Artificial Life, The Second Law of Thermodynamics, and Kolmogorov Complexity

John R. Woodward, Amin Farjudian
School of Computer Science
The University of Nottingham, Ningbo, China
The People's Republic of China
Emails: john.woodward@nottingham.edu.cn, amin.farjudian@nottingham.edu.cn

***Abstract*—One of the basic features of life is replication. Indeed one of the three components of evolution is inheritance, which implies some similarity (both phenotypic and genotypic) between parents and offspring. Life is a process and not a substance (e.g. being carbon-based does not capture what life is), and this therefore justifies an algorithmic definition. Artificial life is concerned with the study of synthetic life, and is implemented on a digital computer. Tierra is a particularly prominent instance, where rudimentary life forms compete for space and time. In this system interesting eco-systems emerge, for example demonstrating parasitic behavior.**

The second law of thermodynamics states that the entropy of an isolated system is nondecreasing. The second law is a consequence of statistics, and that there are many more states with high entropy than low entropy. Entropy has connections with coding theory, data compression and Kolmogorov Complexity, as well as thermodynamics and statistical mechanics. When transmitting a coded message, the length of the message is proportional to the entropy. The shortest computer description is a universal code which is good for all probability distributions. The second law could therefore be restated as the Kolmogorov Complexity of a system is nondecreasing. Kolmogorov Complexity is incomputable, however we can use compression tools to give an approximate upper bound.

We implement a simple digital world consisting of a bit string of 0s and 1s. We implement a system like Tierra, firstly with just thermal noise and no life and observe that the compression size of the world increases according to the second law. Secondly we introduce Tierra-like creatures which replicate, and observe that the size of the compressed world decreases. The question we address in this paper is, does the basic replication mechanism underpinning life cause a decrease in entropy. The contribution of this paper is a demonstration, that in an artificial life scenario, entropy, as measured by a compression algorithm, decreases violating the second law of thermodynamics. Thus we can use this as an algorithmic definition of the process of life.

Keywords- Tierra; life; entropy; compression; complexity; Huffman code; gzip; randomness; thermodynamics.

I. Introduction

Due to space limitations we let the abstract serve as our introduction and directly give the outline of the paper. In section II we discuss some of the characteristics of living things, and state that reproduction is the key feature (both at the level of the organism and the cell). We discuss one implementation in particular of artificial life, Tierra, which is the background for the system we implement in this paper. In section III we examine entropy and the second law of thermodynamics. In section IV we examine the connection between information entropy used in coding and Kolmogorov Complexity. In section V we report on two experiments, one which shows an increase in the compression length of a digital world, and one which shows a decrease. This decrease is due to many copies of the same species accumulating in the world and allowing the compression length to decrease. In section VI we engage in a discussion and suggest intended future work. In section VII we summarize the article.

II. Life and Artificial Life

The notion of life is difficult to define. Living entities move, grow, metabolize, are sensitive to their environment, reproduce and adapt under evolution. There are a number of physical characteristics that superficially link all life forms witnessed on earth, for example, they contain RNA, protein and carbon. The existence of water is also key to the existence of many entities. However, perhaps the ability to reproduce is the key characteristic underlying what it means to be "living". Reproduction is essentially a rewrite rule (i.e. make a copy of me). When individuals reproduce (sexually or asexually) they make a copy of themselves. Though a living organism need not necessarily reproduce it is a characteristic of any species overall. That is, an organism may meet all of the other characteristics of living things, but may not produce any off spring of their own. A definition of life should perhaps include viruses and prions, which have been described as being "at the edge of life", but importantly also demonstrate replication. Life is a process rather than a material substance (i.e. being composed of carbon), and therefore an algorithmic definition of life is appealing. Therefore we turn our attentions to Artificial Life which will allow us to study the algorithmic aspects of life more easily.

Tierra (Spanish for earth) is an artifical life system which was set up to study the dynamics of the synthesis of life. It is the study of open ended evolution, rather than having a particular termination condition in mind, when a solution to a problem is evolved (e.g. as in genetic algorithms). Ray [2] starts with organisms which can already replicate (this is given at the starting point), and the aim is to generate increasing diversity similar to that associated with the Cambrian explosion. A "slice of time" is handed out to each individual in the community approximating parallelism (as long as the time slices are small compared to the time to replicate). Ray

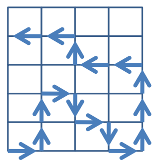

Fig. 1. An illustration of the second law of thermodynamics. A system performs a random walk through the space of states of equal energy.

achieves his aim, and creates an interesting interacting collection of parasites and hyper-parasites. The three components of evolution are present (selection, inheritance and variation). In this paper we have implemented a simple Tierra-like scenario.

III. Entropy and The Second Law of Thermodynamics

In statistical mechanics, entropy is the number of ways a system can be arranged. Entropy is often considered as an indication of "disorder" or uncertainty about a system. The macroscopic description can be considered as a sort of summary of a microscopic state. This definition (see equation 1) describes the number of possible microstates which give rise to the observed macrostate. Statistical mechanics states entropy is statistical in nature and therefore there is a probability that it will increase.

$$S = klnW \tag{1}$$

where k is Boltzmann's constant and W is the number of microstates corresponding to the macrostate. Entropy is proportional to the logarithm of the number of microstates a system could be in for a given macrostate. A system of atoms and molecules is a microscopic system. We observed macroscopic properties such as temperature, pressure and volume. At the microscopic level, everything is reversible (e.g. two atoms colliding). At the macroscopic level, state changes are irreversible, in the probabilistic sense (e.g. a concentration of gas diffuses towards a uniform distribution).

Entropy is a measure of the disorder in a system. The second law of thermodynamics states that a closed system has nondecreasing entropy. This is simply stating that there are more disordered states than ordered states. For example, imagine a jar of marbles which is seperated into black and white marbles. If the jar is shaken, the marbles will mix and become more disordered. However, there is always a small chance that the system can return to its initial state. Interestingly this is the only laws of physics which has a direction of time, all other physical laws remain unchanged when time is reversed.

A system is in a microstate which maps onto a macrostate. The system, under normal thermodynamics, randomly walks from one microstate to another microstate. We assume all microstates are equally likely. Macrostates are just equivalence classes of microstates. As a system meanders around state space, it moves to a state of higher entropy because there are more macrostates that correspond to higher entropy. Typically as humans, we are only aware of (i.e. make observations about) the macrostates. What decides which microstates map to the same macrostate? We will return to this question in the next section.

IV. Coding Theory and Kolmogorov Complexity

Imagine we have a random variable, X, generated by a 4 sided dice with sides labeled $\{a, b, c, d\}$. Let us suppose the dice is not evenly weighted and the probabilities of each face are $(1/2, 1/4, 1/8, 1/8)$ respectively. We generate a sequence of rolls e.g. $acdbac$ and we want to transmit this sequence using a binary sequence of symbols. a, b, c, and d are called the source symbols, and 0 and 1 are called the code symbols. We want to code a sequence of dice rolls so assign the following code words to each face

$$C(a) = 0, C(b) = 10, C(c) = 110, C(d) = 111$$

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i), \tag{2}$$

The entropy, $H(X)$, of X is 1.75 and the expected length $L(C) = El(X)$ of the code is also 1.75 (this is example 5.1.1 from [1]). This is a prefix free code, and can be uniquely and instantaneously decoded into its source symbols. (Think of the 0 representing the end of a codeword, except in the case of d).

If we want a code to be optimal in the sense that we can transmit a message with the shortest message length, we need to assign the shortest code words to the most probable source symbols, as one would intuitively expect. Morse code, for example, is fairly efficient for English as the frequently occurring "E" is represented by a single dot while the infrequently

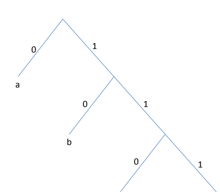

Fig. 2. A Huffman coding tree. Symbols with high probability are assigned short codes e.g. a is assigned 0, and symbols with low probability are assigned long codes e.g. c is assigned 111.

occurring "Q" is coded for by the cumbersome dash dash dot dash.

The expected length L of a $D-ary$ code for a random variable X is greater than or equal to the entropy $H_D(X)$ (see Theorem 5.3.1 [1]); that is

$$L \geq H_D(X) \tag{3}$$

with equality if and only if $D^{-l_i} = p_i$.

There is one concern with Huffman coding, which assumes a probability distribution at the start. Also Huffman coding is not optimal if we drop the symbol-to-symbol constraint. If we toss a fair coin (where $p(H) = 0.5$ and $p(T) = 0.5$, then the entropy is maximum) and obtain a sequence e.g. $HTTHTHHTHTTTHT$, or $HTHTHTHT$..., which also has entropy 1 but appears very regular. If we code this with $C(HT) =$"0" then entropy becomes zero (i.e. we are certain about the outcome), and we can transmit the message as a string of 0s. Let us now think about a different approach to coding messages.

Lets consider using a computer to compress three messages;

000000000000000000000000000000000000000,
000010000100001000010000100001000010000100001,
10101010101000011110001100010101010110.

The first can be transmitted as a program; print "0" n times. The second can be represented by a slight longer program; print "00001" m times. The final is random (in the algorithmic sense), and there is probably no shorter program, and the shortest program simply prints the literal string. The Kolmogorov Complexity of a bit string s is the length of the shortest computer program which outputs s and halts (with respect to machine U).

$$K_U(s) = min_{U(p)=s} l(p) \tag{4}$$

where $l(p)$ is the length of a program p that prints s and halts. This is important when we realize that this definition is independent of the choice of computer up to a constant.

$$K_U(s) = K_V(s) + C \tag{5}$$

where the positive constant C depends on U and V but not on s, thus we can drop the subscript which refers to the machine. This is important because it allows us to define macrostates independently of the computer model. Two microstates fall into the same macrostate (equivalence class) if and only if they have the same complexity;

$$s_1 \in [K(s_2)] \Leftrightarrow K(s_1) = K(s_2)$$

We can take the Kolmogorov Complexity of a string as the measure of disorder and is independent of the choice of computing machine. Thus 0101010101... and 010101000111... have the same entropy (as they were generated with the same probability distribution) but they have different Kolmogorov Complexity. The shortest computer description is a universal code which is good for all probability distributions. The entropy of a system is equal to its Kolmogorov Complexity. The expected length of the shortest binary computer description of a random variable is approximately equal to its entropy.

$$E(1/n)K(X^n) \rightarrow H(X) \tag{6}$$

Thus, we can use the Kolmogorov Compexity as a measure of the disorder in a system. While this is imcomputable, we can use compression tools to measure the compression size of a virtual world as creatures replicate in this world.

V. Experiments

The "world" consists of a bit string of length 10^3 bits. While this is small it is sufficient to demonstrate that life decreases compression length. We conduct two experiments, the first with only thermal noise and the second with replicating virtual creatures. In both experiments thermal noise is modeled by the random flipping of bits. The compression length of the "world" is plotted against time. In both experiments we use gzip, a standard unix utility, to compress the bit string world. This is used to give an upper bound on the Kolmogorov Complexity, and is plotted on the vertical axes figures 3 and 4.

In the first experiment, we begin with an ordered world of 10^3 0s (which is compressible). After the action of thermal noise the world becomes less compressible. We show the results in figure 3. Note that momentarily the compression length does reduce, but overall it increases in accordance with the second law.

In the second experiment two species of creature are present, represented by 0000 and 1111. These creatures are particularly simple. These creatures compete for processor time and space in the bit string world. The creatures replicate each generation (shown on the horizontal) in proportion to their population size. There are two variations of this experiment; non-aggressive and aggressive. In the non-aggressive case, copies of a species are made and may overwrite members of its own species (in whole or in part). In the aggressive case, copies of a species are made and do not overwrite its

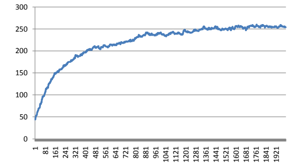

Fig. 3. Experiment 1. Thermal noise causes an increase in the entropy of a system, as measured by the compression length.

own species. The world starts off in a random configuration, and during the replicating process there is background thermal interference causing the occasional perterbation. One figure is presented (due to space limitations). However, in both cases we observe a decrease in the compression length, see figure 4. Again, as with the first experiment, we occasionally see a temporary increase in the compression length, reinforcing the statistical nature of the second law.

VI. Discussion and Further Work

In our implementation we have only implemented a very simple system as "proof of principle" that supports the concept that a reduction in entropy can be used as a definition of life. Further work includes using a more intricate Artificial Life system which could demonstrate open ended evolution. We also intend using a broader range of compression algorithms, though we suspect this is not critical. It would be interesting to see if computer viruses qualify as life under this definition. Neither have we explored the effects of extinction dynamics.

Books on information theory often consider the entropy of English ([1] [4]). The entropy of The Complete Works of Shakespeare is low because English is a predictable (e.g. u often follows q), however what make The Complete Works of Shakespeare have even lower information content is that there are many issues of the collection in print. Thus knowing the contents of one copy allows you to compress the Complete Works of Shakespeare into the single expression "The Complete Works of Shakespeare". That is, once we have compressed one copy, other copies can be referenced by the name alone. This is precisely the mechanism which makes life compressible too; many copies of the same book is analogous to making many copies of an issue of a book. This type of repetition can be exploited by a computer when compressing data (i.e. like a macro).

Freeland [6] states that one of the beauties of implementing a virtual system is that we do not need to remain faithful to the original biology. In our system there is minimal biology. We also state that we have not modeled energy in this system and, while this will play a central role, it is difficult to define what energy means in a virtual world. Of course, underlying an implemented universe is a physical computer so future work intends to take into account the energy requirements needed to support a replicating electronic organism [5].

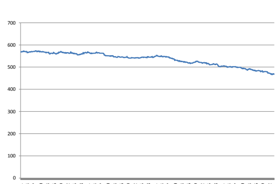

Fig. 4. Experiment 2. With basic replicating life forms, the entropy of the system, measured as compression size, decreases.

VII. Summary

The second law states that entropy is a non-decreasing quantity. One of the interpretations of entropy is a measure of disorder, that is the higher the number of microstates a system can be in, corresponding to a given macrostate, the higher the entropy. The second law is a consequence of the system performing a random walk in the space of microstates and is more likely to move to a state corresponding to higher entropy because there are more corresponding microstates.

Thermodynamic entropy is related to information entropy. Messages with low entropy can be sent with a shorter message length than messages with a high entropy. Indeed, the average message length is equal to the entropy (within rounding errors, as messages are integer length while random variables may take on continuous values). One of the issues with Huffman code and entropy as a measure of disorder is that it assumes a known fixed initial probability distribution. Also, we can generate two sequences with a fair coin (i.e. the probability of head or tails is equal and therefore has maximum entropy) e.g. $HTHTHTHTHT$ and $HHTHHTTT$, however the first clearly has a pattern which can be exploited. Kolmogorov Complexity is therefore a more robust definition of order and is independent of the choice of probability distribution or choice of computing machine. In fact the average Kolmogorov Complexity is equal to the entropy.

The contribution of this paper is the drawing together of the second law and Kolmogorov Complexity to show that entropy, as measured by a compression algorithm, can decrease in a digital world containing replicating entities. Defining life in terms of abstract mathematical concepts (i.e. a reduction in entropy) makes more sense than defining it in terms of physical concepts (e.g. carbon based, or containing RNA). Therefore an algorithmic definition of life is an attractive avenue of future research, and artificial life senarios allow us to embark upon this.

References

[1] Thomas M. Cover and Joy A. Thomas. Elements of information theory. John Wiley and Sons 1991

[2] Ray, T. 1991. An approach to the synthesis of life. In Langton, C. G.; Taylor, C.; Farmer, J. D.; and Rasmussen, S., eds., Artificial Life II. New York: Addison-Wesley. 371.

[3] Information Theory, Inference and Learning Algorithms David J. C. MacKay. Cambridge University Press 2003

[4] The Mathematical Theory of Communication(1948) Claude E Shannon Bell System Tech. J

[5] Feynman Lectures on Computation Richard P. Feynman Westview Press, July 2000

[6] Freeland SJ, Wu T, Keulmann N October 2003. The case for an error minimizing standard genetic code. Orig Life Evol Biosph 33 (4-5): 45777