Question- You must have heard of or used ChatGPT at some point in the last year, maybe even before. It's like Janet from the Good Place, it always has the answers to your question. Unlike Janet, this may not be 100% correct but does get the job done. But it wasn't always like this. Earlier stages of ChatGPT used to give bogus answers all the time, but over time it learnt how to give appropriate answers, relevant to the context, being as accurate as it possibly can. This is done through a method called Reinforcement Learning. For this task, you must understand and explain the working of reinforcement learning. Additionally, list some other examples and explain how they work.

Introduction

Let's start the introduction with an analogy. Suppose you are child playing a game where you have to find a hidden treasure. However, there are several obstacles on your way which will reduce your productivity. In the first try you face an obstacle right at the start of your quest, but you try again and again, and slowly your efficiency in the game is maximized. After trying tons of different ways all the while learning the outcomes, you finally reach the treasure.

Reinforcement Learning is also the same, here the robot or machine is you, the treasure is the reward, the obstacles are punishments.
The robot keeps on trying different ways to reach the reward, remembering every path in it's memory so as to prevent previous mistakes.

Now for the Actual Definition –

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

Reinforcement vs Supervised vs Unsupervised

Along with Reinforcement Learning there is also Supervised and Unsupervised Learning and there are certain differences in each of them –

Unsupervised Learning –
This is when the model is given a dataset without any labels or output. The model itself finds patterns and trends in the dataset usually using Clustering (grouping similar contents together)

Eg. – generating a set of images based on existing ones (generating output on the basis of input)

Supervised Learning-
This is when a model is provided with input as well as correct output or labels corresponding to the input. The model then reads the input provided by the user, matches it with whatever output it was provided with for that input and gives the result.

Eg. Image Recognition or Object Detection is one of its use case

Reinforcement Learning-
In this, it is the model that itself figures out what kind of output it has to give, by learning from previous actions.

Eg. ChatGPT is one of its use cases
(Quote Unquote) In 2016, researchers from Stanford University, Ohio State University, and Microsoft Research used this learning to generate dialogue, like what's used for chatbots. Using two virtual agents, they simulated conversations and used policy gradient methods to reward important attributes such as coherence, informativity, and ease of answering.5 This research was unique in that it didn't only focus on the question at hand, but also on how an answer could influence future outcomes. This approach to reinforcement learning in NLP is now widely adopted and used by customer service departments in many major organizations.

Elements of RL

Reinforcement learning elements are as follows:
Policy
Reward function
Value function
Model of the environment

Policy – Policy defines the learning agent behavior for given time period. Through this the agents decides which action to take the first the model is executed or is met with an unique query.

Reward Function – This is like a feedback system that tells the bot or model how well it has performed. This is for the agent to know whether the output given by it is good or bad.

Value Function - The value of a state is the total amount of reward an agent can expect to accumulate over the future, starting from that state.

Model of the environment – This is the place or the structure or category in which the model has to perform.

Advantages –

1. Can be used to solve complex problems.
2. Can handle environments that are non-deterministic(realistic), where the outcomes of actions are not always predictable. This is useful in real-world applications where the environment condition may change over time.

Disadvantages –

1. Needs a lot of data and computation.
2. Can be difficult to debug and interpret.