Genetic Feature Selection in a Fuzzy Rule-Based Classification System Learning Process

Oscar Cordón
Dept. of Computer Science and
Artificial Intelligence
E.T.S. Ingeniería Informática
University of Granada
18071 - Granada, Spain
ocordon@decsai.ugr.es

María José del Jesus
Dept. of Computer Science
Escuela Politécnica Superior
University of Jaén
23071 - Jaén, Spain
mjjesus@ujaen.es

Francisco Herrera
Dept. of Computer Science and
Artificial Intelligence
E.T.S. Ingeniería Informática
University of Granada
18071 - Granada, Spain
herrera@decsai.ugr.es

Abstract

The inductive learning of a Fuzzy Rule-Based Classification System is made difficult by the presence of a high feature number that increases the dimensionality of the problem to solve.

In this work, we propose a complete Fuzzy Rule-Based Classification System learning process composed of a feature selection, a fuzzy rule generation and selection, and a linguistic tuning processes. For the feature selection stage, we propose two new genetic algorithms with wrapper nature.

The experimentation carried out, using Sonar sample base, shows the increase on simplicity, precision and efficiency achieved by adding the proposed feature selection processes in the learning algorithm.

Keywords: Fuzzy Rule-Based Classification Systems, Inductive Learning, Feature Selection, Fuzzy Reasoning Methods

1 Introduction

The inductive learning of a Fuzzy Rule-Based Classification System (FRBCS) starts from a set of problem instances, and determines a set of fuzzy rules and a fuzzy inference method that generalises the knowledge extracted from the data for the classification of the new instances. Each one of these problem instances or samples is described by a set of features, also called variables.

In the FRBCS design the following problems must be considered by the learning process:

- The determination of the inference method used in the classification stage to establish the class for each pattern of the problem. The inference method employed in most of the FRBCSs [19, 27, 7, 23, 32, 1, 13] uses the information provided by only a fuzzy rule (the fuzzy rule that is most compatible with the example to classify). This fact implies a loss information that can be avoided with the use of alternative inference methods [3, 7, 16, 10].
- With respect to the knowledge extraction process, we can point out two difficulties:
  - the obtaining of a fuzzy rule set with an adequate co-operation level between the fuzzy rules, and
  - the exponential growth of the fuzzy rule search space with the increase of the feature number considered in the learning process.

The fuzzy rule obtaining has been solved by different learning process based on iterative methods [6, 19, 31], Neural Networks [21, 29, 30] or Genetic Algorithms [20, 17, 34, 13, 18] for instance.

In [10, 9] a general definition of the inference method in a FRBCS is presented as well as different proposals for it that improve the behaviour of an FRBCS in the classification stage. In [8] a genetic learning process that considers the fuzzy rule co-operation problem is presented, but the last problem, the need to reduce the problem dimensionality when the number of features is high, is not consider in it.

The design of FRBCS for Classification problems with a high feature numbers implies that the FRBCS learning process must face up to two kind of problems:

- *Memory space problems*, in algorithms as ANFIS [21] that needs to represent in the learning process the complete fuzzy partition for the considered variables in each node of its structure.
- *Efficiency or / and effectiveness problems*, in algorithms that search in the complete fuzzy rule

search space as Genetic Algorithms [17, 34, 13, 18].

The solution for the aforementioned problems is the integration of a Feature Selection process, that determines the most relevant variables before the FRBCS inductive learning process. In this form, the memory space needed for some learning algorithms and the fuzzy rule search space are reduced, the efficiency and effectiveness of the FRBCS learning process is increased as well as the simplicity and interpretability of the FRBCS.

In this work we propose the integration of a feature selection stage in a multistage genetic learning process of FRBCSs. For this task we show two new feature selection methods that can be included in another learning processes.

To carry out this task, in Section 2 some preliminaries are introduced: the FRBCS definition and the description of the multistage genetic learning process for FRBCS. In Section 3, both the integration of the feature selection process in this multistage FRBCS learning process, and two proposals for the feature selection stage, are explained. Section 4 shows the results of the experiments with Sonar sample base. In the last section, the conclusions and future research lines are exposed.

2 Preliminaries

2.1 Fuzzy Rule-Based Classification Systems

An FRBCS is an automatic Classification System that uses as knowledge representation tool the fuzzy rules. This kind of Classification System is made up of two components:

- The **Knowledge Base** (KB) composed of:
  - a *Data Base* (DB) that contains the fuzzy set definitions related to the linguistic terms used in the fuzzy rules, and
  - a *Rule Base* (RB), a set of fuzzy rules with the following structure:

    $$R_k: \quad \text{If } X_1 \text{ is } A_1^k \text{ and } \ldots \text{ and } X_N \text{ is } A_N^k \\ \text{then } Y \text{ is } C_j \text{ with } r^k$$

    where $X_1, \ldots, X_N$ are features considered in the problem, $A_1^k, \ldots, A_N^k$ are linguistic labels employed to represent the values of the variables, and r^k is the certainty degree related to the classification in C_j class for the samples belonging to the fuzzy subspace delimited by the rule antecedent.
- The **Fuzzy Reasoning Method** (FRM), an inference method that, combining the information provided by the fuzzy rules related with the example, determines the class to which it belongs to.

2.2 Multistage Genetic Learning of Fuzzy Rule-Based Classification Systems

In [8] a multistage genetic learning process for FRBCSs is proposed, divided into three stages:

1. A **fuzzy rule generation process**, that obtains a linguistic RB which represents the knowledge extracted from the training samples and verifies the completeness and k-consistency properties [13, 14].

2. A **genetic multiselection process** that generates different KBs. In this process a selection of a rule subset is carried out as well as a learning of a linguistic modifier set, considering the FRM used in the classification stage.

3. A **genetic tuning process** that leads to obtain the best parameter membership function values for the fuzzy rules.

In the following subsections these processes are briefly described. A complete description of them can be found in [8].

2.2.1 Fuzzy Rule Generation Process

The fuzzy rule generation process has two components, a rule generation and an iterative covering methods:

- **The fuzzy rule generation method** obtains, in each iteration, a candidate fuzzy rule set, generating for each training sample the fuzzy rule which better represents the space zone to which it belongs to. From this set of rules, the best rule is selected by means of a multicriteria selection function, which considers criteria related to the rule frequency, completeness and k-consistency.

- **The covering method**, applies the generation method to obtain the best rule for the training samples, and considers the relative covering that this rule provokes in them, eliminating those samples that are covered with a degree higher than a maximum value previously specified, until the training set becomes empty.

2.2.2 Genetic Multiselection Process

The fuzzy rule generation process, which does not consider the relationship among the rules, can obtain an RB with an inappropriate co-operation level among them. To solve this problem, one objective of the genetic multiselection process is the selection of fuzzy rule subsets with optimal co-operation in the classification stage depending on the FRM used. Besides this, the multiselection process allows to increase the precision of the KB selecting a linguistic hedge set for the linguistic terms used in the RB.

This multimodal optimisation problem is solved in this proposal with a Genetic Algorithm (GA) [15, 12] that uses the *sequential niche technique* [5] to induce niches in the search space and obtain different KB definitions by means of the basic genetic selection process.

The *basic genetic selection process* has, as we mentioned previously, a double objective: the selection of a rule subset with a good co-operation among them, considering the FRM, and the selection of a linguistic modifier set related to the fuzzy subsets used by the fuzzy rules. The last learning can be done in two different forms: selecting a linguistic hedge for each fuzzy subset defined in the DB, or determining a linguistic hedge for each fuzzy subset related to each linguistic variable in each fuzzy rule. In this genetic process the fitness associated to each solution (KB definition) is penalised if the completeness property is not verified.

Every time this basic genetic selection process is executed, the solution obtained is optimised using a local search process based on hill-climbing. Finally, the search space zone in which the solution has been obtained is penalised, to get different KBs in posterior executions of the basic genetic selection process.

2.2.3 Genetic Tuning Process

The genetic tuning process leads to optimise the fuzzy partition of the linguistic variables, determining the best membership function parameter values in a common way to all the fuzzy rules.

This process is based on the parametric representation of the membership functions and demands, as the multiselection process does, the verification of the completeness property.

	5 Labels		3 Labels	
	311 Rules		331 Rules	
FRM	Tra.	Test	Tra.	Test
Classic	100	43.27	99.04	75.00
Normalized Sum	100	43.27	98.08	73.08
Arithmetic Mean	100	43.27	96.15	72.11
Quasiarithmetic Mean	100	43.27	99.04	75.00
SOWA Or-Like	100	43.27	98.08	76.92
Badd	100	43.27	99.04	75.00
OWA	100	43.27	98.08	75.96
QuasiOWA	100	43.27	99.04	75.96

Table 1: Results with a KB obtained after the generation process for the Sonar problem

3 Feature Selection in the Fuzzy Rule-Based Classification System Learning

3.1 The incorporation of a Feature Selection process in the multistage genetic learning process of FRBCSs

To show the need of a feature selection stage in the multistage genetic learning process proposed, we have applied it to a sample base with a high feature number, Sonar sample set, 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.

For this problem, if we use five linguistic labels per variable, the search space for the learning process is composed of 5^{60} candidate fuzzy rules. The results obtained after the generation stage are shown in Table 1, columns 2 and 3 with different FRMs which are described in [10, 9]. In this table we can observe that the correct classification percentage is the same independently on the FRM used, due to the wrong classified samples are not classified samples. This problem can be lessened with the use of a more compensated t-norm than the t-norm used, the minimum, or considering fuzzy partitions with a smaller number of linguistic labels. In Table 1, columns 4 and 5, we can see the results obtained by the generation process considering 3 linguistic labels per linguistic variable.

Nevertheless, the results obtained in the first stage of the learning process, in both situations, show that the intervention of the complete set of features leads to the design of an FRBCS overfitted to the training samples that covers only a small proportion of the complete sample space. This fact limits the possibilities of improving for the postprocessing (multiselection and

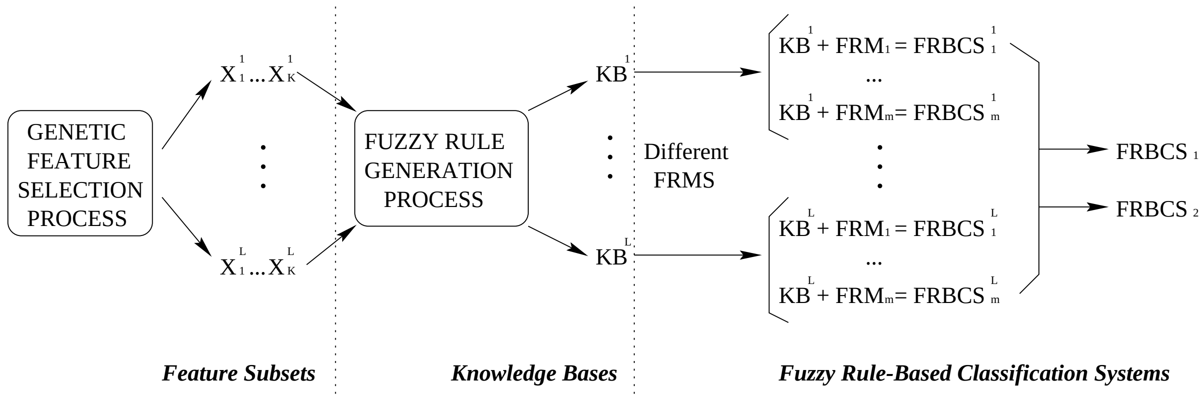

Figure 1: Stages in the FRBCS learning process (I)

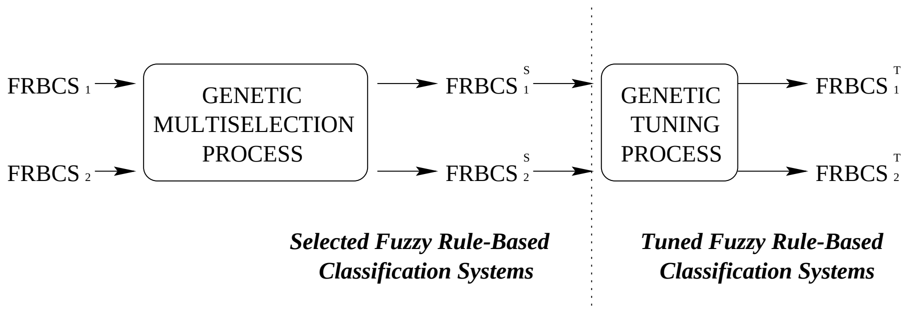

Figure 2: Stages in the FRBCS learning process (II)

tuning) stages.

This problem can be solved with a design process that, using all features, selects the most informative ones for every, or for each one fuzzy rule in the FRBCS inductive learning process. These approaches do not settle the memory space, efficiency and effectiveness problems because of in this form the space of candidate fuzzy rules is not limited, far from it, it is increased.

We propose the integration in the FRBCS multistage genetic learning of a feature selection stage that limits the problem dimensionality by means determining a feature subset for the FRBCS design. The resulting FRBCS learning process is composed of the following steps:

1. A **feature selection process** that gets a feature subset with a fix cardinality previously determined, to learn from it the FRBCS. In this form, we will reduce the problem dimensionality before the FRBCS design. The proposed feature selection process uses a GA as search algorithm and it has wrapper nature [22]. We use a feature selection algorithm with filter nature [24] that searches for a variable cardinality feature subset to obtain the optimal feature number for our proposal of feature selection process. We will explain this process in detail in the next subsection.

2. A **generation process** that obtains an RB independently on the FRM used in the classification stage. We will use this efficient learning stage as an intermediate stage to determine the best feature subset (consequently the best KB) and the best FRM for the problem to solve.

3. A **multiselection process** of different KBs with a good co-operation level among them considering the FRM selected in the previous step.

4. A **tuning process** of the fuzzy partitions for the fuzzy variables in a common way for all rules.

The resulting FRBCS learning process is graphically described in Figures 1 and 2 and is mainly composed of four processes, that is feature selection, generation, multiselection and tuning, and the last three have been briefly described in Section 2.2. In the next subsections we explain our proposals for the feature selection stage.

3.2 Feature Selection Stage

The main objective of the feature selection stage is the problem dimensionality reduction before the supervised inductive learning process. This fact implies that the feature selection algorithm must determine -without the necessity of the FRBCS construction- the best features for its design.

The *filter feature selection algorithms* [24] leak the irrelevant characteristics before the supervised inductive learning process, but as is well known, the feature subsets obtained by them could not be the best features for a specific Classification System design process due to the exclusion in the feature selection process of the heuristic and bias in the inductive learning process.

The *wrapper feature selection algorithms* [22, 24] lead to obtain feature subsets with the best behaviour in the Classification System design because of they use the precision estimation obtained by the Classification System learning process for the candidate feature subset evaluation. The problem of this kind of feature selection algorithms is the inefficiency (because of they must build the FRBCS for each evaluation of a candidate feature subset).

We propose a feature selection stage that combines both kinds of feature selection algorithms in this way:

1. We use two filter feature selection algorithms that looks for feature subsets with variable size considering class separability measures to determine an optimal feature number for an specific classification problem. In this work we employ the following ones:
   - the probabilistic algorithm Las Vegas Filter (LVF) [25, 24] based on the inconsistency rate, proposed by Liu and Setiono, and
   - a greedy algorithm based on a forward selection search using the mutual information (MIFS) developed by Battiti in [4].
2. The results of these feature selection algorithms for the classification problem, provide us an adequate feature subset size for a wrapper feature selection process that determines a feature subset of this cardinality with the best behaviour for the classification problem to solve. To increase the efficiency maintaining the effectiveness of the wrapper feature selection algorithms the proposal uses the precision estimation provided by the k-nearest neighbour rule (k-NN) [11], that is very sensitive to the presence of irrelevant characteristics.

The k-NN rule is not sensible to redundant characteristics. The previous determination of the feature subset size realised by filter algorithms that do not use the k-NN rule, helps to the wrapper selection to select only relevant variables and to reduce effectively the problem dimensionality, in an efficient way.

In the next subsection we describe two proposals for the wrapper feature selection process.

3.3 Steady State Genetic Algorithms for Feature Selection

The first stage of the learning process is carried out by a feature selection algorithm based on a GA with a variant of the pure steady stage reproduction mechanism [33]. This feature selection process is a wrapper feature selection algorithm [22] that uses as evaluation function a precision measure provided by the k-NN considering only the features included in the candidate feature subset.

The GA is described by its components:

1. **Coding scheme.**

   The feature selection process objective is to get an optimal feature subset with a fixed cardinality, so the integer coding using fixed length allows us to represent in a chromosome with length H a candidate subset containing H variables in which, the *ith* gen represents the *ith* selected variable.

   The proposed GA permits the incorporation of available knowledge, that is, features subsets provided by an expert or another feature selection algorithm, in the initial population. The remaining population is randomly generated.

2. **Adaptation Function.**

   To increase the speed of the feature selection stage, the estimation of the reachable precision is calculated by the k-NN. This test precision estimation is obtained by the training random resampling proposed by Kohavi [22] for wrapper feature selection algorithms, with 5 training-test partitions obtained from the original training set, and the adaptation measure calculated by the arithmetic mean of the 5 test correct classification results. In this way, we can estimate the generalisation capability of a feature subset without using the test set employed to validate the finally obtained feature subset.

3. **Reproduction Scheme.**

   The proposed GA uses a variant of the steady state reproduction scheme that does not substitute all the individual from the population in each generation, but a fixed number of them. We propose a reproduction scheme that follows the next steps:

- An intermediate population is generated by assigning probabilities by means of a linear ranking and the universal stochastical sampling [2].
- The crossover and mutation operators are applied to some individuals from this intermediate population. The number of chromosomes to be created will be determined by the crossover and mutation probabilities.
- The new chromosomes substitute to the worst adapted ones from the *original population*.

The generation of more than two new chromosomes leads to have more diversity in the new population than the pure steady stage reproduction scheme. Nevertheless, it maintains the steady state characteristics because the new population only differs from the previous one on these generated chromosomes, which substitute to the worst adapted.

4. **Crossover Operator.**

   We propose two feature selection algorithms that differs only in the crossover operator applied:

   - The algorithm that uses the *partially complementary crossover operator* [26] (which we will identify by SSGA_I). This operator exploits the search space tuning the obtained solutions in the following way: given two chromosomes from the population $P(t)$, $C_v^t = (c_1, \ldots, c_M)$ and $C_w^t = (c'_1, \ldots, c'_M)$, two descendants are generated:

     $$H_1 = (d_1, \ldots, d_k, h_{k+1}, \ldots, h_M)$$
     $$H_2 = (d_1, \ldots, d_k, h'_{k+1}, \ldots, h'_M)$$

     where $d_1, \ldots, d_k$ are the common genes to the two chromosomes selected to be crossed, and $h_{k+1}, \ldots, h_M$, and $h'_{k+1}, \ldots, h'_M$ are genes randomly selected among the remaining.

     In this way, the descendants maintains the parents' common variables and randomly combines the remaining information. They are valid individual and do not need any repairing algorithm.
   - The algorithm that uses the *two point crossover with repair* operator (noted in this paper as SSGA_II).

     An analysis of the SSGA_I lets us observe that, sometimes, it can evolve to a population without enough diversity. To solve this problem, we propose the two-point crossover operator with repair, which not only exploits the information given by the parents, but also introduces diversity in the descendants. This operator works as follows: to obtain the descendants, for each selected chromosome pair, two cross points are determined, and the genes between these points are exchanged. This process can generate non-valid individuals because of the variable repetition. To solve this problem, a repairing algorithm substitutes each repeated gene by a non-selected variable.

     This two-point crossover operator with repair maintains the inheritance and refinement properties of the crossover operators, adding –when the descendants has repeated variables– the exploration property, very suitable in the proposed evolutive process.

5. **Mutation Operator.**

   The *uniform mutation* arbitrarily modifies one or more genes from an individual, removing the corresponding variable, and substituting it for another one which is not present in the chromosome, introducing diversity among the population.

4 Experimentation and Result Analysis

We will show the results obtained using the proposed learning process applied to the Sonar sample base.

As we mentioned before, the feature selection process selects variables sets with a fixed cardinality previously determined. We compute this cardinality by executing classical feature selection algorithms that search for an optimal and minimum feature set, as LVF [25] and MIFS [4]. These algorithms provide us 3 proper feature set sizes: 6, 12 and 15 variables which reduce in 90, 80 and 75 % respectively the fuzzy rule search space for the fuzzy rule generation, selection and tuning processes.

According to the previously exposed multistage learning process, we execute the feature selection algorithms SSGA_I and SSGA_II with these three different set sizes. We build FRBCSs with the generation process -starting with the feature subsets obtained before- and analyse the results to determine two variable subsets (and consequently two KBs) and FRMs with the best behaviour for each cardinality. In this way, and efficiently by the iterative nature of the generation process, we greatly reduce the problem dimensionality because of limiting the fuzzy rule and FRMs space to consider in following stages.

Algorithm	FRM	NR	Tra.	Test
SSGA_I	OWA	55	94.23	89.42
SSGA_II	Arithmetic Mean	58	93.27	90.38

Table 2: Results for an FRBCS built using 6 features

Algorithm	FRM	NR	Tra.	Test
SSGA_I	SOWA Or-Like	183	92.31	94.23
SSGA_II	OWA	45	91.35	90.38

Table 3: Results for an FBRCS built using 12 features

At last, we execute the multiselection and tuning processes to get FRBCSs which reach the results shown in Tables 2, 3 and 4.

With the proposed learning process we have increased in more than a 15 % the correct test classification percentage, and overcame the overfitting and efficiency problems, obtaining a simpler and more interpretable FRBCS.

5 Conclusions

Usually, in the FRBCS design the following problems must be individually or jointly solved:

- The selection of the most relevant features for the considered classification problem.
- The fuzzy partition definitions for the linguistic variables.
- The generation of an RB that represents the samples information and verifies two desired properties in any RB, the completeness and consistency [28].
- The generation of an RB with a good co-operation level among the fuzzy rules with respect to the FRM used in the classification stage.

The proposed multistage genetic learning process of FRBCSs considers these problems in different stages, obtaining a linguistic FRBCS with a good generalisation level, that uses only the most informative features for the problem, with an RB which verifies the completeness and k-consistency properties, having a good co-operation level depending on the FRM, and with an optimised DB.

The inclusion of an feature selection stage in the learning process leads to a limitation, previous to the FRBCS design, of the fuzzy rule space and allows an increase in the efficiency and efficacy of the learning process.

Algorithm	FRM	NR	Tra.	Test
SSGA_I	Classical	125	96.15	94.23
SSGA_II	OWA	94	99.04	92.31

Table 4: Results for an FRBCS built using 15 features

Also, the decomposition of the learning process in multiple stages, permits to progressively reduce the problem dimensionality.

References

[1] S. Abe and R. Thawonmas. A fuzzy classifier with ellipsoidal regions. *IEEE Transactions on fuzzy systems*, 5(3):358–368, 1997.

[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] A. Bardóssy and L. Duckstein. *Modeling with applications to geophysical, biological and engineering systems.* Systems Engineering Series. CRC Press, 1995.

[4] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, July 1994.

[5] D. Beasly, D. R. Bull, and R. R. Martin. A sequential niche technique for multimodal function optimization. *Evolutionary Computation*, 1:101–125, 1993.

[6] Z. Chi, J. Wu, and H. Yan. Handwritten numeral recognition using self-organizing maps and fuzzy rules. *Pattern Recognition*, 28(1):59–66, 1995.

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

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

[9] O. Cordón, M. J. del Jesus, and F. Herrera. Reasoning methods based on OWA operators under fuzzy majority in fuzzy rule-based classification sytems. Technical Report 98121, Dept. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, 1998.

[10] 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.

[11] B.V. Dasarathy. Nearest neighbour (nn) norms: Nn pattern classification techniques. *IEEE Computer Society Press*, 1990.

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

[13] A. González and R. Pérez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96:37–51, 1998.

[14] F. Herrera, M. Lozano, and J. L. Verdegay. A learning process for fuzzy control rules using genetic algorithms. *Fuzzy Sets and Systems*, 100:143–158, 1998.

[15] John H. Holland. *Adaptation in Natural and Artificial Systems*. Universtity of Michigan Press, Ann Arbor, MI, 1975.

[16] H. Ishibuchi, T. Morisawa, and T. Nakashima. Voting schemes for fuzzy-rule-based classification systems. In *Proc. of the Fith International IEEE Conference on Fuzzy Systems*, pages 614–620, 1996.

[17] H. Ishibuchi, T. Nakashima, and T. Murata. A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems. In *Proc. of the Second International IEEE Conference on Fuzzy Systems*, pages 759–764, 1995.

[18] H. Ishibuchi, T. Nakashima, and T. Murata. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Transactions on Systems, Man and Cybernetics*, 29:601–618, 1999.

[19] H. Ishibuchi, K. Nozaki, and H. Tanaka. Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets and Systems*, 52:21–32, 1992.

[20] 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.

[21] Jyh-Shing Roger Jang. Anfis: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):665–685, may-june 1993.

[22] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

[23] L. I. Kuncheva. On the equivalence between fuzzy and statistical classifiers. *International Journal of Uncertainty, Fuzzyiness and Knowledge-Based Systems*, 4(3):245–253, 1996.

[24] Huan Liu and Hiroshi Motoda. *Feature selection for knowledge discovery and data mining*. Kluwer Academic Publishers, 1998.

[25] Huan Liu and Rudy Setiono. A probabilistic approach to feature selection. A filter solution. In L. Saitta, editor, *Proc. of International Conference on Machine Learning (ICML-96)*, pages 319–327. Morgan Kaufmann Publishers, 1996.

[26] Weiquan Liu, Minghui Wang, and Yixin Zhong. Selecting features with genetic algorithm in handwritten digits recognition. In *Proc. of the International IEEE Conference on Evolutionary Computation (ICEC'95)*, volume 1, pages 396–399, 1995.

[27] D. P. Mandal, C. A. Murthy, and S. K. Pal. Formulation of a multivalued recognition system. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4):607–620, 1992.

[28] R. S. Michalski. A theory and methodology of inductive learning. In *Machine Learning: An artificial intelligence approach I*, pages 83–134. Morgan Kaufmann, 1983.

[29] S. Mitra and S. K. Pal. Fuzzy self-organization, inferencing, and rule generation. *IEEE Transactions on Systems, Man and Cybernetics*, 26:608–619, 1996.

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

[31] Sankar K. Pal and Deba Prasad Mandal. Linguistic recognition system based on approximate reasoning. *Information Sciences*, 61:135–161, 1992.

[32] K. S. Ray and J. Ghoshal. Approximate reasoning approach to pattern recognition. *Fuzzy Sets and Systems*, 77:125–150, 1996.

[33] G. Syswerda. A study of reproduction in generational and steady-state genetic algorithms. In G. Rawlins, editor, *Foundations of Genetic Algorithms*. Morgan Kaufmann, 1991.

[34] Yufei Yuan and Huijun Zhuang. A genetic algorithm for generating fuzzy classification rules. *Fuzzy Sets and Systems*, 84:1–19, 1996.