Final Project Defense Paper

Ryan Rieth

Prof. Alexander — CS 370 — SNHU

The pirate agent project is a prime example of analysis on the differences and similarities between human approaches and machine approaches to solve problems, riddles, puzzles, or scenarios. In this case, the pirate agent’s goal is outright and simple: solve the maze and obtain the treasure. The pirate can win by finding the treasure in the maze or lose by solving the maze without the treasure or getting lost in the maze. Arguably, the same realistic outcomes would be true if the pirate were a human thinking for his self rather than a deep-learning machine utilizing a Deep-Q algorithm. However, the scripting and results of this pirate agent neural network showcase wonderful comparisons and contrasts between how humans can approach the maze versus how the machine approaches it.

Regardless of similarities, humans and machines differ widely from one another. A machine has the ability and environmental allowance to consistently repeat the same experiment with fed data and learned experience while a human does not have quite the same level of luxury, if at all. In the case of a maze with treasure and a life-or-death outcome, a human would not be able to learn from their mistakes anywhere near a machine that can simply “restart” with their prior experience saved, recorded, and processed. While this is a substantial and foundational difference, humans — when granted a chance to learn and a shot at a do-over — learn much more rapidly and significantly than even the best deep-learning AI machines (Pedamkar, P. 2023). Humans garner experience more quickly than machines, even if machines can ultimately do more with this experience when it is employed with their deep-learning algorithms.

With respect to the treasure hunt scenario, the pirate agent employs a surprisingly similar mindset to approaching the situation like a human would, albeit with some differences. In the maze, a human would begin more than likely by choosing a random direction to travel in: forward, backward, right, or left as applicable. There is an incredibly little to no chance the human has a clue which direction is more correct, and so randomness leads. As the human comes to dead ends, they backtrack and make another random decision only less random in that the direction they just traveled is understood to be incorrect. The pirate agent in the Deep-Q network begins in a similar fashion, determining its valid actions to take (at most, four directions at a time) and deciding a random direction from there and continuing in very humanlike fashion. However, if a particular choice reveals itself to be greater than an allowed exploration factor and can be combined with existing successful and unsuccessful replay experience already learned, then the machine can more accurately and more strategically predict a correct route to the treasure.

The human and machine differences in approach come to fruition at this junction. Humans are as strategic as they can be, are flexible, can learn deeply very quickly, and are adaptable in many vital ways to changing scenarios or unexpected results. As Volodymyr Minh et al (2015) describe in their deep-learning reinforcement breakdown paper, deep-learning machines — such as the Deep-Q neural network at play with the pirate agent — are complex; they rely much more on learning and experience than humans, resulting in immense rigidity and subsequent lack of adaptability and, arguably, learnability. Portraying this occlusion of shared approach is how the pirate agent handles exploration and exploitation. Exploration can be understood as a strategy-less strategy; the agent explores its options and learns these potential routes with less regard to how significant these routes are to obtaining rewards. Exploitation, on the other hand, is consistently utilizing successful actions, routes, or choices to maximize reward gain. These are two strategies humans can and do employ, true, but deep-learning machines tend to be a bit more binary when it comes to deciding to explore or to exploit while a human can smoothly transfer between the two approaches or even find a “human way” to mesh them together. An ideal proportion of the two is Boltzmann Exploration, a delicate balance of exploration and exploitation (van Heeswijk, W. 2022). In this mindset, the pirate agent begins the maze with an exploration-heavy mindset and finishes the maze with an exploitation-heavy mindset. The agent can learn its surroundings, process potential patterns in the maze’s design, remember successful and unsuccessful routes, and bring all this data together to more accurately and efficiently predict a path that not only leads to the treasure but also to solving the maze.

By utilizing such a mindset as the Boltzmann Exploration method, the agent can strategically and successfully accomplish tasks and problems cemented in reinforcement learning. This kind of deep-, neural network-learning slowly but surely reveals a path to success for, in this case, the pirate agent. Accomplishment is attained so via a classic, humanly psychological system of rewards (understood as positive rewards) and punishment (understood as negative rewards). The neural agent learns an action and whether to continue to implement this action to increase its chances of overall success and its highest level of possible rewards earned should the action render a positive reward gain (Minh, V. et al 2015).

The reinforcement learning in the case of the pirate agent is buttressed particularly by Deep-Q algorithms. The agent analyzes Q-values of every action relevant to the current environment and makes educated predictions of which following action would render a higher success rate or reward gain. In the beginning of the scenario, the agent initializes values that will determine its “mindset” as well as potentially change throughout the scenario. As the agent progresses, it will predict future occurrences based on learned experiences stored as experience replay episodes (learned successes and failures). The agent will then act, whether in explorative or exploitative fashion, and update its Q-values to train and ultimately evolve the neural network as it progresses.

Coincidentally — perhaps, on purpose — Deep-Q learning mirrors in many ways learning accomplished by the human brain. A child repeatedly falls and gets hurt and eventually learns better ways to balance his or herself through growth. Neural networks utilize reinforcement learning in quite similar ways. They perform and learn whether their performance was successful or unsuccessful. From this experience, the deep-learning machine can better its own algorithm and learning procedure to exponentially improve its performance, its efficiency, and its overall success and reliability. Nuances in human and machine approaches widely vary and even be entirely different. However, the core of each approach is simple and shared: learn by doing.

## References

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M.A., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*, 529-533. <https://courses.engr.illinois.edu/cs546/sp2018/Slides/Apr05_Minh.pdf>

Pedamkar, P. (2023, June 5). *Artificial Intelligence vs Human Intelligence: Top 5 Useful Comparisons*. EDUCBA. <https://www.educba.com/artificial-intelligence-vs-human-intelligence/>

van Heeswijk, W. (2022, June 24). *Seven Exploration Strategies in Reinforcement Learning You Should Know*. Medium. <https://towardsdatascience.com/seven-exploration-strategies-in-reinforcement-learning-you-should-know-8eca7dec503b>