Liu, Kevin

16 April 2025

Professor Yaroslav Litvak

CS-370 Current/Emerging Trends

**Design Defense**

There are significant differences between a human and an artificial intelligent agent navigating a treasure hunt. A human navigates a treasure hunt based from prior experiences and would be intrigued by what the possible treasure is. A human would also be able to use clues to better navigate the terrain and find the treasure through inference. On the other hand, artificial intelligence would not have the curiosity of a human being and would tackle a treasure hunt by brute force. It would explore all possibilities of a given location and ignore clues before moving on. An agent is also data-driven and uses reinforcement learning strategy to optimize finding the treasure before the human player. Although there are differences, they also share similarities such as, a goal of finding the treasure and navigating the surroundings through trial and error. Exploring paths, routes, and clues that would assist them in finding the treasure before the other player.

The difference between exploitation and exploration is that exploration is the process of trying different actions even if they are not the best decisions. An example of this would be the agent walking into a trap to learn that traps are bad or walking into a wall and noticing that it stops movement forward. Exploration is a trial-and-error phase for the agent to understand its surroundings. On the other hand, exploitation is the agent using what is already known to find the treasure without trial and error. The agent would exploit the paths it has learned to find the treasure efficiently. The ideal proportion of exploitation and exploration would be the initial phase be mainly exploration for the agent to learn its environment and uses reinforcement learning to understand its environment through rewards and optimization. Then use exploitation to not trigger traps or walk into walls in order to find the most optimal path to treasure.

Reinforcement learning is used in the exploration phase where the agent is learning through trial and error. The agent is rewarded for good behavior and penalized for incorrect behavior like triggering a trap or walking into the wall. The agent then uses this reward and penalization to seek as much rewards as possible while limiting its bad behavior. Over time, this allows the agent to make better decisions, leading to its success in the long run.

This environment is complex and would require Q-learning. In traditional Q-learning, a table of Q-values is maintained, but this becomes impractical as the environment grows larger. Instead, Deep Q-Learning uses a neural network to approximate the Q-function. This network allows the stable learning and prevents the network from overfitting.

References

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. <https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf>