Learning simpler rule sets with multi-objective EAs

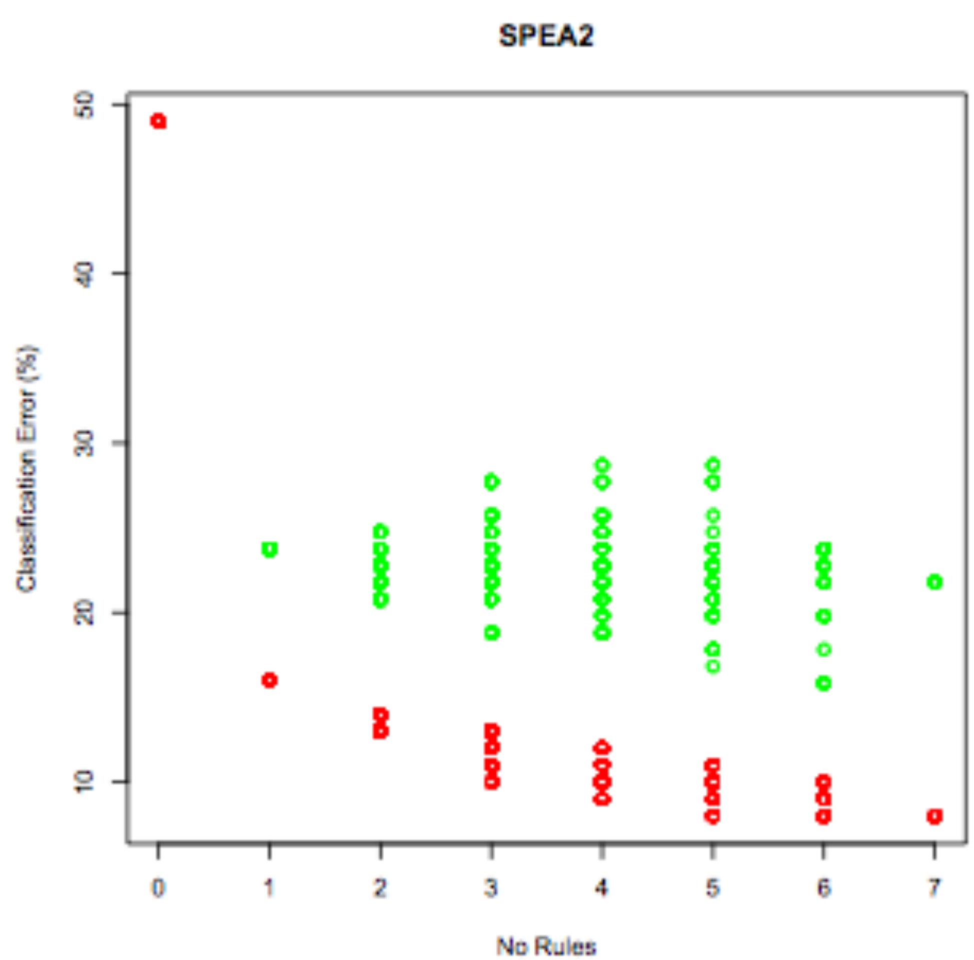
Adam Ghandar*, Zbigniew Michalewicz**, Ralf Zurbruegg*

*University of Adelaide
**also at Polish Japanese Institute of of Computer Science, Polish Academy of Sciences and Polish-Japanese Institute of Information Technology

Summary

- Accuracy interpretability trade-off for optimising prediction/generalisation performance
- Straightforward solution representation that is independent of the optimisation algorithm and partially independent of the rule evaluation implementation
- Experiment results are given for several well known classification test problems which are found to be in the vicinity of good results reported in the literature

Accuracy -interpretability tradeoff

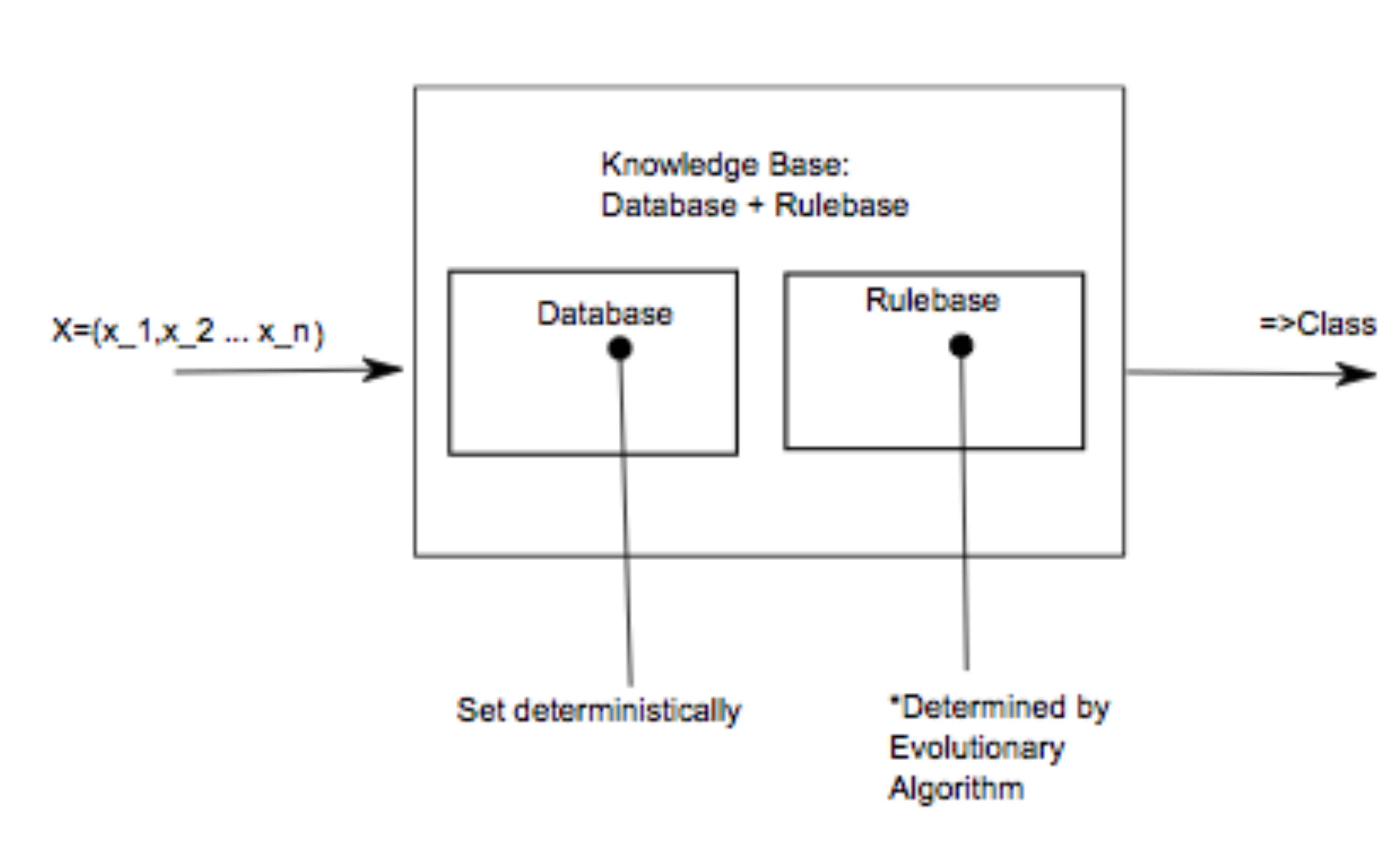


(a) Performance of SPEA2 in and out of the training sample.

```
Error = 33%
petal.width %is% M ,
 classification %is% (0.0,1.0,0.0)
```

```
Error = 2%
sepal.width %is% VL
 && petal.length %is% H,
   classification %is% (0.0,3.0,4.0)
petal.width %is% M ,
 classification %is% (0.0,1.0,0.0)
petal.length %is% H ,
 classification %is% (1.0,1.0,3.0)
petal.width %is% H ,
 classification %is% (1.0,1.0,3.0)
```

Approach


Knowledge Base:
Database + Rulebase
Database
Rulebase
X=(x_1,x_2 ... x_n)
=>Class
Set deterministically
*Determined by
Evolutionary
Algorithm

Solution Representation (genotype)

	in 1	in 2	...	class 1	...	class n
Rule 1	{0,1,..,k}	{0,1,..,k}	...	[0,1]	...	[0,1]
Rule2	{0,1,..,k}	{0,1,..,k}		[0,1]	...	[0,1]
...	...	...	...	...	...	...
Rule M	{0,1,..,k}	{0,1,..,k}	...	[0,1]	...	[0,1]

Rule Representation and evaluation (phenotype)

$$R_k : \text{if } x_1 \text{ is } A_1 \wedge \ldots \wedge x_n \text{ is } A_n; \text{ then } (z_{k,1}, \ldots, z_{k,c})$$

$$z_{k,i} = \frac{\text{Sum of matching degrees of rule k with examples of class i}}{\text{Sum of total matching degrees of rule k for all examples}}$$

$$eval^{\mathrm{TSKIII}}(\mathbf{x}) = \max_{k=1}^{M} \left\{ \prod_{j=1}^{n} \{\mu_j(x_j)\} \right\}$$

- Mu - triangular MFs
- X- feature vector

Evolutionary computation

- Evolutionary Multi-objective Algorithms (EMO):
- Apply mutation and recombination operations on a population of rulebases successively over many generations to find a set of non dominated rulebases (w.r.t to the objectives)
- Recombination involves swapping whole rules, mutation in/decrements inputs at random with a probability to switch off rules and inputs (to limit difference between parents and offspring and bias selection to simpler rule bases)
- The rule consequents are set deterministically to be a measure of confidence the input implies a class
- Multiple solutions (ie a Pareto front is obtained from a single run)
- Work best when limited to around 3-4 objectives, SPEA2 performs better with >3

Objectives

- Number of rules,
- Number of inputs per rule,
- Classification error in training data.

Algorithms

- NSGAII
- SPEA2
- MOCell
- Steady state NSGAII
- FPGA

Experimentation

	Iris	BC	Glass	Ionospher	Diabetes
# features	4	9	10	34	8
# classes	3	2	7	2	2
instances	150	286	214	354	786

Dataset	Reported Error *	Reference
Breast Cancer	4.1 - 6.5	[7, 10]
Iris	0.5 - 4	[3]
Glass	24.4, 32.06	[1, 6]
Ionosphere	13.1 (C4.5 algorithm res. = 5.9), 5-6 (Fung's res.)	[3, 5]
P I Diabetes	26 - 27 (C4.5 was 24.4)	[3]

*decision trees, ANN, etc

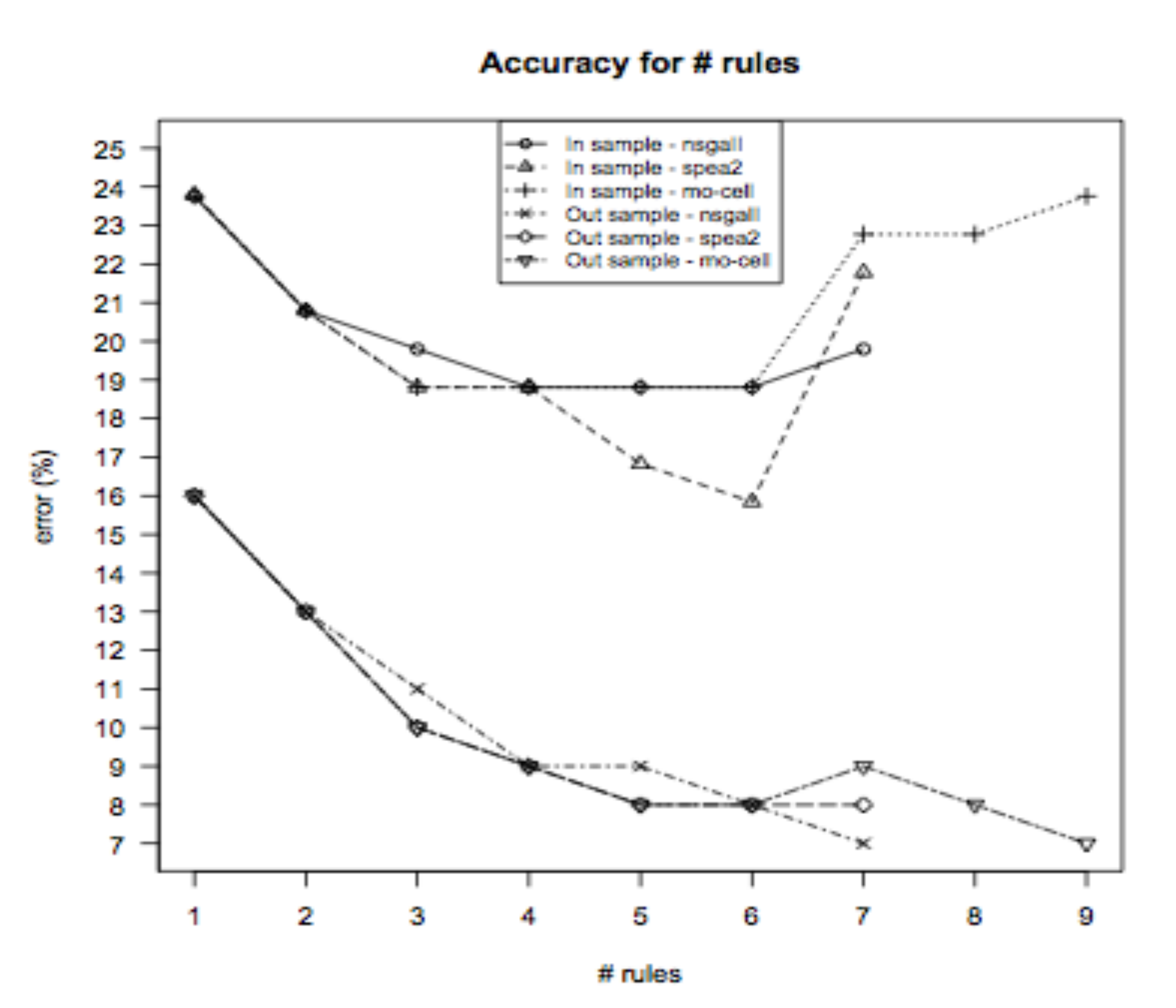
Results

- Best out of sample classification error

	NSGAII	SPEA2	SSNSGA	FPGA	Cell
BC					
FS 1	6.35*	6.44*	7.32	10.53	7.86
FS 2	37.5	39	38.9	39.72	37.87
FS 3	4.47*	4.29*	5.33*	7.11	6.17
FS 4	36.11	38.2	38.9	37.5	39.1
Iris					
FS 1	1.67*	2.29*	4.16	3.33*	5
FS 2	0.83*	2.5*	1.67*	5	3.33*
FS 3	2.08*	1.67*	1.67*	3.33*	4.16
FS 4	4.17	3.33*	1.67*	4.16	1.67*
Glass					
FS 1	36.05	35.19	32.57	33.14	34.3
FS 2	34.88	41.86	34.89	40.69	42.44
FS 3	30.81*	28.49*	30.81*	33.72	30.23*
FS 4	38.95	37.2	37.21	36.05	41.86
Ionosphere					
FS 1	14.084*	15.84	18.66	19.71	23.94
FS 2	60.3	59.32	58.09	64.789	62.67
FS 3	16.54	19.71	20.422	21.83	17.25
FS 4	64.12	63.98	63.38	60.91	63.03
Diabetes					
FS 1	23.7*	23.86*	20.13 *	21.76*	22.89*
FS 2	25.32	25.81	28.24	24.18	28.41
FS 3	22.72*	18.83 *	21.1 *	20.29*	19.96*
FS 4	29.38	26.79	27.59	23.53*	25.32

Share price prediction example

- Divergence of the in and out of sample classification error for prediction share price change direction
- Features: 10 variables measuring price change over time periods 1 to 100 days (P_t/P_{t-period}
- 3 classes increase stay the same or decrease
- Chart shows prediction error and training error



Conclusions

- An approach to learning classifiers was described which showed performance comparable to best results for test problems reported in the literature (sometimes better)
- The internal rulebase representation and associated operators could be used with other rule evaluation methods and other modern heuristic optimisation algorithms
- In addition, there was some specialisation of the approach to the problem (novel variation operators and separate determination of rule consequents)