Applied Soft Computing 12 (2012) 2409–2424

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Chaotic populations in genetic algorithms

Zhanshan (Sam) Ma

Computational Biology and Medical Ecology Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, PR China

ARTICLE INFO

Article history:
Received 25 July 2008
Received in revised form 9 February 2012
Accepted 4 March 2012
Available online 3 April 2012

Keywords:
Genetic algorithm
Chaos theory
Logistic Chaos map
Dynamic populations
Fitness aggregation
Power Law

ABSTRACT

We set two objectives for this study: one is to emulate chaotic *natural populations* in GA (Genetic Algorithms) populations by utilizing the Logistic Chaos map model, and the other is to analyze the population *fitness distribution* by utilizing insect spatial distribution theory. Natural populations are so dynamic that one of the first experimental evidences of Chaos in nature was discovered by a theoretical ecologist, May (1976, *Nature*, 261,459–467)[30], in his analysis of insect population dynamics. In evolutionary computing, perhaps influenced by the stable or infinite population concepts in population genetics, the *status quo* of population settings has dominantly been the fixed-size populations. In this paper, we propose to introduce dynamic populations controlled by the Logistic Chaos map model to Genetic Algorithms (GA), and test the hypothesis – whether or not the dynamic populations that emulate chaotic populations in nature will have an advantage over traditional fixed-size populations.

The Logistic Chaos map model, arguably the simplest nonlinear dynamics model, has surprisingly rich dynamic behaviors, ranging from exponential, sigmoid growth, periodic oscillations, and aperiodic oscillations, to complete Chaos. What is even more favorable is that, unlike many other population dynamics models, this model can be expressed as a single parameter recursion equation, which makes it very convenient to control the dynamic behaviors and therefore easy to apply to evolutionary computing. The experiments show result values in terms of the fitness evaluations and memory storage requirements. We further conjecture that Chaos may be helpful in *breaking* neutral space in the fitness landscape, similar to the argument in ecology that Chaos may help the exploration and/or exploitation of environment heterogeneity and therefore enhance a species' survival or fitness.

1. Background

In almost every major field of evolutionary computing, population size is one of the parameters that a researcher has to deal with. In nature, natural selection as the driving force of evolution acts upon populations, and in computing science, group search is arguably the unique feature of evolutionary computing. Therefore, it was no surprise that the pioneers of evolutionary computing, such as De Jong [6] and Holland [15], and a little bit later by Baker [4], Goldberg [10], etc. had already studied the population sizing from the very beginning of evolutionary computing. What is perhaps surprising is that more than three decades later, we still *largely* depend on the experience or ad-hoc trial-and-error approach to set the population size. In their recent monograph for evolutionary computing, Eiben and Smith [8] indicated: "In almost all EC applications, the population size is constant and does not change during the evolutionary search". Does the *status quo* imply that population size is not an important parameter?

The answer to the previous question seems a definite no. Current practice of manual setting of population size in evolutionary computation is experience-based, but not robust. Too small of populations can lead to premature convergence, and too big of populations can be computationally costly. In particular, in Genetic Programming (GP), big population is often a culprit for the too early occurrence of code bloat, which may cause computation failure or even crash the system. This manual setting by experienced programmers is acceptable for design modeling or ad-hoc applications; it is problematic for real-time applications, which require high predictability and robustness. In the latter scenarios, automatic, adaptive and robust population sizing (also other parameters such as mutation and crossover rates) is necessary. On the practical side, what is equally important is the robustness and adaptability of the algorithms. Furthermore, many of the problems we face are NP-hard problems; efficient population sizing may have significant impacts on the success of the heuristic algorithms. However, an optimal population size effective in exploring fitness space and efficient in utilizing computing resources is theoretically very

---

E-mail address: ma@vandals.uidaho.edu

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2012.03.001

intriguing. Achieving the balance in effectiveness and efficiency is such a dilemma that prompted Harik et al.'s [14] approach to the problem with a gambler's ruin random walk model, one of the most rigorous analyses of the population size problem.

Rylander [33] reviewed the population size issues in GA and stated: "There has yet to be a proven or even generally accepted optimum population size for all problems. Since earlier work indicated that no single optimum existed, there has been little recent work attempting to determine such an optimum. The most notable tradeoffs are as follows: the larger the population the more quickly a solution space can be explored, the smaller the population the more quickly it can converge to optimum once it is found. Unfortunately, there is no agreement as to how large the large is or conversely how small the small is."

Early pioneering studies of population sizing are largely ad-hoc, problem-oriented and empirical. Many of the studies were simple sample comparisons of multiple population sizes or were compared with De Jong [6] research (e.g., Grefenstette [43], Schaffer [34], Gates et al. [9], Odetayo [32], Yuen and Ma [42], Costa et al. [5]). Goldberg instead analyzed the optimal population size theoretically, and found that the optimal population size increases exponentially and is rather large for even moderate chromosome size [10]. Goldberg was also the first to study the relationship between fitness variance and population size, and later derived a conservative bound for the convergence quality of the Gas [11]. Goldberg et al. further analyzed population-sizing based on recombinative mixing, disruption, deception, population diversity, and selective pressure [13]. As argued by Alander, the results seem rather pessimistic [2]. Alander cited Goldberg's own assessment on his theoretical analysis: "too few empirical studies have been performed to know whether the theory provides quantitatively accurate predictions" [2].

Alander [2] reasoned from linear algebra that to cover each point of a given n dimensional space, at least n base vectors are needed. Alander further argued that for small values of n, the optimal population size may be approximately equal to n. Alander [2] also mentioned that in nature, large populations are more stable and resistant to evolution than small populations that may be founded by a few colonists, or the "founder effect". Although the founder effect implies that the population is small initially, and it may not tell much information about the population changes in later stages. In fact, the Chaos population model implies that initial population size has little values in predicting population dynamics because the extreme sensitivity to initial population size is a signature of chaotic populations. Alander [2] also attempted to correlate population size with problem complexity, but the characterization was solely dependent on the experiment curves and contained no biological/algorithmic arguments. Arabas et al. [3] pointed out that there are two important issues in the genetic search: population diversity and selective pressure, both of which are strongly related to each other. Both are also influenced by population sizes. It is important to achieve a balance between diversity and selective pressure. Arabas et al. [3] introduced the concept of "age" of a chromosome, which is equivalent to the number of generations the chromosome stays "alive". The age of the chromosome replaces the concept of selection. Their experiments showed that to achieve high fitness, a large population has to be maintained. Arabas et al. [3] concluded that while the variable population seemed favorable over the fixed population, the major problem of determining the lifetime parameter remained unresolved.

Dynamic population sizing in multi-objective evolutionary algorithms (MOEA) is considered an open problem (Khor et al. [17]). Khor et al. [17] approached the dynamic size for MOEA with the following intuitive concept: in an m-dimensional objective space, assuming that the desired population size at generation n is, $\mathbf{dps}(n)$, and the population density per unit volume is $\mathbf{ppv}$, the approximated trade-off hyper-area $\mathbf{A}(n)$ discovered by the population at generation n can be defined as, $lowbps \le \mathbf{dps}(n) = \mathbf{ppv}^* \mathrm{A}(n) \le upbps$, where $lowbps$ and $upbps$ are the lower and upper bounds for the $\mathbf{dps}(n)$. They developed a formula for computing $\mathrm{A}(n)$ based on the average of the longest and shortest diameters of the hyper-area (Tan et al. [36,37]). Lu and Yen [19] pointed out that the estimation of $\mathbf{dps}(n)$ by previous authors is a heuristic approximation and its robustness has to be tested with various initial population sizes, and in some occasions may lead to premature convergence. They proposed an alternative approach in which a multi-objective optimization problem is converted into a bi-objective optimization problem in terms of individual's rank and density values (Lu and Yen [19,20]).

Ma and Krings [26] compared *simple* dynamic populations (such as random, increasing, decreasing, bell-shaped, or inverse bell-shaped fluctuating schemes) with the fixed-size populations and found that the dynamic populations outperform the fixed-size populations and the results consistently outperformed the fixed-size populations with the same average population size. There was not any mathematical model used in that simple dynamic population, and the five schemes were somewhat more similar to *fine-tuned static* populations, although the term "dynamic population" was used. For example, the "decreasing population" scheme was adopted to perform more intensive exploration in the early stage of GA. In this study, we greatly expand the simple dynamic population schemes with the mathematical models derived from modeling of the natural populations. Ma [27] briefly reviewed biological/ecological principles and models that can be inspirational for developing a unified population dynamics theory in evolutionary computing.

In stochastic approach, Harik et al.'s [14] "gambler's ruin" formulation, as well as the further follow-up study by Ahn and Ramakrishna's [44] seem to be among the most rigorous analytic approaches to the population-sizing problem. Despite this excellent analytic exploration, it appears that further relaxation of the assumptions used in the studies is needed to apply the results in GA practice. However, further relaxation of the assumptions may easily lead to stochastic problems that are intractable analytically.

In summary, it appears that one of the earliest consensuses is that relative small population size in the magnitude of chromosome length is a feasible choice (e.g., De Jong [6], Schaffer [34], Alander [2], Goldberg [10]), but later studies found that population size should exponentially increase with chromosome (string) size (e.g., Goldberg [10,12], Gates [9]). This latter conclusion paints a very pessimistic picture about the efforts of obtaining optimum population size. It might partially explain why three decades after De Jong's [6] and Holland's [15] first examination of the issue, the fixed-sized populations are still dominantly used in practice as indicated by Eiben and Smith [8], despite the extreme significance of this parameter. Rigorous analysis with random walk model by Harik et al. [14], Ahn and Ramakrishna [44] brought fresh insights to the field; however, the problem is still far from solved.

2. Ecological theory: population dynamics and Logistic Chaos map model

2.1. Population dynamics

Population dynamics or the spatial-temporal change of the population size or density is the central topic of population ecology. The mathematical modeling of population dynamics can be traced back to Thomas Malthus's (1798) "*An Essay on the Principle of Populations*" (cited in Kot [18]), in which Malthus proposed now well-known Malthusian population growth model [Eq. (2) below]. It is well documented that Malthus' work had significant influence