REINFORCEMENT LEARNING: AN ANALYTICAL OVERVIEW AND OUTCOMES

[1]Deepak Kumar
Assistant Professor (Comp. Sc.)
Govt. PG College Sector-1, Panchkula
[2]Parul
Master of Technology (CSE)
Maharshi Dayanand University Rohtak

Abstract:*Reinforcement learning is an area of Machine Learning. Reinforcement. 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 behaviour or path it should take in a specific situation. Reinforcement learning differs from the supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task.Reinforcement learning (RL) is a field of research that uses dynamic programming, among other approaches, to solve sequential decision-making problems [1]. In the absence of training dataset, it is bound to learn from its experience.*

Keywords: Machine learning, Supervised Learning, Software and machines etc.

Introduction: In the last few years, Reinforcement Learning (RL), also called adaptive (or approximate) dynamic programming (ADP), has emerged as a powerful tool for solving complex sequential decision-making problems in control theory. Although seminal research in this area was performed in the artificial intelligence (AI) community, more recently, it has attracted the attention of optimization theorists because of several noteworthy success stories from operations management [2]. In general terms, AI refers to a broad field of science encompassing not only computer science but also psychology, philosophy, linguistics and other areas. AI is concerned with getting computers to do tasks that would normally require human intelligence. Having said that, there are many points of views on AI and many definitions exist. Below some AI definitions which highlight key characteristics of AI.

Reinforcement learning solves a different kind of problem. In RL, there's an agent that interacts with a certain environment, thus changing its state, and receives rewards (or penalties) for its input. Its goal is to find patterns of actions, by trying them all and comparing the results, that yield the most reward points.

One of the key features of RL is that the agent's actions might not affect the immediate state of the environment but impact the subsequent ones. So, sometimes, the machine doesn't learn whether a certain action is effective until much later in the episode.Aiming to maximize the numerical reward, the agent has to lean toward actions that, it knows, lead to positive results and avoid the ones that don't. This is called **exploitation** of the agent's knowledge.

However, to find out which actions are correct the first place it must try them out and run the risk of getting a penalty. This is known as **exploration**.Balancing exploitation and exploration is one of the key challenges in Reinforcement Learning and an issue that doesn't arise at all in pure forms of supervised and unsupervised learning.

Some of the practical applications of reinforcement learning are:

Inventory Management

A major issue in supply chain inventory management is the coordination of inventory policies adopted by different supply chain actors, such as suppliers, manufacturers, distributors, so as to smooth material flow and minimize costs while responsively meeting customer demand.Reinforcement learning algorithms can be built to reduce transit time for stocking as well as retrieving products in the warehouse for optimizing space utilization and warehouse operations.

Delivery Management

Reinforcement learning is used to solve the problem of Split Delivery Vehicle Routing. Q-learning is used to serve appropriate customers with just one vehicle.

Power Systems

Reinforcement Learning and optimization techniques are utilized to assess the security of the electric power systems and to enhance Microgrid performance. Adaptive learning methods are employed to develop control and protection schemes. Transmission technologies with High-Voltage Direct Current (HVDC) and Flexible Alternating Current Transmission System devices (FACTS) based on adaptive learning techniques can effectively help to reduce transmission losses and CO2 emissions.

RL, known as a semi-supervised learning model in machine learning, is a technique to allow an agent to take actions and interact with an environment so as to maximize the total rewards. RL is usually modelled as a Markov Decision Process (MDP).

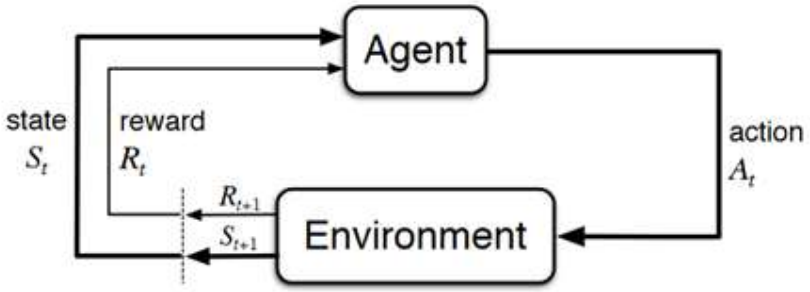

Figure: The Agent- Environment interaction in Markov decision making process.

The study of RL is to construct a mathematical framework to solve the problems. For example, to find a good policy we could use valued-based methods like Q-learning to measure how good an action is in a particular state or policy-based methods to directly find out what actions to take under different states without knowing how good the actions are.

However, the problems we face in the real world can be extremely complicated in many different ways and therefore a typical RL algorithm has no clue to solve. For example, the state space is very large in the game of GO, environment cannot be fully observed in Poker game and there are lots of agents interact with each other in the real world. Researchers have invented methods to solve some of the problems by using deep neural network to model the desired policies, value functions or even the transition models, which therefore is called Deep Reinforcement Learning.

Resources management in computer clusters

Designing algorithms to allocate limited resources to different tasks is challenging and requires human-generated heuristics. The paper "Resource Management with Deep Reinforcement Learning" [3] showed how to use RL to automatically learn to allocate and schedule computer resources to waiting jobs, with the objective to minimize the average job slowdown. State space was formulated as the current resources allocation and the resources profile of jobs. For action space, they used a trick to allow the agent to choose more than one action at each time step. Reward was the sum of (-1/duration of the job) over all the jobs in the system. Then they combined REINFORCE algorithm and baseline value to calculate the policy gradients and find the best policy parameters that give the probability distribution of actions to minimize the objective.

Traffic Light Control

In the paper "Reinforcement learning-based multi-agent system for network traffic signal control"[4], researchers tried to design a traffic light controller to solve the congestion problem. Tested only on simulated environment though, their methods showed superior results than traditional methods and shed a light on the potential uses of multi-agent RL in designing traffic system.

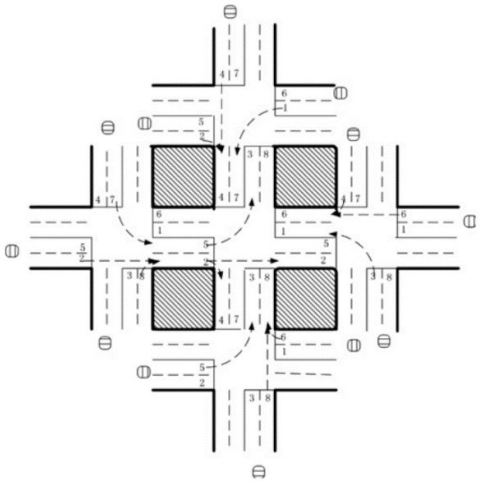

Figure: Five-intersection traffic network

Web System Configuration There are more than 100 configurable parameters in a web system and the process of tuning the parameters requires a skilled operator and numerous trail-and-error tests. The paper "A Reinforcement Learning Approach to Online Web System Auto-configuration" [5] showed the first attempt in the domain on how to do autonomic reconfiguration of parameters in multi-tier web systems in VM-based dynamic environments.

Robotics

There are tremendous work on applying RL in Robotics. Readers are referred to [6] for a survey of RL in Robotics. In particular, [7] trained a robot to learn policies to map raw video images to robot's actions. The RGB images were fed to a CNN and outputs were the motor torques. The RL component was the guided policy search to generate training data that came from its own state distribution.

Used in Machine Learning:

Reinforcement Learning is one of the major topics in Machine Learning and is currently in trend and is a major source of attraction for many researchers and developers.

- **Used in creating training systems and self-operating systems:**
  reinforcement learning is being used to create different self-operating systems like self-driving cars,

automated arms, House cleaning agents, etc. Apart from these, many training systems are also designed using it like the test conducting systems, systems which are able to have a human-like communication, systems which are able to receive any call and reply as per the going conversation, etc.

- **Used in Data Processing:**
  In Intelligent agents and expert systems which work in a dynamic and partially observable environment, the reinforcement learning is an effective and widely used way for Data Processing, as the conditions with uncertainty can be easily handled using it.

Limitations and Scope

Most of the reinforcement learning methods we consider in this book are structured around estimating value functions, but it is not strictly necessary to do this to solve reinforcement learning problems. For example, methods such as genetic algorithms, genetic programming, simulated annealing, and other optimization methods have been used to approach reinforcement learning problems without ever appealing to value functions. These methods evaluate the "lifetime" behaviour of many non-learning agents, each using a different policy for interacting with its environment, and select those that are able to obtain the most reward [8].

We call these evolutionary methods because their operation is analogous to the way biological evolution produces organisms with skilled behaviour even when they do not learn during their individual lifetimes. If the space of policies is sufficiently small, or can be structured so that good policies are common or easy to find—or if a lot of time is available for the search—then evolutionary methods can be effective. In addition, evolutionary methods have advantages on problems in which the learning agent cannot accurately sense the state of its environment. Our focus is on reinforcement learning methods that involve learning while interacting with the environment, which evolutionary methods do not do (unless they evolve learning algorithms, as in some of the approaches that have been studied) [9].

It is our belief that methods able to take advantage of the details of individual behavioural interactions can be much more efficient than evolutionary methods in many cases. Evolutionary methods ignore much of the useful structure of the reinforcement learning problem: they do not use the fact that the policy they are searching for is a function from states to actions; they do not notice which states an individual pass through during its lifetime, or which actions it selects. In some cases, this information can be misleading (e.g., when states are misperceived), but more often it should enable more efficient search. Although evolution and learning share many features and naturally work together, we do not consider evolutionary methods by themselves to be especially well suited to reinforcement learning problems. For simplicity, in this book when we use the term "reinforcement learning method" we do not include evolutionary methods.

Conclusion

Reinforcement learning is also different from what machine learning researchers call unsupervised learning, which is typically about finding structure hidden in collections of unlabelled data. The terms supervised learning and unsupervised learning appear to exhaustively classify machine learning paradigms, but they do not. This article just showed some of the examples of RL applications in various industries. They should not limit your RL use case and as always, you should use first principle to understand the nature of RL and your problem [10].Although one might be tempted to think of reinforcement learning as a kind of unsupervised learning because it does not rely on examples of correct behaviour, reinforcement learning is trying to maximize a reward signal instead of trying to find hidden structure. Reinforcement learning takes the opposite tack, starting with a complete, interactive, goal-seeking agent. All reinforcement learning agents have explicit goals, can sense aspects of their environments, and can choose actions to influence their environments. Moreover, it is usually assumed from the beginning that the agent has to operate despite significant uncertainty about the environment it faces.

References:

[1] Nir Levine, Tom Zahavy, Daniel J. Mankowitz, Aviv Tamar: Shallow Updates for Deep Reinforcement Learning.
[2] Abhijit Gosavi: Reinforcement Learning: A Tutorial Survey and Recent Advances, http://web.mst.edu.
[3] H.Mao, Alizadeh, M. Alizadeh, Menache, I.Menache, and S.Kandula. Resource Management With deep Reinforcement Learning. In ACM Workshop on Hot Topics in Networks, 2016.
[4] I. Arel, C. Liu, T. Urbanik, and A. Kohls, "Reinforcement learning-basedmulti-agent system for network traffic signal control,"IETIntelligentTransport Systems, 2010.
[5] X. Bu, J. Rao, C. Z. Xu. A reinforcement learning approach to online web systems auto-configuration. In Distributed Computing Systems, 2009. ICDCS'09.29th IEEE International Conference on. IEEE , 2019.
[6] J. Kober, J. A. D. Bagnell, J. Peters. Reinforcement Learning in Robotics: A survey. Int. J. Robot. Res. Jul. 2013.
[7] S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end Training of Deep Visuomotor Policies. arXiv preprint arXiv:1504.00702, 2015.
[8] Reinforcement Learning: An Introduction Second edition, Richard S. Sutton and Andrew G. Barto c 2014, 2015.
[9] Reinforcement Learning: An Introduction Second edition, Richard S. Sutton and Andrew G. Barto c 2014, 2015.
[10] https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12\