International Workshop
on Mathematical Models
and its Applications

DEVELOPMENT OF ADAPTIVE GENETIC ALGORITHMS FOR NEURAL NETWORK MODELS MULTICRITERIA DESIGN

Christina Brester

Motivation

- Conventional GA's performance depends on the settings of genetic operators:

At the Workshop on Evolutionary Algorithms, Minnesota, October 21 – 25, 1996,

Lawrence "David" Davis (as the most recognized practitioner of Evolutionary Algorithms at that time) said that all theoretical results in the area of Evolutionary Algorithms were of no use to him. If a theoretical result indicated that the best value of some parameter was such-and-such, he would **<u>never use</u>** the recommended value in any real-world implementation of an evolutionary algorithm!

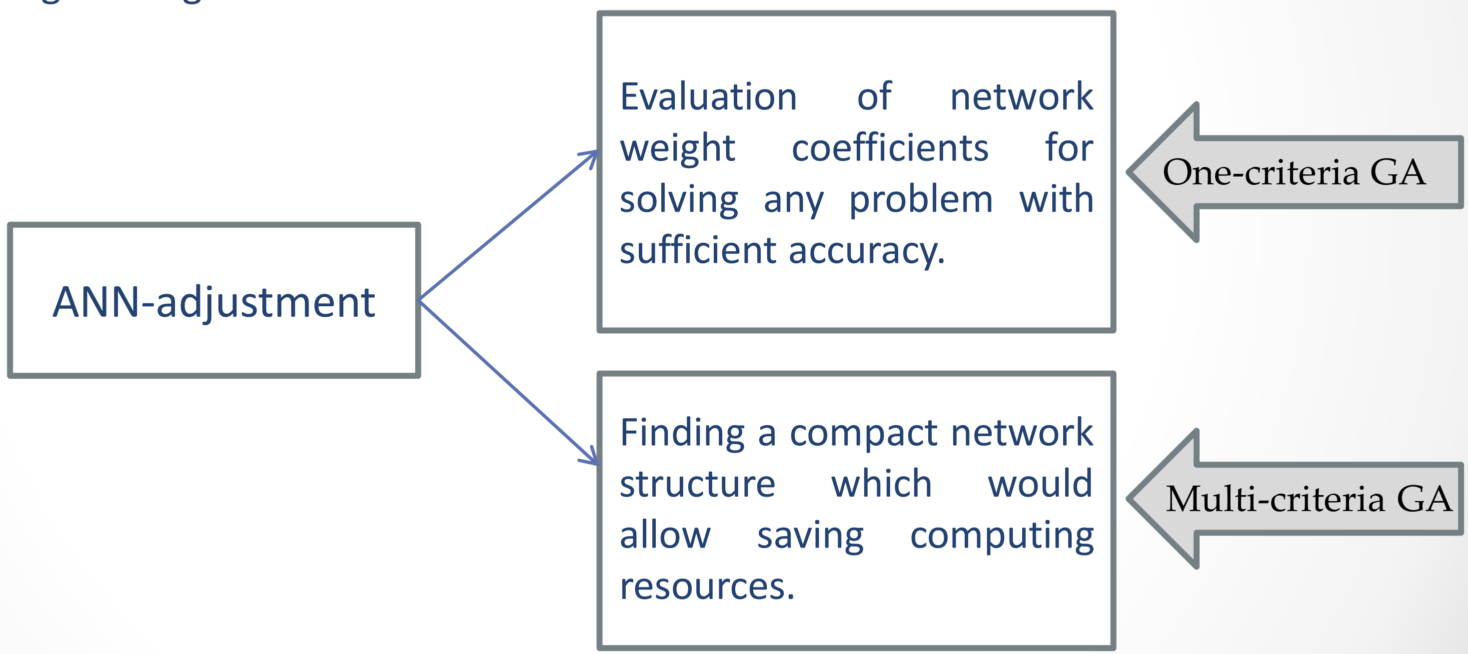
(Zbigniew, Michalewicz "Quo Vadis, Evolutionary Computation? On a Growing Gap between Theory and Practice")

✓ Solution:

Self-adaptive algorithms

Motivation

- Most of real problems are constrained and multi-criterion:
  - ✓ Development of multi-objective self-adaptive GA
- Artificial Neural Networks (ANN) are exploited in different applications: data analysis systems, speech or images recognition and so on.



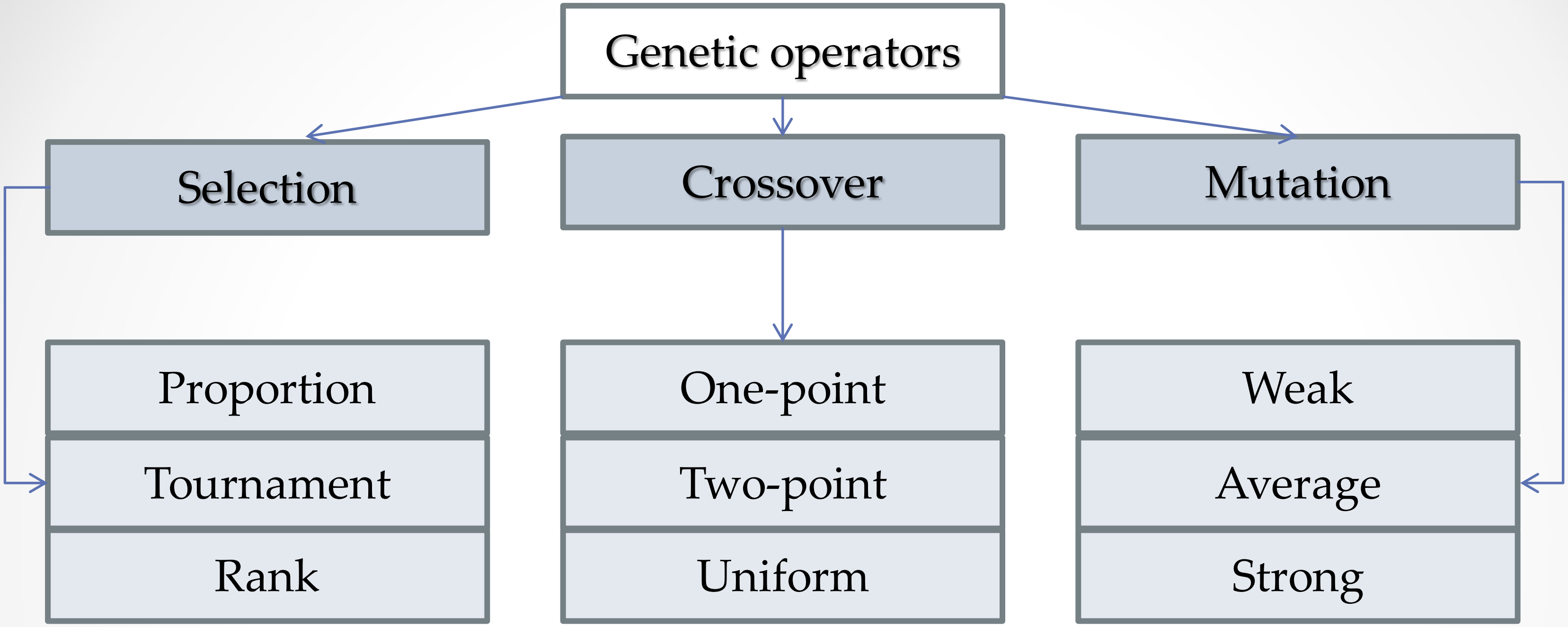
✓ A combination of self-adaptive GAs

Research Plan

Realization of this approach was implemented through three key steps:

- Design, realization and efficiency investigation of one-criteria self-adaptive GA;
- Design, realization and efficiency investigation of multi-objective self-adaptive GA;
- Design, realization and efficiency investigation of an approach for ANN-adjustment via a combination of one- and multi-criteria self-adaptive GAs.

Adaptive one-criteria genetic algorithm



❖ Key concept – "application probabilities"

- 1^{st} generation "equal probabilities": $q_i^k = 1/n^k$, where q_i^k is the application probability for the *i*-th variant of the *k*-th operator, n^k is the number of different variants of a certain genetic operator, $i,k=\overline{1,3}$.
- *T-th* generation "probabilities recalculation":

$$q_i^k = \frac{0.2}{n^k} + 0.8\cdot\frac{ratio_i^k}{scale^k},\quad scale^k = \sum_i ratio_i^k$$

, where «$ratio_i^k$» is a fitness sum of individuals generated with the *i*-th variant of the *k*-th operator.

Adaptive one-criteria genetic algorithm

Testing results of an adaptive single-objective GA

№	Test function	The number of cases when an adaptive GA lost to a conventional GA
1	Rastrigin function	9
2	Rosenbrock function	4
3	Katkovnik function	3
4	Griewank function	4
5	Multiplicative potential function	0
6	Additive potential function	7
7	Rastrigin function (ravine function with rotated axis)	1
8	«Foxholes»	8

The efficiency of an adaptive genetic algorithm is not lower than the efficiency of an «average» GA and is comparable with «best» GA in many cases.

Multi-objective GA

Strength Pareto Evolutionary Algorithm (SPEA)

- Adaptive mutation: $p_m = \frac{1}{240} + \frac{0.11375}{2^t}$,

  where *t* is the current generation number.

- Self-configurable crossover operator:

  The main idea: *the population is divided into parts and each part is generated with a certain type of crossover.*

❖ *Conceptual apparatus*

«Fitness» is proportional to the number of non-dominated individuals generated with a certain type of crossover and stored in the outer set.

«Penalty» is a parameter which denotes the amount of resources which the genetic operator with lower «fitness» gives the genetic operator with higher «fitness».

Decreasing of resources must be limited with a parameter named «*social card*».

Multi-objective GA

- An example of test problems:

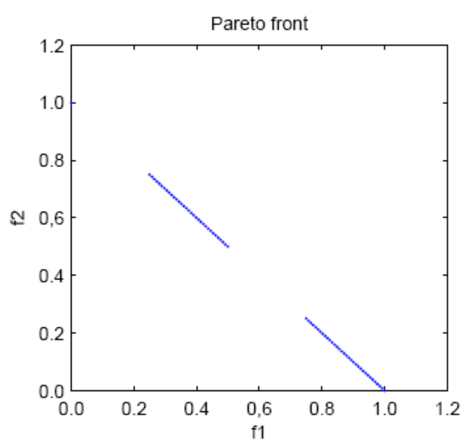
The two objectives to be minimized:

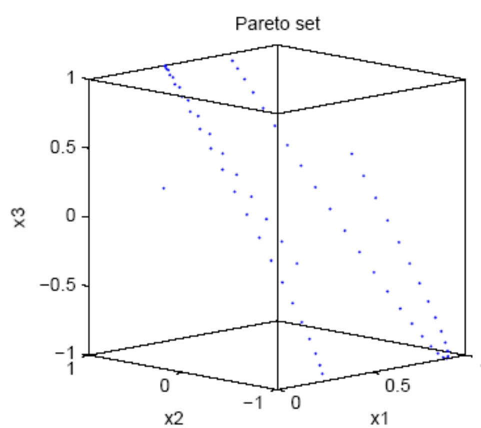
$$f_1 = x_1 + \max\{0, 2(\frac{1}{2N} + \varepsilon)\sin(2N\pi x_1)\} + \frac{2}{|J_1|}(4\sum_{j\in J_1} y_j^2 - 2\prod_{j\in J_1}\cos(\frac{20y_j\pi}{\sqrt{j}}) + 2)$$

$$f_2 = 1 - x_1 + \max\{0, 2(\frac{1}{2N} + \varepsilon)\sin(2N\pi x_1)\} + \frac{2}{|J_2|}(4\sum_{j\in J_2} y_j^2 - 2\prod_{j\in J_2}\cos(\frac{20y_j\pi}{\sqrt{j}}) + 2)$$

where $J_1 = \{j|j \text{ is odd and } 2 \le j \le n\}$, $J_2 = \{j|j \text{ is even and } 2 \le j \le n\}$, and

$$y_j = x_j - \sin(6\pi x_1 + \frac{j\pi}{n}), j = 2, \ldots, n.$$





Multi-objective GA

The adaptive SPEA's testing results

Number of a test function	A value of IGD-metrics	The number of cases when the adaptive SPEA's lost to a conventional SPEA
1	0,10579	2
2	0,04160	3
3	0,0049	1
4	0,04493	2
5	0,34105	2
6	0,0088	2
7	0,11738	1

Self-adaptation is an alternative to random choice of genetic operators or multiple runs of MOP-GA for each variant of settings.

❖ **Comments:**

The results of algorithms work were estimated with the *IGD*-metrics:

$$IGD(A,P^*) = \frac{\sum_{v \in P^*} d(v,A)}{\left|P^*\right|},$$

where P^* is a set of points uniformly distributed along the PF, v is the point of P^*, $d(v, A)$ is the minimum Euclidean distance between v and A, A is the approximate set of the PF.

Application of adaptive GAs for the ANN-design

- ANN-structures representation (MLP):

$$\underbrace{\boxed{1}\boxed{0}\boxed{0}\boxed{\ldots}}\boxed{1}\boxed{0}\boxed{0}$$

The index of an activation function in the finite set F;
$F = \{$sinus, sigmoid, Heaviside function, linear function, hyperbolic tangent,…, triangular function$\}$.

- Weights for every ANN-structure also have a binary representation: it is a sequence of binary codes which correspond to certain weights.

➢ Benchmark problems

«German credits» and «Australian credits»

(all samples were borrowed from the machine learning repository)

Features (24 and 14): age, gender, marital status, credit history records, job, etc.

Sample sizes: 1000 and 690 examples respectively.

Application of adaptive GAs for the ANN-design

Approximations of the Pareto's front for benchmark problems

Australian Credits			German Credits		
The number of neurons		Relative error	The number of neurons		Relative error
Layer 1	Layer 2		Layer 1	Layer 2	
8	7	0,093525	8	4	0,2200
7	6	0,107914	7	2	0,2250
6	6	0,115108	5	3	0,2350
4	6	0,129496	6	1	0,2550
			5	1	0,3000

Results of alternative approaches

Method	Australian Credits (relative error)	German Credits (relative error)	Method	Australian Credits (relative error)	German Credits (relative error)
SCGP	0,0978	0,205	Bayesian approach	0,153	0,321
MGP	0,1015	0,2125	Boosting	0,24	0,3
2SGP	0,0973	0,1985	Bagging	0,153	0,316
GP	0,1111	0,2166	RSM	0,148	0,323
Fuzzy classifier	0,109	0,206	CCEL	0,134	0,254
C4.5	0,1014	0,2227	CART	0,1256	0,2435
LR	0,1304	0,2163	MLP	0,1014	0,2382

Conclusion

- In this research two self-adaptive GAs were developed:

  one for the one-criterion optimization and another one for the multi-criterion optimization.

- There is also GAs-cooperation for automated design of ANN-based classifiers:

  classifiers designed in this way are enough accurate and also have simple structures.

- Directions of the future research can be divided into three groups:

  The first of them is the improvement of GAs self-adaptation abilities.

  The second direction is the modification of ANN-based models automated design through including different ANNs types.

  And the third is an expansion of adaptive GAs application areas, e.g., automated design of fuzzy systems, decision trees, etc.