Aprendizaje Supervisado de Reglas Difusas mediante un Sistema Clasificador Evolutivo Estilo Michigan

Albert Orriols-Puig[1,2]

Jorge Casillas[2]

Ester Bernadó-Mansilla[1]

[1]Grup de Recerca en Sistemes Intel·ligents
Enginyeria i Arquitectura La Salle, Universitat Ramon Llull

[2]Dept. de Ciencias de la Computación e IA
Universidad de Granada

Motivation

- **Michigan-style LCSs for supervised learning.** ***Eg. UCS***
  - Evolve ***online*** highly accurate models
  - Competitive to the most-used machine learning techniques
    - (Bernadó et al, 03; Wilson, 02; Bacardit & Butz, 04; Butz, 06; Orriols & Bernadó, 07)

- **Main weakness: Intepretability of the rule sets**
  - Continuous attributes represented with intervals: $[l_i, u_i]$. *Semantic-free variables*
  - Number of rules or classifiers
    - Reduction schemes

    (Wilson, 02; Fu & Davis, 02; Dixon et al., 03, Orriols & Bernadó, 2005)

Motivation

- **Jorge's Proposal:**
  - Let's "fuzzify" UCS
    - Change the rule representation to *fuzzy rules*
- **Framework on Michigan-style Learning Fuzzy-Classifier Systems (LFCS)**
  - (Valenzuela-Radón, 91 & 98)
  - (Parodi & Bonelli, 93)
  - (Furuhashi, Nakaoka & Uchikawa, 94)
  - (Velasco, 98)
  - (Ishibuchi, Nakashima & Murata, 99 & 05): *First LFCS for pattern classification*
  - (Casillas, Carse & Bull, 07) → *Fuzzy-XCS*

Aim

- **Propose Fuzzy-UCS**
  - *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 rules
  - We expect:
    - Achieve similar performance than UCS
    - Higher interpretability
  - Plus new opportunities:
    - Mine in uncertain environments

Outline

1. Description of Fuzzy-UCS
2. Experimental Methodology
3. Results
4. Conclusions

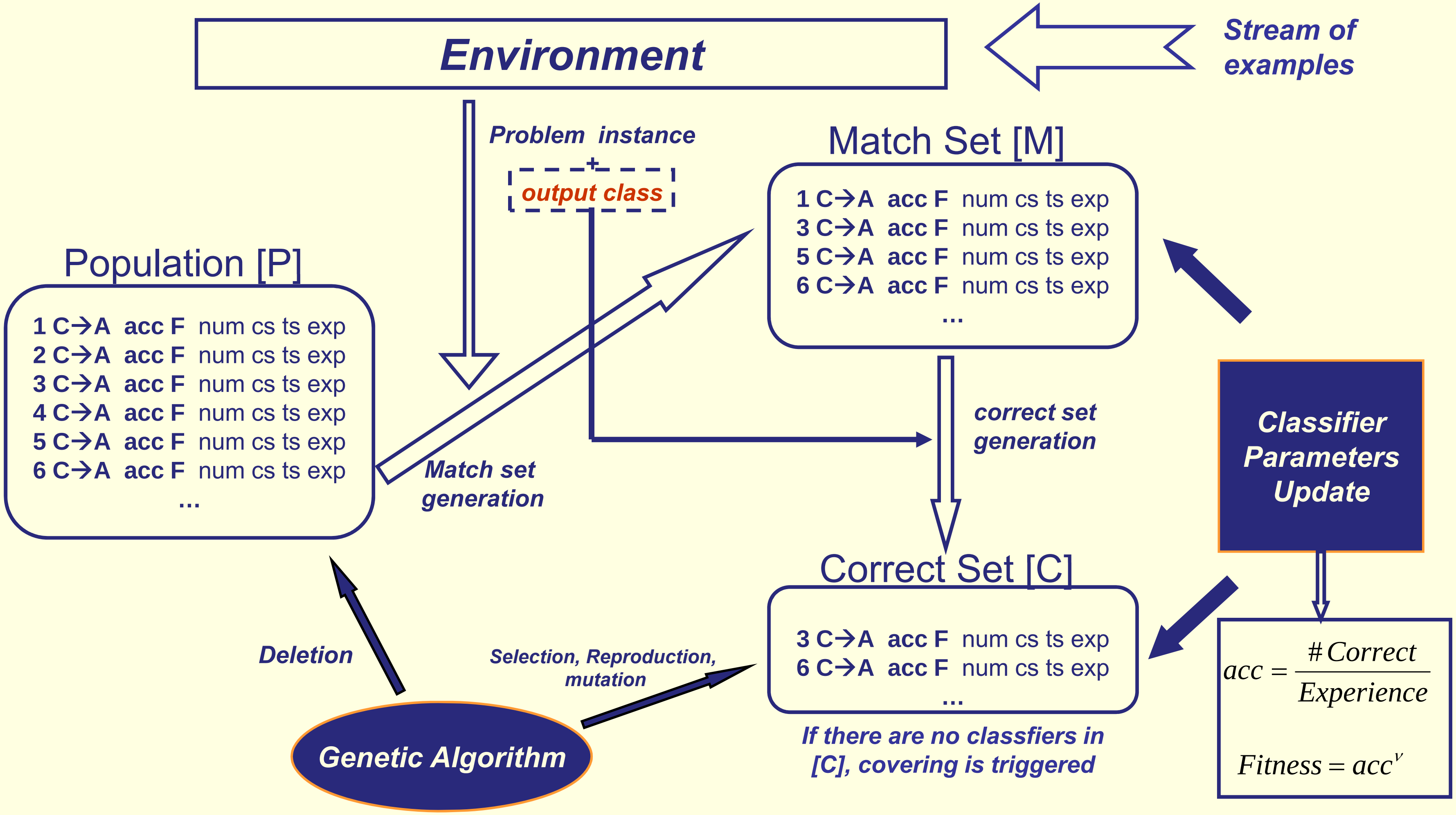
Description of UCS

- **Michigan-style LCS's (Holland, 1975):**
  - Derived from XCS (Wilson, 1995), a reinforcement learning method.
  - Designed specifically for supervised learning
- **Rule representation:**
  - Continuous variables represented as intervals: $[l_i, u_i]$
  - Eg:

    IF x_1 Є $[l_1, u_1]$ ^ x_2 Є $[l_2, u_2]$ … ^ x_n Є$[l_n, n_n]$ **THEN** $class_1$

  - *Matching instance e:* for all e_i: $l_i \leq e_i \leq u_i$
  - Set of parameters: Accuracy, Fitness, Numerosity, Experience, Correct set size

Description of UCS



Description of Fuzzy-UCS

- **Describe the different components**
  1. Rule representation and matching
  2. Learning interaction
  3. Discovery component
  4. Fuzzy-UCS in test mode

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 fuzzy terms

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

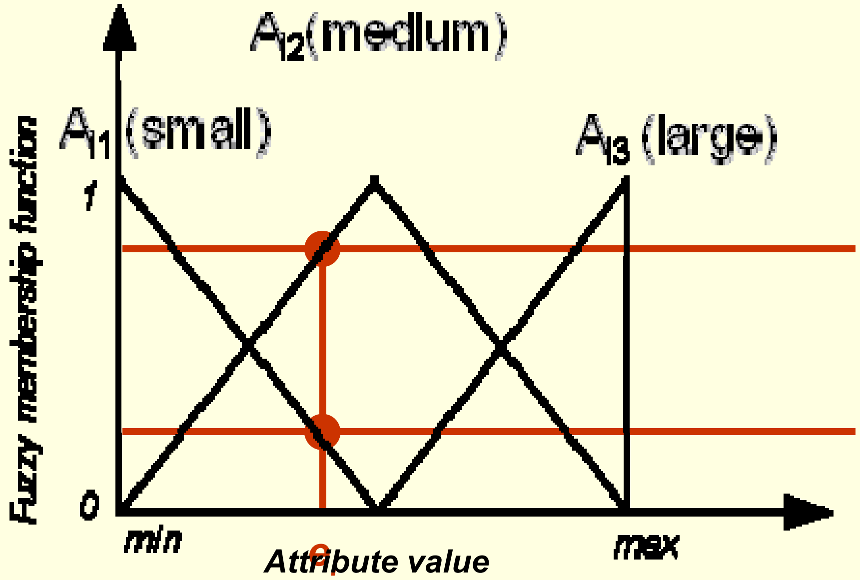
  IF x_1 **is** small **and** x_2 **is** medium or large **THEN** $class_1$

- Codification:

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

Description of Fuzzy-UCS

- **How do we know if a given input is small, medium or large?**
  - Each linguistic term defined by a *membership function*



Belongs to *medium* with a degree of 0.8

Belongs to *small* with a degree of 0.2

Triangular-shaped membership functions

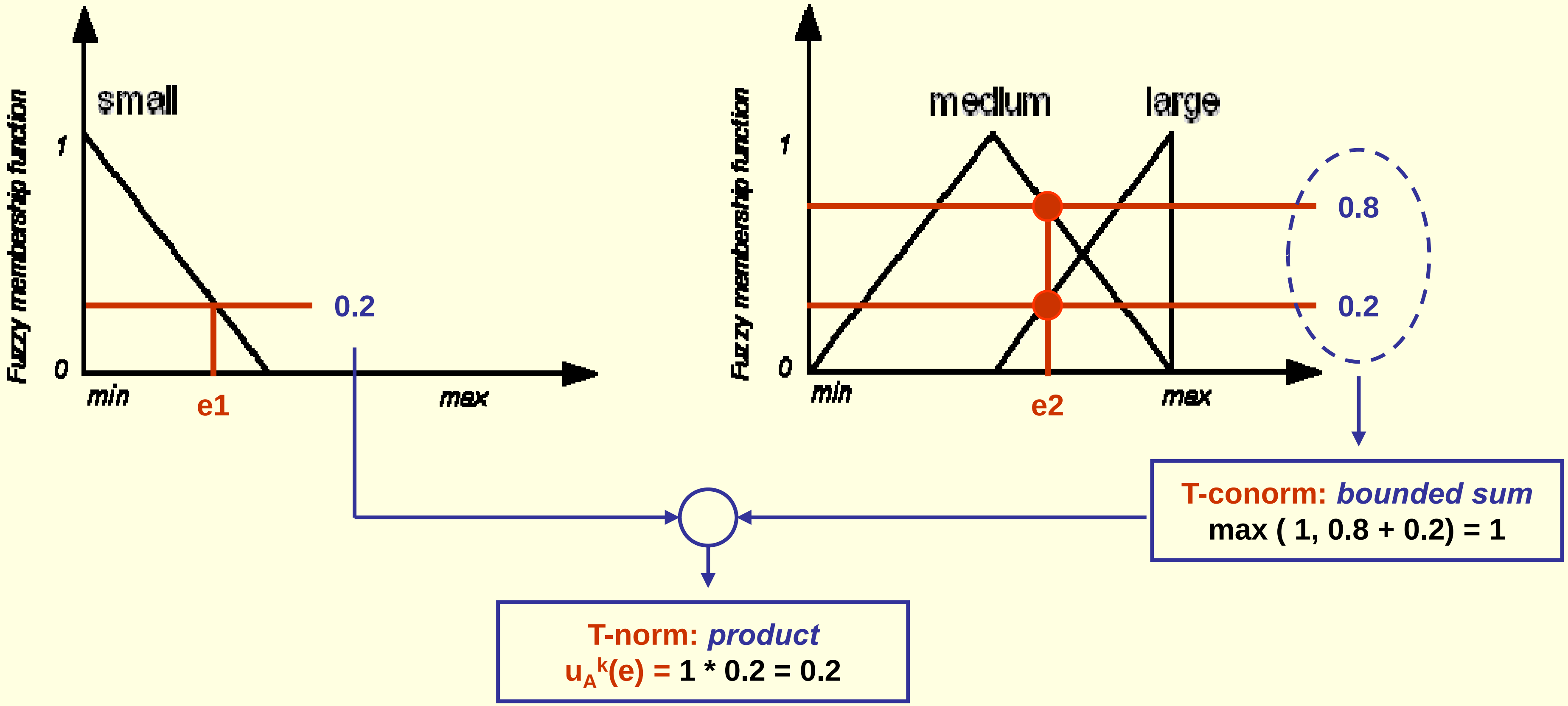
3 linguistic labels $A_i = \{A_{i1}, A_{i2}, A_{i3}\} = \{small, medium, large\}$

Description of Fuzzy-UCS

- ## Matching degree $u_A^k(e)$ → [0,1]

k: **IF** x_1 **is** small **and** x_2 **is** medium or large **THEN** $class_1$

Example: (e1, e2)



Description of Fuzzy-UCS

The role of matching changes:

- **UCS: A rule matches or not an example (binary function)**
- **Fuzzy-UCS: A rule matches an example with a certain degree**

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.**
   - **The *class value is dynamic* and corresponds to the class j with higher w_j**
2. **Fitness:**
   - **Quality of the rule**
3. **Other parameters directly inherited from UCS:**
   - **numerosity**
   - **Experience**

Description of Fuzzy-UCS

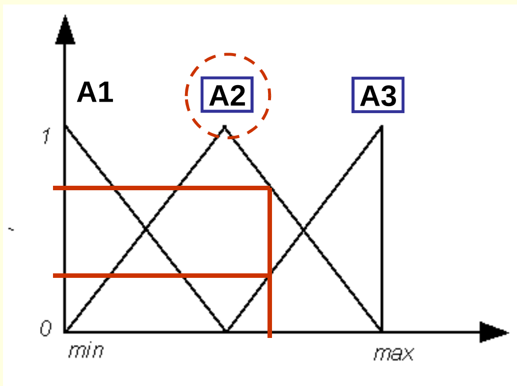
- **Describe the different components**
  1. Rule representation and matching
  2. Learning interaction
  3. Discovery component
  4. Fuzzy-UCS in test mode

Description of Fuzzy-UCS

Learning interaction:

- The environment provides an example *e* and its class *c*
- Match set creation: all classifiers that match with $u_A^k(x) > 0$
- Correct set creation: all classifiers that advocate *c*
- Covering: if there is not a classifier that maximally matches e
  - Create the classifier that match the input example with maximum degree.
  - Generalize the condition with probability $P_{\#}$

For each variable:



Description of Fuzzy-UCS

Parameters' Update

- Experience:

$$exp_{t+1}^{k} = exp_{t}^{k} + \mu_{A^k}(e)$$

- Sum of correct matching per class *j* cm_j:

$$cm_j^k = cm_j^k + m(k,j)$$

$$m(k,j) = \begin{cases} \mu_{A^k}(e) & \text{if j=c} \\ 0 & \text{otherwise} \end{cases}$$

Description of Fuzzy-UCS

Parameters' Update

– Use *cm* to update of the weights per each class:

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

- **Rule that only matches instances of class c:**
  - **$w_c = 1$**
  - **For all the other classes j: $w_j = 0$**
- **Rule that matches instances of all classes:**
  - **All weights w_i ranging [0, 1]**

– Calculate the fitness

$$F^k_{t+1} = w^k_{max_{t+1}} - \sum_{j|j \neq max} w^k_{j_{t+1}}$$

Pressuring toward rules that correctly match instances of only one class

Description of Fuzzy-UCS

- **Describe the different components**
  1. Rule representation and matching
  2. Learning interaction
  3. Discovery component
  4. Fuzzy-UCS in test mode

Description of Fuzzy-UCS

- **Discovery component**
  - Steady-state niched GA
  - Roulette wheel selection

Instances that have a higher matching degree have more opportunities of being selected

$$\boxed{p_{sel}^{k} = \frac{(F^k)^{\nu} \cdot \mu_{A^k}(e)}{\sum_{i \in [C]} (F^i)^{\nu} \cdot \mu_{A^k}(e)}}$$

Description of Fuzzy-UCS

Discovery component

- Crossover and mutation applied on the antecedent
  - 2 point crossover

    IF [100 | 011] **THEN** $class_1$
    IF [101 | 100] **THEN** $class_1$

  - Mutation:
    - Expansion: **IF** [100 | 011] **THEN** $class_1$ → **IF** [10**1** | 011] **THEN** $class_1$
    - Contraction: **IF** [100 | 011] **THEN** $class_1$ → **IF** [100 | 0**0**1] **THEN** $class_1$
    - Shift: **IF** [100 | 011] **THEN** $class_1$ **IF** [**01**0 | 011] **THEN** $class_1$

Description of Fuzzy-UCS

- **Describe the different components**
  1. Rule representation and matching
  2. Learning interaction
  3. Discovery component
  4. Fuzzy-UCS in test mode

Description of Fuzzy-UCS

- **Class inference of a test example *e***
  - ***Combining the information of all rules yields better results than taking a single rule for reasoning (Cordon et al. 1998)***
    - *Inference:*
      - All experienced rules vote for the class they predict as: $u_A{}^k(e) \cdot F^k$
      - The most voted class is returned.

Outline

1. Description of Fuzzy-UCS
2. Experimental Methodology
3. Results
4. Conclusions

Experimental Methodology

Evaluating Fuzzy-UCS' performance

- Compare Fuzzy-UCS' accuracy to:
  - Three non-fuzzy learners: UCS, SMO, and C4.5
  - Two fuzzy learners: Fuzzy LogitBoost and Fuzzy GP
- Default configuration for all methods
- Fuzzy-UCS configuration:

  $iter = 100{,}000$, $N = 6400$, $F_0 = 0.99$, $v=10$, $\{\theta_{GA}, \theta_{del}, \theta_{sub}\} = 50$, $x = 0.8$, $u=0.04$, $P_{\#}=0.6$
- Fuzzy learners: *5 linguistic labels* per variable
- 10-fold cross-validation
- Averages over 10 runs

Experimental Methodology

Data domains

	#Inst	#Fea	#Re	#In	#No	#Cl	%Min	%Max	%MisAtt
annealing	898	38	6	0	32	5	0.9	76.2	0
balance	625	4	4	0	0	3	7.8	46.1	0
bupa	345	6	6	0	0	2	42	58	0
glass	214	9	9	0	0	6	4.2	35.5	o
heart-c	303	13	6	0	7	2	45,5	54.5	15,4
heart-s	270	13	13	0	0	2	44.4	56.6	0
iris	150	4	4	0	0	3	33.3	33.3	0
wbcd	699	9	0	9	0	2	34.5	65.5	11,1
wine	178	13	13	0	0	3	27	39.9	0
zoo	101	17	0	1	16	7	4	40.6	0

Outline

1. Description of Fuzzy-UCS
2. Experimental Methodology
3. Results
4. Conclusions

Results

- **1st objective:** Competitive in terms of performance

Problema	SMO	UCS	C4.5	Fuzzy LBoost	Fuzzy GP	Fuzzy-UCS
annealing	97,24 %	98,83 %	**98,90 %**	76,20 %	77,86 %	98,08 %
balance	86,89 %	78,11 %	77,42 %	80,93 %	69,73 %	**88,85 %**
bupa	58,28 %	**66,75 %**	62,31 %	65,97 %	56,62 %	61,99 %
glass	56,93 %	**69,34 %**	66,15 %	66,29 %	48,89 %	63,23 %
heart-c	**85,15 %**	81,46 %	78,45 %	57,19 %	73,98 %	83,48 %
heart-s	**84,44 %**	74,33 %	79,26 %	58,26 %	73,70 %	81,11 %
iris	**96,67 %**	95,13 %	94,00 %	96,00 %	94,47 %	95,40 %
wbcd	**96,85 %**	96,54 %	94,99 %	95,44 %	93,31 %	95,99 %
wine	**99,44 %**	94,89 %	93,89 %	81,59 %	82,91 %	94,89 %
zoo	96,50 %	**96,66 %**	92,81 %	41,89 %	71,18 %	96,27 %
Ranking	**2,3**	**2,45**	**3,8**	**4,3**	**5,4**	**2,75**

Results

- **2nd objective: Improve the interpretability**

Example of rules evolved by UCS for iris

1. **IF** $x_1 \in [4.30,\ 7.76]$ and $x_2 \in [2.91,\ 4.40]$ and $x_3 \in [1.00,\ 3.77]$ and $x_4 \in [0.10,\ 2.19]$ **THEN** Iris-setosa
2. **IF** $x_1 \in [4.30,\ 6.06]$ and $x_2 \in [2.00,\ 2.53]$ and $x_3 \in [4.30,\ 6.90]$ and $x_4 \in [0.10,\ 2.50]$ **THEN** Iris-virginica
3. **IF** $x_1 \in [4.30,\ 7.71]$ and $x_2 \in [2.00,\ 3.90]$ and $x_3 \in [2.63,\ 6.90]$ and $x_4 \in [0.10,\ 1.39]$ **THEN** Iris-versicolor

Example of rules evolved by Fuzzy-UCS for iris

- Linguistic terms: {XS, S, M, L, XL}

1. **IF** x_3 is {XS, S or L} and x_4 is {XS or S} **THEN** Iris-setosa **WITH** $F = 1$
2. **IF** x_1 is not M and x_4 is {L or XL} **THEN** Iris-virginica **WITH** $F = 1$
3. **IF** x_3 is M and x_4 is not L **THEN** Iris-versicolor **WITH** $F = 1$

Further work

- **Still large rule-sets!**

	Fuzzy-UCS	UCS
annealing	2769	4494
balance	1212	2177
bupa	1440	2961
glass	2799	3359
heart-c	3574	2977
heart-s	2415	3735
iris	480	1039
wbcd	3130	2334
wine	3686	3685
zoo	773	1291

- **Solution: New inference schemes**

Further work

- **Still large rule-sets!**

	Fuzzy-UCS best rule	Fuzzy-UCS	UCS
annealing	36	2769	4494
balance	75	1212	2177
bupa	39	1440	2961
glass	36	2799	3359
heart-c	46	3574	2977
heart-s	62	2415	3735
iris	7	480	1039
wbcd	28	3130	2334
wine	26	3686	3685
zoo	10	773	1291

- **Solution: New inference schemes**

Outline

1. Description of Fuzzy-UCS
2. Experimental Methodology
3. Results
4. Conclusions

Conclusions and Further Work

- **Conclusions**
  - We proposed a Michigan-style LFCS for supervised learning
  - Competitive with respect to:
    - Some of the most-used machine learners: UCS, SMO, and C4.5
    - Recent proposals of Fuzzy-learners: Fuzzy LogitBoost and Fuzzy GP
  - Improvement in terms of interpretability with respect to UCS

- **Further work**
  - Evolve more reduced populations
  - Enhance the comparison with new real-world problems
  - Compare to other LFCS
  - Exploit the incremental learning approach to dig large datasets

Aprendizaje Supervisado de Reglas Difusas mediante un Sistema Clasificador Evolutivo Estilo Michigan

Albert Orriols-Puig[1,2]

Jorge Casillas[2]

Ester Bernadó-Mansilla[1]

[1]Grup de Recerca en Sistemes Intel·ligents
Enginyeria i Arquitectura La Salle, Universitat Ramon Llull

[2]Dept. de Ciencias de la Computación e IA
Universidad of Granada