

LINKAGE LEARNING VIA PROBABILISTIC MODELING IN THE ECGA

Georges Harik

IlliGAL Report No. 99010
January 1999

PREVIOUS WORKS

- Linkage learning in GAs is the identification of building blocks to be conserved under crossover.
- Past approaches have focused on evolving problem representation along with its solution. But this is a hard work. (messy GA, fast messy GA)
- Recently, a number of algorithms replace the GA's population and crossover operator with a probabilistic representation and generation method. (compact GA, MIMIC)

TWO MAIN ASSERTIONS

- Learning a “good” probability distribution is equivalent to learning linkage.
- One “good” distribution can be found by searching for a jointly small representation of two components:
 - ① the compressed representation of the population under the given distribution
 - ② the distribution’s representation given the problem encoding

PARAMETERS OF GA

- Uniform crossover.
- No mutation.
- Block selection: the best fraction $\frac{1}{5}$ of the population is to be retained after selection.
- Stop criterion: the population converges.
- L: chromosome length.

IDEA OF cGA

- Assume the population size n approaches infinite, and we want to perform simple GA on this population.
- Randomly pick S individual. The winner replace other $S-1$ loser.
- Perform n selection to generate n individuals.
- Repeat 2 and 3 until the population converges.

PROCEDURE OF cGA

- ① Initialize an L -dimensional probability vector $P[]$ to 0.5
 $P = \{0.5, 0.5, 0.5, \dots, 0.5\}$
- ② S solutions are generated by polling this vector
- ③ The gene positions of the fittest of these S solutions are rewarded
- ④ Repeat step 2 and 3 until the $P[]$ vector implies a single solution

EXAMPLE OF CGA

- ① Assume $L = 4, S = 4, E = 0.01$. Initialize $P = \{0.5, 0.5, 0.5, 0.5\}$. The fitness function is OneMax.
- ② Generate 2 chromosomes:
0111 fitness=3
1010 fitness=2
1000 fitness=1
0100 fitness=1
- ③ Since best chromosome is 0111,
 $P = \{0.5 - 0.02, 0.5 + 0.02, 0.5 + 0.02, 0.5 + 0.03\}$
- ④ Repeat until all values in P are zeroes or ones.

PROBABILISTIC OPTIMIZATION AND LINKAGE LEARNING

- ① GA's population can be interpreted as representing a probability distribution over the set of future solutions to be explored.
- ② In the sense, the role of crossover can be played by a more direct representation of the distribution itself, like what the cGA does.
- ③ Therefore, a good probability distribution is equivalent to linkage learning.

LINKAGE LEARNING AND DECEPTIVE PROBLEMS

- ① We define a “deceptive” version of OneMax problem. A 40 dimensional problem is composed of 10 little problems, each of which is 4 dimensional.
- ② The score of a little problem is the summation of 1s, but the score is 5 if the solution is 0000.
- ③ Compact GA will tend to gravitate towards a suboptimal solution 1111.
- ④ To learn linkage, order-1 probability model is not enough. The probability distribution needs to recognize that these four genes are related.

HIGH ORDER GAS

- ① MIMIC investigates pairwise inter-gene correlations and generating a distribution. Order-2 correlation is better than order-1.
- ② How about order-L?
- ③ What order is most appropriate for the question?
- ④ Given all other things are equal, simpler distributions are better than complex ones.

MINIMUM DESCRIPTION LENGTH MODELS

Mitchell, *Machine Learning*:

By reliance on Occam's Razor, good distributions are those under which the representation of the distribution using the current encoding, along with the representation of the population compressed under that distribution, is minimal.

IMPLEMENTS OF MDL MODELS

Model Complexity =

$$\log N \times \sum_{i=1}^{n_s} 2^{s_i}$$

Compressed Population Complexity = $N \sum_{i=1}^{n_s} E_i$

$$E_i = \sum_{j=1}^{2^{s_i}} -p_j \log_2 p_j$$

- E_i : the entropy of i th subset
- s_i : the size of i th subset
- N : population size

MARGINAL PRODUCT MODEL (MPM)

A marginal probability model partitions genes into several subsets.
Each subset has its own distribution.

Here is an example of MPM over four genes:

	[0,3]	[1]	[2]
00	0.5	0	0.6
01	0	1	0.4
10	0		
11	0.5		

COMBINE MPM AND MDL MODEL

Population: 1010, 1101, 0010, 1001

MPM 1						MPM 2					
[0,3]		[1]		[2]		[0,1]		[2]		[3]	
00	1	0	3	0	2	00	0	0	2	0	2
01	0	1	1	1	2	01	1	1	2	1	2
10	1					10	2				
11	2					11	1				

- Model complexities of both distributions are $\log_2 4(2^2 + 2^1 + 2^1) = 16$.
- The population complexity of second distribution is 3.5, and the population complexity of first distribution is 3.3133.

PROCEDURE OF ECGA

- ① Generate a random population of size N .
- ② Perform tournament selection with tournament size s .
- ③ Model the population using a greedy MPM search.
 - Initial: assume that all genes are independent.
 - Attempt to merge all pairs of subsets, choosing the best merging result as new probability distribution.
 - Repeat until there's no improvement.
- ④ Building-block-wise crossover.
- ⑤ Calculate fitnesses.

COMPARE SIMPLE GA AND ECGA (1/2)

- No mutation.
- Simple GA with uniform crossover.
- Both are tournament selection.
- Tournament size = 16.
- Objective function: the trap function.
 - 0000 \rightarrow 5
 - 0001 \rightarrow 1
 - 0010 \rightarrow 1
 - 0011 \rightarrow 2
 - \vdots
 - 1111 \rightarrow 4
- Chromosome length = 40.
- Fitness = the sum of 10 smaller problem.

COMPARE SIMPLE GA AND ECGA (2/2)

Population size	Subfunctions Solved	Objective Evaluations
100	3.9	740
500	5.2	5100
1000	6.1	15600
5000	6.8	100000
10000	7.3	248000
20000	8.0	614000
50000	7.9	1560000
100000	8.8	3790000

Table 2: Simple GA complexity on deceptive subproblems.

Population size	Subfunctions Solved	Objective Evaluations
100	4.0	750
200	5.2	1460
300	7.1	2610
500	9.3	4000
600	9.9	5040
1000	10.0	7300

Table 3: ECGA complexity on deceptive subproblems.

FUTURE WORKS (1/2)

- The ECGA is simple to parallelize, by replacing the migration step of standard parallel GAs by one of probability model exchange.
- Reduce the computational cost of MPM model. For example, the MPM search could be biased somewhat to the encoder's original linkage specification.
- Another approach to reducing the complexity of the MPM model search is to sample from the population.
- Extend ECGA to more complex structures such as permutations, or programs.

FUTURE WORKS (2/2)

- Although BOA is more complex, the function evaluations are less.
- Algorithmic complexity analysis.
- The role of selection is not quite clear.
- Use different GA operators.
- Extend ECGA to real-world problems.

INTEGER ECGA

- Some problems may occur if integers are represented as binary strings.
 - If the integer has lower bound and upper bound, there may exist a gap in mapping from an integer to a binary string.
 - Can ECGA handle two level building blocks?
 - Is bit-level crossover is suitable for integers?
 - An extreme example: nonexistent linkage
 $f(0) = 5, f(1) = 1, f(2) = 1, f(3) = 2, \dots, f(15) = 4$
- iECGA is not designed to defeat the ECGA, but to make up the ability of solving integer problems.



