Andrew Heard

Project Two

CS 370-R4883

SNHU

Design Defense

While machine learning may have initially been designed to mimic biological thinking and problem solving, it has evolved into something much different. A human could approach the problem of navigating a maze a few different ways. One approach would be to simply pick a side, or wall, and follow that path until the end is reached. This method has a 100% certainty of success, but they might possibly span the entire maze, making it inefficient. The most likely way a human would go through is either marking or memorizing paths already explored to avoid repeating steps and simply exploring the maze randomly. Repeated exposures to the same maze would mean the human could remember the quickest path leading to the most efficient result. The machine would take a different form of decision making. The machine would evaluate each step and each direction heavily against its past experiences. Each action would be assigned a value based on the likely hood of getting closer to receiving the reward at the end of the maze. While both players are focused on reaching the end as quickly and efficiently as possible, the human will be leaning more towards an exploratory method and leaning more on luck while the machine is calculating each action to use exploitation to achieve the most mathematically likely correct path.

This is not to say that the machine will not use exploration decision making at all. In fact, when exploitation is failing or no possible actions net a positive enough chance of reward, the machine would default back to exploratory actions to try to find new, unexperienced solutions. Exploitation is simply plan A in this design. With the implementation of reinforcement learning, exploitation can lead to the most mathematically sound and safest plan of action. Evaluating past actions against their past results and assigning those experiences positive or negative values is textbook exploitation. When the exploitation fails, or falls into a loop, exploration can be engaged. Exploration is riskier behavior within unknown outcomes but can randomly lead to net positive experiences. The machine can assign values to these new experiences, compare it to other past experiences, and change its strategy. Once these new ideas are implemented, exploitation can again be enabled, and the cycle continues like so until the best path is calculated. The portion of exploitation versus exploration that this agent used is to heavily rely on exploitation first while seldom falling back to exploration for needed breakthroughs, which for problems such as mazes has been shown to be very efficient.

Another way to describe the way this agent handles decision making is the use of Deep Q-Learning with neural networks. The machine saves each instance of the board they are currently experiencing, called states, and each action they took at each state in a table called a Q-value table. The machine also remembers the outcome of the game for each of those actions for each state. Using these three variables, a neural network is implemented to efficiently calculate the values of each action in the table and the likely hood that the action will move the agent closer to the reward. All the agent’s past experiences start to pile up and make for a mountain of data that the network can parse and give helpful predictions.