Interaction Detection by NFE Estimation

A Practical View of Building Blocks

Kai-Chun Fan
Department of Electrical
Engineering
National Taiwan University
Taipei 106, Taiwan
r98921070@ntu.edu.tw

Tian-Li Yu
Department of Electrical
Engineering
National Taiwan University
Taipei 106, Taiwan
tianliyu@cc.ee.ntu.edu.tw

Jui-Ting Lee
Department of Electrical
Engineering
National Taiwan University
Taipei 106, Taiwan
r98921050@ntu.edu.tw

ABSTRACT

Multi-variate estimation of distribution algorithms (EDAs) build models via detecting interactions between genes and estimate the distributions to solve problems. EDAs have been applied for real world applications, but whether the models given by EDAs match what are really needed to solve the problems is yet unknown. This paper proposes using the number of function evaluation ($\mathcal{N}_{fe}$) to measure the performance of models and defines the optimal model to be the one that consumes the fewest $\mathcal{N}_{fe}$ for EDAs to solve the problem. Then the building blocks (BBs) that construct the optimal model are defined as the correct BBs. The capabilities of some existing interaction-detection metrics are compared based on this definition. This paper also proposes a test problem by utilizing Bézier curve. We find that all the mentioned metrics fail to identify the correct BBs for the proposed problems intrinsically. This paper then proposes a new metric directly based on the idea of $\mathcal{N}_{fe}$ to enhance the existing interaction-detection mechanisms. Empirical results show that the new metric is able to build the optimal models. The preliminary success suggests another view on learning linkage.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Algorithms, Design

Keywords

Genetic Algorithm, Linkage Learning, EDA, Building Blocks

1. NEW DEFINITION OF BB: FEWEST $\mathcal{N}_{FE}$

Since the importance of building blocks (BBs) has been addressed, many different linkage-learning techniques have been developed for genetic algorithms (GAs). One fast-growing development that falls into this category is the estimation of distribution algorithm (EDA). EDAs identify linkages among genes by utilizing interaction-detection metrics, such as non-linearity, entropy, simulnateity, and DMC and

Figure 1: $\mathcal{N}_{fe}$ **for OneMax and** $(3,3)$**-trap with 21147 possible models.**
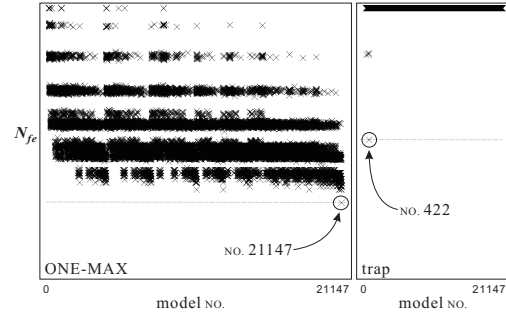
$l = 9$	Optimal Model	# of Model
OneMax	[1][2][3][4][5][6][7][8][9]	21147
$(3,3)$-trap	[1-2-3][4-5-6][7-8-9]	422

build models based on the views on BBs. However, the meanings of BBs have not been defined clearly and debated in many researches till now. To give an unambiguous definition, this paper proposes adopting the number of function evaluation ($\mathcal{N}_{fe}$), which means the product of convergence time and population size, as a metric to estimate models.

Definition 1. If a group of genes are considered together, such group is considered as a BB. Such genes are considered interacting with each other. If a gene does not interact with any others, it constructs a building block itself.

Definition 2. A model is a set that contains all BBs. If a model consumes the fewest $\mathcal{N}_{fe}$ for EDAs to solve a specific problem, then the model is the optimal model for such problem. The BBs that construct the optimal model are defined as the correct BBs.

Figure 1 shows the $\mathcal{N}_{fe}$ of all possible models for solving the OneMax problem and the trap function. Every possible model is mapped to a unique positive integer. The optimal models for the two problems are listed below the figure. The empirical results show that the optimal model for the OneMax problem is that all genes construct BBs itself for the OneMax problem; and the optimal model for the trap function is that a group of genes construct a BB if and only if they are in the same sub-problem.
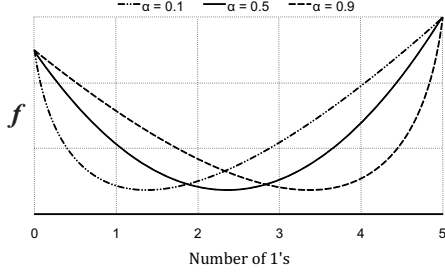
$(m, k) = (2, 5)$	Optimal Model
$\alpha \leq 0.3$	[1][2][3][4][5][6][7][8][9][10]
$\alpha \geq 0.4$	[1-2-3-4-5][6-7-8-9-10]

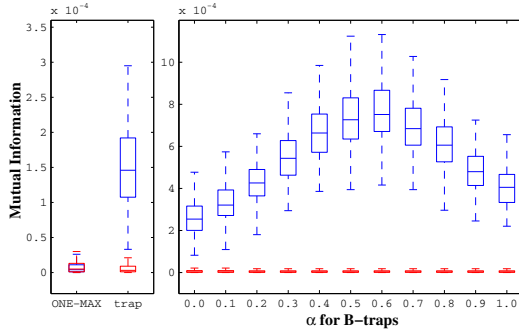**Figure 2: Fitnesses and optimal models of B-traps.**

Figure 3: The distributions of mutual information for two groups of relations among gene pairs with various problems.

2. FAILURE OF EXISTING METRICS

To examine existing metrics more precisely, this paper proposes another test function developed from Bézier curve, named B-trap. The general form is given by

$$B(t) = (1 - t)^2(0, 0.9) + 2(1 - t)t(\alpha, 0) + t^2(1, 1).$$

Figure 2 shows the fitness curves and optimal models for B-traps. We could separate the pair-wise relations into two group: one contains the pairs which the two genes locate in the same sub-problem; the other one contains the rest. Then, we observe the distributions of interaction-detection metrics for the two groups to examine the capabilities of the metrics. Such separation is reasonable because if and only if every gene in the optimal model constructs a BB itself, the characteristics of two distribution will be the same.

Figure 3 shows the results of mutual information. For the OneMax problem, the two distributions are almost the same. For the trap function and B-traps, the two distributions are all different. Simultaneity and DMC perform similarly to mutual information. This paper does not do the experiments for the non-linearity metric because the results are so trivial that can be calculated directly. The results show that all mentioned metrics can identify the optimal models of OneMax and the trap function intrinsically. However, they fail to detect correct interactions for B-traps with $\alpha \leq 0.3$.
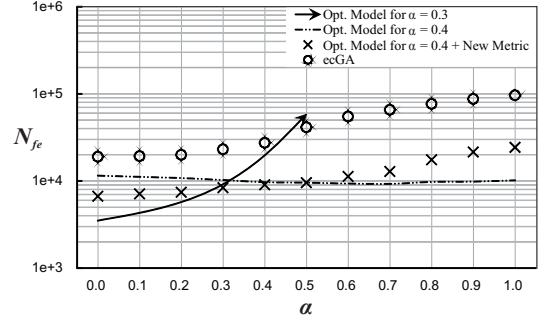
Figure 4: $\mathcal{N}_{fe}$ of B-traps, where $(m, k) = (30, 5)$.

3. PROPOSED METRIC

The previous section shows that commonly used interaction-detection metrics lack of abilities to deal with the B-traps in some cases. To overcome the difficulties, this paper proposes a new metric directly based on the new definition of BBs. The proposed metric is detailed as follow:

Algorithm. The proposed metric for real applications.

1. $b_{ij} \leftarrow$ true. Implement the original interaction-detection mechanism. If interaction is not detected, then $b_{ij} \leftarrow$ false and jump to step 5.

2. Find out the schema which should be increased.

3. Detect the convergence direction. If the direction is incorrect, then $b_{ij} \leftarrow$ true and jump to step 5.

4. If $(\mathcal{N}_{fe,i} \leq \mathcal{N}_{fe,ij}) \wedge (\mathcal{N}_{fe,j} \leq \mathcal{N}_{fe,ij})$, then $b_{ij} \leftarrow$ false.

5. Interaction detected between genes (i, j) iff $b_{ij} =$ true.

The empirical results show in Figure 4. The new metric successfully identify the optimal models for B-traps intrinsically.

4. CONCLUSIONS

This paper proposes using the number of function evaluation ($\mathcal{N}_{fe}$) to estimate the performance of models and then defines the one which consumes the fewest $\mathcal{N}_{fe}$ as the optimal model that GAs really need to solve problem. The proposed definition of BBs is practical and unambiguous. Based on the new definition, we found that some existing interaction-detection metrics fail to detect the correct interactions with the B-traps. To overcome the problem caused by B-traps, this paper proposes a new metric directly based on the idea of $\mathcal{N}_{fe}$ to enhance the existing interaction-detection mechanisms. Empirical results show that the proposed metric reach the goal of recognizing the correct BBs intrinsically. For more details, readers are referred to [1].

5. REFERENCES

[1] K.-C. Fan, T.-L. Yu, and J.-T. Lee. Interaction detection by nfe estimation: A practical view of building blocks. *TEIL Report No. 2011001*, National Taiwan University, Taipei, Taiwan, February 2011.