

The Channel Capacity of Evolution

Chris Watkins

Dept. of Computer Science, Royal Holloway
Egham, Surrey TW20 0EX, UK
C.Watkins@cs.rhul.ac.uk

Categories: F.m—Miscellaneous

General Terms: Theory

Keywords: Evolutionary Computation, Genetic Algorithms, Physical Information, Information Theory

1. INTRODUCTION

Animals and plants are intricately adapted to their environments, and much genomic information is needed to construct them. In each generation, genomic information is degraded by mutation, and it is also in some sense restored by selection. It is reasonable to ask how “information from selection” may be defined. Given a suitable definition, we may then ask whether there are limits on how much information can result from selection, and if so what determines the limits? Which types of genetic encoding are most efficient, in the sense that they enable selection to maintain the greatest amount of information in the genome? Do sexual and asexual reproduction have the same informational limits? Does the informational efficiency of a genetic code depend on whether reproduction is sexual or asexual?

We define information from selection in a way that is closely related to the concept of “physical information” proposed by [?]. For some simple genetic models, including genetic algorithms, we show how to compute the maximum achievable amount of “physical information” – genomic information that can result from selection. This is new, and the major contribution of this paper. The computations are very simple, using models well known in population genetics, as described in for example [?]. Similar models have been repeatedly proposed in the evolutionary computation and the wider computer science community, for example by [?]. The calculations are simple and very well known: our information-theoretic interpretation of the results of the calculations is new, and it gives a fundamental insight into the types of genetic architecture that can code for complex organisms.

The results are striking. We consider the case where we keep the mutation-rate u fixed, and we vary the genome length L , and a finite population size (of order $1/u$). It turns out that:

- there is a clear limit on the amount of genomic information that can result from selection
- for sexual reproduction, different methods of genetic encoding have dramatically different informational efficiencies: highly diffuse, error-correcting codes are far more efficient than compact codes
- for sexual reproduction, with error-correcting genetic codes, the amount of information resulting from selection approaches a limit proportional to $\frac{1}{u^2}$
- for asexual reproduction, there is little difference in informational efficiency between error-correcting and compact coding systems, and the amount of information from selection that can be maintained in the genome is of order $\frac{1}{u}$

Information-theoretic arguments, though abstract, are powerful because they are basically counting arguments. Our results show that, given an underlying mutation rate, the complexity of organisms, or of evolutionary programs that can be maintained or produced is limited by the mode of reproduction, and by the system of genetic encoding. Our results show that the greatest precision of adaptation for organisms or for evolutionary designs, can in principle be achieved with sexual systems, with long poorly determined genomes that are decoded during development by a mechanism that is in some respects similar to error-correcting codes.