Content Checked For Plagiarism

Working of Reinforcement Learning:
Consider the scenario of teaching new tricks to a dog.
As a dog doesn't understand English or any other human language, we can't tell him directly what to do. Instead, we follow a different strategy.
We emulate a situation, and the dog tries to respond in many different ways. If the dog's response is the desired way, we will give him food.
Now whenever the dog is exposed to the same situation, the dog executes a similar action with even more enthusiasm in expectation of getting more reward(food).
That's like learning that a dog gets "what to do" from positive experiences.
At the same time, the dog also learns what not to do when faced with negative experiences.

RL dog example

Reinforcement Learning Algorithms:
There are three approaches to implement a Reinforcement Learning algorithm.

Value-Based:
In a value-based Reinforcement Learning method, one should try to maximize a value function V(s).
In this method, the agent is expecting a long-term return of the current states under policy π.

Policy-based:
In a policy-based RL method, one should try to come up with such a policy that the action performed in every state helps you to gain maximum reward in the future.
Two types of policy-based methods are:
Deterministic: For any state, the same action is produced by the policy π.
Stochastic: Every action has a certain probability, which is determined by the following equation.Stochastic Policy :
n{a\s) = P\A, = a\S, =S]

Model-Based:
In this Reinforcement Learning method, you need to create a virtual model for each environment.
The agent learns to perform in that specific environment.

Important traits of Reinforcement Learning:

There is no supervisor, only a real number or reward signal
Sequential decision making
Time plays a crucial role in Reinforcement problems
Feedback is always delayed, not instantaneous
Agent's actions determine the subsequent data it receives

Types of Reinforcement Learning
Two kinds of reinforcement learning methods are:
Positive:
It is defined as an event that occurs because of specific behavior.
It increases the strength and the frequency of the behavior and impacts positively on the action taken by the agent. This type of Reinforcement helps you to maximize performance and sustain change for a more extended period. However, too much Reinforcement may lead to over-optimization of state, which can affect the results.
Negative:
Negative Reinforcement is defined as strengthening of behavior that occurs because of a negative condition which should have stopped or avoided.
It helps you to define the minimum standard of performance.
However, the drawback of this method is that it provides enough to meet up the minimum behavior.
Learning Models of Reinforcement
There are two important learning models in reinforcement learning:
Markov Decision Process
Q learning

Markov decision processes

In MDP, there is an agent that interacts with the environment that it is placed in. The agent gets a representation of the current state of the environment sequentially over time. Based on the state, the agent chooses a particular action which leads to a change in the environmental state, and the agent is rewarded as a cause of the action that it took.

There are 5 major components involved in an MaDP:
Agent
Environment
State Space (S)
Action Space (A)
Reward Space (R)

The process of transitioning from one state to another happens sequentially over and over again. Throughout this process, it is the agent's goal to maximize not just the immediate reward, but the cumulative rewards it receives over time

At each time step t, the agent receives a representation of the environment's current state $S_t \in S$. Based on this state, the

agent selects an action At ∈ A. This gives a state-action pair for time t as (St, At).

Time is incremented to the next time step t+1, and the environment is transitioned to a new state St+1 ∈ S. At this time, the agent receives a numerical reward Rt+1 ∈ R for the state-action pair (St, At).
The following diagram (fig 1) illustrates the above-mentioned algorithm.

Markov decision processes

Expected Return

The goal of the agent is to maximize the total rewards which can be represented as the sum of future rewards till the last time step T. This is called the expected return denoted by G. G is given by

G = Rt+1 + aRt+1 + a2 Rt+1 + ...
G = k = 0∞akRt+k+1

Where Rt is the return at time t and a is a number ranging from 0 to 1.

Policies and Value Functions

A policy is a function that maps a given state to probabilities of selecting each possible action from that state. It is denoted by $\boldsymbol{\pi}$. If an agent follows policy $\boldsymbol{\pi}$
at time t, then $\boldsymbol{\pi}$(a|s) is the probability that At= a if St, = s. An optimal policy is a policy that gives the highest expected reward among all the policies.

Value functions give a measure of how good a particular action is. This is achieved by using an action-value function which is denoted by qπ. The value of action A in state S under policy $\boldsymbol{\pi}$ is the expected return from starting from state S at time t, taking action A, and following policy $\boldsymbol{\pi}$ thereafter is given by

qπ(S, A) = Eπ[k = 0∞akRt+k+1 | St, = S, At= A]

An optimal state value function is denoted by q* and is defined as

q* = max qπ(S, A)

q*must always satisfy an equation called as Bellman Optimality Equation which is given as

q*(S, A) = E[Rt+1 + a*max q*(S', A')]

S' and A' denote the next state and action respectively. This equation ensures that the agent is following an optimal policy and the next state S' will be the state from which best possible action A' can be taken at time t + 1.

Q-Learning

Q Learning keeps a lookup table of values Q(s, a) with one entry for every state-action pair [2].
The Q-learning algorithm makes use of the Bellman equation for the Q-value function.

Matched Source

Similarity 25%

Title:Introduction to Machine Learning | Types of Machine ...

· We emulate a situation, and the dog tries to respond in many different ways. If the dog's response is in the desired way, we will give it a treat. If the dog's response is in ...

https://harshal-pawar.medium.com/introduction-to-machine-learning-516322d64628

Similarity 34%

Title:Reinforcement Learning: What is, Algorithms, Applications ...

· In this method, the agent is expecting a long-term return of the current states under policy π. Policy-based: In a policy-based RL method, you try to come up with such a policy that the action performed in every state helps you to gain maximum reward in ...

https://www.guru99.com/reinforcement-learning-tutorial.html

Similarity 7%

Title:www.mindmeister.com › 2004807996 › reinforcementReinforcement Learning problem | MindMeister Mind Map

14.2. There is no supervisor, only a real number or reward signal 14.3. Feedback is always delayed, not instantaneous 14.4. Time plays a crucial role in RL problems 14.5. Agent's action determine the subsequent data it receives

https%3a%2f%2fwww.mindmeister.com%2f2004807996%2freinforcement-learning-problem/

Similarity 5%

Title:Best and No.1 Introduction to Reinforcement Learning!

Time plays a crucial role in Reinforcement problems ; Feedback is always delayed, not instantaneous ; Agent's actions determine the subsequent data it receives; Types of Reinforcement Learning. Two kinds of reinforcement learning methods are: Positive: It is defined as an event, that occurs because of ...

https://writex.today/artificial-intelligence/reinforcement-learning-2/

Similarity 5%

Title:How can machines think ? Machine learning from scratch ...

· It increases the strength and the frequency of the behavior and impacts positively on the action taken by the agent. basically, it helps maximizing performance and sustaining change for a more extended period. Negative: It is defined as strengthening of behavior that occurs because of a negative condition which should have been stopped or ...

https://laabidigh.medium.com/how-can-machines-think-machine-learning-from-scratch-c554e041e88c

---

Similarity 5%

Title:writex.today › artificial-intelligenceBest and No.1 Introduction to Reinforcement Learning!

However, too much Reinforcement may lead to over-optimization of state, which can affect the results. Negative: Negative Reinforcement is defined as strengthening of behavior that occurs because of a negative condition which should have stopped or avoided.

https%3a%2f%2fwritex.today%2fartificial-intelligence%2freinforcement-learning-2%2f/

Similarity 4%

Title:Types of Machine Learning and Associated Algorithms | by ...

· However, the drawback of this method is that it provides enough to meet up the minimum behavior . For example , after being failed , too ...

https://medium.com/ml-with-arpit-pathak/types-of-machine-learning-and-associated-algorithms-cefd6e88fcca

---

Similarity 4%

Title:Reinforcement Learning: What is, Algorithms, Applications ...

https://www.guru99.com/reinforcement-learning-tutorial.html#:~:text=Two%20widely%20used%20learning%20model,given%20sample%20data%20or%20example.

Similarity 3%

Title:Reinforcement Learning - Policies and Value Functions

Sep 27, 2018 — A policy is a function that maps a given state to probabilities of selecting each possible action from that state. We will use the symbol π ...

https://deeplizard.com/learn/video/eMxOGwbdqKY

Similarity 2%

Title:L22_23-DeepRL - Brief introduction to reinforcement learning and ...

View Notes - L22_23-DeepRL.pdf from COMP 424 at McGill University. Brief introduction to reinforcement learning and deep reinforcement learning Vincent ...

https://www.coursehero.com/file/40059995/L22-23-DeepRLpdf/
