**Design Defense**

Humans and machines solve problems differently due to inherent differences in their decision-making processes.

**Human Approach**

A human with the same maze will estimate the possible ways through it, possible barriers, and the place where the treasure is hidden.

Steps might include,

Exploration - Exploring from the beginning and doing it by experimentation.

Memory Usage - Drawing on prior attempts to prevent choosing approaches that have been tried with no success before.

Goal-Oriented Decisions - The first option of deciding to take a path that seems to bring a novice closer to the treasure according to what is seen.

**Machine Approach**

The identified Q-learning agent operates based on reinforcement learning to wander through the given maze, evaluate states and actions, and accumulate suboptimal knowledge about the environments through the rewards.

Steps include,

State Recognition - Describing the environment and states discretely as a grid where states can be represented by a number.

Exploration vs. Exploitation - Juggling between exploratory search, as looking for new avenues or ideas to explore and exploit found knowledge, in an attempt to get the most out of the acquired knowledge.

Reward Optimization - Upsolving resulting in the determination of the best actions for each state under the context of maximum cumulative reward.

**Similarities**

Both strategies involve actions that feedback from the environment updates to make the right decision.

Both entail the acquisition of resources for knowledge and the application of those resources for the right outcome.

**Differences**

Humans on the other hand use emotions and sight while the agent on the other hand uses numbers and computations.

Computers may carry out an over-enthusiastic search of all possibilities while aircrews may select a course of action based on conjecture or intuition.

**Measures Employed by the Intelligent Agent**

The intelligent agent solves the pathfinding problem using the following steps:

Initialize - Describe the kind of environment to be used in the maze, set up the Q-network, and input other parameters such as learning rate, discount factor and exploration rate.

Observe State- The agent starts at start state and they see where they are in the maze.

Select Action,

First, we see the agent wandering randomly around the environment.

In training, the agent utilizes the Q-network to discover which action will provide the most accurate Q-values after training.

Perform Action - The agent also acts in the following way depending on the action that has been selected in the environment formed by the maze.

Receive Feedback - The agent gets a point from the action (e.g. +10 for getting to the treasure, -1 for an obstacle is hit).

Update Q-Values

Q(s,a)←Q(s,a)+α[r+γa′max​Q(s′,a′)−Q(s,a)]

Repeat - The agent repeats the process for multiple episodes until it converges on an optimal policy.

**Purpose of the Intelligent Agent in Pathfinding**

So, the main target of the intelligent agent is to find the way which will take it to the treasure with minimal number of steps. Agents such as this are needed for various applications including robotics, for instance, for the self-navigation of robots and logistics, for instance, for organizing supply chain routes.

**Exploitation vs. Exploration**

Exploration - Exploring unknown territories to find out novel state and action spaces.

Exploitation - The second type utilizes the information we have to select the best-known actions.

Ideal Proportion - If exploring at the beginning using a larger epsilon (e.g., epsilon=1.0) to gain knowledge, then reducing epsilon over time (e.g., epsilon=0.01), after sufficient learning, will allow the agent to act more on the knowledge learnt. This balance enables the working of the agent in such a way that it is free from local optimums and find the global optimum.

**Role of Reinforcement Learning**

Reinforcement learning helps the agent learn an optimal policy by,

Providing penalties for those activities that would take the agent farther from the goal and providing a sense of accomplishment to those that would narrow the gap.

Punishing activities that are all the wrong (for example, hitting the object the child is not supposed to touch).

Updating Q-values in a cyclic manner in order to have the agent grasp the best rewarding way through a sequence of actions.

**Implementation of Deep Q-Learning**

State Representation - Store each current position of the maze as a one-dimensional input.

Action Space - Outline four conceivable operations as: upwards, downwards, leftwards or rightwards.

Neural Network - Employ a feedforward neural network in order to estimate the Q-values for all actions at a given state. The network includes:

Another layer called input layer which should have the same size as the state.

Hidden layers - Two – activation functions: ReLU.

An output layer consisted of DQ for each action.

Training Loop

* Use the agent's experiences (state, action, reward, next state) to train the Q-network.
* Minimize the mean squared error between predicted Q-values and target Q-values: Loss=1n∑i=1n[Qtarget−Qpredicted]2\text{Loss} = \frac{1}{n} \sum\_{i=1}^{n} \left[ Q\_{\text{target}} - Q\_{\text{predicted}} \right]^2Loss=n1​i=1∑n​[Qtarget​−Qpredicted​]2
* Update the network weights using backpropagation and the Adam optimizer.

References

Khurana, R. (2023, June 1). *Differences between artificial intelligence and human intelligence.* Scaler Topics. Retrieved from <https://www.scaler.com/topics/differences-between-artificial-intelligence-and-human-intelligence/>

Nuer, J. (2022, May 20). *A comprehensive guide to deep Q-learning.* Medium. Retrieved from <https://medium.com/@jereminuerofficial/a-comprehensive-guide-to-deep-q-learning-8aeed632f52f>

Singhal, R. (2018, December 13). *Reinforcement learning: A beginner’s guide.* Medium. Retrieved from <https://medium.com/@rsinghal757/reinforcement-learning-a-beginners-guide-56de9c7fe1d9>