

- Things I changed:
 1. Changing 'replay' method's name to 'train': At first, since it was doing the same thing over and over again 'replay' was a good name, but because we want to train the model to keep its balance for longer amount of time, 'train' is a better name.
 2. Changing the 'for' loop inside the 'reply' method: 'for' loops are relatively slow, especially when we are working with large batches. Therefore, I tried doing it by converting the entire minibatch into a NumPy array and slicing columns. Since it is a more efficient way, I think I will see a speed boost. (Remember that the speed of program also depends on CPU/GPU too.)

Honestly, I didn't see a huge speed boost at all! It was only about 10 minutes?? Maybe since CartPole is not an advanced project, and basically is 'hello world!' of RL area I didn't see a significant speed boost. Take note that I did this before applying the 3rd change.

But surprisingly, I think it effected the overall result of agent! My max score was 13000

3. Using truncated in 'run' method, in 'while not terminated' loop: Honestly, I do not know why I haven't used it, I have no clue. I noticed that I am getting scores more than 500 while the max score for CartPole is 500, but I never thought it could be because of not using 'truncated'.

After applying the change: Now the max is score is 500, I have to admit that I was deceived by the appearance of the code! I thought since it is a complicated code, it must have a complicated solution, but I was wrong. However, it was a good experience to know that you can even get scores more than 10000 in CartPole! I think removing the cap is a good way to test the full potential of agent, checking the efficiency of reward function the long run, testing exploitation and exploration, and finally checking the convergence.

I have to mention that results not being consistent was an issue here too.

Other problems that I faced during applying the changes:

- I. When there was no score caps applied, the agent was facing some major issues. The results went up and down sharply, especially under 2000 episodes. Even after 2000th episode the results were not consistent enough. I think I need to change the reward function for better results. And maybe I try different exploitation & exploration numbers.
- II. Since I am doing reverse coding, basically going from top of the triangle to bottom of it, I have to record everything on my screen, every search I do, every change I apply, I also need to record the date of doing them. However, doing it using 2 separate monitors is not an efficient way to do it. Either I have to buy a 49-inch monitor to be able to show my browser, Pycharm, and word at the same time, or I just stick with my current 27-inch(vertical) + 34-inch and record the browser and Pycharm on 34-inch, and taking notes on vertical monitor, all happening while I explain everything, I using a mic. (Since the 49-inch monitor is so expensive, I might not buy it.) I also want to buy a camera to record my body while I am working, something like streamers do.
- III. Blue screen problem of my computer has been finally solved! I thought it was a system file issue or a there was a problem with SSD, but actually my rams got loosen! Probably because of moving from Songdo to here, during the shipping they got loosen. Current specs are: i5 13400F, RTX 4060TI 8GB, 32GB of DDR5 Ram. I was thinking of upgrading the graphic card to RTX 5070TI, I have the money, but I am just not sure whether I need it or not. I was thinking of buying a better system just for training agent. Should I do it?