VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018

A TECHNICAL SEMINAR REPORT ON

"REINFORCEMENT LEARNING"

Submitted in partial fulfillment for the award of the degree in

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE & ENGINEERING

Submitted by:

SHIVA THAPA

USN: 1BH18CS114

Under the guidance of

Mr. Praveen S R
(Assistant Professor)

Bangalore Technological Institute

Department of Computer Science & Engineering

Off-Sarjapura Road, Kodathi,

Bangalore - 560035.

BANGALORE TECHNOLOGICAL INSTITUTE

(An ISO 9001:2015 Certified Institute)

(Affiliated to Visvesvaraya Technological University, Belagavi)

Department of Computer Science & Engineering

CERTIFICATE

It is certified that the Technical Seminar work entitled **"REINFORCEMENT LEARNING"** carried out by **SHIVA THAPA (1BH18CS114)** was bonafide student of **BANGALORE TECHNOLOGICAL INSTITUTE**, **Bangalore** in partial fulfillment for the award of **Bachelor of Engineering in Computer Science and Engineering** of the **Visvesvaraya Technological University**, **Belagavi** during the year 2022. Thus, it is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report submitted to the Department of Computer Science and Engineering. The technical seminar report has been approved as it satisfies the academic requirements in respect of technical seminar prescribed for the said degree.

Mr. Rajsheker A,	**Dr. Sohan Kumar Gupta,**
Technical Seminar Co-ordinator	H.O.D,
Department of CSE	Department of CSE

.

External Viva

Name of the examiners	Signature with date

BANGALORE TECHNOLOGICAL INSTITUTE

(An ISO 9001:2015 Certified Institute)

(Affiliated to Visvesvaraya Technological University, Belagavi)

Department of Computer Science & Engineering

DECLARATION

I am the student of eighth semester **B.E. COMPUTER SCIENCE AND ENGINEERING, BANGALORE TECHNOLOGICAL INSTITUTE, BENGALURU**, hereby declare that the technical seminar entitled **"REINFORCEMENT LEARNING"** has been independently carried out by me at Bangalore Technological Institute, Bengaluru and submitted in partial fulfilment of the requirements for the award of the degree in **Bachelor of Engineering in Computer Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the academic year 2022.

I also declare that, to the best of my knowledge and believe the work reported here does not form or part of any other dissertation on the basis of which a degree or award was conferred on an early occasion of this by any other student.

PLACE: BENGALURU

DATE:

Acknowledgement

Any achievement, be it scholastic or otherwise does not depend solely on the individual efforts but on the guidance, encouragement and cooperation of intellectuals and elders. I would like to take this opportunity to thank them all.

First, I would like to thank my guide **Mr. Praveen S R**, for giving me the opportunity to express myself.

I also would like to thank all the people that worked along with me with their patience and openness they created an enjoyable working environment. It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I hereby extend my gratitude to coordinator, **Mr. Rajsheker A** for his valuable advice, encouragement and suggestion given to me in the course of presentation preparation. I convey my gratitude to her for having constantly monitored the development of the work and setting up precise deadlines.

I would like to express my gratitude to Head of Department, **Dr. Sohan Kumar Gupta**, for his unfailing encouragement and suggestions given to me in the course of my presentation preparation.

I would take this opportunity to express my gratitude to the Principal, **Dr. H S Nanda**, for giving me this opportunity to enrich my knowledge.

I am grateful to the President **Dr. Prabhakara Reddy** and Secretary **Sri. C L Gowda** for having provided me with a great well-environment.

Finally, a note of thanks to the Department of Computer Science and Engineering, both teaching and non-teaching staff for their co-operation extended to us.

Last but not the least, I acknowledge the support and feedback of our parents, guardians and friends, for their indispensable help at all times.

SHIVA THAPA USN: 1BH18CS114

ABSTRACT

Reinforcement learning is a learning paradigm concerned with learning to control a system so as to maximize a numerical performance measure that expresses a long-term objective. What distinguishes reinforcement learning from supervised learning is that only partial feedback is given to the learner about the learner's predictions. Further, the predictions may have long term effects through influencing the future state of the controlled system. Thus, time plays a special role. The goal in reinforcement learning is to develop efficient learning algorithms, as well as to understand the algorithms' merits and limitations. Reinforcement learning is of great interest because of the large number of practical applications that it can be used to address, ranging from problems in artificial intelligence to operations research or control engineering.

Table of Contents

CHAPTER 1

INTRODUCTION

Reinforcement learning (RL) refers to both a learning problem and a subfield of machine learning. As a learning problem, it refers to learning to control a system so as to maximize some numerical value which represents a long-term objective. A typical setting where reinforcement learning operates is shown in Figure 1: A controller receives the controlled system's state and a reward associated with the last state transition. It then calculates an action which is sent back to the system. In response, the system makes a transition to a new state and the cycle is repeated. The problem is to learn a way of controlling the system so as to maximize the total reward. The learning problems differ in the details of how the data is collected and how performance is measured.

Reinforcement Learning is an auxiliary learning technique an agent looks at the space of possible procedures and gets an input on the results of the alternatives made. These alternatives are found out through experimental communications with a dynamic situation. It is likewise characterized by contrasting the problem with the different controls of study in machine learning. There are two principal approaches for solving reinforcement learning problems. The first methodology is to look at the space of practices in order to discover one that performs well in the environment. This methodology has been taken to work in the genetic calculations and programming as well as some more novel search strategies. The second methodology is to utilize the factual strategies and dynamic programming techniques to compute the utility of taking activities in conditions of the world.

1.1 Basic Model of RL

A RL model consists of

a) A discrete set of environment states S

b) A discrete set of agent action A

c) A set of scalar reinforcement signals {0, 1} or the real numbers.

A reinforcement learning agent is autonomous which means that its behavior is determined by its own experience. Learning is the mechanism through which an agent

can increase its intelligence while performing operations. What is outside the agent is considered the environment. The states are parameters or features that describe the environment. An RL agent senses the environment and learns the optimal policy or near optimal policy by taking actions in each state of the environment. The agent must be aware of the states while interacting with the environment. An agent learns from reinforcement feedback received from its environment known as either reward or punishment signal. RL agents try to maximize the reward or minimize the punishment. Actions could affect the next state of the environment and subsequent rewards and have the ability to optimize the environment's state. Continuous learning and adapting through interaction with environment help the agent to learn online in terms of performing the required task and improving its behavior in real time.

Figure 1.1 shows the basic model of the reinforcement signal. Stepwise execution of basic model of RL

1. Intelligent agent receives as input 'i', some indication of current state (s1) of the environment
2. The intelligent agent then chooses an action from the set of actions (A) to generate as output
3. (a) The action changes the state of environment (s2)

   (b) The value of this state transition is communicated to the agent through a scalar reinforcement signal (r)

One of the significant issues is how to improve an autonomous agent's capability and how to enhance the agent's intelligence. The ability of an agent can increase its intelligence through learning while in operation. Knowledge gained by an RL agent is specific to the environment in which it operates and cannot be easily transferred to another agent, even if the environments are very similar. The general knowledge obtained from one agent could benefit the other, but no efficient methods exist for transfer of that knowledge.

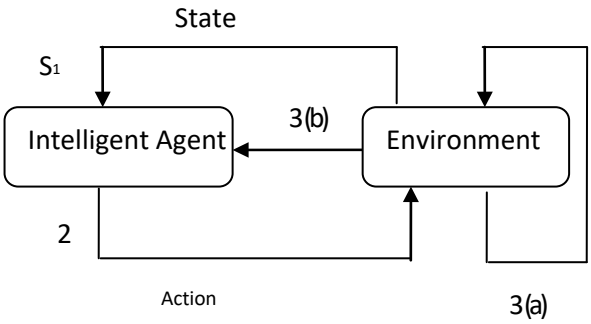

Figure 1.1: A basic model of RL

An RL agent's design is based on the characteristics of the problem to be solved. The problem must be clearly defined and analyzed so that the purpose of designing the agent can be determined. The RL agent is a decision-maker in the process which takes an action that influences the environment. The agent acquires knowledge of different actions and eventually learns to perform actions which are most rewarding in order to attain a certain goal. Another key element of RL is an action policy which defines the agent's behavior at any given time and it maps the states to the actions to be taken [6].

1.2 Challenges in Reinforcement Learning

The system developed using RL or self-learning requires being intelligent enough to take decisions according to the environment changes. The two most commonly and critical applications of self-learning (RL) are: robot soccer and mars rover. These require accuracy, agility actions, and flexible strategy. The conditions are dynamic, uncertain and complex and have to take decisions as humans. These systems require general control subsystem, visual subsystem, decision subsystems, communication subsystem and the robot vehicle subsystems. The hardware required is robot vehicle system, the image acquisition device, host computer(s), wireless launchers. These multi-agent system (MAS) uses robotics, artificial intelligence, intelligent control, computer technology, sensor technology. The systems are critical as they have to train themselves for unknown environment about which less or no information is available. Mars rover had objective of investigating mars' habitability, studying its climate and geology and collecting data for a manned mission to mars. It's difficult to train the system so that the system developed can fulfil the targeted objective. These system developments require the in-depth study of reinforcement learning or self-learning.

CHAPTER 2

LITERATURE REVIEW

As highlighted by Musumeci et al, 2018 Machine Learning is one of the most suitable ways to handle the approach that performs the analysis of network data and fault management. As indicated by Lewis et al, 2008 that in the field of artificial intelligence reinforcement learning can take care of issues ideally by cooperating with its condition and furthermore by changing its control arrangements. Busoniu et al, 2009 clarifies that reinforcement learning is utilized to locate the best arrangement to expand the reward. Primarily it got from the instrument of regular learning and furthermore dependent on incentives and penalties on the environment. Flore, 2015 stated that Reinforcement learning is the problem that is faced by an agent and that has to be learned by trial-and-error interfaces through the dynamic environment. According to Sutton, 1992 Reinforcement learning always depend upon delayed result and experimentation. Tiwana et al., 2014 highlighted that reinforcement learning based framework for Quality of Services (QoS) used for the enhancement in Fourth generation (4G) Networks. Hou et al, 2017 proposed a technique that is powerful for taking care of choice issues which are consecutively enhanced. Reinforcement learning is identified with programming that depends on Markov decision process. Olafati, 2006 states that reinforcement learning algorithms are used to represents the unavoidably and social procedure. It uses state activity learning parameters that builds the components of the factors exponentially. Vidhate et al, 2016 stated that reinforcement learning is a methodology that is utilized for the improvement of multi-agent learning and furthermore a structure that have new strategies which demonstrates that simulated results and also acquire more results. Carlucho et al, 2017 proposed that on steady Q-learning procedure for portable robots a versatile PID control. This procedure doesn't require and earlier knowledge but it can understand the procedure that is differ from the conventional procedure. Hung et al, 2017 also recommended that by applying Q-learning algorithms to a small flocking fixed-wing UAVs to figure out how to run and it is also simulated that it can fly in the non-stationary stochastic environment.

CHAPTER 3

METHODOLOGY

3.1 Agent

Agent is machine that learns from the close interaction with the environment that it senses the state in which it is in and

takes the action and which causes the environment to change.

There are basically three components of Agent in Reinforcement Learning:

i. Policy

ii. Value Function

iii. Model

3.1.1 Policy

Say, we are looking for a short way and you have 2 options at a particular time, left or right. Since, you don't have any idea which way to do, you will assign a random probability to each direction. This is known as Stochastic Policy as for every decision you have the probability given the current state. Over experience your agent will learn and will go for the option that gives higher probability of success.

$\Pi(a|s) = P[A=a \mid S=s]$

So, it is defined as the probability of taking an action (a) on a particular current state (s). „π" refers to the policy chosen at the time of decision.

On the other hand, say if you are taking the decision by doing proper procedure, you would be knowing that which direction to choose at any point of time. So, it is basically a function that tracks the state to actions. This is known as Deterministic Policy.

3.1.2 Value Function

Value Function defines how good it is to be in a particular state.

Considering the previous example. This time you have a rough idea about the two ways i.e., you now know the obstacles that will come in each way and the estimate time

required in each path. So, you to calculate which path will for practical for you getting both heavy rewards and obstacles. This is known as Value Function.

3.1.3 Model

Model is the agent's representation of the environment.

We again consider the example of two roads (left and right). You are standing of the crossing. You can see a little bit of the path ahead in each road. For example, the road to the left does not have proper lighting and also, the roads are not quite good. While the other road has proper lighting and the roads are also concrete road. From that point of view, you will predict the road ahead and accordingly we will decide which path to go for. So, for our agent it will predict the dynamic environment with the data he has and accordingly it will choose the path. It is known as Transition model.

3.2 Environment

Environment is nothing but the surrounding in which the agent moves or perform its task i.e., it is the physical world in which the agent operates. Here the agent performs its task and gain reward. The environment is the place where we can find out whether our agent is getting positive reward of negative reward. The environment consists of number of obstacles designed for the agent to perform its task. Environment plays an important role in reinforcement learning. Without environment the reinforcement learning is incomplete.

3.3 Reward

The feedback sends by the environment to determine the last action. In reinforcement learning the rewards are received on the basis of the agent's behavior in the environment and the state that the agent is in. If the agent is in the state in which he should not be then it will receive a negative reward but if the agent is the state in which it should be then it will receive a positive reward. The amount of positive or negative reward a agent receives according to it the accuracy of the agent is decided. The reward is decided on the basis of the value of gamma. While the agent is in testing process gamma's value is given 0, if the agent is in perfect state the value of gamma given is 1. Even if the value we are not sure about the agent's action the gamma is kept between 0 and 1. The positive reward indicates that the agent is learning properly and the task given to agents is being done properly. If the agent has more negative reward, then the positive reward then the

agent is not working properly. So, to make it work properly we have to make some changes in the agent's program.

3.4 State

State is something that the machine it is in. For example, a car is halted at the signal. Currently the signal is red so the car will stop. Stop is the current state of the car which it is in and when the signal turns green the state is changed and the car should move now. The change of state should be done which will result in some action taken by the car.

3.5 Action

Action is taken when state changes. Considering the above example, when the signal turns red the state is stop and car stops. The action taken by the car is to stop and when the signal turns green the car should move and the action taken is to move. When the car moves it will receive a positive reward +1 and if doesn’t it will receive a negative reward -1. This is how an agent learns from the action.

CHAPTER 4

STRENGTHS AND LIMITATIONS

4.1 Strengths

a) Reinforcement learning is Adaptable, unlike supervised learning algorithms, reinforcement learning doesn't require retraining because it adapts to new environments automatically on the fly.

b) It's Innovative, unlike reinforcement learning supervised learning is actually imitating whoever provided the data for that algorithm.

c) Bias Resistance, if there is bias in the way the data is labeled then a supervised learning algorithm will pick up that bias and learn inherited bias.

d) Goal-oriented, Reinforcement learning can be used for sequences of actions while supervised learning is mostly used in an input-output manner.

e) Reinforcement learning doesn't require large labeled datasets.

f) Reinforcement learning methods have been used to train neural networks, to control dynamic channel assignment in communications networks, and to construct fuzzy logic rule bases for fuzzy control systems.

g) Online Learning, reinforcement learning runs in real-time reinforcement learning combines exploration when the machine tests new approaches on the fly to find better solutions and exploitation when the machine exploits the best solutions which it has found thus far.

4.2 Limitations

a) Reinforcement learning needs tons of knowledge and tons of computation, it's data-hungry, that's why it works rather well in video games because one can play the sport again and again and again, so getting many data seems feasible.

b) The tradeoffs between exploration and exploitation have to be found. RL agents try to reach a goal as quickly as possible (exploitation); they also seek to learn more information about their environment in order to enhance the future performance (exploration).

c) The exploration/exploitation dilemma is analogous to the tradeoff between system control and system identification in the field of optimal control.

d) You cannot get very specific about the definition of the info sorting and therefore the output this often because the info utilized in unsupervised in unsupervised learning is labelled and not known, it's employment of the machine to label and group the data before determining the hidden patterns.

e) Too much reinforcement learning can cause an overload of states which may diminish the results.

f) Less accuracy of the results, this is often also because the input file isn't known and not labelled by people beforehand, which suggests that the machine will got to do that alone. The results of the analysis can't be ascertained. There's no prior knowledge within the unsupervised method of machine learning.

g) Knowledge transfer from one agent to another is another difficulty when considering RL systems. This is due to the fact that RL is a global learning method that contains all of the information learned about the environment in a single value function.

CHAPTER 5

STEPS FOR RL PROBLEM

These steps are applied to Reinforcement Learning problem before it came into existence.

5.1 Understanding your Problem

Reinforcement learning isn't essentially required every problem. There should be check for each and every problem before applying reinforcement learning algorithm the following attributes are taken into consideration,

a) trial-and-error an experimental technique

b) delayed rewards

c) can be displayed as MDP

d) to check that the problem is a controlled problem or not.

5.2 A Simulated Environment

Before applying the Reinforcement Learning algorithms, it is necessary that numbers of iterations are to be calculated. To represent the real-world objects effectively a simulated program is required.

5.3 Markov Decision Process (MDP)

For each and every problem the formulation of the problem has to be done by following these steps. Firstly, problem is to be formulated into a MDP and then design the state space, action space, reward function etc. An agent will do what is to be compensated under the requirements.

5.4 Algorithms

There are various Reinforcement Learning algorithms that are accessible and used to discover the best policy or to become familiar with the value function.

CHAPTER 6

APPLICATIONS

Coming up next there are some application zones in which Reinforcement learning is utilized as activities in a domain to amplify some thought of combined reward.

6.1 Traffic Forecasting Service

There is rapid growth in the quantity of vehicles running on the roads, management of traffic appears to a gigantic issue. To conquer this issue machines can be prepared and used to tackle this issue. Machines that overlay gauge about future traffic conditions on an advanced traffic stream map. These frameworks can likewise be utilized to know the present and future traffic states of an area and furthermore furnish clients with steering alternatives dependent on that data.

6.2 Robotics

Robots can perform amazing errands under human control, including medical procedures and family tasks. In this field it is impossible to foresee that there is a state that is totally discernible or not. This learning framework won't be able to predict the information related to different states that may look similar. Multi-Robot Systems can frequently be utilized to satisfy the tasks that are hard to be cultivated by the single robot, particularly within the sight of vulnerabilities, incomplete data, conveyed control and non-concurrent calculations.

6.3 Computer Games

The trade related to gaming has been developed enormously in the ongoing years. Artificial Intelligence agents are reused to make intelligent gaming knowledge for the players. These agents can assume various jobs like player's adversaries, teammates or other non-player characters. Aside from communicating with the human players a game requires to fulfill a large group of different necessities like the sound and special visualizations.

6.4 Machinery Applications

Reinforcement learning is a kind of Machine Learning calculation that permits programming agents and machines to expressly decide the ideal conduct inside a

particular setting and to expand its performance. These applications can't be modified. It comprises of Manufacturing, Inventory Management, Delivery Management, Power Systems and Finance Sector.

6.5 Stock Market Analysis

The stock exchange market and its patterns continuously changing day by day to make benefits and for the endurance in the stock market legitimate comprehension is essential. Machine learning has been fundamentally utilized for forecast of monetary markets. Suitable algorithms like support vector machine and reinforcement learning have been viable in following the stock market and augmenting the benefit of investment opportunity at low risk. It likewise consolidates the analysis of market that considers the options of the general financial investors who invests globally in stock market to foresee the day-to-day stock pattern.

6.6 Semantic Annotation of Learning Environments

In the today's world of functional learning is picking up significance in each and every aspect of life. It helps in procuring down the practical knowledge as well as gives the better comprehension of the item. The utilization of semantic calculations as a major aspect of an abilities that are based on learning environment that is extremely helpful. Simulations of genuine circumstances helps in the advancement of viable aptitudes like decision making, communication, team working and problem solving.

CHAPTER 7

CONCLUSION

In today's world people are constantly looking for an agreeable life that is the major reason that we have always relied on machines to accomplish our work more effortlessly in a faster and well-organized way. In the past years, machines have been utilized to reduce the physical work, but in the present era with the advancement of Artificial Intelligence humans seek to make machines that are not only strong but also intelligent therefore the concept of machine learning has come into existence and that become an area of study that is gaining momentum day by day. In this paper categories of machine learning i.e., supervised learning, unsupervised learning, recommender system and reinforcement learning are discussed that also displays the various applications under machine learning. Reinforcement learning is one of the types of machine learning that expounds its application and research to the wider region of control and various decision problems that are not usually handled by supervised or unsupervised learning techniques. To deal with such type of situations Reinforcement Learning has come into existence and become one of the most intelligent agents due to its various characteristics like online learning, self-improving and very less programming effort. Reinforcement learning provides a technically and mathematically solid solution for optimal decision making in many challenging tasks having multi-dimensional, noisy data, complex nonlinear dynamics, sequential decision procedures with delayed rewards. This work intends to present a comprehensive report of RL applications to many important areas including healthcare, robotics, communications and networking, natural language processing, internet of things, computer vision, games and scheduling.

REFERENCES

[1] Dr. Amanpreet Kaur and Kumar Gourav, "A Study of Reinforcement Learning Applications and its Algorithm" (2020)

[2] Rashmi Sharma, Manish Prateek and Ashok K. Sinha, "Use of Reinforcement Learning as a Challenge: A Review" (2013)

[3] Mahesh Joshi, Lav Sharma, Arvind Pawar, Omkar Ghadge and Prof. Vijaya Chavan, "Reinforcement Learning" (2018)

[4] Muddasar Naeem, Syed Tahir Hussain Rizvi, And Antonio Coronato, "A Gentle Introduction to Reinforcement Learning and Its Application in Different Fields" (2020)