Journal Report 7
10/14/19-10/28/19
Pawan Jayakumar
Computer Systems Research Lab
Period 5, White

Daily Log

Note

Since I spent most of my time thinking, I'll provide a summary of the different ideas I went though and hope you believe it is a representation of either 45 or 90 minutes of thought. On PSAT day, I spent 90 minutes on senior research as if there was class.

Monday October 14

I started developing my algorithm that finds the structure of the hierarchy automatically without the need for hyper-parameters specifying how many layers and options are in the hierarchy. The two biggest problems, accordingly, are how to determine whether another layer or option should be made at each time step. Each option has a set of initiation states. Every time the agent encounters a new state, the agent has to determine whether an existing option is sufficient or a new option should be made. I can't just use another neural network to learn which option to choose because the number of options changes over time. I had to find some metric to judge which option a new state should go in.

Wednesday October 16

I thought maybe I could create a metric to measure an options effectiveness. I had to define what effectiveness is. Some traits associated with effectiveness is efficiency, accuracy, and relevance to the current state and goal. I thought it could be as simple as how many times was an option was successful in reaching the goal divided by the number of times it was called. It seemed like a good idea until I realized it didn't really help for determining whether a new option should be made. I then thought of similarity. A lot of states are effectively the same because enough conditions are the same that slight alterations in the positioning of the agent or other objects in the environment don't matter.

Friday October 18

Determining similarity between states is much harder than determining similarity between images or audio files because the state space is massive and its not clear what makes two states similar. Sure, anyone could use cosine similarity and say that is how similar they are. The hard part is determining thresholds that allow us to group states together for the purpose of creating options. The set of all states has to be partitioned into subsets that can be used as the initiation set for each option. I thought to solve this problem I needed specifics about the environment I was working on. At first, I thought I would use MuJoCo, but then I found out about Microsoft's Project Malmo. This is a modified version of Minecraft built for AI research. When I discovered this environment, I knew I had to use it. Minecraft is a well known game and would draw many people's attention. It's cooler to see an agent playing Minecraft than watching an ant move around in abstract space.

Monday October 21

I looked at the documentation for Malmo and found that the world state is represented as an object which contains the pixel data of the screen, the position of the agent (x,y,z), and the agents pitch and yaw (for the head position). As I said earlier, I am only interested in significantly different world states. For this reason, the agent's pitch, yaw doesn't really matter for options not in the lowest layer of the hierarchy. The video frame coupled with the agent's position can give us good insight into the key features of the state. These features are something that need to be learned and can be used to decide what option to use or whether a new one should be made. It seems like a CNN might be involved in the answer. I understand more about the problem, but not enough.

Wednesday October 23

I turned my attention back to options. Since options produce sub-goals for lower levels, why not use that as both the initiation state and the goal state? This way, we can just focus on comparing sub-goals rather than the current state the agent is in. This way, the agent doesn't make options tied to regions in the world. The agent might be in a mountain or a river, but if the sub-goal is just finding grass, then the option shouldn't care about where it starts. The number of unique sub-goal states produced is much less than the total number of states and more importantly, they are by nature of being check points, going to be further apart in state space.

Friday October 25

Comparing goal states is an interesting idea, I think for now I am going to try a simpler idea and see if it works. You would expect that the more an option is called, the more it is capable of doing. The more it is capable it is doing, the more support it should get. New options should be made for the most popular option in every level. This will happen after every episode and only if there is an option being called way more than the others (the frequency that other options are being called collectively ¡ the most frequently called option).

Timeline

Date	Goal	Met
Today minus 2 week	Adapt the DDQN to play atari breakout	No, I changed my goal midweek
Today minus 1 week	Iron out the details of my new internal reward function	No, I changed my goal again
Today	Figure out when to make new options	I have an interesting idea,but I'd have to test it, so I guess met for now.
Today plus 1 week	Figure out when to make a new layer in the hierarchy	
Today plus 2 weeks	Run the Malmo environment with a dummy agent	

Reflection

I changed my idea again. I think internal reward is a topic for another day. I don't have enough experience to come up with a new and better one yet. Over the course of first quarter, I reduced the scope and complexity of my research project because by learning more from reading papers, I realized my project idea was way too ambitious. Now my project idea is to modify an existing idea that will hopefully make it better by making it more flexible for learning multiple tasks. I also think I have way too much work to actually test this out by the Regeneron deadline. I think to preserve my mental health, sleep, and grades in other classes, I am not going to enter the competition. It just isn't realistically possible. I sound like I am down in the dumps, but I think I am just being more realistic about what I can and can't do. With that being said, I am still confident that I can at least accomplish this new idea. It takes a while to come up with solutions to problems, but I think I can do it.