ПЛЕНАРНІ ДОПОВІДІ

SELECTION NOISE IN GENETIC ALGORITHMS

Gulayeva N., Borrego-Díaz J., Sancho-Caparrini F.

ШУМ ВІДБОРУ В ГЕНЕТИЧНИХ АЛГОРИТМАХ

Гулаєва Н. М.[1,2], Боррего-Діаз Х.[2], Санчо-Капарріні Ф.[2]
[1]National University of Kyiv-Mohyla Academy / [1]Національний університет «Києво-Могилянська академія», [2]University of Seville / [2]Університет Севільї
[1]NaUKMA, dept. of Informatics, 2 Skovorody street, Kyiv, Ukraine, 04070, tel. (038) 044 463 69 85
[2]University of Seville, dept. of Computer Science and Artificial Intelligence, avenue Reina Mercedes, Seville, Spain, 41012, tel. (034) 954 55 68 17 / [2]Universidad de Sevilla, Escuela Técnica Superior de Ingeniería Informática, Departamento de Ciencias de la Computación e Inteligencia Artificial, avda. Reina Mercedes s/n, Sevilla, España, 41012, tlfno (034) 954 55 68 17.
E-mail: gulayeva@ukma.edu.ua.

Ефективність генетичних алгоритмів значною мірою визначається налаштуванням їх параметрів. Відбір, на думку багатьох авторів, є фундаментальною силою еволюційного процесу. Існуючі схеми відбору відрізняються за такими характеристиками як тиск відбору, шум відбору (генетичний дрейф), обчислювальна складність тощо. Втім, досі не проведено детального порівняльного аналізу схем відбору за формальними характеристиками. В роботі вивчається шум найпоширеніших схем відбору (пропорційних, за рангом, турнірних) за допомогою двох метрик, а саме шумового часу поглинання η та чистої швидкості репродукції PRR. Досліджено вплив розміру популяції, довжини хромосоми, а також параметрів схем відбору на шум відбору. Порівняння схем відбору, що є еквівалентними за тиском, показало, що в більшості випадків відповідні схеми різняться за шумом. В роботі поєднані теоретичний та експериментальний підходи.

Genetic algorithms (GAs) are metaheuristic algorithms inspired by the biological evolutionary process; these algorithms are used to solve various scientific and practical problems including those arising in artificial intelligence systems. GAs are characterized by a set of parameters, and the choice of parameter values significantly affects GA performance.

Selection is often considered as a fundamental force in evolutionary process, and several dozens of selection schemes (SSs) and their modifications have been developed by now. These SSs differ in terms of their convergence properties, selection pressure, selection noise (genetic drift), computational complexity, and other characteristics. Despite the progress in understanding GA parameters achieved in recent years, choosing an appropriate SS as well as tuning other GA parameters is still a challenge [1]. We believe that providing explicit numerical characteristics of SSs and conducting comparative analysis of SSs based on these characteristics could contribute greatly to choosing an appropriate SS by practitioners.

In our study, we concentrate on the selection noise characteristic of SSs. Recall that genetic drift, a well-known phenomenon in population genetics, is observed in GAs due to the stochastic nature of SSs. In a finite size population, a random selection among individuals of equal fitness leads to a disproportion between the expected and actual number of copies of an individual in the mating pool. We study selection noise of the most popular SSs used in generational GAs. These are proportional SSs, namely, stochastic universal selection (SUS), roulette-wheel selection (RWS), and these schemes using fitness scaling techniques such as linear scaling, power low scaling, and sigma truncation; linear and exponential ranking SSs based on SUS and RWS; deterministic and probabilistic tournament SSs with, with partial, and without replacement. Note that ranking SSs as well as tournament SSs have scheme specific parameters (c for exponential ranking, β for linear ranking, tournament size t, and tournament winning probability p). Recall that in a generational GA the entire population of size N is replaced during each generation: N individuals are selected to the mating pool by use of a SS, and these individuals produce N offspring that replace entirely the population of parents.

To estimate selection noise of SSs, we use the *noise takeover time* η and *pure reproduction rate PRR* measures from [2]. Following our understanding of genetic drift in natural selection processes, η is defined as the number of generations needed for a population to become homogeneous on a flat fitness function when the evolutionary process is driven by selection operator only, and PRR is defined as the fraction of individuals selected into the mating pool when the evolutionary process is driven by selection operator only and effect of selection pressure is excluded. The lower the selection noise, the larger η and PRR, and vice versa.

For theoretical estimates of PRR values, see Table 1. Note that PRR for probabilistic tournaments can be obtained from PRR for deterministic tournaments at *t*=2. Values of PRR for ranking SSs based on RWS are bounded from above by $1-\left(\frac{N-1}{N}\right)^N$. Details can be found in [2].

Table 1. Theoretical estimates of PRR values.

Selection scheme	**PRR**
SUS, SUS scaled	1
RWS, RWS scaled	$1-\left(\frac{N-1}{N}\right)^N$
Deterministic tournament without replacement	$1-\left(\frac{N-k}{N}\right)^t * \left(\frac{N-i}{N}\right), k=\left\lfloor\frac{N}{t}\right\rfloor, i=N-k*t$
Deterministic tournament with partial replacement	$1-\prod_{i=0}^{t-1}\frac{(t-1)*N-i}{t*N-i}$
Deterministic tournament with replacement	$1-\left(\frac{N-1}{N}\right)^N$
Linear ranking based on SUS, $1<\beta\leq 2$	$1-\frac{\beta-1}{4}$
Exponential ranking based on SUS, $0<c<1$	$1-\left(\frac{1-\ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha}-\frac{\alpha}{\alpha-1}\right), \alpha=c^N$

The following empirical approach is used in our study of genetic drift. We conduct multiple runs of GA with a SS under analysis; no other genetic operator is used, thus, the population of offspring matches the mating pool. Runs are conducted on a flat fitness function, hence, the effect of selection pressure is eliminated. GA stops when population becomes homogeneous or when allocated time resources are spent. For each run, the number of iterations performed by the algorithm and PRR average over generations are saved. Afterwards, the average and standard deviation of number of iterations and of PRR average are calculated for the runs when the algorithm stops under the population homogeneity condition. Based on these values, η and PRR are estimated.

Our theoretical estimates of PRR are supported by the experiments. We revealed the same effect of scheme specific parameters on both the PRR and η metrics. We confirmed empirically that there is a strong relationship between η and population size *N*. Moreover, we found that for RWS and RWS scaled SSs as well as for all tournament and linear ranking SSs the value of η increases proportionally to *N*.

We have also compared selection pressure equivalent schemes to be selection noise equivalent. We found differences in selection noise for most of the known selection pressure equivalent schemes. Therefore, the choice of a SS for a practical problem should be based on the selection pressure as well as on the selection noise scheme characteristics [2].

1. Bäck, T. H. W., Kononova, A. V., Van Stein, B., Wang, H., Antonov, K. A., Kalkreuth, R. T., De Nobel, J., Vermetten, D., De Winter, R., & Ye, F. (2023). Evolutionary algorithms for parameter optimization – Thirty years later. *Evolutionary Computation*, *31*(2), 81–122. https://doi.org/10.1162/evco_a_00325
2. Gulayeva, N., Borrego-Díaz, J., & Sancho-Caparrini, F. (in press). Analysis of selection noise in genetic algorithms. *Soft Computing.*

ОСНОВНІ ВИКЛИКИ СТВОРЕННЯ УЗАГАЛЬНЕНОГО АГЕНТА ШІ: ПРОЄКТ SIMA Глибовець М.М., Бачинський Н. П.
Національний університет «Києво-Могилянська академія»
Email: glib@ukma.edu.ua, n.bachynskyi@ukma.edu.ua