A case study on Morphological Data from Eimeria of Domestic Fowl using a Multiobjective Genetic Algorithm and R&P for Learning and Tuning Fuzzy Rules for Classification

Edward Hinojosa C.
Dept. Informatics Engineering
Pontifical Catholic University of Peru
Lima, Peru
ehinojosa@pucp.pe

Cesar A. Beltran C.
Dept. Informatics Engineering
Pontifical Catholic University of Peru
Lima, Peru
ebeltran@pucp.pe

Abstract

In this paper, we use fuzzy rule-based classification systems for classify cells of the Eimeria of Domestic Fowl based on Morphological Data. Thirteen features were extracted of the images of the cells, these features are genetically processed for learning fuzzy rules and a method reward and punishment for tuning the weights of the fuzzy rules. The experimental results show that our classifier based on interpretability fuzzy rules has a similar classification rate to that of a non-parametric and non-interpretability method.

1 Introduction

The fuzzy systems were proposed by Zadeh at 1965 (Zadeh, 1965) and they are systems based on the theory of the fuzzy sets and logic fuzzy. A of the most important types of fuzzy systems are the Fuzzy Rule Based Classification Systems (FRBCSs) (Herrera, 2005) (Herrera, 2008). Classification problem is studied in the machine learning, data mining, database, and information retrieval communities with applications in a several domains.

The rules are a paradigm for representing knowledge and they have the capacity to build a linguistic model interpretable to the users. The learning (or automatic generation) and tuning of the fuzzy rules in FRBCSs from data sample is a difficult task (Herrera, 2008). This task can be considered as an optimization or search process that can be managed by using Evolutionary Algorithms (EAs). The Genetic Algorithms (GAs) is one of the most know and highly used of EAs. The FRBCSs are defined as Genetic Fuzzy Rule-Based Systems (GFRBSs) when the GAs are used to learn or tuning FRBCSs. The GFRBSs continue to be researched and used in recent years (Nojima and Ishibuchi, 2013), (Chen et al., 2013), (Jalesiyan et al., 2014).

Generally a FRBCSs is composed of two components (Herrera, 2005), the Knowledge Base (KB) and the Inference Mechanism (IM). The KB is composed of two components too, the Data Base (DB) and the Rule Base (RB). This paper is concerned with the genetic learning of the RB.

The most commonly used approaches for rule learning in FRBCSs using GAs are Pittsburgh, Michigan, Iterative Rule Learning (IRL) and Genetic Cooperative-Competitive Learning (GCCL). In the Pittsburgh approach, each chromosome encodes a set of fuzzy rules, after the genetic process the RB is a better chromosome (De Jong et al., 1993). In the Michigan approach, each chromosome encodes a single rule, after the genetic process the RB is the set of chromosomes or rules of the population (Holland and Reitman, 1978). In the IRL approach, each chromosome encodes a single rule too, but after the genetic process, the better rule is selected and inserted to the RB, this process is repeated iteratively until a condition is satisfied (Gonzalez and Perez, 2012). The GCCL approach is a hybrid of the Pittsburgh and Michigan approaches, the rules or chromosomes cooperate among themselves based on Pittsburgh approach and the rules or chromosomes compete among themselves based on Michigan approach (Giordana and Neri, 1995).

This paper is based in the IRL approach using a

Multiobjective Genetic Algorithms (MOGAs). We use MOGAs because in the process of learning fuzzy rules in FRBCSs are considered two objectives: accuracy and interpretability. This objectives are considered contradictory (Casillas and Carse, 2009) and we search a trade-off of them. The accuracy is measured by the classification rate and the interpretability is measured for many features of the FRBCSs, for example, quantity of the rules or quantity of the conditions of each rule. We use specifically the well-known algorithm called Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002). After the learning the fuzzy rules, we use a Reward and Punishment(R&P) method for the tuning the factors or weights of the rules (Nozaki et al., 1996) to improve the accuracy of the FRBCS.

We use the proposed method for classify cells of the Eimeria of Domestic Fowl. The Eimeria genus comprises a group of protozoan parasites that infect a wide range of hosts. A total of seven different Eimeria species infect the domestic fowl, causing enteritis with severe economic losses. We use three groups of morphological features: geometric measures, curvature characterization, and internal structure quantification (Beltran, 2007).

This paper is organized as follows: we present in Section 2 the basic concept of classification and FRBCSs employed in this paper. In Section 3 we describe the genetic algorithm multiobjetivo called NSGA-II used in this paper. The proposed method for learning the RB and tuning the factor of each rule is detailed in Section 4. The Section 5 shows the results of the classification on morphological features of the Eimeria genus. The conclusions of this work are presented in Section 6.

2 Fuzzy Rule Based Classification Systems

Classification problem is studied in the machine learning, data mining, database, and information retrieval communities with applications in a several domains, such as medical (Kumar et al., 2013), target marketing (Yongzhi et al., 2013), biology (Silla and Kaestner, 2013), among others.

Any classification problem has a set of examples $E = \{e_1, e_2, ..., e_p\}$ and a set of classes $C = \{C_1, C_2, ..., C_m\}$, the objective is labeled each example $e_q \in E$ with a class $C_j \in C$. Each e_q is defined by a set of features or characteristics $e_q = \{a_{q1}, a_{q2}, ..., a_{qn}\}$.

A FRCS resolves classification problems using rules usually with the follow structures:

R_i: **IF** V_1 **IS** T_{1l_1} **AND** V_2 **IS** T_{2l_2} **AND** ... **AND** T_n **IS** T_{nl_n} **THEN** Class = C_j **WITH** CF_i

where:

R_i	: Index of the fuzzy rule i.
$V_1, V_2, ..., V_n$	: Linguistic variables or features of each example e_q.
$T_{1l_1}, T_{2l_2}, ..., T_{nl_n}$	: Linguistic terms or fuzzy sets used for representing the class C_j.
C_j	: The class of the fuzzy rule R_i.
CF_i	: The certainty grade (i.e. rule weight) of the rule R_i.

Usually a FRBCS has two main components (Herrera, 2005): The Knowledge Base (KB) and the Inference Mechanism (IM), these are detailed below:

1. The Knowledge Base: The KB is composed of two components:
   (a) The Data Base: The DB contains the membership functions, fuzzy sets or linguistic terms for each linguistic variable of the classification problem.
   (b) The Rule Base: The RB contains the collection of fuzzy rules representing the knowledge.
2. The Inference Mechanism: The IM is the fuzzy logic reasoning process that determines the outputs corresponding to fuzzified inputs (Lakhmi and Martin, 1998). The most common fuzzy inference method for fuzzy classification problems are the classic and general reasing methods (Cordon et al., 2013). This paper uses the classic method.

3 Non-dominated Sorting Genetic Algorithm Multiobjetive II

Is a new version of the NSGA (Srinivas and Deb, 1994), the NSGA-II was proposed by Deb in 2002 (Deb et al., 2002) and it is computationally more efficient, elitist and doesnt need to define additional parameters.

In the NSGA-II, the population Q_t (size N) is generated using the parent population P_t (size N). After this, the two populations are combined for generating the population R_t (size $2N$). The population R_t is sorted according the dominance of the solutions in different Pareto fronts (Pareto, 1896) and the crowding distance. A new population P_{t+1} (size N) is generated with the bests Pareto fronts F_1, F_2, F_3 and so forth, until the P_{t+1} size equals to the value of N. The solutions in the Pareto fronts under this limit are removed. After P_{t+1} is a new P_t and the process is repeated until a conditions is satisfied. The figure 1 shows the process of evolutions of the solutions in the NSGA-II. More details on NSGA-II can be found at (Deb et al., 2002).

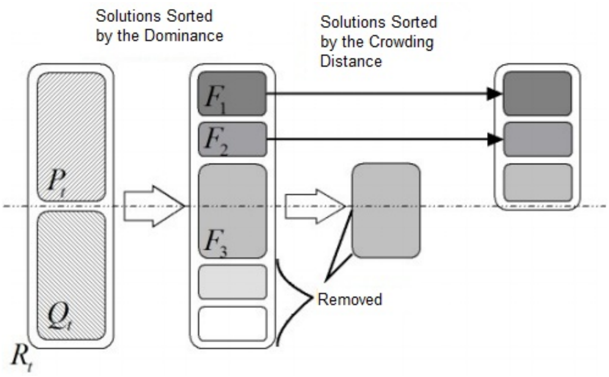

Figure 1: Evolutions of the Solutions in the NSGA-II

4 Proposed Method

This section presents the proposed methods for learning fuzzy rules using the IRL approach and a MOGA, and tuning the weights of the fuzzy rules using a R&P method. In the next subsections each method is detailed.

4.1 Learning Fuzzy Rules

The proposed method for learning fuzzy rules is based in the iterative multiobjective genetic method described in (Hinojosa and Camargo, 2012) and uses a MOGA for learning a single fuzzy rule in each iteration of the MOGA. The main difference with the proposed method is the module for defining the order of the class for learning. This method proposed is illustrated in the Figure 2.

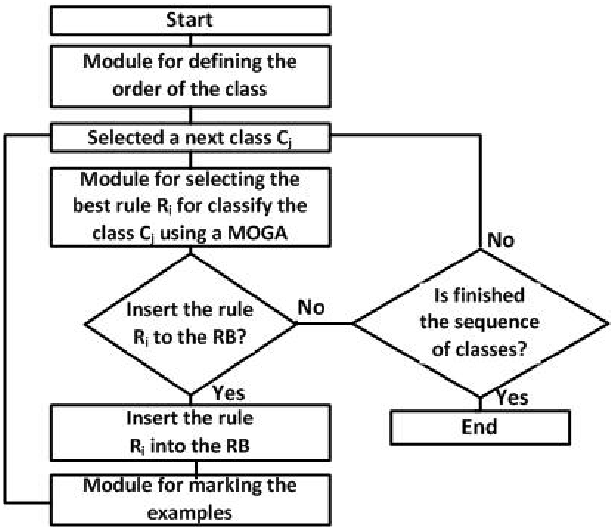

Figure 2: Proposed Method for Learning Fuzzy Rules

A set of examples is used as the set of training. The proposed IRL method starts when is defined the order of the class for learning. After that, a class is selected and the module for generate the best rule that used a MOGA is executed. The MOGA considers two objectives for minimization: accuracy and interpretability. The accuracy is determined by the integrity and consistency of each rule (Gonzalez and Perez, 1999) and the interpretability is defined by the quantity of conditions of each rule. When the best rule in the Pareto front improves the rate of classification of the RB, this rule is inserted into the RB, some examples are marked and the process of learning a fuzzy rule starts again. When the best rule in the Pareto front doesnt improve the rate of classification of the RB, the process verifies that all the sequence of class was learned, if the sequence is not learned a new class is selected and the process of learning a fuzzy rule starts again, else the process finishes and the set of the best rules is the RB.

In the process detailed above all rules has a weight equals to one. These weights can be tuning for improve the rate classification. This tuning is detailed in the next subsection.

4.2 Tunning Weights of the Fuzzy Rules

We use the method proposed in (Nozaki et al., 1996) for this task. This method rewards or increases the weight the a fuzzy rule R_i when a example e_q is correctly classified by this rule according to the next equation:

$$CF_i^{new} = CF_i^{old} + n_1 \left(1 - CF_i^{old}\right) \quad (1)$$

And this method punishes or decreases the weight of the fuzzy rule R_i when a example e_q is misclassifed by this rule according to the next equation:

$$CF_i^{new} = CF_i^{old} - n_2 CF_i^{old} \quad (2)$$

In the experimental study detaild in Section 5 we used the values n1=0.001 and n2=0.1 and the tuning procedure for 500 iterations.

5 Experimental Study

The experimental study is aimed to show the application of the proposed method and the comparation with the classification with non-parametric method for classifying cells of the Eimeria of Domestic Fowl based on Morphological Data. The Emeira genus comprises a group of protozoan parasites that infect a wide range of hosts, seven different Emeira species infect the domestic fowl, causing enteritis with several economic losses. This protozoan morphology was represented by 13 features: mean of curvature, standard deviation of curvature, entropy of curvature, major axis (lenght), minor axis (width), symmetry through major axis, symmetry through minor axis, area, entropy of internal structure, second angular moment, contrast, inverse difference moment, entropy of co-occurrence matrix; these features are used as the input pattern for the classification process.

The Table 1 shows the class and the number of examples or instances of each class. More detail how the features were extracted or about the Eimeria genus can be found at (Beltran, 2007).

The Table 2 shows the parameters for learning fuzzy rules and the NSGA-II (the MOGA used in this paper).

After the process of learning fuzzy rules, the process of tuning the weights starts. The Table 3 shows the result of the classification or dispersion matriz after the process tuning the weights.

After the proposed method, the result is a set of rules similar of the set shows in the Figure 3. This rules has a high level of interpretability for the expert users.

Class Number	Class Name	# of Examples
1	E. acervulina	636
2	E. maxima	321
3	E. brunetti	418
4	E. mitis	757
5	E. praecox	747
6	E. tenella	608
7	E. necatrix	404

Table 1: Distribution of Classes

Parameter	Value
Size the population	50.0
Crossover rate	1.0
Mutation rate	0.2
Number of generations	500.0
Mark value	0.3

Table 2: Parameters of the Proposed Method

```
IF (major_axis IS Small_4) THEN Class IS 4 WITH 1.0
IF (contrast IS Large_4) THEN Class IS 4 WITH 0.1
IF (mean_of_curvature IS Large_4) THEN Class IS 4 WITH 0.7
IF (major_axis IS Small_3) THEN Class IS 4 WITH 0.4
IF (symmetry_through_major_axis IS Large_1) THEN Class IS 6 WITH 1.0
IF (area IS Large_2) THEN Class IS 2 WITH 1.0
IF (area IS Large_3) THEN Class IS 2 WITH 1.0
IF (minor_axis IS Small_3) THEN Class IS 1 WITH 0.3
IF (mean_of_curvature IS Small_40) THEN Class IS 4 WITH 1.0
```

Figure 3: Proposed Method for Learning Fuzzy Rules

We compared the proposed method with the method non-parametric classifier proposed in (Beltran, 2007) with the same set of examples. The Table 4 shows the classification rate by each class of both classifiers. These results shows that the proposed method (PM) has a similar rate classification (overall 77.17) that the non-parametric method (NPM) (overall 80.24), but with a high degree of interpretability. The non-parametric method does not consider the interpretability.

6 Conclusions

In this article, we proposed a iterative multiobjective genetic method to learn fuzzy classification rules. The fuzzy rules are learned in each iteration depend of the sequencia of class. After that, the weights of the each fuzzy rules are tuned using a R&P method. The results obtained have indicated that FRBCSs

	ACE	MAX	BRU	MIT	PRA	TEN	NEC
ACE	**85.06**	0.00	0.00	2.38	0.13	0.16	7.18
MAX	0.00	**98.44**	0.72	0.00	0.00	0.00	0.00
BRU	0.00	1.56	**87.08**	0.00	6.29	1.48	0.74
MIT	1.73	0.00	0.00	**86.79**	4.95	1.64	4.70
PRA	2.67	0.00	5.50	6.87	**69.34**	10.53	24.01
TEN	7.86	0.00	6.70	1.19	16.06	**82.73**	37.62
NEC	2.67	0.00	0.00	2.77	3.21	3.45	**25.74**

Table 3: Results of the Proposed Method

Class Name	PM	NPM
E. acervulina	85.06	87.70
E. maxima	98.44	96.12
E. brunetti	87.08	94.98
E. mitis	86.79	86.27
E. praecox	69.34	64.46
E. tenella	82.73	76.53
E. necatrix	25.74	55.60

Table 4: The PM vs. NPM

have better interpretability and similar accuracy than a non-parametric method for classify the Eimeria of domestic fowl.

References

Zadeh L. A. 1965. Fuzzy Sets. *Information and Control*, 8(3):338–353.

Herrera F. 2005. Genetic fuzzy systems: Status, critical considerations and future directions. *International Journal of Computational Intelligence Research*, 1(1):59–67.

Herrera F. 2008. Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1(1):27–46.

Nojima, Y. and Ishibuchi, H. 2013. Multiobjective genetic fuzzy rule selection with fuzzy relational rules. *IEEE International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS)*, 1:60–67

Chen, S.-M, Chang, Y.-C., Pan, J.-S. 2013. Fuzzy Rules Interpolation for Sparse Fuzzy Rule-Based Systems Based on Interval Type-2 Gaussian Fuzzy Sets and Genetic Algorithms. *IEEE Transactions on Fuzzy Systems*, 21(3):412–425.

Jalesiyan, H., Yaghubi, M., Akbarzadeh, T.M.R. 2014. Rule selection by Guided Elitism genetic algorithm in Fuzzy Min-Max classifier. *Conference on Intelligent Systems*, 1:1–6.

De Jong KA, Spears WM, Gordon DF. 1993. Using genetic algorithms for concept learning. *Mach Learn*, 13:161–188.

Holland, J. and Reitman, J. 1978. *Cognitive Systems Based on Adaptive Algorithms*, ACM SIGART Bulletin

Gonzalez, A. and Perez, R. 1999. SLAVE: A genetic learning system based on an iterative approach. *IEEE Transactions on Fuzzy Systems*, 7:176–191.

Giordana, A. and Neri, F. 1995. Search-intensive concept induction. *Evol Comput*, 3:375–416.

Casillas, J. and Carse, B. 2009. Special issue on Genetic Fuzzy Systems: Recent Developments and Future Directions. *Soft Comput.*, 13(5):417–418.

Deb, K. and Pratap, A. and Agarwal, S. and Meyarivan, T. 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *Trans. Evol. Comp.*, 6(2):182–197.

Nozaki, k., Ishibuchi, H., Tanaka, H. 1996. Adaptive fuzzy rule-based classification systems. *IEEE Trans. Fuzzy Systems*, 4(3):238–250.

Beltran, C. 2007. Anlise e reconhecimento digital de formas biolgicas para o diagnstico automtico de parasitas do gnero Eimeria. *PhD. Teses - USP - Brazil*.

Kumar, S.U., Inbarani, H.H., Kumar, S.S. 2013. Bijective soft set based classification of medical data. *International Conference on Pattern Recognition, Informatics and Mobile Engineering*, 1:517–521.

Yongzhi, Ma., Hong Gao, Yi Ding, Wei Liu. 2013. Logistics market segmentation based on extension classification. *International Conference on Information Management, Innovation Management and Industrial Engineering*, 2:216–219.

Silla, C.N. and Kaestner, C.A.A. 2013. Hierarchical Classification of Bird Species Using Their Audio Recorded Songs. *IEEE International Conference on Systems, Man, and Cybernetics*, 1:1895–1900.

Lakhmi C. and Martin N. 1998. *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications (International Series on Computational Intelligence)*. CRC Press

Cordon, O.,Herrera, F., Hoffmann, F., Magdalena, L. 2001. *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases* World Scientific

Srinivas, N. and Deb, K. 1994. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2:221–248.

Pareto, V. 1896. *Cours d'Economie Politique*. Droz, Genve.

Hinojosa, E. and Camargo, H.A. 2012. Multiobjective genetic generation of fuzzy classifiers using the iterative rule learning. *International Conference on Fuzzy Systems*, 1:1–8.

Gonzalez, A. and Perez, R. 2009. Improving the genetic algorithm of SLAVE. *Mathware and Soft Computing*, 16:59–70.