Ortus: an Emotion–Driven Approach to (artificial) Biological Intelligence

Andrew W.E. McDonald, Sean Grimes and David Breen

Drexel Unversity, Philadelphia, PA 19104
awm32@cs.drexel.edu

Abstract

Ortus is a simple virtual organism that also serves as a framework for developing biologically based artificial intelligence. Born from a goal to create complex virtual intelligence and an initial attempt to model C. elegans, Ortus implements a number of mechanisms observed in organic nervous systems, and attempts to fill in unknowns based upon plausible biological implementations, psychological observations. Implemented mechanisms include excitatory and inhibitory chemical synapses, bidirectional gap junctions, Hebbian learning with its Stentian extension. We present an initial experiment that showcases Ortus fundamental principles; specifically, a cyclic respiratory circuit, and emotionally–based associative learning with respect to an input stimulus. Finally, we discuss the implications and future directions for Ortus and similar systems.

Introduction

While much work has been done to develop artificial intelligence (AI) systems that borrow principles from organic nervous systems, far less has been done that specifically targets the intersection of biology and artificial intelligence such that biological principles—rather than a specific applicability of the technology—are of primary concern, with the main goal being a virtual system that exhibits biological intelligence (BI). As our understanding of organic nervous systems and access to computation power both increase, widespread interest in systems that do exaclty this is greatly increasing; as evidenced by DARPA's recent L2M project, in search of machines that learn throughout their lives (DARPA, 2017).

Researchers in the realm of computational biology and neuroscience have started making progress toward devloping systems that model specific organisms or neural circuits, such as the nematode Caenorhabditis elegans (C. elegans) (Izquierdo and Beer, 2016), though these systems have the potential to require too much focus on organism–specific details to achieve proper functionality, shifting focus away from creating more generalized neurologically–inspired intelligent systems.

On the other hand more traditional (application focused) AI research has started taking more inspriation from human learning, such as developing an auto–encoder augmented by Hebbian learning, decreasing the need for an initial superivised–like learning period (Bowren et al., 2016). Further, Marblestone et al. (2016) discusses ways that artificial neural networks (ANNs) can more closely approximate neural functionality. In the context of biologially–inspired AI, the frameworks underlying these approaches may be too constraining for full exploration of the potential for that field of study.

Recent work at the intersection of these two areas includes Sinapayen et al. (2016), which investigates the applicability and biological plausibility of spiking neural networks learning by "stimulation avoidance". Perhaps the project most closely aligned to Ortus is a biologically inspired neural network modeled off of a honey bee's visual system, which merges biological mechanisms and neural networks (Roper et al., 2017).

Ortus is an initial implementation of, and framework for creating, virtual life aimed at approximating the intelligence of living organisms. Bourn from the study and analysis of C. elegans' connectome and behavior, it aims to strike a balance between biological abstraction, retention of biological fidelity, and computation scalability in order to as closely as possible approximate biological intelligence and learning. At its core, Ortus is a network of biologically–inspired, non–spiking neurons, capable of forming excitatory, inhibitory, and electrical synapses. Similar to the way the structure of C. elegans' 302 neuron nervous system is capable of complex behaviors including toxin avoidance, reflexively withdrawing from a "tap", and "remembering" the temperature it found food (Jarrell et al., 2012), Ortus' "connectome" (neural structure) enables its inherent functionality. Once running, Ortus refines its network—similar to the way organic nervous systems adjust themselves, based upon Ortus' intrinsic "understanding" that certain things are "good" and others are "bad", with regard to its own longevity. This understanding is derived from the structure of the nervous system it generates for itself from a set of input definitions given to it.

The remaning sections of this paper outline Ortus' design and implementation, describe an initial experiement, discuss the implications of this framework, and analyize its shortcomings.

System Design

As Ortus aims to be a virtual analogy to intelligent life, we tried to only implement functions that either had a known analogous biological process, or which may have an analogous biological implementation that is unknown, but can be defended based off anecdotal evidence. Following each Ortus design element (ODE) below, is its biological rationale (BR):

1)

As stated by Verma et al. (2015), "emotions, motivations, and reinforcement are a closely related, evoluntionarily–conserved phenomena maintaining the integrity of an individual and promoting survival in a natural environment". Which, along with Gore et al. (2015), and other anecdotal evidence, suggests that building a virtual organism driven by emotional states is a sound approach. In Ortus the idea of "emotions", are simply the rise and fall in activation of different neurons or groups of neurons, tied to very fundamental behaviors—such as "breathing"—as seen in **(NOTE: FIG SOMETHING GRAPH!)**. The concepts of "good" and "bad" sensations or emotions only carry meaning to us because of their associations to circuits that are either fundamentally desirable or undesirable from a longevity/survival perspective.

It is well known that emotions are a driving force in motivation and behavior of intelligent organisms (cite). While at the human level, the emotional "part" of the brain is quite complex, it is not unreasonable to assume that as organismic complexity (and thereby intelligence) decreases, the complexity of emotions decreases. If one follows this line of thought, the possibility emerges that organisms like C. elegans may, in fact, be driven by "emotions" as well. For example, C. elegans is capable of toxin avoidance, a tap–withdrawal response, as well as learning that it found food at a certain temperature (citations?) – one must ask how this can be. There is nothing external that tells it what is good vs. bad, but it wants to avoid certain things, while being attracted to others. The premise behind Ortus, is that these behaviors are a result of a *very* simple emotional subsystem that forms the basis for C. elegans' behavior.

While Ortus' current implemention is a vastly simplified approach to a virtual organism, there are a few underlying principles. First, its core, least mutable circuit, is its respiratory circuit, which naturally cycles between inhalation of O_2, and exhalation of CO_2 (show graph). Second, as a part of its "life", a constant amount of O_2 is decreased, while a constant amount of CO_2 is introduced—this causes inhalation, which increases O_2. Third, all sensory inputs, are, after input consolitation, tied to all emotional centers that exist. The `FEAR` center, has a causal relationship between increased CO_2 and increased `FEAR` activation. This enables associative learning between an input, and an emotional state.

Synapic plasticity, or Hebbian learning, as well as (find out the opposite to it...) are implemented. Synapses strengthen when a presynaptic neuron and postsynatptic neuron or muscle activate (relative to the synapse's mutability index), and weaken gradually according to (insert equation) both over time, and when the presynapic neuron reaches its threshold, but the postsyaptic element (neuron or muscle) does not. Synaptic habituation

Ortus currently implements Hebbian learning, and the Stentian extension to Hebbian learning, as described by Kutsarova et al. (2016). Specifically, for each chemical synapse, on each kernel iteration, if activity is sufficiently synchronous, the synapse strengthens according to **PUT EQUATION HERE**, and if it is sufficiently asynchronous, it weakens according to **PUT EQUATION HERE**.

the ORT filetype is a specification of "rules", akin to very simplified "virtual DNA" that define elements (neurons and muscles), and relationships between them, currently: A causes B, A is correlated with B, A opposes B, and A dominates B. Each realtionship has a number of paramters, such as Age, Mutability, Threshold, and either a chemical synapse (CS) weight (unidirectional), or a gap junction (GJ) weight (bidirectional) (NOTE: equations!). Ortus' neurons are based upon C. elegans' neurons, and are non–spiking, but do not transmit any "activation" (footnote about why activation is used instead of voltage? or use 'potential?) below a given threshold, as in (Graubard, 1978). The equations for CS and GJ synapses were derived (simplified) from Wicks' 1996 paper (cite), and appear to have been used in recent C. elegans' models (cite Beer's 2013).

Ortus bulids its connectome based upon the rules specified in a supplied ORT file, based upon guidelines that aim to approximate the way neural circuits form in organic brains.

It is essentially wired to behave in certain ways, similarly to mammals, and how genes cause circuits to form in certain ways via gene expression (**?**). Schröter et al. (2017) provides evidence for the existence of organizational "motiefs" found in C. elegans that may underly more complex neteworks in larger brains; this also lends credence to Ortus' rule–based development approach.

In an analgous manner, the instructions in the ORT filetype, such as:

1. **element:** `sO2:` `{type=sense, affect=pos}`
2. **element:** `sCO2:` `{type=sense, affect=neg}`
3. **element:** `eFEAR:` `{type=emotion, affect=neg}`
4. **element:** `ePLEASURE:` `{type=emotion, affect=pos}`

5. **element:** `mINHALE: {type=motor}`
6. **element:** `mEXHALE: {type=motor}`
7. **element:** `LUNG: {type=muscle}`
8. `sO2` **opposes** `sCO2`
9. `mINHALE` **opposes** `sCO2`
10. `+A` **causes** `+B`
11. `A` **dominates** `B`

¡++¿

Sensory stimulation is sensed by sensory neurons that are topographically arrainged, similar to those in organic systems. A recent study suggesting that the brain is organized based upon an n = 2something... and something about permutations lends credence to this approach (cite n something 2 permutation paper).

ODE: Emotional learning is achieved by strengthening synapses of interneurons that form junctions between consolidated sensory information, and interneurons that are tied to each emotional "center" by GJ connections. The GJ connection allows an emotion to propagate over the entire system once activated. The effect of this, is that if a certain sensory input is causing the activation of a given emotion, the introduction of another sensory input will cause the synapses at the junction of the newly introduced sensory input and elevated emotion to strengthen. (graph here)

BR: Specifically for the mammalian opamine circuit, there exists a small locus of dopaminergic cells, but the they recieve inputs from diverse sources, as well as project to divers parts of the brain, as discussed by **?**. From a psychological perspective, it is clear that when one enters an emotional state, the emotion is encompassing, not localized. It is also clear that a stimulus or activity is likley to be "colored" by the emotional state one was in when the stimulus occurred.

ODE: Each interneuron that acts as an "emotion extension interneuron" (EEI) that branches away from the primary neuron for a given emotion, is connected to its primary emotion neuron via a GJ. It also has a CS connection back to the sensory consolidatory interneuron that has a CS into it This creates an emotional feedback circuit, allowing various emotional states to cause Ortus to "remember" other things that are related to that emotional state, by *slightly* activating the sensory consolidatory interneuron that would trigger the emotional response in the presence of the actual stimulus. In Ortus' current state, to "remember" means a given neuron or neural circuit has a measurably elevated activation, as a result of stimulation by an interneuron that would not have caused an increase in activation had associative learning not occured (diagram of S -¿ SEI -¿ SCI =¿ EEI —-— E thing).

BR: Gore et al. (2015) suggests that associative learning is funneled through innate behavior circuits to assign positive or negative emotions to neutral sensory stimuli. **(NOTE: This is precisely what the experiment below shows)**. Also psychological...

ODE: Sensory consolidatory interneurons have weights set initially such that the total incoming weight to any given SCI is the same; so the weight of any individual presynapse is $\frac{1}{\#inputs}$. This allows simple snesory, error–less sensory consolidation. However one current issue is if we have SCIs for sensory neurons `A`, and `B`, we will have `iA`, and `iB`, and `iAB`. If `iAB` is activated, that means that `iA` and `iB` will also be activated. A solution might be for more complex SCIs to inhibit less complex ones, though that would perhaps unneccesarily add to computational complexity.

BR: Barral and D Reyes (2016) suggests that synaptic strength scales inversely with the number of connections, "K as $\frac{1}{\sqrt{k}}$".

Talk about the combinatorial explosion, but then discuss how that can be pruned away... similar to babies. The $2^n - 1$ paper suggested reducing the resolution of inputs, so, instead of each sensor being able to connect with every other sensor, funnel 3 sensors (for example) into 1, to create a "functional connectivity module" (FCM). That approach might not be bad, however, it could be severely limiting. Perhaps this can be cut down by looking at relationships between sensors, and getting rid of things that don't matter?

Choices are: 1) an array of 9 sensors, if 0,1,2 ¿ A, 2,3,4 ¿ B, 4,5,6 ¿ C, 6,7,8 ¿ D, then sum nCr(k,n) for n from 1 to k is $2^n - 1$. It is clear that to get a fully connected network, capable of expressing all possible combinations, even from a living organism perspective, things become unfeasible.

2) Rather, it is probably the case that our brains aren't quite as connected as they appear to be. (e.g., within a certain radius, two places on your skin feel like the same location. visual system either might be an exception, however it is well known that visual information gets interpolated.) There must be some sort of neural processing shortcuts ... approximation. e.g., if you are quickly reading, you might read "New York", when the only word there was "York"

There is evidence that neighboring neurons in the same "layer" are connected in C. elegans (Azulay et al., 2016), so it is not unreasonable to assume a certain amount of information consolidation occurs.

Schröter et al. (2017) suggests that C. elegans' neurons may have multiplexed functions, meaning that one neuron may contribute to more than one behavior. This is another possible way to decrease the number of necessary neurons in Ortus.

floor so that learned stuff can't be unlearned beyond a certain point – e.g., as a relationship ages, it's mutability should tend toward 0.

ODE: Certain emotional states preside over (dominate) others. In Ortus, fear dominates pleasure.

BR: The idea of a hierarchical system, where certain emotional states dominate others, is supported by the inhibition of fear in mice, in favor of searching for food when hungry (blood glucose levels falling causing the release of hormones), but greater concern for safety when not hungry (Verma et al., 2015). Further, Leknes and Tracey (2008) discusses the Motivation–Decision Model, which suggests that anything that is more important for survival than pain should inhibit the feeling of pain.

The chemical synapse and gap junction activation transfer equations were simplified from those described by Wic (1996) to ignore physical properties of neurons (such as neuron length).

? talks about topographic organization

Implementation Details

Ortus is written in C++ and OpenCL. Each iteration of the OpenCL kernel constitutes one timestep, during which, each neuron sums incoming (positive or negative) "activation" from presynaptic cells via chemical synapses (CSes) and gap junctions (GJs). **NOTE: maybe this is a good spot for the eqations!**

Implementation Details - C++, OpenCL - each neuron operates on its incoming connections, so it only does "postsynaptic" work, leaving its "presynaptic" work to neurons that it is presynaptic to (see "shortcomings") - sensory consolidation causes resolution loss, currently maximum resolution for any sensor array (e.g., not just O2) has a maximum resolution of 2 for intra–sensor consolidation, and 3 for inter–sensor consolidation. Single neuron sensors have no minimum for either. (NOTE: sensor array not implemented yet... take this out if it doesn't get implemented)

Discussion

While the implementation presented is quite simple, the initial results presented pave the way for quick advances to Ortus' capability with regard to "behaviors" resulting from emotional amalgamation as well as the introduction of more complex sensory input.

While Ortus' eventual goal is to develop complex "neural" functionality akin to that observed in complex systems, we

Kutsarova et al. (2016) indicates that, relative to synaptic strengthening and weakening, axonal branch tips emerge to form new synapses. synchronous activity stabilizes synapses and prolongs axonal branches. This is one way for Ortus to alter its structure in addition to synaptic weights.

? suggests that in mice, while synchronous activity (correlation) is necessary for synaptic plasticity, it is not sufficient; neuromodulatory signaling is also required. This could be a way to control some of the unstable behavior seen as a result of the purely correlation–based learning rules implemented.

Need to implement habituation

From an artificial intelligence standpoint, the

1. less connections
2. ability to specify initial behavior
3. don't need a massive traning set
4. 3 types of interactions: inhibitory CS, excitatory CS, and GJ.

Together, these afford artificial neural networks far greater flexibility and thereby power. It is possible that this will enable more nuanced behavior.

From an artificial life standpoint: people are modeling networks with less complexity than this to attempt to recreate C. elegans' behavior from its connectome (cite beer.. and newer). This work pivoted from that approach, because in attempting to model the entire connectome, it became clear that there were too many unknowns. However, it also became clear that there was a certain logic to the way the neurons were wired; a notion backed up by (cite c elegans wiring works... have some in mendeley).

A simple idea is, how can C. elegans ¡DO SOMETHING¿ – point here is to describe the requirement that certain things inhibit while others excite, in order to allow a certain behavior to happen.

Shortcomings and Issues. There is no temporal differentiation between CS and GJ transmission. On each timestep, neurons collect all incoming "activation" from.

There is minimal presynaptic processing – (NOTE:how true is this?)

No location–based axon growth, as is seen in, relative to hebbian plasticity (cite)

Conclusion

We have presented Ortus, an initial approach to a framework that exhibits basic emotionally–based learning (fear conditioning), self–sustaining inherent behavior (cyclic breathing), and a topographical approach to sensory processing that enables intersensory associative learning, with or without an emotional component.

Ortus also has a system to map sensory inputs such as "breathe" to the stimulation of O2 sensors.

Acknowledgements

This work was supported by NSF grant No. PHY-9723972.

References

(1996). A Dynamic Network Simulation of the Nematode Tap Withdrawal Circuit : Predictions Concerning Behavioral Criteria of the Nematode Tap Withdrawal Synaptic Function Using. 16(12):4017–4031.

Azulay, A., Itskovits, E., and Zaslaver, A. (2016). The C. elegans Connectome Consists of Homogenous Circuits with Defined Functional Roles. *PLOS Computational Biology*, 12(9):e1005021.

Barral, J. and D Reyes, A. (2016). Synaptic scaling rule preserves excitatoryinhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, 19(12):1690–1696.

Bowren, J., Pugh, J., and Stanley, K. (2016). Fully Autonomous Real-Time Autoencoder-Augmented Hebbian Learning through the Collection of Novel Experiences. *Proceedings of the Artificial Life Conference 2016*, (Alife 15):382–389.

DARPA (2017). Toward machines that improve with experience.

Gore, F., Schwartz, E. C., Brangers, B. C., Aladi, S., Stujenske, J. M., Likhtik, E., Russo, M. J., Gordon, J. A., Salzman, C. D., and Axel, R. (2015). Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. *Cell*, 162(1):134–145.

Graubard, K. (1978). Synaptic transmission without action potentials: input-output properties of a nonspiking presynaptic neuron. *Journal of Neurophysiology*, 41(4):1014–1025.

Izquierdo, E. J. and Beer, R. D. (2016). The whole worm: Brain-body-environment models of C. elegans. *Current Opinion in Neurobiology*, 40:23–30.

Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., and Emmons, S. W. (2012). The Connectome of a Decision-Making Neural Network. *Science*, 337(July):437–444.

Kutsarova, E., Munz, M., and Ruthazer, E. S. (2016). Rules for shaping neural connections in the developing brain. *Frontiers in Neural Circuits*, 10(January):111.

Leknes, S. and Tracey, I. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience*, 9(4):314–320.

Marblestone, A. H., Wayne, G., and Kording, K. P. (2016). Towards an integration of deep learning and neuroscience. *bioRxiv*, 10(September):1–61.

Roper, M., Fernando, C., and Chittka, L. (2017). Insect Bio-inspired Neural Network Provides New Evidence on How Simple Feature Detectors Can Enable Complex Visual Generalization and Stimulus Location Invariance in the Miniature Brain of Honeybees. *PLOS Computational Biology*, 13(2):e1005333.

Schröter, M., Paulsen, O., and Bullmore, E. T. (2017). Micro-connectomics: probing the organization of neuronal networks at the cellular scale. *Nature Reviews Neuroscience*, 18(3):131–146.

Sinapayen, L., Masumori, A., and Ikegami, T. (2016). Learning by Stimulation Avoidance: A Principle to Control Spiking Neural Networks Dynamics. page 17.

Verma, D., Wood, J., Lach, G., Herzog, H., Sperk, G., and Tasan, R. (2015). Hunger Promotes Fear Extinction by Activation of an Amygdala Microcircuit. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 41(2):431–439.