**A Seminar Report**

**on**

**“Reinforcement Learning”**

Submitted to the

Savitribai Phule Pune University

In partial fulfillment for the award of the Degree of Honours\*

in

Artificial Intelligence and Machine Learning

by

**Akshay Balasaheb Lasunkute**

(B150318509 / 47040 & BE IT B)

Under the guidance of

**Vandana Dixit**

****

Department Of Information Technology

P.E.S’s Modern College Of Engineering

Shivajinagar Pune - 411005 , Maharashtra, India

**2021-2022**

**I**

****

**CERTIFICATE**

This is to certify that the project based seminar report entitled **“Reinforcement Learning”** being submitted by **Akshay Balasaheb Lasunkute (B150318509 / 47040 & BE IT B)** is a record of bonafide work carried out by him under the supervision and guidance of **Vandana Dixit** in partial fulfillment of the requirement for **BE Honors in Artificial Intelligence and Machine Learning** of Savitribai Phule Pune University, Pune in the academic year 2021-2022

Date: 14/05/2022

Place: Pune

**Vandana Dixit Sarita Deshpande**

Guide Head of the Department

This project based seminar report has been examined by us as per the Savirtibai Phule Pune University, Pune, requirements at P.E.S’s Modern College of Engineering on

15 May 2022.

(Name & Signature) (Name & Signature)

Internal Examiner External Examiner

**II**

**ACKNOWLEDGEMENT**

It gives me pleasure in presenting the seminar report on Reinforcement Learning.

Firstly, I would like to express my indebtedness and appreciation to my guide Vandana Dixit. Her constant guidance and advice played a very important role in the successful completion of the report. She always gave me her suggestions, which were crucial in making this report as flawless as possible.

I would like to express our gratitude towards Prof. Dr. Mrs S. Deshpande, Head of Information Technology Engineering Department, PES Modern College of Engineering for her kind co-operation and encouragement which helped me during the completion of this report.

Also, I wish to thank our Principal, Prof. Dr. Mrs. K. R. Joshi and all faculty members for their wholehearted cooperation in the completion of this report. I also thank our laboratory assistants for their valuable help in the laboratory.

Last but not the least, the backbone of my success and confidence lies solely on the blessings of dear parents and lovely friends.

**Akshay Balasaheb Lasunkute**

(Students Name & Signature)

**III**

**ABSTRACT**

The field of reinforcement learning has greatly influenced the neuroscientific study of conditioning. This report provides an introduction to reinforcement learning. Reinforcement learning algorithms have been some of the most influential computational theories in [neuroscience](https://www.sciencedirect.com/topics/neuroscience/neurosciences) for behavioral learning that is dependent on reward and penalty.

RL in the artificial intelligence perspective is a field of machine learning aimed to find computational solutions to a class of problems closely related to the psychological paradigms described in the case of instrumental conditioning (Sutton & Barto, 1998).

This report highlights what is Reinforcement Learning? Its applications and uses and some basic types of RL learnings.

**IV**

**Contents**

Certificate I

Acknowledgement I

Abstract III

Index IV

List of Tables V

List of Figures VI

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Sr.** | | **Chapter** | | **Page No** | | **1.** |  | **Introduction** | |  | |  | 1.1 | Introduction | |  | |  | 1.2 | Motivation | |  | |  | 1.3 | Aim of the topic | |  | |  | 1.4 | Brief Theory | |  | | **2.** |  | **Literature Survey of Reinforcement Learning** | |  | | **3.** |  | **Introduction to Idea** | |  | |  | 3.1 | What is Reinforcement Learning? | |  | |  | 3.2 | Terms used in Reinforcement Learning | |  | |  | 3.3 | Key Features of Reinforcement Learning | |  | |  | 3.4 | Approaches to implement Reinforcement Learning | |  | |  | 3.5 | Types of Reinforcement learning | |  | |  | 3.6 | Reinforcement Learning Applications | |  | |  | 3.7 | How does Reinforcement Learning Work? | |  | |  | 3.8 | The Bellman Equation | |  | |  | 3.9 | Difference between Reinforcement Learning and Supervised Learning | |  | | **4.** |  | **Conclusion** | |  | | **5.** |  | **References** | |  | |  | | | |  | | | |  | | |   **V**  **LIST OF FIGURES**  **Sr. No. Figure Name Page No.**   1. Global RL market overview 10 2. Reinforcement Learning Interaction 13 3. Reinforcement Learning Applications 16 4. Reinforcement Learning Process 17 5. The Bellman Equation 20 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

VI

**LIST OF TABLES**

**Sr. No. Table Name Page No.**

1 Literature Survey 11

2 Terms used in RL 13

3 Difference Between SL and RL 22

**CHAPTER 1**

**INTRODUCTION TO TOPIC**

* 1. **Introduction to Project**

Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.

In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike [supervised learning.](https://www.javatpoint.com/supervised-machine-learning)

Since there is no labeled data, so the agent is bound to learn by its experience only.

RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as game-playing, robotics, etc.

The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.

The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way. Hence, we can say that "Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that." How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.

* 1. **Motivation behind project topic**

The global reinforcement learning market is estimated to grow at a CAGR of ~44% over the forecast period, i.e., 2022 – 2030. The growth of the market can be attributed to the increasing adoption of machine learning (ML) and [artificial intelligence](https://www.researchnester.com/reports/artificial-intelligence-in-automotive-market/1970) (AI) systems, and growing application areas of reinforcement learning. In 2020, more than 10% companies around the world were using ML, whereas more than 40% companies adopted AI. Almost 60% of the enterprises currently have machine learning initiatives in the production stage. So it’s important to have at least basic knowledge of Reinforcement learning. This report provides that basic knowledge required to understand concept of Reinforcement Learning.

Graphical user interface

Description automatically generated with low confidence

* 1. **Aim of the work**

The aim of this report is to attract the attention of people towards the increasing growth of Reinforcement Learning. There are different types of Machine learning approaches are present but supervised Machine Learning trains models based on known answers, Reinforcement Learning, and researchers train the model through an agent, which interacts with the environment.

* 1. **Brief Background**

Reinforcement Learning (RL) is the science of decision making. It is about learning the optimal behavior in an environment to obtain maximum reward. This optimal behavior is learned through interactions with the environment and observations of how it responds, similar to children exploring the world around them and learning the actions that help them achieve a goal.

In the absence of a supervisor, the learner must independently discover the sequence of actions that maximize the reward. This discovery process is akin to a trial-and-error search. The quality of actions is measured by not just the immediate reward they return, but also the delayed reward they might fetch. As it can learn the actions that result in eventual success in an unseen environment without the help of a supervisor, reinforcement learning is a very powerful algorithm.

**CHAPTER 2**

**LITERATURE SURVEY OF RAINFORCEMENT LEARNING**

|  |  |
| --- | --- |
| Recurrent Reinforcement Learning: A Hybrid Approach | [Xiujun Li](https://arxiv.org/search/cs?searchtype=author&query=Li%2C+X), [Lihong Li](https://arxiv.org/search/cs?searchtype=author&query=Li%2C+L), [Jianfeng Gao](https://arxiv.org/search/cs?searchtype=author&query=Gao%2C+J), [Xiaodong He](https://arxiv.org/search/cs?searchtype=author&query=He%2C+X), [Jianshu Chen](https://arxiv.org/search/cs?searchtype=author&query=Chen%2C+J), [Li Deng](https://arxiv.org/search/cs?searchtype=author&query=Deng%2C+L), [Ji He](https://arxiv.org/search/cs?searchtype=author&query=He%2C+J) |
| Benchmarking Deep Reinforcement Learning for Continuous Control | [Yan Duan](https://arxiv.org/search/cs?searchtype=author&query=Duan%2C+Y), [Xi Chen](https://arxiv.org/search/cs?searchtype=author&query=Chen%2C+X), [Rein Houthooft](https://arxiv.org/search/cs?searchtype=author&query=Houthooft%2C+R), [John Schulman](https://arxiv.org/search/cs?searchtype=author&query=Schulman%2C+J), [Pieter Abbeel](https://arxiv.org/search/cs?searchtype=author&query=Abbeel%2C+P) |
| A Concise Introduction to Reinforcement Learning | [Ahmad Hammoudeh](https://www.researchgate.net/profile/Ahmad-Hammoudeh) ([Université de Mons](https://www.researchgate.net/institution/Universite-de-Mons)) |
| A Review Paper on Implementing Reinforcement Learning Technique in Optimising Games Performance | [Mohd Azmin Samsuden](https://ieeexplore.ieee.org/author/37087098299); [Norizan Mat Diah](https://ieeexplore.ieee.org/author/37399982300); [Nurazzah Abdul Rahman](https://ieeexplore.ieee.org/author/37546728300) |
| [Curiosity-driven Exploration in Deep Reinforcement Learning via Bayesian Neural Networks](http://arxiv.org/abs/1605.09674) | R. Houthooft et al., *arXiv*, 2016. |

**CHAPTER 3**

**INTRODUCTION TO IDEA**

**3.1 What is Reinforcement Learning?**

* Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
* In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike [supervised learning.](https://www.javatpoint.com/supervised-machine-learning)
* Since there is no labeled data, so the agent is bound to learn by its experience only.
* RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as game-playing, robotics, etc.
* The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.
* The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way. Hence, we can say that "Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that." How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.
* It is a core part of [Artificial intelligence](https://www.javatpoint.com/artificial-intelligence-tutorial), and all [AI agent](https://www.javatpoint.com/agents-in-ai) works on the concept of reinforcement learning. Here we do not need to pre-program the agent, as it learns from its own experience without any human intervention.
* Example: Suppose there is an AI agent present within a maze environment, and his goal is to find the diamond. The agent interacts with the environment by performing some actions, and based on those actions, the state of the agent gets changed, and it also receives a reward or penalty as feedback.
* The agent continues doing these three things (take action, change state/remain in the same state, and get feedback), and by doing these actions, he learns and explores the environment.
* The agent learns that what actions lead to positive feedback or rewards and what actions lead to negative feedback penalty. As a positive reward, the agent gets a positive point, and as a penalty, it gets a negative point.

Diagram

Description automatically generated

## 3.2 Terms used in Reinforcement Learning

|  |  |
| --- | --- |
| Agent() | An entity that can perceive/explore the environment and act upon it. |
| Environment() | A situation in which an agent is present or surrounded by. In RL, we assume the stochastic environment, which means it is random in nature. |
| Action() | Actions are the moves taken by an agent within the environment. |
| State() | State is a situation returned by the environment after each action taken by the agent. |
| Reward() | A feedback returned to the agent from the environment to evaluate the action of the agent. |
| Policy() | Policy is a strategy applied by the agent for the next action based on the current state. |
| Value() | It is expected long-term retuned with the discount factor and opposite to the short-term reward. |
| Q-value() | It is mostly similar to the value, but it takes one additional parameter as a current action. |

## 3.3 Key Features of Reinforcement Learning

* In RL, the agent is not instructed about the environment and what actions need to be taken.
* It is based on the hit and trial process.
* The agent takes the next action and changes states according to the feedback of the previous action.
* The agent may get a delayed reward.
* The environment is stochastic, and the agent needs to explore it to reach to get the maximum positive rewards.

## 3.4 Approaches to implement Reinforcement Learning

There are mainly three ways to implement reinforcement-learning in ML, which are:

1. Value-based:  
   The value-based approach is about to find the optimal value function, which is the maximum value at a state under any policy. Therefore, the agent expects the long-term return at any state(s) under policy π.
2. Policy-based:  
   Policy-based approach is to find the optimal policy for the maximum future rewards without using the value function. In this approach, the agent tries to apply such a policy that the action performed in each step helps to maximize the future reward.  
   The policy-based approach has mainly two types of policy:
   * Deterministic: The same action is produced by the policy (π) at any state.
   * Stochastic: In this policy, probability determines the produced action.
3. Model-based:

In the model-based approach, a virtual model is created for the environment, and the agent explores that environment to learn it. There is no particular solution or algorithm for this approach because the model representation is different for each environment.

## 3.5 Types of Reinforcement learning

There are mainly two types of reinforcement learning, which are:

* Positive Reinforcement
* Negative Reinforcement

Positive Reinforcement:

The positive reinforcement learning means adding something to increase the tendency that expected behavior would occur again. It impacts positively on the behavior of the agent and increases the strength of the behavior.

This type of reinforcement can sustain the changes for a long time, but too much positive reinforcement may lead to an overload of states that can reduce the consequences.

Negative Reinforcement:

The negative reinforcement learning is opposite to the positive reinforcement as it increases the tendency that the specific behavior will occur again by avoiding the negative condition.

It can be more effective than the positive reinforcement depending on situation and behavior, but it provides reinforcement only to meet minimum behavior.

## 3.6 Reinforcement Learning Applications

Diagram

Description automatically generated

1. Robotics:

RL is used in Robot navigation, Robo-soccer, walking, juggling, etc.

1. Control:

RL can be used for adaptive control such as Factory processes, admission control in telecommunication, and Helicopter pilot is an example of reinforcement learning.

1. Game Playing:

RL can be used in Game playing such as tic-tac-toe, chess, etc.

1. Chemistry:

RL can be used for optimizing the chemical reactions.

1. Business:

RL is now used for business strategy planning.

1. Manufacturing:

In various automobile manufacturing companies, the robots use deep reinforcement learning to pick goods and put them in some containers.

1. Finance Sector:

The RL is currently used in the finance sector for evaluating trading strategies.

## 3.7 How does Reinforcement Learning Work?

To understand the working process of the RL, we need to consider two main things:

* Environment: It can be anything such as a room, maze, football ground, etc.
* Agent: An intelligent agent such as AI robot.

Let's take an example of a maze environment that the agent needs to explore. Consider the below image:

Calendar

Description automatically generated

In the above image, the agent is at the very first block of the maze. The maze is consisting of an S6 block, which is a wall, S8 a fire pit, and S4 a diamond block.

The agent cannot cross the S6 block, as it is a solid wall. If the agent reaches the S4 block, then get the +1 reward; if it reaches the fire pit, then gets -1 reward point. It can take four actions**:** move up, move down, move left, and move right.

The agent can take any path to reach to the final point, but he needs to make it in possible fewer steps. Suppose the agent considers the path S9-S5-S1-S2-S3, so he will get the +1-reward point.

The agent will try to remember the preceding steps that it has taken to reach the final step. To memorize the steps, it assigns 1 value to each previous step. Consider the below step:

Calendar

Description automatically generated

Now, the agent has successfully stored the previous steps assigning the 1 value to each previous block. But what will the agent do if he starts moving from the block, which has 1 value block on both sides? Consider the below diagram:

Calendar

Description automatically generated

It will be a difficult condition for the agent whether he should go up or down as each block has the same value. So, the above approach is not suitable for the agent to reach the destination. Hence to solve the problem, we will use the **Bellman equation**, which is the main concept behind reinforcement learning.

## 3.8 The Bellman Equation

The Bellman equation was introduced by the Mathematician **Richard Ernest Bellman in the year 1953**, and hence it is called as a Bellman equation. It is associated with dynamic programming and used to calculate the values of a decision problem at a certain point by including the values of previous states.

It is a way of calculating the value functions in dynamic programming or environment that leads to modern reinforcement learning.

The key-elements used in Bellman equations are:

* Action performed by the agent is referred to as "a"
* State occurred by performing the action is "s."
* The reward/feedback obtained for each good and bad action is "R."
* A discount factor is Gamma "γ."

The Bellman equation can be written as:

1. V(s) = max [R(s,a) + γV(s`)]

Where,

**V(s)= value calculated at a particular point.**

**R(s,a) = Reward at a particular state s by performing an action.**

**γ = Discount factor**

**V(s`) = The value at the previous state.**

In the above equation, we are taking the max of the complete values because the agent tries to find the optimal solution always.

So now, using the Bellman equation, we will find value at each state of the given environment. We will start from the block, which is next to the target block.

**For 1st block:**

V(s3) = max [R(s,a) + γV(s`)], here V(s')= 0 because there is no further state to move.

V(s3)= max[R(s,a)]=> V(s3)= max[1]=> **V(s3)= 1.**

**For 2nd block:**

V(s2) = max [R(s,a) + γV(s`)], here γ= 0.9(lets), V(s')= 1, and R(s, a)= 0, because there is no reward at this state.

V(s2)= max[0.9(1)]=> V(s)= max[0.9]=> **V(s2) =0.9**

**For 3rd block:**

V(s1) = max [R(s,a) + γV(s`)], here γ= 0.9(lets), V(s')= 0.9, and R(s, a)= 0, because there is no reward at this state also.

V(s1)= max[0.9(0.9)]=> V(s3)= max[0.81]=> **V(s1) =0.81**

**For 4th block:**

V(s5) = max [R(s,a) + γV(s`)], here γ= 0.9(lets), V(s')= 0.81, and R(s, a)= 0, because there is no reward at this state also.

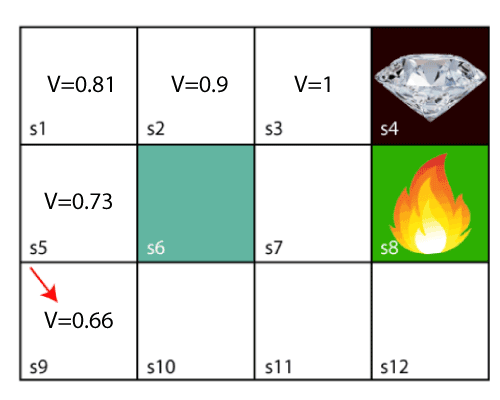
V(s5)= max[0.9(0.81)]=> V(s5)= max[0.81]=> **V(s5) =0.73**

**For 5th block:**

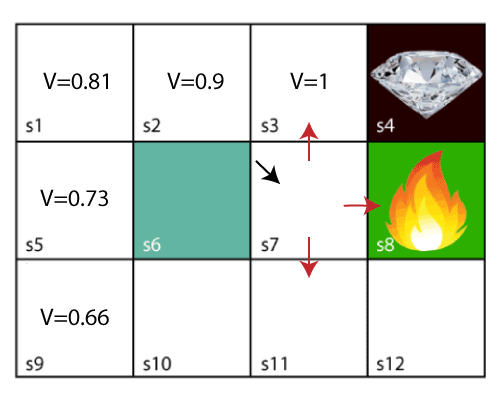
V(s9) = max [R(s,a) + γV(s`)], here γ= 0.9(lets), V(s')= 0.73, and R(s, a)= 0, because there is no reward at this state also.

V(s9)= max[0.9(0.73)]=> V(s4)= max[0.81]=> **V(s4) =0.66**

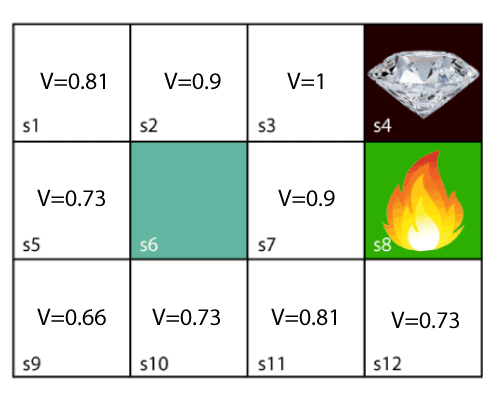
**Consider the below image:**



Now, we will move further to the 6th block, and here agent may change the route because it always tries to find the optimal path. So now, let's consider from the block next to the fire pit.



Now, the agent has three options to move; if he moves to the blue box, then he will feel a bump if he moves to the fire pit, then he will get the -1 reward. But here we are taking only positive rewards, so for this, he will move to upwards only. The complete block values will be calculated using this formula. Consider the below image:



## 

## 3.9 Difference between Reinforcement Learning and Supervised Learning

The Reinforcement Learning and Supervised Learning both are the part of machine learning, but both types of learnings are far opposite to each other. The RL agents interact with the environment, explore it, take action, and get rewarded. Whereas supervised learning algorithms learn from the labeled dataset and, on the basis of the training, predict the output.

The difference table between RL and Supervised learning is given below:

|  |  |
| --- | --- |
| Reinforcement Learning | Supervised Learning |
| RL works by interacting with the environment. | Supervised learning works on the existing dataset. |
| The RL algorithm works like the human brain works when making some decisions. | Supervised Learning works as when a human learns things in the supervision of a guide. |
| There is no labeled dataset is present | The labeled dataset is present. |
| No previous training is provided to the learning agent. | Training is provided to the algorithm so that it can predict the output. |
| RL helps to take decisions sequentially. | In Supervised learning, decisions are made when input is given. |

**CHAPTER 4**

**CONCLUSION**

From the above discussion, we can say that Reinforcement Learning is one of the most interesting and useful parts of Machine learning. In RL, the agent explores the environment by exploring it without any human intervention. It is the main learning algorithm that is used in Artificial Intelligence. But there are some cases where it should not be used, such as if you have enough data to solve the problem, then other ML algorithms can be used more efficiently. The main issue with the RL algorithm is that some of the parameters may affect the speed of the learning, such as delayed feedback.

**CHAPTER 5**

**REFERENCES**

1. Sutton, Richard S. and Barto, Andrew G., [Reinforcement Learning: An Introduction](http://www-anw.cs.umass.edu/~rich/book/the-book.html), MIT Press, 1998

1. [The Reinforcement Learning Repository](http://www-anw.cs.umass.edu/rlr/), University of Massachusetts, Amherst
2. Tesauro, Gerald, [Temporal Difference Learning and TD-Gammon](http://www.research.ibm.com/massive/tdl.html), Communications of the Association for Computing Machinery, March 1995 / Vol 38, No. 3
3. Perez, Andres, [Reinforcement Learning and Autonomous Robots](http://www.geocities.com/fastiland/rlrobots.html) - collection of links to tutorials, books and applications