Aryan Luthra

CS285 FA20

September 23, 2020

Partial Curriculum Based IRL

CS285 Final Project

Often in RL we want to our rational agents to be just as smart or smarter than humans. Hence, for certain tasks, we would like to have a computer simulate the an adult brains decision making process. However, Alan Turing had a nice way to simplify the problem: “Instead of trying to produce a program to simulate the adult brain, why not try to produce one that simulated the child’s? If this were subjected to an appropriate course of education, one would obtain the adult brain.” This quote is in essence the basis for a history of wanting to incorporate demonstration into RL algorithms, where “experts” demonstrate actions and trajectories and then the agent improves its policy based on the demonstration.

Another famous reinforcement learning method/add-on is called “Curriculum Learning” where RL agents, (lets say Q-learning agents, for example) are presented with situations from the same problem/objective, but often are characterized by different initial state distributions *p(S0)*. This, with enough practice acts as “training wheels” to get the RL agent started along the right policy before developing advanced nuances within it. The best example of this would be with sparse reward functions, such as a rewards of 100 for completing the maze and a reward of 0 for not. In this case it make be useful to first start the agent close to the end of the maze, and then slowly move it towards the start of the maze. By adopting this method, the agent will have to explore less to find a trace of the reward function, and we will save many iterations till convergence.

Often when teaching children, teachers adopt a method of demonstration, but in order for an agent, but they also follow a curriculum. Teachers teach to add before they teach to multiply, and teach algebra before calculus. This project aims to do the same with reinforcement learning agents. Not only will agents be shown demonstration data from an “expert”, the demonstrations will be shown in a “curriculum”, that would allow for much quicker convergence. This project will work with clasical RL examples, such as Lunar Lander and OpenGym’s version of Doom (aka VizDoom), and compare the time/iterations to convergence using a model trained in all four following scenarios:

* No curriculum + No demonstration
* No curriculum + Demonstrations
* Curriculum + No demonstration
* Curriculum + Demonstrations

The hypothesis to test would be does an agent that is shown demonstrations at each step of the curriculum perform better or learn quicker than an agent that has received neither. If this hypothesis is shown to be correct, then this opens up new ways for humans to teach robots, and makes learning from demonstration much more feasible and concise. Imagine if you could train a robot arm to to be versatile in its function based on the demonstrations you showed it. If curriculum based learning substantially reduces the number of demonstrations needed, then it may suddenly be feasible for humans to teach agents to, for example, “bring the coffee” or “take the plate to the kitchen”, in only a few demonstrations.