Deep Q-Learning and Hierachical Reinforcement Learning in Exploratory Games

Lauryn Brough, Emmett Tomai, Roberto Rivas, Michael Aldaba and Abhinna Adhikari
Department of Computer Science, University of Texas Rio Grande Valley

Previous Research & Background

Machine Learning in Game Environments provides a perfectly controlled simulation of events that could be relatable to real-world environments. In years past, machine learning has been applied to games such as Go (Silver et al., 2016) and classic Atari games (Mnih et al., 2015) and has shown to outperform human opponents. However, these environments are designed to work specifically there and networks are trained to be rigid in their application. With these environments focused on skill, a vast amount of games that focus on creativity and building untouched.

The typically repetitive nature of exploratory games, commonly referred by gamers as 'grinding', should prove to be an easy first task. There are a number of objectives that could be used, from gathering materials to crafting items, exploring maps to battling opponents. A common objective follows the basic 'gather-craft' cycle as it is very repetitive and easy to predict.

Goals

In an exploratory game environment that is harder to gauge success, the aim is to develop an agent that can accomplish a goal aside from purely winning and achieve more human-like behavior. For example, an agent that might decide to explore the map to find the best spot to gather from trees, with plenty of trees nearby, instead of going for the first tree in sight.

The first goal, however, is to ensure basic learning from the machine as applied to the crafting task which implies gather materials and then crafting instead of crafting when no materials are in the inventory. To check basic learning, multiple recipes can be used.

Methods

By employing a options framework (Sutton, Precup, and Singh 1999), we look at the basic gather-craft cycle, gathering materials to craft a more useful item, from exploratory video games. Using trained options, the network might be able to determine the necessary subtasks to accomplish its overall goal as soon in Figure 2. We apply this options model alongside an Double Dueling DQN or A3C network and an experience buffer to train all the network for our goals.

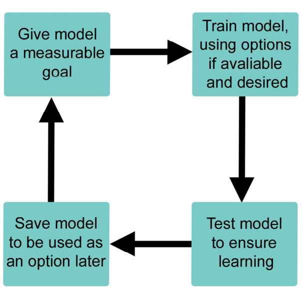


Figure 1. Basic Cycle of working with the Options Model

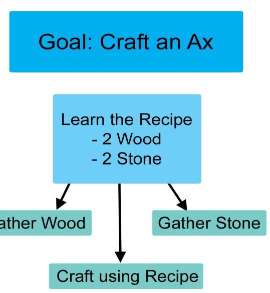


Figure 2. Generation of Subgoals Example

Double Dueling DQN

- Chooses Actions and Generates Q-Values in seperate networks after seperating Advantage and Value
- Advantage: Best Action to take compared to others
- Value: How good of a state we are in

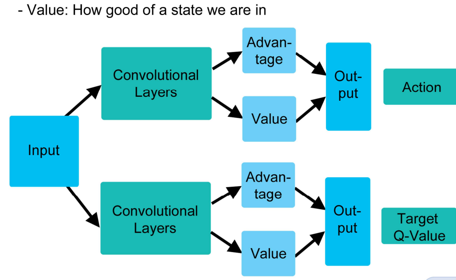


Figure 4. Visual Breakdown of Double Dueling DQN

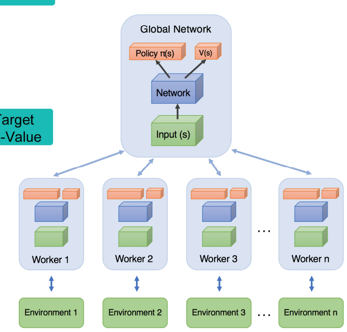


Figure 3. Visual Breakdown of the A3C Network

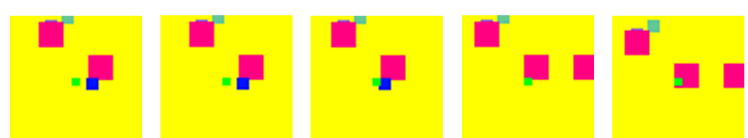
A3C Network

- Runs Asyncronously, multiple instances happening at once that all train from and update the global network
- Policy: chances that an action is a good action to take
- Value: how good of a state we are in

Results & Future Work

Only recently, we were able to get the A3C network working with our environment and as such are unable to present any results with this network. However, as similar tests with the Double Dueling DQN have shown, that the options model is providing a steady improvement in the ability of the network to learn and a clear transference of knowledge. Figure 11 shows the total rewards for the model across multiple episodes using a crafting goal. As shown, the fewer options we give the model, the more steps it takes for the network to properly identify it's goal so it may begin gaining positive rewards.

We hope to expand this project beyond the basic gather-craft cycle to later include map exploration, where the agent is only able to see smaller parts of a big map at a given time, to be applied to searching objectives such as find the treasure.



Figures 5-10. Steps as the network identifies and moves toward a gatherable object, gathers it, and moves onto its next objective

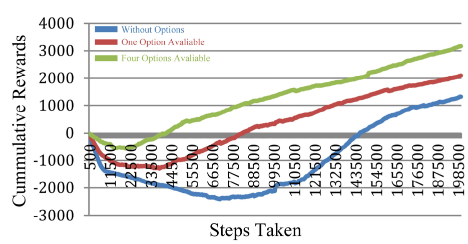


Figure 11. Graph showing reward growth over steps for different numbers of options avaliable

Summer 2018