



ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 8.4
IJAR 2023; SP4: 136-139

Palak Kalsi
Lecturer, DPG Polytechnic
Sec-34, Gurgaon, Haryana,
India

Sucheta Mor
Lecturer, DPG Polytechnic
Sec-34, Gurgaon, Haryana,
India

(Special Issue)

**“National Conference on
Multidisciplinary research for sustainable development”**

Role of reinforcement learning in machine learning

Palak Kalsi and Sucheta Mor

Abstract

In today's world where everything is related to AI (Artificial Intelligence), more focus is put on machine learning. In the hierarchy of AI, it consists of deep learning and further covers ML. So, maybe we can say that machine learning is a sub-part of artificial intelligence, which is basically defined as the scientific study of algorithms that help programs improve themselves. In the normal world, learning is led by whether there is a supervisor to teach or without a supervisor by assuming. There is also another way in which anyone can learn by experience by counting good and bad experiences. But machine learning uses all types of learning, such as supervised, unsupervised and reinforcement learning. According to reinforcement learning, an agent placed in an environment must learn how to behave successfully in that environment. In this paper, we introduce the concept of reinforcement learning and how it plays a crucial role in machine learning and what are the major advantages of reinforcement learning in the field of machine learning.

Keywords: ANN (Artificial neural networks), reinforcement learning, machine learning

Introduction

1. Machine Learning

In a normal world, learning means “to gain knowledge and understanding “and “Modification of behavioural tendency by experience [3]”. As related to machines, Machine learns whenever change in structure, program and data in such a manner that expected future performance improves. These types of changes can include additional records to databases. It is usually referring to changes in systems that perform tasks associated with AI [3]. These tasks can be recognition, diagnosis, planning, prediction etc. Changes can be enhanced to present performing system or ab in to synthesis to new systems and these changes consider as learning, Different learning mechanisms might be depended on which subsystem is being changed [3].

2. Reinforcement Learning

As we know supervised learning agent needs to be told correct move for each state it encounters but it is not possible that this feedback is always available. In absence of feedback, agent must need to learn to own its own and can predict but agent don't know which is good or bad move. Agent needs to know what is good or what is bad this kind of feedback is known as reward or reinforcement. For e.g. - When infant starts to crawl any forward motion is an achievement or we can say that it is a reward. An agent needs to focus on reward any sensory input it sensed whether that is a reward or normal input. The Task of reinforcement learning is to use observed rewards to learn optimal policy [2]. In moreover domains, reinforcement learning is feasible way to train program to perform at higher levels. For e.g. In a game it is difficult for humans to provide accurate and consistent evaluation of positions then there is a need to train evaluation function with examples. Instead of told whether game is won or lost it can use information to learn evaluation function and give probability from any state to win [3].

Correspondence Author:
Palak Kalsi
Lecturer, DPG Polytechnic
Sec-34, Gurgaon, Haryana,
India

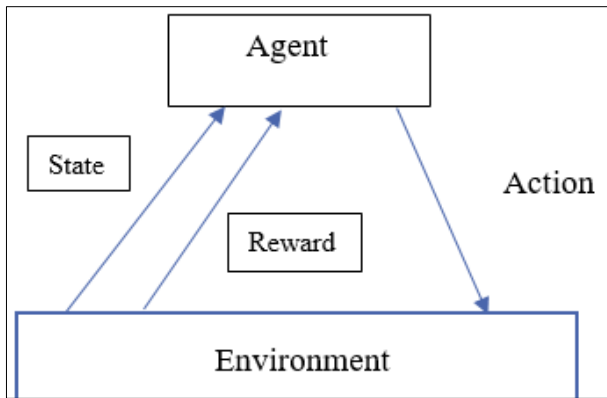


Fig 1: Basic Diagram of Reinforcement Learning-the agent exists environment described by some set of possible states(S) and it can perform various number of actions (A), each time agent performs an action(A) in some state(S) where agent will receive reward

How reinforcement learning superior from supervised and unsupervised learning

1. **Reinforcement learning and supervised learning:** Supervised learning is a learning from a training set of labelled examples provided by knowledgeable external supervisor. Each example under this type of learning is a set of examples with description of situation with a specification of correct action that system should take in order to manage the situation [1]. But in interactive problems, it is impractical to obtain examples of desired behaviour that are both correct and representative of all situations in which agent has to act. In this type of situation, we need a type of learning where agent must be able to learn from its own experiences which means reinforcement learning [1].
2. **Reinforcement Learning and Unsupervised Learning:** Unsupervised learning is generally about finding structure hidden in collection of unlabeled data. Whereas Reinforcement Learning is trying to maximize a signal instead of finding a hidden structure.

Basis	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	Learning from a training set of labelled data examples including problem and its solutions.	Learning finding structure that hidden in unlabeled data.	Learning from experiences. Basically, depends on rewards-negative and positive.
Supervision	Under supervision of knowledgeable external supervisor	No supervisor required	No supervisor required
Type of data	Labeled data	Unlabeled data	No predefined data
Usefulness	Do not predict in interactive problems. So, not useful during interactive problems.	Predict by understanding pattern and structure.	Follows up the trial-and-error method including exploration and exploitation.

1. **How Reinforcement Learning different from other approximation tasks in several important respects:**
 - a. **Delay Reward:** In general learning methods agent need to learn a target function that maps from current state to optimal action but in Reinforcement Learning, trainer only provide a sequence of immediate reward values as the agent executes its sequence of actions. The agent

- might faces a problem of temporal credit assignment: determine which of action in its sequence are to be added to produce positive rewards [2].
- b. **Exploration and Exploitation:** There are two challenges that reinforcement learning must face-exploration and exploitation [2], to obtain better result it is important to try those actions that has already experienced in past, this part is known as exploitation. Whereas, agent also have to explore in order to make better action selection in future. Agent always pursues progressively in order to appear best.
- c. **Partially Observable States:** In general, it is better if agent’s sensors can perceive the entire state of environment at each time but in practical situations sensors provide only partial information, in this type of situations agent need to consider its previous observations with its current sensor data when choosing actions [2].
- d. **Long term learning:** Agent able to learn several related tasks. In reinforcement learning it requires that agent learn related tasks within same environment using same sensors [2].

Elements of Reinforcement Learning

1. **Policy:** In reinforcement learning, a policy refers to the strategy or approach used by the learning agent to determine the appropriate action to take in a given state of the environment. It is essentially a mapping between the states that the agent perceives and the actions that it should take in response to those states. Therefore, a policy defines the behavior of the agent at any given moment in time by specifying the action that should be taken based on the current state of the environment.
2. **Reward:** In reinforcement learning, the reward signal is a crucial component that specifies the objective or goal of the learning problem. At each time step, the environment provides the agent with a scalar value known as the reward, which indicates the performance of the agent based on the action it has taken. The primary aim of the agent is to maximize the cumulative reward it receives over the entire learning process [2]. The reward signal essentially serves as a guide for the agent, helping it to distinguish between favorable and unfavorable outcomes, and ultimately, to learn an optimal policy that maximizes the total reward received over time [2].
3. **Value Function:** In reinforcement learning, a value function plays a critical role in determining the long-term desirability or utility of a given state in the environment. Specifically, the value of a state is defined as the expected amount of reward that an agent can accumulate over the future starting from that state [2]. Unlike rewards, which only provide immediate feedback on the desirability of states, the value function takes into account the expected future rewards and states that are likely to follow, providing a more comprehensive view of the long-term desirability of states. In essence, the value function allows the agent to evaluate the potential of different states in achieving its overall goal of maximizing cumulative rewards.
4. **Model of Environment:** The fourth and final element of some reinforcement learning systems is a model of the environment. This is something that mimics the behavior of the environment, or more generally, that

allows inferences to be made about how the environment will behave [2].

Learning Task for Reinforcement Learning

The learning task for reinforcement learning is to learn a policy that maximizes the total expected reward in a sequential decision-making process [4]. The agent interacts with an environment and takes actions based on the current state to receive rewards or penalties. The goal of the agent is to learn a policy that maps states to actions such that the expected sum of rewards over time is maximized [4].

The learning process can be broken down into the following steps:

1. The agent observes the current state of the environment.
2. The agent selects an action based on its policy.
3. The agent receives a reward or penalty based on the action taken and the resulting state.
4. The agent updates its policy based on the observed reward and the expected rewards of the next state.

This process is repeated over many episodes until the agent converges to an optimal policy. The challenge in reinforcement learning is to balance exploration and exploitation, where the agent needs to explore new actions to discover better policies while exploiting known actions to maximize the expected reward.

Markov Decision Process (MDP)

MDP stands for Markov Decision Process. It is a mathematical framework used in reinforcement learning and other areas of machine learning to model decision-making in a stochastic (probabilistic) environment. In an MDP, an agent interacts with an environment consisting of states, actions, rewards, and transition probabilities. The agent's goal is to learn a policy that maximizes the cumulative reward received over time.

The agent can perceive at set(s) of distinct states of its environment and a set of actions which it can be perform.

At each discrete time step(t) agent senses current state (s_t) choose a current action (a_t) and perform it and environment respond by giving reward $r_t = r(s_t, a_t)$ and produce a

succeeding state = s_{t+1} = value function (s_t, a_t) where value function and a_t are part of environment known by agent. The task of agent is to learn a policy is to select its next action based on current state. To specify which policy agent can learn. So, this is a way of cumulative reward [2].

Most Important Algorithm of Reinforcement Learning- Q-learning

Q-learning is a popular reinforcement learning algorithm that is used to learn the optimal action-value function, also known as the Q-function. The Q-function maps a state-action pair to an estimate of the expected cumulative reward that can be obtained by taking that action in that state and then following an optimal policy.

Q-learning is a model-free, off-policy algorithm, meaning that it does not require knowledge of the transition probabilities of the environment, and it can learn from data generated by any policy, even one that is not the optimal policy. The algorithm uses a table to store Q-values for each state-action pair, and it updates these values iteratively using the Bellman equation.

In Q-learning, at each time step, the agent observes the current state of the environment and selects an action to take based on its current policy. It then receives a reward and observes the next state of the environment. The Q-values in the table are updated based on this experience using the following update rule:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Where s is the current state, a is the action taken, r is the reward received, s' is the next state, a' is the optimal action in the next state, α is the learning rate, and γ is the discount factor that controls the trade-off between immediate and future rewards.

Over time, the Q-values in the table converge to their optimal values, and the agent can use them to select the action that maximizes the expected cumulative reward in each state, thus following an optimal policy. Q-learning is widely used in applications such as robotics, game AI, and recommendation systems.

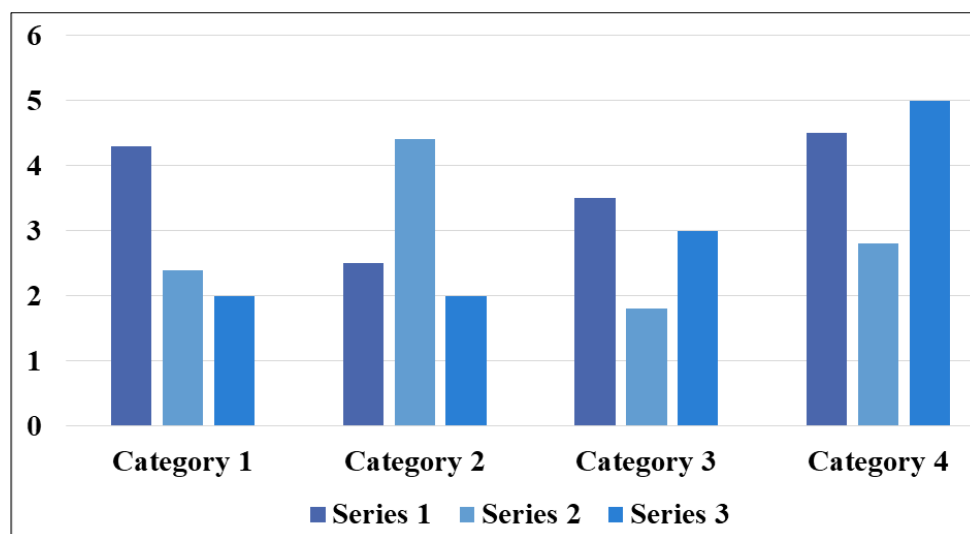


Fig 2: Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.

For more information about all elements of APA formatting, please consult the APA Style Manual, 6th Edition.

Conclusion

Reinforcement learning is a vast and complex field of study with a rich history, a diverse range of applications, a sophisticated theoretical foundation, impressive achievements, innovative algorithms, and numerous challenges yet to be solved. Recent research in artificial intelligence is focused on finding general principles of learning, decision-making, and search that can be applied across different domains, while also incorporating specialized knowledge relevant to specific areas. Reinforcement learning research is at the forefront of this effort to develop more efficient and effective methods for artificial intelligence that rely less on general principles and more on tailored approaches that can be adapted to specific problems and applications. Q-learning algorithms are off policy reinforcement learning algorithms that try to perform the most profitable action given the current state.

References

1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach, 3rd ed. Pearson. 2009.
2. Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction 2nd ed. The MIT press. 2018.
3. Christopher M. Bishop F.R. Eng, Pattern Recognition and Machine Learning, Springer. 2006.
4. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd ed. Springer. 2003.
5. Ahmad Hammoudeh. A Concise Introduction to Reinforcement Learning. 2018.
6. John Anderson, Ranan Banerji, Gary Bradshaw, Jaime Carbonell, Thomas Dietterich. Norman Haas Machine Learning: An Artificial Intelligence Approach, Morgan Kaufmann Publishers, INC. 1983.
7. Csaba Szepesvári. Algorithms for Reinforcement Learning. 2014.
8. Beakcheol Jang. Q-learning Algorithms: A Comprehensive Classification and Applications, IEEE Access. 2017.
9. Christopher JCH Watkins. Technical Note Q-Learning. 1989.