Observing Integrated Information in Artificially Evolved Neural Networks

Jordan Perr-Sauer

November 17, 2009

Contents

Abstract

Evolution has proven to be a wildly successful autonomous process for creating intelligent systems in the natural world and in simulation. Since the early 1960s, researchers have used artificial evolution to find ingenious and novel solutions to complex problems such as series prediction and flight control. Recently, artificial evolution has been applied to neural networks with the aim of evolving more robust artificial intelligence. Several metrics have been proposed to chart the emergence of intelligence in these evolved networks. This work analyzes the behavior of a new metric, integrated information (ϕ). Observed data is analyzed, interpreted, and compared to more conventional properties of the artificial neural network. The data analysis shows that ϕ increases over evolutionary time and is therefore promising as a heuristic measure.

1 Introduction

This work introduces essential concepts in artificial evolution, key tools for its modern study, and new experiments and data regarding the use of heuristic analysis in neural networks generated through artificial evolution. This section describes past works and known concepts. Later sections discuss the process and results of my own experiments and connect them to ideas in the field.

1.1 Artificial Evolution

Evolutionary programming is a technique modeled after biological evolution used to create complex and well-designed programs. This method of programming was inspired by the success of evolution in the development of complex biological organisms on earth. If, over millions of years, creatures as intelligent as humans could arise from seeming randomness via evolution, why not apply this same process to yet unsolved problems? Must a "conscious" mind be involved in the solution of a complex problem? Or can a machine be programmed to *evolve* the solutions autonomously?

These questions were being tackled back in the late 1960s by researcher and computer scientist Lawrence J. Fogel. In that early dawn of evolutionary programming, Fogel fought for it's viability by conducting rigorous scientific experiments. In a series of "prediction experiments," Fogel et al. created software that studied a string of semi-random symbols with the goal of predicting the rest of the symbols in the series. These programs worked by testing a set of randomly produced finite state machines against the given series, and then mutating and re-trying the most-fit machines and discarding the worst performers. This simple process arrived at the best fit finite state machine for the given input sequence much faster than by a "brute force" approach. [2]

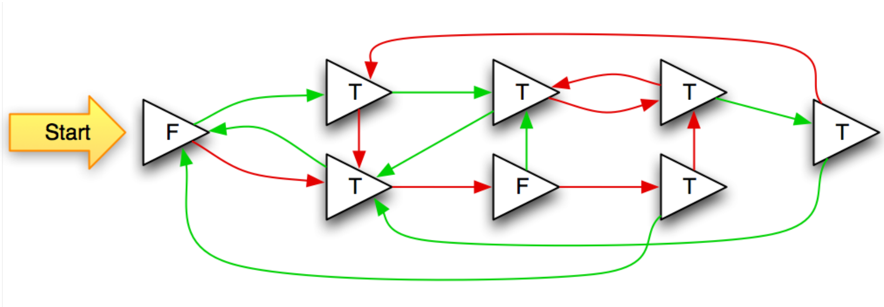

Figure 1: Diagram of a fully evolved, automatically generated finite state machine. Each triangle represents a condition. Red directional arrows indicate the FALSE path while green arrows indicate the TRUE path. Output values are indicated by the character inscribed within each node.

Figure 1 shows a finite state machine that was evolved using my own software (available online at `http://svn.jperr.com/intel/src/fogelfsm` [7]). This simulation produced an 8 node finite state machine capable of accurately predicting a repeating sequence of about a dozen binary symbols. Efficiency aside, the most intriguing aspect was how different each resultant finite state machine turned out to be. Each resultant machine was able to solve the given symbol prediction problem, yet did so in unique ways.

As I stared down at the wiring diagram (Figure 1) from my simulator's output, I realized: I could never have come up with these solutions using my own mind. It was at this instant that I understood the vast philosophical and scientific implications of a program *designing* and *creating* new solutions to a stated problem. Financiers and economists spend their entire lives trying to solve similar symbol prediction problems. Here I was, staring at the screen of an unconscious laptop which had effectively solved such a problem by itself.

1.2 Polyworld

Polyworld is a modern descendant of Lawerence J. Fogel's finite state machine evolution environment. It is an exciting tool for research in neural complexity and evolved intelligences due to its sophistication, dedication to scientific accuracy and result reproducibility. Using the power of modern day personal computers, Polyworld accurately simulates the evolution and learning in haploid agents within a virtual environment. Polyworld's agents (or Poly-

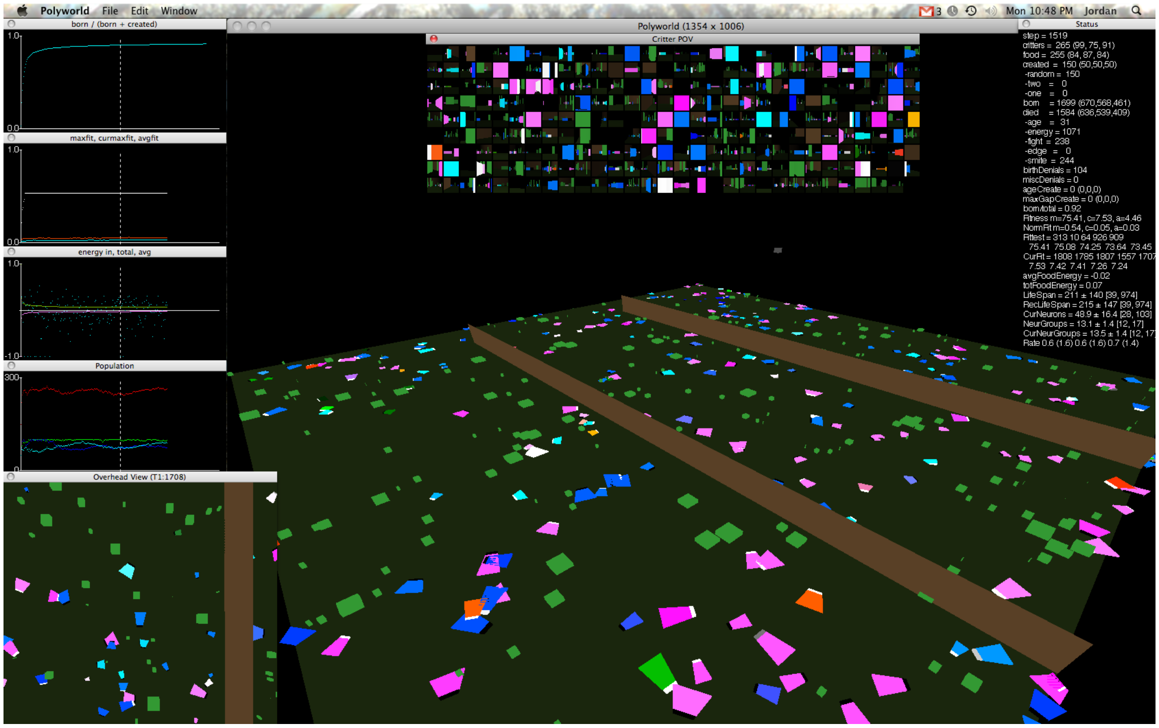

Figure 2: A typical Polyworld simulation. The green plane represents the boundaries of the simulated world. Green bricks represent food. The trapezoidal figures represent individual Polyworldians. The various panes surrounding the main simulation window provide with live streams of data, such as population size.

worldians) are governed by an artificial neural network (ANN) "brain" which is encoded in their genome. The ANN of a Polyworldian takes visual input from the rendered virtual environment (pixels), and can cause the creature to physically interact with the environment. ANNs in polyworld have some degree of plasticity and can perform Hebbian learning throughout the agent's life [11]. For more information, please refer to [8].

Much like the finite state machines in the experiments of Fogel et al. [2], unfit agents are less likely to pass their genes (and therefore their ANN) to future generations. Conversely, agents who are more fit are are able to pass their genes on to future generations.

2 Related Work

Larry Yaeger, at the School of Informatics at Indiana University, published a paper that describes his own efforts using Polyworld to observe changes in Fuctional Complexity in evolving neural networks. Yaeger showed that complexity increases for the first 5,000 time steps before tapering off. His work lent evidence to the widely held belief that biological organisms tend to increase in complexity as they evolve. This work (in a similar vein to Lizier) showed that complexity metrics follow predictable patterns as networks evolve. [10]

Joseph T. Lizier published a paper in which different topographical metrics of evolved networks such as assortativity, modularity, clustering coefficient, and closeness centrality were analyzed over time. He noticed that the neural networks tended to become more "small world" or "integrated" as evolution progressed. This observation supports the "integration" facility of Integrated Information (discussed in Section 4), yet this was unknown to Lizier. [6]

3 Significance

In order to conduct more efficient simulations of evolution, it is essential to understand exactly how neural networks change while evolving. In the past, trends in even the most basic properties of evolved neural networks were merely theoretical and were hotly debated. As computers became powerful enough to run more complicated simulations, research was conducted to scientifically measure the trends of these properties.

This work extends the findings of a recent flurry of related work (discussed in Section 2) by adding Integrated Information (discussed in Section 4) to the pool of analyzed features in evolved neural networks. My findings will help future researchers determine how and where to use Integrated Information in their own simulations. A growing school of thought is to use certain heuristic metrics as feedback to the simulator itself. In this way, the simulator might be able to "push" evolving systems down a more intelligent or perhaps even *conscious* evolutionary path.

Should these metrics prove useful in optimizing future simulations, they may become widely used "guides" in improving existing systems. While the most direct application is the development of more realistic artificial intelligence, these metrics could be used to aid the design of virtually any system. Content delivery networks could be restructured to more intelligently distribute files on servers throughout the world. Top-down corporate and military structuring and economic risk analysis could be improved by evolving models that use Integrated Information as a guiding heuristic.

4 Integrated Information

Integrated Information (ϕ) is the measure of how much of a neural network's information is generated "synergistically" among its parts. The actual algorithm used to calculate ϕ is beyond the scope of this paper (see [1, 4, 9]). The total amount of information in a network is defined as the reduction in possible states a given network experiences by the choosing of one particular future state. This is a mouthful, but is actually quite intuitive. Picture an unabridged dictionary with millions of definitions to choose from. By picking a particular definition to study, you have eliminated the millions of other definitions from your focus. With a smaller and more concise dictionary, the number of definitions excluded from consideration is much smaller when a definition in particular is chosen.

The second component to ϕ is how synergized the network is. To understand synergy, consider the following thought experiment: A human and an array of photosensitive diodes (a camera) are placed in front of a projector screen. The human is told to push a button whenever the screen is white, and to release the button when the screen turns black. The camera indicates a similar thing. When the screen turns from bright to dark (and vice versa), both subjects will correctly indicate the change. Most would agree that the human subject consciously chose to press their button while the camera did not. But *why* is the human uniquely conscious?

One answer according to Tononi et. al [?] is that the human mind contains more *Integrated Information* than the camera's sensor. When the human subject sees the screen change, their mind explores a virtually limitless number of new possible states. They might question whether the screen will flash green, whether the experiment will be over soon, or whether he or she had parked in an illegal parking space. The human mind has a vast amount of information and is incredibly skilled at connecting and organizing it. In other words, the human mind contains a relatively large amount of Integrated Information.

The camera, on the other hand, is simply allowing photons emitted by the projector to cause electrons locked within its light sensitive conductor to be freed. Each photodiode can experience two discrete states, on and off. Though the camera contains many bits of total information due to its large number of photodiodes, the camera's ϕ remains low due to the fact that each photodiode is on its own circuit and one photodiode does not affect the state of another. This isolation of nodes drives the cameras integrated information to zero and shows how synchronicity affects ϕ. What, then, is the trend in ϕ for a neural networks evolved in Polyworld?

5 Experimentation and Results

To calculate this trend, I used the opensource Consciousness project developed by Virgil Griffith [3]. This software uses the most efficient algorithm known for calculating ϕ [4]. Despite this efficiency, Consciousness operates in *factorial time.* Calculating ϕ, using a modern laptop, for a network of 11 nodes takes about 5 seconds. Calculating ϕ for 12 nodes takes 30 seconds, and calculating ϕ for 13 nodes can take longer than 5 minutes.

In a typical Polyworld simulation, over 2,000 individual agents are created with neural networks that can contain over 150 nodes. Calculating ϕ for even one standard agent (no less 2,000 of them) was simply not feasible. For this experiment, I took a page out of Dr.

Eric Kandel's playbook [5] and decided upon a reductionist approach. To do so, I created a Polyworld simulation definition (or worldfile [8]) that severely limited the number of nodes in each Polyworldian network and the number of agents in the simulation. In all simulations conducted for my analysis of ϕ, the typical Polyworldian's network contained less than 20 nodes.

The small brain anatomical neural network data generated using Polyworld was then processed for ϕ over the course of two weeks. On average, each network took approximately 60 seconds to process. Any calculation taking longer than four minutes was restarted with a lossy optimization that calculates the estimated lower bound on the actual ϕ. Because I am seeking an increasing trend in ϕ, this lower bound estimation does not affect the integrity of my results. This optimization produces, if anything, a more *conservative* trend. This faster and more conservative calculation was given exactly one hour to complete.

My automation script (available online at `http://svn.jperr.com/intel/src/pwd_calc` [7]) intelligently randomized the order in which each network was processed. By doing this, I could halt the process midway through and still salvage a full timespan trend from the data processed up to that point. With this feature, I was able to analyze a diverse set of data from many different Polyworld simulations without fully processing any one simulation. By choosing this route, I decided that data diversity was more important than dwelling on the accuracy of each individual simulation.

5.1 Analysis

Figure 3 displays observed values of ϕ using a bin size of 200 time steps. Error bars represent the standard error in each sample. The error bars are centered vertically at the mean ϕ value for their representative sample. Each bin contains an average of 78 unique agents.

An increasing trend in Integrated Information over the first 4,000 units of evolutionary time is clearly visible. The scientific community can now be certain that, for whatever reason, the amount of integrated information contained in a neural network *increases* with evolu-

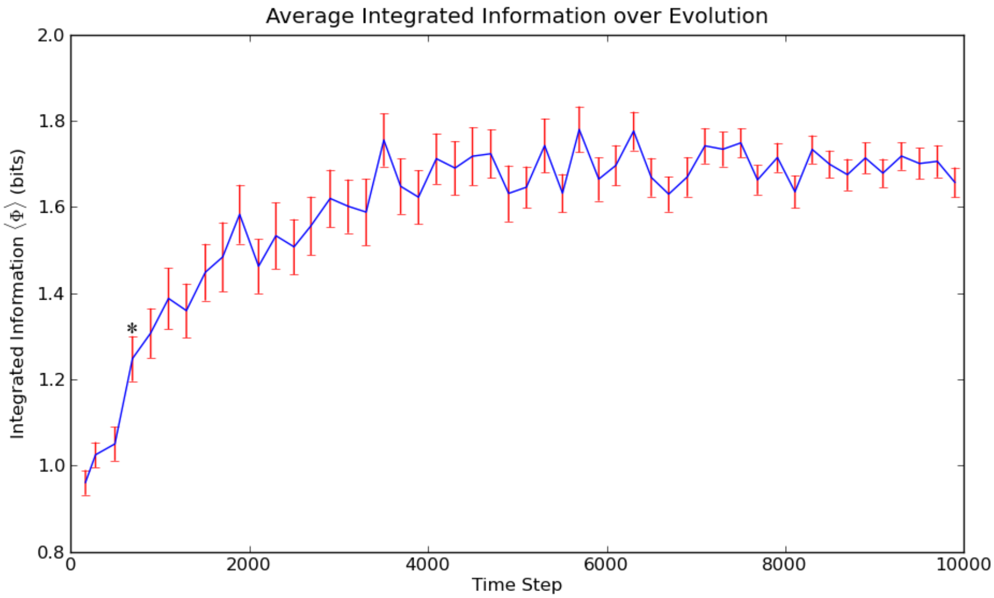

Figure 3: Results from Integrated Information experiments. The data set includes anatomical neural network sampled from 5,202 Polyworldians, at death, from three genetically independent simulations. Each data point represents a 200 time step bin size. *First sample for which $p < 0.01$

tion. This statement is consistent with early thought experiments that brought Integrated Information into existence. The results also give credence to the notion that more "evolved" biological organisms (such as humans) are more complex than lesser evolved organisms [1]. In conclusion, this data is consistent with scientific thought and accepted experimental knowledge of this time and serves to strengthen current beliefs held about the importance and usefulness of Integrated Information.

To extend these results even further, it is interesting to note the similarity between the trend of Funcational Complexity over evolutionary time as shown by Yaeger [10] and ϕ. Both metrics increase before the 3,000 and 5,000 time step mark but then taper off. In his analysis of this pattern in Complexity, Yaeger suggests, "This is consistent with the observation that the seed population is too simple to sustain its numbers and must evolve or become extinct." To my best judgment ϕ undergoes this same trend for similar reasons. It is curious that these two metrics follow similar trends, as they are both very different measurements.

6 Future Work

Integrated Information is certain to become a useful metric for researchers of artificial evolution. The results presented in this paper are a great foundation upon which future researchers can base their decisions concerning the use of Integrated Information.

As discussed in Section 4, these simulations used for analysis of Integrated Information used agents with smaller brains than normal which resulted in behavioral deficiencies exhibited by the Polyworldians during simulation. For example, most agents abandoned interesting behavioral patterns in favor of a much simpler "run around in circles" strategy. This was expected and, quite honestly, *is* the most intelligent behavior that such limited Polyworldians could develop. The fact that Integrated Information was still shown to increase during these simulations only serves to strengthens its applicability. As noted in Section 4, the best known algorithm to calculate Integrated Information operates in *factorial time*. This is a major drawback of Integrated Information when dealing with larger neural networks. Should a better algorithm be discovered, trends for more realistic sets of data could be analyzed and researchers could begin to apply Integrated Information as a feedback function for real world applications.

Another huge stepping stone on the path to evolving more intelligent machines is the analysis of naturally occurring biological networks for heuristics such as Integrated Information. We have come to a critical point in time where our technology has allowed us to peer into working brains of biological organisms and observe how they function. The scientific community hopes to soon obtain a full functional map of the nervous system of *C. Elegans*, a small yet independently living nematode worm. This creature is a promising specimen for this kind of full mapping because it has one of the simplest and smallest nervous systems of any creature on Earth. The anatomy of *C. Elegans'* nervous system has already been fully mapped, yet a full functional map of the creature's brain at work has yet be obtained. Once the scientific community has access to the full functional network of this specimen, researchers will undoubtedly test different types of heuristic measures, such as Integrated

Information, with its naturally occurring biological network. How will Integrated Information behave when tested on real biological networks? What might that mean for artificial analogs of *C. Elegans* that hold similar heuristic values? Should they be considered living systems?

The future of this field is bright. This work should be viewed as one small, yet important inquiry into the evolution of neural networks. The scientific community now knows how Integrated Information changes as neural networks evolve. It increases. What new simulation techniques and advancements might come from this newfound data? How will these techniques aid society and advance modern technology? In such a bright and open field, such questions are numerous. Their tantalizing answers are what keep researcher like me interested in our work.

References

[1] David Balduzzi and Giulio Tononi. Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Comput Biol*, 4(6):e1000091, 06 2008.

[2] Lawrence J. Fogel. *Intelligence Through Simulated Evolution: Forty Years of Evolutionary Programming (Wiley Series on Intelligent Systems)*. Wiley-Interscience, 1999.

[3] Virgil Griffith. Consciousness: Computing measures for the information integration theory of consciousness, 2008-2009.

[4] Virgil Griffith. An information-based measure of synergistic complexity based on phi. 2009.

[5] Eric R. Kandel. *In Search of Memory: The Emergence of a New Science of Mind*. Norton Paperback, 2006.

[6] Joseph T. Lizier. Functional and structural topologies in evolved neural networks. 2009.

[7] Jordan Perr-Sauer. Code references. Available online `http://svn.jperr.com/intel`, 2009. Automated Metric Calculation for Polyworld Result Files, Fogel Finite State Machine Simulation Replica, and other code referenced throughout this work.

[8] Proceedings of the Artificial Life III Conference. *Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or polyworld: life in a new context.*, 1994.

[9] Giulio Tononi. Measuring information integration. *BMC Neuroscience*, 2003.

[10] Larry Yaeger. Passive and driven trends in the evolution of complexity. *Journal of Artificial Life*, 2008.

[11] Larry Yaeger. How evolution guides complexity. *Online Journal*, 2009.