DOI 10.18699/MM-HPC-BBB-2018-12

Estimates from evolutionary algorithms theory applied to gene design

A. Eremeev[1, 2]*, A. Spirov[1, 3]
[1] *The Institute of Scientific Information for Social Sciences RAS, Moscow, Russia*
[2] *Omsk Branch of Sobolev Institute of Mathematics SB RAS, Omsk, Russia*
[3] *The I.M. Sechenov Institute of Evolutionary Physiology and Biochemistry RAS, St. Petersburg, Russia*
** e-mail: eremeev@ofim.oscsbras.ru*

Key words: runtime analysis, SELEX procedure, Royal Road function, binding site, promoter, in silico gene design, synthetic biology

Motivation and Aim: The field of evolutionary algorithms (EAs) emerged in the area of computer science as a transfer of ideas from biology and developed independently for several decades, enriched with techniques from probability theory, complexity theory and optimization methods. Our aim is to consider how some recent results in theory of EAs may be transferred back into biology.
Results: It has been noted in [1] that the EAs optimizing Royal Road fitness functions may be considered as models of evolutionary search for the gene promoter sequences "from scratch". Here we consider the main known approaches to design the synthetic promoters from the EAs methodology viewpoint. This is the problem to find a tight cluster of the supposedly unknown motifs from the initial random (or partially random) set of DNA sequences using SELEX-type approaches. On the positive side, we apply the upper bounds from [2] on expected hitting time of a target area of genotypic space by EA (the EA runtime) to upper-bound the expected time to finding a sufficiently efficient series of motifs (e.g. binding sites for transcription factors) in a SELEX-type procedure. On the negative side, the pessimistic results from [3] yield upper bounds on expected proportion of the DNA sequences with sufficiently high fitness at a given iteration of SELEX-type procedure.
Conclusion: Our results suggest that some of the theoretically provable EA runtime bounds may be used, at least in principle, for a-priory estimation of efficiency of SELEX-based approaches. Further research is required to find out the properties of fitness landscape around the peaks of fitness function corresponding to separate conserved motifs in biologically meaningful fitness functions of Royal Road type.
Acknowledgements: Supported by the Russian Science Foundation (grant No. 17-18-01536).

References

1. Spirov A., Holloway D. (2012) New approaches to designing genes by evolution in the computer. In: Real-World Applications of Genetic Algorithms (ed. by O. Roeva) InTech. pp. 235-260. DOI 10.5772/36817.
2. Corus D., Dang D.-C., Eremeev A.V., Lehre P.K. (2017) Level-based analysis of genetic algorithms and other search processes. IEEE Transactions on Evolutionary Computation, Published online. DOI 10.1109/TEVC.2017.2753538.
3. Eremeev A.V. (2017) On proportions of fit individuals in population of genetic algorithm with tournament selection. Evolutionary Computation. Published online. DOI 10.1162/EVCO_a_00210.