

Data classification by a fuzzy genetic system approach

R. P. Espíndola & N. F. F. Ebecken
COPPE/Federal University of Rio de Janeiro, Brazil

Abstract

This paper presents a fuzzy genetic approach to perform data classification. The method aims to obtain small fuzzy classifiers by means of optimization of fuzzy rules bases using a genetic algorithm. It is shown how a fuzzy rules base is generated from a numerical database and how its best subset is found by the genetic algorithm. The classifiers are evaluated in terms of accuracy, cardinality and amounts of features employed. The results obtained are compared to a known study in the literature and with an academic decision tree tool. The method was able to produce small fuzzy classifiers with very good performance.

1 Introduction

Classification (Gordon, [1]) is an important task in each knowledge field. It consists of classifying elements described by a fixed set of attributes into one of a finite set of categories or classes. For example, to diagnose a disease of a person by his medical exams or to identify a potential customer of a product by his purchases. Several artificial intelligence approaches have been applied to this problem like artificial neural networks, decision trees and production rules systems.

A classifier is not evaluated based solely by their ability of classifying new elements correctly. Another important aspect is the knowledge representation obtained from the datasets. There are situations in which a specialist is not present to elucidate the information necessary to perform the task. Sometimes new pieces of information are obtained and need to be presented in an adequate way. Moreover the correct application of the classifier is favored by the comprehension of what is being done.

One of the most efficient forms to represent knowledge is employing production rules. The quality of a rule base may be measured by the amounts of conditions and conclusions in its rules and by its cardinality – amount of rules it is formed. Other important aspects are the amount of features used and the presence of linguistic qualifiers to describe them. The former aspect is approached by the feature selection methods. Feature selection (Liu & Motoda, [2]) is a task related to classification and its aim is to identify and to select only the most important features of a problem, discarding irrelevant or redundant ones. Due to the uncertainty and ambiguity present in linguistic expressions, the latter aspect is suitably treated by fuzzy logic.

In this work, feature selection and classification are performed by a fuzzy genetic system. The fuzzy rules are generated automatically from the datasets and the genetic algorithm is applied to find the shortest and most accurate subset of rules. As each rule employs only one feature, the final subset possibly uses few features. In this kind of rule, the decision is done along with a value which estimates the degree of relationship between the condition and the class.

This work was inspired in the work of Ishibuchi et al [3] which employed a genetic algorithm to obtain a fuzzy classification system using Mamdani's model of rules. This method suffers from combinatorial explosion of rules when applied to other than very simple problems. Espíndola & Ebecken [4] applied the same genetic algorithm to optimize decomposed zero-order Takagi-Sugeno-Kang (TSK) fuzzy rules bases. This kind of rule uses only one feature on the antecedent part and concludes about the class of a new pattern. The gains in processing time, simplicity of implementation and comprehension are notable. Espíndola & Ebecken [5] improved the fuzzy genetic system in order to perform feature selection as well.

The focus of this study is on the way the initial population of the genetic algorithm is generated. On previous work, each allele had the same attribution probability (section 2.3). Then each individual had closely half loci with allele 1, which indicates the presence of a rule. It means that each individual represents a rule subset with half size of the rule base and this causes a delay on performing the fitness function at initial generations. Anyway the best fuzzy classifiers found by the genetic algorithm were formed by less than 10% of the rule bases size. Based on these facts it is intended to speed up the evolutive process by reducing the attribution probability of allele 1, allowing initial individuals with fewer rules. It is also investigated if this change affects the population diversity and average fitness. Some databases from UCI Machine Learning Repository were studied and the results compared to others obtained from a decision tree tool applied to the same problems.

In the next section, the fuzzy genetic system is shown. In section 3, the experiments realized are presented and commented. In order to verify the quality of the study, an efficient decision tree tool was employed to this problem. In the last section, final comments are exposed.

2 The fuzzy genetic system

2.1 Rule base generation

The process of rule generation was presented in Evsukoff et al. [6] which applied the decomposition scheme proposed by Kosko [7] to zero-order TSK fuzzy rules. In this work, each feature space was normalised and divided in five partitions defined by triangular membership functions associated to the linguistic labels small, medium small, medium, medium large and large. They are:

$$\mu_{\text{small}}(x) = \max\{0, 1 - 4 \cdot |x|\} \quad (1)$$

$$\mu_{\text{medium_small}}(x) = \max\{0, 1 - 4 \cdot |x - 0.25|\} \quad (2)$$

$$\mu_{\text{medium}}(x) = \max\{0, 1 - 4 \cdot |x - 0.5|\} \quad (3)$$

$$\mu_{\text{medium_large}}(x) = \max\{0, 1 - 4 \cdot |x - 0.75|\} \quad (4)$$

$$\mu_{\text{large}}(x) = \max\{0, 1 - 4 \cdot |x - 1|\} \quad (5)$$

The rules are constructed in such a way that they inform, besides the class, an output value which can be obtained by two ways. On the first view, this value is interpreted as the influence of each linguistic label of an attribute in the occurrence of a determined class. This influence is a frequency estimation of a class given a record or a pattern. Given training patterns $x^m = (x_1^m, \dots, x_n^m)$ with classes $y^m = (y_1^m, \dots, y_K^m)$, the influence value ϕ_{kj}^i is obtained from the following expression:

$$\phi_{kj}^i = \frac{\sum_{m=1}^M \mu_{X_{ij}}(x_i^m) \cdot y_k^m}{\sum_{m=1}^M y_k^m} \quad (6)$$

in which k is the class influenced by the linguistic label j of the feature X_i .

Alternatively, the output value can be computed by the fuzzy subethood measure. Given training patterns $x^m = (x_1^m, \dots, x_n^m)$ with classes $y^m = (y_1^m, \dots, y_K^m)$, the degree to which the set of the antecedent $X_{i,j}$ is subset of the set of class k is:

$$\phi_{kj}^i = \frac{\sum_{m=1}^M \mu_{X_{ij}}(x_i^m) \cdot y_k^m}{\sum_{m=1}^M \mu_{X_{ij}}(x_i^m)} \quad (7)$$

Thus, the rules have the following structure:

Rule $R_{k,j}^i$: If x_i is $X_{i,j}$ then class = k with output value $\pi_{k,j}^i = \phi_{k,j}^i$

in which $k = 1, \dots, K$ e $j = 1, \dots, 5$.

Considering a database with n attributes and K possible classes, a rule base generated in this way has $5 \cdot K \cdot n$ elements.

2.2 Classification of new patterns

Given a pattern $(x_1, \dots, x_n)$, to determine its class requires the execution of the following steps:

1. for each feature i , combine the outputs of those rules related to the same class k

$$\pi_k^i = \frac{\sum_{j=1}^5 \phi_{kj}^i \cdot \mu_{X_{i,j}}(x_i)}{\sum_{j=1}^5 \mu_{X_{i,j}}(x_i)} = \sum_{j=1}^5 \phi_{kj}^i \cdot \mu_{X_{i,j}}(x_i), \text{ in which } k = 1, \dots, K. \quad (8)$$

2. for each class, aggregate the previous combined outputs by the following function:

$$\pi_k = \min\{\pi_k^i / i = 1, \dots, n\} \quad (9)$$

3. the class with the highest value π_k is associated to the pattern.

2.3 The genetic algorithm

In many problems, to use all the rule base on the classification of a new pattern can not be appropriate due to, among other reasons, an excessive delay to the pattern evaluation and the rule base may not produce the correct class due to the negative influence of rules with low confidences. The aim of this genetic algorithm is to select a small subset of rules with the greatest accuracy and which employs possibly few features.

Each rule subset is a candidate solution and is represented as a chromosome using the binary alphabet $\{0,1\}$. Each rule is represented by a gene in the chromosome. If the rule is present on the subset, its correspondent gene receives the allele 1. Otherwise, it receives the allele 0.

The amount of rules of a candidate solution S acts as a penalty factor on the fitness function, since small subsets are desirable. The same occurs with the amount of features employed. The ability of correctly classifying patterns from a

database by S is more important than the previous penalty factors. Thus, that characteristic must have a distinct weight on the constitution of the fitness function defined by:

$$f(S) = W_{NCP} \cdot NCP(S) - W_C \cdot C(S) - W_F \cdot F(S) \quad (10)$$

in which:

- $NCP(S)$ is the amount of patterns correctly classified by S and W_{NCP} is its weight;
- $C(S)$ is the amount of rules – cardinality - of S and W_C is its weight;
- $F(S)$ is the amount of features employed by the rules of S and W_F is its weight.

Due to higher importance of the accuracy, $0 < W_C, W_F \ll W_{NCP}$. In the experiments, W_{NCP} was set to 1000, W_S and W_F to 1, that is, the cardinality of a rule subset is as important as the amount of features employed.

As genetic operators, it was applied the uniform recombination with probability of 0.5 and the selection of the best chromosome pairs without restitution. In other words, all the individuals are selected in pairs in a decreasing fashion to perform recombination.

Considering the aim of reducing the amount of rules in each individual, two strategies of mutation were defined. The first one, $P_m(1 \rightarrow 0) = 0.1$, executes the change of the allele 1 to allele 0. The other one, $P_m(0 \rightarrow 1) = 0.0075$, changes the allele 0 to allele 1. Also, in each generation, elitism was applied in order to maintain the best individual found by the search process.

It was worked with a population of 20 individuals and all of them were initiated randomly. At each locus, two values of attribution probability of allele 1 during the generation of initial population were employed, 0.50 and 0.25, the latter being the focus of this work. When an evolutive cycle had produced 500 generations the algorithm was stopped.

3 Experimental results and analysis

3.1 Experiments performed

In order to evaluate the performance of the fuzzy genetic system, eight datasets obtained in www.ics.uci.edu/~mllearn/MLRepository.html were studied. Table 1 shows the datasets, their dimensions and the amount of rules generated. Moreover, a decision tree tool was applied to have its results compared. This decision tree tool is part of a system called Weka which is developed at the University of Waikato in New Zealand (www.cs.waikato.ac.nz/ml/weka). Weka contains an implementation of the very known decision tree algorithm C4.5 revision 8. Before working with the datasets, some changes were made to allow the analysis of the method. Repeated records or records with incomplete information were eliminated and qualitative features were converted to discrete quantitative features.

Table 1: Summary of datasets' characteristics.

Database	Valid features	Classes	Valid records	Training examples	Test examples	Rules generated
Car evaluation	6	4	1728	1152	576	120
Credit card application approval	15	2	653	416	237	150
Glass identification	9	7	214	143	71	315
Iris plant	4	3	150	150	150	60
Pima indian diabetes	8	2	768	510	258	80
Wine recognition	13	3	178	120	58	195
Wisconsin breast cancer	9	2	451	300	151	90
Wisconsin diagnostic breast cancer	30	2	569	379	190	300

As genetic algorithms are based on probabilistic operations, the results obtained from two executions with given parameters setting are not necessarily equal. Then, 20 executions of the genetic algorithm were performed to each setting and the medium performance was calculated. To each database, 4 experiments with the fuzzy genetic system were performed with both interpretations of the output value – influence degree and fuzzy subsethood – and both genetic algorithm initiation strategies.

3.2 Results analysis

In the following tables, the letters R, F and A are related to amount of rules, amount of features employed and accuracy (%), respectively. The use of influence degree is represented by the term “inf” and fuzzy subsethood by “sub”. The values after these terms indicate the attribution probability of allele 1: “50” and “25” denote probability of 0.50 and 0.25, respectively.

Tables 2 and 3 present the results obtained by the fuzzy genetic system and Table 4 by the decision tree tool on the databases studied. As it can be noticed, there is no significant divergence on the results obtained by the fuzzy genetic system on the four experiments performed on each dataset. The use of lower attribution probability of allele 1 did not prevent the fuzzy genetic system to find classifiers as good as those obtained with the previous value.

Besides, the average performance of the classifiers obtained by the fuzzy genetic system is close to that achieved by the decision trees generated. Except

for iris plant identification and car evaluation problems, the worst fuzzy classifiers are better than the decision trees in terms of accuracy.

Table 2: Results obtained by the fuzzy genetic system – part I.

Settings	Fuzzy classifiers obtained – Iris plant								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	4	2	97.3	4.1	2.1	97.4	6	3	98.0
inf-25	4	2	97.3	4.1	2.1	97.4	6	3	98.0
sub-50	5	2	97.3	4.3	2.1	97.4	8	4	98.0
sub-25	5	2	97.3	4.2	2.0	97.3	4	2	97.3
Settings	Fuzzy classifiers obtained – Wine recognition								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	8	6	94.8	7.3	5.3	98.8	5	4	100
inf-25	13	7	96.6	8.2	5.6	98.9	6	3	100
sub-50	10	6	96.6	7.2	4.8	99.2	5	3	100
sub-25	11	8	98.3	6.5	4.5	99.5	5	3	100
Settings	Fuzzy classifiers obtained – Wisconsin breast cancer								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	11	5	98.7	9.5	5.4	99.4	9	5	100
inf-25	8	5	98.7	8.5	4.9	99.5	9	5	100
sub-50	9	4	98.7	9.5	5.4	99.6	9	4	100
sub-25	8	5	98.7	8.7	5.2	99.5	8	3	100
Settings	Fuzzy classifiers obtained – Wisconsin diagnostic breast cancer								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	16	10	94.2	11.2	8.7	94.9	9	9	95.8
inf-25	10	6	93.7	10.4	8.8	94.7	14	12	95.8
sub-50	13	11	94.2	10.5	8.2	94.5	10	7	95.3
sub-25	10	10	93.2	9.8	7.8	94.6	9	6	96.8

Table 3: Results obtained by the fuzzy genetic system – part II.

Settings	Fuzzy classifiers obtained – Credit card application approval								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	4	4	88.2	5.6	5.0	89.7	8	5	92.0
inf-25	6	5	88.6	6.3	5.6	90.0	9	7	92.0
sub-50	5	3	89.0	7.9	5.8	90.5	9	8	91.6
sub-25	4	3	88.6	8.0	5.8	90.6	10	6	92.0
Settings	Fuzzy classifiers obtained – Car evaluation								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	10	5	79.7	7.3	4.1	84.3	7	4	88.0
inf-25	6	3	74.7	7.4	4.4	83.5	7	4	88.0
sub-50	7	4	81.8	6.1	3.1	81.8	6	3	81.8
sub-25	7	4	81.8	6.4	3.4	81.8	6	3	81.8
Settings	Fuzzy classifiers obtained – Pima Indian diabetes								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	7	2	81.8	8.9	5.4	83.8	9	6	84.9
inf-25	9	6	80.6	8.2	5.3	83.2	9	6	84.9
sub-50	6	3	81.8	7.3	4.2	84.1	8	5	85.7
sub-25	4	4	81.4	7.2	4.4	84.2	8	5	85.7
Settings	Fuzzy classifiers obtained – Glass identification								
	Worst			Average			Best		
	R	F	A	R	F	A	R	F	A
inf-50	12	8	70.4	10.7	6.3	75.7	12	6	80.3
inf-25	8	5	70.4	11.0	6.2	74.9	10	5	80.3
sub-50	9	6	73.2	9.8	6.2	77.3	11	7	81.7
sub-25	9	5	70.4	10.3	5.8	76.7	10	6	83.1

To compare the populations generated by the two attribution probabilities, Table 5 presents the average results. The last column shows the division of the average fitness of the first generation with probability 0.25 by the one with 0.50.

Table 4: Decision trees obtained on all databases.

Dataset	R	F	A
Car evaluation	5	2	98.0
Credit card application approval	5	4	87.9
Glass identification	7	5	97.4
Iris plant	8	6	90.5
Pima indian diabetes	18	12	86.9
Wine recognition	42	6	86.6
Wisconsin breast cancer	14	7	77.5
Wisconsin diagnostic breast cancer	21	7	69.0

Table 5: Average initial populations generated by the genetic algorithm.

Dataset	Attribution probability	Rules diversity	Amount of rules	Average fitness ratio
Car evaluation	0.25	119.9	609.3	> 95%
	0.50	120.0	1216.0	
Credit card approval	0.25	149.5	765.1	> 96%
	0.50	150.0	1506.2	
Glass identification	0.25	311.5	1629.2	> 98%
	0.50	315.0	3161.7	
Iris plant	0.25	59.5	299.7	> 92%
	0.50	60.0	599.1	
Pima indian diabetes	0.25	79.3	397.9	> 97%
	0.50	80.0	807.4	
Wine recognition	0.25	194.8	991.4	> 97%
	0.50	195.0	1971.1	
W. breast cancer	0.25	89.5	464.9	> 98%
	0.50	90.0	911.4	
W. diagn. breast cancer	0.25	300.0	1529.0	> 98%
	0.50	299.1	3038.7	

As it can be noticed, the amount of rules used at first generation was drastically reduced and the population diversity was maintained. It can be seen from the ratios that the average fitness was not significantly reduced – less than 8% on the worst case, the iris plant problem.

4 Final considerations

As observed, the fuzzy genetic system obtained very good results in every case studied when compared to a very known decision tree algorithm. The quality of a classifier was estimated by some characteristics considered important in this work such as accuracy, comprehensibility, amount of rules and features employed. The objective of finding accurate and relatively short fuzzy classifiers and feature selectors was successful. The results obtained from the eight problems suggest that the reduction of the attribution probability allowed the speeding of the evolutive cycle without sacrificing the system performance. Higher dimension problems may emphasize the relevance of this setting. Future studies may consider other kinds of linguistic terms and alternative ways of performing the genetic operations.

Acknowledgements

This research was supported by the Brazilian Petroleum National Agency under the program PRH-ANP/MME/MCT.

References

- [1] Gordon, A.D., *Classification*, Chapman and Hall: London, 1981.
- [2] Liu, H. & Motoda, H., 1998, *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers: Boston, 1998.
- [3] Ishibuchi, H., Murata, T. & Tanaka, H., Construction of Fuzzy Classification System with Linguistic If-Then Rules Using Genetic Algorithms (Chapter 11). *Genetic Algorithms for Pattern Recognition*, eds S.K. Pal & P.P. Wang, CRC Press: New York, pp. 227-251, 1996.
- [4] Espíndola, R.P. & Ebecken, N.F.F., Evolving TSK fuzzy rules for classification tasks by Genetic Algorithms. *Proceedings of Second International Conference on Data Mining*, eds N.F.F. Ebecken & C.A. Brebbia, WIT Press: Southampton, pp. 467-476, 2000.
- [5] Espíndola, R.P. & Ebecken, N.F.F., Seleção de atributos e classificação por regras fuzzy TSK otimizadas por algoritmos genéticos. *Proceedings of 22nd Iberian Latin-American Congress On Computational Methods In Engineering*, São Paulo, 2001.
- [6] Evsukoff, A., Branco, A.C.S. & Gentil, S., A Knowledge Acquisition Method for Fuzzy Expert System in Diagnosis Problems. *Proceedings of Sixth IEEE International Conference on Fuzzy Systems*, Barcelona, 1997.
- [7] Kosko, B., *Neural Networks and Fuzzy Systems - A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall: New Jersey, 1992.