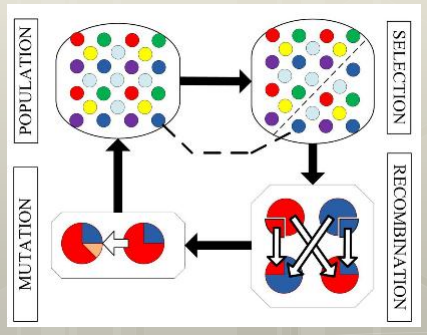


ICOR 2014: International Conference on Operations Research

GENETIC ALGORITHMS WITH DIVERSITY MEASURES FOR MULTIPLE CLASSIFIER SYSTEMS

Authors:
Leidys Cabrera Hernández
Alejandro Morales Hernández
Gladys M. Casas Cardoso
Yailen Martínez Jiménez

Universidad Central «Marta Abreu» de Las Villas
Santa Clara, Villa Clara, Cuba

1

Introduction

- The combination of classifiers is an active research area of the machine learning communities.
- Diversity measures are used to decide what base classifiers combine.
- Genetic algorithm are emerging as tools for solving complex search and optimization problems.

2

The Objective

- To combine
  - diversity measures results
  - accuracy classifiers results
- using Genetic Algorithms to obtain a good multiclassifier.

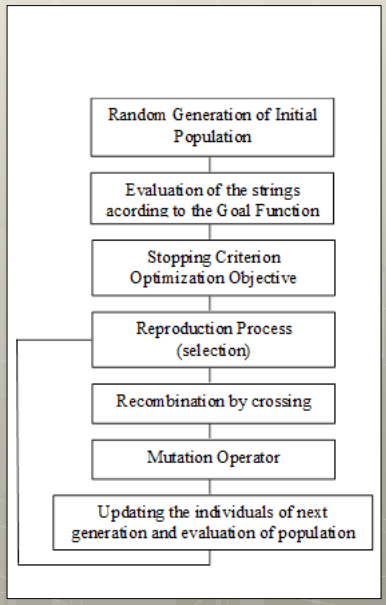
3

Diversity in Classifiers Ensembles

- **The Disagreement Measures (D)**[1]
  - Diversity within the three or more base classifiers are calculated by averaging all pair of classifiers.
- **The Double Fault Measure (DF)** [1]
  - The diversity decreases when the value of the double fault measure increases.
- **The Correlation Coefficient Measure (ρ)** [1]
  - The diversity of two predictors is inversely proportional to the correlation between them.
- **The Q Statistic** [1]
  - Classifiers that tend to recognize the same objects correctly will have positive values of Q. It can be proved that $|\rho| \leq |Q|$

1 Kuncheva, L. "Combining Pattern Classifiers: Methods and Algorithms", 2004

4



Functional diagram of a Genetic Algorithm

5

Description

- Goal Function
  - Max f(X)
  - X is the chromosome
  - f = the evaluation of the multiple classifier system + the evaluation of the diversity measures
- Chromosome configuration
  - **Gen**: binary variable: It takes the value 1 if the classifier belongs to the combination.
  - **Chromosome**: the set of all base classifiers

6

Description

- Configuration of the population
  - The population size can be calculated as $2^{\frac{Sc}{2}}$, where Sc denotes the size of the chromosome
- Probability of mutation: 0.2
- Probability of crossover: 0.75

7

Description

- Selection operator
  - The roulette method was implemented.
- Input parameters
  - Generations: 100
  - Diversity Measures:
    - D (The Disagreement Measures (D))
    - DF (The Double Fault Measure)

8

Real Aplication: Hepatitis [1]

Classifier	Accuracy
Naivesbayes	0.83
Multilayer Perceptron	0.81
Lazy.IBK	0.79
Trees.J48	0.79
Functions.Logistic	0.84
Random Tree	0.83
Trees.LMT	0.84
Functions.SMO	0.88
Lazy.KStar	0.90
Functions.SGD	0.88

1 UCI Repository of Machine Learning Databases

9

Real Aplication: Hepatitis [1]

Hepatitis	Measure Value	Goal Function	Chromo-some	Multiple Classifier System
D	0.17	1.11	0100010101	0.94
DF	0.94	1.88	0100010101	0.94
D + DF	0.56	1.50	0100010101	0.94

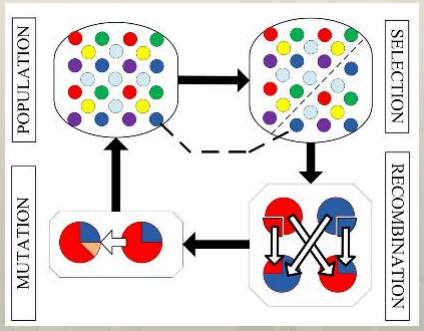
1 UCI Repository of Machine Learning Databases

10

Conclusions

- This paper shows a novel technique using genetic algorithms to find a good subset of diverse classifiers.
- A case of study of Hepatitis data was used to exemplify the contribution.
  - Ten base classifiers were applied and their individual accuracies were not higher than 90%.
  - By using the proposed genetic algorithm with diversity measures we obtain a multiclassifier with a better accuracy: 94%.

11



ICOR 2014: International Conference on Operations Research

GENETIC ALGORITHMS WITH DIVERSITY MEASURES FOR MULTIPLE CLASSIFIER SYSTEMS

Authors:
Leidys Cabrera Hernández
Alejandro Morales Hernández
Gladys M. Casas Cardoso
Yailen Martínez Jiménez

Universidad Central «Marta Abreu» de Las Villas
Santa Clara, Villa Clara, Cuba

12