

Sub-Structural Niching in Estimation of Distribution Algorithms

Kumara Sastry, Hussein A. Abbass, David E. Goldberg, D.D. Johnson

GECCO 2005

Introduction

- Estimation Distribution Algorithms (EDAs) is a class of competent GAs.
- EDAs replace the traditional variation operators with probability models and sample the model to generate new candidate solutions.
- To solve multimodal, hierarchical, dynamic, and multiobjective optimization problems, an efficient niching method is required.
- The purpose of a niching method is to maintain the diversity of population throughout the search.

Niching Methods

Niching methods allow EDAs to

- identify multiple optima when solving multimodal and multiobjective problems.
- identify the global optimum by deciding successfully between substructures.
- rapidly identify global solutions when changes occur in non-stationary problems.

The contribution of this paper

- Traditional niching methods usually maintain diversity at the level of individuals. They don't exploit the underlying working mechanism of EDAs and other linkage learning algorithms.
- In this paper, we propose a niching method that respect problem decomposition, and maintain diversity at sub-structure level in a stable manner.
- The proposed method consists of three components:
 - 1 Sub-structure identification.
 - 2 Sub-structure fitness estimation.
 - 3 Sub-structure niche preservation.
- The proposed method is compared with RTS (restricted tournament selection) that used in hBOA.

The idea of ECGA

- Define a deceptive sub-structure as following:

$$f(x_1x_2x_3x_4) = \begin{cases} 5, & \text{if } x_i = 0 \text{ for } i = 1, 2, 3, 4 \\ x_1 + x_2 + x_3 + x_4, & \text{otherwise} \end{cases}$$

- Define the objective function g :

$$g(x_1x_2 \dots x_n) = f(x_1x_2x_3x_4) + \dots + f(x_{n-3}x_{n-2}x_{n-1}x_n)$$

- Traditional GAs cannot find global optimal of $g(x)$, because GAs can't view four bits as one building block.
- Can we view the chromosome as sets of building blocks?

Probability Model

- If the chromosome contains four genes, [1,2][3][4] and [1,3][2][4] are both probability models. Which one is better?
- In the domain of information theory, minimum description length models assumes that smaller entropy is better.
- Model complexity:

$$C_m = \log_2(n) \sum_{i=1}^m (2^{k_i} - 1)$$

- Compress population complexity:

$$C_p = n \sum_{i=1}^m \sum_{j=1}^{2^{k_i}} -p_{ij} \log_2(p_{ij})$$

Building the Probability Model

- Assume the length of chromosome is four bits.
- Initial model is $[1][2][3][4]$. We perform a greedy algorithm on the merge of building blocks.
- Calculate the entropy of following models:
 $[1,2][3][4], [1,3][2][4], [1,4][2][3], [2,3][1][4], [2,4][1][3], [3,4][1][2]$
Find the model with minimum entropy, and continue merging building blocks.

The process of ECGA

- 1 Initialize
- 2 Evaluate
- 3 Selection
- 4 Model Estimation
- 5 Crossover
- 6 Replacement
- 7 repeat steps 2-6 until the termination criteria are met.

Idea of niching

- Traditional niching methods maintain diversity at the level of individuals.
- The critical steps for GA success is problem decomposition, but the decomposition is achieved indirectly.
- EDAs work by first decomposing the search problems into sub-structures and then creating new solutions by exchanging different sub-structures.
- It might be advantageous to maintain diversity at the level of building blocks.

Three Key Elements

- Sub-structure identification:
In this study, we use the models built by ECGA.
- Sub-structure fitness estimation:
We must decide which substructure to preserve and in what proportion.
- Sub-structure niche preservation:
To keep the diversity.

Fitness Estimation

- For example, for a four-bit problem, whose model is $[1,3][2][4]$, the schema are $0^*0^*, 0^*1^*, 1^*0^*, 1^*1^*, *1^{**}, *0^{**}, ***1, ***0$.
- For each sub-structure (schema) h , the fitness of h is defined as the difference between the average of individuals contains the schema and the average of all individuals.
- If there is no individual that contains the schema, the fitness of the schema is zero.

Niche Preservation

- After the fitness of each niche is decided, we need to decide on a methodology for preserving the sub-structures.
- Since no one method is better than the other, we use fitness-proportionate method.
- For each building block, there are some competing alleles like 00**, 01**, 10**, and 11**.

$$p(00 **) = \frac{f_s(00 **)}{f_s(00 **) + f_s(01 **) + f_s(10 **) + f_s(11 **)}$$

The process of ECGA

- 1 Initialize
- 2 Evaluate
- 3 Selection
- 4 Model Estimation: generate the probability model
- 5 Fitness Estimation: compute the fitness of building blocks
- 6 Crossover: choosing building blocks to mix
- 7 Replacement
- 8 repeat steps 2-7 until the termination criteria are met.

Test Functions: deceptive trap problems

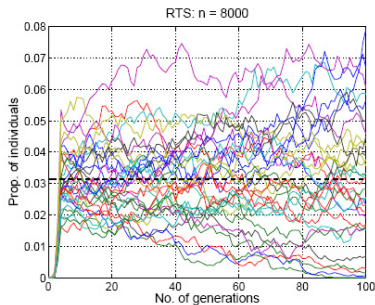
- A m-k deceptive trap problems is composed of m sub-problems, and each sub-problem is a k-bits deceptive problem.
- We use a modified m-4 trap problem where 0000 and 1111 have equal fitness. Therefore there are 2^m global optima with an identical fitness.
- Each k-bit trap is defined as follows:

$$\text{trap}_k(u) = \begin{cases} 1, & \text{if } u = k \\ 1, & \text{if } u = 0 \\ 0.75(1 - \frac{u}{k-1}), & \text{otherwise} \end{cases}$$

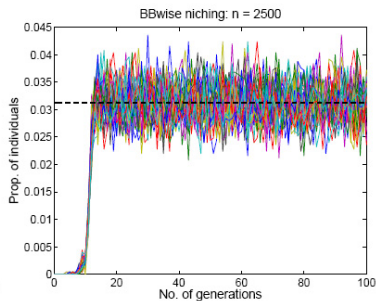
,where u is the number of 1s.

Result 1: the proportion of each optimum

The purpose of defining the test function is to test the ability of maintaining global optima, thus we watch the proportion of each global optimum. For a 5-4 deceptive problem, we use different colors to represent 32 global optimum in the figure.



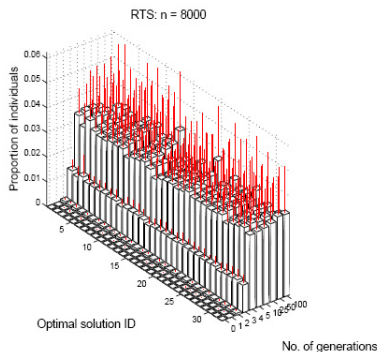
(a) RTS



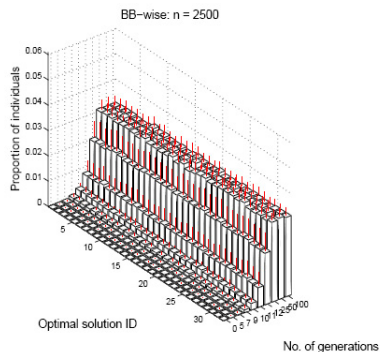
(b) Sub-structural niching

Result 2: the average market share

The figure shows the average market share of each global optimum (each has a ID from 1 to 32). Red line represents the standard deviation.



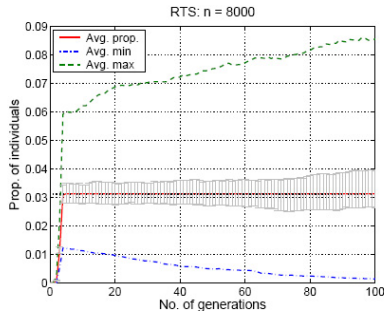
(c) RTS



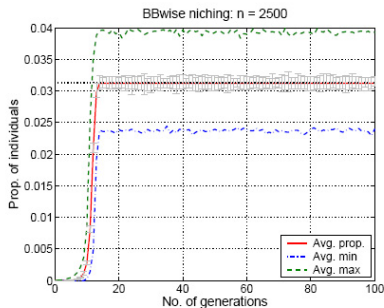
(d) Sub-structural niching

Result 3: observe one optimum

In contrast, we can say RTS converges fast.



(e) RTS

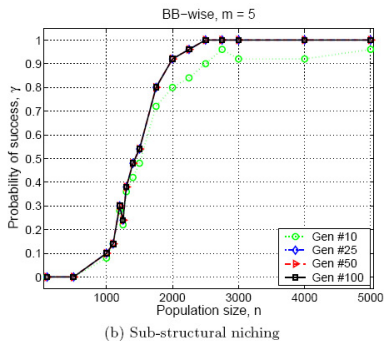
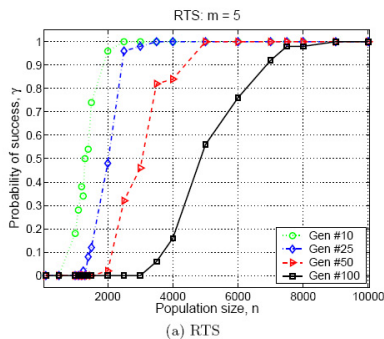


(f) Sub-structural niching

Result 4: the effect of population size n

γ : the probability of maintaining at least one copy for all global optima in generation x .

RTS requires larger population size to maintain the global optimum.



Any Question?