Analytics Vidhya

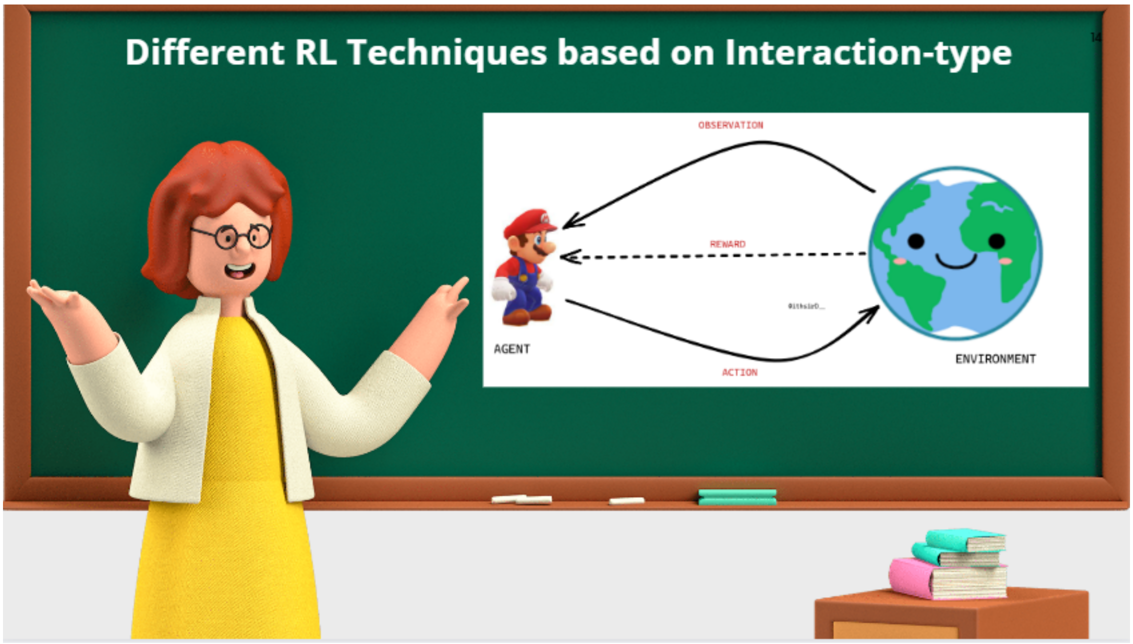
Reinforcement Learning Techniques Based on Types of Interaction

BEGINNER MACHINE LEARNING REINFORCEMENT LEARNING

This article was published as a part of the Data Science Blogathon.

Introduction

With the ubiquitous adoption of deep learning, reinforcement learning (RL) has seen a sharp rise in popularity, scaling to problems that were intractable in the past, such as controlling robotic agents and autonomous vehicles, playing complex games from pixel observations, etc.



Source: Canva

This article will cover what reinforcement learning is and different types of reinforcement learning paradigms based on the types of interaction.

Now, let's begin…

Highlights

- Reinforcement Learning (RL) is a general framework that enables an agent to discover the best way to maximize a given reward signal through trial and error using feedback from its actions and experiences, i.e., actively interacting with the environment by taking actions and observing the reward.

- In online RL, the agent is free to interact with the environment and must gather new experiences with the latest policy before updating.

- In off-policy RL, an agent interacts with the environment and appends its new experiences to a replay buffer, which can then be sampled to update the policy. This paradigm allows for the reuse of prior experiences while relying on a steady stream of fresh ones.

- In offline RL, a behavior policy is used to gather experiences that are used to collect experiences that are stored in a static dataset. Then a new policy is learned without any further interactions with the environment.

What is Reinforcement Learning?

Reinforcement Learning (RL) is a general framework for adaptive control that enables an agent to learn to **maximize a specified reward** signal through trial and error using feedback from its actions and experiences, i.e., **actively interacting with the environment** by taking actions and observing the reward.

Essentially, the agent encounters a sequential decision-making problem where, at each time step, it observes its state, takes action, receives a reward, and then moves to a new state (See Figure1). The RL agent learns a good policy by trial and error based on observations and numeric reward feedback on the previously performed action, where the goal is to maximize the sum of the rewards by minimizing the penalties, which leads to the behavior of achieving a given task successfully.

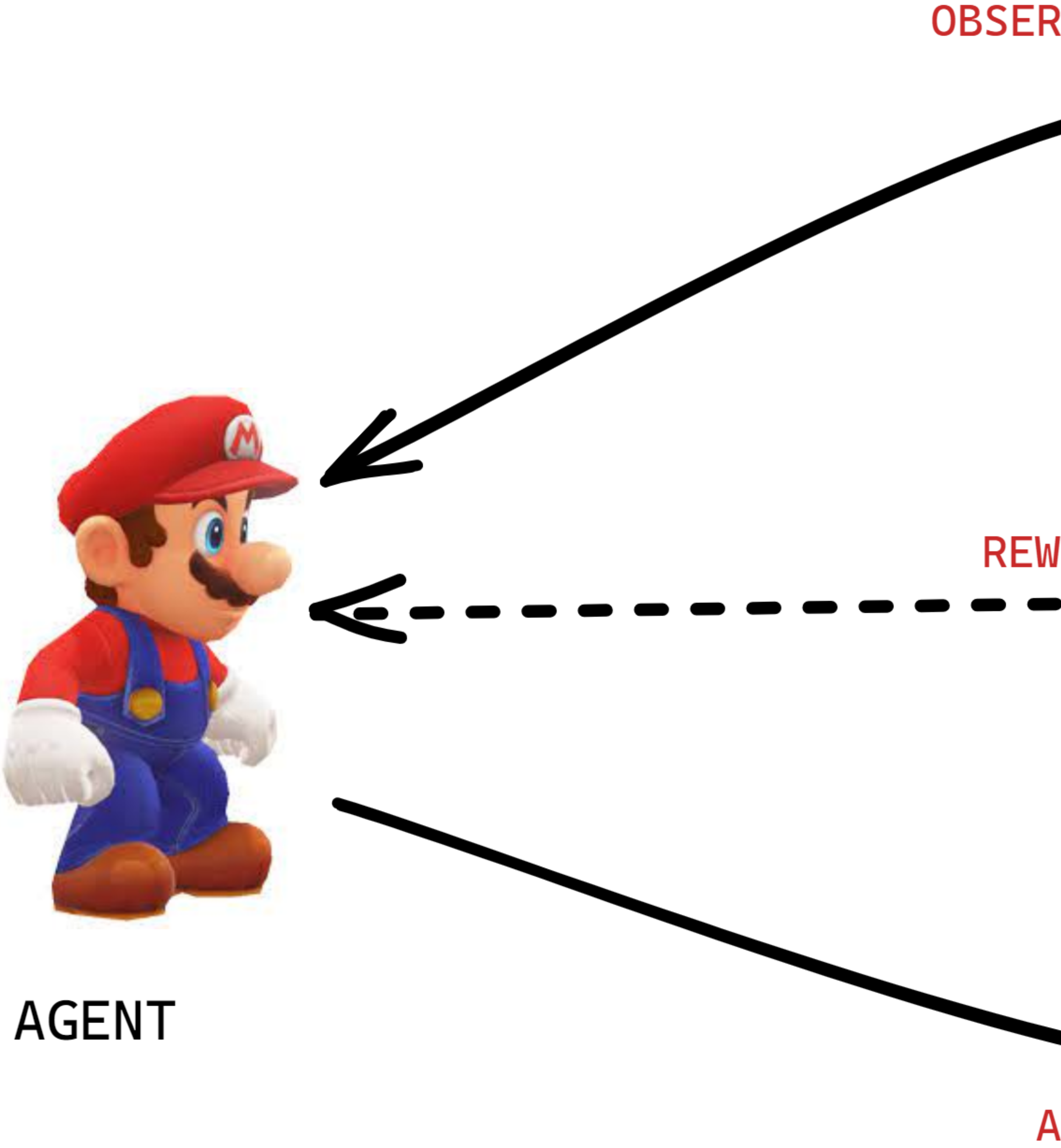


Figure 1: Diagram illustrating Reinforcement Learning

In figure 1, we consider that the agent interacts with the environment. Even though the agent and the environment are separately drawn, we can also picture the agent to be somewhere existing inside the environment. Imagine a huge world in which the agent exists somewhere inside that world and interacts with it, e.g., the Super Mario game.

For instance, in the animation shown below, Mario exists inside the game environment and can move and jump in both left and right directions. When Mario interacts with the flower, he gets a positive reward; however, if Mario comes in contact with the monster, he gets penalized (negative reward). So Mario learns by actively interacting with the environment by taking actions and observing the rewards. Furthermore, it learns to maximize positive rewards through trial and error.

Animation 1: Example of Reinforcement Learning

In essence, reinforcement learning essentially provides a mathematical formalism for learning-based control. By using reinforcement learning, we can automatically develop near-optimal behavioral skills, represented by policies, to optimize user-specified reward functions. The reward function determines what an agent should do, and a reinforcement learning algorithm specifies how to do it.

Now that we are familiar with Reinforcement Learning, let’s explore different RL paradigms based on the type of interaction.

Different RL Techniques Based on the Type of Interaction

In this section, we will focus on the following types of Reinforcement Learning techniques based on interaction-type:

i) Online/On-policy Reinforcement Learning

ii) Off-policy Reinforcement Learning

iii) Offline Reinforcement Learning

i) Online/On-policy RL: The reinforcement learning process involves gathering experience by interacting with the environment, generally with the latest learned policy, and then using that experience to improve the policy. In online RL, the agent is free to interact with the environment and must gather new experiences with the latest policy before updating.

Figure 2, shown below, illustrates online reinforcement learning, where the policy π_k is updated with streaming data collected by π_k itself.

Figure 2: Diagram illustrating Online Reinforcement Learning (Source: Arxiv)

ii) Off-policy RL: In off-policy, the agent is still free to interact with the environment. However, it can update its current policy by leveraging experiences gathered from any previous policies. As a result, the sample efficiency of training increases because the agent doesn't have to discard all of its prior interactions and can rather maintain a buffer where old interactions can be sampled multiple times.

In Figure 3, shown below, the agent interacts with the environment and appends its new experiences to a data buffer (also called a replay buffer) D. Each new policy πk collects additional data, such that D comprises samples from π_0, π_1, . . . , π_k, and all of this data is used to train an updated new policy π_{k+1}.

Figure 3: Diagram illustrating Off-policy Reinforcement Learning (Source: Arxiv)

iii) Offline RL/ Batch RL: In offline RL, a behavior policy is used to gather experiences that are used to collect experiences that are stored in a static dataset. Then a new policy is learned without any further interactions with the environment. After learning an offline policy, one can opt to fine-tune their policy either via online or off-policy RL methods, with the added benefit that their initial policy is likely safer and cheaper to interact with the environment than an initial random policy.

In figure 4, the offline reinforcement learning employs a dataset D gathered by some (potentially unknown) behavior policy π_β. The dataset is collected once and is not altered during training, which makes it feasible to use large previously collected datasets. The training process doesn't interact with the **MDP**, and the policy is only deployed after being fully trained.

Figure 4: Diagram illustrating Offline Reinforcement Learning (Source: Arxiv)

Caveats in Reinforcement Learning

1. The offline RL paradigm can be incredibly beneficial in settings where online interaction is impractical, either due to data collection being expensive (e.g., in healthcare, educational agents, or robotics) or dangerous (e.g., in autonomous driving, etc). Furthermore, **even in domains where online interaction is viable, one might still prefer to utilize previously collected data for improved generalization in complex domains.**

2. While one of the main advantages of **offline RL** is learning from a static dataset, it is also what makes it very challenging for existing online RL algorithms. Theoretically, any **off-policy** method could be leveraged to learn a policy from a dataset of previously gathered experiences. However, these methods often fail when exclusively working with offline data since they were devised under the presumption that further online interactions are possible, and algorithms can typically depend on these interactions to rectify erroneous behavior.

Striking a balance between increased generalization and avoiding unwanted behaviors outside of distribution is one of the core issues of
offline RL. Moreover, this issue is further exacerbated by the ubiquitous use of high-capacity function approximators. To navigate this, most offline RL algorithms address this issue by proposing different losses or training methods that can reduce distributional shifts.

3. **After learning an offline policy, one can still opt to tune the policy online,** with the added benefit that their initial policy is more likely safer and cheaper to interact with the environment than an initial random policy.

Conclusion

To sum up, in this article, we learned the following:

1. Reinforcement Learning (RL) is a general framework for adaptive control that enables an agent to learn to maximize a specified reward signal through trial and error using feedback from its actions and experiences, i.e., actively interacting with the environment by taking actions and observing the reward.

2. In online RL, the agent is free to interact with the environment and must gather new experiences with the latest policy before updating.

3. In off-policy RL, an agent interacts with the environment and appends its new experiences to a replay buffer, which can then be sampled to update the policy. This paradigm allows for the reuse of prior experiences while relying on a steady stream of fresh ones.

4. In offline RL, a behavior policy is used to gather experiences that are used to collect experiences that are stored in a static dataset. Then a new policy is learned without any further interactions with the environment.

5. The offline RL paradigm can be incredibly beneficial in settings where online interaction is impractical due to expensive or dangerous data collection. Also, even in domains where online interaction is viable, one might still prefer to utilize previously collected data for improved generalization in complex domains.

6. After learning an offline policy, one can still opt to tune the policy online, with the added benefit that their initial policy is more likely safer and cheaper to interact with the environment than an initial random policy.

That concludes this article. Thanks for reading. If you have any questions or concerns, please post them in the comments section below. Happy learning!

The media shown in this article is not owned by Analytics Vidhya and is used at the Author’s discretion.

Article Url - https://www.analyticsvidhya.com/blog/2022/09/reinforcement-learning-techniques-based-on-types-of-interaction/

Drishti Sharma