Reinforcement Learning

Popular Reinforcement Learning Algorithms

Learning Objectives

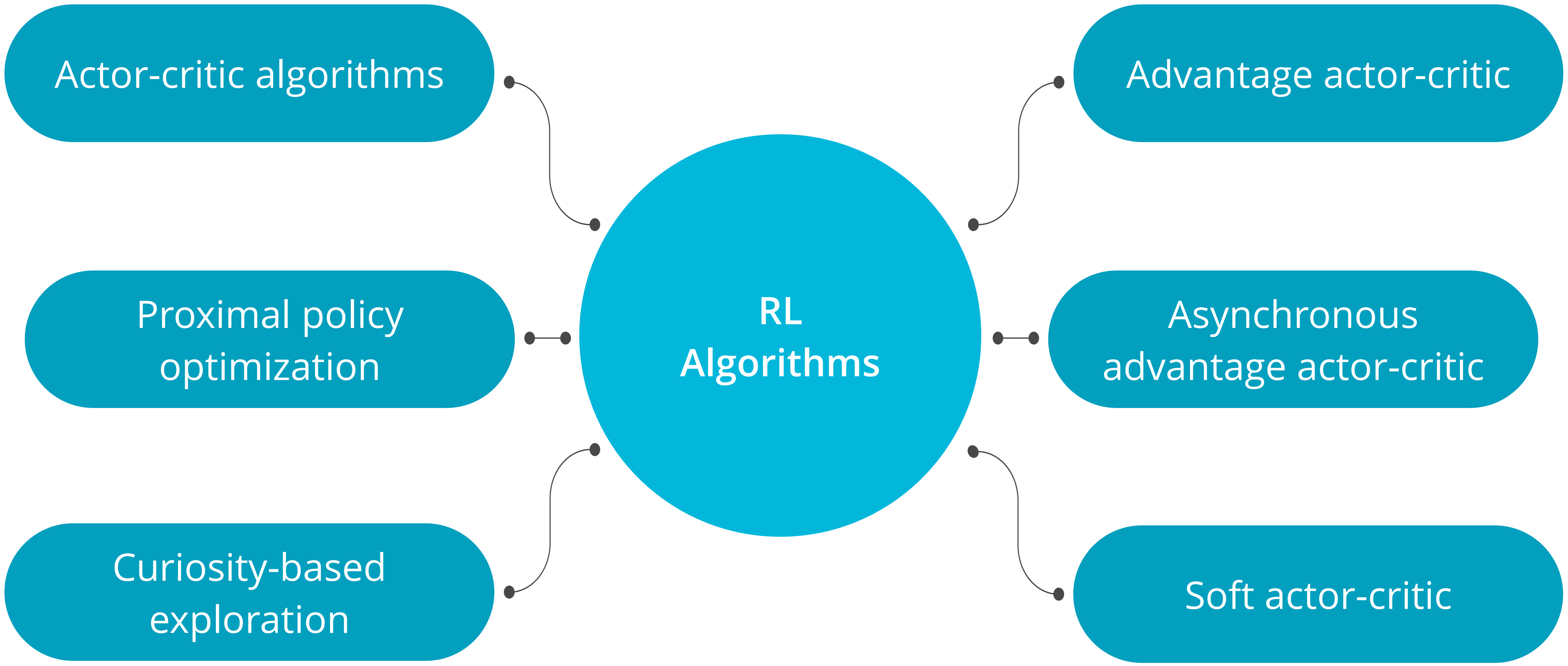
By the end of this lesson, you will be able to:

- Explore various popular RL algorithms
- Describe algorithms like actor critic, proximal policy optimization, curiosity-based algorithms, etc
- Discuss the advantages of these algorithms

Overview of Popular RL Algorithms

Popular Reinforcement Learning Algorithms

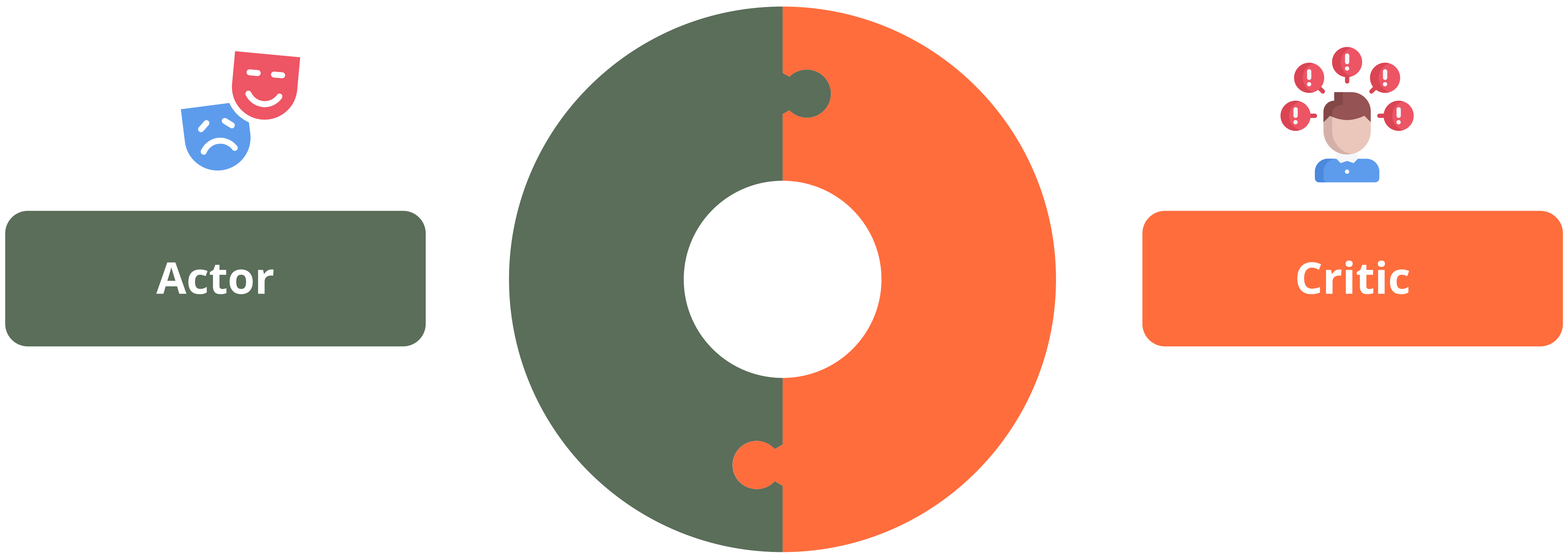
Several reinforcement learning (RL) algorithms are popular. A few of these algorithms are mentioned here:



Actor-Critic Algorithms

Actor-critic algorithms are a class of reinforcement learning (RL) methods that combine elements of both value-based and policy-based approaches.

The main idea is to have two components working together:



Actor-Critic Algorithms

Actor-critic algorithms implements combination of strategies by integrating the concept of value-based learning (DQN) and policy-based strategies, aiming to leverage the strengths of both.

Policy network (actor)

Directly maps states to actions, deciding the best course of action based on the current policy.

Actor-Critic Algorithms

Actor-critic algorithms implements combination of strategies by integrating the concept of value-based learning (DQN) and policy-based strategies, aiming to leverage the strengths of both.

Value network (critic)

Critiques the actions taken by the actor based on the q-values, guiding the actor to improve its policy.

Actor-Critic Algorithms

Actor-critic algorithms implements combination of strategies by integrating the concept of value-based learning (DQN) and policy-based strategies, aiming to leverage the strengths of both.

Faster and more stable learning

The critic's guidance helps the actor learn more robust policies faster than traditional policy gradient methods.

Asynchronous Advantage Actor-Critic

Asynchronous advantage actor-critic or A3C is a reinforcement learning algorithm designed for training deep neural network policies in an asynchronous and distributed manner.

The key features include:

Distributed agents

Multiple agents operate in their own instance of the environment, promoting diverse experience collection and faster learning.

Asynchronous Advantage Actor-Critic

Asynchronous advantage actor-critic or A3C is a reinforcement learning algorithm designed for training deep neural network policies in an asynchronous and distributed manner.

The key features include:

Asynchronous updates

Agents periodically push updates to a central model and fetch the latest model, ensuring a continuous and diverse learning process.

Asynchronous Advantage Actor-Critic

Asynchronous advantage actor-critic or A3C is a reinforcement learning algorithm designed for training deep neural network policies in an asynchronous and distributed manner.

The key features include:

Advantage function

Focuses on how much better an action is compared to the average, helping in more nuanced policy updates.

Advantage Actor-Critic

Advantage actor-critic builds upon the asynchronous advantage actor-critic (A3C) algorithm but focuses on a synchronous implementation, making it simpler and more stable for single-machine training.

The key features are:

Synchronization for efficiency

A2C synchronizes the update steps of all parallel agents, making it more efficient on modern hardware by effectively utilizing batch processing.

Advantage Actor-Critic

Advantage actor-critic builds upon the asynchronous advantage actor-critic (A3C) algorithm but focuses on a synchronous implementation, making it simpler and more stable for single-machine training.

The key features are:

Batch updates

Larger, synchronized updates provide more stable and reliable gradient estimates, improving the learning process.

Soft Actor-Critic

Soft actor-critic (SAC) is a state-of-the-art reinforcement learning algorithm designed for training agents in continuous action spaces.

The key features are:

Entropy as a goal

A

SAC adds an entropy term to the reward, encouraging the actor to explore more diverse strategies.

Soft Actor-Critic

Soft actor-critic (SAC) is a state-of-the-art reinforcement learning algorithm designed for training agents in continuous action spaces.

The key features are:

It maintains a balance between exploring new strategies and exploiting known ones, leading to more robust and effective policies.

B

Balancing exploration and reward

Soft Actor-Critic

Soft actor-critic (SAC) is a state-of-the-art reinforcement learning algorithm designed for training agents in continuous action spaces.

The key features are:

C

Efficiency and stability

Known for its sample efficiency, stability, and ability to solve a wide range of challenging tasks.

Proximal Policy Optimization

Proximal policy optimization (PPO) is a RL algorithm introduced by open to address issues related to policy optimization.

It aims to strike a balance between sample efficiency, stability, and ease of implementation.

Clipped objective

PPO modifies the policy gradient objective to limit the changes made to the policy at each step, preventing drastic updates that can harm learning.

Proximal Policy Optimization

Proximal policy optimization (PPO) is a RL algorithm introduced by open to address issues related to policy optimization.

It aims to strike a balance between sample efficiency, stability, and ease of implementation.

Balancing old and new policies

Ensures the new policy doesn't deviate too much from the old, maintaining a stable and gradual improvement.

Proximal Policy Optimization

Proximal policy optimization (PPO) is a RL algorithm introduced by open to address issues related to policy optimization.

It aims to strike a balance between sample efficiency, stability, and ease of implementation.

Wide adoption

Due to its simplicity and effectiveness, PPO has become a popular choice for a variety of tasks, from games to robotics.

Curiosity-Based Exploration

Unlike traditional methods that rely solely on extrinsic rewards from the environment, curiosity-based exploration emphasizes the learning process itself as a source of motivation.

Intrinsic motivation

Models an agent's curiosity by rewarding it for exploring actions that lead to unexpected or less predictable results.

Curiosity-Based Exploration

Unlike traditional methods that rely solely on extrinsic rewards from the environment, curiosity-based exploration emphasizes the learning process itself as a source of motivation.

Predictive model

Employs a predictive model of the environment's dynamics, finding interest in areas where the prediction error is high.

Curiosity-Based Exploration

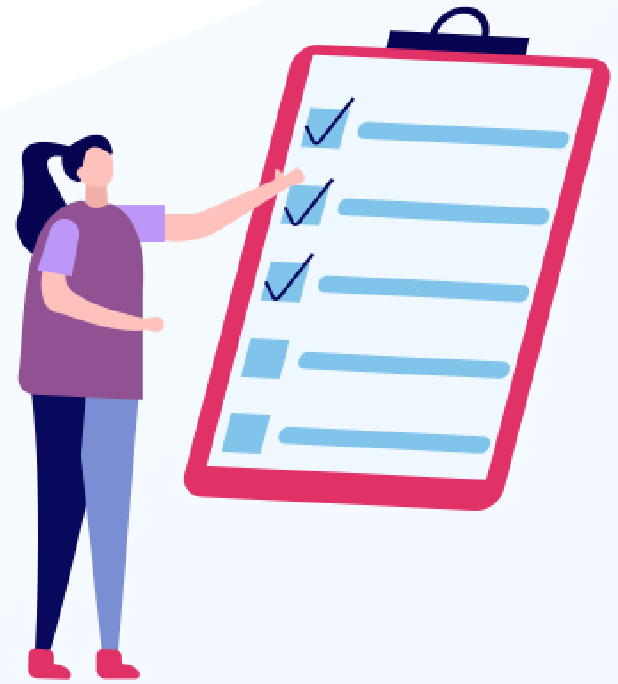
Unlike traditional methods that rely solely on extrinsic rewards from the environment, curiosity-based exploration emphasizes the learning process itself as a source of motivation.

Overcoming sparse rewards

Particularly effective in environments where external rewards are rare or difficult to obtain, encouraging the agent to explore and learn from the intrinsic structure of the environment.

Key Takeaways

- Actor-critic blends value-based learning (DQN) and policy-based strategies, harnessing the strengths of both approaches.
- A3C is a reinforcement learning algorithm designed for training deep neural network policies in an asynchronous and distributed manner.
- Advantage actor-critic improves on A3C with a synchronous approach, simplifying and stabilizing single-machine training.
- Proximal policy optimization (PPO) aims to strike a balance between sample efficiency, stability, and ease of implementation.
- Curiosity-based exploration emphasizes the learning process itself as a source of motivation.



Knowledge Check

Knowledge Check 1

Which of the following RL algorithms is characterized by the incorporation of both policy and value functions, providing a more stable and efficient learning process?

A. A3C (asynchronous advantage actor-critic)

B. PPO (proximal policy optimization)

C. Actor-critic algorithms

D. Soft actor-critic

Knowledge Check 1

Which of the following RL algorithms is characterized by the incorporation of both policy and value functions, providing a more stable and efficient learning process?

A. A3C (asynchronous advantage actor-critic)

B. PPO (proximal policy optimization)

C. Actor-critic algorithms

D. Soft actor-critic

The correct answer is **C**

The actor-critic algorithm combines the advantages of both policy-based (actor) and value-based (critic) methods, offering a balance between stability and efficiency in learning.

Knowledge Check

2

Which algorithm is designed to address the issue of high variance in policy gradient methods by using a trust region constraint?

A. Actor-critic algorithms

B. PPO (proximal policy optimization)

C. Curiosity-based exploration

D. A3C (asynchronous advantage actor-critic)

Knowledge Check 2

Which algorithm is designed to address the issue of high variance in policy gradient methods by using a trust region constraint?

A. Actor-critic algorithms

B. PPO (proximal policy optimization)

C. Curiosity-based exploration

D. A3C (asynchronous advantage actor-critic)

The correct answer is **B**

PPO addresses policy gradient method variance by using a trust region constraint, ensuring the new policy doesn't deviate significantly from the old one, promoting stability in learning.

Thank You