Some Techniques for the Measurement of Complexity in Tierra

Russell K. Standish

High Performance Computing Support Unit
University of New South Wales
Sydney, 2052
Australia
R.Standish@unsw.edu.au
http://parallel.hpc.unsw.edu.au

Abstract. Recently, Adami and coworkers have been able to measure the information content of digital organisms living in their Avida artificial life system. They show that over time, the organisms behave like Maxwell’s demon, accreting information (or complexity) as they evolve. In Avida the organisms don’t interact with each other, merely reproduce at a particular rate (their fitness), and attempt to evaluate an externally given arithmetic function in order win bonus fitness points. Measuring the information content of a digital organism is essentially a process of counting the number of genotypes that give rise to the same phenotype.

Whilst Avidan organisms have a particularly simple phenotype, Tierran organisms interact with each other, giving rise to an ecology of phenotypes. In this paper, I discuss techniques for comparing pairs of Tierran organisms to determine if they are phenotypically equivalent. I then discuss a method for computing an estimate of the number of phenotypically equivalent genotypes that is more accurate than the “hot site” estimate used by Adami’s group. Finally, I report on an experimental analysis of a Tierra run.

1 Introduction

The issue of what happens to complexity in an evolving system is of great interest. In natural (biological) evolution, the naive view is that life started simple, and evolved ever more complex life forms over time, leading to that pinnacle of complexity, homo sapiens. The end points of that process are of course fixed. In the beginning, life must be simple. In our present era, there must exist intelligent organisms (namely us) pondering over the mystery of how we came to be. So the anthropic principle fixes the present day as having complex lifeforms. There is nothing within the Modern Synthesis of Darwinism that implies a steady interpolation between these two end points. In fact it is even plausible that more complex organisms than us existed in the past, but have since vanished into obscurity. However, examinations of the fossil record over the Phanerozoic (the last 550 million years of the Earth’s history) indicate almost no growth in complexity by a number of different measures over that period, apart from an initial large jump at the Cambrian explosion.[1]
The interesting thing is to ask what one might see if looking at another evolutionary system apart from the one in which we evolved. Would we see any growth in complexity at all? Since we don't have an extra terrestrial biology to observe (a few Martian meteorites aside), the only other systems available are Artificial Life systems evolving within a digital computer such as Tierra or Avida. The Avida group has reported measuring the information content (complexity) of individual avidan organisms[2], or rather a lower bound of the organism's complexity. Their results are that this lower bound increases over time for the maximally fit organism, thus showing information accumulating as time progresses. One important critique of this work, however, is that organisms do not interact directly with each other, and in order to prevent evolution stagnating, an externally imposed task (e.g., computing a logical operation) is added to the system. Organisms are given “fitness points” depending on how well they perform this task. This heavily weights the system in favour for accruing information.

By contrast, in the Tierra system, the organisms interact with each other, providing a rich array of possible (intrinsic) tasks for the organisms to exploit. Since this is an evolving ecology with no externally imposed task, the above critique does not apply. However, the downside is that determining whether two genotypes are phenotypically equivalent is considerably more complex. In some work a couple of years ago[3], I studied the phenotypic properties of Tierran organisms to build up a picture of the genotype to phenotype landscape. A Tierran organism’s phenotype can be characterised by a couple of numbers for each possible pairwise interaction in the ecology. Multiway interactions are ignored in this study, as experience has shown them to be relatively rare.

2 Complexity of a Digital Organism

The information content of a string is given by the difference between the maximal Shannon entropy of that string (i.e., considering the string to be random, or devoid of information), and the entropy given by assuming that the string codes for some phenotype \( p \):[2, 4]

\[
I(g) = H(g) - H(g|p) = \ell - \log_{32} N
\]

where \( \ell \) is the length of the genotype (in instructions), and \( N \) is the number of genotypes that give rise to the same phenotype \( p \). The base, 32, refers to the number of instructions in the Tierra instruction set. If \( N \approx 32^\ell \) (i.e., a completely random sequence), then \( I(g) = 0 \). Similarly, if \( N = 1 \) (there is only one genetic sequence encoding a genotype, or no redundancy), then \( I(g) = \ell \).

The most obvious way to compute \( N \) is to search all \( 32^\ell \) genotypes for equivalent phenotypes. However, this is an enormous number of strings to check, and computationally infeasible. Adami recognised this problem, and took the approach of counting the number of volatile sites \( v \) (sites that vary amongst phenotypic equivalents), and approximating \( N \approx 32^v \). In one sense this is an overestimate of \( N \), so they argue that this gives a lower bound to the information \( I(g) \). In another sense, however, it is not strictly a lower bound. If it turns out