Information Sciences 230 (2013) 162–182

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Linkage learning by number of function evaluations estimation: Practical view of building blocks

Kai-Chun Fan, Tian-Li Yu [*], Jui-Ting Lee

Taiwan Evolutionary Intelligence Lab, Department of Electrical Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan

ARTICLE INFO

Article history:
Received 17 October 2011
Received in revised form 16 November 2012
Accepted 30 December 2012
Available online 10 January 2013

Keywords:
Estimation of distribution algorithm
Linkage learning
Building block
Genetic algorithm
Number of function evaluations

ABSTRACT

Estimation of distribution algorithms (EDAs) identify linkages among genes and build models which decompose a given problem. EDAs have been successfully applied to many real-world problems; however, whether their models indicate the optimal way to decompose the given problem is rarely studied. This paper proposes using the number of function evaluations ($\mathcal{N}_{fe}$) as the performance measure of EDA models. As a result, the *optimal model* can be defined as the one that consumes the fewest $\mathcal{N}_{fe}$ on average for EDAs to solve a specific problem. Based on this concept, *correct* building blocks (BBs) can be defined as groups of genes that construct the optimal model. Similarly, linkages within a BB are defined as the *correct* linkages of which the specific problem consists. The capabilities of four commonly used linkage-learning metrics, nonlinearity, entropy, simultaneity and differential mutual complement, are investigated based on the above definitions. For certain partially separable problems, none of the above metrics yields difference that is statistically significant between linear and nonlinear gene pairs. Although an optimal threshold still exists to separate linear and nonlinear gene pairs, most existing EDA designs today have not yet characterize such threshold. Based on the idea of $\mathcal{N}_{fe}$ estimation, this paper also proposes a metric enhancer, named *eNFE*, to enhance existing linkage-learning techniques. Empirical results show that *eNFE* improves BB identification by eliminating spurious linkages which occur often in most existing EDAs.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Since the importance of building blocks (BBs) was addressed [17], many different linkage-learning techniques have been developed for genetic algorithms (GAs) [17,11]. Other than a few uni-metric GAs such as LLGA, many such developments adopt additional metrics to identify BBs and then build models for GA operators to process.

There exist several commonly used linkage-learning metrics, such as *nonlinearity*, *entropy*, *simultaneity* and *differential mutual complement* (DMC). Nonlinearity analyzes the fitness differences by perturbing values of genes to learn linkages among them. If the fitness contribution of a gene depends on another, these two genes are considered *nonlinear*. GEMGA [18], OmeGA [19], LINC/LINC-AN/LIMD [24], D^5 [32] and ILI [6] are some typical GAs which adopt nonlinearity. Estimation of distribution algorithms (EDAs) or probabilistic model building GAs (PMBGAs), such as ecGA [16], BMDA [27], EBNA [8], BOA [25], DSMGA [35] and LTGA [26] are GAs that identify linkages by utilizing external metrics, such as entropy [30], simultaneity [2] and DMC [9]. GAs that adopt additional linkage-learning metrics to build models have been widely and successfully applied to many real-world problems. However, BBs and linkages have not been clearly defined, and their meanings and existences have caused major disputes in past decades. Until now, this issue is still inconclusive and could potentially hinder GA development.

---

[*] Corresponding author.
E-mail addresses: r98921070@ntu.edu.tw (K.-C. Fan), tianliyu@cc.ee.ntu.edu.tw (T.-L. Yu), r98921050@ntu.edu.tw (J.-T. Lee).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.12.032

The purpose of this paper is threefold: (1) to define BBs in a practical and unambiguous way, (2) to investigate existing linkage-learning metrics based on the proposed definition, and (3) to develop a technique that helps existing metrics identify *correct* BBs.

First, we propose using the number of function evaluations ($\mathcal{N}_{fe}$) to define BBs. To make this concept work, we use $\mathcal{N}_{fe}$ as the performance measure of EDA models. By adopting a specific model, if the EDA consumes fewer $\mathcal{N}_{fe}$ to solve a specific problem, that model is considered more appropriate for this problem. Under this premise, we define that the *optimal* model for a problem is the one which yields the fewest $\mathcal{N}_{fe}$ to solve the problem. The way that the optimal model decomposes the problem defines BBs of the problem.

To examine metrics more precisely, we design a test function by using Bézier curves [15], named B-trap. The idea behind the design of B-trap is to allow researchers to generate arbitrary degree of dependency among genes within one sub-function with a parameter. With B-trap and several other test problems, we categorize the problem space into several sub-spaces and investigate the capabilities of nonlinearity, entropy, simultaneity and DMC.

Finally, we propose a metric enhancer, named *eNFE*, inspired from the classification and based on the fundamental theory of $\mathcal{N}_{fe}$, to enhance the linkage-learning ability of existing metrics.

The rest of this paper is structured as follows. Section 2 reviews the BB hypothesis and linkage-learning techniques. Section 3 defines what we mean by BBs and examines the capabilities of commonly used linkage-learning metrics. Section 4 describes our linkage-learning enhancer, *eNFE*. Section 5 experiments *eNFE* on several test problems. Finally, Section 6 concludes this paper.

2. GAs, building blocks, and linkage learning

GAs are stochastic search and optimization methods which require only the quality information of solution candidates. The simple GA (SGA) is a specific class of GAs which does not involve learning problem structures, and has been widely applied to many different fields for optimization [11]. However, recombination without concerning problem structures has its limitation, and on some problems like the deceptions, the recombination results in inferior offspring [10].

Centered around the notion of BBs, GA researchers have developed a class of GAs which utilizes linkage-learning techniques to solve nearly decomposable problems with bounded difficulty [12]. These GAs are often called *competent GAs*, to distinguish from SGA. Though different competent GAs learn problem structures by different mechanisms, similarly they use the information of problem structures to customize the recombination for specific problems.

In this section, we first review the concept of BBs—atomic components for problem decomposition. We then discuss several important existing linkage-identification metrics, where we see that detected BBs are actually defined by those metrics, and could be somewhat different from those in the mind of GA researchers.

2.1. Building block hypothesis

The concept of BBs was first revealed in Holland's famous schema theorem [17], which states that minimal, sequential, and superior schemata grow under roulette-wheel-selection, one-point crossover, and simple bitwise mutation. Starting from Holland's notion of BBs, Goldberg thought the key for GAs to solve difficult problems is decomposition. He viewed BBs as atomic components during problem decomposition. This idea was known to be Goldberg's GA design theory, first proposed in 1992 [13], and then explained in greater details in his 2002's book [12]. Goldberg's GA design theory suggested that GA research should focus on solving nearly decomposable problems with bounded difficulty. To solve those problems, GAs should first identify atomic components (or BBs in Goldberg's notion), and then effectively mix those components without disrupting them. The idea that GA research should be centered around BBs, or in other words, that problems of interest are composed of BBs, is later often addressed as the building block hypothesis (BBH) [28].

Since then, BBH has drawn lots of attention in the GA field. The idea of centering GA research around problem decomposition is interesting, while the lack of clear definitions of BBs has also raised many intensive debates and challenges. Nevertheless, researchers have started developing various GAs based on the concept of decomposition, and many of these so-called competent GAs have shown superior problem-solving ability than SGA [10]. However, given that the definition of BBs is still vague, a fair comparison of different BB-identification methods is virtually impossible.

2.2. Linkage-learning metrics

Competent GAs achieve problem decomposition via linkage-learning techniques. One of the most popular techniques is to extract structure information of problems through additional linkage-learning metrics. In other words, they identify linkages among genes from the current population after selection and then recognize BBs for recombination. In this section, we introduce four existing metrics, nonlinearity, entropy, simultaneity and DMC.

2.2.1. Nonlinearity

Nonlinearity [24] (or interdependency [7]) is a metric that detects fitness difference by flipping the value of genes to identify linkages among them. In other words, it identifies linkages by examining that whether the fitness contribution of one

Download English Version:

https://daneshyari.com/en/article/392391

Download Persian Version:

https://daneshyari.com/article/392391

Daneshyari.com