International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN(Online): 2347-5552, Volume-12, Issue-2, March 2024
https://doi.org/10.55524/ijircst.2024.12.2.21
Article ID IJIRD-1282, Pages 119-125
www.ijircst.org

Reinforcement Learning: A Comprehensive Overview

Priyanka Makkar

Assistant Professor, Department of Computer Science & Engineering, Amity University, Gurugram, Haryana, India

Correspondence should be addressed to Priyanka Makkar; pmakkar@ggn.amity.edu

Received 5 March 2024; Revised 18 March 2024; Accepted 27 March 2024

ABSTRACT- Machine Learning is one of the most essential parts of Artificial Intelligence. Machine learning now exists as an important innovation and has a sufficient number of uses. Reinforcement Learning is one of the largest Machine Learning applications that enable machines and software agents to work more precisely and resolve behaviors within a specific context in order to maximize their performance. The self-improvement feature, web-based learning, and minimal effort of Strengthening Learning helped the machines become smart agents in basic technology. With the development of robust and effective algorithms, there is still a lot of work to be done. Therefore, the main purpose of this study is to provide Confirmation Learning reviews and applications using various algorithms from a machine learning perspective.

KEYWORDS- Artificial Intelligence, Machine Learning, Reinforcement Learning, Algorithms, Applications

I. INTRODUCTION

Technology has played an important role in our daily lives since the last few years and somehow, we all rely on them for greater profit and comfort. We live in the era of Information Technology i.e., the era of intellectual innovation in which everyone is associated with this renaming either intentionally or unintentionally. Artificial Intelligence (AI) is one of the most popular inventions in recent years as the demand for Artificial Intelligence is growing day by day. Artificial Intelligence is a dynamic research sector that is often used in parallel with machine learning to enable machines to work as human beings and have the ability to react to certain processes. Machine learning is an advanced approach in the field of Artificial Intelligence. Science in which computers act as human beings without being clearly customizable. Machines are taught mechanical learning techniques to deal with information accurately. To interpret machine learning, data sets or machine learning algorithms are used to view that data pattern and information. The machine learning algorithms are broadly divided into four categories namely Supervised Reading, Unchecked Reading, Enhanced Reading and Recommended Reading Program[4][7][6]. Supervised learning is a learning activity for a machine that articulates the power of knowledge. This process includes a database of divided information for preparation information and a model database for testing. The modified information has yield characteristics that should be expected or noted and the test model uses hidden experimental information to achieve model validity. In other words, the agent is used to measure the objective values of each information and store everything in memory for further reference. This method is also used to detect bugs and to correct the output errors you want. Some of the most commonly used surveillance algorithms are Decision Tree, Native Bayes and Vector Support Machines.

In Unsupervised Learning, ideas are not received by the agent. The agent needs to learn by himself about receiving and delivering a particular contribution as a set of instructions. Here collections are used to represent information effectively as groups have reduced the standard setting. This is why this learning process is mainly used to integrate and highlight the reduction process. The two basic algorithms for combining and reducing size are: K-Means Clustering and Key Component Analysis.

In the reality of learning the choice depends on the movement to be made as a result. Thus, Enhanced Learning is based on literature between the agent performing the task and its setting providing a positive or negative response. The goal is achieved through a combination of testing and context. To become a powerful machine learning system, strengthening learning joins supervised learning fields and flexible programs.

Reinforcement Learning is used successfully in many fields such as Game theory, performance research, Robot, Economics, information theory, control theory, simulation- based improvements, Mathematics and a lot more. In the Recommender program to learn clients especially internet users who can customize their sites such as client and requirement, there are two ways users can capture their information which is a content-based recommendation and a collaborative suggestion. It allows users to recreate their informative, intelligent and novel recommendations. As discussed above strengthening learning is seen as another form of machine learning algorithms as well as unlearned learning supervised learning and compulsory learning set of programs. Recent research suggests that Reinforcement Learning is an agent-based Artificial Intelligence learning algorithm and can be used in a variety of programs. In addition, this paper also highlights Reinforcement Learning by using a few algorithms based on machine learning perspective.

II. REINFORCEMENT LEARNING

Reinforcement Learning is a machine-based machine learning method in which the agent learns local behaviour by doing actions and seeing the results of actions. For every good deed, the agent receives a positive response, and for every bad deed, the agent receives a negative response or penalty.

In Reinforcement Learning, the agent automatically reads using answers without any labeled data, unlike supervised reading. Since there is no labeled data, the agent is therefore obliged to learn only from his own experience.

Reinforcement Learning (RL) solves a kind of problem where decisions are sequenced, and the goal is long-term, like playing a game, robots, etc. The agent interacts with the environment and self-evaluates. The main goal of an agent in strengthening learning is to improve performance by earning high quality rewards.

The agent learns the process of beating and trying, and based on experience, learns to do the job better. Therefore, we can say that "Strengthening learning is a form of machine learning in which the intelligent agent (computer program) communicates with nature and learns to do within that." The way a robotic dog learns to move its arms is an example of learning with Confidence[6].

It is an integral part of artificial intelligence, and every AI agent works with the concept of enhanced learning.

Example: Suppose an AI agent is present in a maze, and his goal is to find a diamond. An agent deals with nature by performing certain actions, and based on those actions, the agent's status is changed, and he or she receives a reward or compensation in response.

The agent continues to do these three things (take action, change / stay in the same position, and get feedback), and by doing these actions, he reads and explores the environment. The agent learns which actions lead to positive feedback or rewards and which actions lead to negative feedback. As a good reward, the agent gets a good point, and as a penalty, gets a bad point.

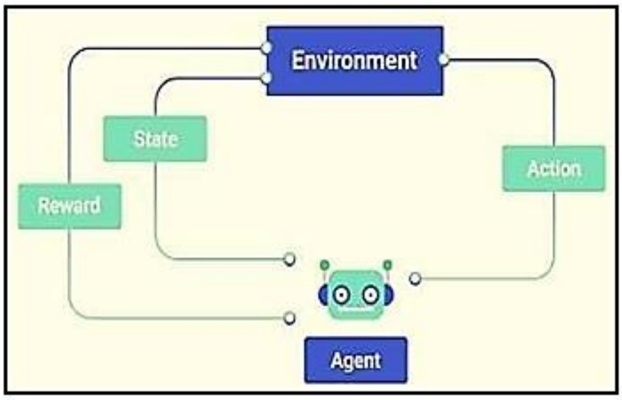

Figure 1: Structure of Reinforcement Learning [11]

A. Terms Used in Reinforcement Learning

- **Agent ():** A business that is able to see / explore the environment and do something about it.
- **Environment ():** The state in which an agent is or is surrounded by. In RL, we take the place of stochastic, which means it is not naturally programmed.
- **Action ():** Actions are actions taken by an agent within the environment.
- **Country ():** The state is a state that is restored by nature after each action taken by the agent.
- **Reward ():** Response returned to the agent from the location to check the agent's action.
- **Policy ():** Policy is the strategy used by the agent in the next action based on the current situation.
- **Amount ():** Long-term reimbursement is expected with the discounted and opposite short-term reward.
- **Q-value ():** Very similar to value, but takes one additional parameter as the current action (a).

B. Key Features of Reinforcement Learning

- In Reinforcement Learning, the agent is not educated about nature and what steps to take.
- It is based on beatings and the temptation process.
- The agent takes the next action and the changes mean according to the response to the previous action.
- The agent may receive a delayed prize.
- Nature is stochastic, and the agent needs to test it in order to achieve it in order to get the best rewards.

C. Approaches to implement RL

In particular, there are three ways to implement Reinforcement Learning (RL) in Machine Learning, namely:

- **Based on Price:** The value-based approach will now find high value work, which is the highest value in the area under any policy. Therefore, the agent expects a long-term return in any case under policy π.
- **Policy Based:** The policy-based approach is to find the right policy for the many rewards for the future without having to spend a fortune on it. In this way, the agent is trying to apply the policy that the action taken in each step helps to increase future rewards.
- **Model Based:** In a model-based approach, the visual model is created by nature, and the agent explores the area to read it. There is no specific solution or algorithm for this method because the model representation varies for different locations.

D. Types of Reinforcement Learning

There are basically two types of Reinforcement Learning:

- **Positive Reinforcement:** A good reading of reinforcement means adding something toincrease the likelihood that the expectedbehaviour will recur. It has a positive effect on the behaviour of the agent and increases morale.
- **Negative Reinforcement:** Learning negative reinforcement is the opposite of positivereinforcement as it increases the likelihood that positive behaviour will recur by avoiding a negative situation. It can work better than positive reinforcement depending on the situation and behaviour, but it provides reinforcement to meet the minimum behaviour.

E. Steps for Reinforcement Learning Problem

These steps apply to the Pre-Strengthening Learning Problems:

- **Understanding the problem:** Learning to tighten is not really necessary for every problem. There shouldbe an assessment of each problem before installing a learning algorithm to validate the following attributes are considered:
  - evaluation and error assessment method

- delayed rewards
- can be indicated as MDP
- to assess whether the problem is a controlled problem or not.

- **A Simulated Environment:** Before using Reinforcement Learning Algorithms, it is necessary to count the repetition numbers. Successful representation of real-world objects requires a simulated environment.
- **Markov Decision Process (MDP):** In every case the problem-solving process should be done following these steps. First the problem has to be done with MDP and then design the region, the site, the award work etc. The agent will do what is to becompensated under the requirements.
- **Algorithms:** There are various Reinforcement Learning algorithms available and used to find the best policy or familiarity with value work.

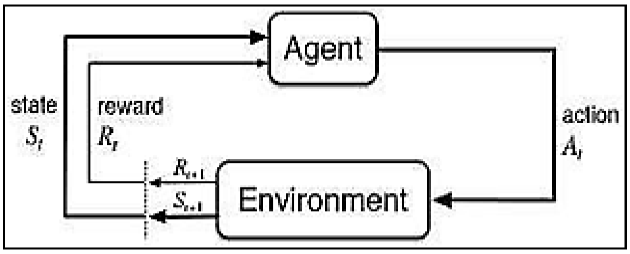

Figure 2: The agent-environment interaction in MDP [10]

III. REINFORCEMENT LEARNING ALGORITHMS

Agents are an important perspective in reinforcement learning, also known as decision makers or learners. Anything outside the agent is considered as the agent's domain. By using trial and error interactions, this area of the reinforcement learning framework becomes a powerful agent for mapping all situations to activities[1][6]. These activities are performed using single-agent and multi- agent frameworks with different characteristics. The multi-agent framework also uses other coordinating measures that destabilize the environment and violate the Markov property that traditional single-agents rely on. Below are various reinforcement learning algorithms used in multi-agent systems.

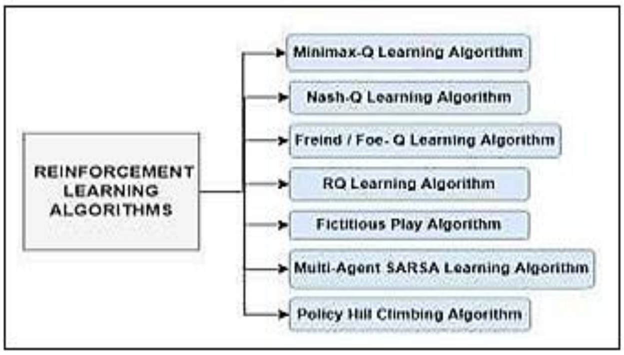

Figure 3: Various Reinforcement Learning Algorithms

A. Minimax-Q Learning Algorithm

Minimax-Q learning algorithm for zero-sum game situations where the learning player extends its payments to any situation. The player's enthusiasm for the game is the opposite. Initially, it has great potential to support reading calculation. In this algorithm the player tries to elevate his or her normal motive even with the worst possible decision of the enemy's possible mission.

B. Nash-Q Learning Algorithm

Hu and Wellman, in 2003 proposed a zero-sum game-based Minimax-Q learning algorithm to integrate common games and build a Nash-Q learning algorithm for learning strategies that strengthens multiple agents[1][6]. In order to spread Q- learning across the various multi-agency regions, there are several joint activities of participating agents instead of individual actions that need to be considered. Because of this huge difference between one agent and many learning agents that reinforces this teaching process it needs to keep up with Q values for both the student himself and the other players. The main reason for finding Nash equity in each province is to find Nash balancing methods for reviewing Q values. The Nash-Q learning algorithm is used only after Nash Q's value is defined. This defined value is considered the average value of the prize pool in which all employees are required to follow the projected Nash rating policies going forward. Continuing Hu and Wellman, in 2003 also highlighted that this learning algorithm in a multiplayer mode meets Nash's rating strategies under certain conditions and additional expectations in payment structures.

C. RQ-Learning Algorithm

RQ - Learning algorithm is designed to deal with major search space problems. In this r-state algorithm and the set of actions must be clearly defined in the first paragraph. The r- state starts with a set of first order relationships like front goal, left team robot, etc. while r-action is preceded by a collection of pre- and post-conditions with regular action. In order to properly describe the action of r, there is a condition to be satisfied, the condition is that —if the action of r is appropriate for a particular r-state event, it must be appropriate for all such occurrences. r-condition. This algorithm is suitable for large search space problems which can be very difficult to define r-state and r-set set correctly, especially in the case of insufficient information. In addition, Morales, 2003 states that in the r-state area there is no guarantee that a number of well-defined actions are acceptable to obtain the best planning of archaeological works and problem-solving approaches.

D. Fictitious Play Algorithm

In Nash-equilibrium-based learning there is a problem finding Nash equilibria results and using a fictional play algorithm and providing an alternative management system with a multi-agent framework. Cao, 1997 and Suematsu, 2002 pointed out that in this algorithm some strategies represent the distribution of tests and players are required to keep their Q values associated with combined activities and measured by their conviction. The fiction play algorithm converts different Q readings each to different players' standings and non-stop player modes. modeling or integrated games where players can learn their Q values from their shared activities known as Joint Action Learner[11-14].

E. Multi-Agent SARSA Learning Algorithm

Nash-Q and Minimax-Q learning algorithms are actually known as off-strategy Reinforcement Learning algorithms because these algorithms change the main operator of a specific Q algorithm for their outstanding response and that the reaction is known as the Nash equity policy[2]. In Teaching Strengthening a continuous non-functional learning algorithm attempts to combine the best Q values of a good strategy no matter which strategy should be used. According to Sutton, the 1998 SARSA algorithm is one of the Policy Reinforcement Learning algorithms that attempts to convert to the appropriate Q values of the currently used strategy[15-16].

F. Friend-or-Foe Q-Learning (FFQ) algorithm

In each FFQ algorithm each draft agent is known as a 'friend' or 'enemy'. Here equality can be distinguished by way of connection or by way of argument. The FFQ learning algorithm can provide a strong assurance of integration compared to the Nash-Q learning algorithm.

G. Policy Hill Climbing (PHC) Algorithm

This algorithm updates Q values in the same way as a fictionalplay algorithm, but maintains a mixed policy also known as stochastic policy by performing mountain climbing in the area of these policies. Bowling and Velso, 2002 proposed the PHC WOLF (Win or Learn Fast) algorithm by adopting the concept of Win or Learn Fast and using a flexible reading level. While using this algorithm it quickly leads to a learning agent if he does not perform effectively and with caution. These changes in levels of learning will be helpful in unifying the inclusion of different agents of evolving strategies.

IV. REINFORCEMENT LEARNING APPLICATIONS

Artificial intelligence methods are widely used in various fields and are becoming more and more popular. However, reinforcement learning approaches still have limited application. The reason for the limited use of reinforcementmethods is that they can be applied where a clear reward can be applied.

For example, algorithms may allocate limited resources to different tasks as long as they have a common goal to be achieved. The goal in this situation is to save time or resources. With this in mind, we can use this method in thefollowing cases.

A. Resource Management in Computer Clusters

It is difficult to design an algorithm that allocates limited resources to different tasks and requires human-generated heuristics. However, reinforcement learning allows you to automatically allocate and schedule computer resources for waiting jobs with the goal of minimizing the slowdown of the average job [6]. The paper "Resource Management by Deep Reinforcement Learning" formulates the state space as the current resource allocation and the resource profile of the job. For the action space, they used a trick to allow the agent to select multiple actions at each time step. The reward was the sum of all the jobs in the system (1 / job duration). Next, they combined the REINFORCE algorithm with the baseline to calculate the policy gradient and found the optimal policy parameter showing the probability distribution of the action tominimize the goal.

B. Traffic Forecasting

With the rapidly increasing number of vehicles on the road, the traffic control seems to be a huge task. To resolve this issue, we can develop and use a machine to address this issue. A machine that overlays information about future traffic conditions on an enhanced traffic flow map. We can also use these frameworks to understand current and future traffic conditions in our area and provide our customers with alternative routing for this data. In the study of "Reinforcement Learning-based Multi-Agent System for Network Traffic Signal Control", researchers sought to design a traffic light controller to solve the congestion problem [10]. However, these methods were only tested ina simulated environment and showed better results than traditional methods, highlighting the potential use of multi-Agent RL in the design of transportation systems. Five agents were placed in the traffic network at five intersections, and one RL agent was placed at the central intersection to control the traffic signal. The state is defined as an 8-dimensional vector, where each element represents the relative traffic flow in each lane. The agent has eight choices, each representing a combination of phases, and the reward function was defined as a reductionin lag compared to the previous time step. The author used DQN to learn the Q value of {state, action} pairs.

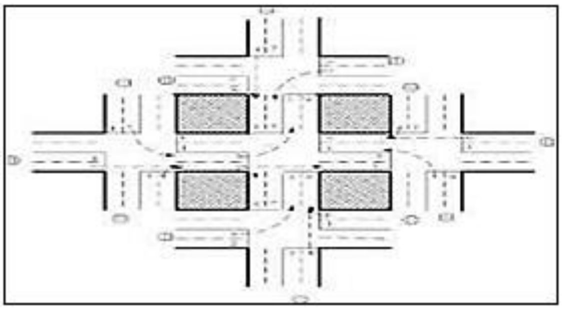

Figure 4: Five intersection traffic network [10]

C. Robotics

Robots can perform amazing tasks under human control, such as medical procedures and household chores. In this area it is impossible to predict if there will be a fully recognizable state. This learning framework cannot predict information about different states that may be similar [6]. Multirobot systems are often used to perform tasks that are difficult for a single robot to perform, especially in the face of vulnerabilities, incomplete data, delegated control, and non- simultaneous computation.

D. Machinery Applications

Reinforcement learning is a type of machine learning calculation that allows programming agents and machines to target ideal behavior in a particular environment and improve performance [10]. These applications cannot be modified. This includes the manufacturing, inventory management, supply management, power systems, and finance sectors.

E. Semantic Annotation of Learning Environments

In today's world, functional learning is becoming more important in all areas of life. It helps to acquire practical knowledge and gives a better understanding of the subject. It is very useful to use semantic calculations as an essential aspect of learning environment-based skills [10]. Simulating real-world situations helps develop useful skills such as decision making, communication, teamwork, and problem solving.

F. Web System Configurations

The web system has over 100 configurable parameters, and the process of setting the parameters requires an experienced operator and numerous trails and error tests. The paper "Reinforcement Learning Approach to Automatic Web System Configuration" shows that the domain was the first to try to perform autonomous parameter reconstruction in a multi-layered web system in a VM-based dynamic environment [10]. The reconstructionprocess can be formulated as a finite MDP. The state space is the system configuration, the action space is the {increase, decrease, maintain} of each parameter, and the reward is defined as the difference between the specified target response time and the measured response time. The author solved the task using a model-free Q-learning algorithm. The author used different techniques, such as policy initialization, instead of possible combinations of RL and neural networks to address the problematic state space and computational complexity, but with pioneering research. It is believed that the path to future research in thisarea has been paved.

G. Stock Market Analysis

The stock market and its ever-changing patterns bring benefits, and a legitimate understanding is essential to stay ahead of the stock market. Machine learning has basically been used to predict money markets[6]. Appropriate algorithms such as support vector machines and reinforcement learning have been proven to track the stock market and increase the value of low-risk investment opportunities. It also integrates market analysis that takes into account the opportunity for the average financial investor investing in stock markets around the world to predict daily stock patterns.

H. Computer Games

Reinforcement learning is very well known these days because it is the mainstream algorithm used to solve various games and sometimes achieve superhuman feats. The trade in gaming have been very well developed in recent years. Artificial intelligence agents are reused to provide players with intelligent gaming knowledge. These agents can take on a variety of tasks, such as a player's opponents, teammates, or other non-player characters. In addition to communicating with human players, games must meet a variety of requirements, including sound and special visuals[9][10].

- **Examples Of Rl In Mobile Games**

As already mentioned, machine learning algorithms have been successfully used in training modern mobile games. Due to its slow process and high resource cost, the most efficient algorithm is one that can adapt well to multiple environments. Q-learning type reinforcement learning is perfectly suited as it requires learning to discredit action spaces with near zero fault tolerance.

Therefore, value-based methods like Q-learning are better suited than probabilistic algorithms. Coming up next there are some examples which describes the implementation of reinforcement learning for mobile video games such as Flappy Bird, Subway Surfers, and Crossy Road[9].

- **Flappy Bird**

Flappy Bird is a Vietnamese mobile video game developed in a short period of time, released in 2013. This is an arcade game in which the player controls a Fabi bird and has to continue to the right and through the set of tubes. The tubes are uniform in size and the brakes are randomly positioned. The player has to touch the screen to make the bird fly. Otherwise, it will automatically crash.

One point is awarded for successfully passing a pair of pipes. The game is over when Fabi hits a pipe or the ground. Generally, players can earn bronze medals at 10 points or more, silver medals at 20 points, and gold medals at 30 points. The highest prize is the Platinum Medal, which can be obtained by earning 40 points. This game can be difficult to play. Fast and accurate reaction times can be frustrating.

Agents must go through the first pipe to receive the initial reward. This task can be challenging for inexperienced players using randomized studies. However, simple and clearvisuals make general politics easy to learn.

Many early stages are possible without subsequent rewards. This means that the agent cannot learn because it does not receive feedback. After going through the first pipe, agents can finally learn how to improve their tactics.

Also, the Flappy new game is prone to misbehavior. For example, performing the evaluation with a random-action epsilon of 0.001 results in a reward of 143 instead of 420.

Reinforcement learning aims to maximize the expected value of the total payoff or the expected return. The Q-learning approach, in which a neural network is used to approximate the Q-function.

According to the test, Flappy Bird was trained at 30FPS witha frame-skip of 2 (15 Steps-Per-Second) for a total of 25M steps. Evaluation Score: Average 420 points, Max 1363 (10 Eval Episodes)[9].

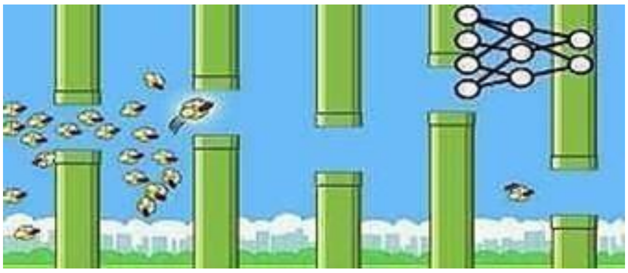

Figure 5: Flappy Bird [9]

- **Subway Surfers**

Subway Surfers is a running game developed and released in Denmark in 2012. In this endless runner video game, the player controls a graffiti artist who draws art on a railroad wall, eventually running across the railroad tracks and trying to escape from an inspector and his dog.

The goal of the game is not to get on the subway and collect as many coins as possible. Trains and other objects create obstacles for runners. To avoid collisions, players can jump over obstacles and ride a hoverboard. The game ends when our character hits an obstacle, gets hit by a train, or gets caught by a swordsman with a dog. However, one can continue playing the game using special keys. Weekly Hunts are bonus events that offer additional rewards and other characters. The player must swipe the screen in different directions to avoid obstacles along the way. Some gestures, such as a quick swipe, serve as special options to speed up our run. The faster the player runs, the more coins he can collect. Also, among the coins, runners can acquire bonus items. In this case, coin counting is not an essential part of growth, but a necessary trigger for the learning process by adding a minus reward of 1 coin to motivate agents to play smoothly. By adding a small negative reward for each action taken, agents don't abuse their bonus items. During training, the game is trained at 30 fps with 4 frame drops (7.5 steps per second) over 25 million steps. The action space is varied with 4 actions (swipe up/down/left/right) and a "noop action". Evaluation Score: 142 averages, 434 max(30 views)[9].

Figure 6: Subway Surfers [9]

- **Crossy Roads**

Crossy Road is an arcade game first released in Australia in 2014. The main goal of this game is to lead the character safely through the obstacles of the endless path. Chicken is the default character. However, the game has other available characters and various vivid environments. Obstacles are also different. One will encounter rivers, busy roads, trains and other themed objects. To control the character, the player moves by tapping the screen and swiping in the desired direction. Time is very important to successfully overcome obstacles. If it fails, the game is over. A reward in the form of points earned for the next forward move. Players can also collect coins that can later be used to purchase bonus items and unlock new characters[9].
In general, players should anticipate additional moves and act on time. Obstacles such as floating logs, moving trains and trucks must also be considered. In this, the player earns coins each time the chicken advances. Evaluation score: Average of 41 points, maximum of 139 points (30 ratings).

Figure 7: Crossy Road [9]

V. CONCLUSION

In today's world of instant gratification, we usually rely on technology to get our work done faster and more efficiently. In the past, machines were used to reduce physical activity, but in modern times with the development of Artificial Intelligence people want to make not only robust but also smart devices so the concept of machine learning now exists and has become a rapidly growing learning environment. Learning to reinforce another type of machine learning that explains its use and research in a wide range of control areas and various decision-making problems that are rarely handled by supervised or supervised learning strategies. Dealing with these types of situations Reinforced Teaching has come into being and has become one of the smartest agents due to its various features such as online learning, self-improvement and very little planning effort.
This paper reviews various reinforcing learning algorithms that can reduce the amount of regional space, improve learning productivity in the initial testing environment and speed up integration. The Multi-Agent Q-learning algorithm is also discussed which is used to build a possible field of practical intelligence to set Q values based on prior knowledge. Here reinforcement learning applications are also discussed that highlight how reinforcement learning has evolved into an impressive technology that is capable of achieving amazing results over a set of challenging problems. In addition to the development of powerful and effective algorithms, efforts have been made to provide specific solutions to these challenges. Therefore, there is a need to solve problems related to learning strategies, decay, measurement methods and integration of real-life bias-related problems through the latest processes in teaching reinforcement and innovative approaches.

REFERENCES

[1] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," J. Artif. Intell. Res., vol. 4, pp. 237-285, 1996.
[2] M. A. Wiering and M. Van Otterlo, "Reinforcement learning," Adaptation, Learning, and Optimization, vol. 12, no. 3, pp. 729, 2012.
[3] C. Szepesvári, "Algorithms for reinforcement learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 4, no. 1, pp. 1-103, 2010.
[4] A. G. Barto and R. S. Sutton, "Reinforcement learning," Handbook of Brain Theory and Neural Networks, pp. 804-809, 1995.

[5] A. Nowé, P. Vrancx, and Y.-M. De Hauwere, "Game theory and multi-agent reinforcement learning," in Reinforcement Learning. Springer, Berlin, Heidelberg, 2012, pp. 441-470.

[6] Y. Li, "Reinforcement learning applications," arXiv preprint arXiv:1908.06973, 2019.

[7] L. M. Gambardella and M. Dorigo, "Ant-Q: A reinforcement learning approach to the traveling salesman problem," in Machine Learning Proceedings 1995. Morgan Kaufmann, 1995, pp. 252-260.

[8] A. Barreto et al., "Successor features for transfer in reinforcement learning," in Advances in neural information processing systems, vol. 30, 2017.

[9] F. Fernández and D. Borrajo, "Two steps reinforcement learning," International Journal of Intelligent Systems, vol. 23, no. 2, pp. 213-245, 2008.

[10] H. R. Tizhoosh, "Reinforcement learning based on actions and opposite actions," in International Conference on Artificial Intelligence and Machine Learning, vol. 414, 2005.

[11] D. Singh, S. Sinha, and V. Thada, "Review of attribute based access control (ABAC) models for cloud computing," in 2021 International Conference on Computational Performance Evaluation (ComPE), IEEE, 2021.

[12] S. Sinha, A. Jain, and S. S. Agrawal, "Empirical analysis of linguistic and paralinguistic information for automatic dialect classification," Artificial Intelligence Review, vol. 51, pp. 647-672, 2019.

[13] A. C. Jaiswal, S. Sinha, and P. Makkar, "Using TRPO to control quadruped gait behaviors," Journal of Integrated Science and Technology, vol. 11, no. 4, pp. 574-574, 2023.

[14] S. Mehrotra, S. Sinha, and S. K. Sharma, "Overview of the Internet of Things and Ubiquitous Computing," in Blockchain Technology for Data Privacy Management. CRC Press, 2021, pp. 1-19.

[15] P. Makkar, S. Sikka, and A. Malhotra, "Optimization of software quality attributes using evolutionary algorithm," International Journal of Electrical and Electronics Research, vol. 10, no. 2, pp. 131-137, 2022.

[16] P. Makkar, S. Sikka, and A. Malhotra, "A Multi-Objective Approach for Software Quality Improvement," Journal of Physics: Conference Series, vol. 1950, no. 1, IOP Publishing, 2021.