A Test Function with Full Controllability over Overlapping

[Estimation of Distribution Algorithms]

Jen-Hao Chang
Department of Electrical Engineering
National Taiwan University
r99921037@ntu.edu.tw

Chung-Hsiang Hsueh
Department of Electrical Engineering
National Taiwan University
r99921044@ntu.edu.tw

Hsuan Lee
Department of Electrical Engineering
National Taiwan University
r99921048@ntu.edu.tw

Tian-Li Yu
Department of Electrical Engineering
National Taiwan University
tianliyu@cc.ee.ntu.edu.tw

Tsung-Yu Ho
Department of Electrical Engineering
National Taiwan University
d98921029@ntu.edu.tw

ABSTRACT

This work proposes a test function to study overlapping. The test function provides full controllability over overlapping. To achieve full controllability, the building block assigning problem is reduced to a bipartite matching problem which allow us to directly assign extent of overlapping to each gene. At the end, an experiment on overlapping shows that to four chosen crossover methods, the problem difficulty increases exponentially with the extent of overlapping.

Categories and Subject Descriptors

G.1.6 [**NUMERICAL ANALYSIS**]: Optimization—*Global optimization, unconstrained optimization*

General Terms

Design, Experimentation

Keywords

Genetic algorithm, building block, overlap, test function

1. NOTATIONS

A chromosome C of length l is represented as a series of genes, $C = g_1 g_2 ... g_i ... g_l$, where the subscripts are the index of gene. The fitness of C is defined as $f(C) = \sum_{i=1}^{m} f_i(G_i)$, where m is the total number of sub-problems, or building blocks (BBs), f_i is the fitness function of i^{th} sub-problem, and G_i is an ordered set of genes related with f_i. A problem is said to be with overlaps if a gene belongs to two or more sub-problems. We define ω as the number of BBs a gene belongs to and $\bar{\omega}$ as the average of ω of all genes in a chromosome.

2. TSUJI ET AL.S' TEST FUNCTION AND ITS INSUFFICIENCIES

Tsuji *et al.* [1] proposed the first practical test function with controllable extent of overlaps. The ordered set of genes related to a sub-problem, G_i, is defined as: $G_i = (N(3i, \sigma^2) \bmod l, N(3i, \sigma^2) \bmod l, ..., N(3i, \sigma^2) \bmod l)$, where $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2. A gene cannot be in G_j more than once. Both σ and μ can control the test function. As the definitions in Section 1, $\bar{\omega} = \frac{k}{\mu}$.

Although Tsuji *et al.*'s test function provides adjust-ability of $\bar{\omega}$, because it decides problem structure by sampling a distribution, it lacks the ability to assign ω to each gene or even to construct a homogeneous overlapping problem structure, where ω of every gene is almost the same. Even when $\sigma \to \infty$, homogeneity is not guaranteed. When experimenting on overlapping, Tsuji *et al.*'s test function is good but not good enough.

3. PROPOSED TEST FUNCTION

A useful test function for overlapping experiments requires not only controllability over $\bar{\omega}$ but also the ability to construct a homogeneous problem structure, where all ωs of all genes are almost the same. Various set of ωs, or heterogeneity in problem structure, makes the difficulty of a problem hard to expect. If those unexpected overlaps are not noticed and handled well, experiments on those heterogeneous problems will be hard. We propose a test function which allows us to assign ω to every gene and still keeps randomness of the problem structure.

3.1 Full controllability over overlapping

A test function with full controllability over overlapping is proposed. Full controllability means we can directly assign ω to each gene. It provides not only intuitive control of overlapping but also the ability to construct a homogeneous overlapping structure. To achieve full controllability, the building block assigning problem is reduced to a bipartite matching problem. By setting parameters, expected overlapping structures are constructed. Bipartite matching problem can be easily solved by finding maximal flow. The procedure to create a heterogeneous structure is trivial so is omitted. We introduce the procedure to construct a homogeneous problem structure, where all ωs of all genes are almost the same. Suppose every BB contains $k \in \mathbb{N} - \{0\}$ different genes, and the chromosome length is $l \in \mathbb{N} - \{0\}$.

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

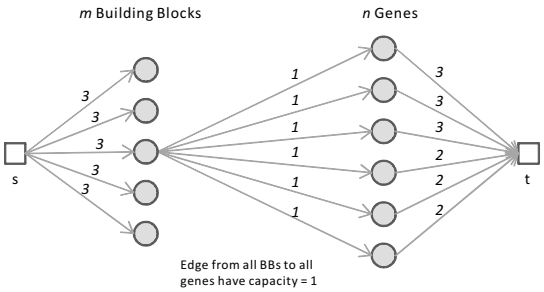

Figure 1: The flow graph of $l = 6, k = 3, \omega_{Desired} = 2.5$. Six nodes represent genes, and $floor(\frac{l\omega_{Desired}}{k}) = 5$ nodes represent BBs. All edges from source to BBs have capacity equal to $k = 3$. Edge from all BBs to all genes have capacity = 1.

The desired ω of each gene is $\omega_{Desired} \in \mathbb{R}, \omega_{Desired} \geq 1$. All $k, l, \omega_{Desired}$ are given. The size of a BB, m, is set to be floor of $\frac{l\omega_{Desired}}{k}$. When difference between any two flow from genes to the target is at most 1, homogeneity is achieved. To ensure homogeneity, difference between any two capacities from genes to the target is at most 1, and the total capacities from genes to the target should be equal to maximal flow; that is $mk = l\bar{\omega}$. The capacities of edges from $mk \ mod \ l$ genes to the target are set at ceiling of $\frac{mk}{l}$, and others are set at floor of $\frac{mk}{l}$. The capacity of minimal cut, ml, should be larger than mk to ensure every BB is fully assigned. An example of $l = 6, k = 3, \omega_{desired} = 2.5$ is shown in Figure 1.

When maximal flow is achieved, ω_i of all genes are either $\omega_{Desired}$ or $\omega_{Desired} - 1$. Homogeneity is achieved. Because of multiple solutions of maximal flow, randomness is kept. Figure 2 shows statistics of 1000 different constructions. It shows that the average ω of all genes is always close to $\omega_{Desired}$. The number of BBs a BB overlaps is also drawn, which shows there still exist randomness in the structures. Therefore, the proposed test function can construct a homogeneous structure without loss of randomness.

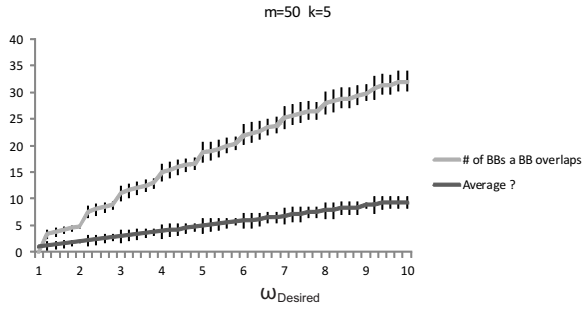

Figure 2: This figure shows the relationship between $\omega_{Desired}$ and $\bar{\omega}$ of structures constructed by the proposed test function. Standard deviation is used to draw error bars. Each point is a result of 1000 different assignments. It shows that $\bar{\omega}$ is always close to $\omega_{Desired}$, and the maximal difference of ω between any two genes is 1.

4. EXPERIMENTS ON OVERLAPPING

Without lose of generality, all BBs are $trap_k^{one}$ with $k = 5$. With full information of BBs, we compare minCut [Yu2005], minCut$^+$ [1], and the crossover proposed by Yu *et al.* [Yu2009], which we called it strength based sampling, SBS. When information of BBs is not provided, we use DSMGA with these crossovers to compare with hBOA. Each point is a result of 5 independent bisections with 10 successive runs to find the global optimum. The results are shown in Figure 3.

When the information of BBs is given, SBS outperforms minCut and minCut$^+$. When the information of BBs is unknown, hBOA performs best. SBS outperforms minCut and minCut$^+$ only when σ gets larger. Both result show that the difficulties of problems with overlaps are highly related to ω. The number of function evaluations, nfe, required to solve this problem approaches an exponential function of ω for most of the existing methods.

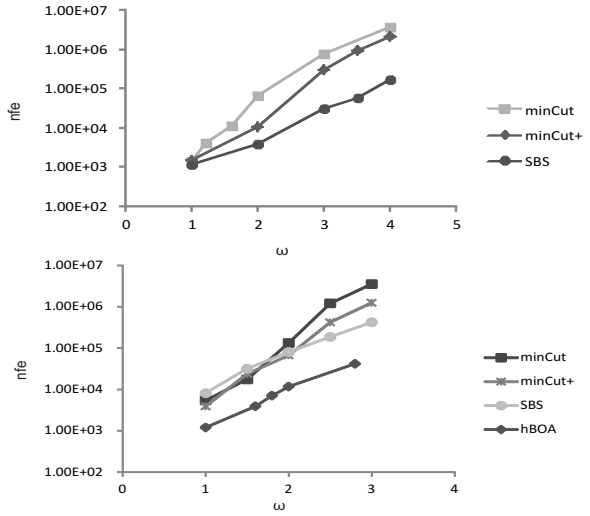

Figure 3: Comparison on the proposed test function. The chromosome length is 30. (a)with perfect BB information. (b) without perfect BB information.

5. CONCLUSION

This work tries to pave the way for future researches on overlapping. A test function with full controllability over overlapping is proposed. By using the proposed test function, the number of BBs a gene belongs to, or ω, of each gene can be directly assigned without loss of randomness of problem structure. Experiments on any desired extent of overlapping can be implemented, making researches on overlapping easier. Four crossover methods are compared on different extent of overlapping. Results show the problem difficulties to these crossover methods increase exponentially with the extent of overlapping.

6. REFERENCES

[1] M. Tsuji, M. Munetomo, and K. Akama. A crossover for complex building blocks overlapping. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, GECCO '06, pages 1337–1344. ACM, 2006.