



4-1 Discussion: Applying Reinforcement Learning

[Listen](#)[Back to Topic](#)

4-1 Discussion: Applying Reinforcement Learning

Joel De Alba posted Aug 4, 2023 12:42 AM [Subscribed](#)

Reinforcement Learning is a branch of machine learning that deals with a process of learning and controlling strategies

to interact with a complex environment. RL is essentially a framework for learning from experience based on The Agent,

The Policy, $\pi(s,a)$ State s , Action a , Reward r , and Environment.

In simple terms, Compared to humanity; its like like learning as humans do when we learn by trial and error.

Good behavior is rewarded, while bad behavior or otherwise is decreased or negative.

We start with:

1. State
2. Agent
3. Environment
4. Action
5. Policy
6. Rewards

In RL the State is essentially like a snapshot of the environment at a particular moment, this means it represents

any relevant information that the agent requires in order to make descisions.

For example, if we consider a simple game like Tic-Tac-Toe, the state would be the current configuration of the game

board, showing where X's and O's are placed. The state tells the agent where things currently stand in the game.

Similarly, in more complex scenarios like training an AI to play a car racing game, the state would include information like the car's current position on the track, its speed, and the direction it's facing. It may also include details about the track layout and any obstacles nearby. A good example of this is illustrated by videos circulating the web of developers utilizing AI to learn how to finish a race in the video game "TrackMania" a PC videogame that involves a plethora of track customization, formula one racing vehicles, timed races, etc. The computer learns to drive by committing hundreds of errors and falling off the racing platform, or crashing several times, not making it in the limit of time, etc. A positive reward is given when the car passes through a checkpoint or reaches the finish line. Negative rewards may be given when the car goes off-track or crashes. The magnitude of the rewards can depend on factors such as speed, accuracy, and distance from the optimal path.

The Agent here is the vehicle itself.

The Environment is the track that is selected to learn from.

(The Agent takes certain actions to interact with the Environment it is given.)

Action is related to a variety of variables like to move forward, backward, left, right, and to also study the amount of threshold or how fast to go, what direction is optimal and does not cause failure, etc.

The Policy is essentially a type of strategy utilized to achieve goals with maximum reward gains.

The Rewards in RL are real-valued and represent how well the car is performing.

This form of learning in TrackMania is known as Supervised learning.

In supervised learning, the training process involves using labeled data, where each input (state) is paired with the corresponding correct output (action). The objective is for the model to learn a mapping from inputs to outputs based on the provided labeled data. During training, the model tries to minimize the error between its predictions and the actual outputs.

In the Trackmania example, the model is trained using labeled data consisting of states and corresponding actions, this is supervised learning. On the other hand, The Markov Decision Process (MDP) is a

framework for reinforcement learning, where the agent learns to make decisions by interacting with an environment and receiving rewards or penalties based on its actions.

The differences are:

Supervised vs. Reinforcement Learning

MDP is a reinforcement learning framework that works by having the agent learn from interactions with the environment, whereas this type of Supervised learning works by implementing labeled data and calculating optimal actions for each state.

Model Predictions vs. Policy Learning:

In the Trackmania example, the model predicts optimal actions directly from states. In MDP, the agent learns a policy, which is a mapping from states to actions based on expected rewards.

Deterministic vs. Stochastic Transitions:

In the Trackmania example, transitions between states are deterministic because the model is trained using specific actions for each state. In MDP, transitions are probabilistic, meaning the agent's actions may lead to different states with certain probabilities

Continuous vs. Discrete Actions:

The Trackmania example uses regression training to predict continuous action values for precise control of the car's movements. In MDP, the agent often deals with discrete action spaces, making it more challenging to control continuous movements.

Sources

Steve Brunton. (2021, February 12). Reinforcement Learning: Machine Learning meets control Theory [Video]. YouTube. <https://www.youtube.com/watch?v=0MNVhXEX9to>

Robins, A. (2021, December 22). Stochastic vs Deterministic Models. Understand the Pros and Cons. Retrieved July 26, 2023, from <https://blog.ev.uk/stochastic-vs-deterministic-models-understand-the-pros-and-cons>

Leonelli, M. (2021, April 19). 1.2 Types of simulations | Simulation and Modelling to Understand Change. https://bookdown.org/manuele_leonelli/SimBook/types-of-simulations.html

Das, A. (2017). The very basics of reinforcement learning. Medium. Retrieved from <https://becominghuman.ai/the-very-basics-of-reinforcement-learning-154f28a79071>

Beysolow, I. T. (2019). Applied reinforcement learning with python : With openai gym, tensorflow, and keras. Apress L. P..

Lamba, A. (2018). A brief introduction to reinforcement learning - We've moved to freeCodeCamp.org/news - Medium. Medium. <https://medium.com/free-code-camp/a-brief-introduction-to-reinforcement-learning-7799af5840db>

Reply to Thread

Filter by: [All Posts](#) ▼ | [Clear filters](#)

Show: [Threaded](#)

There are no replies in this thread

Reply to Thread

 [Reflect in ePortfolio](#)




Activity Details

Well done! You have contributed to the discussion

Assessment

0 / 30 F

 Available on Jul 15, 2023 10:59 PM. **Submission restricted before availability starts.**



Discussion Rubric:
Undergraduate

Last Visited Aug 4, 2023 11:37 AM