UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

BioInformatics Research Group

Workshop

KOWLEDGE EXTRACTION BASED ON EVOLUTIONARY ALGORITHMS

HIDER: Method and Natural Coding

Raúl Giráldez

School of Engineering - Division of Computer Science

Pablo de Olavide University of Seville

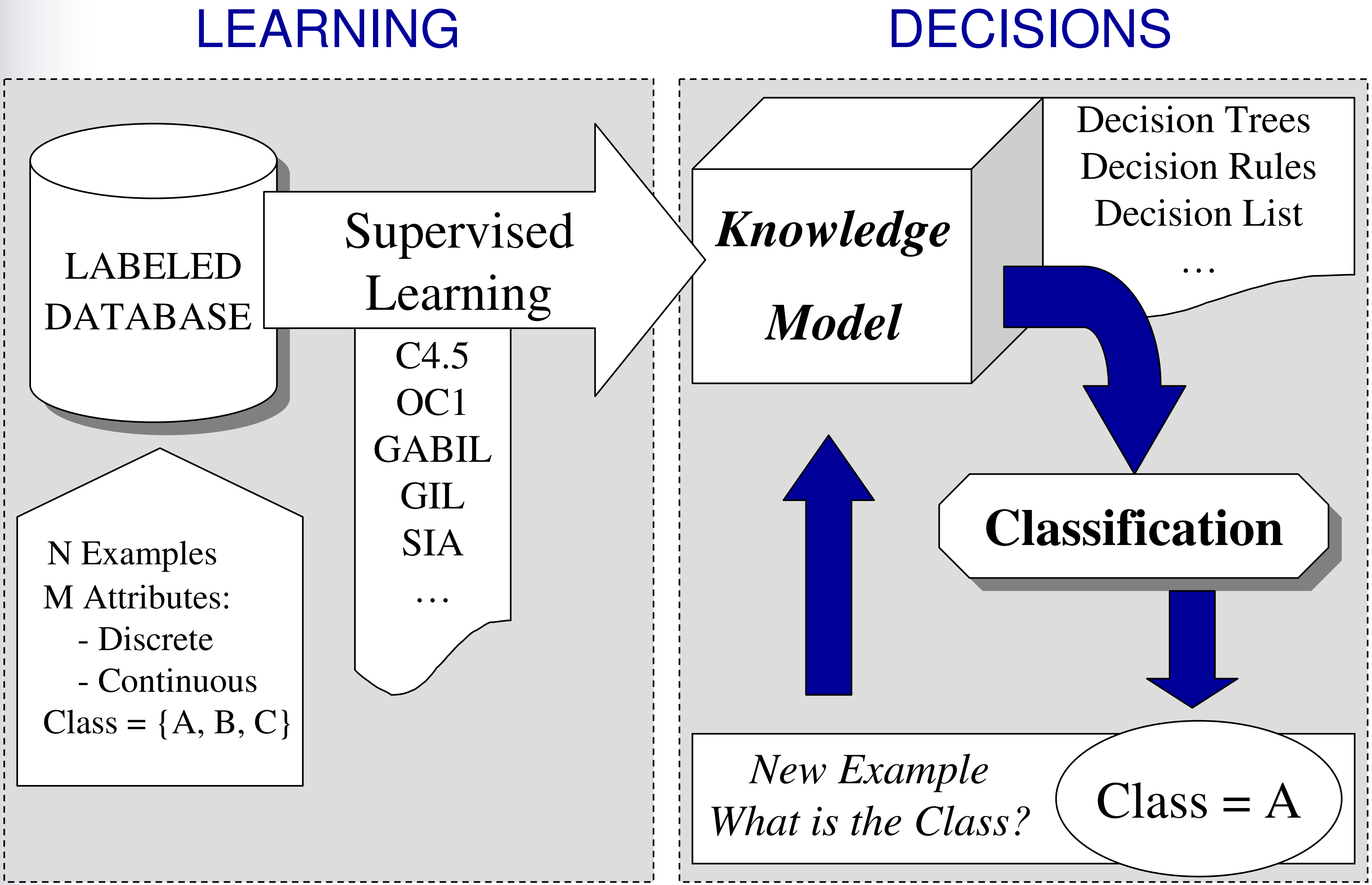
May 15, 2005 – Granada, Spain

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

BioInformatics Research Group

Contents

Introduction: The context



Introduction: Preliminaries

Representations

- Decision Trees
- **<u>Decision Rules</u>**
- Fuzzy Rules
- etc

Methods

- Induction
- **<u>Evolutionary Algorithms</u>**
- Divide & Conquer
- etc

Hierarchy

Hybrid Coding

HIDER (Hierarchical Decision Rules)
[Aguilar-Riquelme-Toro-03]

Natural Coding

Other Improvements

[Giraldez-05]
[Giraldez-Aguilar-Riquelme-05]

HIDER*
[Giraldez-Aguilar-Riquelme-07]

Introduction: The problem

HIDER → Hybrid Coding
HIDER* → Natural Coding

LABELED DATABASE

HIDER

Evolutionary Algorithm
- **Coding**
  Individual = Encoded Rule
- Evaluation
  F(goals, error, ...)

Set of Hierarchical Decision Rules

R1: **If** $Cond_{11}$ **And** $Cond_{12}$ **And** ... $Cond_{1n}$ **Then** $Class_1$
R2: **Else If** $Cond_{21}$ **And** $Cond_{22}$ **And** ... $Cond_{2n}$ **Then** $Class_2$
... ...
Rm: **Else** *"Unknown Class"*

$$Cond_{ij} = \begin{cases} a_j \in [l_i, u_i] & \textit{if } a_j \textit{ es continuous} \\ a_j \in \{v_1, v_2, \ldots, v_k\} & \textit{if } a_j \textit{ es discrete} \end{cases}$$

Method: Fitness Function

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

$$\Phi(r) = N - CE(r) + G(r) + coverage(r)$$

- r is an individual, that is a encoded rule;
- N = Number of Examples
- CE(r) = Class Error (n. of missclassified examples)
- G(r) = Goals (n. of examples correctly classified)
- coverage(r) = Space covered by the rule

Method: Algorithm


```
Procedure HIDER*(E, R)
    R := ∅
    while |E|>0
        r:=EvoAlg(E)
        R:=R ⊕ r
        E:=E-{e ∈ E | e ⊆ Δr}
    End while
fin HIDER*
```

E = Training File
R = Set of Rules
r = Rule
$\oplus$ = Insertion
Δ_r = Coverage of r

P = Population

```
Function EvoAlg(E)
    InitializePopulation(P)
    for i:=1 to num generations
        Evaluate(P)
        next_P:=SelectTheBestOf(P)
        next_P:=next P+Replicate(P)
        next_P:=next P+Recombine(P)
        P:=next_P
    end for
    Evaluate(P)
    return SelectTheBestOf(P)
fin EvoAlg
```

Hider: Hybrid Coding I

Binary Coding

- Suitable for discrete domains
- Loss of accuracy in continuous domains

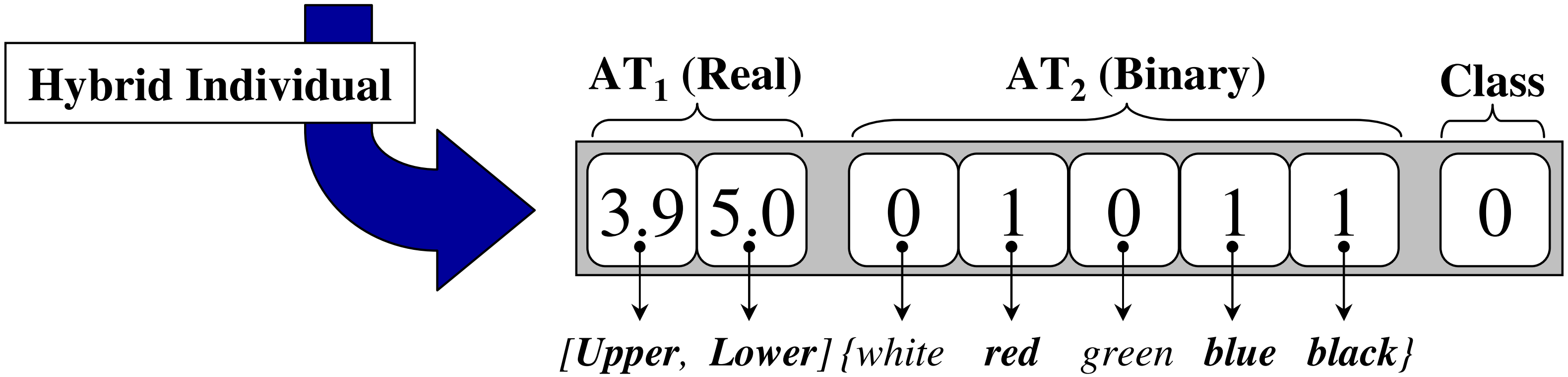
Real Coding

- Suitable for continuous domains
- Very long size of search space

Hybrid Coding (Binary + Real)

- Discrete Attributes ⇒ Binary Coding
- Continuous Attributes ⇒ Real Coding

Rule: If $AT_1 \in [3.9, 5.0]$ and $AT_2 \in \{red, blue, black\}$ Then Class = 0



Hider: Hybrid Coding II

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

Problems:

- Real encoding → Very large size of search space
- Binary encoding → Length of the individuals

Reduction of the search space
+
Reduction of size of individuals

Hider*:
Natural Coding

Hider*: Natural Coding I

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA BioInformatics Research Group

Continuous Attributes

USD cutpoints = {1.4, 2.5, 3.9, 4.7, 5.0, 6.2} [Giraldez-Aguilar-Riquelme-02]

(Very Robust [Aguilar-Bacardit-Divina-04])

Cutpoints	2.5	3.9	4.7	5.0	6.2
1.4	**1**≡[1.4, 2.5]	**2**≡[1.4, 3.9]	**3**≡[1.4, 4.7]	**4**≡[1.4, 5.0]	**5**≡[1.4, 6.2]
2.5	-	**7**≡[2.5, 3.9]	**8**≡[2.5, 4.7]	**9**≡[2.5, 5.0]	**10**≡[2.5, 6.2]
3.9	-	-	**13**≡[3.9, 4.7]	**14**≡[3.9, 5.0]	**15**≡[3.9, 6.2]
4.7	-	-	-	**19**≡[4.7, 5.0]	**20**≡[4.7, 6.2]
5.0	-	-	-	-	**25**≡[5.0, 6.2]

Simple Arithmetic operations

Genetic Operators

- *Mutation(x)* = Shift: up, down, left or right
- *Crossover(x, y)* = Intersection between row and column of *x* and *y*

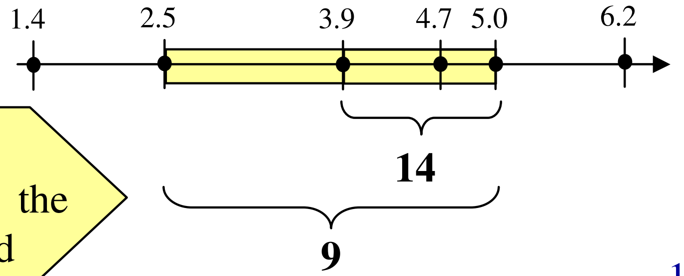
Example:

Mutation(14) = {9, 13, 15, 19}

Mutation(5) = {4, 10}

Crossover(5, 14) = {4, 15}

Shift up=
= Extension of the
lower bound



Hider*: Natural Coding II

Discrete Attributes

Discrete Values					Natural Coding
white	*red*	*green*	*blue*	*black*	
0	0	0	0	0	**0**
0	0	0	0	1	**1**
0	0	0	1	0	**2**
...	...	...	...	...	**...**
0	1	0	1	1	**11**
...	...	...	...	...	**...**
1	1	1	1	1	**31**

Based on Binary Coding

Simple Arithmetic operations

Genetic Operators

- *Mutation(x)* = Change some bits in *x*
- *Crossover(x, y)* = {Mutations(x) ⊕ x} ∩ {Mutations(y) ⊕ y}

mple:

- 11 = 01011 ⇒ Mutations(11) = {10, 9, 15, 3, 27}

11 = 0101**1** = a∈{red, blue, black} → 10 = 0101**0** = a∈{red, blue}

- 19 = 10011 ⇒ Mutations(19) = {18, 17, 23, 27, 3}
- Crossover(19, 11) = {Mutations(19) ⊕ 19} ∩ {Mutations(11) ⊕ 11} = {3, 27}

Hider*: Natural Coding III

Hybrid Coding vs. Natural Coding

Example
- AT_1 (Continuous)
- AT_2 (Discrete)

AT_1

Cutpoins	2.5	3.9	4.7	5.0	6.2
1.4	1	2	3	4	5
2.5	-	7	8	9	10
3.9	-	-	13	14	15
4.7	-	-	-	19	20
5.0	-	-	-	-	25

AT_2

Discrete Values					Natural Coding
white	*red*	*green*	*blue*	*black*	
0	0	0	0	0	0
0	0	0	0	1	1
0	0	0	1	0	2
...	...	...	...	...	...
0	1	0	1	1	11
...	...	...	...	...	...
1	1	1	1	1	31

Rule: If $AT_1 \in [3.9, 5.0]$ and $AT_2 \in \{red, blue, black\}$ Then Class = 0

Hybrid Individual

3.9	5.0	0	1	0	1	1	0

$\underbrace{3.9 \quad 5.0}_{AT_1} \quad \underbrace{0 \quad 1 \quad 0 \quad 1 \quad 1}_{AT_2} \quad \underbrace{0}_{Class}$

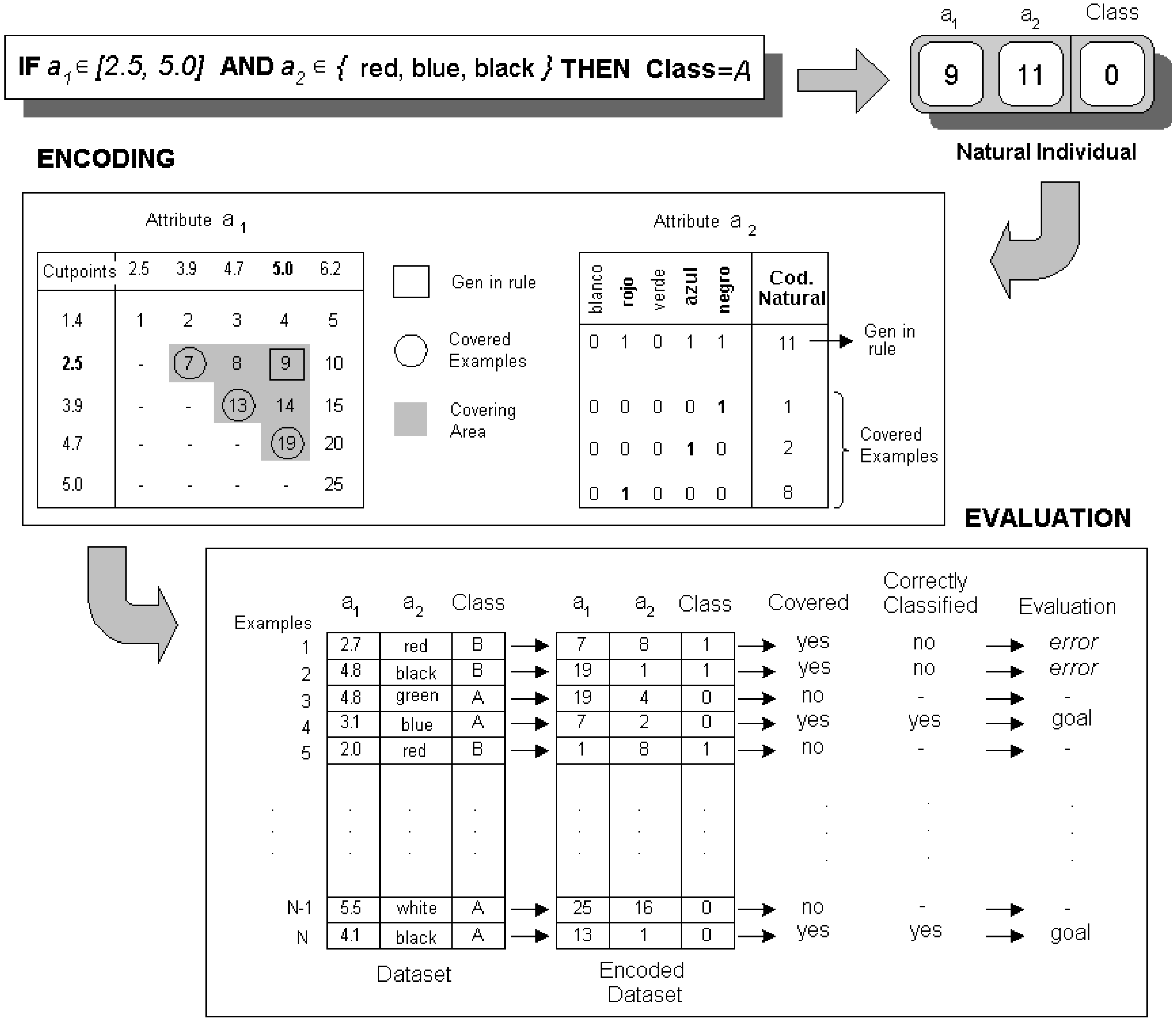
Natural Individual

14	11	0

$\underbrace{14}_{AT_1} \quad \underbrace{11}_{AT_2} \quad \underbrace{0}_{Class}$

Hider*: Natural Coding IV

Evaluation: The encoding of examples is carried out to speed up the evaluation process, since to decode all of the rules in each evaluation implies higher computational cost.



Hider*: Natural Coding V

¿Does Natural Coding imply greater computational cost than Hybrid Coding?

- The arrays are not stored in memory:
  - For continuous attributes only the set of cutpoints are stored.
  - For discrete attributes only the set of values are stored.
- All the operations are defined as simple algebraic expressions:
  - Encode
  - Decode
  - Genetic operators
- The encoding of examples is carried out to speed up the evaluation process

(See [Aguilar-Giraldez-Riquelme-07])

Natural Encoding does not add computational cost

what's more, it guarantees the same (*or better*) performance (error rate and number of rules) with less computational cost

Experiments

- Two kind of experiments [Aguilar-Giraldez-Riquelme-07]:
  1. Firstly, the Hybrid Coding (Hider) is compared to the Natural Coding (Hider*).
  2. Secondly, the results of the natural coding, C4.5 and C4.5Rules are analyzed, regarding the error rate and the number of rules, by using two sets of datasets:
     - 16 datasets with standard size (UCI).
     - 5 large datasets, with several thousands of examples and the number of attributes ranging from 27 to 1558.
- 10-fold stratified cross-validation

Experiments

Comparing Hybrid and Natural Coding

- HIDER and HIDER* have been run with a set of datasets from the UCI Repository.
- Both tools use the same EA
- Including the same fitness function
- Both were run with the same crossover and mutation parameters, but with a different number of individuals and generations:

Parameter		***HIDER (Hybrid)***	***HIDER* (Natural)***
Population Size		**100**	**70**
Number of Generations		**300**	**100**
Replication		20%	
Recombination		80%	
Individual Mutation Probability		0.5	
Gene Mutation Probability		1/‖attributes‖	
Discrete-value Mutation Probability		1/‖values‖	

Experiments

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

Comparing Hybrid and Natural Coding: Performance

	HIDER (hybrid)		HIDER* (natural)		Improvement	
Database	ER	NR	ER	NR	ϵ_{er}	ϵ_{nr}
breast-c	4.3	2.6	4.06	2	1.06	1.3
bupa	35.7	11.3	37.35	4.2	0.96	2.69
cleve	20.5	7.9	25.33	5.9	0.81	1.34
german	29.1	13.3	27.4	8	1.06	1.66
glass	29.4	19	35.24	11.7	0.83	1.62
heart	22.3	9.2	21.85	4.3	1.02	2.14
hepatiti	19.4	4.5	16.67	3.7	1.16	1.22
horse-co	17.6	6	20	11.1	0.88	0.54
Iris	3.3	4.8	3.33	3.2	0.99	1.5
Lenses	25	6.5	25	4.5	1	1.44
mushroom	0.8	3.1	1.18	3.5	0.68	0.89
pima	25.9	16.6	25.66	5.1	1.01	3.25
vehicle	30.6	36.2	33.81	19.7	0.91	1.84
vote	6.4	4	4.42	2.2	1.45	1.82
wine	3.9	3.3	8.82	5.6	0.44	0.59
zoo	8	7.2	4	7.9	2	0.91
Average	17,64	9,72	18,38	6,41	1,02	1,55

ER: Error rate
NR: Number of rules

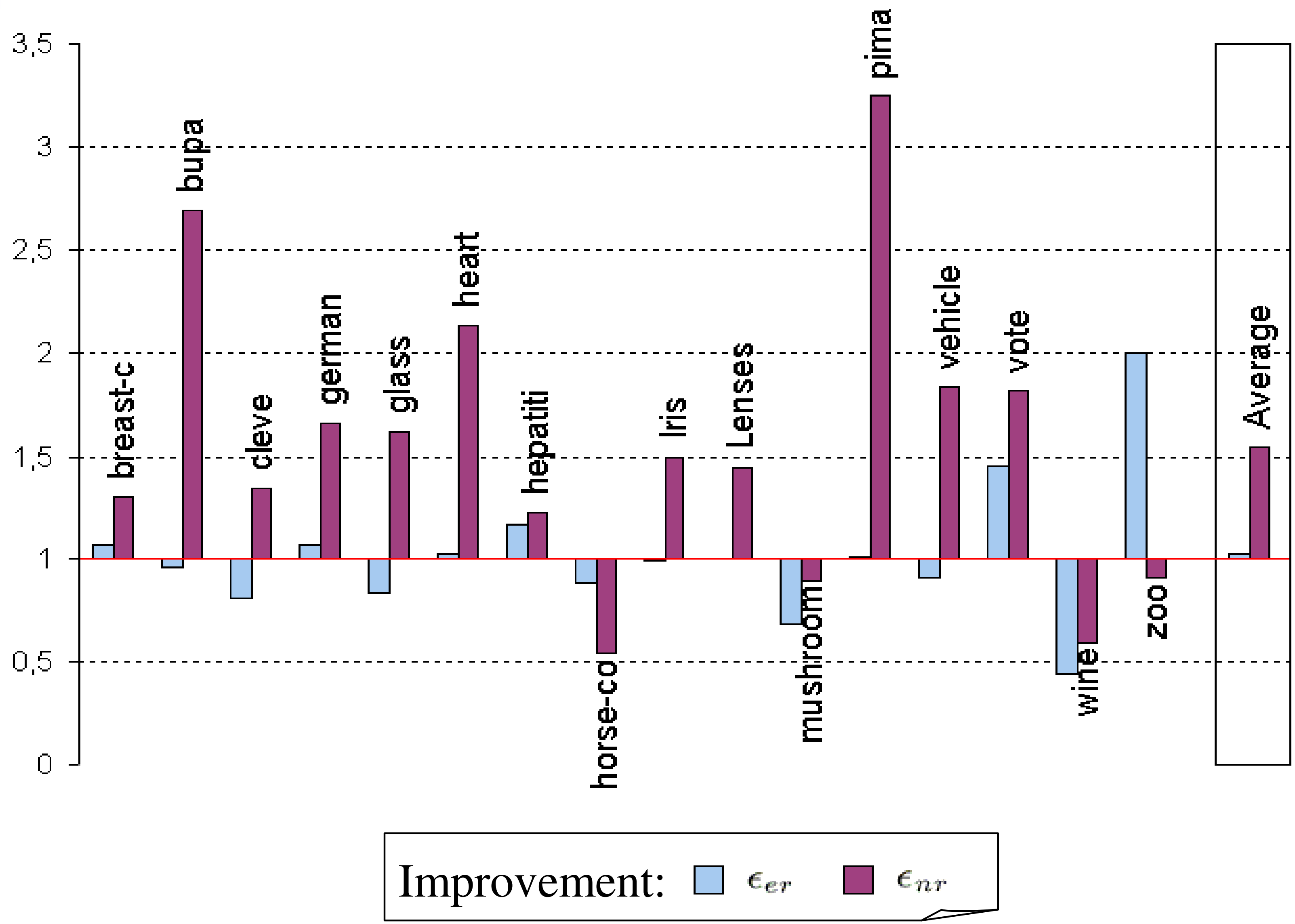
ϵ_{er} : Improvement ER
ϵ_{nr} : Improvement NR

Experiments

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

BioInformatics Research Group

Comparing Hybrid and Natural Coding: Performance



Experiments

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

Comparing Hybrid and Natural Coding: Length of Individuals

Database	NC	ND (NV)	Hybrid	Natural
breast-c	9	-	18	9
bupa	6	-	12	6
cleve	6	7 (19)	31	13
german	7	13 (54)	68	20
glass	9	-	18	9
heart	13	-	26	13
hepatiti	6	13 (26)	38	19
horse-co	7	15 (55)	69	22
Iris	4	-	8	4
Lenses	-	4 (9)	9	4
Mushroom	-	22 (117)	117	22
Pma	8	-	16	8
tic-tac-	9	-	18	9
vehicle	18	-	36	18
vote	-	16 (48)	48	16
wine	13	-	26	13
zoo	-	16 (36)	36	16
Average			34,94	13

Hybrid = 2 x NC + NV

Natural = NC + ND

→ Average Reduction greater than 63%

NC: Number of Continuous Attributes
ND: Number of Discrete Attributes (NV: Number of values)

Experiments

Comparing Hybrid and Natural Coding: Runtime

Dataset	HIDER (hybrid)	HIDER* (natural)	Percent
BC	133	39	29%
BU	160	26	16%
CL	222	73	33%
GE	1832	495	27%
GL	258	71	28%
HD	223	43	19%
HE	104	36	35%
HC	342	275	80%
IR	21	6	29%
LE	5	1	20%
MU	3778	2203	58%
PI	736	95	13%
VE	3862	942	24%
VO	196	43	22%
WI	52	43	83%
ZO	80	40	50%
Average			35.3%

HIDER* was faster than HIDER in all of experiments.

On average, HIDER*'s runtime was only 35% of the time taken by HIDER, about three times faster.

Experiments

HIDER* *vs.C4.5 and C4.5 Rules*

DS	HIDER* ER± σ	HIDER* NR± σ	C4.5 ER± σ	C4.5 NR± σ	HIDER* vs. C4.5 ER	HIDER* vs. C4.5 NR	C4.5Rules ER± σ	C4.5Rules NR± σ	HIDER* vs. C4.5Rules ER	HIDER* vs. C4.5Rules NR
BC	4.1± 2.0	2.0± 0.0	6.3± 3.0	22.9± 3.0	+	+•	4.9± 2.4	9.6± 1.1	+	+•
BU	37.3±11.4	4.2± 0.8	33.7± 9.3	29.7± 5.1	−	+•	32.0± 6.2	11.9± 2.2	−	+•
CL	25.3±10.5	5.9± 0.9	23.5± 7.0	38.3± 5.8	−	+•	25.9±14.7	12.2± 4.6	+	+•
GE	27.4± 3.9	8.0± 1.4	32.9± 4.3	204.2± 8.5	+•	+•	28.8± 3.1	6.2± 2.2	+	−
GL	35.2± 7.8	11.7± 1.6	30.8±11.4	25.8± 2.0	−	+•	18.5± 5.9	15.0± 2.8	−•	+•
HD	21.9± 8.8	4.3± 0.5	25.5± 5.1	33.5± 4.5	+	+•	20.7± 7.0	11.5± 2.0	−	+•
HE	16.7±11.0	3.7± 0.7	19.4± 7.0	15.5± 1.7	+	+•	16.9± 6.1	6.3± 3.6	+	+•
HC	20.0± 7.7	11.1± 1.9	19.0± 7.7	44.4± 3.8	−	+•	17.5± 8.2	5.0± 1.9	−	−•
IR	3.3± 4.7	3.2± 0.4	5.3± 6.9	5.7± 0.5	+	+•	4.7± 7.1	5.0± 0.0	+	+•
LE	25.0±26.4	4.5± 0.9	30.0±21.9	5.2± 0.9	+	+	16.7±22.2	4.1± 0.3	−	−
MU	1.2± 0.6	3.5± 0.5	0.0± 0.0	17.6± 1.0	−•	+•	0.7± 2.3	17.9± 1.9	−	+•
PI	25.7± 3.4	5.1± 0.7	26.1± 5.4	24.4± 8.1	+	+•	29.7± 3.8	8.3± 3.1	+	+•
VE	33.8± 7.4	19.7± 3.3	27.5± 3.6	74.8±10.0	−•	+•	57.6±14.7	4.1± 4.1	+•	−•
VO	4.4± 2.9	2.2± 0.4	6.2± 3.1	15.9± 3.0	+	+•	5.3± 3.7	7.5± 0.7	+	+•
WI	8.8± 4.2	5.6± 0.8	6.7± 7.8	6.5± 0.9	−	+•	6.7± 7.8	5.6± 0.7	−	−
ZO	4.0± 5.2	7.9± 0.9	7.0±10.6	10.9± 1.8	+	+•	29.8±20.7	6.3± 2.0	+•	−•
AVG.	18.4± 7.4	6.4± 1.0	18.7± 7.1	36.9± 3.8			19.7± 8.5	8.5± 2.1		

Symbols + and - mean that Hider* is better or worse, respectively.
If the symbol • appears next, then + or - are statistically significant.

Experiments

HIDER* *vs.C4.5 and C4.5 Rules*

Both algorithms tie regarding the error rate, although C4.5 is significantly better in 2 out of 7 datasets and Hider* in only one.

As regards the number of rules, HIDER* outperforms C4.5 for all the datasets, where the improvement is significant in 15 out of 16 datasets.

			C4.5		HIDER* vs. C4.5		C4.5Rules		HIDER* vs. C4.5Rules	
			σ	NR± σ	ER	NR	ER± σ	NR± σ	ER	NR
			3.0	22.9± 3.0	+	+•	4.9± 2.4	9.6± 1.1	+	+•
			9.3	29.7± 5.1	−	+•	32.0± 6.2	11.9± 2.2	−	+•
			7.0	38.3± 5.8	−	+•	25.9± 14.7	12.2± 4.6	+	+•
			4.3	204.2± 8.5	+•	+•	28.8± 3.1	6.2± 2.2	+	−
			11.4	25.8± 2.0	−	+•	18.5± 5.9	15.0± 2.8	−•	+•
HD	21.9± 8.8	4.3± 0.5	25.5± 5.1	33.5± 4.5	+	+•	20.7± 7.0	11.5± 2.0	−	+•
			7.0	15.5± 1.7	+	+•	16.9± 6.1	6.3± 3.6	+	+•
			7.7	44.4± 3.8	−	+•	17.5± 8.2	5.0± 1.9	−	−•
			6.9	5.7± 0.5	+	+•	4.7± 7.1	5.0± 0.0	+	+•
			21.9	5.2± 0.9	+	+	16.7± 22.2	4.1± 0.3	−	−
			0.0	17.6± 1.0	−•	+•	0.7± 2.3	17.9± 1.9	−	+•
			5.4	24.4± 8.1	+	+•	29.7± 3.8	8.3± 3.1	+	+•
			3.6	74.8± 10.0	−•	+•	57.6± 14.7	4.1± 4.1	+•	−•
			3.1	15.9± 3.0	+	+•	5.3± 3.7	7.5± 0.7	+	+•
			7.8	6.5± 0.9	−	+•	6.7± 7.8	5.6± 0.7	−	−
			10.6	10.9± 1.8	+	+•	29.8± 20.7	6.3± 2.0	+•	−•
AVG.	18.4± 7.4	6.4± 1.0	18.7± 7.1	36.9± 3.8			19.7± 8.5	8.5± 2.1		

Symbols + and - mean that Hider* is better or worse, respectively.
If the symbol • appears next, then + or - are statistically significant.

Experiments

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

HIDER* *vs.C4.5 and C4.5 Rules*

	HIDER*						C4.5Rules		HIDER* vs. C4.5Rules	
DS	ER± σ	NR± σ	I				± σ	NR± σ	ER	NR
BC	4.1± 2.0	2.0± 0.0	6				± 2.4	9.6± 1.1	+	+•
BU	37.3±11.4	4.2± 0.8	33				± 6.2	11.9± 2.2	−	+•
CL	25.3±10.5	5.9± 0.9	23				±14.7	12.2± 4.6	+	+•
GE	27.4± 3.9	8.0± 1.4	32				± 3.1	6.2± 2.2	+	−
GL	35.2± 7.8	11.7± 1.6	30				± 5.9	15.0± 2.8	−•	+•
HD	21.9± 8.8	4.3± 0.5	25.5± 5.1	33.5± 4.5	+	+•	20.7± 7.0	11.5± 2.0	−	+•
HE	16.7±11.0	3.7± 0.7	19.				± 6.1	6.3± 3.6	+	+•
HC	20.0± 7.7	11.1± 1.9	19.				± 8.2	5.0± 1.9	−	−•
IR	3.3± 4.7	3.2± 0.4	5.				± 7.1	5.0± 0.0	+	+•
LE	25.0±26.4	4.5± 0.9	30.				±22.2	4.1± 0.3	−	−
MU	1.2± 0.6	3.5± 0.5	0.				± 2.3	17.9± 1.9	−	+•
PI	25.7± 3.4	5.1± 0.7	26.				± 3.8	8.3± 3.1	+	+•
VE	33.8± 7.4	19.7± 3.3	27.				±14.7	4.1± 4.1	+•	−•
VO	4.4± 2.9	2.2± 0.4	6.				± 3.7	7.5± 0.7	+	+•
WI	8.8± 4.2	5.6± 0.8	6.				± 7.8	5.6± 0.7	−	−
ZO	4.0± 5.2	7.9± 0.9	7.				±20.7	6.3± 2.0	+•	−•
AVG.	18.4± 7.4	6.4± 1.0	18.7± 7.1	36.9± 3.8			19.7± 8.5	8.5± 2.1		

There is also a tie for the error rates, although in two cases HIDER* outperforms C4.5 Rules significantly

HIDER* improves C4.5 Rules in 11 datasets, being significant in 10 cases
C4.5 Rules provides smaller number of rules in 6 cases, of which only 3 are significant.

Symbols + and - mean that Hider* is better or worse, respectively.
If the symbol • appears next, then + or - are statistically significant.

Experiments

Experiments with High-Dimensional Datasets

Finally, we present a set of experiments in order to show the performance of our approach with five high-dimensional datasets:

ID	Database	NE	NA	Classes
AD	ADS	3279	1558	2
KR	KR-VS-KP	3196	36	2
MK	MUSK	6598	166	2
SI	SICK	3772	27	2
SP	SPLICE	3190	60	3

NE: Number of Examples
NA: Number of Attributes

• HIDER*, C4.5, and C4.5 Rules were run with these datasets.

• The results obtained by C4.5 are not comparable with those produced by the other classifiers (C4.5 produces models with a very low error rate, although with a complexity extremely greater than HIDER* or C4.5 Rules).

• For instance, for the Musk dataset, C4.5 obtains an average of 118 rules for a error about 4%.

Experiments

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

BioInformatics Research Group

Experiments with High-Dimensional Datasets

	HIDER*		C4.5Rules		HIDER* vs. C4.5Rules	
DS	ER± σ	NR± σ	ER± σ	NR± σ	ER	NR
AD	5.2± 1.7	7.6± 4.2	3.1± 1.3	12.0± 3.5	−•	+•
KR	5.6± 1.4	3.3± 0.4	9.4± 1.8	8.4± 0.9	+•	+•
MK	11.4± 2.2	10.7± 1.4	11.7± 2.7	10.2± 9.7	+	−
SI	1.8± 0.6	3.0± 0.0	4.4± 1.7	3.7± 0.7	+•	+•
SP	17.9± 2.7	7.2± 1.1	18.8± 13.3	23.2± 23.5	+	+•
Average	8.4± 1.7	6.4± 1.4	9.5± 4.2	11.5± 7.7		

HIDER* improves C4.5 Rules in 4 datasets, being significant in 2 cases.

HIDER* improves C4.5 Rules in 4 datasets, being significant in all of them.

Conclusions I

- In this work, Natural Coding for EA-based decision rules generation is described and tested.

- *Main features of Natural Coding*:

  - NC transforms the attributes domain (continuous and discrete) in a finite set of natural numbers.

  - The genetic operators (mutation and crossover) are defined as algebraic expressions.

  - EA works from the beginning to the end with natural numbers.

  - All the examples from the database are encoded into the search space, making the evaluation process very fast.

  - NC reduces the size of search space → The algorithm converges more quickly.

  - NC encodes each attribute with only one gene → The length of individual is reduced.

Conclusions II

- Empirical results:
  - The quality of this coding has been tested by applying the same evolutionary learning tool with natural (HIDER*) and hybrid (HIDER) coding, also improving the computational cost.
    - HC needed **300** generations and a population with **100** individuals in order to obtain a suitable model.
    - NC obtained better results with a number of generations as little as **100** and only **70** individuals per population.
  - HIDER* has been compared with C4.5 and C4.5 Rules and the experimental results show an excellent performance, mainly with respect to the number of rules, maintaining the quality of the acquired knowledge model.

Future and Related Works

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA
BioInformatics Research Group

"Standby" work:

- **Feature influence** [Giraldez-Aguilar-05]: an approach that deals with the feature selection problem. While the traditional feature selection is based on the attribute relevance, this work present a new concept, called *feature influence*, which includes two main aspects:
  - First, the selection is done during the evolutionary learning process, i.e., it is a *dynamic* approach.
  - Second, the selection is *local*, i.e., the algorithm selects the best features from the best space region to learn at a given time of the exploration process.

Future work:

- Learning from Unbalanced Data → Rule set with one class label

Related work:

- Genetic-Based Machine Learning Systems Are Competitive for Pattern Recognition [Orriols-Casillas-Bernado-08]

References

[Giraldez-Aguilar-Riquelme-02]
R. Giráldez, J.S. Aguilar-Ruiz, and J.C. Riquelme, “Discretization oriented to decision rules generation”, in *Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies (KES’02). IOSPress*, 2002, pp. 275–279.

[Aguilar-Riquelme-Toro-03]
J. S. Aguilar-Ruiz, J.C. Riquelme, and M. Toro, “Evolutionary learning of hierarchical decision rules,” *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 33, no. 2, pp. 324–331, 2003.

[Aguilar-Bacardit-Divina-04]
J. S. Aguilar-Ruiz, J. Bacardit, and F. Divina, “Experimental evaluation of discretization schemes for rule induction,” in *Genetic and Evolutionary Computation – GECCO-2004*, Seattle, US, June 2004, LNCS, pp. 828–839, Springer-Verlag.

[Giraldez-Aguilar-Riquelme-05]
Raul Giraldez, Jesus S. Aguilar-Ruiz, Jose C. Riquelme, “Knowledge-based Fast Evaluation for Evolutionary Learning”, *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol 35, Nº 2, pp. 254-261.

[Giraldez-05]
Raul Giraldez, “Improving the Performance of Evolutionary Algorithms for Decision Rule Learning”, AI Communications, IOS Press. Vol 18, Nº 1, pp. 63-65

[Giraldez-Aguilar-05]
Raul Giraldez, Jesus S. Aguilar-Ruiz, “Feature Influence for Evolutionary Learning”, in Proc. of Genetic and Evolutionary Computation Conference (GECCO’05), ACM Press, pp. 1139-1145.

[Giraldez-Aguilar-Riquelme-07]
Jesús S. Aguilar-Ruiz, Raul Giraldez, José C. Riquelme, “Natural Encoding for Evolutionary Supervised Learning”, *IEEE Transactions on Evolutionary Computation*, Vol 11, Nº 4, pp. 466-479

[Orriols-Casillas-Bernado-08]
Albert Orriols-Puig, Jorge Casillas, Ester Bernad´o-Mansilla, “”, Evolutionary Intelligence (submitted).

Workshop

KOWLEDGE EXTRACTION BASED ON EVOLUTIONARY ALGORITHMS

HIDER: Method and Natural Coding

Raúl Giráldez

School of Engineering - Division of Computer Science

Pablo de Olavide University of Seville

May 15, 2005 – Granada, Spain