Scalability in GBML, Accuracy-Based Michigan Fuzzy LCS, and New Trends

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

1. Our Research Group (SCI2S, University of Granada, Spain)
2. Genetic Learning and Scaling Up
3. Fuzzy-XCS: An Accuracy-Based Michigan-Style Genetic Fuzzy System
4. Advances Toward New Methods and Problems
SCI2S Research Group

- Website (members, research lines, publications, projects, software, etc.): http://sci2s.ugr.es

Main research lines:
- KDD and Data Mining with Evolutionary Algorithms
- Fuzzy Rule-Based Systems and Genetic Fuzzy Systems
- Genetic Algorithms and other Evolutionary Algorithms
- Bioinformatics (http://www.m4m.es)
- Intelligent Information Retrieval and Web-access Systems
- Image Registration by Metaheuristics
- Decision Making with Linguistic / Fuzzy Preferences Relations
Outline

1. Our Research Group (SCI2S, University of Granada, Spain)
2. Genetic Learning and Scaling Up
3. Fuzzy-XCS: An Accuracy-Based Michigan-Style Genetic Fuzzy System
4. Advances Toward New Methods and Problems
2.1. Motivation and Objectives

- **How do learning algorithms behave when the data set size increases?**
  - The extraction of rule-based models in large data sets by means of classical algorithms produces acceptable predictive models but the interpretability is reduced.

- C4.5 in KDD Cup’99 (494,022 instances, 41 attributes, 23 classes) data set produces models with \(~99.95\)% of accuracy but with at least 102 rules and 10.52 antecedents per rule in average.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Size</th>
<th>Ant</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4.5 Min</td>
<td>99.96</td>
<td>252</td>
<td>13.34</td>
</tr>
<tr>
<td>C4.5</td>
<td>99.95</td>
<td>143</td>
<td>11.78</td>
</tr>
<tr>
<td>C4.5 Max</td>
<td>99.95</td>
<td>102</td>
<td>10.52</td>
</tr>
</tbody>
</table>

- C4.5 is a fast learning algorithm.
- Large data set → good model but low interpretability.
2.1. Motivation and Objectives

- The Evolutionary Instance Selection reduces the size of the rule-based models (for example, decision trees [CHL03])

- The combination of Stratification and Evolutionary Instance Selection addresses the Scaling Problem which appear in the evaluation of Large Size data sets
2.1. Motivation and Objectives

- **Proposal**: Combination of Evolutionary Training Set Selection with Stratified Strategy to extract rule sets with adequate balance between interpretability and precision in large data sets.

- **Objective**: Balance between Prediction and Interpretability.

- **Quality measures** of the rule sets:
  - Test Accuracy
  - Number of Rules
  - Number of Antecedents

Prediction → Interpretability
2.2. Proposal

Training Set Selection Process

- Data Set (D)
- Training Set (TR)
- Test Set (TS)
- Instance Selection Algorithm
- Training Set Selected (TSS)
- Data Mining Algorithm (C4.5)
- Decision Tree
2.2. Proposal

Stratified Instance Selection for Training Set Selection

Data Set (D)

\[ D_1 \rightarrow DS_1 \]
\[ D_2 \rightarrow DS_2 \]
\[ D_3 \rightarrow DS_3 \]
\[ \ldots \]
\[ D_t \rightarrow DS_t \]

PSA: Prototype Selection Algorithm

Stratified Training Set Select. (STSS\(_i\))

Test Set (TS\(_i\))

Data Mining Algorithm (C4.5)

Decision Tree
2.2. Proposal

Evolutionary Prototype Selection:

- **Representation**

- **Fitness Function**

  \[
  \text{Fitness (TSS)} = \alpha \cdot \text{clasper (TSS)} + (1 - \alpha) \cdot \text{percred (TSS)}
  \]

  \[
  \text{percred (TSS)} = 100 \cdot \frac{|TR| - |TSS|}{|TR|}
  \]
2.3. Experiments

Experimental Methodology:

• **Data Set:**

<table>
<thead>
<tr>
<th></th>
<th>Instances</th>
<th>Attributes</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kdd Cup’99</td>
<td>494,022</td>
<td>41</td>
<td>23</td>
</tr>
</tbody>
</table>

• **Prototype Selection Algorithms and Parameters:**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cnn</td>
<td></td>
</tr>
<tr>
<td>Ib2</td>
<td>Acceptance=0.9, Drop=0.7</td>
</tr>
<tr>
<td>Ib3</td>
<td></td>
</tr>
<tr>
<td>EA-CHC</td>
<td>Population=50, Evaluation=10000, α=0.5</td>
</tr>
<tr>
<td>C4.5</td>
<td>Minimal Prune, Default Prune, Maximal Prune</td>
</tr>
</tbody>
</table>

• **Number of strata: t=100. (4,940 instances per strata)**
## 2.3. Experiments

<table>
<thead>
<tr>
<th></th>
<th>% Red</th>
<th>Input Data Set size for C4.5</th>
<th>Accuracy</th>
<th>Size</th>
<th>Ant</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4.5 Min</td>
<td></td>
<td>444,620</td>
<td>99.96</td>
<td>252</td>
<td>13.34</td>
</tr>
<tr>
<td>C4.5</td>
<td></td>
<td>444,620</td>
<td>99.95</td>
<td>143</td>
<td>11.78</td>
</tr>
<tr>
<td>C4.5 Max</td>
<td></td>
<td>444,620</td>
<td>99.95</td>
<td>102</td>
<td>10.52</td>
</tr>
<tr>
<td>Cnn st 100</td>
<td>81.61</td>
<td>90,850</td>
<td>96.43</td>
<td>83</td>
<td>11.49</td>
</tr>
<tr>
<td>Ib2 st 100</td>
<td>82.01</td>
<td>88,874</td>
<td>95.05</td>
<td>58</td>
<td>10.86</td>
</tr>
<tr>
<td>Ib3 st 100</td>
<td>78.82</td>
<td>104,634</td>
<td>96.77</td>
<td>74</td>
<td>11.48</td>
</tr>
<tr>
<td>EA-CHC st 100</td>
<td>99.28</td>
<td>3,557</td>
<td>98.41</td>
<td>9</td>
<td>3.56</td>
</tr>
</tbody>
</table>
2.4. Conclusions

- **Precision**: The Evolutionary Stratified Instance Selection shows the best accuracy rates among the instance selection algorithms.

- **Interpretability**: The Evolutionary Stratified Instance Selection produces the smallest set of rules with the minimal number of rules and antecedents.

- **Precision-Interpretability**: The Evolutionary Stratified Instance Selection offers high balance between precision and interpretability.

2.4. Conclusions

How to apply genetic learning to large data sets?

- **Problems:**
  - Chromosome size
  - Data set evaluation

- **Solutions:**
  - Stratification?
  - Partial Evaluation?
  - Parallel Learning?
  - ...
**Outline**

1. Our Research Group (SCI2S, University of Granada, Spain)
2. Genetic Learning and Scaling Up
3. **Fuzzy-XCS: An Accuracy-Based Michigan-Style Genetic Fuzzy System**
4. Advances Toward New Methods and Problems
3.1. Objectives

- Fuzzy classifier system (Michigan style): 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. [M. Valenzuela-Rendón, 1998]
  3. [A. Parodi, P. Bonelli, 1993]
  4. [T. Furuhashi, K. Nakaoka, Y. Uchikawa, 1994]
  5. [K. Nakaoka, T. Furuhashi, Y. Uchikawa, 1994]
  6. [J.R. Velasco, 1998]
  7. [H. Ishibuchi, T. Nakashima, T. Murata, 1999]
  8. [A. Bonarini, 1996]
  9. [A. Bonarini, V. Trianni, 2001]
  11. [M.C. Su, et al., 2005]
  12. [C.-F. Juang, 2005]

Except [10], all of them are based on strength. In [10], the output is discrete and generality is not considered.
3.1. Objectives

- Objective: an accuracy-based fuzzy classifier system

- This kind of system would have some important advantages:
  - The use of fuzzy rules allow us to describe in a very legible way state-action relations and to deal with uncertainty
  - The proposal returns continuous output by using linguistic variables also in the consequent
  - It tries to obtain optimal generalization to improve compacity of the knowledge representation. This involves to avoid overgeneral rules
  - It tries to obtain complete covering map

3.2. Difficulties

- To develop an accuracy-based fuzzy classifier system has the following difficulties:
  - Since several rules fire in parallel, 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.
3.3. Competitive Fuzzy Inference

- These problems are in part due to the interpolative reasoning performed in fuzzy systems where the final output results from aggregate the individual contribution of a set of rules.

- However, LCS does not consider the interaction among rules as a cooperative action but rather each rule competes with the rest to be the best for a specific input vector.

- To perform a “competitive” inference we only need to change the roles of the fuzzy operators:
  - ALSO operator → Intersection (T-norm): Minimum
  - THEN operator → logical causality (S-implications)
    - Kleene-Dienes: \[ \mu_R(x, y) = \max \{1 - \mu_A(x), \mu_B(y)\} \]
    - Łukasiewicz: \[ \mu_R(x, y) = \min \{1, 1 - \mu_A(x) + \mu_B(y)\} \]
3.4. Fuzzy-XCS

Parameter updates: error ($\varepsilon$), prediction ($p$), fitness ($F$), and exp

Environment

Match Set [MS]

(S,M)⇒L 43 .01 99 2
(SM,S)⇒M 18 .02 92 5
(*,M)⇒M 24 .17 15 23

Candidate Subsets [CS]

(S,M)⇒L 43 .01 99 2
(SM,S)⇒M 18 .02 92 5
(*,M)⇒M 24 .17 15 23

Action Set [AS]

(S,M)⇒L 43 .01 99 2
(SM,S)⇒M 18 .02 92 5
(*)⇒M 24 .17 15 23

Effectors

8.6

Population

Detectors

maximum weighted mean

discount

delay

EA = selection + crossover + mutation

Apply EA?

environment: the focus is on the interaction between detectors and effectors through candidate subsets and action sets, with parameter updates. The diagram illustrates the process of matching, selecting, and applying evolutionary algorithms for optimization.
3.4. Fuzzy-XCS
3.4.1. Generalization representation

- The disjunctive normal form (DNF) is considered:

\[
\text{IF } X_1 \text{ is } \tilde{A}_1 \text{ and } \ldots \text{ and } X_n \text{ is } \tilde{A}_n \text{ THEN } Y_1 \text{ is } B_1 \text{ and } \ldots \text{and } Y_m \text{ is } B_m
\]

\[
\tilde{A}_i = \{A_{i1} \lor \ldots \lor A_{il}\}, \lor = \min\{1, a + b\}
\]

- **Binary coding for the antecedent** (allele ‘1’ means that the corresponding label is used) and **integer coding in the consequent** (each gene contains the index of the label used in the corresponding output variable)

\[
\text{IF } X_1 \text{ is } P \text{ and } X_2 \text{ is } \{M \lor G\} \text{ THEN } Y_1 \text{ is } M \text{ and } Y_2 \text{ is } G
\]

\[
[100 \mid 011 \mid 23]
\]
3.4. Fuzzy-XCS

3.4.2. Performance component

1. **Match set composition** [M]: a matching threshold is used to reduce the number of fired rules

2. **Computation of candidate subsets** [CS]:
   - Equivalent to the prediction array computation in XCS. XCS partitions [M] into a number of mutually exclusive sets according to the actions
   - In Fuzzy-XCS, several “linguistic actions” (consequents) could/should be considered together
   - In our case, different groups of **consistent** and **non-redundant** fuzzy rules with the maximum number of rules in each group are formed
   - We perform an exploration/exploitation scheme with probability 0.5. On each exploitation step, only those fuzzy classifiers sufficiently experienced are considered. On exploration step, the whole match set is considered

3. **Action set selection** [AS]: It chooses the consistent and non-redundant classifier subset with the highest mean prediction
3.4. Fuzzy-XCS
3.4.3. Reinforcement component

- The prediction error ($\varepsilon_j$), prediction ($p_j$), and fitness ($F_j$) values of each fuzzy classifier $C_j$ are adjusted by the reinforcement learning standard techniques used in XCS: Widrow-Hoff and MAM (modified adaptive method).

- However, an important difference is considered in Fuzzy-XCS: the credit distribution among the classifiers must be made proportionally to the degree of contribution of each classifier to the obtained output.

- Therefore, firstly a weight is computed for each classifier (according to the proposed fuzzy inference) and then parameters are adjusted.

- Fuzzy-XCS acts on the action set.
3.4. Fuzzy-XCS

3.4.2. Reinforcement component (Credit Distribution)

- Let be the scaled output fuzzy set generated by the fuzzy rule $R_j$:

$$B'_j = I(\mu_{A_{R_j}}(x), B_j)$$

- Let $R_1$ be the winner rule, with the highest matching degree $\mu_{A_{R_j}}(x)$
- The process involves analyzing the area that the rival fuzzy rules bite into the area generated by the winner rule $R_1$:

$$W_1 = \frac{\int \bigcap_{j=1}^{\text{|AS|}} \mu_{B'_j}(y) \, dy}{\int \mu_{B'_i}(y) \, dy}$$

- The remaining weight is distributed among the rest of rules according to the area that each of them removes to the winner rule $R_1$:

$$w_j = (1 - w_1) \cdot \frac{\int \mu_{B'_i}(y) \, dy - \int \mu_{B'_i}(y) \land \mu_{B'_j}(y) \, dy}{\sum_{i=2}^{\text{|AS|}} \left( \int \mu_{B'_i}(y) \, dy - \int \mu_{B'_i}(y) \land \mu_{B'_j}(y) \, dy \right)}$$
3.4. Fuzzy-XCS
3.4.2. Reinforcement component (Adjustment)

- Firstly the P (payoff) value is computed:
  \[ P = r + \gamma \cdot \max_{i \in CS} \sum_{j} w_i^j p_j^i \]

- Then, the following adjustment is performed for each fuzzy classifier belonging to the action set using Widrow-Hoff:
  1. **Adjust prediction error values** (with MAM):
     \[ \varepsilon_j \leftarrow \varepsilon_j + \beta \cdot w_j \cdot |AS| \cdot (|P - p_j| - \varepsilon_j) \]
  2. **Adjust prediction values** (with MAM):
     \[ p_j \leftarrow p_j + \beta \cdot w_j \cdot |AS| \cdot (P - p_j) \]
  3. **Adjust fitness values**:
     \[ F_j \leftarrow F_j + \beta \cdot (k'_j - F_j) \]

\[ k'_j = \frac{k_j}{\sum_{R_i \in AS} k_i}, \quad k_j = \begin{cases} \left(\frac{\varepsilon_j}{\varepsilon_0}\right)^{-\nu}, & \varepsilon_j > \varepsilon_0 \\ 1, & \text{otherwise} \end{cases} \]
3.4. Fuzzy-XCS

3.4.4. Discovery component

- **Standard two-point crossover operator:**
  - Only acts on the antecedent
  - Prediction, prediction error, and fitness of offspring are initialized to the mean values of the parents

```
[10 001 1 | 23]  [10 101 1 | 23]
```

```
[10 101 0 | 12]  [10 001 0 | 12]
```
3.4. Fuzzy-XCS

3.4.4. Discovery component

- **Mutation:**
  - If the gene to be mutated corresponds to an input variable:
    - **Expansion**
      
    \[
    [100 \mid 011 \mid 23] \rightarrow [101 \mid 011 \mid 23]
    \]
    
    - **Contraction**
      
    \[
    [100 \mid 011 \mid 23] \rightarrow [100 \mid 010 \mid 23]
    \]
    
    - **Shift**
      
    \[
    [100 \mid 011 \mid 23] \rightarrow [010 \mid 011 \mid 23]
    \]
  
  - If the gene to be mutated corresponds to an output variable:
    - The index of the label is increased or decreased by 1
      
    \[
    [100 \mid 011 \mid 23] \rightarrow [100 \mid 011 \mid 22]
    \]
3.5. Experimental Results
3.5.1. Laboratory Problem: Specification

- First experiment in a laboratory 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
3.5. Experimental Results

3.5.1. Laboratory Problem: Specification

- The problem is actually a function approximation (with real inputs and outputs) where we know the optimal (regarding state/action map and generality) solution.
3.5. Experimental Results

3.5.1. Laboratory Problem: Results

**Competitive distribution and inference**
- **Implication**: Łukasiewicz
- **Aggregation**: minimum
3.5. Experimental Results

3.5.1. Laboratory Problem: Results

Cooperative distribution and inference
- **Implication**: minimum
- **Aggregation**: maximum
### 3.5. Experimental Results

#### 3.5.1. Laboratory Problem: Results

<table>
<thead>
<tr>
<th></th>
<th>Fuzzy-XCS (competitive)</th>
<th>Fuzzy-XCS (cooperative)</th>
<th>Pittsburgh GFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.9</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>R&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.0</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>R&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>R&lt;sub&gt;4&lt;/sub&gt;</td>
<td>1.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>R&lt;sub&gt;5&lt;/sub&gt;</td>
<td>1.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>suboptimal rules</td>
<td>0.8</td>
<td>1.1</td>
<td>7.6</td>
</tr>
<tr>
<td>non-suboptimal rules</td>
<td>1.1</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>MSE analyzed examples</td>
<td>0.000302</td>
<td>—</td>
<td>0.001892</td>
</tr>
<tr>
<td></td>
<td>60,000</td>
<td>60,000</td>
<td>4,212,864</td>
</tr>
</tbody>
</table>
3.5. Experimental Results
3.5.2. Real-World Mobile Robot Problem: Specification

- Second experiment in a real-world problem
- It consists on an *on-line learning* of the wall-following behavior for mobile robots (Nomad 200 model)

- **Input variables (4):** relative right-hand distance, right-left hand distance coefficient, orientation, and linear velocity

- **Output variables (2):** linear velocity and angular velocity

- Variables are computed *using uniquely the sensors* of the robot, so it is more realistic

- **Reward:**

\[
r(RD, LV, \theta_{\text{wall}}) = 1 - \left( \alpha_1 \frac{|RD - 1|}{3} + \alpha_2 |LV - 1| + \alpha_3 \frac{\theta_{\text{wall}}}{45} \right)
\]
3.5. Experimental Results

3.5.2. Real-World Mobile Robot Problem: Results
3.5. Experimental Results

3.5.2. Real-World Mobile Robot Problem: Results
3.6. Conclusions

- A fuzzy classifier system for **real-valued output** that tries to generate the **complete covering map with optimal generalization** is proposed.

- It is the **first algorithm with such characteristics** (at least as far as we known).

- Current work involves investigating the behavior of the proposal in **well-know continuous (state and action) reinforcement learning problems**.
Outline

1. Our Research Group (SCI2S, University of Granada, Spain)
2. Genetic Learning and Scaling Up
3. Fuzzy-XCS: An Accuracy-Based Michigan-Style Genetic Fuzzy System
4. Advances Toward New Methods and Problems
Some Ideas on Advances in LCS/GBML

Fishing Ideas from Machine Learning...

- **Subgroup Discovery**
  - It is a form of supervised inductive learning which is defined as follows: given a population of individuals and a specific property of individuals we are interested in, find population subgroups that are statistically “most interesting”, e.g., are as large as possible and have the most unusual distributional characteristics with respect to the property of interest

Some Ideas on Advances in LCS/GBML

Fishing Ideas from Machine Learning...

- **Learning from Unbalanced Data Sets**
  - In the classification problem field, we often encounter the presence of classes with a very different percentage of patterns between them: classes with a high pattern percentage and classes with a low pattern percentage. These problems receive the name of “classification problems with unbalanced classes” and recently they are receiving a high level of attention in machine learning.

Some Ideas on Advances in LCS/GBML

Taking Advantages from Evolutionary Algorithms...

- GAs are very flexible to deal with mixed coding schemes, combinatorial and continuous optimization, ...

- New Model Representations
  - With fuzzy rules: disjunctive normal form, weighted rules, linguistic modifiers, hierarchical models, ...

J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.)
*Interpretability issues in fuzzy modeling.*

J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.)
*Accuracy improvements in linguistic fuzzy modeling.*
Some Ideas on Advances in LCS/GBML

Taking Advantages from Evolutionary Algorithms...

- Multiobjective EAs: one of the most promising issues and one of the main distinguishing marks in evolutionary computation

- **Evolutionary Multiobjective Learning**
  - *Supervised Learning*: to use multiobjective EAs to obtain a set of solutions with different degrees of accuracy and interpretability, support and confidence, etc.
  - *Reinforcement Learning*: to consider several rewards and integrate multiobjective selection/replacement strategies to deal with that

Some Ideas on Advances in LCS/GBML

Facing Up to Open (Awkward) Problems…

- **Efficient learning** with high dimensional data sets
- Solutions to: noisy data, sparse data, incomplete data, vague data, …
- Dealing with realistic robotics/control problems: variables captured from actual sensors, real-valued actions, …
KEEL Project

- We are working in a national research project to:
  - Research on most of the topics previously mentioned, among others
  - Develop a software for KDD by evolutionary algorithms (preprocessing, learning classifier systems, genetic fuzzy systems, evolutionary neural networks, statistical tests, experimental setup, ...)

- **KEEL**: Knowledge Extraction by Evolutionary Learning
  - 5 Spanish research groups
  - About 50 researchers

- Website: [http://www.keel.es](http://www.keel.es)