

[19] L. Jourdan, C. Dhaenens and El-Ghazali Talbi, *A Genetic Algorithm for Feature Selection in Data-Mining for Genetics*, MIC'2001 - 4th Metaheuristics International Conference: 29-34, Portugal, 2001.

[20] Kohavi, R., John, G.H., *Wrappers for Feature Subset Selection*, Artificial Intelligence 97: 273-324, 1997.

[21] Kuncheva, L.I., *Fuzzy Classifier Design*, Studies in Fuzzines and Soft Computing - Springer-Verlag, 2000.

[22] J. Li, M. Balazs, G.T.. Parks, P.J. Clarkson, *A species conserving genetic algorithm for multimodal function optimization*, Evolutionary computation 10(3): 207-234, 2002.

[23] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Editorial Kluwer Academic Publishers, 1998.

[24] S. W. Mahfoud, *Niching methods for Genetic Algorithms*, Ph.D. Thesis, University of Illinois, Urbana-Champaign, 1995.

[25] C.K. Oei, D.E. Goldberg and S.J. Chang *Tournament selection, niching and the preservation of diversity*, Technical Report 91011. Illinois Genetic Algorithms Laboratory. Universidad de Illinois, Urbana, 1991.

[26] I.S. Oh, J.S. Lee, B.R. Moon, *Hybrid Genetic Algorithms for Feature Selection*, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11): 1424-1437, 2004.

[27] E. Pérez, F. Herrera and C. Hernández, *Finding multiple solutions in job shop scheduling by niching genetic algorithms*, Journal of Intelligent Manufacturing, 14: 323-341, 2003.

[28] A. Pétrowski, *Clearing procedure as a niching method for genetic algorithms*, In Proceedings of the IEEE Third International Conference on Evolutionary Computation: 798-803, Nagoya , Japan, 1996.

[29] B. Sareni and L. Krahenbuhl, *Fitness sharing and niching methods revised*, IEEE Transactions on Evolutionary Computation, 2: 97-106, 1998.

[30] W. Siedlecki, J. Sklansky, *A note on genetic algorithm for large-scale feature selection*, Pattern Recognition Letters 10: 335-347, 1989.

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

[32] N. Xiong and P. Funk *Construction of fuzzy knowledge bases incorporating feature selection*, Soft Computing, 10(9): 796-804, 2006.

[33] L. Xu, A. Krzyzak and C.Y. Suen, *Methods of combining multiple classifiers and their application to handwriting recognition*, IEEE Transactions on Systems, Man and Cybernetics, 22: 418-435, 1992.

[34] P. Zhang, B. Verma, K. Kumar, *Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection*, Pattern Recognition Letters 26(7): 909-919, 2005.