High Classification Accuracy does not Imply Effective Genetic Search

Tim Kovacs† and Manfred Kerber††
The University of Bristol†
The University of Birmingham††
kovacs@cs.bris.ac.uk
http://www.cs.bris.ac.uk/~kovacs
Subjects of talk:

- The high classification accuracy achievable with the initial random rule population of a classifier system

- Problems with using classification accuracy on small tasks to evaluate genetic search

- The utility of certain other performance metrics
Aims

Aims of talk:

- Demonstrate ease of widely-used small multiplexer tasks using standard performance metrics

- Demonstrate use of a more discerning metric: %[O]

- **not** to demonstrate a practical technique (scales poorly)
Overview

1. Learning Classifier Systems

2. Ease of small multiplexers

3. %[O]: a more powerful metric

4. Conclusion
Learning Classifier Systems

LCS are:

- learning systems combining evolutionary and supervised, unsupervised or reinforcement learning algorithms

- applicable to classification, clustering and control tasks

- much more complex than a genetic algorithm or a tabular Q-learning system

- nonetheless a relatively simple framework with interesting dynamics
XCS learns if-then rules in Disjunctive Normal Form. E.g.

\[
\begin{align*}
&\text{if } \# 0 1 \text{ then } 1 \\
&\text{if } \### \text{ then } 0
\end{align*}
\]

Each rule predicts the reward it will receive, and its fitness is based on its prediction accuracy.

Conflict resolution is a vote of all matching rules weighted by their prediction and accuracy.
Rule Discovery in XCS and XCS-NGA

XCS generates rules in 3 ways:

1. We generate an initial population of rules with random conditions and actions

2. When no rule matches an input, we generate a random matching rule ("covering")

3. A steady state niched genetic algorithm

In XCS-NGA (No Genetic Algorithm), only the first two occur
1. Learning Classifier Systems

2. **Ease of small multiplexers**

3. \%[O]: a more powerful metric

4. Conclusion
Training Procedure

- Multiplexers: series of Boolean functions

- generate random binary string as input to learning system

- reward: high for correct response, low otherwise

- plot performance as a measure of classification accuracy: proportion of last 50 inputs correctly classified

- also plot population size: number of rules in rule base
• In XCS-NGA rules have prediction, fitness adapted but *no* genetic search occurs

• XCS-NGA reaches 100% classification: GA not needed
XCS vs. XCS-NGA on the 11 Multiplexer

- XCS-NGA with 3200 rules reaches 98% classification (which would probably already overfit many training sets)
**Discussion**

**Observation:**

- Enough random rules will perform very well on Boolean functions

**Conclusion:**

- Measuring classification accuracy on small functions (or when using large populations) is not an adequate test of rule discovery

However, it is adequate to compare alternative parameterisations of XCS
Classification Accuracy of XCS on 6 Multiplexer

- Performance advantage of initial population discernible

- Thus classification accuracy is useful for this comparison
Overview

1. Learning Classifier Systems

2. Ease of small multiplexers

3. %[O]: a more powerful metric

4. Conclusion
%[O]: A More Powerful Metric

- The 6 multiplexer can be represented by a set of 8 optimally general rules

- plot %[O]: the proportion of this set present in population
%[O] on the 11 Multiplexer

- %[O] consists of 16 rules
- Clearly XCS-NGA cannot find accurate general rules
- Unlike classification accuracy, %[O] is a valid test of the GA
Overview

1. Learning Classifier Systems

2. Ease of small multiplexers

3. $\%[O]$: a more powerful metric

4. Conclusion
Conclusion

- Weighted voting by random rules (XCS-NGA) can perform very well

- Classification accuracy on small functions or with large populations is not a good measure of GA performance

- %[O] is a good measure of GA performance