Compact Rulesets from XCSI

Stewart W. Wilson
Prediction Dynamics, Concord, MA 01742
Department of General Engineering
University of Illinois at Urbana-Champaign IL 61801
wilson@prediction-dynamics.com

Why compact rulesets?

- For data inference, should have both:
  - Performance
  - Knowledge visibility

- Classifier systems (XCS) have exhibited high performance and knowledge visibility (accurate, general rules) in several domains.

- BUT, while performance of XCSI on Wisconsin Breast Cancer problem (WBC) was high, evolved populations had $>$ 1000 classifiers!

- Question: Can we extract a minimal subset (a 'compact ruleset') that maintains performance while showing the knowledge more clearly?

What is the WBC problem?

- 699 instances (individual cases), 65.5% Benign, 34.5% Malignant.

- Nine attributes (Clump Thickness, Marginal Adhesion, etc.), each with values from 1 to 10.

Example of an instance matched by a classifier:

```
|    *     |*         |*         |*         | *        |*         |  *       |*         |*         |  0
|0000000...|0000000000|0000000000|00........|00........|0000000000|0000000000|0000000000|0000000000|  0 1000
```

How does XCSI differ from XCS?

- Assume input $x = (x_1, ..., x_n)$ is a vector of integers.

- Classifier condition is a concatenation of "interval predicates" $int_i = (l_i, u_i)$.

- A match occurs iff $l_i \le x_i \le u_i$ for all x_i in x.

- Crossover (two point) as in XCS.

  A crossover point can occur between any two alleles, i.e., within an interval predicate or between predicates.

- Mutation is different.

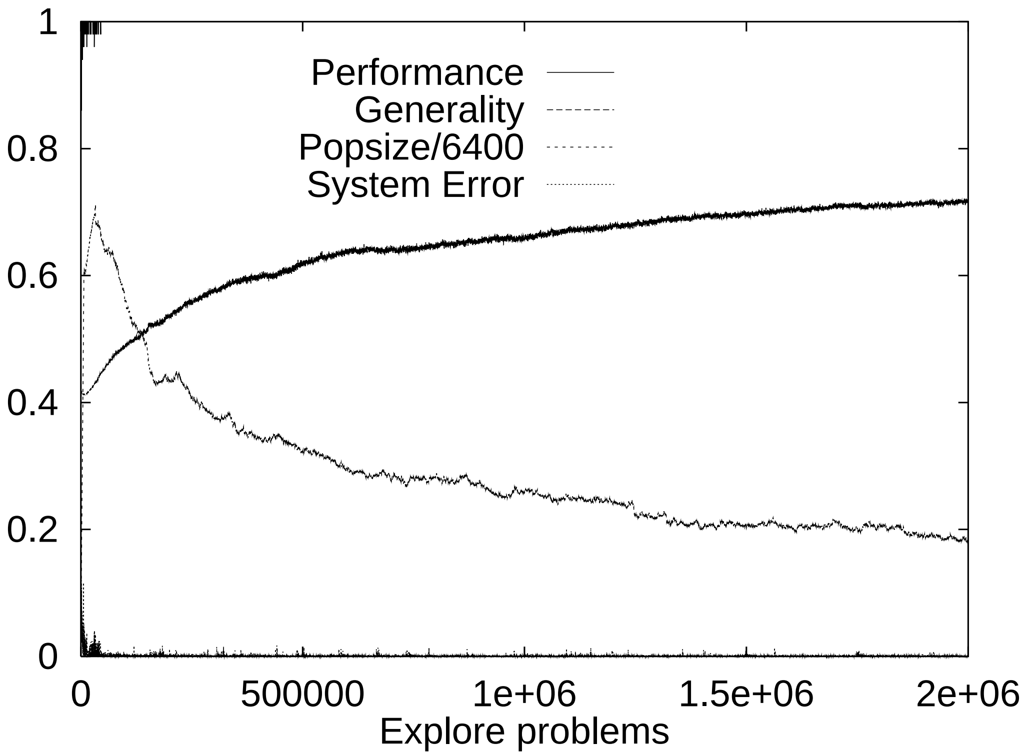
  Mutate an allele by $\pm rand(m_0)$, with m_0 a fixed integer, $rand$ picks randomly from $[0, m_0)$, and the sign is chosen randomly. Resulting values are bounded.

- Condition of a covering classifier is $l_1, u_1, ..., l_n, u_n$ where $l_i = x_i - rand_1(r_0)$, limited by the minimum input value, and $rand_1$ picks randomly from $(0, r_0)$.

  Similarly for u_i.

- Subsumption is similar to XCS, but adapted to interval predicates.

Results on full dataset (from “Oblique” paper)



Note: still 1155 rules at 2,000,000 problems!

How to get the "compact ruleset"?

- Order the population with the "best" classifiers at the top. Call the result $\mathbf{M}$.

- Starting at the top of $\mathbf{M}$ and proceeding downwards, find the smallest sequential subset $\mathbf{M}_n$ that achieves 100% performance on the dataset. Call this $\mathbf{M}_{n^*}$.

- Remove classifiers from $\mathbf{M}_{n^*}$ that did not increase performance as you went down. Call the resulting set $\mathbf{B}$.

- Process $\mathbf{B}$ as follows.

  - Find the classifier in $\mathbf{B}$ that matches the largest number of instances of the dataset D. Place it in an initially empty set $\mathbf{C}$.

  - Remove the matched instances from D.

  - Again find the classifier in $\mathbf{B}$ that matches the most instances of D, add it to $\mathbf{C}$, and remove the matched instances from D.

  - Iterate until D is empty.

- Call the final $\mathbf{C}$ a *compact ruleset* for $\mathbf{M}$, and denote it by $\mathbf{M}_{comp}$.

How do you define “best” classifiers (in forming $\mathbf{M}$)?

- High numerosity — due to frequent selection by the GA, which results from high fitness (accuracy) and high generality.

- High experience — also a good measure.

Each such ordering of the population according to a classifier property is called a *macrostate*. So we speak of a *numerosity macrostate*, an *experience macrostate*, etc.

What does the Compact Ruleset (CR) look like?

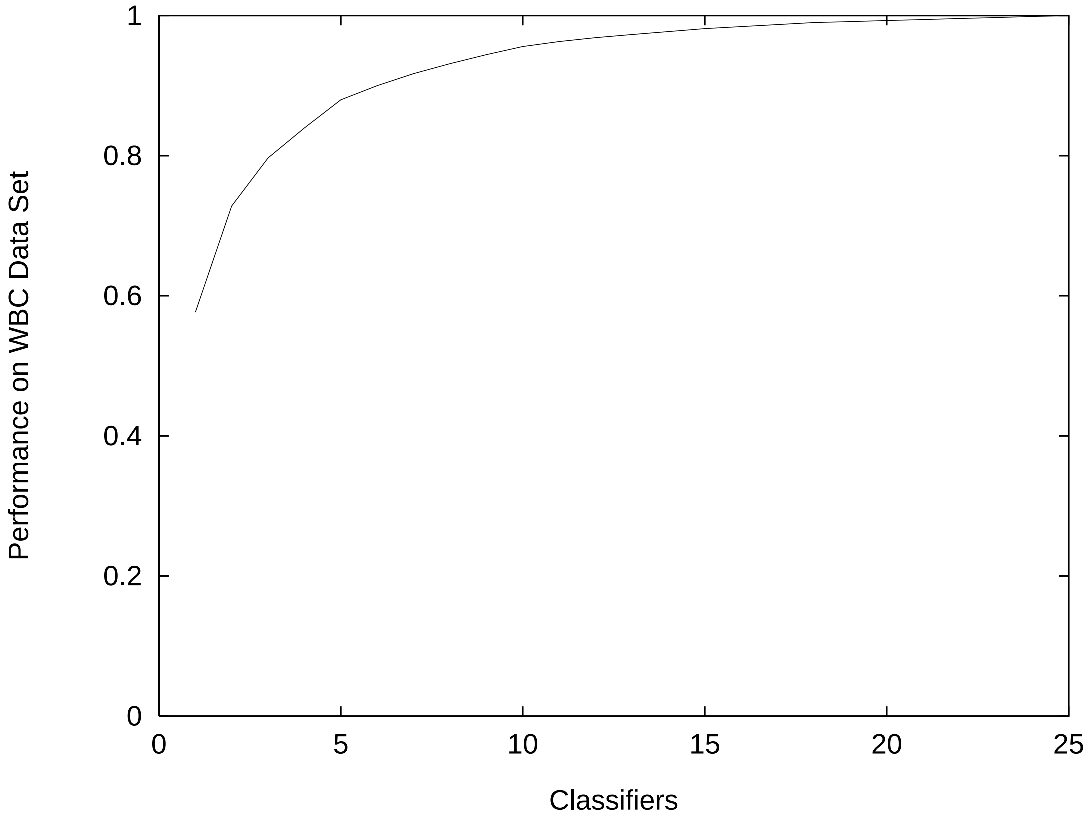
```
 0. |0000000000|0000000000|0000......|0000000000|0000000000|00........|0000000000|00........|0000000000|  0 1000
 1. |......0000|....000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|  1 1000
 2. |0000000000|0000000000|0000000000|0000000000|0000000000|.....00000|....000000|0000000...|0000000000|  1 1000
 3. |0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|.........0|0000000000|  1 1000
 4. |0000000...|0000000000|0000000000|00........|00........|0000000000|0000000000|0000000000|0000000000|  0 1000
 5. |00000.....|0000000000|.....00000|.000000000|0000000000|0000000000|0000000000|0000000000|0000000000|  1 1000
 6. |......0000|0000000000|0000000000|0000000000|0000000000|.......000|0000000000|0000000000|0000000000|  1 1000
 7. |000000....|0000000000|000000000.|000000000.|..00000000|000000....|000000000.|.0000000..|00........|  0 1000
 8. |0000000000|0000000000|..00000000|...0000000|0000000000|0000000...|000000....|00000000..|0000000000|  1 1000
 9. |00000000..|0000000000|0000000000|0000000000|0000000000|0000......|00........|0000000000|0000000000|  0 1000
10. |0000000000|.00000....|0000000000|0000......|..00000000|.....00000|0000000000|0000000000|0000000000|  1 1000
11. |........00|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|  1 1000
12. |0000000000|000.......|0000000000|00000.....|0000000000|..0000000.|0000000000|.00000....|0000000000|  1 1000
13. |.....00000|0000......|.000000000|0000000000|....000000|000.......|0000000000|0000000000|0000000000|  1 1000
14. |00000000..|0000000000|000000000.|....0.....|..00000000|...0000000|000000....|0000000...|000000000.|  0 1000
15. |0000000000|....000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|.000000000|  1 1000
16. |0000000000|0000000000|..00000000|.....00000|0000000000|0000000000|000000....|00000000..|0000000000|  1 1000
17. |000000....|...00000..|.000000...|000000....|..00000...|...00000..|0000......|..00000000|000000....|  0 1000
18. |0000000000|0000000000|0000000000|.000000000|0.........|...0000...|0000000000|0000000000|0000000000|  1 1000
19. |0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|.......000|0000000000|0000000000|  1 1000
20. |.....000..|.000000000|0000000000|.000000...|0000......|0000......|0000000000|0000000...|0000......|  0 1000
21. |....000000|000.......|0000000000|0000000000|00000.....|.000000...|...0......|0000......|000.......|  1 1000
22. |0000000000|0000000000|0000......|0000000000|0000000000|0000......|000.......|0.........|0000000000|  0 1000
23. |....000000|.000......|...0000...|....0.....|00000000..|...0000000|.....00000|0000000000|0000000...|  0 1000
24. |.00000000.|0000......|000000....|......0000|000000000.|..0000....|000000....|00000000..|000000....|  0 1000
```

From 1155 to 25 classifiers, a reduction of 97.8%.

"If Uniformity of Cell Shape $<$ 5 and Bare Nuclei $<$ 3 and Normal Nucleoli $<$ 3 then Benign."
"If Clump Thickness $>$ 6 and Uniformity of Cell Size $>$ 4 then Malignant".
...etc.

Rules "make sense medically" — William H. Wolberg, MD, who compiled the WBC dataset.

What is the marginal performance of each rule in the CR?



First five rules cover 90.0% of D, first ten cover 95.6%.

How would a physician use the CR?

Software displays the rules that match a given case:

```
        1          2          3          4          5          6          7          8          9

   |  *       |*         |*         |*         | *        | *        |  *       |*         |*         |  0
0. |0000000000|0000000000|0000......|0000000000|0000000000|00........|0000000000|00........|0000000000|  0 1000
1. |0000000...|0000000000|0000000000|00........|00........|0000000000|0000000000|0000000000|0000000000|  0 1000
2. |0000000000|0000000000|0000......|0000000000|0000000000|0000......|000.......|0.........|0000000000|  0 1000

   |         *|  *       |    *     | *        |  *       |    *     |   *      |         *| *        |  1
0. |0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|.........0|0000000000|  1 1000
1. |........00|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|0000000000|  1 1000

   |      *   |   *      |   *      |*         | *        |**********|     *    |  *       |*         |  1
0. |0000000000|0000000000|0000000000|0000000000|0000000000|.....00000|....000000|0000000...|0000000000|  1 1000
1. |......0000|0000000000|0000000000|0000000000|0000000000|.......000|0000000000|0000000000|0000000000|  1 1000

   |*         |*         |*         |*         |   *      |  *       |*         |*         |*         |  0
0. |00000000..|0000000000|0000000000|0000000000|0000000000|0000......|00........|0000000000|0000000000|  0 1000
1. |0000000000|0000000000|0000......|0000000000|0000000000|0000......|000.......|0.........|0000000000|  0 1000
```

- Shows attributes that are important.
- Different perspectives.
- Develops intuition.

How about CR performance compared with the full population?

Stratified ten-fold cross-validation experiments:

Fold	Pop. @ 50,000	Pop. @ 2 mill.	nCR of Pop. @ 2 mill.	eCR of Pop. @ 2 mill.
1	0.9714	.9714	.9429	.9714
2	0.9857	.9857	.9714	.9857
3	0.9286	.9143	.9000	.9143
4	0.9429	.9429	.9000	.9286
5	0.9286	.9429	.9286	.9286
6	0.9143	.9429	.9571	.9429
7	1.0000	.9429	.9143	.9286
8	0.9857	.9857	.9286	.9571
9	0.9420	.9565	.9420	.9275
10	0.9571	.9429	.9429	.9286
Mean	0.9556	.9528	.9328	.9413

Summary

- For data inference, population knowledge can be made visible and manageable using a post-processing algorithm such as the CRA.

- Compact rulesets can be useful to practitioners in a variety of ways.

- Compared with the raw populations, performance on new cases is slightly lower.

- A practical software system would include
  - The raw population—for high performance on new cases.
  - The Compact Ruleset—as a summary of "the knowledge", a tool for quick evaluations, and an aid to intuition.