Evolving Neural Network Agents to Play Atari Games with Compact State Representations

Adam Tupper
Department of Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
adam.tupper@pg.canterbury.ac.nz

Kourosh Neshatian
Department of Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
kourosh.neshatian@canterbury.ac.nz

ABSTRACT

Recent success in solving hard reinforcement learning problems can be partly credited to the use of deep neural networks, which can extract high-level features and learn compact state representations from high-dimensional inputs, such as images. However, the large networks required to learn both state representation and policy using this approach limit the effectiveness and benefits of neuroevolution methods that have proven effective at solving simpler problems in the past. One potential solution to this problem is to separate state representation and policy learning and only apply neuroevolution to the latter. We extend research following this approach by evolving small policy networks for Atari games using NEAT, that learn from compact state representations provided by the recently released Atari Annotated RAM Interface (Atari ARI). Our results show that it is possible to evolve agents that exceed expert human performance using these compact state representations, and that, for some games, successful policy networks can be evolved that contain only a few or even no hidden nodes.

ACM Reference Format:
Adam Tupper and Kourosh Neshatian. 2020. Evolving Neural Network Agents to Play Atari Games with Compact State Representations. In *Genetic and Evolutionary Computation Conference Companion (GECCO '20 Companion), July 8–12, 2020, Cancún, Mexico.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3377929.3390072

1 INTRODUCTION

Before the adoption of deep neural networks for reinforcement learning, the only solvable problems were those for which low-dimensional, high-quality state representations could be constructed. Now, deep reinforcement learning has led to success in solving more complex problems, such as video games, by combining feature extraction and state representation learning with policy learning. This end-to-end learning approach removes the need for feature engineering, but, due to the large networks required, rules out many topology and weight evolving neuroevolution algorithms, such as NEAT [5], that are effective at finding solutions for simpler domains.

GECCO '20 Companion, July 8–12, 2020, Cancún, Mexico

ACM ISBN 978-1-4503-7127-8/20/07…$15.00
https://doi.org/10.1145/3377929.3390072

One potential solution that could allow us to exploit the benefits of both deep neural networks and evolutionary reinforcement learning is to separate end-to-end learning into two components: state representation learning and policy learning. With this approach, gradient-based methods, such as auto-encoders, might be used for feature extraction and for learning condensed state representations, enabling evolutionary methods to be used for policy learning.

In this work, we expand on prior work combining deep learning and neuroevolution [1, 4] by investigating the plausibility of evolving Atari agents using NEAT that learn to play from compact state representations, similar to those we might expect to be learned via state representation learning. The state representations we use are provided by the recently released Atari Annotated RAM Interface (Atari ARI) [2]. This interface identifies the specific bytes of RAM that store states variables that are important for playing each game, reducing the size of the input space by up to 93% when compared to using the entire contents of RAM.

2 METHOD

This section provides an outline of our method. Our source code, and the full details and discussion of our experiments and results, can be found at evolvingatari.adamtupper.nz.

2.1 State Representations and Game Selection

We use 14 of the 22 games supported by the Atari ARI in our evaluations. We selected this subset of games because our inspection of the objectives for each game revealed inadequacies in some of the state representations provided. We categorise the supported games into three categories: *poor*, *fair* and *good*; based on the perceived quality and completeness of the state representations provided by the Atari ARI, and use only games with either *fair* or *good* representations in our evaluations. These representations are deemed to include enough information to develop an adequate or optimal strategy from respectively.

2.2 Proposed Neuroevolution

We use our own implementation of NEAT to evolve time-delayed recurrent neural networks that propagate the outputs of each neuron forward at each time step. Although this means the initial outputs are not influenced by the inputs, these networks do not require a topological ordering of nodes or the specification of whether lateral connections are recurrent. Our evolved networks also include explicit biases. To better fit the Atari domain, we also adapt NEAT to support negative fitness values. This is achieved by using the difference between the individual's fitness and the lowest fitness in the population as an adjusted fitness value.

2.3 Experimental Setup

A single set of hyperparameter values were used to evolve a separate policy for each game. These were chosen based on informal experimentation on a subset of three games: Asteroids, Boxing, and Pong. For each game, three evolutionary runs, each lasting 200 generations, were performed, using a population size of 130. Each agent's fitness was defined as the average total accumulated reward over three episodes of gameplay. The length of an episode was capped at 20,000 frames (approximately five minutes of real-time gameplay) to ensure that all episodes terminate. The best policy from each run was evaluated for 100 episodes, and the policy with the highest average reward is reported in our results. The human-normalised scores are calculated using the expert human scores published alongside DQN [3]. The only exception to this is for Berzerk, a game not included in their evaluations. For Berzerk, we use the record human score listed on www.twingalaxies.com.

3 OVERALL PERFORMANCE

Fig. 1 shows the human-normalised performance of each of the best agents; as can be seen, performance varies substantially between games. While the agents for Video Pinball, Boxing and Bowling exceed or match expert human performance, for other games the agents perform far worse. Although most of the best performing games (Boxing, Bowling and Freeway) have *good* state representations, the highest performing agent was found for Video Pinball. This illustrates that for some games, good strategies can still be discovered with imperfect information.

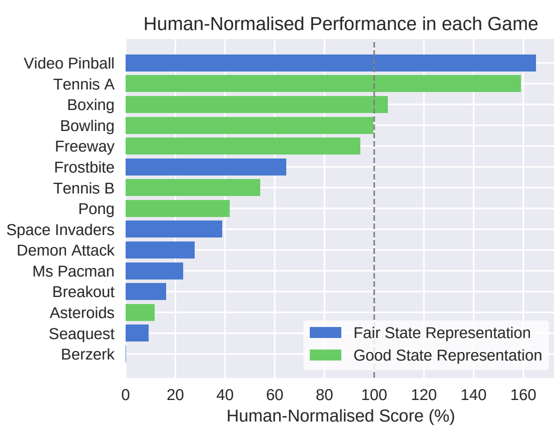

Figure 1: The human-normalised performance of the best agent for each game. The bar colour denotes the quality of the state representation provided by the Atari ARI.

When initially evolving policies for Tennis, we found that the population quickly converged on a strategy of waiting for the frame cap to be reached by refusing to serve the ball, because early strategies performed poorly and accumulated negative rewards. To address this, we set the reward for agents that reached the frame cap to the minimum possible reward (-24). We report the results for the original and modified settings as Tennis A and B, respectively.

4 EVOLVED ARCHITECTURES

Inspecting the architectures of high-performing agents reveals some surprising simplicity. None of the solutions to any games evolved many hidden nodes (the maximum was six for Asteroids), even the networks for high-performing solutions were very simple. This may have been a consequence of the particular hyperparameter values chosen, but it shows that simplicity is not the sole explanation of poor performance. The best-performing agent for Freeway (shown in Fig. 2) epitomises this simplicity, utilising only a single hidden node. Interestingly, the inputs for some cars are disabled, indicating that this information can be derived from the positions of the others.

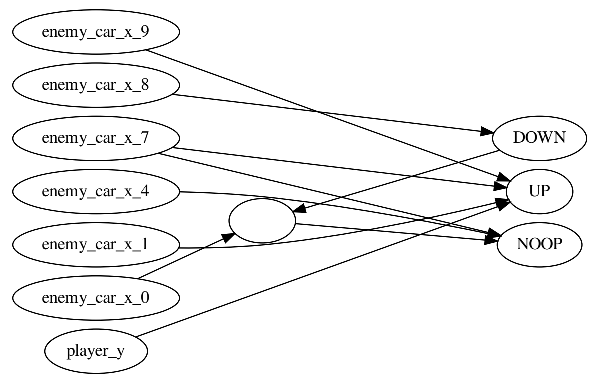

Figure 2: The evolved network architecture for the top performing Freeway agent. Unused inputs are excluded.

5 CONCLUSIONS

Although evolved policies only exceeded or were competitive with expert human performance in a handful of games, we discovered that surprisingly simple and small neural networks could play these games effectively. We plan to extend our work by incorporating NEAT in a separated state representation and policy learning framework to evaluate the overall effectiveness of this approach.

REFERENCES

[1] S. Alvernaz and J. Togelius. 2017. Autoencoder-Augmented Neuroevolution for Visual Doom Playing. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. 1–8.
[2] A. Anand, E. Racah, S. Ozair, Y. Bengio, M. A. Côté, and R. D. Hjelm. 2019. Unsupervised State Representation Learning in Atari. In *Advances in Neural Information Processing Systems 32*. 8766–8779.
[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. 2015. Human-Level Control through Deep Reinforcement Learning. *Nature* 518 (Feb. 2015), 529–533.
[4] A. P. Poulsen, M. Thorhauge, M. H. Funch, and S. Risi. 2017. DLNE: A Hybridization of Deep Learning and Neuroevolution for Visual Control. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. 256–263.
[5] K. O. Stanley and R. Miikkulainen. 2002. Evolving Neural Networks through Augmenting Topologies. *Evolutionary Computation* 10, 2 (June 2002), 99–127.