Exploring Coevolution using the OpenAI Gym Multiagent Particle Environment and NEAT

Joseph Pickens

An independent study under the advisement of Jordan Pollack
DEMO Lab, Brandeis University

Spring 2021

Abstract

Previous work in the field of Coevolution has established that interactions between multiple competing and cooperating populations can drive the evolution of complexity. This paper presents initial work to replicate such findings within a 2D spatial environment. I adapt the NeuroEvolution of Augmenting Topologies (NEAT) algorithm for use in a coevolutionary context and integrate it with the OpenAI Gym Multiagent Particle Environment to coevolve neural network topologies and visualize the evolved behavior.

1 Introduction

Evolutionary algorithms are used within the fields of Artificial Life and Artificial Intelligence to find solutions to various problems through hill climbing a fitness landscape. Such algorithms of course draw on biological evolution for inspiration. Importantly, however, the fitness landscapes that biological organisms are embedded within are not static. The environment of an organism is comprised at least in part of other organisms, all of which are dynamically co-evolving. As one organism changes, the organism's surrounding ecosystem adapts to the change, and vice versa, in an ongoing dance. This dance involves competition, as in the evolutionary arms race, but also cooperation, as in the multitude of symbiotic relationships seen throughout nature.

Those who study open-ended evolution take seriously the idea that the open-ended nature of this evolutionary dance is of crucial importance to the emergence of increasingly complex forms over time, and to the emergence and function of intelligence itself.

In this paper I briefly cover the background work in the field of Coevolution that investigates the conditions under which the open-ended evolution of complexity is possible, and I present my initial work to expand this investigation

within a simulated domain closer resembling that of the biological, integrating the OpenAI Gym Multiagent Particle Environment with the NEAT algorithm to do so. I share the results of my experiments as well as some suggestions for ways to build on my work moving forward.

2   Background

In his 2018 dissertation, Nick Moran shows that interactions between multiple competing and cooperating populations can drive the evolution of complexity beyond what is possible in either strictly competitive or strictly cooperative interactions alone [2]. Moran defines six different "ecosystems" through which he investigates this idea (Figure 1). Each ecosystem specifies the number of species (distinct populations) and the types of interactions between those species, either competitive or cooperative. Agents within these species are represented using finite-state machines. This provides Moran with a useful complexity metric for each agent, measured as the number of states in the minimized finite-state machine, by which to assess complexity growth across time. Moran pairs up agents from each population in a simple 2D matrix prediction game.

Results from Moran's experiments quantitatively show that at least one species in the 3-species mixed and 4-species mixed ecosystems exhibits a much higher rate of ongoing complexification relative to species in the ecosystems in which strictly one type of interaction is present.

3   Creating 2D Spatial Ecosystems

The goal of this independent study is to translate Moran's ecosystems to a 2-dimensional spatial context. This brings the experimental context into closer continuity with a biological one, in which agents physically move about in an environment so as to achieve certain goals. This also affords a more intuitive and qualitative means for assessing the complexity of an agent through the simple observation of its mobile behavior within its environment. I accomplish this by integrating the Gym Multiagent Particle Environment with the NeuroEvolution of Augmenting Toplogies (NEAT) algorithm, including a special Ecosystem class that adapts NEAT for use with Moran's ecosystem framework.

3.1   Multiagent Particle Environment

The Multiagent Particle Environment from OpenAI inherits from the more general OpenAI Gym environment class, adapting it for multiple agents [1]. The environment is a small 2D world that includes a basic physics engine to simulate the dynamics of agent momentum and collisions. Important for the purposes of this study, this environment can be rendered to visualize the evolved behavior of my agents.

At each time step, agents in the Multiagent Particle Environment take as input an observation from their environment and output an action, and are

each assigned a payoff based on the resulting state of the agents and their environment. For my coevolutionary context, the cumulative payoff over all time steps and all agent interactions gets translated into a fitness score, which then differentially impacts the agents' chances of reproduction.

3.2   Agent Representation and NEAT

Moran uses a finite-state machine representation for his agents, which has the benefit of a useful complexity metric. But while such a representation may work for Moran's discrete, binary prediction games, the continuous and complex nature of my 2D spatial world renders such a representation computationally intractable for my purposes.

I instead use a neural network representation for my agents, and resort to evaluating behavioral complexity qualitatively. To evolve the neural network topologies, I use the NeuroEvolution of Augmenting Topologies (NEAT) genetic algorithm [3]. NEAT's proven effectiveness depends on three important factors. The first is that incremental mutations affect not only the network weights, but its structure as well, thus allowing for open-ended complexification of the network topology. Secondly, NEAT uses a sequence encoding for the neural networks, with history markers that enable a coherent means for genetic crossover. Lastly, NEAT allows for speciation within the evolving population, which has been shown to be essential for preserving innovations within the population.

3.3   The Ecosystem Class

In keeping with Moran's ecosystem framework, I create an Ecosystem class which adapts NEAT for evaluating fitness in a coevolutionary context, that is, where multiple populations are interacting and evolving in tandem. The class is parameterized by a list of multiagent environments, a list of agent populations, and a list for specifying which populations will be evaluated for fitness within each environment. Together, this is sufficient to define each of the six ecosystems in Figure 1.

4   Initial Experiments

4.1   Global Observation Scheme

I experimented with two kinds of observation schemes over the course of my study. The first I refer to as the "global" observation scheme. In this scheme, each agent observes at each time step the relative x- and y-coordinates of all other entities in its environment. This gives each agent a "god's-eye view" of its environment.

I used a pursuer-evader game to define the competitive interaction. In this game, one of the agents is the pursuer, the other the evader. The pursuer is assigned a positive fitness payoff proportional to its proximity to the evader. The evader loses payoff if brought into contact with the pursuer, but gains

payoff proportional to its proximity to a special landmark. If in contact with the landmark, the evader is "safe" from the pursuer, so long as it is able to maintain its position.

For the cooperative interaction, I define a game where agents need to coordinate to occupy two landmarks. The agents are rewarded for occupying both landmarks simultaneously, and receive a lesser reward for occupying only one landmark between the two of them. Ultimately, both agents are assigned the same fitness score, equal to their cumulative reward.

For these experiments, a simple feed-forward network was used. The network was initialized with no hidden nodes, and with a probability of 0.5 that a given connection between an input and an output node would be initially present.

The global observation scheme together with the feed-forward network yield a relatively straightforward mapping from agent observations onto actions. For both the 2-species competitive and 2-species cooperative ecosystems, there is a clear gradient for each agent to follow to a point of roughly optimal behavior, after which not much of interest occurs. These results can be found in the supplementary videos (Animations 1 and 2, respectively). All animations showcase the best genomes from each evolved population playing a series of games, each lasting 100 time steps.

Next, I consider the 3-species mixed ecosystem. It became clear early on in running these experiments that the results for the 3-species ecosystem hinge entirely on whether the pursuer or the evader is evaluated also in the cooperative game, given the asymmetry between the two agents in the competitive game. If the evader is evaluated in the cooperative game, it does well by simply continuing its strategy of moving to a landmark. These results are shown in the supplementary Animation 3. The pursuer, however, has the almost impossible task of, on the one hand, catching the evader in the competitive game, and, on the other, moving to a landmark in the cooperative game. As seen in Animation 4, the pursuer (red) struggles to find an optimal behavior for both games. Doing so would require the pursuer evolving some means by which to discern which game it is playing. This could presumably occur through some means of agent signaling in the cooperative game. That being said, if agent actions are restricted to spatial movement only, some form of memory and time-dependent action is necessary for any meaningful signal to be conveyed to the other agent. This is an important point that I will return to.

4.2 Local Observation Scheme

The second set of experiments involved a "local" observation scheme. This scheme was devised for the sake of greater biological plausibility in my experiments. Organisms in nature do not observe their environment from a god's eye perspective. Rather, they are stuck with a limited frame by which to sense the world around them. As a result, organisms must actively move around and explore in order to perceive various features of their environment.

In the local observation scheme, the agent can sense the proximity of only those entities that are within a 120° observation window, and within a 0.5-

meter range. The observation window is divided into 8 observation "wedges," 15° each, so as to give higher resolution to the agent's sensory domain. An additional action is incorporated in this scheme to allow for the agent to turn either clockwise or counterclockwise (or to not turn at all).

Initial experiments with this scheme resulted in stationary agents that simply turned incessantly in one direction, moving only if they happened to have been spawned sufficiently close to a relevant stimulus. One reason for this may have been the high risk of exploration. There are many directions an agent can explore, including off the map altogether, many of which result in substantially lower fitness than simply remaining stationary. To solve this, I first introduced walls around the border of the environment. Rather than wander off the map (and be penalized for it), agents simply bounce off the walls, thereby putting a helpful constraint on the search space. Additionally, walls were included as observable entities by the agents. However, walls alone were insufficient to incentivize exploration. One explanation for this is that in the absence of all stimulus, the only signal propagating through the feed-forward network is due to the network's bias. This severely limits the potential of the network's output.

To address this issue, it is necessary to incorporate some kind of time-dependent action, just as would be necessary for agent signaling to evolve. There are multiple ways of accomplishing this. The first is to maintain a feed-forward representation and to simply expand the observation space to include a time-dependent feature of the agent or its environment, such as observations or actions from previous time steps. The second option is to abandon the feed-forward representation for a recurrent one. A recurrent neural network has a state that changes over time depending on its input. Fortunately, NEAT allows for recurrent connections in its neural networks.

My experiments using networks with recurrent connections yielded the most promising results in terms of agent exploratory behavior. For both the 2-species competitive and 2-species cooperative ecosystems, one agent evolved to explore, as can be seen in Animations 5 and 6, respectively. However, the second agent remains stationary, and subsequent experiments proved this behavior to not evolve consistently. Increasing the probability of recurrent connections could be a solution to this problem. Unfortunately, there is no built-in means within NEAT by which to adjust the mutation rate for recurrent connections specifically. Future work will need to adapt NEAT for this purpose before experimenting with the 3-species and 4-species ecosystems.

5   Conclusion

Integrating the Multiagent Particle Environment with NEAT allows for a promising way to visualize and qualitatively assess the evolution of behavioral complexity in a coevolutionary context closely resembling that of the biological world. Initial experiments show that evolving meaningful agent behavior is possible within this environment, but also reveal the limitations of strictly feed-forward agent representations. Perhaps the most important lesson learned from my

experiments is that complex behavior requires some form of time-dependent action, that is, memory. This is the advantage of Moran's finite-state machine representation, and the disadvantage of a strictly feed-forward neural network representation. It is important that future work on this project start with this in mind, as meaningful results from experiments on Moran's ecosystems will depend on it. While recurrence in NEAT is possible, it will need to be adapted in future work so as to include parameters for fine-tuning the probability of recurrent connection mutations directly.

References

[1] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6382–6393, Red Hook, NY, USA. Curran Associates Inc.

[2] Moran, N. (2018). Complexity in coevolution and deep learning.

[3] Stanley, K. O. and Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evol. Comput.*, 10(2):99–127.
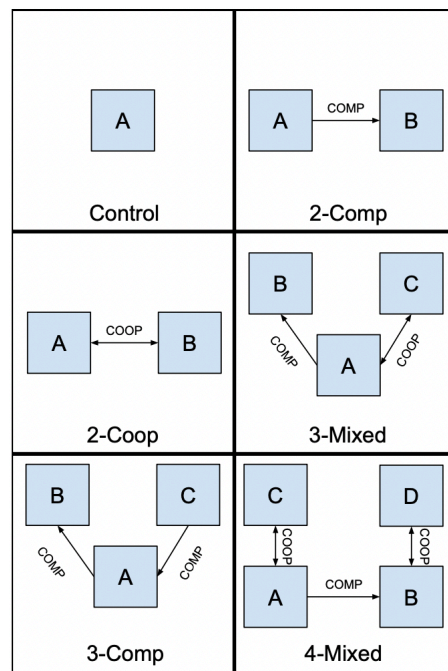
Figure 1: Interactions between species in Moran's six ecosystem models. Connections between species indicate an interaction, which is labeled as either competitive or cooperative.