I Workshop on Knowledge Extraction based on Evolutionary Learning

Introduction to Multi-objective Learning: Some results on the use of MOEAs for tuning FRBSs parameters

Rafael Alcalá
Dept. of Computer Science and Artificial Intelligence
University of Granada
18071 – Spain
http://www.ugr.es/~ralcala/
alcala@decsai.ugr.es

In collaboration with
María José Gacto, Jesús Alcalá-Fdez and Francisco Herrera

Granada, May 16, 2008

University of Granada

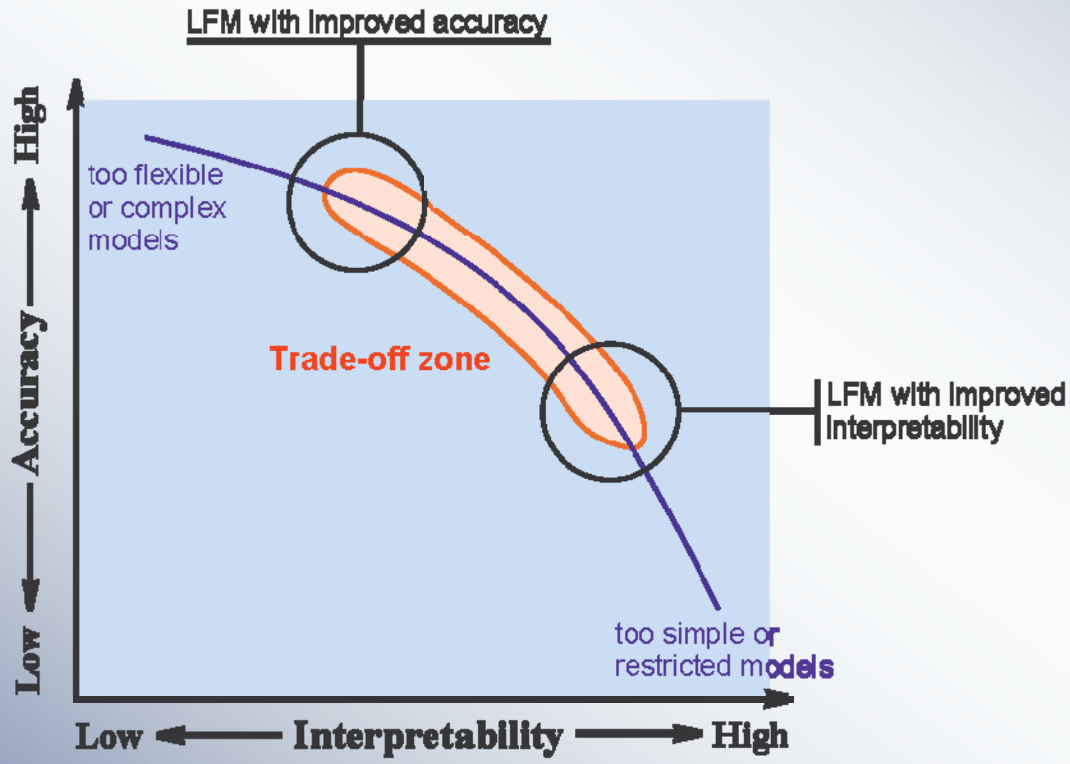
Research Group on Soft Computing and Intelligent Information Systems
http://sci2s.ugr.es

Multiobjective Genetic Tuning of Fuzzy Models: Present results and future directions

1. MOTIVATION AND STATE OF THE ART
2. PARETO FRONTIER TYPE
3. STUDY ON THE APPLICATION OF MOEAs FOR RULE SELECTION AND TUNING OF FUZZY MODELS
4. FURTHER DIRECTIONS

MOTIVATION AND STATE OF THE ART

- **One of the aims that focus the research in the Linguistic Fuzzy Modeling (LFM) area is the trade-off between interpretability and accuracy.**

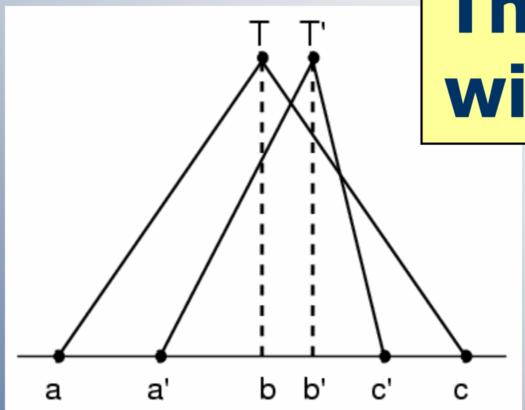


- **A way to improve the accuracy of the LFM is to apply rule selection together with a tuning of MFs.**

MOTIVATION AND STATE OF THE ART

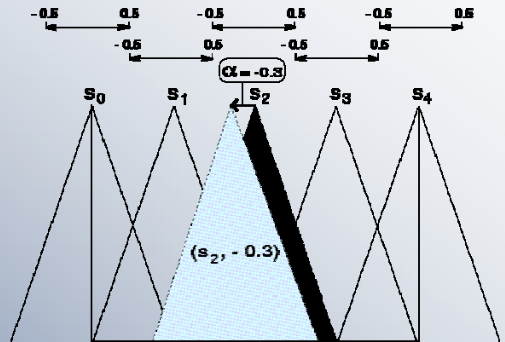
- **In the last years we applied different tuning approaches obtaining good results in different problems**

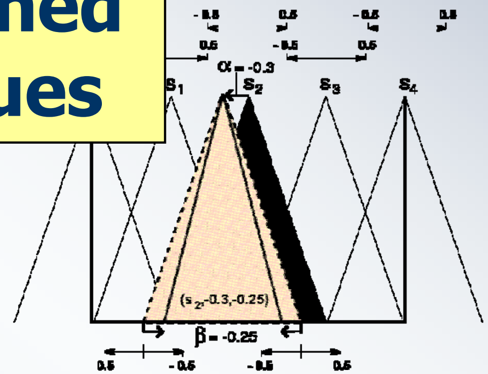
Classic Tuning



They have been also combined with rule selection techniques

(IEEE TFS 2007)

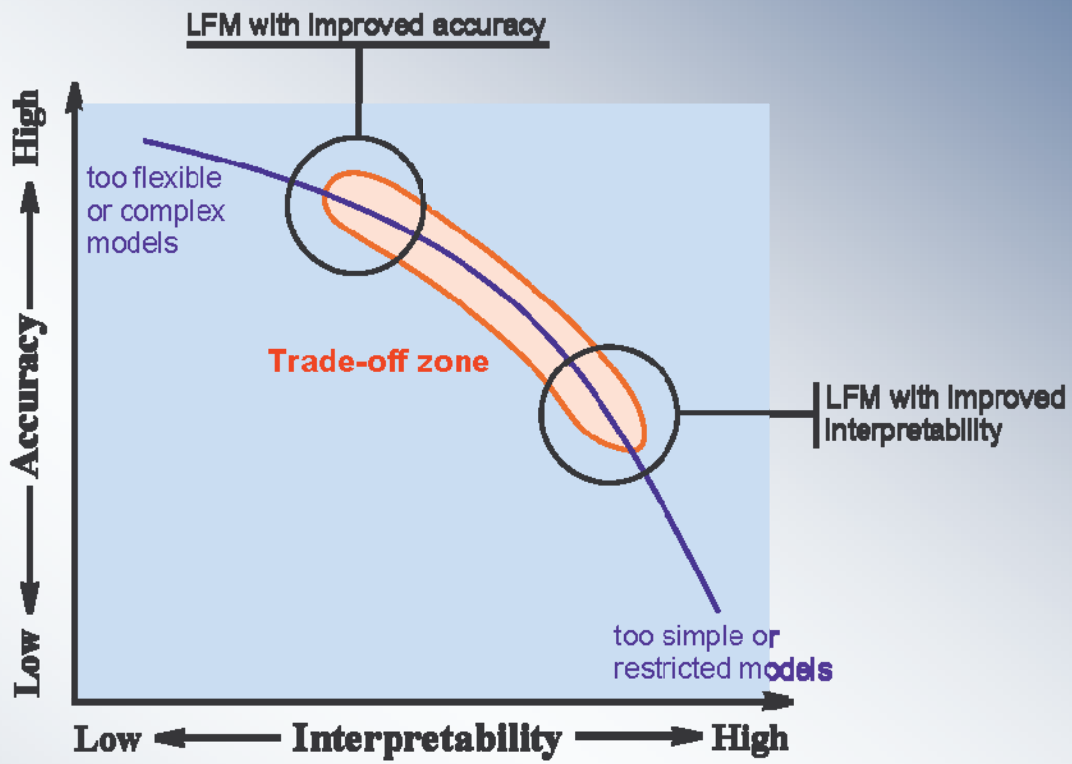




Lateral and Amplitude Tuning (Soft. C. 2007)

MOTIVATION AND STATE OF THE ART

- **What are we searching for?**



- **What we want to see is if it is possible to improve the model interpretability without losing the improved accuracy that a tuning of membership functions together with a rule selection could give**

MOTIVATION AND STATE OF THE ART

- Our main aim was to attain the desired balance by maintaining the improved accuracy that a tuning of MFs could give but trying to obtain more compact models respect with algorithms only guided by accuracy measures.

- Since this problem presents a multi-objective nature the use of Multi-Objective Genetic Algorithms (MOGAs) to obtain a set of solutions with different degrees of accuracy and interpretability is an interesting way to work.

- To do that, we propose the use of MOEAs for rule selection together with the tuning of MFs tries to obtain accurate solutions with the least number of possible rules.

MOTIVATION AND STATE OF THE ART

(State of the Art)

There are not many MOEAs considering interpretability and accuracy as objectives

[1] H. Ishibuchi, T. Murata, I.B. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, Fuzzy Sets and Systems 89:2 (1997) 135–150.

[2] O. Cordon, F. Herrera, M.J. del Jesus, P. Villar, A multiobjective genetic algorithm for feature selection and granularity learning in fuzzy-rule based classification systems, Proceedings of IX IFSA World Congress and XX NAFIPS International Conference (Vancouver, Canada, 2001) 1253–1258.

[3] H. Ishibuchi, T. Nakashima, T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, Information Sciences 136 (2001) 109–133.

[4] H. Ishibuchi, T. Yamamoto, Interpretability Issues in Fuzzy Genetics-Based Machine Learning for Linguistic Modelling, in: J. Lawry, J.G. Shanahan, A.L. Ralescu (Ed.), Modelling withWords: Learning, Fusion, and Reasoning within a Formal Liguistic Representation Framework, Lecture Notes in Computer Science 2873, Springer-Verlag (Berlin, Heidelberg, 2003) 209–228.

[5] H. Ishibuchi, T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Systems 141:1 (2004) 59–88.

[6] K. Narukawa, Y. Nojima, H. Ishibuchi, Modification of evolutionary multiobjective optimization algorithms for multiobjective design of fuzzy rule-based classification systems, Proceedings of 2005 IEEE International Conference on Fuzzy Systems (Reno, USA, 2005) 809-814.

[7] H. Ishibuchi, Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning 44:1 (2007) 4–31.

[8] F. Jimenez, A.F. Gomez-Skarmeta, H. Roubos, R. Babuska, A Multi-Objective Evolutionary Algorithm for Fuzzy Modelling, Proceedings of 9th IFSA World Congress and 20th NAFIPS International Conference (Vancouver, Canada, 2001) 1222–1228.

[9] H.L. Wang, S. Kwong, Y.C. Jin, W. Wei, K.F. Man, Multi-Objective Hierarchical Genetic Algorithm for Interpretable Fuzzy Rule-Based Knowledge Extraction, Fuzzy Sets and Systems 149:1 (2005) 149–186.

[10] H.L. Wang, S. Kwong, Y.C. Jin, W. Wei, K.F. Man, Agentbased evolutionary approach for interpretable rule-based knowledge extraction, IEEE Trans. On Systems, Man, and Cybernetics - Part C: Applications and Reviews 35:2 (2005) 143–155.

MO

No Tuning was used to obtain linguistic models and there exists only one approach for regression problems

Use of MOEAs for finding the accuracy-interpretability trad

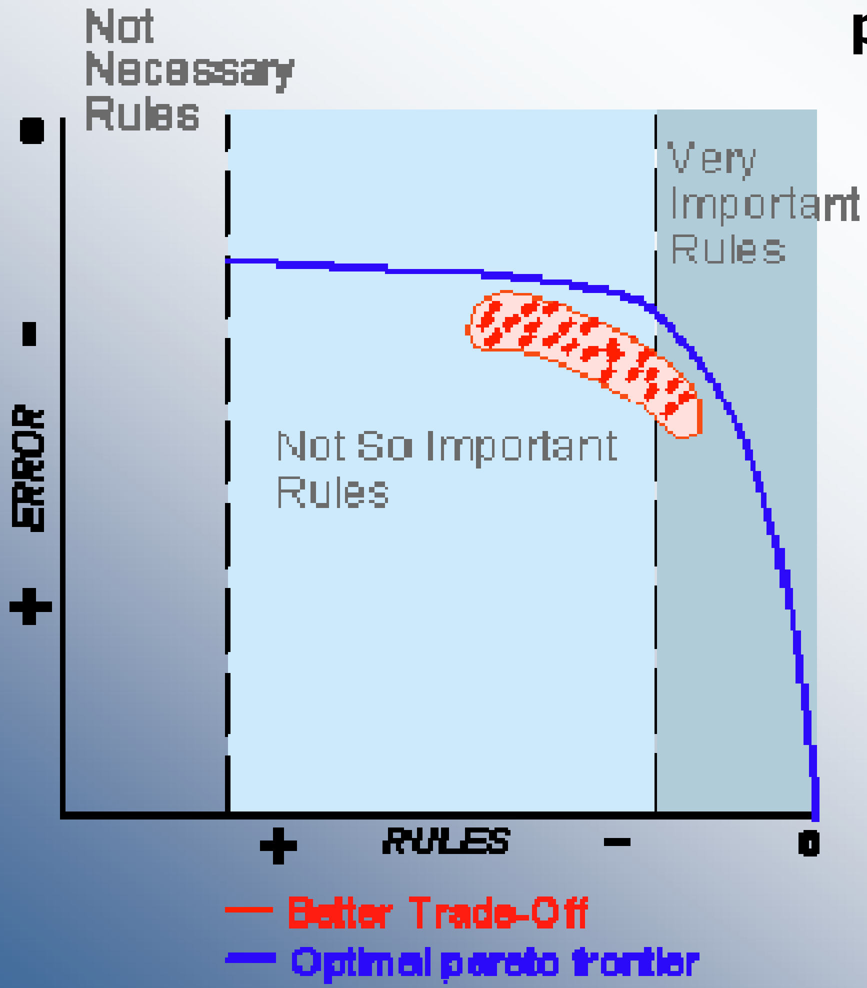
Year	Ref.	Problem	MOEA/Gen.	#Objs.	RS	FS	RL	LP
MAMDANI LINGUISTIC MODELS (closer to interpretability)								
1995/7/8	[1]	Classification	MOGA/1^{st}	2	√	√	-	-
2001	[2]	Classification	MOGA/1^{st}	2	-	√	√	-
TAKAGI-SUGENO MODELS (closer to accuracy)								
2001	[8]	Regression	Specific/1^{st}	3	-	-	√	√
2005	[9]	Regression	MOGA*/1^{st}	5	√	√	√	√
2005	[10]	Regression	NSGA-II*/2^{nd}	5	√	√	√	√

RS = Rule Selection, **FS** = Feature Selection, **RL** = Rule Learning, **LP** = Learning/Tuning of parameters.
* based on that algorithm

Our first aim was to study the applicability of MOEAs for Rule Selection and Tuning of linguistic models in regression problems

STUDY ON THE APPLICATION OF MOEAs FOR RULE SELECTION AND TUNING OF FUZZY MODELS

PARETO FRONTIER TYPE



Taking into account that a tuning is also performed:

- **Not necessary or bad rules could have been included in the initial rule base**
- **Not so important rules, that could be removed without greatly decrease the accuracy could have been also included**
- **Finally, some essential rules should be maintained**

This pareto front could be discontinuous

STUDY ON THE APPLICATION OF MOEAs FOR RULE SELECTION AND TUNING OF FUZZY MODELS

- **However some problems arise when combining rule selection with parameter tuning in a multiobjective framework:**
  - **There exist a lot of different subsets of rules with more or less the same number of rules but representing really different or alternative tuning possibilities**
  - **It is easier to decrease the number of rules than to reduce the system error. This provokes a faster tuning of the simplest solutions**
  - **The obtained parameters (in general) tends to be optimal for these premature solutions**
  - **It is difficult to obtain very accurate solutions by favoring the crossing of solutions with very different rule configurations**

IT IS PRACTICALLY IMPOSSIBLE TO OBTAIN THE COMPLETE OPTIMAL PARETO

STUDY ON THE APPLICATION OF MOEAs FOR RULE SELECTION AND TUNING OF FUZZY MODELS

- **To perform the study we have applied six different approaches based on the two most known and successful MOEAs:**
  - **Application of *SPEA2* and *NSGA-II***
  - **Two vers** **I-II_A and *NSGA-II_U***
  - **Two exte** **and *$SPEA2_{Acc2}$***
- **Two objective** **ber of Rules**
- **Proper operat**
- **The determi** **ze becomes an important issue. Specially in the case of NSGA-II**

Method	Description
WM	Wang & Mendel algorithm
T	Tuning of Parameters
S	Rule Selection
TS	Tuning & Selection
Application of standard MOEAs for general use	
TS-SPEA2	Tuning & Selection by SPEA2
TS-NSGA-II	Tuning & Selection by NSGA-II
TS-NSGA-II_A	Tuning & Selection by NSGA-II_{angle}
TS-NSGA-II_U	Tuning & Selection by NSGA-$II_{utility}$
Extended MOEAs for specific application	
TS-$SPEA2_{Acc}$	Accuracy-Oriented SPEA2
TS-$SPEA2_{Acc^2}$	Extension of $SPEA2_{Acc}$

- **In the next these algorithms are described**

ALG1: SPEA2 BASED ALGORITHM

Scheme of the algorithm of E. Zitzler

1. Generate an initial population P_0 and create the empty external population $P^e_0 = \emptyset$
2. Calculate fitness values of individuals in P_t and P^e_t
3. Copy all non-dominated individuals in P_t U P^e_t to P^e_{t+1}. If $|P^e_{t+1}| > N^e$ apply truncation operator. If $|P^e_{t+1}| < N^e$ fill with dominated in P_t U P^e_t
4. If $t \geq T$, return A and stop
5. Perform binary tournament selection with replacement on P^e_{t+1} in order to fill the mating pool
6. Apply recombination and mutation operators to the mating pool and set P_{t+1} to the resulting population. Go to step 2 with $t = t + 1$

ALG1: SPEA2 BASED ALGORITHM

Coding Scheme and Initial Population

- **A double coding scheme ($C^p = C^p_S + C^p_T$)**

  Rule Selection C_S (Binary-Coding):

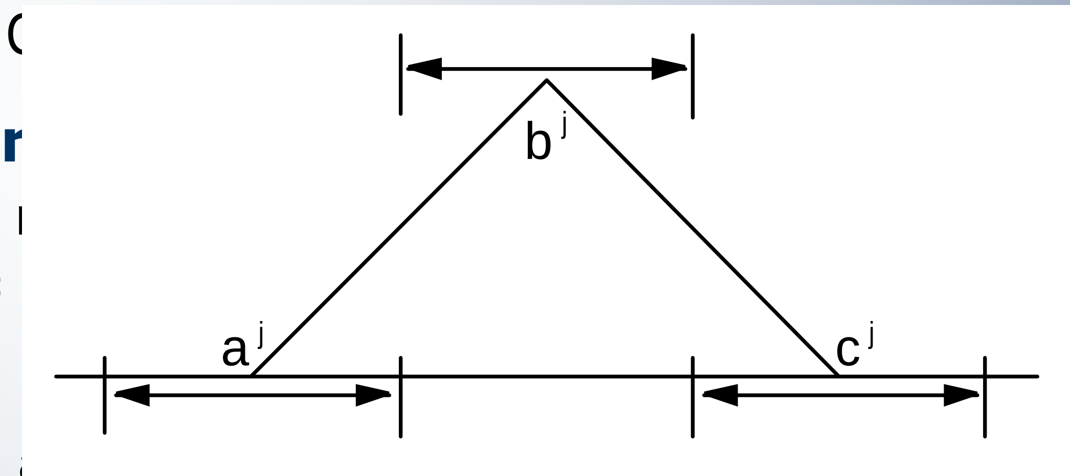
  ***m**:* number of initial rules

  C^p_S = (C_{s1}, ..., C

  Tuning of parameter

  ***n**:* variables; ***m**^i*:

  C^P_T = $C_1 C_2 ... C_n$ with C_i =



- **Initial Population**

Part C_T The initial DB is included

vals

Rules will be progressively removed in order to explore from the probably most accure solutions to the least ones

$[I^l_{c^j}, I^r_{c^j}]$ ... $^j - (c^j - b^j)/2, c^j + (c^j - b^j)/2]$

Part C_S All genes take value 1 in all the individuals of the initial population

ALG1: SPEA2 BASED ALGORITHM

Crossover and Mutation Operators

- **Crossover operator**

In the C_T part, the BLX-0.5 crossover is used.

In the C_S part, the HUX crossover is used.

Four offspring are generated by combining the two from the C_S part with the two from the C_T part (the best two replace to their parent).

- **Mutation operator (with probability P_m)**

In the C_S part it changes a gene value at random

In the C_T part it changes a gene value at random

ALG 2: NSGA-II BASED ALGORITHM

Scheme of the algorithm of K. Deb

1. A combined population R_t is formed with the initial parent population P_t and offspring population Q_t (initially empty).

2. Generate all non-dominated fronts $F = (F_1, F_2, \ldots)$ of R_t.

This algorithm uses the same coding scheme, population initialization and genetic operators described for the previous algorithm

7. Check the next front for inclusion.

8. Sort in descending order using crowded-comparison operator.

9. Choose the first $(N - |P_{t+1}|)$ elements of F_i.

10. Use selection, crossover and mutation to create a new population Q_{t+1}.

11. Increment the generation counter.

NSGA-II FOR FINDING KNEES

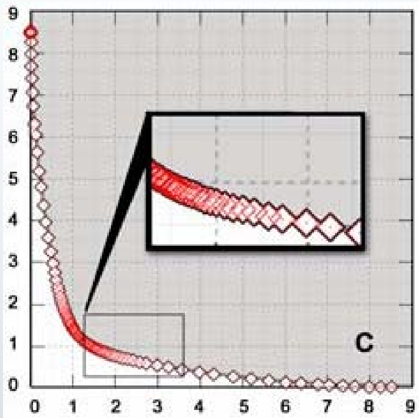
J. Branke, K. Deb, H. Dierolf, and M. Osswald, “Finding Knees in Multi-objective Optimization,” Proc. Parallel Problem Solving from Nature Conf. - PPSN VIII, LNCS 3242, (Birmingham, UK, 2004) 722–731.

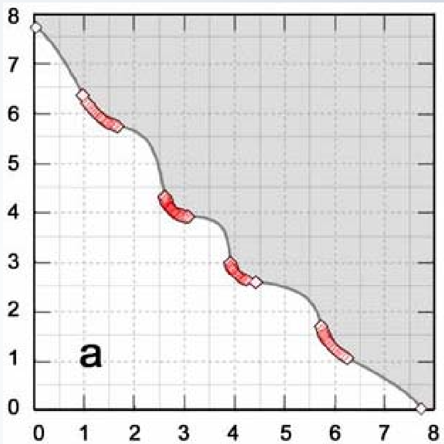
- **A variation of NSGAII in order to find knees in the Pareto front by replacing the crowding measure by either an angle-based measure or an utility-based measure**

Two different approaches

Angle Based Approach

Utility Based Approach



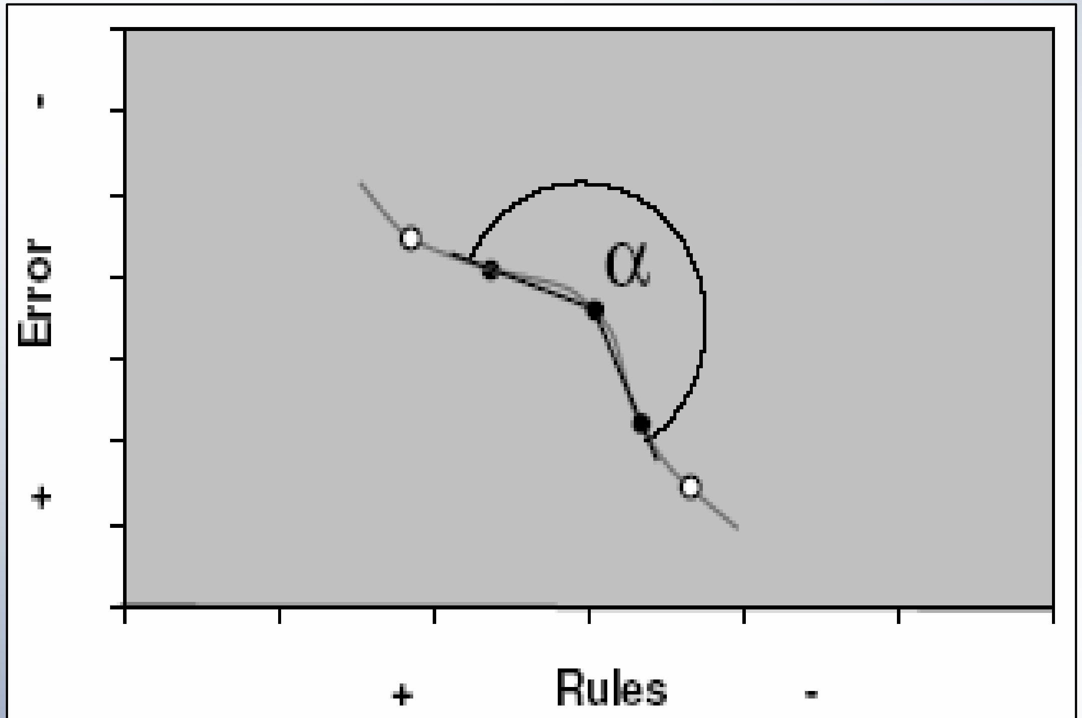


- **In our case, a knee could represent the best compromise between accuracy and number of rules.**

FIRST APPROACH
ALG 3: NSGA-II WITH ANGLE MEASURE (NSGA-IIA)

- **The angle between an individual and its two neighbors can be regarded as an indication of whether this individual is at a knee or not.**
- **The same coding and operators proposed for SPEA2 and NSGA-II are used**

Individuals with a larger angle are prefered



SECOND APPROACH
ALG 4: NSGA-II WITH UTILITY MEASURE (NSGA-IIu)

- **The marginal utility U'(C, λ') of an individual is the additional cost a decision maker would have to accept if that particular individual would not be available and he/she would have to settle for the second best**
- **The same coding and operators proposed for SPEA2 and NSGA-II are used**

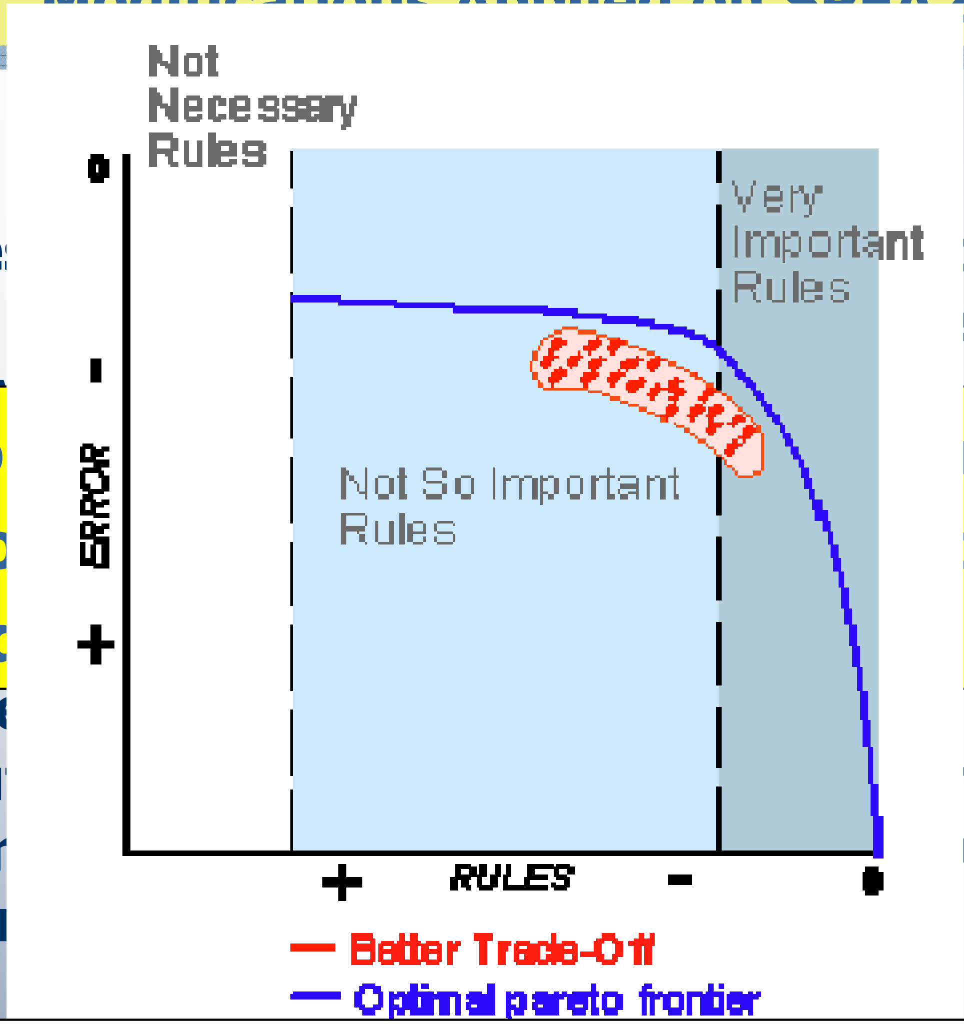
As greater is marginal utility of an individual nearer is of a knee

$$U'(C^i, \lambda') = \begin{cases} \min_{j \neq i} U(C^j, \lambda') - U(C^i, \lambda') & : i = \arg\min U(C^j, \lambda') \\ 0 & : otherwise \end{cases}$$

$$U(C, \lambda) = \lambda f_1(C) + (1 - \lambda) f_2(C)$$

ALG 5: ACCURACY-ORIENTED SPEA2 (SPEA2$_{Acc}$)

Modifications Applied on SPEA2



- A res ... tly at the mid of the ... most accurate indi... in the external ... ls ... of

This algo ... cheme, populatio ... perators described

- In e... the number of solu... considered to form... ely reduced, by focu... accuracy.

R. Alcalá, M.J. Gacto, F. Herrera, J. Alcalá-Fdez, A Multi-objective Genetic Algorithm for Tuning and Rule Selection to Obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15:5 (2007) 539-557*

ALG 6: Extension of $SPEA2_{Acc}$ ($SPEA2_{Acc2}$)

A New Crossover Operator for the Rule Part

- **Objective:** to improve the search with a more intelligent operator replacing the HUX crossover in $SPEA2_{ACC}$
- Once BLX is applied a normalized euclidean distance is calculated between the centric point of the MFs used by each rule of the offpring and each parent
- The closer parent determines if this rule is selected or not for this offpring

EXPERIMENTS

Two different problems

Estimating the maintenance costs of the medium voltage electrical network in a town

Four variables

Predicting the age of an Abalone specimen

Eight variables

THE ELECTRICAL PROBLEM

Problem Description

Estimating the maintenance costs of the medium voltage electrical network in a town

- Four input variables: total length of the streets, total area, edified area, supplied energy.
- One output variable: maintenance costs of the medium voltage electrical network.
- Sample data with 1,059 towns
- 5 Data partitions (80% - 20%) are considered
- Initial linguistic partitions were comprised by five linguistic terms (strong fuzzy partitions) for each variable.

EXPERIMENTS

Studied Methods

Method	Description	Pop. size
WM	Wang & Mendel algorithm	-
T	Tuning of Parameters	61
S	Rule Selection	61
TS	Tuning & Selection	61
TS-SPEA2	Tuning & Selection by SPEA2	200/61
TS-NSGA-II	Tuning & Selection by NSGA-II	61
TS-NSGA-II$_A$	Tuning & Selection by NSGA-II with angle-measure	61
TS-NSGA-II$_U$	Tuning & Selection by NSGA-II with utility-measure	61
TS-SPEA2$_{Acc}$	Accuracy-Oriented SPEA2	200/61
TS-SPEA2$_{Acc2}$	Extension of SPEA2$_{Acc}$	200/61

Different population sizes were probed for this MOEAs showing better results when the population used for parent selection has similar sizes than those considered by single objective oriented algorithms.

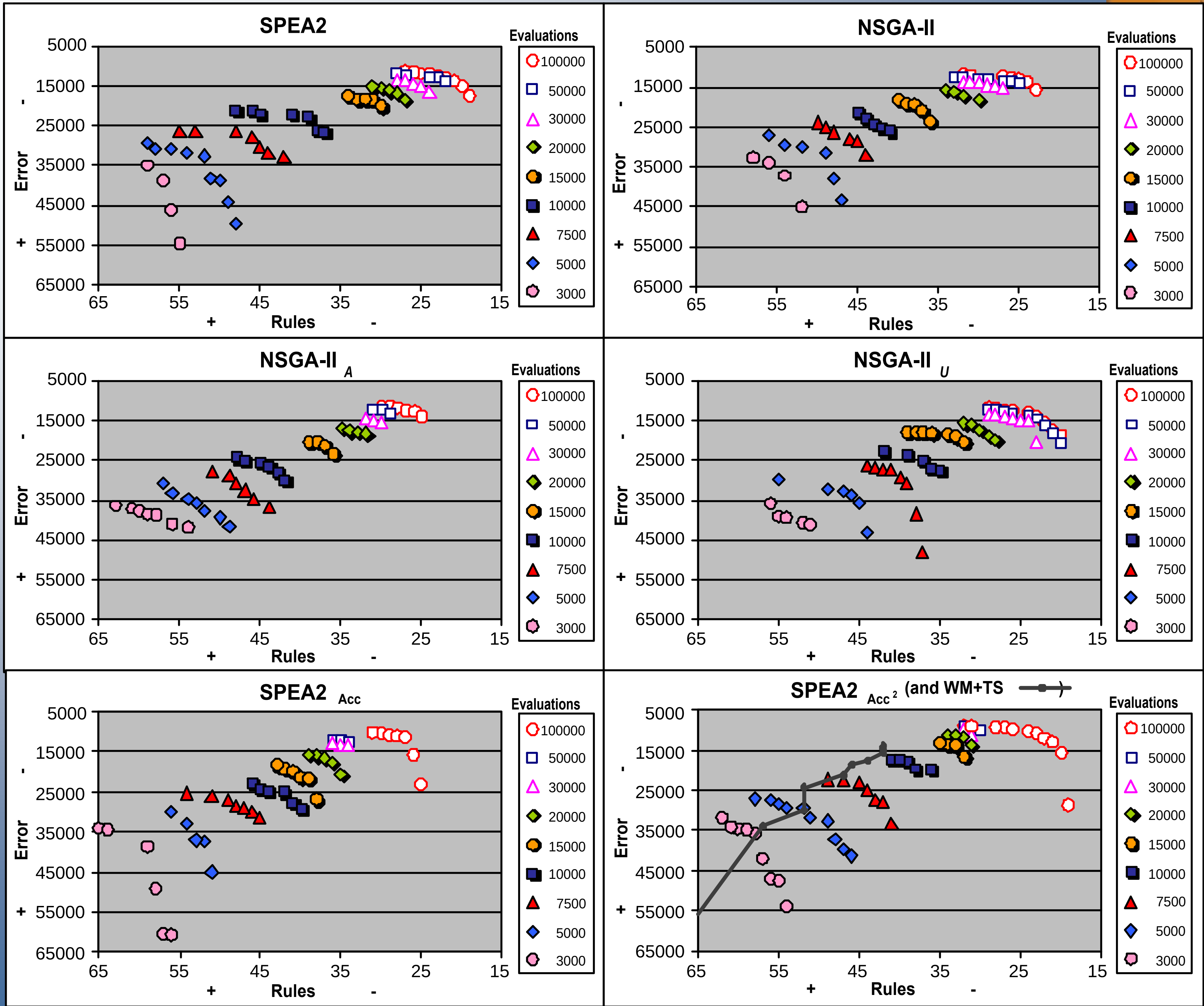
RESULTS IN THE ELECTRICAL PROBLEM

Averaged results of a total of 30 runs

Method	#R	MSE_{tra}	σ_{tra}	t	MSE_{tst}	σ_{tst}	t
100,000 evaluations							
WM	65.0	57605	2841	**+**	57934	4733	+
T	65.0	17020	1893	+	21027	4225	+
S	40.9	41158	1167	+	42988	4441	+
TS	41.3	13387	1153	+	17784	3344	+
TS-SPEA2	28.9	11630	1283	**+**	15387	3108	+
TS-NSGA-II	31.4	11826	1354	+	16047	4070	+
TS-NSGA-II$_A$	29.7	11798	1615	+	16156	4091	+
TS-NSGA-II$_U$	30.7	11954	1768	+	15879	4866	+
TS-SPEA2$_{Acc}$	32.3	10714	1392	=	14252	3181	=
TS-SPEA2$_{Acc2}$	29.8	10325	1121	*	13935	2759	*

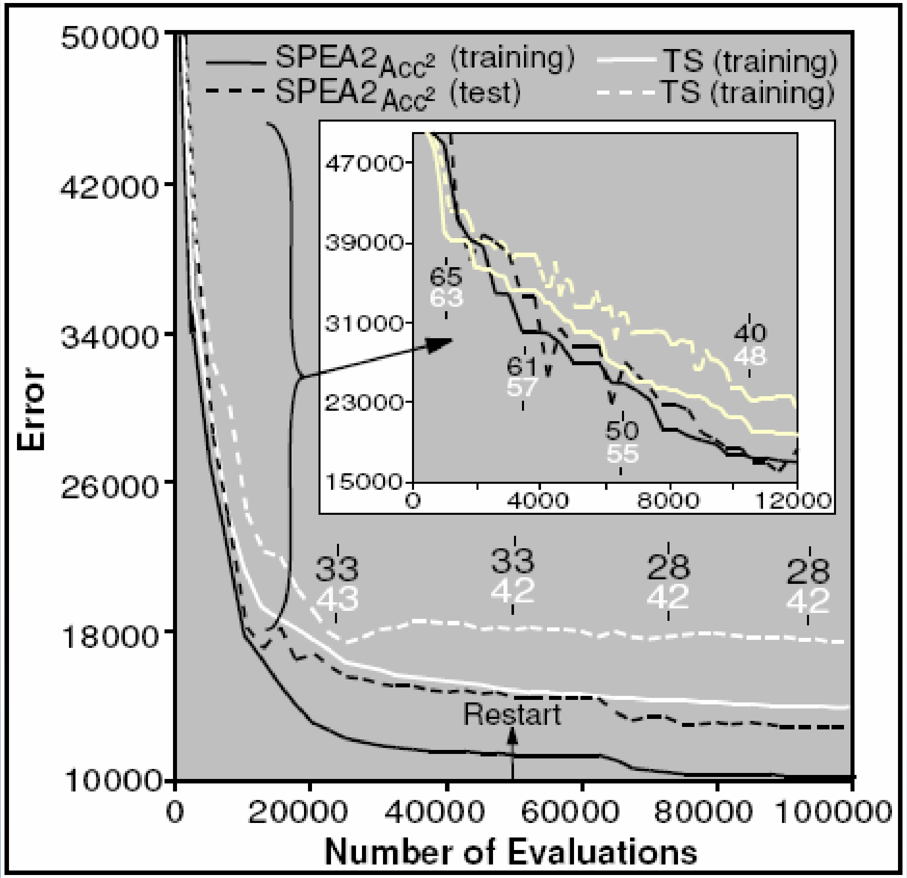
The most accurate solution in each Pareto has been considered to compute these values for each MOEA

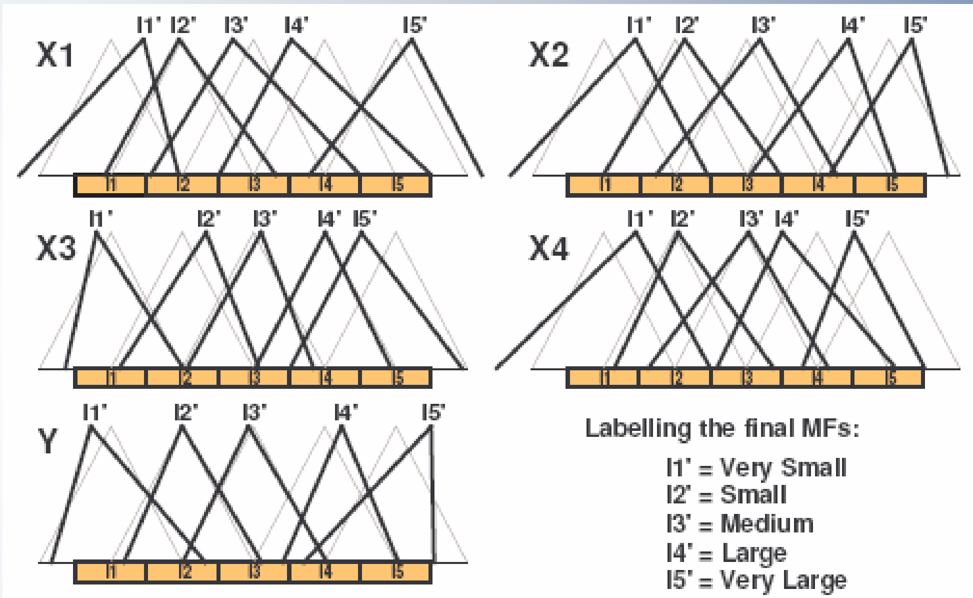
5 Data partitions 80% - 20%
6 Runs per partition
A total of 30 Runs
Test t-student α = 0.05


SPEA2
NSGA-II
NSGA-II A
NSGA-II U
SPEA2 Acc
SPEA2 Acc 2 (and WM+TS
Evaluations
100000
50000
30000
20000
15000
10000
7500
5000
3000
Error
-
+
5000
15000
25000
35000
45000
55000
65000
Rules
65
55
45
35
25
15

RESULTS IN THE ELECTRICAL PROBLEM

Convergence and an example model





Labelling the final MFs:

l1' = Very Small
l2' = Small
l3' = Medium
l4' = Large
l5' = Very Large

#R: 28 *MSE-tra: 8232* *MSE-tst: 14670*

X1	X2	X3	X4	Y
l1'	l1'	l1'	l1'	l1'
l1'	l1'	l1'	l2'	l2'
l2'	l1'	l1'	l1'	l1'
l2'	l1'	l1'	l2'	l2'
l2'	l1'	l2'	l2'	l2'
l2'	l2'	l2'	l1'	l2'
l2'	l3'	l3'	l1'	l3'
l3'	l2'	l1'	l1'	l1'
l3'	l2'	l1'	l2'	l2'

X1	X2	X3	X4	Y
l3'	l2'	l1'	l3'	l2'
l3'	l2'	l2'	l3'	l3'
l3'	l3'	l2'	l2'	l2'
l3'	l3'	l3'	l2'	l3'
l3'	l4'	l3'	l3'	l3'
l4'	l2'	l2'	l2'	l2'
l4'	l3'	l2'	l1'	l2'
l4'	l3'	l2'	l3'	l3'
l4'	l3'	l2'	l4'	l3'

X1	X2	X3	X4	Y
l4'	l3'	l3'	l2'	l3'
l4'	l3'	l3'	l3'	l4'
l4'	l4'	l3'	l2'	l3'
l4'	l4'	l4'	l2'	l4'
l4'	l4'	l4'	l4'	l5'
l4'	l5'	l4'	l2'	l3'
l4'	l5'	l5'	l3'	l5'
l5'	l2'	l2'	l5'	l4'
l5'	l2'	l3'	l2'	l3'
l5'	l4'	l3'	l5'	l5'

THE ABALONE PROBLEM

Problem Description

Predicting the age of an Abalone specimen
(a type of shellfish)

Available from the UCI Machine Learning Repository
(www.ics.uci.edu/mlearn/MLRepository.html)

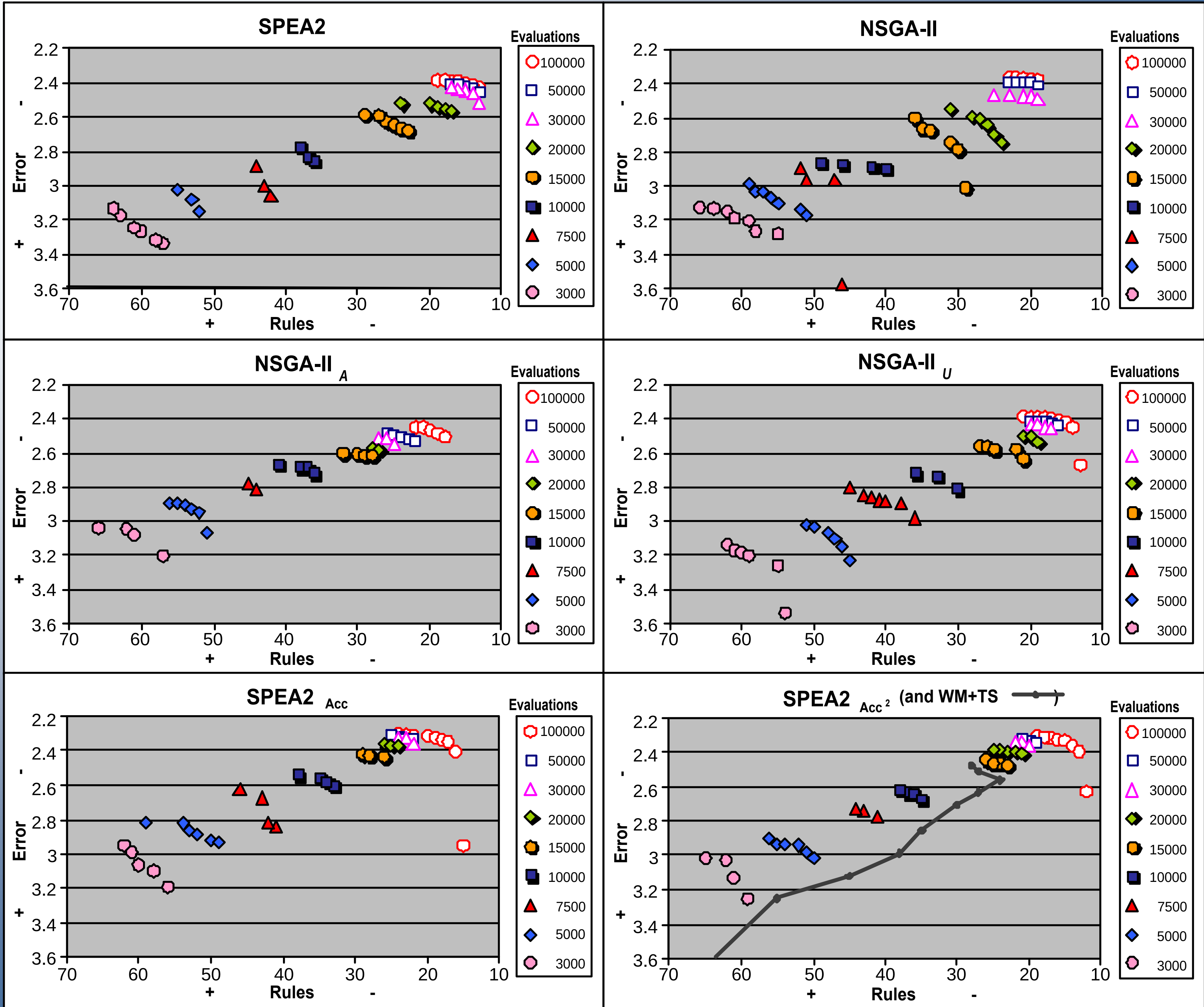
- Eight input variables (seven continuous and one nominal): sex (nominal), length, diameter, height, whole weight, shucked weight, viscera weight, and shell weight.
- One output variable: Number of rings of the Abalone (proxy for age)
- Sample data with 4177
- 5 Data partitions (80% - 20%) are considered
- Initial linguistic partitions were comprised by three linguistic terms (strong fuzzy partitions) for each variable.

RESULTS IN THE ABALONE PROBLEM

Averaged results of a total of 30 runs

Method	#R	MSE_{tra}	σ_{tra}	t	MSE_{tst}	σ_{tst}	t
			100,000 evaluations				
WM	68.2	8.407	0.443	+	8.422	0.545	+
WM+T	68.2	2.688	0.063	+	2.770	0.242	+
WM+S	18.0	4.825	1.078	+	4.795	1.165	+
WM+TS	28.4	2.473	0.097	+	2.582	0.290	=
TS-SPEA2	20.0	2.383	0.078	=	2.518	0.246	=
TS-NSGA-II	22.4	2.398	0.084	=	2.526	0.242	=
TS-NSGA-II$_A$	22.1	2.404	0.098	=	2.535	0.265	=
TS-NSGA-II$_U$	21.8	2.407	0.082	=	2.520	0.237	=
TS-SPEA2$_{Acc}$	22.2	2.368	0.085	*	2.511	0.263	*
TS-SPEA2$_{Acc2}$	18.6	2.372	0.075	=	2.517	0.230	=

5 Data partitions 80% - 20%
6 Runs per partition
A total of 30 Runs
Test t-student α = 0.05


SPEA2
NSGA-II
NSGA-II A
NSGA-II U
SPEA2 Acc
SPEA2 Acc 2 (and WM+TS)
Evaluations
100000
50000
30000
20000
15000
10000
7500
5000
3000
Error
Rules
2.2
2.4
2.6
2.8
3
3.2
3.4
3.6
70
60
50
40
30
20
10
+
-

STUDY ON THE APPLICATION OF MOEAs FOR TUNING OF FUZZY MODELS

Some Conclusions

- The **models obtained by MOEAs presented a better trade-off** than those obtained by only considering performance measures

- These results show that **the use of experience based knowledge** in the MOEAs design process **can** significantly **improve the search ability** of these algorithms.

RESULTS IN THE ELECTRICAL PROBLEM

New hopeful solutions can arise

Method	#R	MSE_{tra}	σ_{tra}	t	MSE_{tst}	σ_{tst}	t
100,000 evaluations							
WM	65.0	57605	2841	+	57934	4733	+
T	65.0	17020	1893	+	21027	4225	+
S	40.9	41158	1167	+	42988	4441	+
TS	41.3	13387	1153	+	17784	3344	+
TS-SPEA2	28.9	11630	1283	+	15387	3108	+
TS-NSGA-II	31.4	11826	1354	+	16047	4070	+
TS-NSGA-II_{A}	29.7	11798	1615	+	16156	4091	+
TS-NSGA-II_{U}	30.7	11954	1768	+	15879	4866	+
TS-SPEA2_{Acc}	32.3	10714	1392	+	14252	3181	+
TS-SPEA2_{Acc2}	29.8	10325	1121	+	13935	2759	+
$\text{TS-SPEA2}_{E/E}$	29.8	9688	942		12671	1796	

Still not published results

FURTHER DIRECTIONS

To find appropriate interpretability measures in order to preserve the meaning associated to the membership functions

Inherent problems:

- **This is a very difficult task since the interpretability is a subjective concept**
- **MOEAs do not work so well when more than three objectives are considered**

I Workshop on Knowledge Extraction based on Evolutionary Learning

Introduction to Multi-objective Learning: Some results on the use of MOEAs for tuning FRBSs parameters

Rafael Alcalá

Dept. of Computer Science and Artificial Intelligence
University of Granada

18071 – Spain

http://www.ugr.es/~ralcala/
alcala@decsai.ugr.es

In collaboration with

María José Gacto, Jesús Alcalá-Fdez and Francisco Herrera

Granada, May 16, 2008

University of Granada

Research Group on Soft Computing and Intelligent Information Systems
http://sci2s.ugr.es