International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 4, May-June 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Review on Introduction to Reinforcement Learning

Shreya Khare[1], Yogeshchandra Puranik[2]

[1]PG Student, [2]Assistant Professor,
[1,2]Affiliated to Department of (MCA), P.E.S.'s Modern college of Engineering, Pune, Maharashtra, India

ABSTRACT

This paper aims to introduce, review and summarize the basic concepts of reinforcement learning. It will provide an introduction to reinforcement learning in machine learning while covering reinforcement learning workflow, types, methods and algorithms used in it.

How to cite this paper: Shreya Khare | Yogeshchandra Puranik "A Review on Introduction to Reinforcement Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.1096-1099, URL: www.ijtsrd.com/papers/ijtsrd42498.pdf

IJTSRD42498

INTRODUCTION:

It was in the 1940s when ENIAC (Electronic Numerical Integrator and Computer), was designed. It was the first manually operated computer system. During that period "computer" was being used as a name for a human with intensive numerical computation capabilities, so, ENIAC was called a numerical computing machine. The idea was to build a machine that has the ability to match human thinking and learning. The researchers observed that computers could recognize patterns and developed a theory that machines can learn and can be automated to perform a specific task.

The term Machine Learning was framed in the year 1959 by **Arthur Samuel**. He described Machine Learning as "**Field of study that gives computers the ability to learn without being explicitly programmed".** Machine learning began to pick up speed in 90's and was separated from artificial intelligence and became a unique field rooted in statistical modelling and probability theory. Machine learning is a discipline where we study the computer algorithms that provide better results based on information given to the computer and its experience. One of the aspects of Machine Learning is Reinforcement learning. Reinforcement Learning in Machine Learning is the training of machine learning models to take suitable actions based on the reward and feedback they receive for those actions. The models learn by interacting with the environment and experience successes and failures while performing actions. The history of reinforcement learning comes from different directions. Firstly, learning by trial and error and started in the psychology of animal learning. This is used in some of the earliest work in artificial intelligence and led to the revival of reinforcement learning in the early 1980s. Second is the problem of optimal control and its solution using value functions and dynamic programming. This did not involve learning for most of the part. Although these two directions have been largely independent, the exceptions revolve around a third direction, concerning temporal-difference methods such as used in the tic-tac-toe example. All these came together in the late 1980s to produce the modern field of reinforcement learning. Another neural-network learning machine was designed to learn by trial and error by Farley and Clark. It was in the 1960s when the terms "reinforcement" and "reinforcement learning" were used for the first time in the engineering literature (e.g., Waltz and Fu, 1965; Mendel, 1966; Fu, 1970; Mendel and McClaren, 1970).

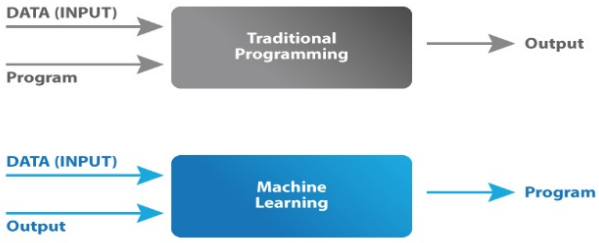
Machine Learning:

Machine learning is a subset of AI. Machine learning algorithms are used by the computers to learn from the data and past experiences to provide better results without the need to program manually. In attempt to make machines more human like and improve their behaviour and decision making, machines are given the ability to learn and develop their own programs. The learning process is automated and improved based on the learning experience of the machine.

Data is given to the machines, and machines are trained on this data using various machine learning algorithms.

The choice of algorithm to be used is based on the type of data and task that needs to be automated. In traditional programing, input and a well written program is given and an output is produced. But in machine learning, data and

output is given to the machine and it produces a program by itself based on its previous results.



Types of Machine Learning:
Following are the types of machine learning

Supervised Learning:
In supervised learning, model is monitored or supervised in sense that we already know the output and the algorithms are corrected every time to improve the results. The algorithm identifies the mapping function between input and output variables. Supervised learning problems can be grouped as regression problems (model is trained with historical dataset and used to predict future values) and classification problems (labelled dataset trains algorithm to identify and categorize items).

Unsupervised Learning:
In unsupervised learning, training model has only input parameters and the output is unknown. The algorithm learns by itself and finds the structure in the dataset. Unsupervised learning can be grouped as clustering (finding a pattern in uncategorised data) and association (discovering existing relationships between variables).

Reinforcement learning:
Reinforcement learning is an area of machine learning where agents take actions in an environment in order to maximize the reward. Machines train themselves on reward and punishment mechanism. Here, an agent is built that can study and interact with the environment in which it is placed and take actions. The agent in an interactive environment, learns from its own actions and experience through trial and error method.

Exploration and exploitation:
Reinforcement learning uses the technique exploration and exploitation. The agent explores the sample space and learns new and better strategies and exploits by greedily using the best available strategy to obtain the results. Since exploration is costly in terms of resource, time and opportunity, this raises a question about how much to exploit and how much to explore. The agent has to balance between greedily exploiting what it has learnt so far to yield the maximum reward and continuously explore the environment to acquire more information and achieve higher value in long term.

Components of Reinforcement Learning:
Some basic terms in Reinforcement Learning

- Agent: an entity that makes the decision to get maximum reward and learns by interacting with the environment.
- Environment: sample space where agent decides the action to be performed.
- Action: steps performed by the agent which is based on state of the environment.
- State: the situation in which the agent is present in the particular instance of time.
- Reward: a scalar value as a feedback from the environment.
- Policy: strategy prepared by the agent to map current state to the next action.
- Model: different dynamic states of an environment and how these states lead to a reward.
- Value Function: estimates the value of state which shows the achieved reward of being in a state

Workflow:
Firstly, you need to define the environment for the agent and an interface between agent and the environment.

Next, define a reward to measure the performance of the agent against the actions.

Then to create an agent, you have to decide a method to represent the policy and the select a training algorithm for the agent.

Then train the agent to adapt to the policy. Training the agent in reinforcement learning is an iterative process.

Validate the trained policy after training ends

And lastly deploy the trained policy.

Characteristics:
1. The reward signals act as the feedback.
2. The decision making is sequential.
3. As there can be multiple solutions for a problem, many outcomes are possible.
4. The agent's action decides the succeeding data it will receive.
5. Delayed feedback.

Types:

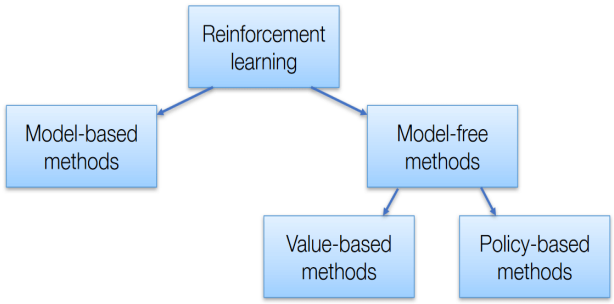
- **Positive Reinforcement:**

When an event occurs due to a specific behaviour, which strengthens the frequency of behaviour and has a positive impact on the actions taken by the agent, it is known as positive reinforcement. Positive reinforcement helps to endure change for a long time and increases performance. But too much positive reinforcement may overload the states and limit the results.

- **Negative Reinforcement:**

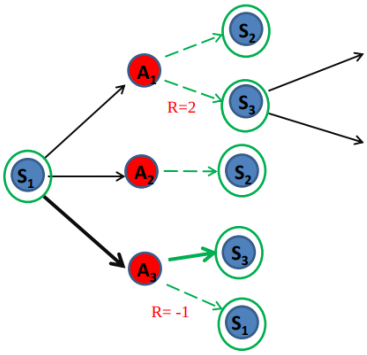
When an event occurs due to removal of negative stimuli or negative condition that strengthens the behaviour of the agent, it is known as negative reinforcement. It helps agent to meet a particular level of performance.

Reinforcement Learning Methods:
3 methods of reinforcement learning are:
1. Model based
2. Value based
3. Policy based



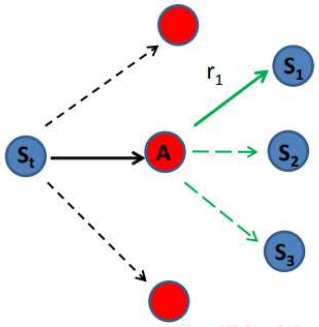
The problem we often face in reinforcement learning is you might not necessarily know the next state you'll end up in.

- **Model based method**:



In Model-based approach, you either have the access to the model or you build the model. Either way you can determine the probability distribution over states you end up in.

- **Model free method:**



In Model-free approach, you are not given a model and you don't try to figure out how it works. Optimal policy is derived through experience and interaction with the environment.

In simple words, the agent exploits a previously learned model to accomplish the task in model based learning, whereas in model free learning, the agent simply relies on trial and error experience for taking action.

- **Value based method**:

The value based method is used to find the maximum value of a state under any policy. We find the optimal value function. Here the agent acts by choosing the best action in the state. Here, exploration is necessary.

- **Policy based method:**

Policy based method is used to find the optimal policy that maps state to action without using value function (select action without using a value function). Surely we can use the value function to optimize the policy parameters, but there is no need of value function to select an action.

- Deterministic: Policy that defines clear and defined action for every state.
- Stochastic: Policy that defines probability distribution for the actions to take from that state.

Markov Decision Process:

Markov Decision process validates the reinforcement learning problems. It is used to mathematically describe the interaction between the agent and the controlled environment.

Elements needed to represent Markov Decision Process are State, Action, Reward and Probability.

The agent and the environment interacts at definite time t where t = 0, 1, 2, 3…At each time step, the agent gets information about the environment state S_t. Based on the environment state at instant t, the agent chooses an action A_t. In the following instant, the agent also receives a numerical reward signal R_{t+1}. This thus gives rise to a sequence like S_0, A_0, R_1, S_1, A_1, R_2…The random variables R_t and S_t have well defined discrete probability distributions. These probability distributions are dependent only on the preceding state and action. Let S, A, and R be the sets of states, actions, and rewards. Then the probability that the values of S_t, R_t and A_t taking values s`, r and a with previous state s is given by, $p(s', r|s, a) = P\{S_t = s', R_t = r\ |S_{t-1} = s, A_{t-1} = a\}$ The function p controls the dynamics of the process.

Markov Property

If an agent is in state S_1 and moves to next state $S_2$by performing an action A_1, then the state change from S_1 to S_2 depends only on the current state and future action and is independent of past state, action or reward.

Finite MDP:

A finite MDP is where there are finite states, finite actions and finite rewards.

RL considers only finite MDP.

Markov Chain:

Also known as Markov Process, Markov chain is a memory less process of transitioning from one state to another state according to some probabilistic rules. To determine the probability distribution of current state, we only need the knowledge of previous state. But the probability of moving from one state to another may change with time.

Reinforcement learning algorithm:

Reinforcement learning algorithms are categorised as model based and model free, which are further classified as on-policy and off-policy. In model based algorithm, you need a model and store all the states and actions data in the memory.

The model free algorithm works on trial and error basis, so you don't need to store states and actions in the memory. On-policy and off-policy learning, both involve a function Q(s, a). This function predicts the future reward by learning future state and reward. It also include terms Target Policy and Behaviour Policy. Target policy is the policy that an agent is trying to learn. Behaviour policy is being used by the agent for interacting with the environment and select action.

On-Policy Learning:

On-policy involves learning from current state and actions. On-policy methods estimates the value of policy while using it for control. In short, [Target Policy == Behaviour Policy].Some examples of On-Policy algorithms are Policy Iteration, Value Iteration, Monte Carlo for On-Policy, Sarsa, etc.

- It tries to estimate or improve the policy that is used to make decisions,
- often use soft action choice, i.e. $\pi(s,a)>0, \forall a \pi(s,a)>0, \forall a$,
- it always keeps exploring and try to find the best policy that still explores

Off-Policy Learning:

Off-policy involves learning from random state and actions. In off-policy method, policy used to generate behaviour may be unrelated to the policy that is evaluated. In short, [Target Policy != BehaviourPolicy].Some examples of Off-Policy learning algorithms are Q learning.

- tries to evaluate the greedy policy while following a more probing scheme
- the policy used for behaviour should be soft
- policies may not be sufficiently similar
- may be slower, but remains more flexible if alternative routes appear.

Advantages and Disadvantages of Reinforcement Learning:

Advantages:

- It can solve complex problems and the obtained solutions will be accurate. We can build various problem solving models.
- It is similar to human learning, therefore it might give perfect results.
- The model undergoes through a rigorous training which takes time and helps correct errors. The model learns continuously and the mistake made earlier is unlikely to be repeated.
- The best part is, even if no training data is provided, the model will learn through experience gained.

Disadvantages:

- As the model is generally used to tackle complex problems, we will be wasting unnecessary processing power and space by using it for simpler problems.
- Reinforcement Learning models require a lot of training data to develop accurate results which consumes time and lots of computational power.
- While building real world models, we would require a lot of maintenance for hardware and software.

Conclusion:

The main purpose of this review paper is to establish a basic concept of Reinforcement Learning. Reinforcement Learning addresses the problem of finding optimal policy with least or no data. Reinforcement Learning learns continuously from its experience and gets better and better. Reinforcement Learning is one of the interesting and useful part of Machine Learning.

References:

[1] R. S. Sutton and A. G. Barto, Reinforcement learning : an introduction, 2nd ed. Cambridge, MA: Mit Press, 2017.

[2] Lectures of Stanford's CS234 by Emma Brunskil- CS234: Reinforcement Learning | Winter 2019

[3] http://www.google.co.in

[4] https://stackexchange.com/