T.C.
ALTINBAS UNIVERSITY
ELECTRICAL AND COMPUTER ENGINEERING

AUGMENTED RANDOM SEARCH APPLIED IN ARTIFICIAL INTELLIGENCE

OTHMANE EL MEZIANI

Master Thesis

SUPERVISOR:
Asst. Prof. Dr. OĞUZ ATA

ISTANBUL, 2019

AUGMENTED RANDOM SEARCH APPLIED IN ARTIFICIAL INTELLIGENCE

by
ELMEZIANI OTHMANE

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science

ALTINBAŞ UNIVERSITY
2019

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

____
Assoc. Prof. Dr OĞUZ ATA

Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Prof. Dr. Hasan Hüseyin BALIK	Air Force Academy, National Defense University	____________________
Asst. Prof. Dr. Oguz ATA	School of Engineering and Natural Science, Altinbas University	____________________
Prof. Dr.Osman Nuri UÇAN	School of Engineering and Natural Science, Altinbas University	____________________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

____
Asst. Prof. Dr ÇAĞATAY AYDIN

Head of Department

____
Asst. Prof. Dr OĞUZ BAYAT

Director

Approval Date of Graduate School of
Science and Engineering: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

EL MEZIANI OTHMANE

ABSTRACT

AUGMENTED RANDOM SEARCH
APPLIED IN ARTIFICIAL INTELLIGENCE

El meziani, Othmane,
M.S, ECE, Altınbaş University,
Supervisor: Ass.Prof.Dr.OGUZ ATA

Date: 04/2019
Pages: 49

Random search algorithms are helpful for several ill-structured world improvement issues with continuous or distinct variables. Usually, random search algorithms give a perfect optimality for locating an honest resolution quickly with convergence. Random search algorithms embrace simulated annealing, tabu search, genetic algorithms, biological process programming, particle swarm improvement, pismire colony improvement, cross-entropy, random approximation, multi-begin and bunch algorithms, to call some. They will be classified as world versus native search, or instance-based versus model-based. However, one feature these ways share is that the use of likelihood in determinative their repetitious procedures. This text provides a summary of those random search algorithms, with a probabilistic read that ties them along. Augmented Random Search is actually one among the foremost mind-blowing algorithms where the fundamental is using a systematic approach particularly Augmented Random Search. It is a newly born methodology for Reinforcement Learning which suppose to use the strategy for limited contrasts and it can do precisely the same that Google Deep Mind did in their achievement a year ago, in other words an AI to walk and keep running over a field. With the ongoing selection of standard benchmark suites, a huge assemblage of late research has connected RL strategies for constant control within recreation situations.

Keywords: Artificial intelligence, ARS, Reinforcement Learning, Random Search, Mujoco, Pybullet.

TABLE OF CONTENTS

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

In this thesis we are going to introduce the main content which is augmented random search. It display background of the study, problem statement and justification, the purpose and objectives of study, the research questions which the study attempts to answer, significance, scope, limitation and organization of the study.

1.1 GENERAL INTRODUCTION

A common belief in model-free reinforcement learning is that strategies supported random search within the parameter benchmark exhibit considerably sample complexness than those who explore benchmark of actions. we tend to dispel such beliefs by introducing a random search technique for coaching static, linear policies for continuous management issues, matching progressive sample potency on the benchmark MuJoCo tasks. Our technique conjointly finds the virtually best controller for a difficult occurrence of the LQR (Linear-Quadratic-Regulator), once the dynamics aren't familiar.

Computationally, our random search algorithmic rule is a minimum of fifteen times and the quickest competitor model-free strategies on these benchmarks. we tend to profit of this Simulation potency to judge the performance of our technique over many random seeds and lots of totally different hyper-parameter configurations for every benchmark task. Our simulations improve a high variability in performance, implying that normally used estimations of sample potency don't adequately measure the performance of RL algorithms.

Machine learning is a science which was found and developed as a subfield of artificial intelligence in the 1950s. The first steps of machine learning goes back to the 1950s but there were no significant researches and developments on this science. However, in the 1990s, the researches on this field restarted, developed and have reached to this day. It is a science that will improve more in the future. The reason behind this development is the difficulty of analysing and processing the rapidly increasing data. Machine learning is based on the principle of finding the best model for the new data among the previous data thanks to this increasing data. Therefore, machine learning researches will go on in parallel with the increasing data. This research includes the history of machine learning, the methods used in machine learning, its application fields, and the researches on this field. The aim of this study is to transmit the knowledge on machine learning, which has become very popular nowadays, and its applications to the researchers.

This thesis demonstrates random search algorithms that comprise 2 classes. The primary class is regionally best sampling based mostly mechanical phenomenon optimization strategies. The second is ARS methodology with supervised learning concept. This thesis also presents a varied dynamic programming that could be a random search mechanical phenomenon optimization methodology, derived from the differential dynamic programming rule. This permits the statistics to be recomputed from sampled knowledge rather than utilizing differentiation to get them.

The thesis additionally presents ways in which to regularize the SaDDP rule with efficiency.ARS methodology bestowed during this thesis alter of sophisticated systems, like physics-based 3D characters. The strategies perform a receding horizon golem search and use the information created by the surroundings search to show machine learning models the way to higher seek for the actions within the future. The incontestable combination of receding horizon search and supervised learning is quick to converge and yields sturdy learning. The ARS bestowed during this thesis combines data from multiple sources.

This thesis presents the way to mix the data from varied sources in such some way that the search adapts to the data sources agreeing or disagreeing. Augmented random search is a vital algorithm in several application areas. It is an example a central tool in artificial intelligence. Several widely used strategies like differential dynamic programming (DDP) are supported differentiating the dynamics of the controlled systems and also the objective. The logic that one would have right to access a various model of the whole system doesn't hold for several systems of interest as an example, collisions break this assumption. During this case one needs to resort to random search (Augmented) algorithms.

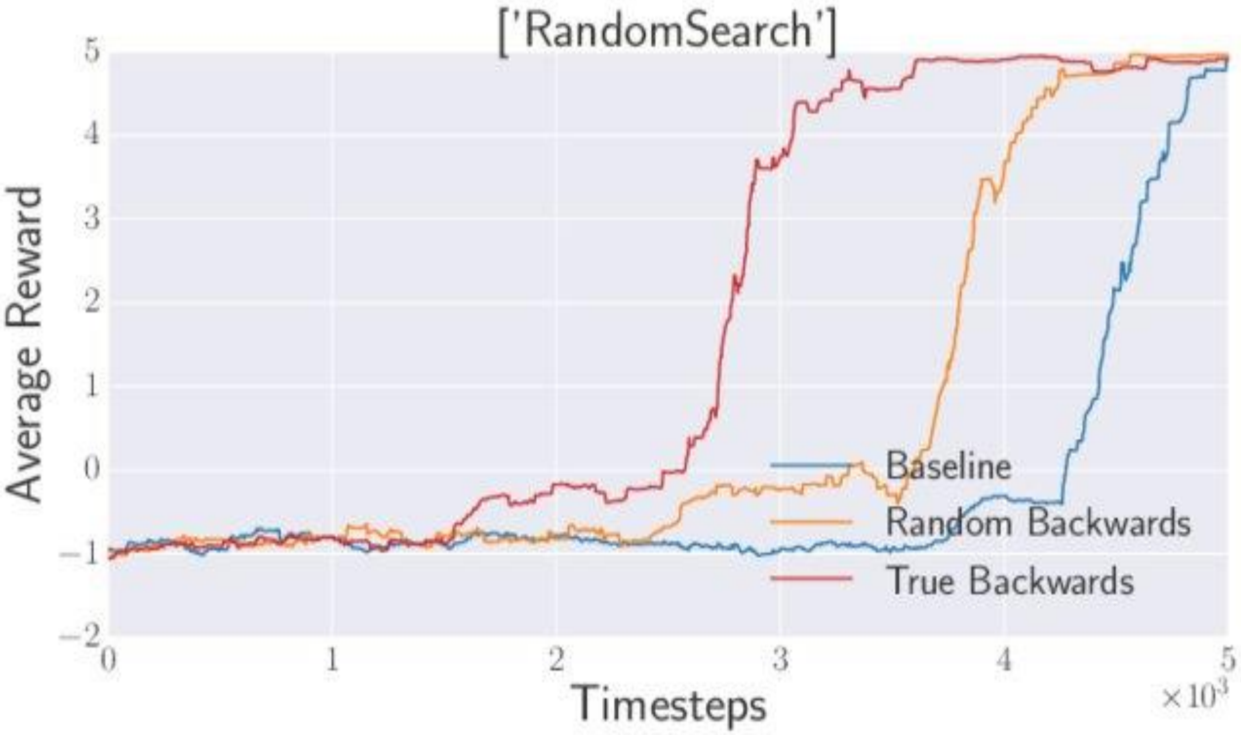


Figure 1.1: Training curves from the Four Room Environment for the Actor-Critic baseline [4]

1.2 LITTERATURE REVIEW

When dealing with ARS, there are three main updates that involve scale update step by standard deviation of rewards, online normalization of states and discarding directions that yield the lowest rewards(figure 1).The first step of scaling the updates requires that one divides the coefficient with the standard deviation of the rewards. The next step is normalizing the states whereby it is possible to have different input values that produce different weights. The resulting weights are responsible for producing the output values. Therefore, it implies that the weights are responsible for determining the output values that will be obtained from the input values. When utilizing the method of finite differences to deal with augmented random search, the process can be divided into two layers namely the input layer and the output layer as illustrated in the figure below.

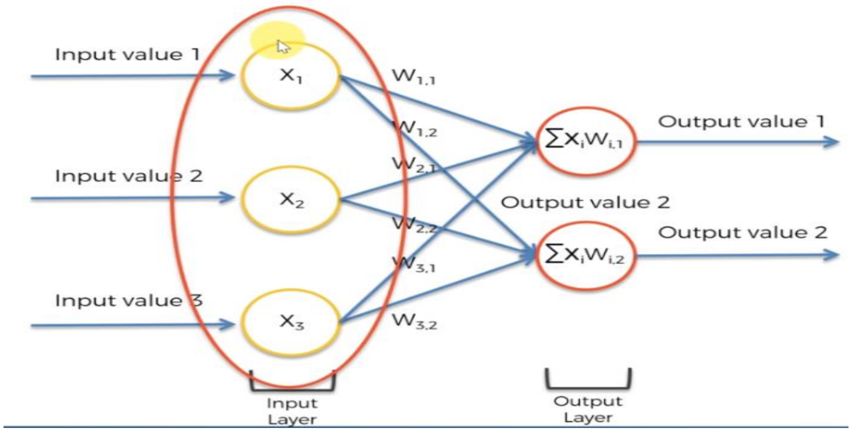


Figure 1.2: Input layer and the output layer

It is necessary to normalize inputs (figure 2)which is also referred to as widening of states because when the weights are adjusted, they have a significant impact on the output values that will be achieved. The reason why normalization is necessary is because undesired results will be achieved when the input values are varied. After the states have been normalized, the third update is discarding the states that produce the lowest rewards. An example of this process is illustrated in the figure below.<

In general terms, artificial intelligence is simply a form of intelligence that is shown by software or machines. It may not be possible to discuss Augmented Random search without considering controlling engineering because it forms the basis of how the agents are being trained and how they function. This is because it is essential to make machines learn from human actions so that they can be able to solve problems. However, it should be noted that intelligent control is mainly concerned with dealing with real life situations which are in the case of ARS used to design environments where the agents will be trained. Although AI is divided into many categories that may not entirely correlate, it is clear that there are common and main objectives which are perception, reasoning, communication, knowledge, learning and planning. The field of AI is interdisciplinary whereby it involves the interaction of many professionals ranging from neuroscientists, control engineers, philosophers, computer scientists, linguists, mathematicians and psychologists. The reason why AI is so interdisciplinary is because it falls in the categories of being both specialized and technical.

AI has many tools that include mathematical optimization, search, probability methods and logic just to mention a few. Some of the tools that are essential for one to understand when considering ARS and how it relates with simple random search and linear controllers are systems based on knowledge, ambient-intelligence, fuzzy logic, reasoning based on cases, acquisition of automatic knowledge genetic algorithms and neural networks. AI specialists understand that it may not be practical to capture the entire capability of the human brain and train a machine to emulate and therefore efforts are directed at finding ways of tapping into at least fractions of human intelligence and transferring the same to machine. The transfer of human intelligence to machines involves simulation processes. Although the idea of having machines that can think on their own may present benefits, it is without a doubt that issues of whether developing this technology has resulted in the questioning of whether it is an ethical practice.

Machine intelligence began as a simple idea of venturing into a journey of making or building a machine that resembled a child. This is because a child is born without knowledge and goes ahead to keep on learning and improving the way the child thinks. The learning process is facilitated by the environment that the child is in. it was in 1940 that the field of machine intelligence was proposed and when the Second World War ended, different researchers embarked on a journey of exploring how the idea can be converted into practical research. Among the first researchers that actively produced results in terms of presenting ideas and predictions are Zadeh and Alan Turing. In the lecture that was given by Turing in 1947, he predicted that by the time the century ended, there would be in existence what is known as intelligent computers. In the 1950 publication titled "Thinking Machines- A New Field in Electrical Engineering," Turing and Zadeh presented a discussion that involved the criteria that could be used to determine whether a machine could be considered to be intelligent. Turing advocated for the idea that if there should arise a situation where a machine is capable of pretending to be a human and perceived to be so by an observer that is knowledgeable, then it would be fair to consider the machine intelligent.

The term "Artificial Intelligence" was proposed by John McCarthy during a gathering in 1956 that consisted of computer scientists. The gathering was held in New Hampshire at the Dartmouth College. The debate that was done during the meeting formed the foundation of how machines can be simulated so that they can be able to adopt human intelligence through improvement by imitating the human cognition process. The discussion was keen on inventing and creating ways that could describe learning as well as other human intelligence aspects. From the discussion, McCarthy came up with a definition of Artificial Intelligence whereby he defined it as "the science and engineering of making intelligent machines." For a machine to exhibit intelligence, it is necessary to adopt procedures that lead to the combination of advanced technologies which are responsible for enabling the machine to show traits of learning, adapting, making decisions and displaying new behaviors. There are many applications of artificial intelligence such as:

a) Robotic manipulation
b) Robotic programming
c) Human-computer interaction
d) Walking robots
e) Computer vision
f) Wheelchair assistance
g) Assembly

Simple Random Search and the Linearization Principle mean When considering simple random search and the linearization principle, there are many variants that arise that could lead to undesired results. However, it is worth noting that linear models can be used as indicators of the behavior of machine learning algorithms. It is so because their behavior in these models directly represents the behavior in non-linear models that exhibit complexity. Therefore, the linearization principle is essential since it can be used in the decomposition of complex issues so that they can be dealt with as simple problems during research where the problems are tractable. Although the application of linearization principle in reinforcement learning may not be adequate, it provides a foundation for understanding machine learning. It is notable that it is not possible to have equal generalization even when dealing with global minimizers. This can be understood better by considering a situation where the number of data points is half the number of parameters. It implies that it is possible to achieve a zero error using a training set that has N degrees of freedom. This information can be used to create training data that can still give a zero error even after adding a model that has a zero error in training. This is possible through interpolation of random labels on the perturbed data. The resulting model can be considered to be a global minimizer of the true training set for the training error but still the model would be incapable of generalizing effectively. From the experiment using the model that we have created, it is evident that effectiveness is exhibited when the shallow minimizer is translated to a larger margin.

When dealing with simple random search in the tune of linearization, it is notable that when batch norm layers are inserted, the overall effect is that SGD is sped up. It remains unclear why this effect is exhibited but it has been proven that when standardization is done, the SGD becomes accelerated. However, an explanation can be provided by considering the effect of whitening in the case of data matrices because it leads to improvement of covariance of the data. This improvement is responsible for improving the rate at which convergence takes place of SGD. If the model has less data points than parameters, the resulting factor is that there will be many local minimizers for the training error. From previous research, it has been indicated that a type of regularization is achieved when perturbing back-propagation is randomized although dropout is different in simple random search compared to the same process in deep models. In the case of linear models, deep nets are known to have the ability of memorizing random labels and at the same time generalize.

Linear Controllers in artificial intelligence is an important tool in the field of control engineering since robotics is one of its beneficiaries. The technology is also applied in other instances like cars and wheelchairs. In application of this know-how in artificial intelligence, it is necessary to have inner control loops that are responsible for manipulating the dynamic systems so that they can complete the desired objectives. Therefore, it is essential to design controllers so that desired tasks can be completed by the concerned stakeholder and in this case the agent in the learning process. An important consideration is the design of interfacing systems as well as actuators and sensors.

Inner control loops in artificial intelligence technology is applied in designing of inner control loops because they are designed in a manner similar to the nervous systems in animals. The control loops are designed so that they have the ability to take action without necessarily doing so consciously. This ability is equated to natural functions such as the regulation of the rate of heart beat, respiratory rate and pupillary response. In the case of animals, a similar example would be the situation of adrenaline rush which initiates an animal's action of either fighting or fleeing as a defensive mechanism. Considering the mechanism, the nervous system is autonomous and comprises of two systems namely the parasympathetic and sympathetic nervous system. The first is a slow activated dampening system that is activated whereas the latter is a quick response system that is responsible for quick reactions. In artificial intelligence, inner control loops are responsible for controlling the robotic agents quickly so that they can react to the information relayed by the slow acting sensors responsible for the monitoring processes.

For Outer control loops these are different from inner control loops although they work in collaboration to achieve results. It is not possible for the inner control loops to function by themselves because as it has been explained in the previous section, they are similar to the nervous system in animals. In the case of outer control loops, they do not require inputs which act as profiles or reference points for them to produce results. Outer control loops are equated to the brains in animals because in their functioning, they do not necessarily have to be predictable automatically and therefore they tend to be more conscious. It is normal for the brain to be regarded as the highest center for serving the controlling function because it is responsible for processes such as swallowing, walking and talking. The brain is responsible for thinking functions and therefore understanding how it functions is important because it allows designing of outer control loops that enable the functioning of Artificial Intelligence (AI) in and specifically in the case of Augmented Random Search (ARS).

Automatic controle is Considering the technology that is used in Artificial intelligence where a robot or a robotic mechanism is designed in a manner that it is able tom function on itself without intervention from a human, it is correct to deduce that an automatic control system has been utilized in the process. The technology that encompasses automatic control is not a new phenomenon since the invention was recorded over 2000 years ago in ancient Egypt. This is the mechanism that was used in the construction of the water clock that was named *Ktesibiosis.* There were also other inventions that were made like the furnace that was built by Drebbel which had a temperature regulator. The furnace was built in 1620 and in the 17th century, James Watt explored the technology and found a way of regulating the speed of steam engines by building a centrifugal fly ball governor. Owing to this invention, many control systems during this era utilized the governor mechanisms for regulation of different processes. In order to understand the governor mechanism, it is necessary to use differential equations that aid in understanding the dynamics of the systems and comprehend how the same can be stabilized to achieve the desired results as in the case of ARS. This knowledge is useful because it enables measuring of the output performance using sensors that are used when building a device such as a robotic agent in Reinforcement Learning. The measurements that are obtained from the sensors are useful in providing the necessary feedback that is relayed to the actuators which in turn are responsible for making corrections that are aimed at achieving the desired results.

1.3 WHY REINFORCEMENT LEARNING

Reinforcement learning describes the set of learning issues wherever Associate in nursing agent should take actions in Associate in nursing atmosphere so as to maximize some outlined reward perform. Unlike supervised deep learning, massive amounts of labeled information with the right input output pairs don't seem to be expressly bestowed. Most of the training happens on-line, i.e. because the agent actively interacts with its atmosphere over many iterations, it eventually begins to find out the policy describing that actions to require to maximize the reward.

The method reinforcement learning models the matter needs many conditions: we can quantify all the variables the atmosphere describes and have access to those variables at on every occasion step, or state. Neither is also the case within the real world; additional usually than not you simply have access to partial info. the knowledge that you just do have access to itself will be inaccurate and in would like of any extrapolation, since it's measured from Associate in Nursing egocentric purpose of read (at least within the case of a golem interacting with Associate in Nursing unknown environment).

Since learning is preponderantly on-line, you have got to run trials several over and over so as to supply an efficient model. This is often acceptable once the task at hand is straightforward, actions are distinct, and knowledge is quickly offered. however in several cases, the matter formulation is considerably additional advanced and you need to balance the preciseness of your machine with each coaching time and period of time performance constraints .It is thanks to these limitations that recent successes in reinforcement learning have happened nearly entirely in simulated, controlled environments (think DeepMind's analysis on Atari, AlphaGo). There is still tremendous analysis required in overcoming these limitations and adapting deep RL to figure effectively in period of time agents.

1.4 PROBLEM STATEMENT

The deep RL community has invested a big part of time and energy on a set of benchmarks, occurred by OpenAI and positioned the MuJoCo simulator. The main control problem is to grab the simulation of a legged character or Robot to move as much and quickly as achievable in one or different direction. A number of the tasks are easy enough, however some are quite hard just like the sophisticated robot models with twenty two degrees of freedom.

The dynamics of leg-like humanoid are specified by Hamilton-Equations, however coming up with locomotion from these models is difficult as a result of it's not clear the way to best style the target perform and since the model is piecewise linear. The model changes whenever a part of the automaton comes into contact with a solid object, and this a standard force is introduced that wasn't antecedently acting upon the automaton. Therefore, obtaining robots to figure while not having to handle sophisticated non-convex nonlinear models seems to be a solid and attention-grabbing challenge for the RL paradigm.

1.5 RESEARCH OBJECTIVES

1.5.1 General Objectives

In this research, we adopted a qualitative research approach, where the main method which is the literature review using a systematic approach particularly Augmented Random Search. The dynamics of legged robots are well defined by Hamiltonian Equations, but planning Movement from these models is complex challenge because it is not clear how to get the best design character function.

1.5.2 Specific Objectives

- To determine how Augmented Random Search model can be used to anticipate the number of perceptrons.
- scale each update step by the standard deviation of the rewards collected for computing that update step
- Use different environments with different benchmark rules.
- Design, implement and test the ARS model for determining the performance and result instead of other algorithms results.
- Normalize the system’s states by online estimates of their mean and standard deviation

1.6 THESIS CONTRIBUTIONS

The principal purpose of this thesis is designing and implementing an interactive humanoid that takes into account the different inputs that can be generated from the local environment to grant loans to non-risky and profitable customers. This tool can recommend the suitable customers and determining the credit worthiness of customers according to some factors such as age, gender, pregnancy, Income, Job, Housing, Education, Family numbers and occupations. We passed through three phases toward building the proposed tool: data-set preprocessing, processing using data-mining techniques and status prediction.

1.7 SCOPE OF THE STUDY

The study was confined to only one selected robot design provided by Pybullet. The algorithm used in the model was improved to minimize the reward score in parallel with the environment.

1.8 RESEARCH QUESTIONS

- What can reinforcement learning learn from random search?
- What makes a good benchmark for RL?
- Can simple random search find linear controllers?
- Does random search break down as we move to harder problems?
- What’s the method of finite Differences?
- How Ars algorithm works?
- What’s the difference between Pybullet and Mujoco?
- How Perceptron Works?

1.9 ORGANIZATION OF THESIS

The study under discussion is structured into 5 chapters. The First Chapter one is about the introduction which has the background of the study, Why Augmented Random Search, Statement of the problem and justification, Scope of the study, Research questions and Organization of the study as its sub-headings.

In the second Chapter, is about a literature review. All Previous related works in the light of artificial intelligence in reinforcement learning. The sections are artificial intelligence and AI techniques (Robotic Process Automation, Natural Language Generation, NLP, Machine learning, Virtual

Agents, Speech recognition,), Processes developed in AI and application of Reinforcement Learning, how AI can assist innovation Research.

Chapter three contains the study methodology. The following headings are discussed. Research design (Overview of ARS, Reward maximization, Method of finite differences), model evaluation methods or we can call it reward model, the difference between AI algorithms process for Reinforcement learning we introduce the use of AI, a part of mysterious optimization algorithms, as an Plan B to famous Reinforcement Learning techniques such as Q-learning and Policy Gradients and privacy and confidentiality.

Chapter four presents system modeling and performance evaluation procedures. It discusses how the models was developed and evaluated and Chapter five is concerned with results and analyses.

2. BACKGROUND

This chapter discusses the relevant literature of other authors who did research into areas where Reinforcement learning techniques are very helpful to solve daily problems since it is considered as part of machine learning paradigms. This chapter is grouped under the following sub-headings:

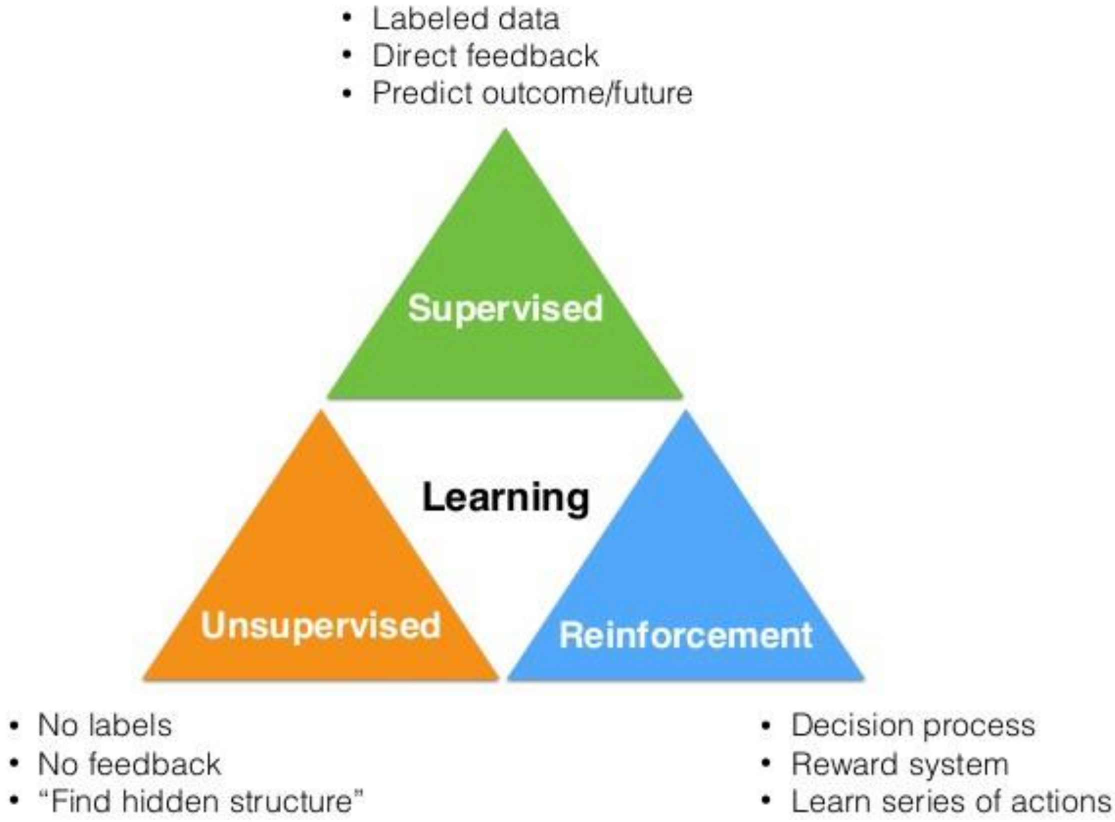


Figure 2.1: Machine learning paradigms [6]

2.1 REINFORCEMENT LEARNING CATEGORIZATION

Reinforcement learning is the process that is undertaken to train a machine referred to as a learner or an agent (figure 2.1)so that it is in a position to make decisions on its own. The training is conducted in an environment characterized by complexity and uncertainty and the agent is required to make a set of decisions that have an influence on the environment. The rewards achieved for every process or step depends on the choices or decisions made by the agent and the overall rewards are the summation of individual step reward. In the reinforcement process, there must be a level of interaction between the agent and the environment whereby the interaction is determined by the policy.

The way that people learn has been considered in different angles whereby in most cases learnng processes have been theorized and based on assumptions made by various scholars. Reinforcement learning approach is keen on looking into learning processes by using computational methods. This approach is concerned with how people learn through interaction. To understand this concept, it is necessary to develop an artificial intelligence expert’s perspective. Reinforcement learning is associated with how machine learning is used using the modern technology to facilitate learning and solving complex problems [1].

Although other machine learning processes are also concerned with learning specific goals, reinforcement learning deals with learning processes that take place due to interaction with situations. Maximizing a reward signal is essential in reinforcement learning whereby problems that exist in a “closed-loop” system are learned and allow mapping of incidences thus solutions to the problems can be recommended or formulated. It is crucial to note the importance of the closed-loop nature of the system which is so because the actions that are taken by the learner or machine after learning are influencing factors of the inputs in later stages of the process [2]. In many instances, reinforcement learning is mistaken for unsupervised learning but that is not the case. The reason for this misconception is because reinforcement learning does not rely on correct behavior. A trade-off exists between exploitation and exploration whereby it necessary for the agent to explore new actions which will produce inputs that will be used in the future. At the same time, it is crucial that the agent exploits the knowledge that it already has in terms of actions. It implies that exploitation is based on the current actions while exploration is based on future actions. Studies have found out that the trade-off between exploitation and exploration does not take place in both unsupervised and supervised learning.

2.2 ASPECTS OF REINFORCEMENT LEARNING

2.2.1 Environment Model

The model is a kind of an environment that has an influential role in determining how inferences will be made. The model has a significant impact on how the behavior will be characterized in the sense that it has the ability to determine the action and state of the learning environment [3]. Their determinacy on the future actions and states makes models facilitate planning.

2.2.2 Policy

The policy is responsible for determining the behavior of the learner or agent at all times and therefore it can be considered to be the map or plan of how the environment will look like [4]. It also determines how the learner will behave when they are in the pre-determined environment. The policy has the most significant impact on the agent because it is capable of determining the behavior of the agent on its own.

2.2.3 Reward Signal

Association Rule Mining developed much later than machine learning and is directed to greater control from the investigation field of databases. Although association rule mining was first presented as a market basket analysis agent, it has since enhanced one of the most important means for performing unsupervised exploratory data analysis over a comprehensive range of investigation and investment fields. Usually, Association is one of the essential approaches of data mining that is used to find out the familiar patterns, new relationships among a set of data items in the data repository [19][20].

2.2.4 Value Function

The value function is part of the model in away because it determines the number of rewards that the learner may expect to achieve. It is a determination of the best achievable results that the agent can get as well as the optimum number of rewards that the agent may be expected to get in order to achieve good results.

2.2.5 Multi-arm Bandits (Example)

This element of reinforcement learning is aimed at dealing with how the agent makes decisions given that there are many options to choose from. In a conventional model, the number of actions available to the learner is denoted as *n*. It represents the number of choices that have to be made whereby after every choice that the learner makes, there is a reward attached to the decision before proceeding to the next action [6]. The multi-arm scenario can best be explained by considering a slot machine where the player or user operates it severally with the overall outcome depending one very operation. The results obtained in every operation are considered to be the rewards which in this case are the payoffs. Therefore, the total rewards or payoffs will be the accumulated rewards for all the operations represented by *n*.

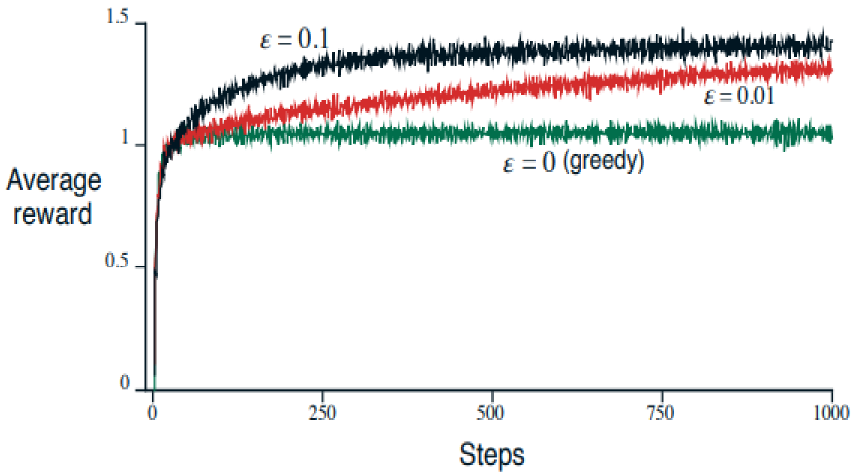


Figure 2.2: Graph of average rewards in a multi-arm bandit model [9]

The agent is required to interact with the environment so that learning can take place. If the agent fails to interact in accordance with the policy, then the expected rewards may not be achieved check (figure 4). The interaction process is a continuous process whereby the agent is required to make decisions or select desired actions while at the same time the environment is also expected to respond to the actions taken by the agent [7]. It is the environment’s response to the current actions that result in the formation of new situation for the agent in the next selection process.

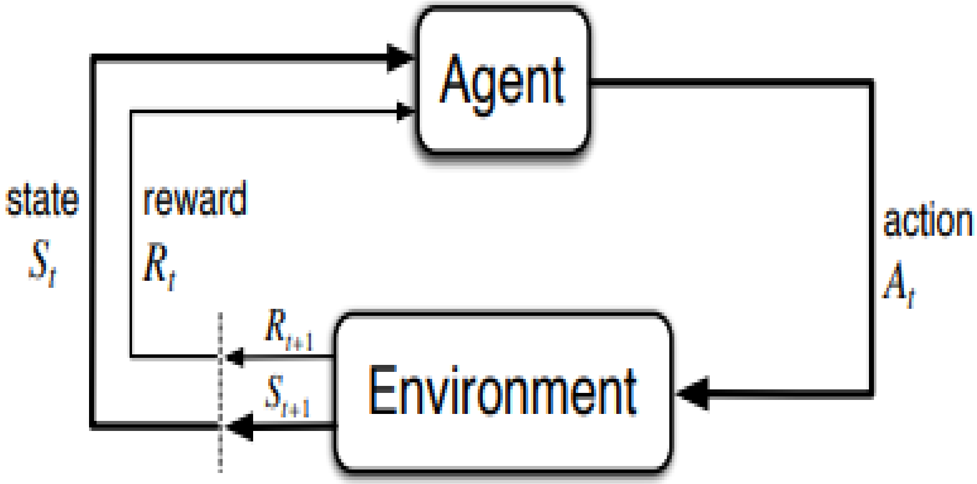


Figure 2.3: Agent-environment interaction [8]

Every complete interaction between the agent and the environment is referred to as a task denoted as *t.* the tasks are numbered in succession with the rewards labeled in accordance to the task (figure 2.3).

As stated earlier, reinforcement learning is involved with dealing with complex situations that have uncertain results. It means that it may be possible to predict the exact outcome or the rewards that will be achieved by the agent. Reinforcement learning relies on using models in predetermined environment that the agent is expected to interact. However, it becomes difficult to achieve the expected results when the events are transferred to the real world where the outcomes may be entirely different from those obtained in the model.

Although reinforcement learning may be an expensive venture that has its fair share of challenges, it is a crucial process that presents possible solutions to complex matters that are experienced in real-world. A proper understanding of reinforcement learning requires knowledge in the policy, environment and the model. It takes place when the agent or learner interacts with these elements to achieve rewards.

As stated earlier, reinforcement learning is involved with dealing with complex situations that have uncertain results. It means that it may be possible to predict the exact outcome or the rewards that will be achieved by the agent. Reinforcement learning relies on using models in predetermined environment that the agent is expected to interact. However, it becomes difficult to achieve the

expected results when the events are transferred to the real world where the outcomes may be entirely different from those obtained in the model.

Although reinforcement learning may be an expensive venture that has its fair share of challenges, it is a crucial process that presents possible solutions to complex matters that are experienced in real-world. A proper understanding of reinforcement learning requires knowledge in the policy, environment and the model. It takes place when the agent or learner interacts with these elements to achieve rewards.

2.3 OVERVIEW OF AUGMENTED RANDOM SEARCH

In augmented random search, reinforcement learning is utilized when the process of finite differences is applied. It allows for adjustment of weights that allow learning so that tasks can be completed.

2.3.1 Procedure

It is essential for the weights to be adjusted appropriately and this is done through randomizing of matrix that has small values. These tiny values are then added to the weights. After the values are added, the artificial intelligence then adds the random matrix with the same numbers but the numbers are negated so that negative weights are obtained. The process is repeated numerous times so that the agent is able to perform the task but each time with different weights [8]. An example of the process is illustrated in the figure below.

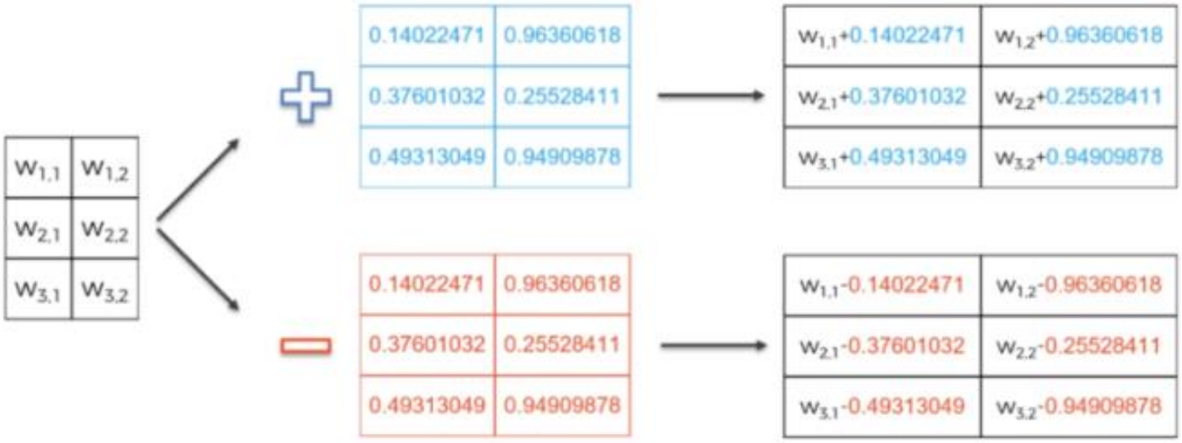

w1,1 w1,2
w2,1 w2,2
w3,1 w3,2
0.14022471 0.96360618
0.37601032 0.25528411
0.49313049 0.94909878
w1,1+0.14022471 w1,2+0.96360618
w2,1+0.37601032 w2,2+0.25528411
w3,1+0.49313049 w3,2+0.94909878
0.14022471 0.96360618
0.37601032 0.25528411
0.49313049 0.94909878
w1,1-0.14022471 w1,2-0.96360618
w2,1-0.37601032 w2,2-0.25528411
w3,1-0.49313049 w3,2-0.94909878

Figure 2.4: Finite differences process [10]

When each weight in (figure 2.4) is configured, there is a corresponding reward achieved and it is notable that when the tasks are extracted from the environment, some weights are found to be higher than others. According to reinforcement requirement, it is essential that the highest rewards are achieved. This is possible when the ARS adjusts the weights so that in line with the weight configurations. In order to achieve maximum results, it is necessary to adjust the weights more [9]. When the weights are adjusted less, the rewards obtained are also less. The formula used to calculate the rewards is illustrated below (figure 2.5).

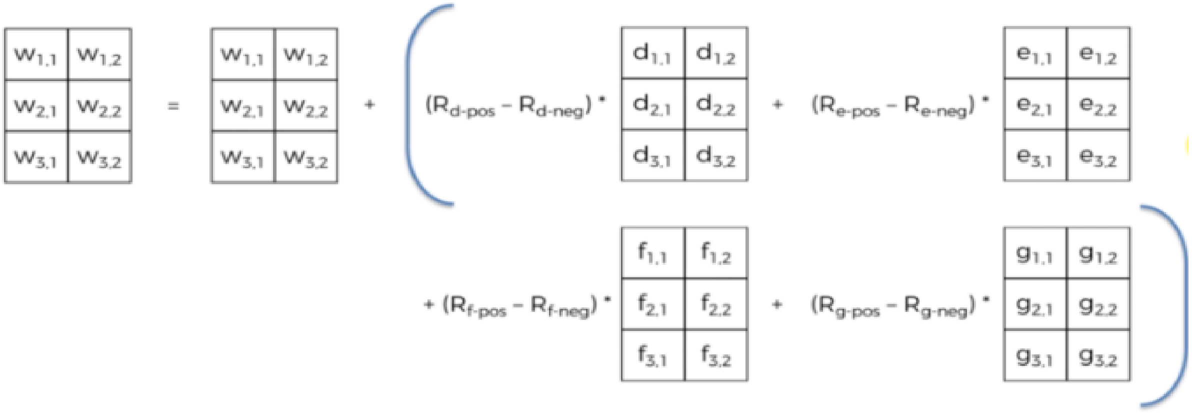


Figure 2.5: Augmented random calculation for Ars algorithm [10]

From the figure, it is possible to note that there are four weight configurations whereby the coefficient is the difference between the negative configuration and the positive configuration of the corresponding weight. From the equation, the low rewards were not included because they were discarded. Therefore, the upper k configurations were used to ensure that only the maximum rewards were achieved and used. Discarding low rewards facilitates using less time to compute thus enhancing computational power. Apart from algorithms, Augmented Random Search also explores policy spaces and therefore it does not rely on action spaces. It follows that the agent analyzes the rewards only after a number of actions have been taken rather than analyzing the rewards after every action.

Rather than using a deep neural network, ARS makes use of perception to run the process. In order to ensure that maximum rewards are achieved, tiny values are added to the weights but negative values are used in the configuration process [11]. By so doing, it becomes possible to obtain bigger rewards. It is also notable that when the rewards are bigger for individual weights, the influence on the adjustment process also becomes large.

Although reinforcement learning may be an expensive venture that has its fair share of challenges, it is a crucial process that presents possible solutions to complex matters that are experienced in real-world. A proper understanding of reinforcement learning requires knowledge in the policy, environment and the model and what takes place when the agent or learner interacts with these elements to achieve rewards [12].

2.4 ARS SUMMARY

Neural networks which are usually called artificial neural networks, are mathematical models that have their root from biological neural networks. They possess the capability of processing data in a non-linear form using appropriate statistical tools [13]. Neural networks contain a number of linked parts usually called neurons, units, or nodes. The units in neural networks collaborate with one another to come out with an output function. Thus the units of neural networks are interconnected together in such a way that they execute input information by adopting a connectionist style of calculation. Neural networks possess the capability of identifying relationship between a set of variables (data) that are extremely difficult to discover using the human brain or some other methods using computer. For the reason that neural networks use connectionist style of calculation, they are able to work even when some units are not working properly [14].

3. PYBULLET

Physics has a branch that is concerned with deep reinforcement learning, simulation and robotics that utilizes a simple python module that is referred to as Pybullet. This interface utilizes the Bullet physics SDK that makes it possible for the user to load file formats, SDF and URDF that are articulated bodies [15]. Pybullet is very useful in the engineering field because it offers a variety of functions ranging from ray intersection queries, forward dynamics simulation, inverse and forward kinematics, collision detection and forward dynamics simulaton.th functions of Pybullet are not only limited to these but it also offers virtual reality headsets support, OpenGL visualization and CPU rendering. These are some of the functionalities that are offered by Pybullet.

Although it is a powerful tool, it is remarkably easy to use as well as to install. Under the physical objects category, Pybullet software can be found for free and therefore one can be able to conduct simulation of different bodies that interact with each other. After the project has undergone complete simulation, it was initially shared on SoureForge.net before it was sent to Google Code under finally licensed under zlib. Currently, the Pybullet simulation project is shared on GitHub. This type of physics technology is attributed to Erwin Coumans who is the lead author. His founding works was the Havok Project. Pybullet uses the TinyRenderer for the rendering process as well as binding [16].

Pybullet has features that make it suitable for carrying out simulations. It works with different shapes when it comes to handling collisions. These shapes are mesh triangles, spheres, convex hull, rectangular parallel piped and cylindrical shapes. Pybullet has the capacity to calculate distance algorithm which it does using the GJK. It can also be able to continuously detect collisions. It can also conduct physics support using COLLADA [17]. Another useful feature of Pybullet is that there are different modules available that makes it possible to customize different physics requirements.

4. MUJOCO

This is a system that is meant for testing various physics engineering fields such as machine learning, biomechanics, mechanical technology and dynamics frameworks that are complex. MUJOCO is an abbreviation for Multi-Joint Contact whereby the system was started in 2009 and it enhances innovation in the aforementioned fields [18]. Mujoco enables reproduction that is fast and quick.

Before it was approved the initial devices were determined to be incapable of sufficient ad effective exploration of frameworks, ideal control and estimation of states. Its approval was done at the University of Washington in its Movement Control Laboratory(figure 8). Although it was conducted as an experiment at the laboratory, its use has expanded with its growing popularity to the extent that it is currently applied by customers. It is essential to conduct stream lining of the numerical order and Mujoco makes it possible.

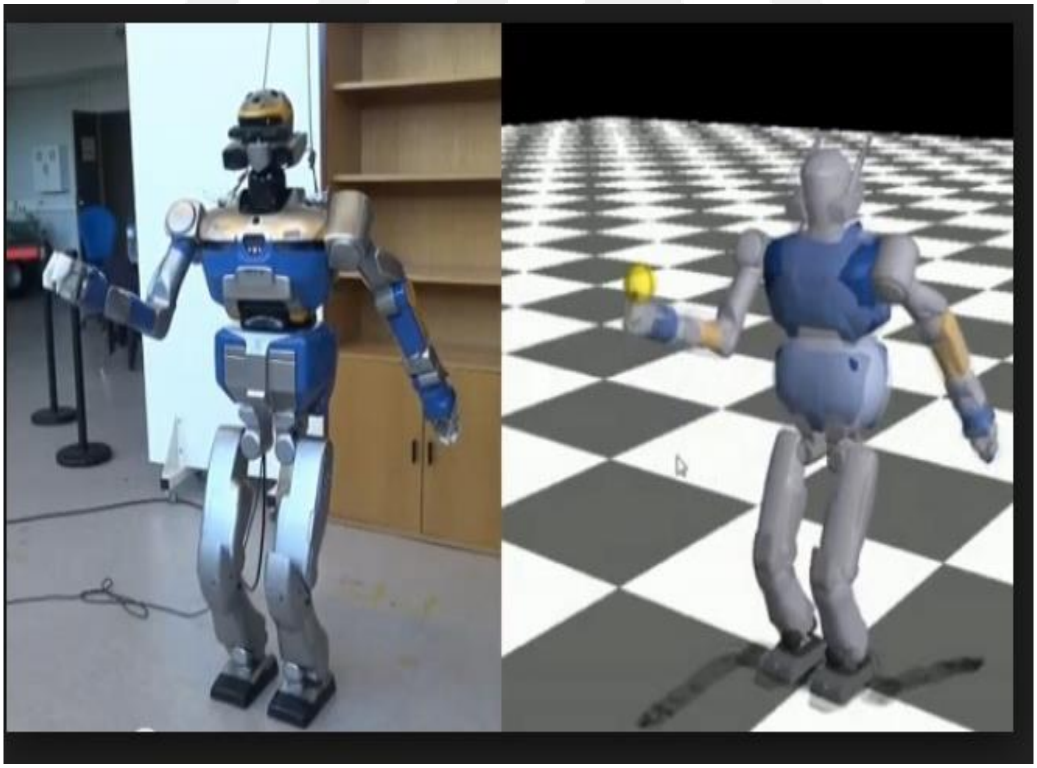


Figure 4.1: Balancing and Reaching with Model Predictive Control [5]

Due to its success in this physics engineering field, it was turned into a foundation. The system facilitates creation in material science which requires provisions that are enhanced by its enhancers. Material science requires steadiness and precision and Mujoco offers just that [19]. For these objectives to be realized, it is necessary that applications rely on subsidiaries that provide test element. These functionalities are carried out by motors that act as mechanical frameworks such as Mujoco which has been able to outdo various science motors in the functionality.

4.1 SIMILARITIES

Both systems are useful in running simulations in science engineering especially n the physics field and it is without a doubt that they can be considered successful in their endeavors. There are many systems that provide this functionality although most of them cannot achieve the success that these two systems have. The major similarity in these two systems is that they both the purpose of simulation [20]. They are useful in the fields of robotics, mechanical engineering and they have proven to be useful in improvement of Research and Development. They are both useful when it comes to simulation in physics and detecting collisions. These features are useful in machine learning, VR, robotics, interactive video games and visual effects.

4.2 DIFFERENCES

One of the major differences between Pybullet and Mujoco is the ease of setting up whereby Pybullet has an interface that is easy use. It is so because one does not need to have technical expertise to install the system. The same cannot be said for Mujoco which requires that one must have a certain level of expertise to be able to install and operate.

The other difference arises from the type of file format that is required to run the two systems. In the case of Mujoco, it only uses MJCF file format which is not the case for Pybullet because it utilizes both SDF and URDF when it comes to loading articulated bodies [21]. Another disparity in the two systems is the purpose that they were designed. The design of Pybullet is directed at gaming unlike Mujoco is intended for robotics.

However, it should be noted that Pybullet also has the capabilities of simulation but when it comes to robotics, Mujoco emerges superior. The two systems can also be differentiated when considering damping functionality. In the case of Mujoco, it provides implicit damping functionality which is facilitated by hinge PD controller implementation that is built-in. Pybullet's mode of damping is different from Mujoco’s because it makes use of spring dampers where damping is done at the hinge joints [22].

The raw timing of the two systems is different and although it may not be completely possible to deduce that one system is faster than the other, speed can be considered in terms of functionality. Pybullet is faster when it comes to gaming simulation while Mujoco is faster when it comes to robotics simulation [23].

The other aspect that can be used to differentiate the systems is the consistency. Consistency refers to the time that the CPU takes to carried out one update. In this perspective, Mujoco is far more superior to Pybullet probably this is the reason why its use requires technical expertise because of its ability to run complex simulations faster. However, Pybullet is superior in terms of capsule testing.

5. IMPORTANCE AND APPLICATION OF REINFORCEMENT LEARNING

The reason why research is conduced concerning reinforcement learning is because there are real life situations that give rise to complex problems that require solutions. Reinforcement Learning relies on construction of mathematical frameworks that are capable of solving such problems in virtual situations after which the solutions are applied in real life situation. It is necessary that a good policy is identified which is possible by using methods that have the ability of measuring value of the rewards obtained from specific actions.

Among the many methods that can be used to measure the value of rewards is Q-Learning. Although the problems that are encountered in the real world are complex and complicated. This complexity makes it difficult for typical algorithms to solve. Reinforcement Learning proves adequate since it provides for the possibility of carrying out simulations. These simulations (figure 9) can give an insight into probabilities or possible outcomes that can be anticipated after certain actions have been taken. Reinforcement Learning can be applied in various situations [24].

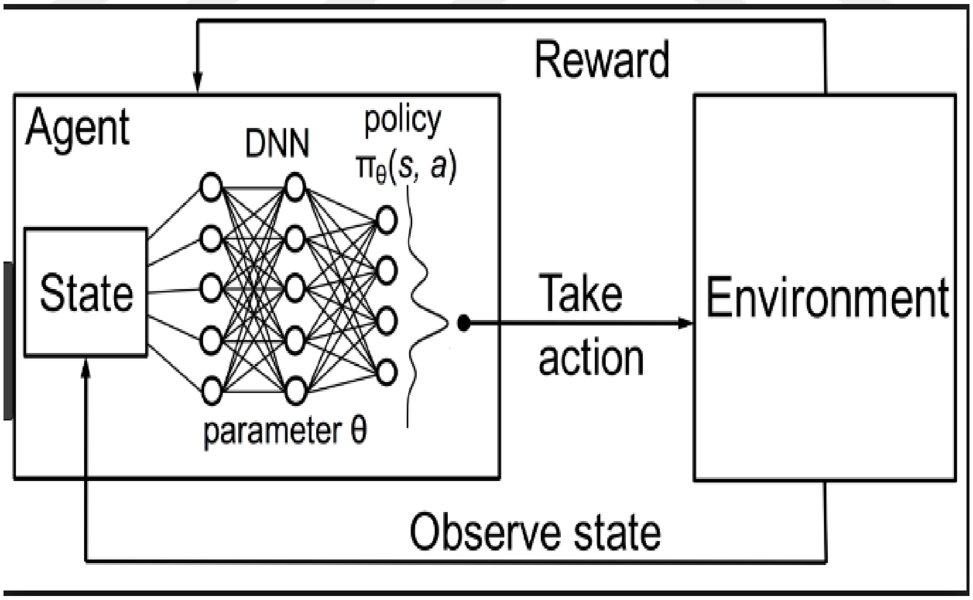


Figure 5.1: Reinforcement Learning with policy represented via DNN [11].

5.1 INDUSTRIAL AUTOMATION AND ROBOTICS

Reinforcement learning has gained popularity and interest in automation industry as well as the academia sector. Technological advancement has favored building of robotic products that are used in various industries. Robotic systems are being used by manufacturers and other business organizations to carry out many heavy tasks that require a lot f human effort. In this way, these stakeholders are able to cut down the cost of production thereby maximizing their output and thereby increasing revenue.

An instance where industrial automation has been applied is the case of Google that utilized DeepMind's Reinforcement Learning. The technology enabled Google to significantly reduce their consumption of energy (HVAC) in the data centers [25]. Another application of reinforcement learning is evidenced in the organization called Bonsai. Bonsai enables other companies to incorporated reinforcement learning techniques. One such instance is operation of tuning machines that are operated by trained personnel.

5.2 TRAINING AND EDUCATION

There are various researches that are ongoing aimed at innovating ways that machine learning can be used for the creation of personalized experiences. The researchers responsible for these case studies are responsible for coming up with creative ways that machine learning can be utilized to facilitate personalized learning and tutoring systems. Reinforcement learning can prove useful because the technology can enable designing of teaching strategies that are customized according to individual learner's needs [26].

Such strategies can make it possible to formulate customized materials and instructions. There are also studies that are underway that are aimed at coming up with statistical methods and reinforcement learning algorithms that will enable simulation that does not require big data. One of the problems associated with RL is that a large amount of data is required so that meaningful outcomes can be achieved.

5.3 MEDICINE AND HEALTH

Reinforcement learning is based on the technology of creating simulations whereby it is possible to virtually create a possible problem that may have many unknown outcomes. The purpose of this is to enable learning how such states can be dealt with by selecting the actions that provide maximum rewards. The technology can be applied in healthcare facilities through application of medicine sciences to obtain the best treatment policies [27]. In the health sector, research is conducted on a daily basis in an effort to find solution to complex problems such as chronic illnesses.

The objective in this case is to learn the diseases traits such as causes, symptoms, predisposing factors among other factors related to specific ailments. The purpose is to learn more about the health condition so that although some of them may be incurable, it becomes possible to formulate optimum treatment policies that can help mitigate the effects of the diseases. Reinforcement learning is utilized for conducting clinical trials, used in medical equipment and medication dosing among other uses [28].

Another medical area that is faced with problems associated with human error is in the intensive care unit. The normal procedure requires that there should be the ventilation process when carrying out operations. The ventilation is done mechanically despite the fact that a lot of precision is required. During ventilation, there is regulation of analgesia and sedation. This process has a history of incurring increased costs to the health facilities that have intensive care units. In addition, there is increased of development of complications in the process due to reliance on mechanical invasive ventilation.

5.4 ADVERTISING AND MEDIA

Reinforcement learning is also useful in media and advertising whereby the technology is used for various benefits depending on the desired outcome. One of the organizations that have benefited from integrating the technology is Microsoft which launched the Decision Service launched on Azure. The internal system is responsible for executing various processes that revolve around advertising and recommendation. The system is designed to deal with matters like debugging, weak monitoring, feedback bias and loops, environmental changes and collection of distributed data among others. The technology has also proven to be useful in display advertising online whereby it is used in optimizing cross-channel marketing and bidding systems that utilize real time notion [29].

5.5 DIALOG SYSTEMS, SPEECH AND TEXT

Companies rely on their ability to understand their customers through collection of data which is analyzed and deductions made. It is essential that these companies are able to isolate unstructured texts. For instance, Al researchers have conducted an investigation that led to creation of summaries from texts that were collected from customers.

The RL technique enabled the researchers to create abstracts from the texts thereby enhancing the data mining process. Reinforcement learning has proved useful in allowing chatbots which are dialogue systems and the information is used to facilitate learning of the systems. Learning the dialogue systems is useful in the determination of how the same can be improved over time. Improving dialogue systems is important for VC investments as well as the field of research both of which rely on collection of data [30].

5.6 MACHINE LEARNING AND DATA SCIENCE

Machine learning is not a new technology and its development has made it easier for researchers to use libraries responsible for machine learning. However, data scientists are still faced with the challenge of decision-making whereby it still remains a challenge selecting the most appropriate model. Reinforcement learning opens up an entire broad field of research through the deep learning process [31].

RL opens up opportunities that can enable data scientists to identify neural networks. Apart from identifying the networks, the RL techniques also makes it possible to find ways of tuning the networks. Suggestions have been made to choose reinforcement learning as the primary method when it comes to designing of architecture of the neural networks nature.

Artificial neural network architectures are divided into two groups as feedforward and back propagation based on the directions of the links between the neurons. In the feedforward networks, the signals go from input layer to output layer on the one-way links. At the same time, in the feedforward networks, the output values of the cells in one layer are transmitted to the following layers as the inputs on the weights. The input layer sends the input to the hidden layer without making any change. Once this information is processed on the hidden and the output layer, its output on the network is determined. Multilayer sensors and learning vector quantity can be examples of feedforward artificial networks. The most important characteristics of the back propagation artificial neural networks is that output value of at least one cell is given to itself or another cell as an input value. The back propagation can be processed on a retardation unit as well as the cells in one layer or among the cells on other layers. Because of this feature, the back propagation artificial neural networks show a dynamic behaviour [12]. Those networks got their name by their function that they can organize the weights backwards in order to minimize the errors occurred on the output layer.

The first researches on artificial intelligence started with the single layer artificial neural networks. The most important feature of the network is classification of the problems which can be selected linear as a layer. After the inputs in the problem are multiplied by the weights and added, the calculated values are classified according to their threshold value as high or low. The groups are shown like -1 and 1 or 0 and 1. During the learning process, both the weights and the weights of threshold value are updated. The output value of the threshold value is 1. Since the single layer artificial neural networks are inefficient for the nonlinear problems, multilayer artificial neural networks have been developed. Today, mostly used artificial neural network is the multilayer artificial neural network. Multilayer networks emerged during the studies to solve the XOR problems. Multilayer networks have 3 layers. Input Layer: This layer gets the information from the outer world, but there is no process on this layer. Interlayers: The information from the input layer is processed on this layer. Mostly one interlayer can be adequate for the solution of the problem. However, if the relations between input and output are not linear or there are some complications, more than one layer can be used. Output Layer: The information from the interlayer is processed on this layer and the outputs which correspond the input are detected. In training the multilayer artificial neural networks, the ‘delta rule’ is used. As the multilayer networks use supervised learning methods, both the inputs and the outputs which correspond the inputs are shown to the network. According to the learning rule, the error margin between the outputs and the expected outputs are distributed to the network in order to minimize the error margin.

There are recommendations to make the architectures easily accessible. Some of these suggestions are Net2Net operations and MetaQNN which is the work of MIT. Another instance that RL has been utilized is the case of Google that has enhanced language modeling and computer vision by using its state-of-the-art neural networks architectures that has been produced using RL

5.7 REINFORCEMENT LEARNING IN ARTIFICIAL INTELLIGENCE

Artificial learning requires an understanding of machine learning it follows that the idea is to place an agent or learner in an environment where the agent can learn from changes in the environment and be able to make decisions that will lead to the most profitable rewards. In artificial learning or machine learning, it is unavoidable that understanding these technologies require the knowledge of three algorithms namely reinforcement learning, unsupervised learning and supervised learning [32].

Supervised learning is the type of algorithm whereby data that is labeled is fed into the data. In real sense, inputting labeled data is informing or explaining to the machine what it is seeing. In unsupervised learning, the machine is not fed labeled data as in the other case. The researcher does not have any specific target or goal in mind (figure 10). Rather, the machine or agent has the sole responsibility of uncovering structural aspects. These aspects are uncovered through the learning process.

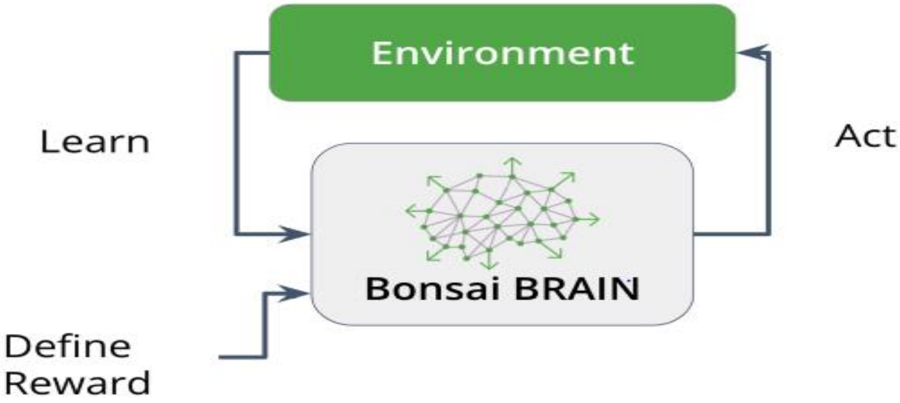


Figure 5.2: Artificial learning process Bonsai brain

The other type of artificial intelligence is reinforcement learning. In this case, the machine is fed with data or information but care is taken to ensure that the information fed to the machine is incorrect. It is a form of supervised learning only that the information fed to the agent is not correct. After the data has been fed to the agent, it has the responsibility of learning from the environment through a series of actions that lead to rewards. The agent is required to learn from every action so that the next action will result in maximum attainable rewards.

5.8 DEEP REINFORCEMENT LEARNING

Deep reinforcement learning is a type of artificial intelligence and therefore it is clear that RL plays a significant part when it comes to AI. As discussed in the previous section, this type of learning is a kind of supervised learning whereby partial information is fed to the agent. Recently, significant strides have been made in the AI field where deep learning has found favor in companies that have to deal with many complex problems that require formidable solutions [33]. Even though, DL is considered as a subset of machine learning algorithms and shares many common aspects with it, it has some differences that diverge it from the traditional machine learning approach. The most significant difference is that machine learning algorithms require the features to be picked manually to feed the algorithm whereas in DL these features are detected automatically by the algorithm. Furthermore, DL adopts a hierarchical learning methodology. After high-scale data sources and more advanced hardware/software opportunities required for DL have been available to public, DL has increasingly been used for medical image analysis, as well. In this section, RNN and CNN, mostly used DL architectures for medical image analysis, will be explained shortly.

The methods mentioned so far are all supervised learning methods. Another major machine learning approach, unsupervised learning, is also an active research topic in the context of medical image analysis by use of DL. For instance, Plis et al. [45], employed Deep Belief Networks (DBN) to extract useful features from MRI images of patients having Huntington disease and schizophrenia. Likewise, Suk et al. [46] used Restricted Boltzmann Machines (RBM) to reveal relationships among different parts of the brain in fMRI images so that patients with Mild Cognitive Impairment (MCI) could be detected.

Deep learning is becoming popular because of its nature whereby the learner is responsible for undertaking a series of actions. Every action taken by the agent yields outcomes known as rewards. The objective in deep learning is to enable the agent to learn from the environment so that it can be able to make better decisions in the next task. This process is useful because basically, it trains the agent so that the machine can learn and prepare for unexpected outcomes.

Rather than being sequential, images are generally considered as a kind of data containing inner correlations and spatial information about pixels. Therefore studies that take biomedical images as non-sequential data more often utilized DNN or CNN instead of RNN [19]. By using improved versions of RNN, researchers have recently paid increased attention to RNN for the purpose of image based recognition. For instance, Multidimensional Recurrent Neural Network (MDRNN) [26] has been applied to three dimensional images. Furthermore, Stollenga et al. [27] implemented a MDRNN based solution to segment neural structures in MRI and three dimensional electron microscope images.

Single Layer and Multilayer Artificial Neural Networks is the first researches on artificial intelligence started with the single layer artificial neural networks. The most important feature of the network is classification of the problems which can be selected linear as a layer. After the inputs in the problem are multiplied by the weights and added, the calculated values are classified according to their threshold value as high or low. The groups are shown like -1 and 1 or 0 and 1. During the learning process, both the weights and the weights of threshold value are updated. The output value of the threshold value is 1. Since the single layer artificial neural networks are inefficient for the nonlinear problems, multilayer artificial neural networks have been developed. Today, mostly used artificial neural network is the multilayer artificial neural network. Multilayer networks emerged during the studies to solve the XOR problems. Multilayer networks have 3 layers. Input Layer: This layer gets the information from the outer world, but there is no process on this layer. Interlayers: The information from the input layer is processed on this layer. Mostly one interlayer can be adequate for the solution of the problem. However, if the relations between input and output are not linear or there are some complications, more than one layer can be used. Output Layer: The information from the interlayer is processed on this layer and the outputs which correspond the input are detected. In training the multilayer artificial neural networks, the 'delta rule' is used. As the multilayer networks use supervised learning methods, both the inputs and the outputs which correspond the inputs are shown to the network. According to the learning rule, the error margin between the outputs and the expected outputs are distributed to the network in order to minimize the error margin .

Feedforward and back propagation artificial neural networks Artificial neural network architectures are divided into two groups as feedforward and back propagation based on the directions of the links between the neurons. In the feedforward networks, the signals go from input layer to output layer on the one-way links. At the same time, in the feedforward networks, the output values of the cells in one layer are transmitted to the following layers as the inputs on the weights. The input layer sends the input to the hidden layer without making any change. Once this information is processed on the hidden and the output layer, its output on the network is determined. Multilayer sensors and learning vector quantity can be examples of feedforward artificial networks. The most important characteristics of the back propagation artificial neural networks is that output value of at least one cell is given to itself or another cell as an input value. The back propagation can be processed on a retardation unit as well as the cells in one layer or among the cells on other layers. Because of this feature, the back propagation artificial neural networks show a dynamic behaviour [12]. Those networks got their name by their function that they can organize the weights backwards in order to minimize the errors occurred on the output layer (Hamzaçebi ve Kutay, 2004).

5.9 PERCEPTRON

Global Definition : A perceptron is basic part of a neural network. It represents a single neuron of a human brain and is used for binary Classifiers. It can be trained to do some basic binary classification and this is how a basic perceptron looks like

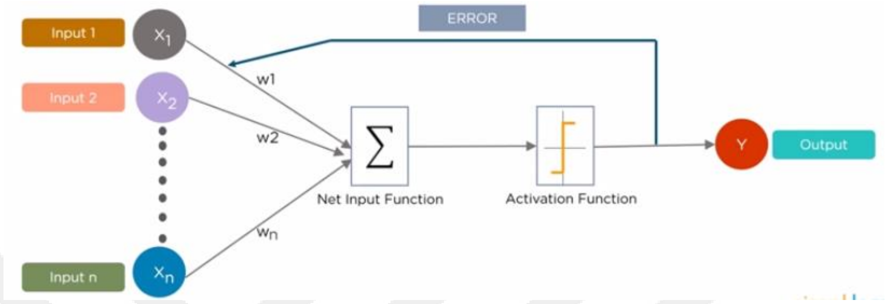


Figure 5.3: Perceptron Learning processes [12]

Referring to (figure 11)We have inputs **X_1 X_{2} … X_n** and there is a summation function and then there is what is known as an activation function (Define input of a node as output for the next neuron or node) and based on this input what is known as the weighted sum the activation function either gets gives an output like a 0 or a 1 so we say the neuron is either activated or not so that's the way it works so you get the inputs these inputs are each of the inputs are multiplied by a weight and there is a bias that gets added and that whole thing is fed to an activation function and then that results in an output and if the output is correct it is accepted if it is wrong if there is an error then that error is fed back and the neuron then adjusts the weights and biases to give a new output and so on and so forth so that's what is known as the training process of a neural network.

Neural networks serve many purposes and they have been utilized to solve several problems and this is the same mechanism that is used in perceptron. The functioning of perceptron is based on the functioning of neural networks. This study aims to deliver a view of DL applications in the field of medical imaging analysis. To this end, relevant studies in the literature, with a focus on the most recent ones, are considered and summarized through the paper. Through this analysis, the advantages of and the problem related to the DL approach when used on medical images will also be discussed as well. Therefore, the study aims to provide a general picture of DL application in the field that covers general trends, advantages, disadvantages, problematic aspects of these applications so that researchers who want to conduct a study in the field may benefit.

They are often referred to as Artificial Neural Networks (ANN) and therefore they work similarly to the central nervous system of humans. Some of the functions of neural networking is classification of data, detecting novelties or anomalies, processing signals and approximating target functions among others [34]. Neural networks work in unison and the result is that complex behaviors are exhibited. The unique fact that makes the functionality of neural networks interesting to machine learning is that they have the ability to learn. The learning takes place in the three learning processes that have been discussed.

There are three ways that a neural network works namely weighing, summing up and activating. During the weighing process, data is inputted in the form of signals. The signal is multiplied using weighted value. It implies that if a neuron has four inputs, then there are four weights that are four weights that can be assigned individually to the inputs. During the learning process, the neural network is responsible for making decisions depending on the errors that have been made in the previous test. After the signals have been weighted, they are weighted to one value which constitutes the summing process. During summing up, a *bias* value is added to the summed up values and the value is not constant as it is changed during the learning process. When the process is initiated all neurons have weights that are randomized and the biases are also randomized. After the learning process has taken place, all weights and biases are adjusted so that they produce an output that is desired. After the two processes have been successfully completed, the activation process takes place. The calculations that have been done by the neurons are then turned into an output signal.

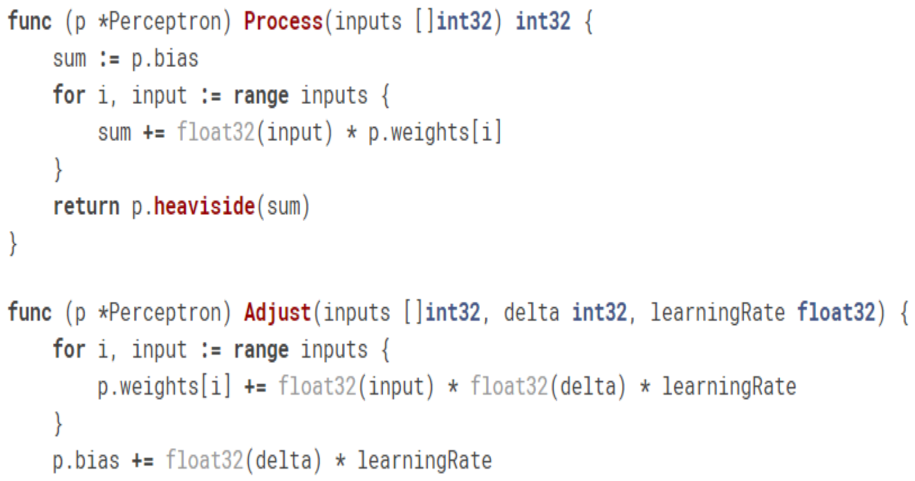
The activation process has two possibilities that comprise a simple binary function that is known as Heaviside Step Function. The function produces zero in the input is negative and if it is positive it produces a result of 1 if the input is either one or zero. One perceptron is capable of solving many problems [35].

For instance, when a vector is considered as a point's coordinates, if the vector has *n* elements, then the dimensional space would be *n*. to understand the mechanism better, one can consider the function as a plane paper that has two sets of data as illustrated in the figure below.

Machine Learning Application Areas The previous section includes the theoretical background of the machine learning algorithms. In this section, information about the areas and studies in which the machine learning are used nowadays will be given. Today, the use of machine learning has increased considerably. Although it is though that it can only be done in large studies, many people face machine learning in their daily life.

These studies and applications are as follows: Education: One of the most important application fields is education in which there have been some studies in order to identify and increase success recently. Despite the projects made in the field of education in recent years, the desired success has not been achieved. There are a lot of factors that influence this failure. However, it has not been determined which factor has more influence on this failure. In this context, by a questionnaire applied to secondary school students, the successes of the students in the lessons were predicted by machine learning models, which resulted with success. Similarly, there are some studies in order to determine the proficiencies of students in higher education.

The model developed by Fukushima in 1980 to mimic human visual system (Neocognitron) can be considered as a simple version of CNN [21]. LeNet, a more successful CNN model developed by Le Cun et al. [22], was utilized to recognize handwritten digits with an architecture made up of 1 input, 3 hidden and 1 output layers.



```
func (p *Perceptron) Process(inputs []int32) int32 {
    sum := p.bias
    for i, input := range inputs {
        sum += float32(input) * p.weights[i]
    }
    return p.heaviside(sum)
}

func (p *Perceptron) Adjust(inputs []int32, delta int32, learningRate float32) {
    for i, input := range inputs {
        p.weights[i] += float32(input) * float32(delta) * learningRate
    }
    p.bias += float32(delta) * learningRate
```

Figure 5.4 : Perceptron process

It is necessary to consider all situations apart from the condition where the line is vertical because this enables having a linear function that can be represented with a linear function equation Eq.5.1 The way this optimization algorithm works(figure 12) is that each training instance is shown to the model one at a time. The model makes a prediction for a training instance, the error is calculated and the model is updated in order to reduce the error for the next prediction.

This procedure can be used to find the set of weights in a model that result in the smallest error for the model on the training data.

$$F(x) = ax + b \tag{5.1}$$

İn Eq.5.1 The b value represents the offset while the *a* value represents the gradient of the line. It represents the steepness of the line. It becomes easy to define whether a certain point is below or above the line by looking at its coordinates. When the value of *y* is larger than the result of f(x), the point is above the line and vice versa. The figure below indicates the situation.

6. CONCLUSION

Reinforcement learning is a type of machine learning that involves an agent that is placed in an environment and expected to learn from the factors that it encounters. There are a series of actions that the agent is involved in and it is expected to make decision on which actions that it will take. The results obtained from an action are used to make the decision on the next assignment. The overall intention is that the agent learns from the decisions that it makes so that the rewards obtained from the next action will result in higher rewards. The agent identifies decisions/actions that do not produce maximum results and rejects them. Only actions that lead to the greatest rewards are chosen. There are two main simulation processes namely Pybullet and Mujoco. Although simulation processes take place in artificial environments, the objective is apply the same criteria to real-life situations. This enables finding solutions to complex problems and therefore reinforcement learning can be applied in many situations. i will focus also on short comparison between Augmented random search an other standard or model of AI algorithms . So first factor is exploration. ARS performs the exploration in the policy space more than other AI which are based on exploration in the action space . the Perceptoron we wait until the agent gets to the end of the episode then we get a reward which we calculate based on adjusted weights.

REFERENCES

[1] R. Sutton and A. Barto, Reinforcement learning. Cambridge, Mass.: MIT Press, 1998.

[2] M. Mozer, M. Jordan and T. Petsche, Advances in neural information processing systems. Cambridge, Mass.: MIT Press, 1997.

[3] P. Nicolas, Scala for Machine Learning. Packt Publishing, 2014, p. 449.

[4] H. Hasselt, Insights in reinforcement learning. Utrecht: Universiteit Utrecht, 2011, p. 10.

[5] S. Whiteson, Adaptive Representations for Reinforcement Learning. Berlin: Heidelberg: Springer-Verlag, 2010, p. 7.

[6] C. Szepesvári, Algorithms for reinforcement learning. San Rafael, Calif. (1537 Fourth Street, San Rafael, CA 94901 USA): Morgan & Claypool, 2010, p. 30.

[7] P. Vrancx, Decentralised reinforcement learning in Markov Games. Brussel: VUBPress, 2010, p. 33.

[8] J. Spall, Introduction to stochastic search and optimization. Hoboken, N.J.: Wiley-Interscience, 2005.

[9] L. Madden and N. Samani, Professional augmented reality browsers for smartphones. Chichester, West Sussex, U.K.: John Wiley & Sons Ltd, 2011.

[10] A. Phadnis, "Augmented Random Search —One of the Best RL Algs + What I Built", Hacker Noon, 2018. [Online]. Available: https://hackernoon.com/augmented-random-search-one-of-the-best-rl-algs-what-i-built-e0e3e765808a. [Accessed: 19- Dec-2018].

[11] R. Hammoud, *Augmented Vision Perception in Infrared*. London: Springer, 2009.

[12] D. Schmorrow and C. Fidopiastis, *Foundations of Augmented Cognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982.

[13] K. Gurney, *An Introduction to Neural Networks*. CRC Press, 2003.

[14] A. Waibel and K. Lee, *Readings in speech recognition*. San Mateo, Calif.: Morgan Kaufmann Publishers, 1990.

[15] M. Watt et al., *Multithreading for Visual Effects*. CRC Press, 2019.

[16] R. Lo and W. Lo, *OpenGL data visualization cookbook*. Packt Publishing Ltd, 2015.

[17] Khronos Group, *COLLADA 1.4 Quick Reference*. Lulu.com, 2014.

[18] S. Saito, W. Yang and R. Shanmugamani, *Python reinforcement learning projects*. Birmingham: Packt Publishing Ltd, 2018.

[19] S. Kakani and A. Kakani, *Material science*. New Delhi: New Age International, 2004.

[20] R. Sutton and A. Barto, *Reinforcement learning*. MIT Press, 2018.

[21] A. Koubâa, *Robot Operating System (ROS)*. Cham: Springer, 2016.

[22] R. Featherstone, *Rigid Body Dynamics Algorithms*. Boston, MA: Springer Science+Business Media, LLC, 2008.

[23] R. Featherstone, *Rigid Body Dynamics Algorithms*. Boston, MA: Springer Science+Business Media, LLC, 2008.

[24] E. Feinberg and A. Shwartz, *Handbook of Markov decision processes*. Boston: Kluwer Academic Publishers, 2002.

[25] S. Dutta, *Reinforcement Learning with TensorFlow*. Birmingham: Packt Publishing, 2018.

[26] M. Kosorok and E. Moodie, *Adaptive treatment strategies in practice*. Philadelphia: Society for Industrial and Applied Mathematics, 2016.

[27] R. Sutton and A. Barto, *Reinforcement learning*. 2018.

[28] V. Rieser and O. Lemon, *Reinforcement Learning for Adaptive Dialogue Systems*. 2011.

[29] P. Dangeti, *Statistics for Machine Learning*. 2011.

[30] V. Rieser and O. Lemon, *Reinforcement Learning for Adaptive Dialogue Systems*. Springer Science & Business Media, 2011.

[31] C. Szepesvári, *Algorithms for reinforcement learning*. [San Rafael, CA]: Morgan & Claypool Publishers, 2010.

[32] S. Sanner and M. Hutter, *Recent advances in reinforcement Learning*. Berlin: Springer, 2012.

[33] S. Kwon, *Artificial neural networks*. New York: Nova Science Publishers, 2011.

[34] K. Seeler, *System Dynamics*. New York, NY: Springer, 2014.

[35] J. Wang, G. Yen and M. Polycarpou, *Advances in neural networks-- ISNN 2012*. Berlin: Springer, 2012.

[36] H.Mao , M.Alizadeh , I.Menache, S.Kandula Resource Management with Reinforcement Learning- Massachusetts Institute of Technology