

A
Seminar Report
on
“Reinforcement Learning and Deep Reinforcement Learning with applications”

Submitted to the
Savitribai Phule Pune University
In partial fulfillment for the award of the Degree of
Bachelor of Engineering
in
Information Technology
by
Aditya Kangune
(33323)

Under the guidance of
Mrs. Usha. A. Jogalekar



Department Of Information Technology
Pune Institute of Computer Technology College of Engineering
Sr. No 27, Pune-Satara Road, Dhankawadi, Pune - 411 043.

A. Y. 2021-2022



CERTIFICATE

This is to certify that the seminar report entitled “**Reinforcement Learning and Deep Reinforcement Learning with applications**” being submitted by **Aditya Kangune (33323)** is a record of bonafide work carried out by him under the supervision and guidance of **Mrs. Usha A. Jogalekar** in partial fulfillment of the requirement for **TE (Information Technology Engineering) – 2019** course of Savitribai Phule Pune University, Pune in the academic year 2021-22.

Date: 14/11/2021

Place: Pune

Mrs. Usha. A. Jogalekar
Guide

Dr.A.M.Bagade
Head of the Department

Dr. R. Sreemathy
Principal

ACKNOWLEDGEMENT

“As our circle of knowledge expands, so does the circumference of darkness surrounding it.”

The seminar has helped me a lot to discover various new things. I extend my gratitude to Pune Institute of Computer Technology for giving me an opportunity to enhance my knowledge through this seminar.

I am extremely grateful to Dr. R. Shreemathy, Principal, PICT, and Dr. Anant Bagade, Head of the Department (Information Technology), for providing all the required resources for the successful completion of my seminar.

My heartfelt gratitude to my seminar guide Mrs. Usha. A. Jogalekar, Department of Information Technology, for her valuable suggestions and guidance in the preparation of the seminar report.

I specially thank my friend Aagaaz for working on this report together with me and constantly being there for helping me with difficulties.

I express my thanks to all the staff members and friends for all the help and coordination extended in bringing out this seminar successfully in time.

I acknowledge with grateful thanks to the authors of the references and other pieces of literature referred to in this seminar.



Aditya Kangune
(33323)

ABSTRACT

Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Deep reinforcement learning is the combination of reinforcement learning (RL) and deep learning. DRL has applications in many fields like medicine, robotics, games, etc. RL works on the Markov Decision Process which leads to Q-learning. MDP provides a mechanism to maximize the reward in a given environment. Combining DL and RL leads to the formation of Deep Q-Networks. RL holds great importance in creating AI for games. Few AIs have managed to defeat world champions in a few games like chess. When we dive deeper into RL, we encounter more complex multi-agent environments. The main challenge in multi-agent environments is to understand the interaction between agents. There are certain techniques to deal with such environments including Graph Convolutional Reinforcement Learning. Particular focus is on exploring these concepts and generalizing DRL and exploring its use in more practical applications like personalized recommendations systems, etc.

Keywords: Reinforcement Learning, Deep Reinforcement Learning, Machine Learning, Artificial Intelligence, MDP, Neural Networks.

CONTENTS

Acknowledgement	I
Abstract	II
List of Tables	IV
List of Figures	IV

Sr.	Chapter	Page No
1.	Introduction	
1.1	History of RL	1
1.2	Motivation	2
1.3	Objective	2
1.4	Introduction to RL and DRL	2
2.	Literature Survey	3
3.	Reinforcement Learning & Deep Reinforcement Learning	
3.1	Introduction	5
3.2	Terminologies	6
3.3	Working of RL	7
3.4	Algorithms	8
3.5	Important traits of RL	8
3.6	Types of RL	9
3.7	Learning models of RL	9
3.8	Markov decision process	10
3.9	Expected return	11
3.10	Policies and value functions	11
3.11	Q-Learning	12
3.12	Deep Q-Learning	12

3.13	Reinforcement Learning vs Supervised Learning	13
4	Applications of RL and DRL	14
5	Advantages and Disadvantages	14
6	Conclusion	15
7	References	16
8	Plagiarism Report	17

LIST OF FIGURES

Sr. No.	Figure Name	Page no.
1	Branches of Automation	1
2	ML Types	1
3	Atari game RL	3
4	RL Terms	5
5	RL Dog example	7
6	Markov decision processes	10

LIST OF TABLES

Sr. No.	Table Name	Page no.
1	Reinforcement Learning vs. Supervised Learning	13

CHAPTER 1

INTRODUCTION

1.1 History of RL:

Artificial Intelligence is a concept that has been discovered and tweaked since around the 1940s. Under AI we have Machine Learning where we work with statistics and algorithms. Finally, we have Deep Learning under ML where we deal with deep concepts like image and speech recognition.

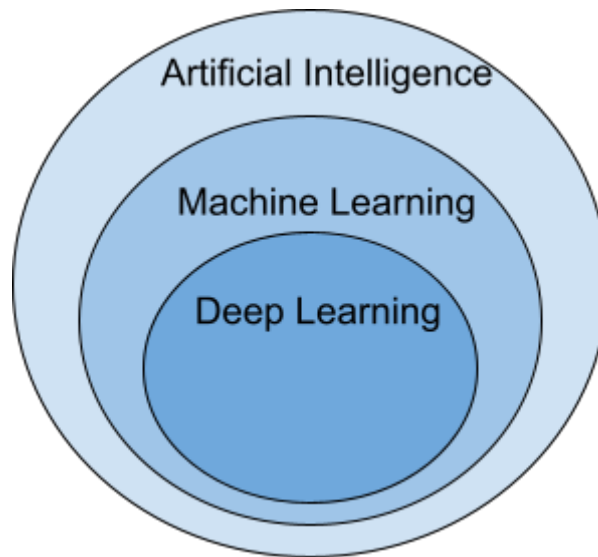


Fig 1- Branches of Automation

There are 3 branches in ML - Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

In RL, the model keeps on developing its performance using Reward and Feedback to learn the behavior and the pattern. These algorithms are particular to a particular problem.

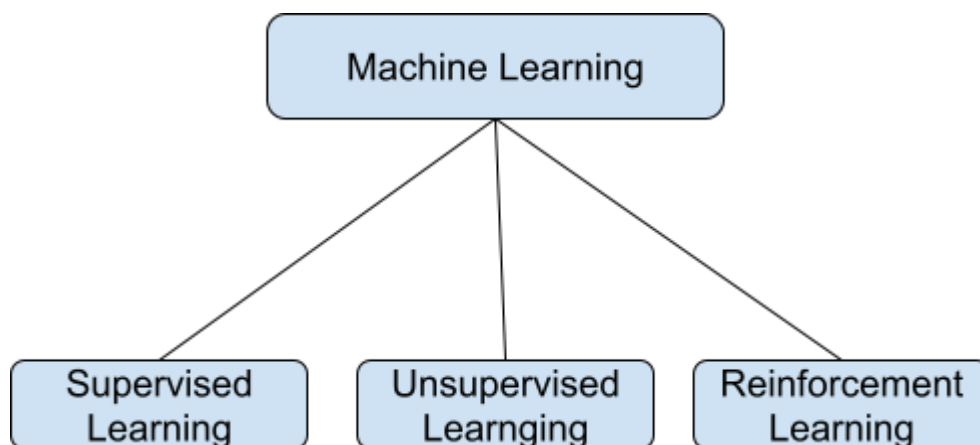


Fig 2 - ML Types

1.2 **Motivation:**

For new age problems in this decade, we need a solution that can create the perfect model to solve that particular problem. That's what the RL aims for.

RL has various interesting applications in the modern world viz medicine, healthcare, self-driving cars, robotics, gaming, marketing, advertising, etc.

1.3 **Objective:**

Aim:

The aim of this seminar is to understand the complex subject of RL and DRL and simplify them with studying about applications of the same.

Objectives:

- Studying basic concepts in Reinforcement Learning.
- Learning how to integrate Reinforcement Learning with Deep Learning or neural networks.
- Exploring the applications of single-agent environments.
- Exploring the techniques involved in multi-agent environments.

1.4 **Introduction to RL and DRL:**

Reinforcement Learning is a subfield of machine learning that teaches linked agents how to choose an action/movement from a set of options in a given environment. This could be one way to maximize rewards over time by mentoring.

Deep Reinforcement Learning is what happens when Deep Learning and Reinforcement Learning are coupled and linked. It can be found in a variety of real-world and funny applications in the United States that go unnoticed, such as games, recommendation engines for well-known e-commerce websites, video-streaming platforms, and so on.

This presentation will mostly cover the implementation of DRL in recommendation systems. From Google Maps recommendations to OTT services like Netflix and Amazon, recommendation algorithms are something we all come across on a daily basis. Self-supervised recommendation systems may be used to increase watch time on video streaming platforms or sales on e-commerce websites. There are other elements to consider, such as the interactions between users and things that form a unit. By focusing on the interaction between the agent, i.e. the user, and the atmosphere (in this case, a website), RL can improve and increase recommendation systems, resulting in an increase in the cumulative and average reward for the agent supporting the interaction with the atmosphere.

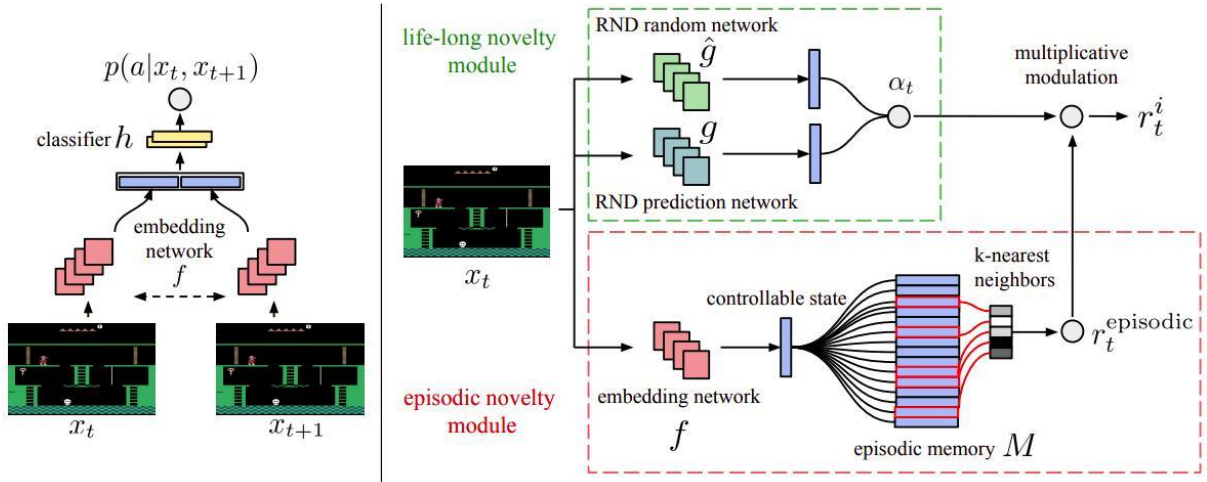
This seminar focuses on the work of reinforcement learning and deep reinforcement learning, as well as their application in two state-of-the-art recommendation frameworks, namely Self-Supervised Q-learning (SQN) and Self-Supervised Actor-Critic Learning (SAC) (SAC).

CHAPTER 2

LITERATURE SURVEY OF RL and DRL

[1]

By understanding directed explorative paths, an RL agent is employed to expose exploration games. To teach the explorative policies, memory-based intrinsic reward is used to maximize the use of knn over fresh experiences. The agent goes over all of the states in the environment once more. The embeddings of the highest neighbor are taught using a self-supervised inverse dynamics model, which biases the new signal towards what the agent can handle. UVFA trains many directed exploration strategies with similar NN in real time. Transfer is no match for major explorative policies leading to effective exploitation since it punishes equivalent NN for n number of degrees of exploration/exploitation. The strategies are implemented using distributed RL agents that collect massive amounts of data from a large number of actors operating in parallel on distinct setting instances. In all, the policy doubles the minimal agent's performance.



[7] Fig 3 - Atari game RL, official - “Never provide Up: Learning Directed Exploration Strategies”

[2]

A CNN is taught to transfer raw pixels from a single front-facing camera onto steering commands. The system learns to drive in traffic on local roads with or without lane markings, as well as on highways, with minimal human intervention. It can be used in areas where visual direction is hazy. Due to the work signal, it automatically learns internal representations of the stated procedure stages such as useful road choices. The model was never explicitly taught to look for things like road definitions. In comparison to particular matter decomposition, this end-to-end approach optimises all procedure steps in real time. Internal pieces may self-optimize to maximise overall system performance, resulting in higher performance. Such criteria are understandably elite for straightforward human interpretation, but they do not ensure the performance of most systems. As a result of the system, smaller networks are possible.

[3]

Virtual assistants are taught with human-annotated lines to learn different language commands. This research describes a toolset for managing novel chemical lines with minimal human intervention. The goal is to translate language into VAPL code using a VAPL and a language programme. A realistic set of input for the brain model is desired by the disembodied spirit. Disembodied spirits use data augmentation, whereas coders use snippets to get information. It is proposed to coach a linguistics programme style principles that build VAPL languages that are adaptable to language translation on the side of the synthesis information. These concepts are being utilised to modify ThingTalk, the language used by the virtual assistant Almond. The disembodied spirit is used to create a primary languages programme that will make compound virtual assistant commands with unquoted free-form parameters easier to execute. On realistic human inputs, it achieves a 62 percent accuracy. Genie's generality is no match by showing a 19 and 31st improvement over the previous state of the art on a music talent, mixture functions, and access management.

[4]

On the internet, where the number of options is vast, there is a need to efficiently transmit relevant material so that the problem of data overload is alleviated. This has resulted in a believable fall for numerous internet users. Recommender systems overcome this annoyance by sifting through a vast amount of constantly generated data. Recommendation systems play a critical role in providing users with personalised content and services. This paper examines the various potentials of the numerous prediction approaches used in recommendation systems. The systems act as a compass for analysis, providing the user with the counsel that he or she desires or anticipates based on data from all other users.

[5]

In recent years, chatbots have advanced rapidly in a variety of industries. Education, Health Care, Cultural Heritage, and Recreation are all examples of the same. The study provides a natural overview of the evolution of human interest in chatbots. The reason why one of chatbots is processed, as well as chatbots' utility in a very specific sector. Furthermore, the influence of societal stereotypes on chatbot style is discussed. The study then carries on to a chatbot classification based on a variety of factors, such as the world of information they communicate with, the requirement they serve, and so on. Furthermore, the overall design of contemporary chatbots is imparted, as well as the most common platforms for their construction.

[6]

The focus of this video series is on reinforcement learning (RL). It is covered how to comprehend the intuition, math, and code involved in RL. An introduction to RL is provided first, followed by a discussion of Markov Decision Processes (MDPs) and Q-learning. Then policy gradients and deep RL deep Q-networks (DQNs) are investigated. There are some great RL projects implemented in code utilising Python, PyTorch, and OpenAI Gym.

CHAPTER 3

Reinforcement Learning & Deep Reinforcement Learning

Reinforcement Learning is a branch of Machine Learning that investigates how software agents should behave in a given environment and assigns appropriate rewards.

Reinforcement Learning is a type of deep learning that allows you to maximise a percentage of the total/average reward. This neural network learning method helps to understand how to optimise a single dimension or how to achieve a complex goal across a number of steps.

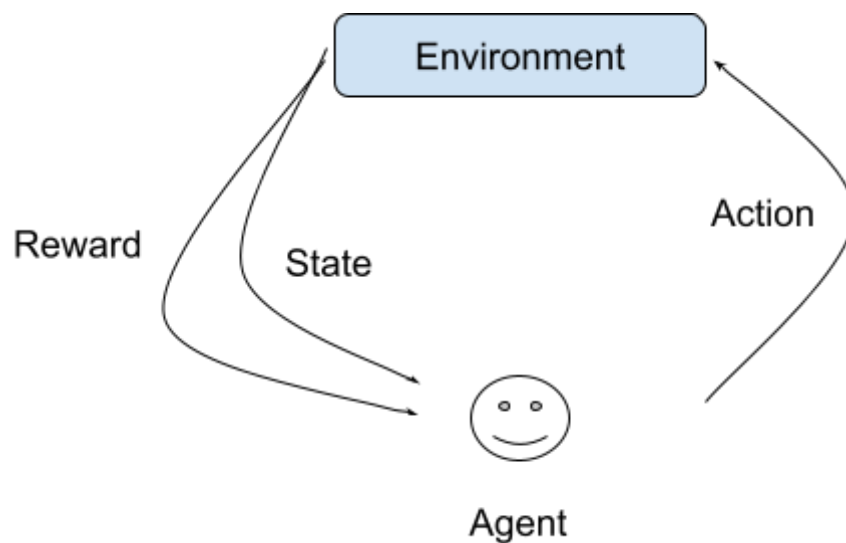


Fig 4 - RL Terms

Consider a self-driving car, in which the car is the agent and the track it must traverse is the environment. Positive benefits include breaking or turning at the proper angle, stopping at a red light, and so on, whereas negative rewards include colliding with another car/person, and so on. In this situation, the agent's clear purpose would be to arrive at the destination with the highest possible payout. The agent has been programmed to make the correct decisions at the appropriate times. We can use Markov decision processes to formalise sequential decision-making.

Terminologies:

- **Agent:** It is an entity which performs actions in an environment in order to gain some reward.
- **State (s):** State is the current situation in the environment by the environment.
- **Policy (π):** It is a strategy which is applied by the agent to decide the next action according to the current state.
- **Environment (e):** A scenario that an agent has to face/surroundings of the assumed agent.
- **Reward (R):** An immediate acknowledgement given to an agent when it performs a particular action or task.
- **Value (V):** In comparison to the short-term reward, it is an expected long-term return with discount.
- **Model of the environment:** This mimics or represents the behavior of the environment. It helps to determine how the environment will behave and make conclusions.
- **Model based methods:** Method for solving RL problems which use model-based methods.
- **Value Function:** Agent which should be expected beginning from that state.
- **Q value or action value (Q):** The only difference between Q value and value is that Q-value takes an additional parameter as a current action.

Working of Reinforcement Learning:

- Consider the situation of training a dog new tricks.
- We can't tell a dog what to do because he doesn't comprehend English or any other human language. Instead, we use a different approach.
- We simulate a situation, and the dog attempts a variety of responses. We will feed the dog if he responds in the desired manner.
- When the dog is exposed to the same situation again, he does a similar activity with even more zeal in the hopes of receiving a larger reward (food).
- That's similar to discovering that positive experiences teach a dog "what to do."
- Simultaneously, the dog learns what not to do when confronted with unpleasant situations.

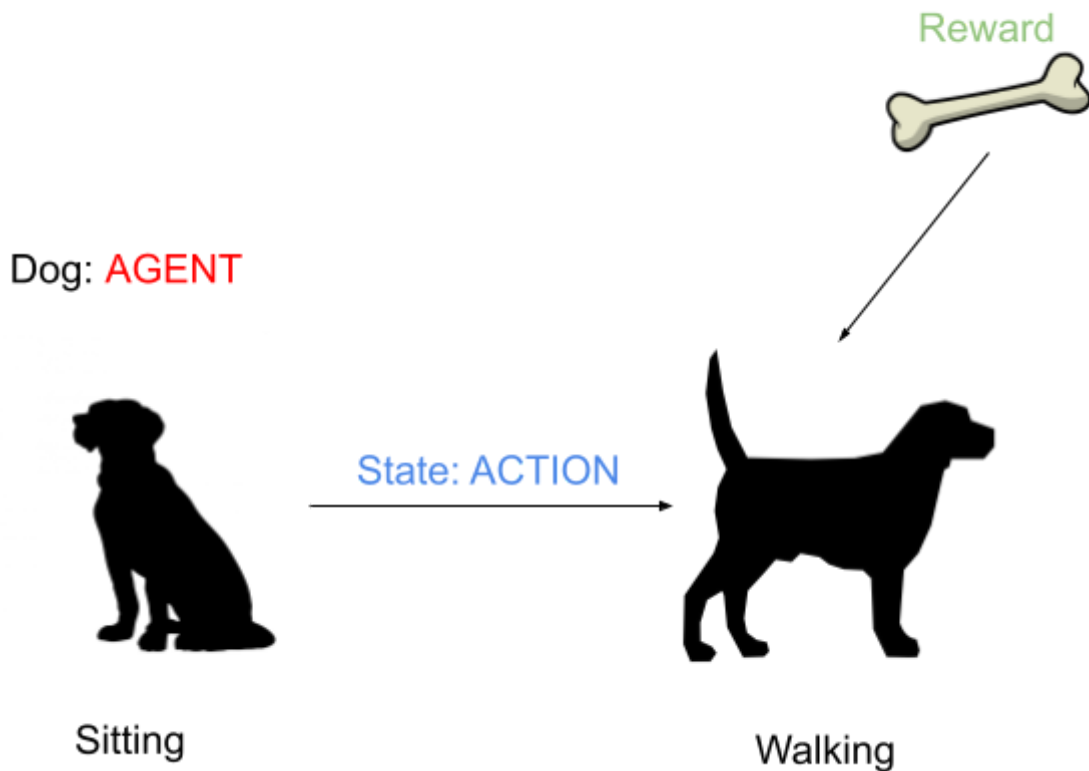


Fig 5 - RL dog example

Reinforcement Learning Algorithms:

There are three approaches to implement a Reinforcement Learning algorithm.

Value-Based:

A value function V should be maximised in a value-based Reinforcement Learning approach (s). In this strategy, the agent anticipates a long-term return of the current policy states.

Policy-based:

In a policy-based RL technique, the goal is to create a policy in which every action taken in each state helps you obtain the most reward in the future.

Two types of policy-based methods are:

- Deterministic: The policy produces the same action for any state.
- Stochastic: Every action has a probability, which may be calculated using the equation below. Stochastic Policy :

$$P(a|s) = P(A_t = a | S_t = s)$$

Model-Based:

You must develop a virtual model for each environment in this Reinforcement Learning approach.

The agent learns how to perform in that particular setting.

Important traits of Reinforcement Learning:

- There is no supervisor, simply a number or a signal of reward.
- Making decisions in a sequential order
- In Reinforcement problems, time is critical.
- Feedback is never immediate; it is always delayed.
- The data that the agent receives is determined by its actions.

Types of Reinforcement Learning

Two kinds of reinforcement learning methods are:

Positive:

It is described as an occurrence that occurs as a result of certain actions. It improves the strength and frequency of the behaviour and has a beneficial impact on the agent's actions.

This type of reinforcement aids in maximising performance and maintaining change for a longer period of time. However, too much reinforcement might lead to state over-optimization, which can have an impact on the outcome.

Negative:

Negative Reinforcement is defined as behaviour strengthening that occurs as a result of a negative condition that should have been avoided or halted. It assists you in establishing a minimal level of performance. The disadvantage of this strategy is that it just gives enough to meet the basic behaviour requirements.

Learning Models of Reinforcement

There are two important learning models in reinforcement learning:

- Markov Decision Process
- Q learning

Markov decision processes

There is an agent in MDP that interacts with the environment in which it is put. The agent successively receives a representation of the current state of the environment through time. The agent picks a specific action based on the state that results in a change in the environmental state, and the agent gets rewarded as a result of the action it performed.

There are 5 major components involved in an MaDP:

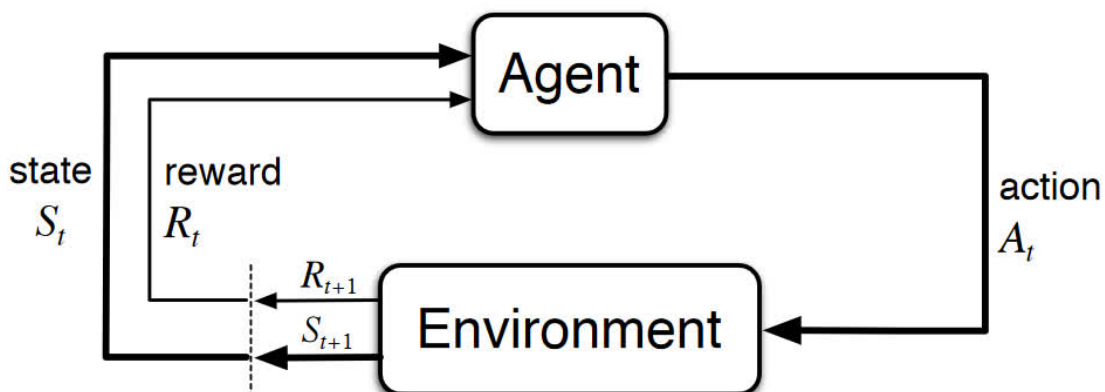
1. Agent
2. Environment
3. State Space (S)
4. Action Space (A)
5. Reward Space (R)

The process of transitioning from one state to another occurs through a series of steps. The agent's purpose throughout this process is to maximise not only the immediate benefit, but also the cumulative rewards it receives over time.

The agent obtains a representation of the current state of the environment S_t at each time step t . The agent chooses an action A_t based on this state. This results in a state-action pair for time t as follows: (S_t, A_t) .

The environment is converted to a new state S_{t+1} , and time is increased to the following time step $t+1$. The agent is given a numerical reward R_{t+1} for the state-action pair at this point (S_t, A_t) .

The following diagram illustrates the above-mentioned algorithm.



Expected Return

The agent's purpose is to maximise total rewards, which are calculated as the sum of future benefits up to time step T. This is known as the expected return, which is denoted by G.

$$G = R_{t+1} + a R_{t+1} + a^2 R_{t+1} + \dots$$

$$G = \sum_{k=0}^{\infty} a^k R_{t+k+1}$$

Where R_t is the return at time t and a is a number ranging from 0 to 1.

Policies and Value Functions

A policy is a function that maps a given state to probabilities of selecting each possible action from that state. It is denoted by π . If an agent follows policy π at time t , then $\pi(a|s)$ is the probability that $A_t = a$ if $S_t = s$. An optimal policy is a policy that gives the highest expected reward among all the policies.

Value functions give a measure of how good a particular action is. This is achieved by using an action-value function which is denoted by q_{π} . The value of action A in state S under policy π is the expected return from starting from state S at time t , taking action A , and following policy π thereafter is given by

$$q_{\pi}(S, A) = E_{\pi} \left[\sum_{k=0}^{\infty} a^k R_{t+k+1} \mid S_t = S, A_t = A \right]$$

An optimal state value function is denoted by q_* and is defined as

$$q_* = \max_{\pi} q_{\pi}(S, A)$$

q_* must always satisfy an equation called as Bellman Optimality Equation which is given as

$$q_*(S, A) = E[R_{t+1} + a * \max_{A'} q_*(S', A')]$$

S' and A' denote the next state and action respectively. This equation ensures that the agent is following an optimal policy and the next state S' will be the state from which best possible action A' can be taken at time $t + 1$.

Q-Learning

Q Learning keeps a lookup table of values $Q(s, a)$ with one entry for every state-action pair [2]. The Q-learning algorithm makes use of the Bellman equation for the Q-value function.

Deep Q-Learning

The use of a deep neural network to estimate the Q-values for each state-action pair in a given environment is called deep Q-Learning and the resultant network is called a deep Q-Network. For each given state input, the network outputs the Q-values for each action that can be taken from that state. The Bellman equation is used for this purpose.

$$q_*(S, A) = E[R_{t+1} + \alpha * \max q_*(S', A')]$$

The loss associated with the neural network is calculated by comparing the output Q-values with the target Q-values. This loss is then used for backpropagation in order to update the weights in the

Reinforcement Learning vs. Supervised Learning:

Parameters	Reinforcement Learning	Supervised Learning
Decision style	Reinforcement learning enables you to make judgments in a sequential manner.	The input given at the start is used to make a choice in this technique.
Works on	focuses on Interacting with the environment is a focus.	Works with examples or data provided as a sample.
Dependency on decision	The decision is dependent on the RL method of learning. As a result, all of the dependent decisions should be labelled.	Supervised learning of judgments that are unrelated to one another, with labels assigned to each decision.
Best suited	Supports and operates better in AI where there is a lot of human involvement.	It is generally controlled by a software system or programmes that are interactive.
Example	Chess game	Object recognition

CHAPTER 4

APPLICATIONS OF RL and DRL

- OTT platforms like Netflix, Amazon prime, etc.
- Social Media (Eg: Facebook suggestions).
- Gaming.
- Google Maps.
- CNN.
- Virtual Assistants.
- Self-driving cars.
- Robotics for industrial automation.
- Business strategy planning.
- Machine learning and data processing.
- It helps you to create training systems that provide custom instruction and materials according to the requirements of students.
- Aircraft control and robot motion control.

CHAPTER 5

Advantages and Disadvantages

Advantages of RL:

- It aids in determining which situations require action.
- Aids you in determining which activity delivers the highest benefit over time.
- Reinforcement Learning also includes a reward function for the learning agent.
- It also enables it to determine the most efficient means of collecting significant rewards.

Disadvantages of RL:

- When enough data is available to solve the problem using a supervised learning method, the reinforcement learning model cannot be used.
- Reinforcement Learning is time-consuming and computationally intensive, especially when the action space is huge.
- Design of the feature/reward, which should be highly involved.
- The pace with which you learn can be influenced by a variety of factors.
- Partially observable environments are possible in realistic settings.
- Too much reinforcement might result in an overabundance of states, lowering the quality of the output.
- Non-stationary situations are possible in realistic settings.

CHAPTER 6

CONCLUSION

- Reinforcement Learning is a form of Machine Learning.
- Aids you in determining which activity delivers the highest benefit over time.
- Three methods for reinforcement learning are 1) Value-based 2) Policy-based and Model based learning.
- Value function, Agent, State, Reward, Environment The terms "model of the environment" and "model-based approaches" are used frequently in the RL learning method.
- Your cat is an agent that is exposed to the environment, which is an example of reinforcement learning.
- The most notable feature of this system is that there is no supervisor involved; instead, a genuine number or incentive signal is used.
- Two types of reinforcement learning are 1) Positive 2) Negative.
- Two widely used learning models are 1) Markov Decision Process 2) Q learning.
- The supervised learning approach works on given sample data or example, but the reinforcement learning method works on interacting with the environment.
- Robotics for industrial automation and business strategy planning are two examples of applications or reinforcement learning methodologies.
- When there is enough data to answer the problem, this method should not be employed.
- The most significant disadvantage of this strategy is that factors can influence learning speed.

CHAPTER 7

REFERENCES







- [1]- Adrià Puigdomènech Badia, “**Never Give Up: Learning Directed Exploration Strategies**” arXiv:2002.06038 [cs.LG], 14 Feb 2020.
- [2] Urs Muller, “**End to End Learning for Self-Driving Cars**” arXiv:1604.07316 [cs.CV], Apr 2016.
- [3] Giovanni Campagna, “**Genie: A Generator of Natural Language Semantic Parsers for Virtual Assistant Commands**” arXiv:1904.09020 [cs.CL], Apr 2019.
- [4] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, “**Recommendation systems: Principles, methods and evaluation**”, Egyptian Informatics Journal, Volume 16, Issue 3, 2015, Pages 261-273.
- [5] Adamopoulou E., Moussiades L. (2020) “**An Overview of Chatbot Technology**”. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31
- [6] Chris and Mandy. “Reinforcement Learning - Goal Oriented Intelligence”. *deeplearning.ai*, Sep 2018
- [7] Fig 3 - Atari game RL, official - “Never provide Up: Learning Directed Exploration Strategies”, <https://vitalab.github.io/article/2020/05/28/NGU.html>
- [8] Fig 6 - Markov decision processes, <https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da>

CHAPTER 8
PLAGIARISM REPORT

Document Information

Analyzed document	33323 Seminar Report(2).pdf (D121244501)
Submitted	2021-12-06T10:59:00.0000000
Submitted by	USHA
Submitter email	uajogalekar@pict.edu
Similarity	19%
Analysis address	uajogalekar.pict@analysis.orkund.com

Sources included in the report

W	URL: https://onlinecoursesschools.com/what-is-reinforcement-learning Fetched: 2021-12-06T11:02:00.0000000		1
W	URL: https://www.guru99.com/reinforcement-learning-tutorial.html Fetched: 2021-12-06T11:02:00.0000000		10
W	URL: https://writex.today/artificial-intelligence/reinforcement-learning-2/ Fetched: 2021-12-06T11:02:00.0000000		4
W	URL: https://www.worldofitech.com/reinforcement-learning/ Fetched: 2021-12-06T11:02:00.0000000		1
W	URL: https://nischalmadiraju.medium.com/reinforcement-learning-333abddae456 Fetched: 2021-12-06T11:02:00.0000000		2
W	URL: https://cementanswers.com/what-is-reinforcement-learning-in-ai/ Fetched: 2021-12-06T11:02:00.0000000		1

Entire Document

A Seminar Report on "Reinforcement Learning and Deep Reinforcement Learning with applications" Submitted to the Savitribai Phule Pune University In partial fulfillment for the award of the Degree of Bachelor of Engineering in Information Technology by Aditya Kangune (33323) Under the guidance of Mrs. Usha. A. Jogalekar Department Of Information Technology Pune Institute of Computer Technology College of Engineering Sr. No 27, Pune-Satara Road, Dhankawadi, Pune - 411 043. A. Y. 2021-2022 P: F-SMR-UG/08/R0

CERTIFICATE This is to certify that the seminar report entitled "Reinforcement Learning and Deep Reinforcement Learning with applications" being submitted by Aditya Kangune (33323) is a record of bonafide work carried out by him under the supervision and guidance of Mrs. Usha A. Jogalekar in partial fulfillment of the requirement for TE (Information Technology Engineering) – 2019 course of Savitribai Phule Pune University, Pune in the academic year 2021-22. Date: 14/11/2021 Place: Pune Mrs. Usha. A. Jogalekar Dr.A.M.Bagade Guide Head of the Department Dr. R. Sreemathy Principal II

ACKNOWLEDGEMENT "As our circle of knowledge expands, so does the circumference of darkness surrounding it." The seminar has helped me a lot to discover various new things. I extend my gratitude to Pune Institute of Computer Technology for giving me an opportunity to enhance my knowledge through this seminar. I am extremely grateful to Dr. R. Shreemathy, Principal, PICT, and Dr. Anant Bagade, Head of the Department (Information Technology), for providing all the required resources for the successful completion of my seminar. My heartfelt gratitude to my seminar guide Mrs. Usha. A. Jogalekar, Department of Information Technology, for her valuable suggestions and guidance in the preparation of the seminar report. I specially thank my friend Aagaaz for working on this report together with me and constantly being there for helping me with difficulties. I express my thanks to all the staff members and friends for all the help and coordination extended in bringing out this seminar successfully in time. I acknowledge with grateful thanks to the authors of the references and other pieces of literature referred to in this seminar. Aditya Kangune (33323)

ABSTRACT

95%

MATCHING BLOCK 1/19

W

[https://onlinecoursesschools.com/what-is-reinf ...](https://onlinecoursesschools.com/what-is-reinf...)

Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Deep reinforcement learning is

the combination of reinforcement learning (RL) and deep learning.

DRL has applications in many fields like medicine, robotics, games, etc. RL works on the Markov Decision Process which leads to Q-learning. MDP provides a mechanism to maximize the reward in a given environment. Combining DL and RL leads to the formation of Deep Q-Networks. RL holds great importance in creating AI for games. Few AIs have managed to defeat world champions in a few games like chess. When we dive deeper into RL, we encounter more complex multi-agent environments. The main challenge in multi-agent environments is to understand the interaction between agents. There are certain techniques to deal with such environments including Graph Convolutional Reinforcement Learning. Particular focus is on exploring these concepts and generalizing DRL and exploring its use in more practical applications like personalized recommendations systems, etc. Keywords: Reinforcement Learning, Deep Reinforcement Learning, Machine Learning, Artificial Intelligence, MDP, Neural Networks. II

CONTENTS Acknowledgement I Abstract II List of Tables IV List of Figures IV Sr. Chapter Page No 1. Introduction 1.1 History of RL 1 1.2 Motivation 2 1.3 Objective 2 1.4 Introduction to RL and DRL 2 2. Literature Survey 3 3. Reinforcement Learning & Deep Reinforcement Learning 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 3.10 3.11 3.12 Introduction 6 Terminologies 7 Working of RL 8 Algorithms 9 Important traits of RL 9 Types of RL 10 Learning models of RL 10 Markov decision process 11 Expected return 12 Policies and value functions 12 Q-Learning 13 Deep Q-Learning 13 3.13 Reinforcement Learning vs Supervised Learning 14 4 Applications of RL and DRL 15 5 Advantages and Disadvantages 15 6 Conclusion 16 7 References 17 8 Plagiarism Report 17 II

LIST OF FIGURES Sr. No. Figure Name Page no. 1 Branches of Automation 1 2 ML Types 1 3 Atari game RL 3 4 RL Terms 6 5 RL Dog example 8 6 Markov decision processes 11 **LIST OF TABLES** Sr. No. Table Name Page no. 1 Reinforcement Learning vs. Supervised Learning 12 IV

CHAPTER 1 INTRODUCTION 1.1 History of RL: Artificial Intelligence is a concept that has been discovered and tweaked since around the 1940s. Under AI we have Machine Learning where we work with statistics and algorithms. Finally, we have Deep Learning under ML where we deal with deep concepts like image and speech recognition. Fig 1- Branches of Automation There are 3 branches in ML - Supervised Learning, Unsupervised Learning, and Reinforcement Learning. In RL, the model keeps on developing its performance using Reward and Feedback to learn the behavior and the pattern. These algorithms are particular to a particular problem. Fig 2 - ML Types 8

1.2 Motivation: For new age problems in this decade, we need a solution that can create the perfect model to solve that particular problem. That's what the RL aims for. RL has various interesting applications in the modern world viz medicine, healthcare, self-driving cars, robotics, gaming, marketing, advertising, etc. 1.3 Objective: Aim: The aim of this seminar is to understand the complex subject of RL and DRL and simplify them with studying about applications of the same.

Objectives: • Studying basic concepts in Reinforcement Learning. • Learning how to integrate Reinforcement Learning with Deep Learning or neural networks. • Exploring the applications of single-agent environments. • Exploring the techniques involved in multi-agent environments. 1.4 Introduction to RL and DRL: Reinforcement Learning is a subfield of machine learning that teaches linked agents how to choose an action/movement from a set of options in a given environment. This could be one way to maximize rewards over time by mentoring. Deep Reinforcement Learning is what happens when Deep Learning and Reinforcement Learning are coupled and linked. It can be found in a variety of real-world and funny applications in the United States that go unnoticed, such as games, recommendation engines for well-known e-commerce websites, video-streaming platforms, and so on. This presentation will mostly cover the implementation of DRL in recommendation systems. From Google Maps recommendations to OTT services like Netflix and Amazon, recommendation algorithms are something we all come across on a daily basis. Self-supervised recommendation systems may be used to increase watch time on video streaming platforms or sales on e-commerce websites. There are other elements to consider, such as the interactions between users and things that form a unit. By focusing on the interaction between the agent, i.e. the user, and the atmosphere (in this case, a website), RL can improve and increase recommendation systems, resulting in an increase in the cumulative and average reward for the agent supporting the interaction with the atmosphere. This seminar focuses on the work of reinforcement learning and deep reinforcement learning, as well as their application in two state-of-the-art recommendation frameworks, namely Self-Supervised Q-learning (SQN) and Self-Supervised Actor-Critic Learning (SAC) (SAC). 9

The seminar's goals include gaining a solid understanding of MDP and applying that knowledge to the two recommendation frameworks described before. 10

CHAPTER 2 LITERATURE SURVEY OF RL and DRL [1] By understanding directed explorative paths, an RL agent is employed to expose exploration games. To teach the explorative policies, memory-based intrinsic reward is used to maximize the use of knn over fresh experiences. The agent goes over all of the states in the environment once more. The embeddings of the highest neighbor are taught using a self-supervised inverse dynamics model, which biases the new signal towards what the agent can handle. UVFA trains many directed exploration strategies with similar NN in real time. Transfer is no match for major explorative policies leading to effective exploitation since it punishes equivalent NN for n number of degrees of exploration/exploitation. The strategies are implemented using distributed RL agents that collect massive amounts of data from a large number of actors operating in parallel on distinct setting instances. In all, the policy doubles the minimal agent's performance. [7] Fig 3 - Atari game RL, official - "Never provide Up: Learning Directed Exploration Strategies" [2] A CNN is taught to transfer raw pixels from a single front-facing camera onto steering commands. The system learns to drive in traffic on local roads with or without lane markings, as well as on highways, with minimal human intervention. It can be used in areas where visual direction is hazy. Due to the work signal, it automatically learns internal representations of the stated procedure stages such as useful road choices. The model was never explicitly taught to look for things like road definitions. In comparison to particular matter decomposition, this end-to-end approach optimises all procedure steps in real time. Internal pieces may self-optimize to maximise overall system performance, resulting in higher performance. Such criteria are understandably elite for straightforward human interpretation, but they do not ensure the performance of most systems. As a result of the system, smaller networks are possible. 11

[3] Virtual assistants are taught with human-annotated lines to learn different language commands. This research describes a toolset for managing novel chemical lines with minimal human intervention. The goal is to translate language into VAPL code using a VAPL and a language programme. A realistic set of input for the brain model is desired by the disembodied spirit. Disembodied spirits use data augmentation, whereas coders use snippets to get information. It is proposed to coach a linguistics programme style principles that build VAPL languages that are adaptable to language translation on the side of the synthesis information. These concepts are being utilised to modify ThingTalk, the language used by the virtual assistant Almond. The disembodied spirit is used to create a primary languages programme that will

make compound virtual assistant commands with unquoted free-form parameters easier to execute. On realistic human inputs, it achieves a 62 percent accuracy. Genie's generality is no match by showing a 19 and 31st improvement over the previous state of the art on a music talent, mixture functions, and access management. [4] On the internet, where the number of options is vast, there is a need to efficiently transmit relevant material so that the problem of data overload is alleviated. This has resulted in a believable fall for numerous internet users. Recommender systems overcome this annoyance by sifting through a vast amount of constantly generated data. Recommendation systems play a critical role in providing users with personalised content and services. This paper examines the various potentials of the numerous prediction approaches used in recommendation systems. The systems act as a compass for analysis, providing the user with the counsel that he or she desires or anticipates based on data from all other users. [5] In recent years, chatbots have advanced rapidly in a variety of industries. Education, Health Care, Cultural Heritage, and Recreation are all examples of the same. The study provides a natural overview of the evolution of human interest in chatbots. The reason why one of chatbots is processed, as well as chatbots' utility in a very specific sector. Furthermore, the influence of societal stereotypes on chatbot style is discussed. The study then carries on to a chatbot classification based on a variety of factors, such as the world of information they communicate with, the requirement they serve, and so on. Furthermore, the overall design of contemporary chatbots is imparted, as well as the most common platforms for their construction. [6] The focus of this video series is on reinforcement learning (RL). It is covered how to comprehend the intuition, math, and code involved in RL. An introduction to RL is provided first, followed by a discussion of Markov Decision Processes (MDPs) and Q-learning. Then policy gradients and deep RL deep Q-networks (DQNs) are investigated. There are some great RL projects implemented in code utilising Python, PyTorch, and OpenAI Gym. 12

CHAPTER 3

Reinforcement Learning & Deep

42%

MATCHING BLOCK 2/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

Reinforcement Learning Reinforcement Learning is a branch of Machine Learning that investigates how software agents should behave in a given environment and assigns appropriate rewards. Reinforcement Learning is a type of deep learning that allows you to maximise a percentage of the total/average reward. This neural network learning method helps to understand how to optimise a

single dimension or how to achieve a complex goal across a number of steps. Fig 4 - RL Terms Consider a self-driving car, in which the car is the agent and the track it must traverse is the environment. Positive benefits include breaking or turning at the proper angle, stopping at a red light, and so on, whereas negative rewards include colliding with another car/person, and so on. In this situation, the agent's clear purpose would be to arrive at the destination with the highest possible payout. The agent has been programmed to make the correct decisions at the appropriate times. We can use Markov decision processes to formalise sequential decision-making. 13

Terminologies: •

79%

MATCHING BLOCK 3/19

W

[https://writex.today/artificial-intelligence/r ...](https://writex.today/artificial-intelligence/r...)

Agent: It is an entity which performs actions in an environment in order to gain some reward. •

64%

MATCHING BLOCK 4/19

W

[https://writex.today/artificial-intelligence/r ...](https://writex.today/artificial-intelligence/r...)

State (s): State is the current situation in the environment by the environment. • Policy (π): It is a strategy which is applied by the agent to decide the next action according to the current state. •

35%

MATCHING BLOCK 5/19

W

[https://writex.today/artificial-intelligence/r ...](https://writex.today/artificial-intelligence/r...)

Environment (e): A scenario that an agent has to face/surroundings of the assumed agent. • Reward (R): An immediate acknowledgement given to an agent when it performs a particular action or task. •

Value (V): In comparison to the short-term

35%

MATCHING BLOCK 6/19

W

[https://writex.today/artificial-intelligence/r ...](https://writex.today/artificial-intelligence/r...)

reward, it is an expected long-term return with discount. • Model of the environment: This mimics or represents the behavior of the environment. It helps to determine how the environment will behave and make conclusions. • Model based methods: Method for solving RL problems which use model-based methods. • Value Function: Agent which should be expected beginning from that state. • Q value or action value (Q): The only difference between Q value and value is that Q-value takes an additional parameter as a current action. 14 Working of Reinforcement Learning: •

Consider the situation of training a dog new tricks. • We can't tell a dog what to do because he

50%

MATCHING BLOCK 7/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

doesn't comprehend English or any other human language. Instead, we use a different approach. • We simulate a situation, and the

dog attempts a variety of responses. We will feed the dog if he responds in the desired manner. • When

52%

MATCHING BLOCK 8/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

the dog is exposed to the same situation again, he does a similar activity with even more zeal in

the hopes of receiving a larger reward (food). • That's similar to discovering that positive experiences teach a dog "what to do." • Simultaneously, the dog learns what not to do when confronted with unpleasant situations. Fig 5 - RL dog example 15

97%

MATCHING BLOCK 9/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

Reinforcement Learning Algorithms: There are three approaches to implement a Reinforcement Learning algorithm. Value-Based: A value

function V should be maximised

62%

MATCHING BLOCK 12/19

W

[https://www.worldofitech.com/reinforcement-lea ...](https://www.worldofitech.com/reinforcement-lea...)

in a value-based Reinforcement Learning approach (s). In this strategy, the agent anticipates a long-term return of the current policy states. Policy-based: In a policy-based RL technique,

the goal is to create a policy in which every action taken in each state helps you obtain the most

50%

MATCHING BLOCK 10/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

reward in the future. Two types of policy-based methods are: • Deterministic: The policy produces the same action for any state. • Stochastic: Every action has a probability, which may be calculated using the equation below. Stochastic Policy : $n(a|s) = P(A = a|S, =S)$ Model-Based: You must develop a virtual model for each environment

in this Reinforcement Learning approach.

50%

MATCHING BLOCK 11/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

The agent learns how to perform in that particular setting. Important traits of Reinforcement Learning: • There is no supervisor, simply a number or a signal

of reward. • Making decisions in a sequential order • In Reinforcement problems, time is critical. • Feedback is never immediate; it is always delayed. • The data that the agent receives is determined by

29%

MATCHING BLOCK 13/19

W

[https://nischalmadiraju.medium.com/reinforceme ...](https://nischalmadiraju.medium.com/reinforceme...)

its actions. 16 Types of Reinforcement Learning Two kinds of reinforcement learning methods are: Positive: It is described as an occurrence that occurs as a result of certain actions. It improves the strength and frequency of the behaviour and has a beneficial impact on the agent's actions. This type of reinforcement aids in maximising performance and maintaining change for a

longer period of time.

40%

MATCHING BLOCK 17/19

W

[https://nischalmadiraju.medium.com/reinforceme ...](https://nischalmadiraju.medium.com/reinforceme...)

However, too much reinforcement might lead to state over-optimization, which can have an impact on the outcome. Negative: Negative Reinforcement is defined as behaviour strengthening that occurs as a result of a negative condition that should have been avoided or halted. It assists you in establishing a minimal level of performance. The disadvantage of this strategy is that it just gives enough to meet the basic behaviour requirements. Learning Models of Reinforcement There are two important learning models in reinforcement learning: • Markov Decision Process •

Q learning 17

Markov decision

processes There is an agent in MDP that interacts with the environment in which it is put. The agent successively receives a representation of the current state of the environment through time. The agent picks a specific action based on the state that results in a change in the environmental state, and the agent gets rewarded as a result of the action it performed. There are 5 major components involved in an MaDP: 1. Agent 2. Environment 3. State Space (S) 4. Action Space (A) 5. Reward Space (R) The process of transitioning from one state to another occurs through a series of steps. The agent's purpose throughout this process is to maximise not only the immediate benefit, but also the cumulative rewards it receives over time. The agent obtains a representation of the current state of the environment S_t at each time step t . The agent chooses an action A_t based on this state. This results in a state-action pair for time t as follows: (S_t, A_t) . The environment is converted to a new state S_{t+1} , and time is increased to the following time step $t+1$. The agent is given a numerical reward R_{t+1} for the state-action pair at this point (S_t, A_t) . The following diagram illustrates the above-mentioned algorithm. 18

[8] Fig 6 - Markov decision processes Expected Return The agent's purpose is to maximise total rewards, which are calculated as the sum of future benefits up to time step T . This is known as the expected return, which is denoted by $G_t = R_t + \gamma V_t$. Where R_t is the return at time t and γ is a number ranging from 0 to 1. Policies and Value Functions A policy is a function that maps a given state to probabilities of selecting each possible action from that state. It is denoted by π . If an agent follows policy π at time t , then $\pi(a|s)$ is the probability that a if $s = s$. An optimal policy is a policy that gives the highest expected reward among all the policies. Value functions give a measure of how good a particular action is. This is achieved by using an action-value function which is denoted by Q . The value of action A in state S under policy π is the π expected return from starting from state S at time t , taking action A , and following policy thereafter is given by $Q^\pi(s, a) = \mathbb{E}_\pi [\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a]$. An optimal state value function is denoted by V^* and is defined as $V^*(s) = \max_a Q^*(s, a)$. It must always satisfy an equation called as Bellman Optimality Equation which is given as $V^*(s) = \max_a [R(s, a) + \gamma V^*(s')]$ 19

and denote the next state and action respectively. This equation ensures that the agent is following an optimal policy and the next state will be the state from which best possible action a' can be taken at time $t + 1$. ' Q-Learning Q Learning keeps a lookup table of values $Q(s, a)$ with one entry for every state-action pair [2]. The Q-learning algorithm makes use of the Bellman equation for the Q-value function. Deep Q-Learning The use of a deep neural network to estimate the Q-values for each state-action pair in a given environment is called deep Q-Learning and the resultant network is called a deep Q-Network. For each given state input, the network outputs the Q-values for each action that can be taken from that state. The Bellman equation is used for this purpose. $Q^*(s, a) = [R(s, a) + \gamma \max_{a'} Q^*(s', a')]$ The loss associated with the neural network is calculated by comparing the output Q-values with the target Q-values. This loss is then used for backpropagation in order to update the weights in the 20

62%

MATCHING BLOCK 14/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

Reinforcement Learning vs. Supervised Learning: Parameters Reinforcement Learning Supervised Learning Decision style Reinforcement learning enables you to make judgments in a sequential manner. The input given at the

start is used to make a choice in this technique. Works on focuses on Interacting with the environment is a focus. Works with examples or data provided as a sample. Dependency on decision The decision is dependent on the RL method of learning. As a result, all of the dependent decisions should be labelled. Supervised learning of judgments that are unrelated to one another, with labels assigned to each decision. Best suited Supports and operates better in AI where there is a lot of human involvement. It is generally controlled by a software system or programmes that are interactive. Example Chess game Object recognition 21

CHAPTER 4 APPLICATIONS OF RL and DRL • OTT platforms like Netflix, Amazon prime, etc. • Social Media (Eg: Facebook suggestions). • Gaming. • Google Maps. • CNN. • Virtual Assistants. • Self-driving cars. •

100%

MATCHING BLOCK 15/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

Robotics for industrial automation. • Business strategy planning. • Machine learning and data processing. • It helps you to create training systems that provide custom instruction and materials according to the requirements of students. • Aircraft control and robot motion control.

CHAPTER 5 Advantages and Disadvantages Advantages of RL: • It aids in determining which situations require action. • Aids you in determining which activity delivers the highest benefit over time. • Reinforcement Learning also includes a reward function for the learning agent. • It also enables it to determine the most efficient means of collecting significant rewards. Disadvantages of RL: • When enough data is available to solve the problem using a supervised learning method, the reinforcement learning model cannot be used. • Reinforcement Learning is time-consuming and computationally intensive, especially when the action space is huge. • Design of the feature/reward, which should be highly involved. • The pace with which you learn can be influenced by a variety of factors. • Partially observable environments are possible in realistic settings. 22

• Too much reinforcement might result in an overabundance of states, lowering the quality of the output. • Non-stationary situations are possible in realistic settings. CHAPTER 6 CONCLUSION •

48%

MATCHING BLOCK 16/19

W

[https://www.guru99.com/reinforcement-learning- ...](https://www.guru99.com/reinforcement-learning-...)

Reinforcement Learning is a form of Machine Learning. • Aids you in determining which activity delivers the highest benefit over time. • Three methods for reinforcement learning are 1) Value-based 2) Policy-based and Model based learning. •

Value function, Agent, State, Reward, Environment The terms "model of the environment" and "model-based approaches" are used frequently in the RL

41%

MATCHING BLOCK 19/19

W

[https://cementanswers.com/what-is-reinforcemen ...](https://cementanswers.com/what-is-reinforcemen...)

learning method. • Your cat is an agent that is exposed to the environment, which is an example of reinforcement learning. • The most notable feature of this system is that there is no supervisor involved; instead, a genuine number or incentive signal

is used. •

Two types of reinforcement learning are 1) Positive 2) Negative. • Two widely used learning models are 1) Markov Decision Process 2) Q learning. • The supervised learning approach works on given sample data or example, but the reinforcement learning method works on interacting with the environment. • Robotics for industrial automation and business strategy planning

are two examples of applications or reinforcement learning methodologies. • When there is enough data to answer the problem, this method should not be employed. • The most significant disadvantage of this strategy is that factors can influence learning speed. 23

CHAPTER 7 REFERENCES [1]- Adrià Puigdomènech Badia, "Never Give Up: Learning Directed Exploration Strategies" arXiv:2002.06038 [cs.LG], 14 Feb 2020. [2] Urs Muller, "End to End Learning for Self-Driving Cars" arXiv:1604.07316 [cs.CV], Apr 2016. [3] Giovanni Campagna, "Genie: A Generator of Natural Language Semantic Parsers for Virtual Assistant Commands" arXiv:1904.09020 [cs.CL], Apr 2019. [4] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation", Egyptian Informatics Journal, Volume 16, Issue 3, 2015, Pages 261-273. [5] Adamopoulou E., Moussiades L. (2020) "An Overview of Chatbot Technology". In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31 [6] Chris and Mandy. "Reinforcement Learning - Goal Oriented Intelligence". deeplearning.ai, Sep 2018 [7] Fig 3 - Atari game RL, official - "Never provide Up: Learning Directed Exploration Strategies", <https://vitalab.github.io/article/2020/05/28/NGU.html> [8] Fig 6 - Markov decision processes, <https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da> 24

25

CHAPTER 8 PLAGIARISM REPORT IV

Hit and source - focused comparison, Side by Side

Submitted text As student entered the text in the submitted document.
Matching text As the text appears in the source.

1/19	SUBMITTED TEXT	33 WORDS	95% MATCHING TEXT	33 WORDS
	Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Deep reinforcement learning is		Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Reinforcement learning is	
	W https://onlinecoursesschools.com/what-is-reinforcement-learning			

2/19	SUBMITTED TEXT	52 WORDS	42% MATCHING TEXT	52 WORDS
	Reinforcement Learning Reinforcement Learning is a branch of Machine Learning that investigates how software agents should behave in a given environment and assigns appropriate rewards. Reinforcement Learning is a type of deep learning that allows you to maximise a percentage of the total/average reward. This neural network learning method helps to understand how to optimise a		Reinforcement Learning?Reinforcement Learning is defined as a Machine Learning method that is with how software agents should take actions in an environment. Reinforcement Learning is a part of the deep learning method that helps you to maximize some portion of the cumulative reward. This neural network learning method helps you to learn how to attain a	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

3/19	SUBMITTED TEXT	17 WORDS	79% MATCHING TEXT	17 WORDS
	Agent: It is an entity which performs actions in an environment in order to gain some reward. •		Agent: It is an assumed entity which performs actions in an environment to gain some •	
	W https://writex.today/artificial-intelligence/reinforcement-learning-2/			

4/19	SUBMITTED TEXT	35 WORDS	64% MATCHING TEXT	35 WORDS
	State (s): State is the current situation in the environment by the environment. • Policy (π): It is a strategy which is applied by the agent to decide the next action according to the current state. •		State (s): State refers to the current situation returned by the environment. • Policy (π): It is a strategy which applies by the agent to decide the next action based on the current state. •	
	W https://writex.today/artificial-intelligence/reinforcement-learning-2/			

5/19	SUBMITTED TEXT	31 WORDS	35% MATCHING TEXT	31 WORDS
	Environment (e): A scenario that an agent has to face/surroundings of the assumed agent. • Reward (R): An immediate acknowledgement given to an agent when it performs a particular action or task. •		Environment (e): A scenario that an agent has to face. • Reward (R): An immediate return given to an agent when he or she performs specific action or task. •	
	W https://writex.today/artificial-intelligence/reinforcement-learning-2/			

6/19	SUBMITTED TEXT	84 WORDS	35% MATCHING TEXT	84 WORDS
	reward, it is an expected long-term return with discount. • Model of the environment: This mimics or represents the behavior of the environment. It helps to determine how the environment will behave and make conclusions. • Model based methods: Method for solving RL problems which use model-based methods. • Value Function: Agent which should be expected beginning from that state. • Q value or action value (Q): The only difference between Q value and value is that Q-value takes an additional parameter as a current action. 14 Working of Reinforcement Learning: •		reward. It is an agent which should be expected beginning from that state. • Model of the environment: This mimics the behavior of the environment. It helps you to make inferences to be made and also determine how the environment will behave. • Model based methods: It is a method for solving reinforcement learning problems which use model-based methods. • Q value or action value (Q): Q value is quite similar to value. The only difference between the two is that it takes an additional parameter as a current action How Reinforcement Learning	
	W https://writex.today/artificial-intelligence/reinforcement-learning-2/			

7/19	SUBMITTED TEXT	20 WORDS	50% MATCHING TEXT	20 WORDS
	doesn't comprehend English or any other human language. Instead, we use a different approach. • We simulate a situation, and the		doesn't understand English or any other human language, we can't tell her directly what to do. Instead, we follow a different strategy. • We emulate a situation, and the	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

8/19	SUBMITTED TEXT	19 WORDS	52% MATCHING TEXT	19 WORDS
	the dog is exposed to the same situation again, he does a similar activity with even more zeal in		the cat is exposed to the same situation, the cat executes a similar action with even more enthusiastically in	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

9/19	SUBMITTED TEXT	14 WORDS	97% MATCHING TEXT	14 WORDS
	Reinforcement Learning Algorithms: There are three approaches to implement a Reinforcement Learning algorithm. Value-Based: A value		Reinforcement Learning Algorithms There are three approaches to implement a Reinforcement Learning algorithm. Value-Based: In a value-	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

12/19	SUBMITTED TEXT	25 WORDS	62% MATCHING TEXT	25 WORDS
	in a value-based Reinforcement Learning approach (s). In this strategy, the agent anticipates a long-term return of the current policy states. Policy-based: In a policy-based RL technique,		In a value-based Reinforcement Learning technique, you should attempt to maximize a value function s). In this strategy, the agent is expecting a long-term return of the present states under π . Policy-based: In a policy-based RL technique,	
	W https://www.worldofitech.com/reinforcement-learning/			

10/19	SUBMITTED TEXT	49 WORDS	50% MATCHING TEXT	49 WORDS
	reward in the future. Two types of policy-based methods are: • Deterministic: The policy produces the same action for any state. • Stochastic: Every action has a probability, which may be calculated using the equation below. Stochastic Policy : $n(a s) = P(A = a S = S)$ Model-Based: You must develop a virtual model for each environment		reward in the future. Two types of policy-based methods are: • Deterministic: For any state, the same action is produced by the policy π . • Stochastic: Every action has a certain probability, which is determined by the following equation. Stochastic Policy : $n(a s) = P(A = a S = S)$ Model-Based: In this Reinforcement Learning method, you need to create a virtual model for each environment.	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

11/19	SUBMITTED TEXT	25 WORDS	50% MATCHING TEXT	25 WORDS
	The agent learns how to perform in that particular setting. Important traits of Reinforcement Learning: • There is no supervisor, simply a number or a signal		The agent learns to perform in that specific environment. Characteristics of Reinforcement Learning Here are important characteristics of reinforcement learning • There is no supervisor, only a real number or reward signal •	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

13/19	SUBMITTED TEXT	55 WORDS	29% MATCHING TEXT	55 WORDS
	its actions. 16 Types of Reinforcement Learning Two kinds of reinforcement learning methods are: Positive: It is described as an occurrence that occurs as a result of certain actions. It improves the strength and frequency of the behaviour and has a beneficial impact on the agent's actions. This type of reinforcement aids in maximising performance and maintaining change for a		its experience. Reinforcements: There are two kinds of reinforcement learning methods: Positive: It is as an event, that occurs because of specific behaviour. It increases the strength and the frequency of the behaviour and impacts positively on the action taken by the agent. This type of Reinforcement helps you to maximize performance and sustain change for a	
	<div>W https://nischalmadiraju.medium.com/reinforcement-learning-333abddae456</div>			

17/19	SUBMITTED TEXT	77 WORDS	40% MATCHING TEXT	77 WORDS
	However, too much reinforcement might lead to state over-optimization, which can have an impact on the outcome. Negative: Negative Reinforcement is defined as behaviour strengthening that occurs as a result of a negative condition that should have been avoided or halted. It assists you in establishing a minimal level of performance. The disadvantage of this strategy is that it just gives enough to meet the basic behaviour requirements. Learning Models of Reinforcement There are two important learning models in reinforcement learning: • Markov Decision Process •		However, too much Reinforcement may lead to over-optimization of state, which can affect the results. Negative: Negative Reinforcement is defined as strengthening of behaviour that occurs because of a negative condition which should have stopped or avoided. It helps you to define the minimum of performance. However, the drawback of this method is that it provides enough to meet up the minimum behaviour. Types of Learning Models of Reinforcement There are two important learning models in reinforcement learning: Markov Decision Process:	
	<div>W https://nischalmadiraju.medium.com/reinforcement-learning-333abddae456</div>			

14/19	SUBMITTED TEXT	24 WORDS	62% MATCHING TEXT	24 WORDS
	Reinforcement Learning vs. Supervised Learning: Parameters Reinforcement Learning Supervised Learning Decision style Reinforcement learning enables you to make judgments in a sequential manner. The input given at the		Reinforcement Learning vs. Supervised Learning Parameters Reinforcement Learning Supervised Learning Decision style reinforcement learning helps you to take your decisions sequentially. In this method, a decision is made on the input given at the	
	<div>W https://www.guru99.com/reinforcement-learning-tutorial.html</div>			

15/19	SUBMITTED TEXT	36 WORDS	100% MATCHING TEXT	36 WORDS
	Robotics for industrial automation. • Business strategy planning. • Machine learning and data processing. • It helps you to create training systems that provide custom instruction and materials according to the requirements of students. • Aircraft control and robot motion control.		Robotics for industrial automation. • Business strategy planning • Machine learning and data processing • It helps you to create training systems that provide custom instruction and materials according to the requirement of students. • Aircraft control and robot motion control	
	<div>W https://www.guru99.com/reinforcement-learning-tutorial.html</div>			

16/19	SUBMITTED TEXT	34 WORDS	48% MATCHING TEXT	34 WORDS
	Reinforcement Learning is a form of Machine Learning. • Aids you in determining which activity delivers the highest benefit over time. • Three methods for reinforcement learning are 1) Value-based 2) Policy-based and Model based learning. •		Reinforcement Learning is a Machine Learning method • Helps you to discover which action yields the highest reward over the longer period. • Three methods for reinforcement learning are 1) Value-based 2) Policy-based and Model based learning. •	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			

19/19	SUBMITTED TEXT	40 WORDS	41% MATCHING TEXT	40 WORDS
	learning method. • Your cat is an agent that is exposed to the environment, which is an example of reinforcement learning. • The most notable feature of this system is that there is no supervisor involved; instead, a genuine number or incentive signal		learning your cat is an agent that is exposed to the environment. The biggest characteristic of this method is that there is no supervisor, only a real number or reward signal.	
	W https://cementanswers.com/what-is-reinforcement-learning-in-ai/			

18/19	SUBMITTED TEXT	53 WORDS	56% MATCHING TEXT	53 WORDS
	Two types of reinforcement learning are 1) Positive 2) Negative. • Two widely used learning models are 1) Markov Decision Process 2) Q learning. • The supervised learning approach works on given sample data or example, but the reinforcement learning method works on interacting with the environment. • Robotics for industrial automation and business strategy planning		Two types of reinforcement learning are 1) Positive 2) Negative • Two widely used learning model are 1) Markov Decision Process 2) Q learning the environment, whereas the supervised learning method works on given sample data or example. • Application or reinforcement learning methods are: Robotics for industrial automation and business strategy planning •	
	W https://www.guru99.com/reinforcement-learning-tutorial.html			