Integrating Neural Pathways for Learning in Deep Reinforcement Learning Models

Varun Ananth
Paul G. Allen School of Computer Science, University of Washington
185 E Stevens Way NE
Seattle, WA 98195 USA
vananth3@uw.edu

Abstract

Considering that the human brain is the most powerful, generalizable, and energy-efficient computer we know of, it makes the most sense to look to neuroscience for ideas regarding deep learning model improvements. I propose one such idea, augmenting a traditional Advantage-Actor-Critic (A2C) model with additional learning signals akin to those in the brain. Pursuing this direction of research should hopefully result in a new reinforcement learning (RL) control paradigm that can learn from fewer examples, train with greater stability, and possibly consume less energy.

Introduction

An important challenge of human decision-making is determining via trial and error which options maximize reward and minimize punishment. In computer science, this problem is framed as reinforcement learning (RL), and particular RL models such as the advantage actor-critic (A2C) have been the subject of extensive research (Niv 2009). I am interested in studying the intersection of neuroscience and AI (NeuroAI) with deep reinforcement learning (RL) to create systems that learn in similar ways to our brain. The first experiment that I would conduct would be to add additional learning signals, like pain, to RL algorithms in the most biologically accurate way possible. This will improve the learning power of deep RL agents so they can be deployed in more complex scenarios. In addition, I hope to engineer solutions that enable agents to learn quickly with fewer examples.

Nociception is instrumental to the survival of complex organisms, and pain is an incredibly efficient learning signal. Though I do not seek to have computers experience pain and suffering, I hope to add new mathematical equations into the A2C model that have heavy basis in human nociception. The improvements to society are widespread. Smarter agents in autonomous robots could perform search-and-rescue missions in locations humans cannot explore safely. Improvements on bottleneck computations through smarter algorithm design discovered by deep RL agents can reduce the compute power needed for under-served communities to access the power of AI. NeuroAI in deep RL is a clear next step for the field of AI.

Background

Many widely used RL algorithms often do not go further than temporal-difference (TD) learning in their mimicry of the brain, despite the fact combining neuroscience and AI has historically improved the performance of these algorithms (Tassa et al. 2018). There are also recent rallying calls to the AI community to put efforts into NeuroAI research (Zador et al. 2023). The MaxPain architecture (Elfwing and Seymour 2017) begins to solve this problem, incorporating a second optimization signal as an *in silico* analog for pain. Its implementation, however, was a loose analogy for how pain is interpreted in the brain and can be better aligned through architecture modifications (Ananth et al. 2023). MaxPain improves the speed at which the agent converges on a path in a maze-game, and also improves its obstacle avoidance for real-world robotics. Clearly, integrating the knowledge gained from the field of neuroscience into RL models has been shown to improve them. Properly trained RL algorithms that *do not* have strong, encoded brain parallels for learning still show incredible performance. AlphaGo (Silver et al. 2016), and AlphaTensor (Fawzi et al. 2022) have displayed incredible creativity in their solutions to the "games" they were presented with. Fusing the power of deep RL and the human brain is sure to push the boundaries of AI technology further than ever before.

Prior Work by the Applicant

My prior work in Ananth et al. (2023) discusses the MaxPain architecture and possible improvements to it and other similar architectures. In this context, an "improvement" is a change that adjusts the model architecture to mimic the brain with greater accuracy. For example, one of my proposed changes was to use a neural network (possibly a recurrent neural network) to estimate a coefficient that affected how much an agent considered painful stimuli at its current timestep. The previous architecture (Wang, Elfwing, and Uchibe 2021) simply used a fixed temperature (τ_w) that determined how the agent would consider pain "important" at that timestep in the context of optimizing towards a goal. This does not mimic how humans weight pain and reward. The importance of one stimuli over the other is never fixed but rather dependent on past context. Adding a neural network that can modify this parameter would fix that problem.

Approach

My approach to creating a better *in silico* pain signal analog for deep RL systems stems from my aforementioned previous work. **I hope to recreate the augmented MaxPain architecture described in Wang, Elfwing, and Uchibe (2021) and replace the fixed τ_w hyperparameter with a recurrent neural network that considers past context to weight pain versus reward at a single timestep.** I plan to use my knowledge from a class I have taken at UW titled "Deep Learning". This class taught me a myriad of training tricks that I hope to implement into my RNN so that it produces useful coefficients for the agent to weight pain by. This project would also be a wonderful time for me to practice good deep learning model design and learn from an expert in the field of deep RL.

Evaluation

As I mentioned before, an application of my research would be in autonomous agents that can navigate complex and dynamic environments that humans would be unsafe in. To this end, I plan to evaluate the efficacy of my agent by having it solve 2D "painful" maze puzzles as shown in Elfwing and Seymour (2017). The model receives some kind of "painful" stimuli if it touches the walls of a maze. I will first see if the architecture I propose has any improvements over the MaxPain architecture, and also over a basic advantage-actor-critic model with respect to episodes until stability and wall avoidance. If the results are promising, I hope to create a series of dynamic "dangerous" mazes that have failure conditions (e.g pits in which the agent would "die" in) to see if the architecture modifications I made do improve agent performance in dynamic environments where threat levels fluctuate. If the agent I designed can reach the end of the maze while the others cannot (over a series of trials), then I will have succeeded.

Discussion

I hope to find that dynamic weighting of pain and reward results in more intelligent obstacle avoidance. It would also be interesting to see how the policies learned by the agents line up with humans navigating the same maze. If I observe and publish either of these results, I hope to see the field of deep RL (and more generally, deep learning) move away from larger compute and big data that makes these models inaccessible to anyone but the largest companies. I hope more deliberate model architectures that mimic the human brain become the norm, and that they process and learn from data with greater efficacy. In addition to autonomous rescue robots, this would benefit society by lowering the barrier for those without GPU clusters to train their own models for deployment.

Conclusion

Current deep RL architectures show great training instability (Nikishin et al. 2018) and do not find paths through environments that can safely be deployed into real-world robotics without additional safety measures (Elfwing and Seymour 2017). However, drawing inspiration from neuroscience by properly integrating a pain signal for the agent can theoretically improve learning speed and obstacle avoidance (Ananth et al. 2023). This will benefit society through better autonomous robotics and lower barriers to training RL models. Additionally, creating a model that can learn to weight pain on-the-fly will be a step towards the greater goal of making AI systems learn in the same way humans do – efficiently. This will have downstream effects by pushing AI over the "wall of compute" that it currently faces.

References

Ananth, V.; Inman, C.; Hong, J.; and Grieskamp, M. 2023. Deinforcement Learning. *Canadian Undergraduate Journal of Cognitive Science*.

Elfwing, S.; and Seymour, B. 2017. Parallel reward and punishment control in humans and robots: Safe reinforcement learning using the MaxPain algorithm.

Fawzi, A.; Balog, M.; Huang, A.; Hubert, T.; Romera-Paredes, B.; Barekatain, M.; Novikov, A.; Ruiz, F. J. R.; Schrittwieser, J.; Swirszcz, G.; Silver, D.; Hassabis, D.; and Kohli, P. 2022. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930): 47–53.

Nikishin, E.; Izmailov, P.; Athiwaratkun, B.; Podoprikhin, D.; Garipov, T.; Shvechikov, P.; Vetrov, D. P.; and Wilson, A. G. 2018. Improving Stability in Deep Reinforcement Learning with Weight Averaging.

Niv, Y. 2009. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3): 139–154. Special Issue: Dynamic Decision Making.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; and Hassabis, D. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484–489.

Tassa, Y.; Doron, Y.; Muldal, A.; Erez, T.; Li, Y.; de Las Casas, D.; Budden, D.; Abdolmaleki, A.; Merel, J.; Lefrancq, A.; Lillicrap, T.; and Riedmiller, M. 2018. DeepMind Control Suite. arXiv:1801.00690.

Wang, J.; Elfwing, S.; and Uchibe, E. 2021. Modular deep reinforcement learning from reward and punishment for robot navigation. *Neural Networks*, 135: 115–126.

Zador, A.; Escola, S.; Richards, B.; Ölveczky, B.; Bengio, Y.; Boahen, K.; Botvinick, M.; Chklovskii, D.; Churchland, A.; Clopath, C.; DiCarlo, J.; Ganguli, S.; Hawkins, J.; Körding, K.; Koulakov, A.; LeCun, Y.; Lillicrap, T.; Marblestone, A.; Olshausen, B.; Pouget, A.; Savin, C.; Sejnowski, T.; Simoncelli, E.; Solla, S.; Sussillo, D.; Tolias, A. S.; and Tsao, D. 2023. Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nature Communications*, 14(1).