Improving the Performance of Evolutionary Algorithms in Deep Reinforcement Learning via Gradient-Based Initialization
Chris Waites, Andrew Shoats, Matthew Prelich

**Introduction**

Gradient-based methods in RL are able to move quickly from regions of initialization to broad regions of lower loss
- Struggle to escape locally optimal solutions

Genetic algorithms place a heightened emphasis on global search and can reach superior solutions
- Take a long time to obtain reasonable candidate solutions

Idea: propose algorithm which takes hybrid approach, combining gradient-based methods in RL with gradient-free evolutionary algorithms

**Methodology**

N DQN models are trained using the following update rule to initialize GA population

\[ \theta_{t+1} = \theta_t - \nabla_{\theta_t} \left( Q_\theta(s_t,a_t) - (r_t + \gamma \max_{a'_t} Q_\theta(s'_t,a'_t)) \right) \]

GA evolves a population of N individuals through what are called generations
The best performing network in the generation is preserved for the next generation. A parameter selected uniformly at random from the top T performing networks is mutated by adding Gaussian noise
In effect, the best performing networks are passed down by generation and thus the networks keep improving as a function of generations

**Experiments**

**Findings**

Overall, our work introduces a hybrid algorithm which effectively surpasses DQN (with Adam optimizer) and GAs alone in the popular RL benchmark of CartPole terms of both expected reward and time complexity.

These results can be even further improved due to the highly parallelizable nature of the algorithm

A 100-fold decrease in the number of forward propagations per generation was seen (this should be more methodically analyzed in future work) in comparison to GA without DQN initializations

1. Apply methods to more challenging environments (Atari, MuJoCo, etc.)
2. Investigate generalization of proposed method which incorporate mix of gradient-based and non gradient-based approaches
3. Strongly supervised learning: Would the same positive results hold? Or would strictly gradient-based methods still reign?