遺伝的アルゴリズムの統計的アプローチ

藤原 義久*

通信総合研究所 知覚機構研

Abstract： 遺伝的アルゴリズム (genetic algorithm, GA) は進化モデル・組み合わせ最適化・探索のための典型的な数理工学的なアルゴリズムである．工学だけではなく，数理生態や進化遺伝モデルとして広く使われてきている．しかし，GA の確率的ダイナミックスが，与えられた「自然淘汰」「突然変異」「交叉」にどのように依存しているのか調べる道具が不足していた．最近，A. Prügel-Bennett や J. Shapiro らにより，統計 (力学) 的アプローチが行われ，非常に有望であると考えられる．そのアプローチの基本的な考え方を紹介し，簡単な問題についてシミュレーションを行った．この方法が，工学的な GA の設計 (淘汰圧と多様性の維持の問題など) について，J. Crutchfield らの中立進化ダイナミックスとの関係などについて議論した．紙面の都合上，詳しくは reference とその引用文献を参照されたい．

Introduction

Genetic Algorithm (GA) has been applied to many practical problems as stochastic search and optimization techniques (see, for a recent review, [1]). It is unique in the population-based parallel algorithm and the genetic operators. Theoretical analysis and understanding, however, is in infancy due to the unique stochastic process of GA. Although some general aspects are clarified by Markov chain analysis, there are too few powerful tools of theoretical analysis. Rather than a problem-independent "general theory", one would need theoretical tools to analyze problem-specific aspects of GA so as to utilize them for planning parameters and for modifying GA algorithms, and in order to study evolutionary dynamics.

A. Prügel-Bennett, J. Shapiro and M. Rattray have extensively developed a theoretical formalism of GA [2]–[4]. It is motivated by statistical mechanics but, basically, is a rather standard analysis in statistics. The formalism predicts statistical quantities such as the mean and the variance of fitness distribution in the population and their time evolution. So far their formalism has been successfully applied to bit-counting, "random field paramagnet", "spin glass" system, subset-sum problems as well as a learning perceptron problem.

In this manuscript with limited space, we shall concentrate on reviewing the basic idea of their approach and its simple application. Our purpose is to show how the theoretical technique can be applied practically in engineering field and study on evolution. Discussion on related things and possible extensions is omitted. Details of calculation can be found in [6].

Basic Idea of the Theory

The basic idea is quite natural in the viewpoint of statistics. One performs simulations with different initial populations and obtains many different realizations in the stochastic process of GA. Those different realizations give an *ensemble of finite populations*.. One needs some statistical quantities in order to compare the theory with the simulations performed. A natural statistics in GA would be the *fitness distribution* of population.

This leads us quite naturally to model the ensemble of finite populations with a *distribution*. Each realization can then be regarded as a set of P individuals independently sampled from a distribution $\rho(F)$. Here and hereafter we denote the size of population by P and the fitness of each individual by F.

This viewpoint is essentially equivalent to the "population and sample" in statistical inference. Nevertheless, it gives us an important bridge between a theory and an experiment. For example, the variance of the fitness distribution in a finite population is defined by $\kappa_2 = (1/P)\sum_i^P F_i{}^2 - (1/p^2)(\sum_i^P F_i)^2$. One takes the average $\overline{\kappa_2}$ over many simulations. On the other hand, it is well known in statistics that the averaged variance over all the different samplings of each set $\{F_i\}_{i=1,\cdots,P}$ from a distribution is related to the variance of the distribution as

$$\overline{\kappa_2} = \left(1 - \frac{1}{P}\right) K_2, \tag{1}$$

where K_2 is the variance of the distribution $\rho(F)$.

We assume that the fitness distribution does not deviate much from a Gaussian distribution. (This assumption is valid in many practical situations in GA with some important exceptions that will be not discussed here.) A standard way to characterize such a distribution is to

† 科学技術庁特別研究員．e-mail: yfujiwar@crl.go.jp address: 〒651-2401 神戸市西区岩岡 588-2

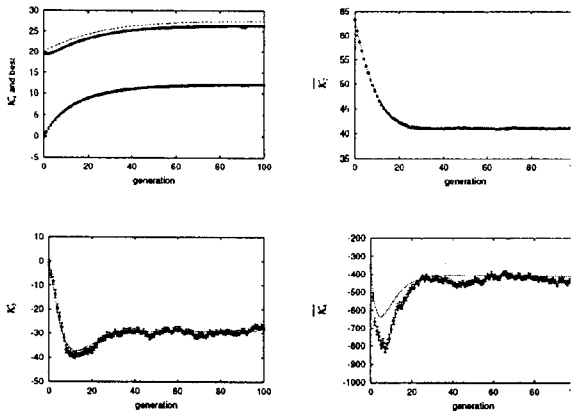

図 1: Simulation (dots) and theory (broken lines). From upper to lower: the best R_i, $\overline{\kappa}_n$ $(n = 1, 2, 3, 4)$.

employ the so-called cumulant expansion (Gram-Charlier expansion). By ignoring higher order cumulants, one has a reduced dynamical space of a few significant cumulants, on which stochastics dynamics can be described.

Similarly as (1), one can easily derive relations for each order of cumulants. Through these relations one can go back and forth from a distribution to an ensemble of finite populations taking into account the effect of the finite size of the population.

The above picture has practical advantage over the starting finite-population picture. Generally speaking, distributions are much simpler to handle than finite populations. This is due to the fact that as one tries to calculate the statistical quantities one necessarily has to treat a strong constraint appearing in the corresponding probability distribution. Since the idea of ensemble and distribution allows us to sample a set of members *independently*, the calculation can be extremely simplified and transparent.[1]

In this way one can follow the time-evolution of the fitness distribution in terms of the change of the cumulants of the probability distribution. One might call such statistical quantities as *macroscopic variables* analogously to statistical mechanics, whose space is the arena of the GA dynamics.[2]

Example for Illustration

We assume a bit string for each gene, whose allele s_α can take either $+1$ or -1. We denote the αth allele of the ith gene by $s_{i\alpha}$. Letting the length of the bit string be L, the problem is to minimize a quadratic potential of bit-counting:

$$F_i = V(R_i) = (R_i - R_0)^2 \tag{2}$$

where $R_i = \sum_\alpha^L s_{i\alpha}$. This problem is studied in [3]. (The theoretical curves for cumulant dynamics given below are obtained by a higher-order calculation made by me.)

Simulation is done and compared in Figure . Population consists of $P = 100$ such genes with $L = 64$ bits. Selection is done by the so-called roulette-wheel selection with Boltzman weight. This means to select each individual (for successive genetic operations) with a probability proportional to its fitness:

$$p_i = \frac{w(F_i)}{\sum_i^P w(F_i)}, \qquad w(F_i) = \exp(-\beta F_i), \tag{3}$$

where β is a parameter representing selection pressure. Mutation is performed by bit-flipping for each allele. Mutation rate is $1/L$ and $\beta = 1/L^2$. The minimum of potential R_0 is $L/2$.

Acknowledgments

I would like to thank Adam Prügel-Bennett and John Shapiro for helpful discussion and their patience. I express my thanks to Dr. Hidefumi Sawai who made an effort to invite them to our institute.

参考文献

[1] M. Mitchell, *An Introduction to Genetic Algorithms*, (MIT Press, 1996).

[2] A. Prügel-Bennett and J. L. Shapiro, The Dynamics of a Genetic Algorithm for Simple Random Ising Systems, *Physica* **D104** (1997) 75.

[3] M. Rattray, The Dynamics of a Genetic Algorithm under Stabilizing Selection, *Complex Systems* 9 (1995) 213.

[4] A. Prügel-Bennett, Lecture Notes on Modelling Micro-Evolution, 1996, available on `http://www.nordita.dk/adam/evol-course`.

[5] E. van Nimwegen, J. P. Crutchfield and M. Mitchell, Finite Populations Induce Metastability in Evolutionary Search *Phys. Lett.* **A 229** (1997) 144.

[6] Y. Fujiwara and H. Sawai, "Weak" Theory of GA and Its Application, 第 10 回自律分散システム・シンポジウム (計測自動制御学会主催)(1998) `ftp://www-karc.crl.go.jp/avis/yfujiwar/research/ga/sice.ps.tar.gz`

[1] This is apparently analogous to the relation between the canonical ensemble and micro-canonical ensemble in statistical mechanics, the latter of which has to handle with the constraint for the total energy of a system.

[2] One can further go beyound this "mean field" picture. Fluctuations may not be ignored for a long-time behavior due to the non-linearity of GA dynamics. See [4] for its treatment.