I Workshop on Knowledge Extraction based on Evolutionary Learning

Incorporating Fuzzy Rules in LCS: Fuzzy-XCS and Fuzzy-UCS

Jorge Casillas
Dept. Computer Science and Artificial Intelligence
University of Granada
http://decsai.ugr.es/~casillas

Granada, May 15, 2008

Outline

1. Fuzzy-XCS (for reinforcement learning)
    1. Introduction
    2. Proposal description
    3. Experiments
    4. Conclusion

2. Fuzzy-UCS (for supervised learning)
    1. Introduction
    2. Proposal description
    3. Experiments
    4. Conclusion

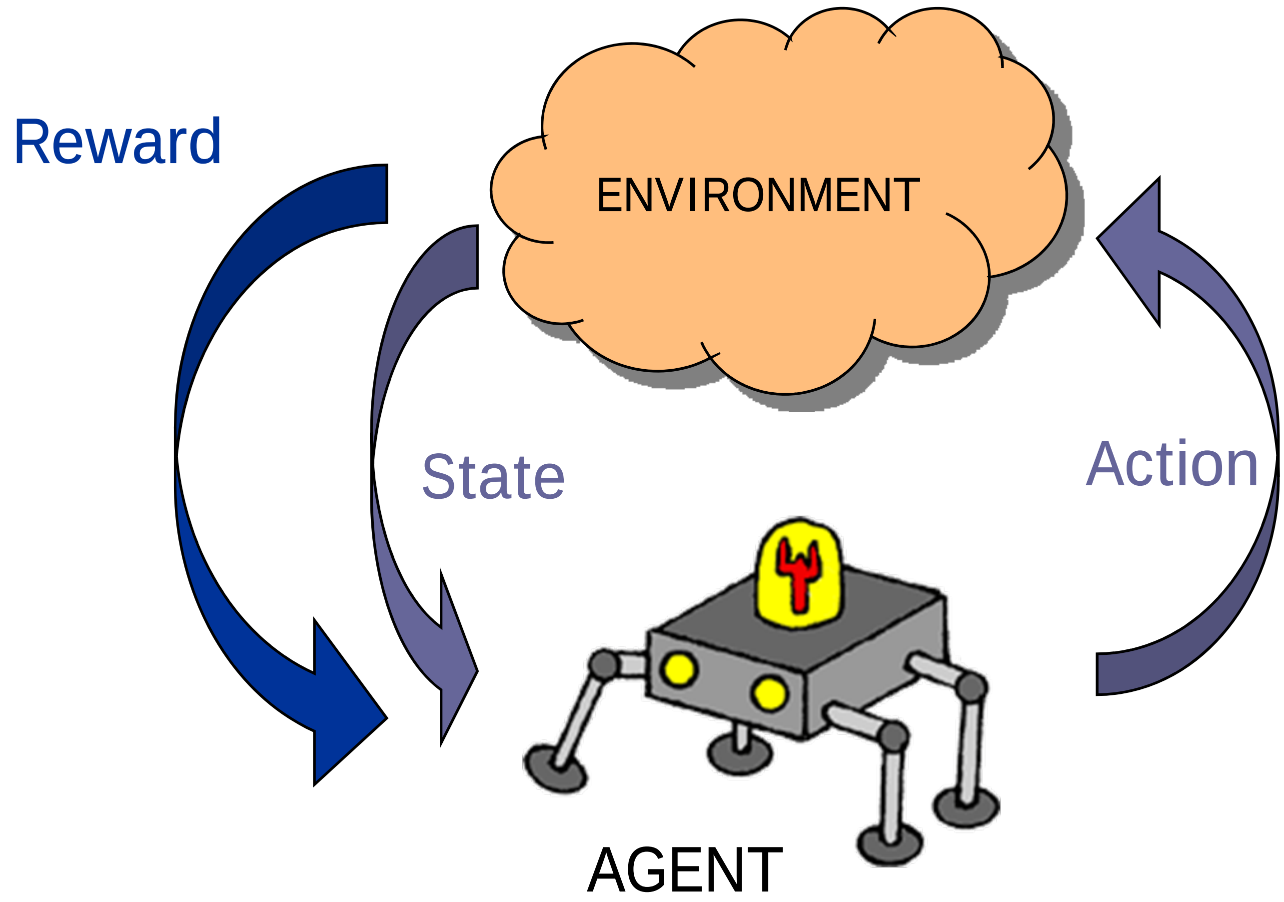
Fuzzy-XCS: A Michigan-style Genetic Fuzzy System for Reinforcement Learning

Jorge Casillas
Dept. Computer Science and Artificial Intelligence
University of Granada
http://decsai.ugr.es/~casillas

In collaboration with
Brian Carse and Larry Bull
University of the West of England, Bristol, UK

Introduction

- Reinforcement learning with fuzzy rule-based systems



Introduction

- Learning Classifier Systems (LCS) [John H. Holland & Judith S. Reitman, 1978]
  - Machine learning paradigm that uses genetic algorithms and other low-level induction operations to keep a population of classifiers wich allow an *Animat* to optimally respond within an environment
  - Classifier: contains a rule composed by the condition (state) and action parts, and some parameters for the learning process
  - Animat: Animal simulated within a constrained environment, with a limited set of detectors and effectors
- Michigan-style classifier systems are the most commonly used (1 individual = 1 rule) for reinforcement learning

Introduction

- Why it is interesting Michigan-style LCS?
  - Designed to do an “on-line” learning
  - The system is build at the same time that stimulus are received
  - This allow the system to adapt itself to variable environment

- Thanks to these characteristics, Michigan-style LCS is ideal for:
  - Robot control
  - Behavior learning
  - Self-governing systems
  - Data mining
  - ...

Introduction

- Traditionally, Michigan-style LCS was based on strength
  - Each classifier accumulates strength during its interaction with the environment through rewards and punishments
  - Strength represents an estimation of the reward that the classifier will receive from the environment
  - Informally speaking, strength is a kind of average of the received reward

- This parameter is used for two objectives:
  - Resolve conflicts among classifiers that are activated simultaneously during the learning process
  - Assess the fitness degree of each classifier for the evolutionary algorithm

Introduction

- In 1995, S.W. Wilson proposes an alternative (XCS) based on accuracy to compute the fitness
  - Those classifiers with a high capability to predict the received payoff are rewarded
  - That is, only those classifiers that obtain a mean reward more or less constant survive
  - Informally speaking, accuracy is a kind of standard deviation of the received reward
  - This shows some advantages: overgeneralization is avoided, classifiers with an optimal generalization are obtained, and the complete cover map is learnt

- Currently, XCS-based LCS is the subject of extensive ongoing research

Introduction

- Michigan-style genetic fuzzy system: it is a LCS composed by fuzzy rules

- There are not many proposals of fuzzy LCS. A list almost exhaustive is the following:
    1. [M. Valenzuela-Rendón, 1991]
    2. [A. Parodi, P. Bonelli, 1993]
    3. [T. Furuhashi, K. Nakaoka, Y. Uchikawa, 1994]
    4. [K. Nakaoka, T. Furuhashi, Y. Uchikawa, 1994]
    5. [A. Bonarini, 1996]
    6. [M. Valenzuela-Rendón, 1998]
    7. [J.R. Velasco, 1998]
    8. [H. Ishibuchi, T. Nakashima, T. Murata, 1999]
    9. [A. Bonarini, V. Trianni, 2001]
    10. [D. Gu, H. Hu, 2004]
    11. [M.C. Su, et al., 2005]
    12. [C.-F. Juang, 2005]

All except [10] are based on *strengh*. In [10], the output is discrete and generality is not considered

Introduction

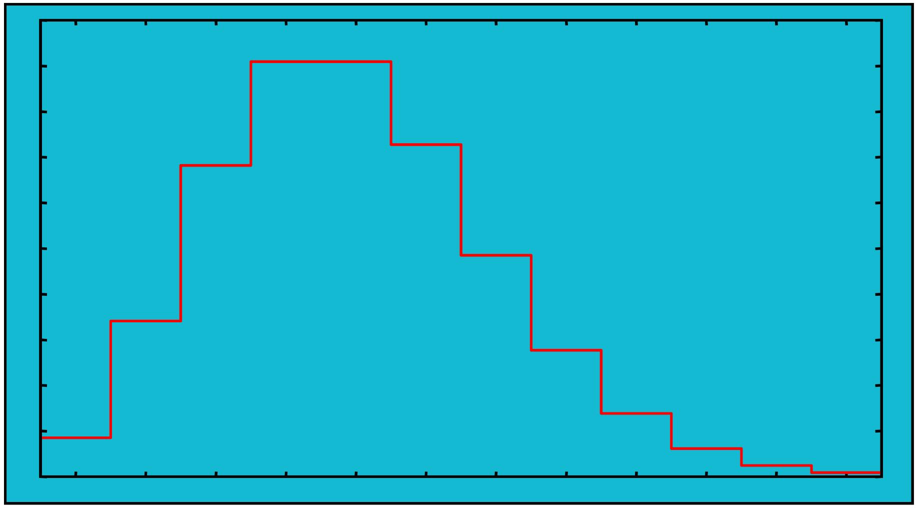
- An accuracy-based Fuzzy Classifier System has the following difficulties:
  - Since several rules fire in parallel and the output is due to a combination of them, credit assignment is much more difficult
  - The payoff a fuzzy rule receives depends on the input vector, an active fuzzy rule will receive different payoffs for different inputs
  - Measuring the accuracy of a rule's predicted payoff is difficult since a fuzzy rule will fire with many different other fuzzy rules at different time-steps, giving very different payoffs

Proposal

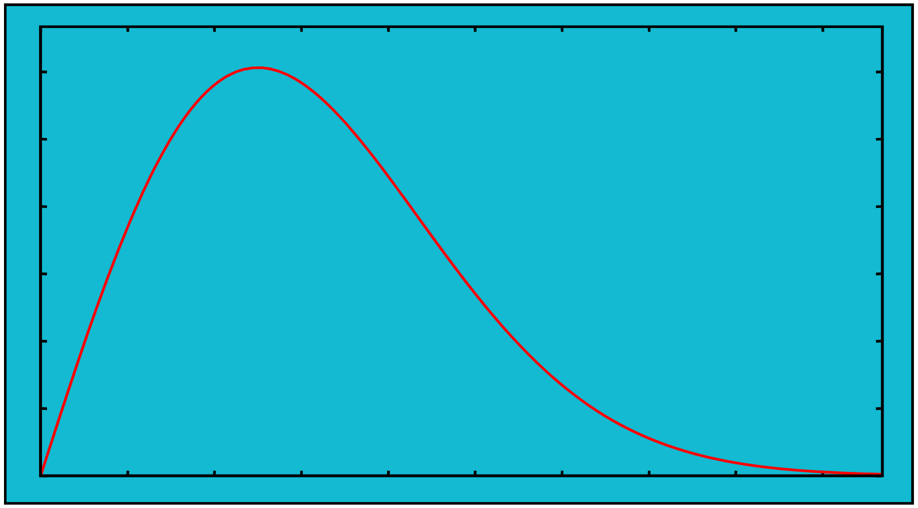
- Fuzzy-XCS: an accuracy-based fuzzy classifier system
- This kind of system would have some important advantages:
  - Compared with LCS:
    - Fuzzy rules are a natural framework for real-valued input and output environments. This is very usual in robotics, for instance
  - Compared with the rest of fuzzy LCS:
    - The accuracy-based scheme allow to obtain rules with optimal generalization. This generalization generates compact rule sets, properly face the course of dimensionality, quick inference, better interpretability, ...
    - Overgeneralized classifiers are avoided
    - It can maintain both consistently correct and consistently incorrect classifiers which allows learning of a complete covering map

Proposal
Continuous output

Discrete output



Continuous output

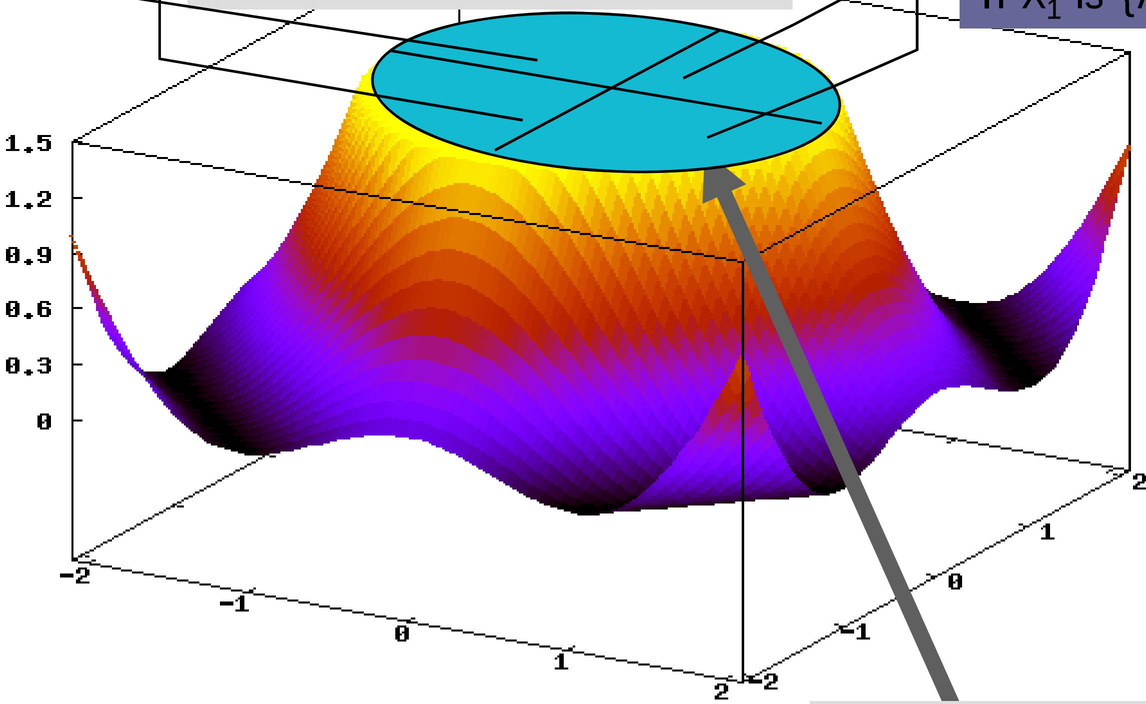


Proposal
Optimal generality

If X_1 is A and X_2 is D $\Rightarrow$ Y is F
If X_1 is A and X_2 is E $\Rightarrow$ Y is F
If X_1 is B and X_2 is D $\Rightarrow$ Y is F
If X_1 is B and X_2 is E $\Rightarrow$ Y is F

Rule with optimal generality
If X_1 is {A,B} and X_2 is {D,E} $\Rightarrow$ Y is F

Over-general rule
If X_1 is {A,B,C} and X_2 is {D,E} $\Rightarrow$ Y is F

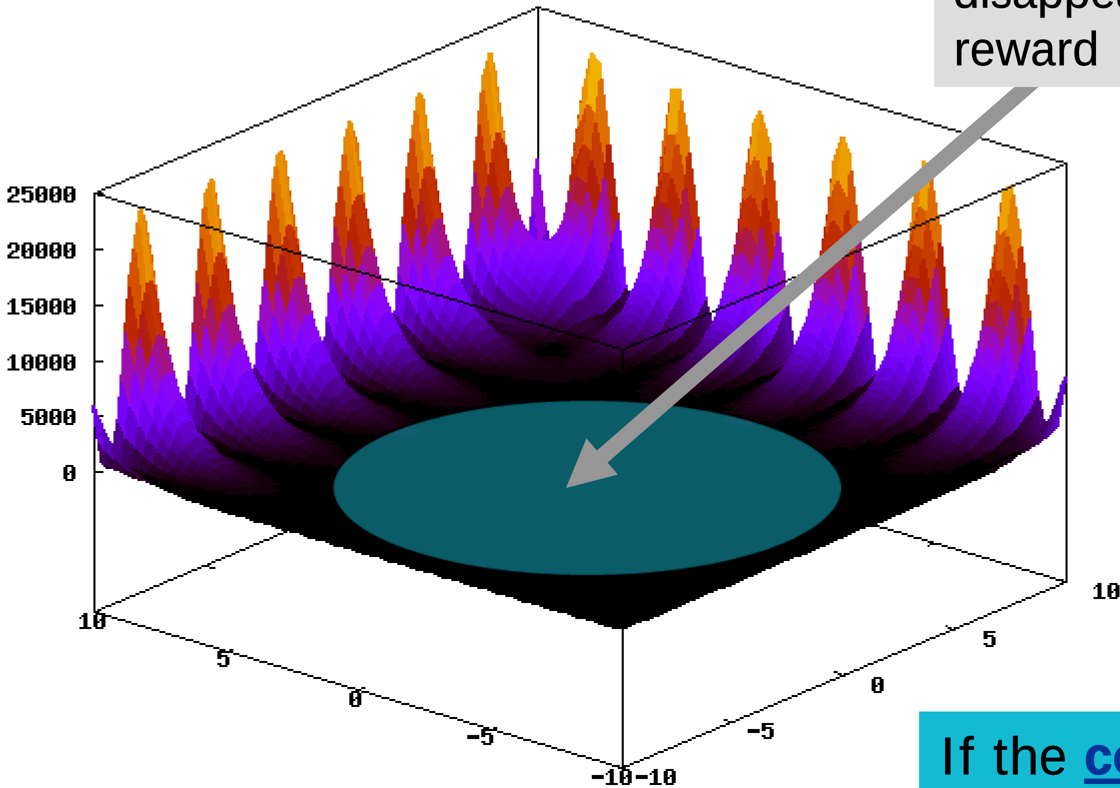


Graphical representation of the reward function

If there is a region with equal or similar reward, it is better to obtain a single rule that describes the complete area

Proposal
Complete covering map

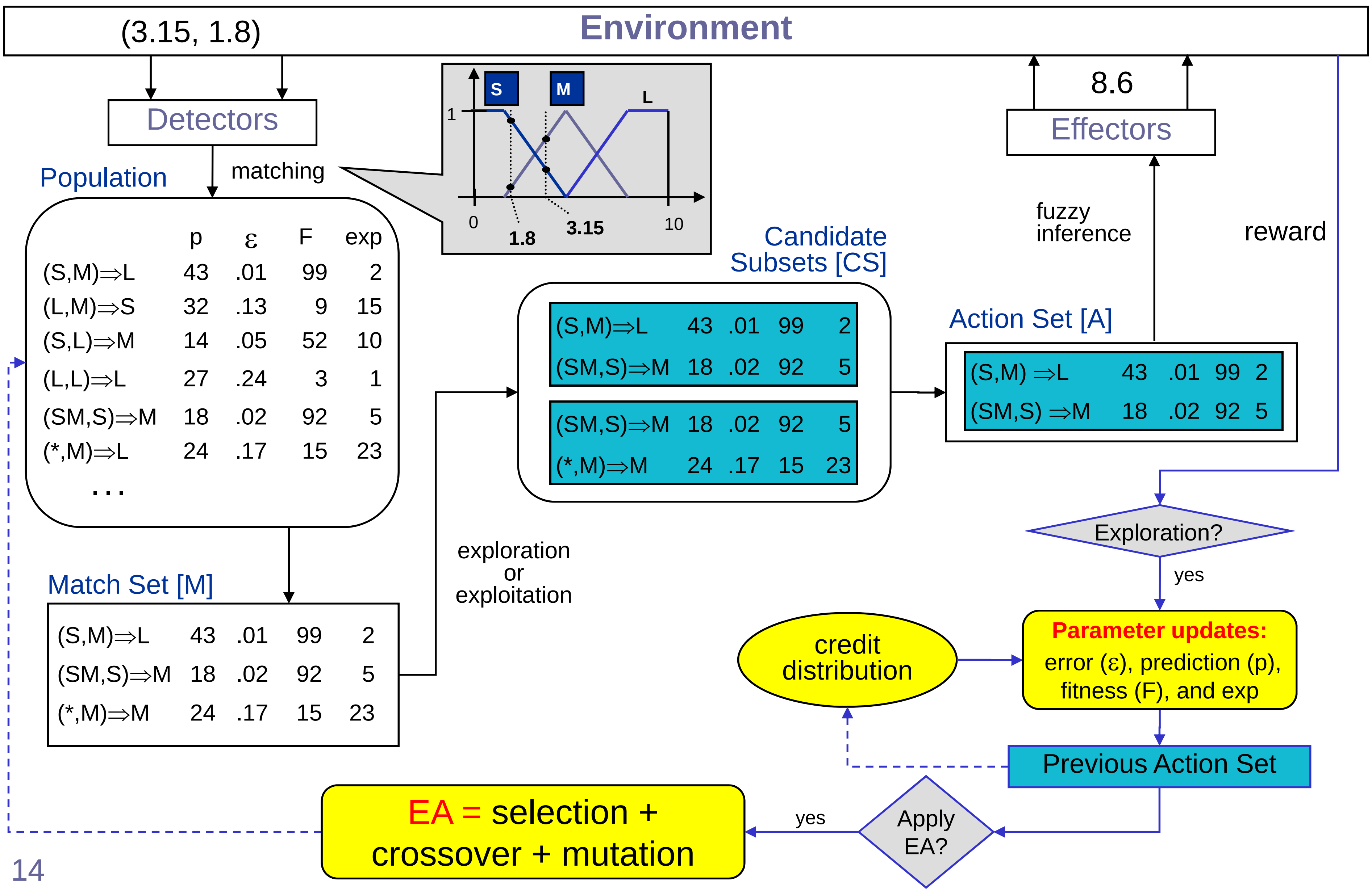
Regions with low reward tend to disappear in favor of regions with high reward



Graphical representation of the reward function

If the **complete covering map** is generated, the system also obtains rules in areas with low reward

Fuzzy-XCS



(3.15, 1.8)
Environment
Detectors
8.6
Effectors
matching
Population
S
M
L
1
0
1.8
3.15
10
p ε F exp
(S,M)⇒L 43 .01 99 2
(L,M)⇒S 32 .13 9 15
(S,L)⇒M 14 .05 52 10
(L,L)⇒L 27 .24 3 1
(SM,S)⇒M 18 .02 92 5
(*,M)⇒L 24 .17 15 23
. . .
Candidate Subsets [CS]
(S,M)⇒L 43 .01 99 2
(SM,S)⇒M 18 .02 92 5
(SM,S)⇒M 18 .02 92 5
(*,M)⇒M 24 .17 15 23
fuzzy inference
reward
Action Set [A]
(S,M) ⇒L 43 .01 99 2
(SM,S) ⇒M 18 .02 92 5
Exploration?
yes
Match Set [M]
(S,M)⇒L 43 .01 99 2
(SM,S)⇒M 18 .02 92 5
(*,M)⇒M 24 .17 15 23
exploration or exploitation
credit distribution
Parameter updates:
error (ε), prediction (p), fitness (F), and exp
Previous Action Set
EA = selection + crossover + mutation
yes
Apply EA?

Experiments Results in a Laboratory Problem

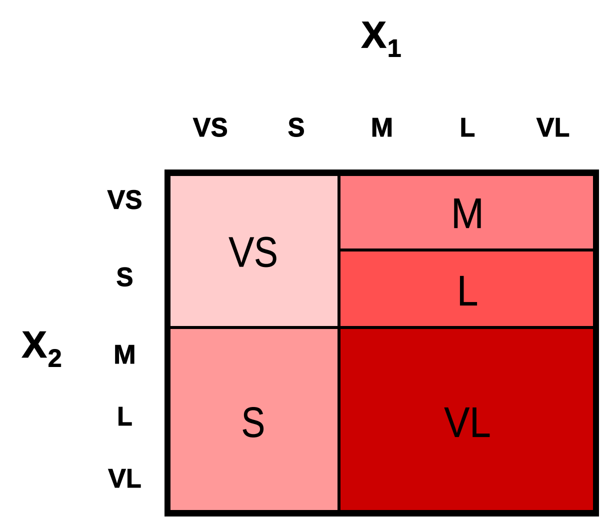
- First experiments on a laboratory problem
- We have developed a laboratory problem to play a similar role as the multiplexer problem for discrete-valued classifier systems
- The problem is build by generating a data set from a fuzzy system previously fixed
- The objective is to obtain the set of rules that best approximate the data with the highest degree of generalization, i.e., a rule base as accurate and compact as possible
- The reward depends inversely on the difference between the inferred and the desired output in a non-linear way

$$\boxed{f(x,y) = \frac{1 - |y - F(x)|}{1 + \dfrac{|y - F(x)|}{\rho}}, \qquad \rho = 0.3}$$

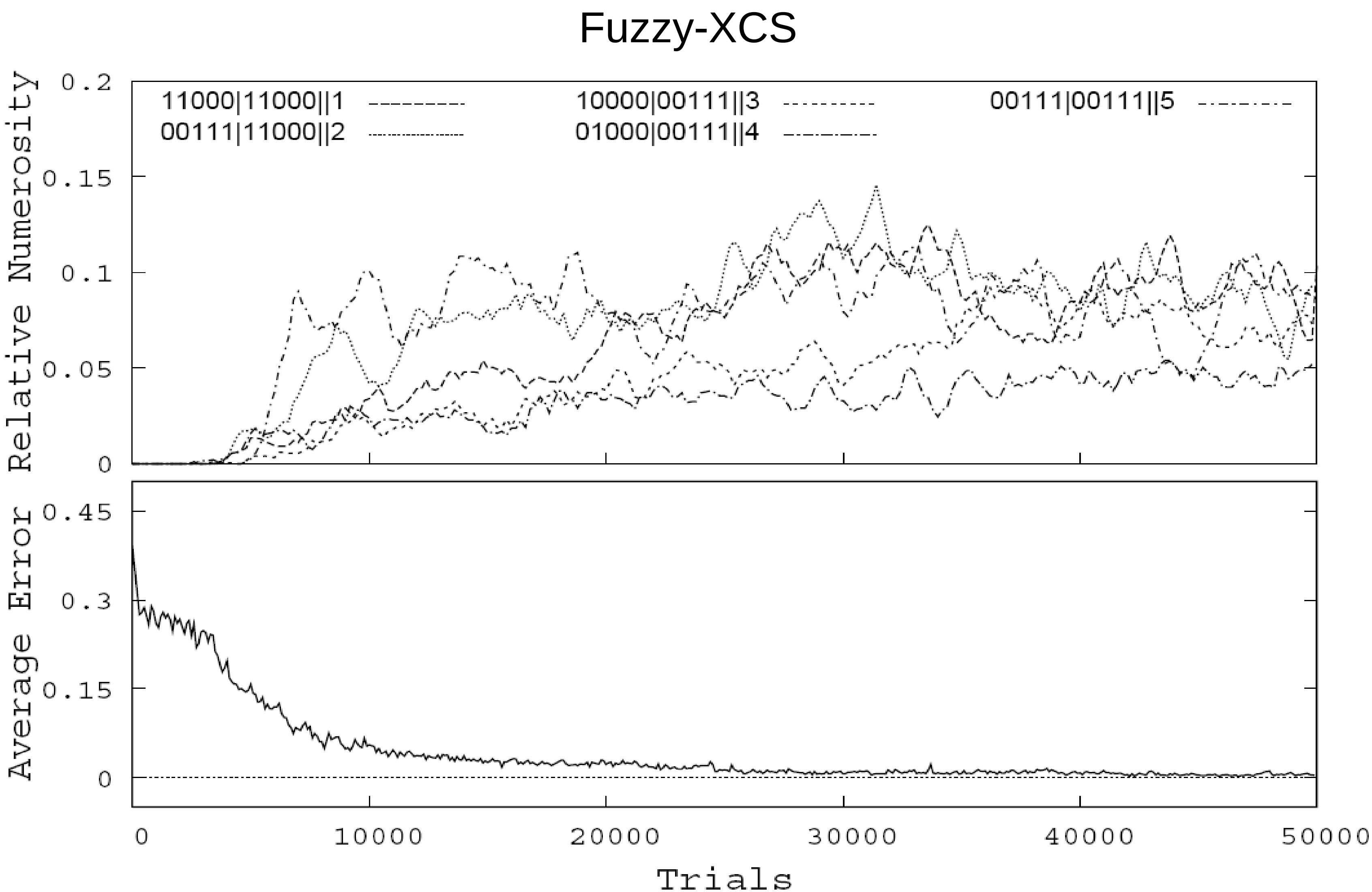
Experiments Results in a Laboratory Problem

- Problem:
  - 2 inputs and 1 output
  - 5 linguistic terms for each variable (triangular-shape fuzzy sets)
  - 5 fuzzy rules of different generality degree
  - 576 examples uniformly distributed in the input space (24 x 24)
  - The output value for each input is the result of the inference with the fixed fuzzy system

	X_1					X_2					Y				
	VS	S	M	L	VL	VS	S	M	L	VL	VS	S	M	L	VL
R_1	X	X				X	X				X				
R_2	X	X						X	X	X		X			
R_3			X	X	X	X							X		
R_4			X	X	X		X							X	
R_5			X	X	X			X	X	X					X

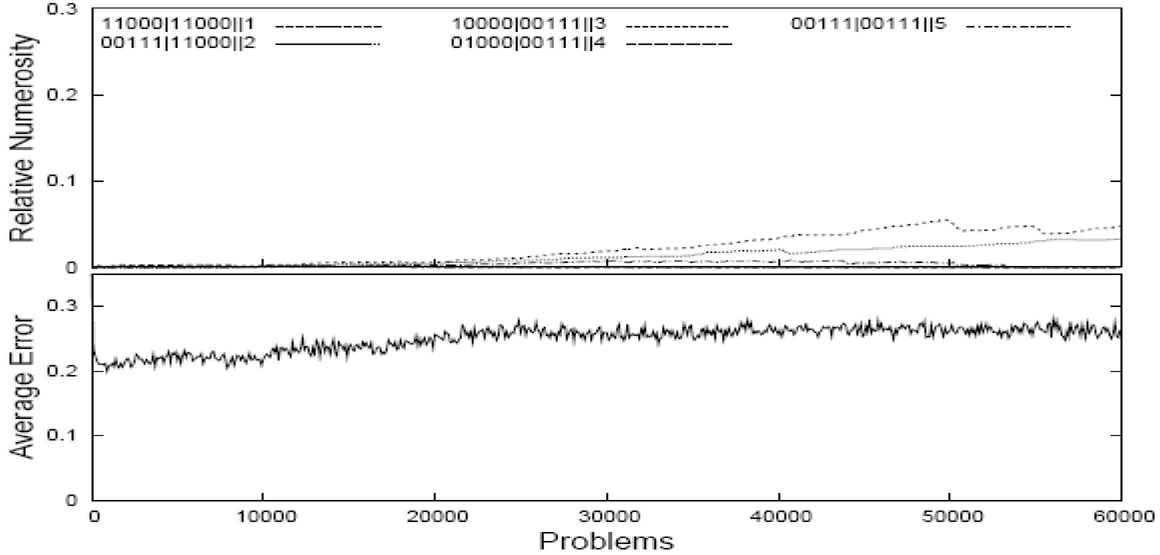


Experiments Results in a Laboratory Problem

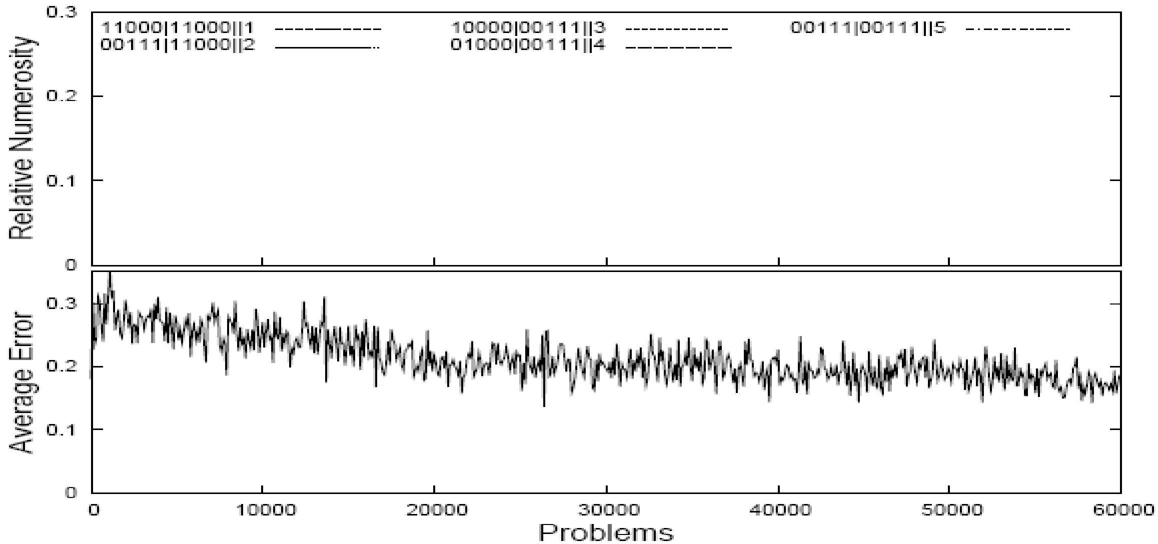


Experiments Results in a Laboratory Problem

Valenzuela-Rendón



Bonarini



Experiments Results in a Laboratory Problem

	Fuzzy-XCS	Valenzuela-Rendón	Bonarini	Pittsburgh GFS
R_1	0.7	0.0	0.0	0.1
R_2	0.9	0.0	0.0	0.2
R_3	0.8	0.0	0.0	0.0
R_4	0.7	0.0	0.0	0.2
R_5	1.0	0.0	0.0	0.1
no. suboptimal rules	0.1	0.0	4.8	7.6
no. non-suboptimal rules	1.3	1.0	20.2	2.0
MSE	0.001614	0.144631	0.052516	0.001892
no. analyzed examples	25,000	60,000	60,000	4,212,864

Experiments Results in the Mobile Robot Problem

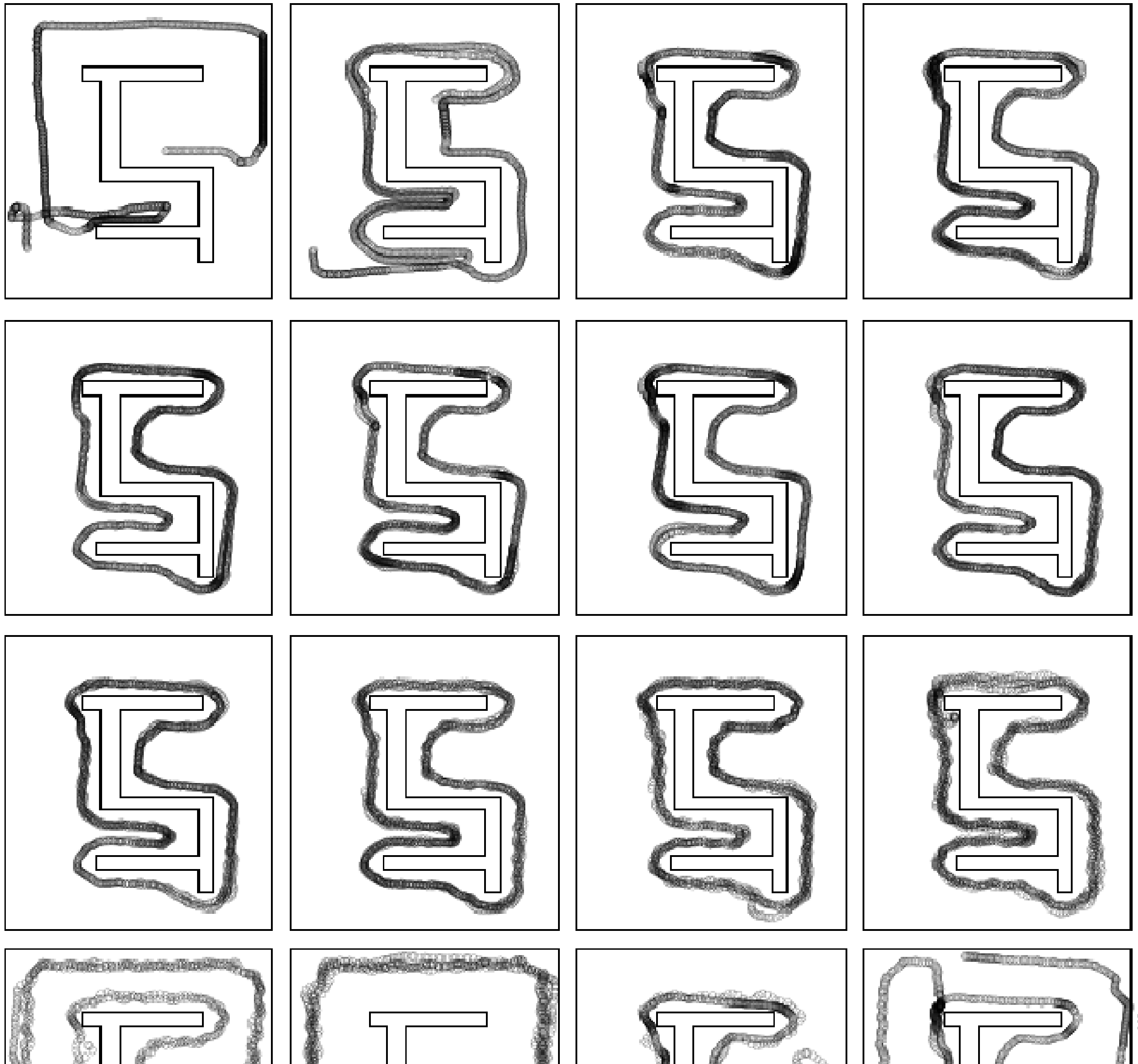
- Real-world problem
- On-line learning of the wall-following behavior for a mobile robot (Nomad 200 model)

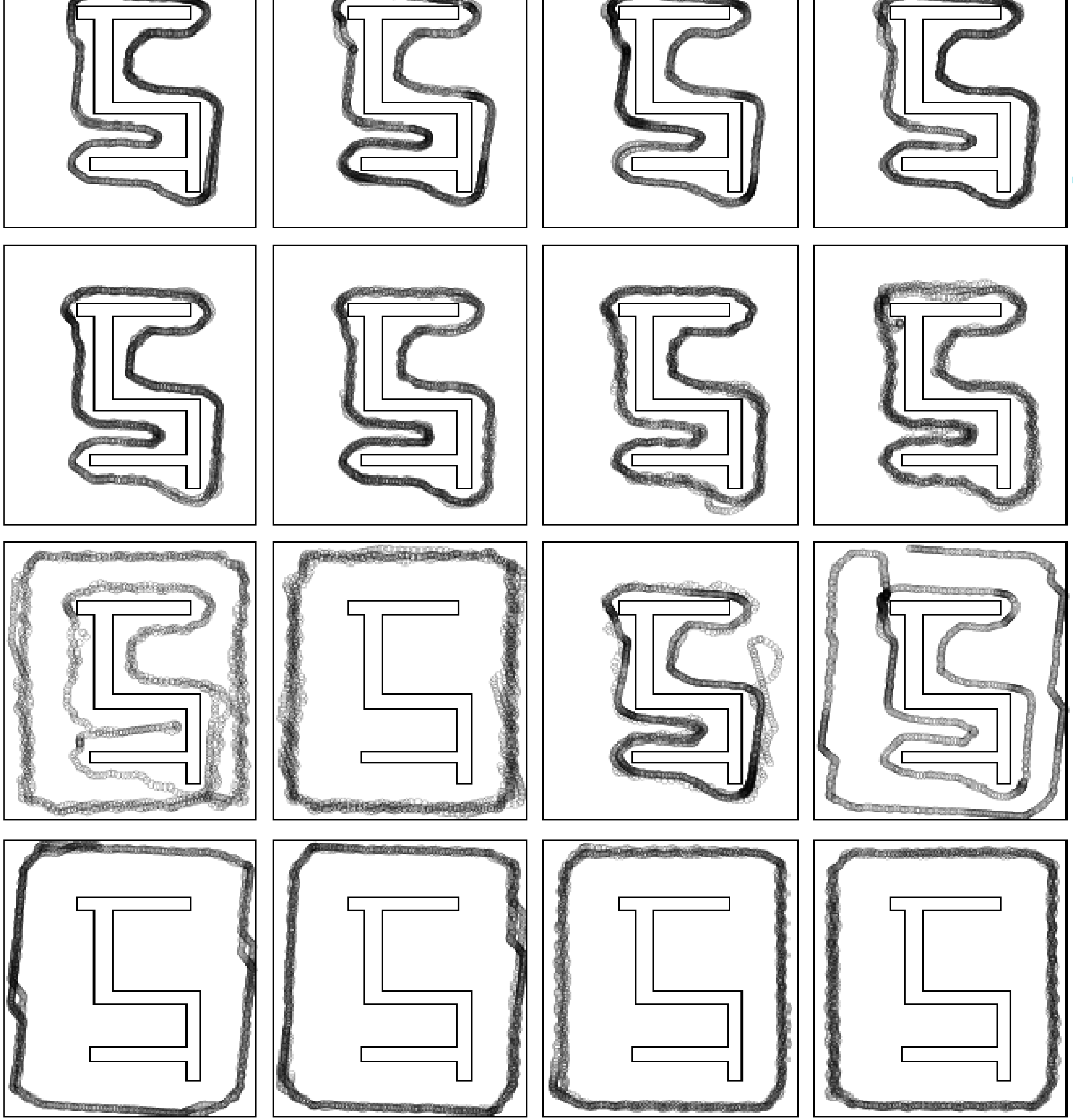
- Inputs: right relative distance, distance coefficient, orientation, and linear velocity
- Variables are computed exclusively from sensorial information of the robot, which is more realistic

- Reward:

$$\boxed{R(DD, VL, \theta_{pared}) = 1 - \left(\alpha_1 \frac{|DD-1|}{3} + \alpha_2 |VL-1| + \alpha_3 \frac{\theta_{pared}}{45} \right)}$$

Experiments Results in the Mobile Robot Problem





Conclusion

- A fuzzy classifier system for real-valued output that properly generates the complete covering map in reinforcement problems is proposed
- It is the first algorithm with such characteristics (at least as far as we known)
- Future work involves investigating the behavior of the proposal in multi-step and real-world problems

Fuzzy-UCS: A Michigan-style Learning Fuzzy-Classifier System for Supervised Learning

Jorge Casillas
Dept. Computer Science and Artificial Intelligence
University of Granada, Spain
http://decsai.ugr.es/~casillas

In collaboration with
Albert Orriols-Puig and Ester Bernadó-Mansilla
Ramon Llull University, Barcelona, Spain

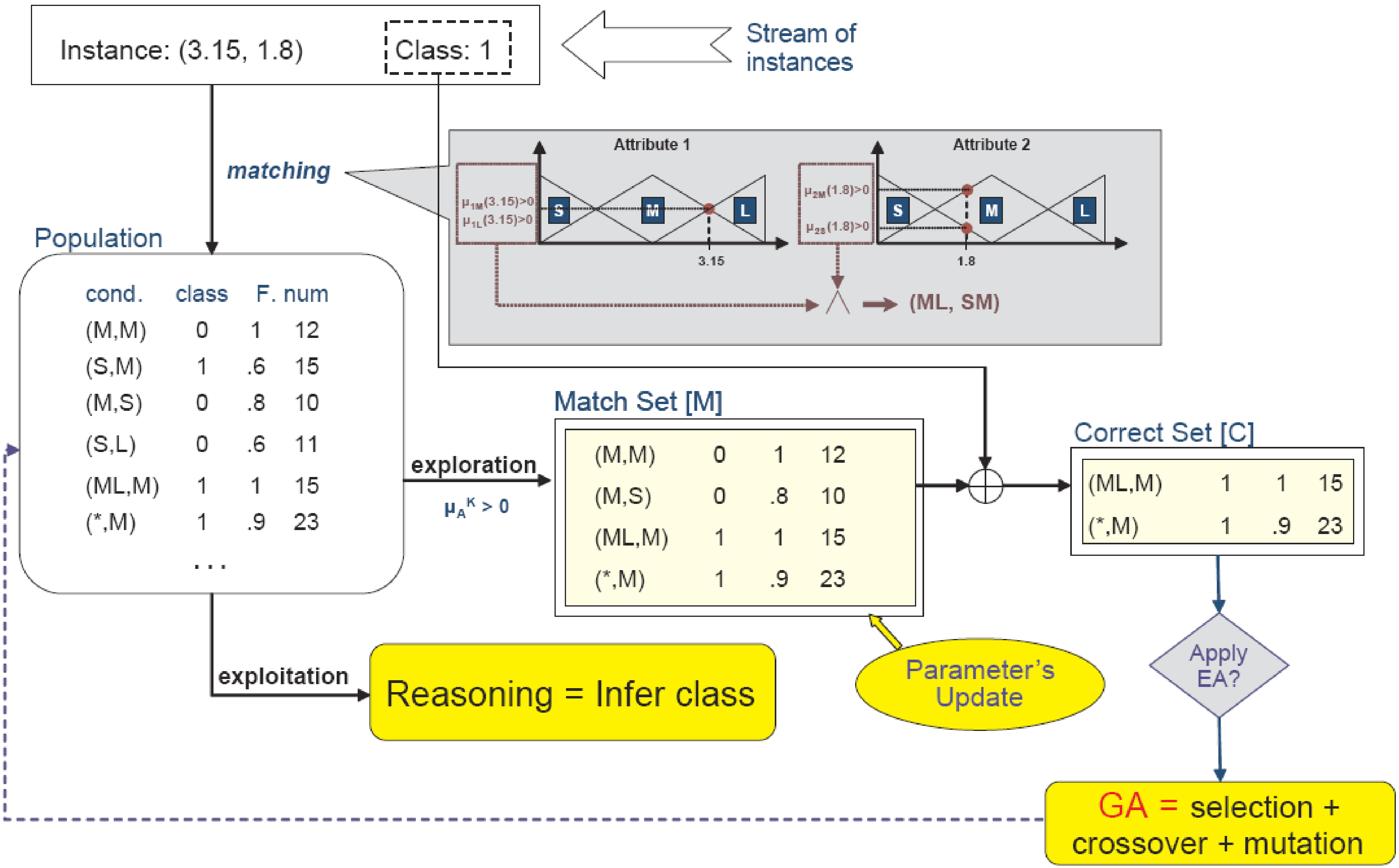
Motivation

- Michigan-style LCSs for supervised learning
  Eg. XCS and UCS
  - Evolve online highly accurate models
  - Competitive to the most-used machine learning techniques
    - [Bernadó et al, 2002; Wilson, 2002; Bacardit & Butz, 2004; Butz, 2006; Orriols & Bernadó, 2007]
- Main weakness: Interpretability of the rule sets
  - Continuous attributes represented with intervals: $[l_i, u_i]$. Semantic-free variables
  - Number of rules or classifiers
    - Tackled with *reduction schemes* [Wilson, 2002; Fu & Davis, 2002; Dixon et al., 2003]

Fuzzy-UCS's Aim

- Accuracy-based Michigan-style LFCS
- Supervised learning scheme
- Derived from UCS [Bernadó & Garrell, 2003]
  - Introduction of a linguistic fuzzy representation
  - Modification of all operators that deal with linguistic rules
- We expect:
  - Achieve similar performance than UCS
  - Higher interpretability since we would deal with linguistic rules
  - Lower number of fuzzy rules in the final population

Description of Fuzzy-UCS



Description of Fuzzy-UCS

- Rule representation
  - Linguistic fuzzy rules
  - E.g.:

IF x_1 **is** A_1 **and** x_2 **is** A_2 … **and** x_n **is** A_n **THEN** $class_1$

Disjunction of linguistic terms

  - All variables share the same semantics
  - Example: A_i = {small, medium, large}

IF x_1 **is** small **and** x_2 **is** medium or large **THEN** class **with** weight

  - Codification:

IF [100 **|** 011] **THEN** $class_1$

Description of Fuzzy-UCS

- Each classifier has the following parameters:

1. Weight per class w_j:
   - Soundness with which the rule predicts the class j.

$$\forall j : w^k_{j_{t+1}} = \frac{cm^k_{j_{t+1}}}{exp^k_{t+1}}$$

   - The class value is dynamic and corresponds to the class j with higher w_j

2. Fitness:
   - Quality of the rule $F^k_{t+1} = w^k_{max_{t+1}} - \sum_{j \mid j \neq max} w^k_{j_{t+1}}$

3. Other parameters:
   - numerosity
   - correct set size
   - experience

Description of Fuzzy-UCS

- Class inference of a test example *e*
  - Weighted average inference
    - All experienced rules vote for the class they advocate:

$$v_k = \mu_{A^k}(e) \cdot F^k$$

    - The votes for each class *j* are added:

$$\forall j : vote_j = \sum_{k|c^k=j}^{N} v_k$$

    - The most-voted class is returned as output
  - Action winner inference
    - Select the experienced rule *k* that maximizes $\mu_{A^k}(e) \cdot F^k$
    - Choose the class of such a rule as output

Description of Fuzzy-UCS

- Rule set reduction
  - Reduction based on weighted average (**wavg**)
    - Remove all the rules that a) are not experienced enough or b) have zero or negative fitness. Weighted average inference is used
  - Reduction based on action winner (**awin**)
    - Only the rules that maximizes the vote v_j for at least one example are kept. Action winner inference is used
  - Reduction based on the most numerous and fitted rules (**nfit**)
    - For each training example, the rule that maximizes the vote weighted by its numerosity (number of copies) is kept:

$$F^k \cdot \mu_{A^k}(e) \cdot num^k$$

    - Weighted average inference is used

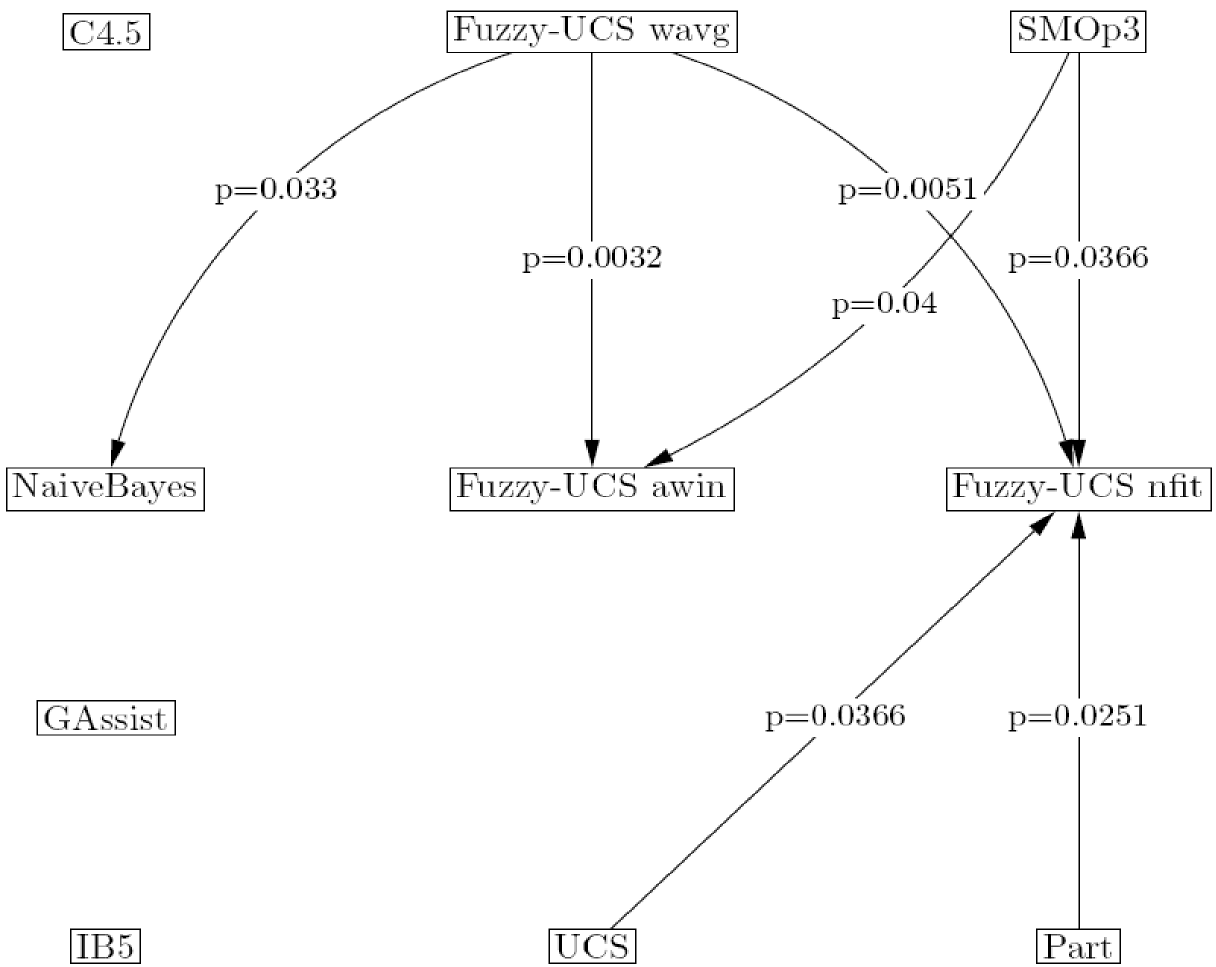
Experimental Methodology

- Compare Fuzzy-UCS to 14 methods (from authors' codes as well as KEEL and Weka software):
  - 6 fuzzy GBMLs (GP, GAP, SAP, AdaBoost, LogitBoost, MaxLogitBoost)
  - 2 state-of-the-art interval-rule-based GBMLs (GAssist, UCS)
  - 6 classical classifiers (C4.5, IBk, Part, NaiveBayes, SMOp3, SMOrbf)
- 21 public data sets (examples with missing values are kept)
- 10-fold cross-validation, averages over 10 runs per partition
- 5 linguistic terms in fuzzy approaches
- Non-parametric statistical tests [Demšar, 2006]: Friedman, Bonferroni-Dunn, and Wilcoxon

Fuzzy-UCS vs. Non Fuzzy Approaches
Performance (test success rate)

Pos.	Algorithm	Ranking
1	SMOp3	4.28
2	UCS	4.65
3	Fuzzy-UCS (wavg)	4.68
4	IB5	5.28
5.5	C4.5	5.95
5.5	Part	5.95
7	GAssist	6.25
8	Fuzzy-UCS (awin)	6.50
9	Naïve Bayes	6.63
10	Fuzzy-UCS (nfit)	7.40
11	SMOrbf	8.95
12	ZeroR	11.50

Wilcoxon signed-ranks test



Fuzzy-UCS vs. Non Fuzzy Approaches
Interpretability

Size of the learnt models

	C4.5	Part	GAssist	UCS	Fuzzy-UCS		
					wavg	awin	nfit
ann	38	15	5	4494	2769	75	36
aut	44	21	7	4565	3872	114	74
bal	45	37	8	2177	1212	114	75
bpa	25	9	6	2961	1440	73	39
cmc	162	168	15	3634	1881	430	271
col	5	9	5	3486	4135	154	81
gls	24	15	5	3359	2799	62	36
h-c	29	21	6	2977	3574	113	46
h-s	17	18	5	3735	3415	117	62
irs	5	4	3	1039	480	18	7
pim	19	7	7	3605	2841	192	62
son	14	8	5	520	3042	178	160
tao	36	17	6	807	111	19	14
thy	8	4	4	1994	1283	37	11
veh	69	32	7	4941	3732	332	147
wbcd	12	10	3	2334	3130	138	28
wdbc	11	7	4	5206	5412	276	101
wne	5	5	3	3685	3686	95	26
wpbc	12	7	4	5299	3772	156	115
zoo	11	8	6	1291	773	16	10

Fuzzy-UCS vs. Non Fuzzy Approaches
Interpretability

C4.5

```
x <= -2.75
|   x <= -3.25: red (308.0)
|   x > -3.25
|   |    y <= 1.75: red (55.0)
|   |    y > 1.75
|   |   |   x <= -3: red (11.0/1.0)
|   |   |   x > -3
|   |   |   |     y <= 4.25: blue (6.0)
|   |   |   |     y > 4.25: red (4.0)
                    ...
```

SMO

```
1.000      * <0.229 0.875 > * X]
0.298      * <0.708 0.437 > * X]
                 ...
```

Part

if x $\leq$ -3.25 **then** red (308)
else if x $>$ 2.75 **then** blue (347/1)
else if y $\leq$ 0 **and** x $\geq$ -1 **then** red (192/1)
...

Fuzzy-UCS vs. Non Fuzzy Approaches
Interpretability

GAssist

if x > 2.72 **and** (y is [0.92,4.61] or y > 5.07) **then** blue
else if (x is [-0.54, 0.54] or x > 2.72) **and** y is [-4.28, -2.57] **then** blue
...
otherwise red

UCS

if x is [-6.00, -0.81] **and** y is [-6.00, 0.40] **then** red **with** acc= 1.00
if x is [2.84, 6.00] **and** y is [-5.26, 4.91] **then** blue **with** acc =1.00
if x is [-6.00, -0.87] **and** y is [-6.00, 0.74] **then** red **with** acc =1.00
...

Fuzzy-UCS

if x is XL **then** blue **with** F=1.00
if x is XS **then** red **with** F=1.00
if x is {XS or S} **and** y is {XS or S} **then** red **with** F=0.87
...

Experiments with a Large Data Set

- Fuzzy-UCS learns from a stream of examples
- The learning can be stalled whenever required. The more learning iterations the system has performed, the more general and accurate the rules
- The cost of the algorithm's runtime increases linearly with the maximum population size, the number of variables per rule, and the number of learning iterations, but it does not depend directly on the number of examples
- We exploit the benefits of online learning to mine large data sets

	#Inst	#Fea	#Re	#No	#Cl	Disp
Kdd'99	494,022	41	35	6	23	$8.3 \cdot 10^{-5}$

Experiments with a Large Data Set

#Iter	wavg		awin		nfit	
	perf.	#rules	perf.	#rules	perf.	#rules
500,000	99.32	1944	99.13	541	99.27	417
1,000,000	99.36	2089	99.07	492	99.25	369
1,500,000	99.37	2178	99.02	460	99.24	350
2,000,000	99.36	2257	99.00	428	99.19	323

Conclusion: Self-analysis

- SWOT analysis:
  - **S**trengths: main advantages of Fuzzy-UCS
  - **W**eaknesses: drawbacks of Fuzzy-UCS
  - **O**pportunities: further works on Fuzzy-UCS
  - **T**hreats: optional approaches considered by other methods than can compete with Fuzzy-UCS

	Positive	Negative
Internal	Strengths	Weaknesses
External	Opportunities	Threats

Conclusion: Self-analysis

Strengths	Weaknesses
• It shows a high performance regarding error rate; comparable with the state-of-the-art in classification • It uses a highly legible knowledge representation based on linguistic fuzzy rules • It performs incremental, on-line learning • It is capable of mining large data sets	• It generates rule sets with size moderate or big (depending on the chosen configuration) • Though it can deal with real, integer or categorical features, it is recommended to be applied only in the two former cases

Conclusion: Self-analysis

Opportunities	Threats
• It is ready to be applied in data streams; further analysis on this problem will be made • Because of the use of fuzzy logic, the algorithm could be adapted to deal with vague and uncertain data • The proposed seminar system opens the door to further works on fuzzy knowledge representation, fuzzy inference engine, and evolutionary operators	• Few interval rules can be more easily interpreted than many linguistic fuzzy rules • Other learning approaches combined with preprocessing can also deal with large data sets