Southern New Hampshire University

6-2 Assignment: Cartpole Revisited

Jehu A. Domenech Ortiz

CS-370-T3352 Current/Emerging Trends in CS 23EW3

02/10/2023

Humans use experience to come up with an answer to a problem they are facing (Joshi, 2022). When compared to Artificial Intelligence who utilize many iterations of the same task to come up with a solution we can use all of our experiences collectively to come up with an answer without having done the task beforehand. In the pirate game example we humans could use maze puzzles as a reference to complete the pirate game within our first try. An untrained AI wouldn’t be able to successfully complete the puzzle on its first try.

The AI in the pirate game uses episodes to train itself. During each episode it uses the experience that it gained from previous episodes to come up with a better solution for the problem by modifying the Q value. The goal for the AI is to maximize its reward this can be done by making better choices and avoiding any penalties.

Because the first approach does not require one to play the maze to come up with an answer and instead it uses past experience and general logic to come with an answer there is no need for trial and error to solve the possible. The second approach does not use logic or past experience from the beginning which requires the model to train itself from zero and gain enough experience to come up with an answer by trial and error.

Because the AI needs to maximize its rewards it needs to have a set of actions in which it can select to search for a different possible action that could potentially improve the reward that it gets (Bisong, The exploration-exploitation trade-off). When the AI uses a different action to explore whether it’ll gain a bigger reward than its previous actions that’s known as exploration. Exploitation on the other hand is when the action that the AI uses is already known to be the maximized result for the current step. Exploitation is necessary to avoid guessing for an answer in each iteration.

For the pirate game I had to implement a Q value algorithm for coming up with an answer to the pirate game. I first started by implementing the epochs which uses a range of epochs that was set up beforehand by the user. A random cell then is selected and the maze is reset alongside the agent. The environment then gets observed and checked for the current game status. The valid actions that can currently be used are checked and saved in a variable. A while loop then is used to check whether the previously saved variable holding the game status is over or not, it will loop until the variable does not equal ‘not\_over’. In the while loop the previous environment state is going to be the current environment state an if statement is then used to set the action to take. The environment state, reward, and game status are set to the action that was selected randomly the episode variable is updated and then the episode is set with the variables of the previous state, action, reward, current state, and game status. The episode is then remembered to use in the next episode. The history is then set and also the lost after this is done then the game status of the episode is checked to see whether it won or lost and the process repeats.

References

Bisong, E. (n.d.). The exploration-exploitation trade-off. dvdbisong.github.io. Retrieved February 16, 2023, from https://ekababisong.org/the-exploration-exploitation-trade-off/

Joshi, N. (2022, August 16). Human vs ai: Who would win in the iconic Mind combat? Forbes. Retrieved February 16, 2023, from https://www.forbes.com/sites/naveenjoshi/2022/08/12/human-vs-ai-who-would-win-in-the-iconic-mind-combat/?sh=78fc60c45aa2