A STUDY OF REINFORCEMENT LEARNING APPLICATIONS & ITS ALGORITHMS

Kumar Gourav, Dr. Amanpreet Kaur

Abstract— Machine Learning is an indispensable part of Artificial Intelligence. It is the investigation of projects that makes computer to express like humans. Machine learning has come into existence as an important innovation with its adequate number of uses. Reinforcement Learning is one of the major application of Machine learning that enables machines and software agents to work explicitly and also resolve the conduct within a definite situation to maximize its performance. Due to the aspect of self-improving, web based learning and less programming effort Reinforcement Learning becomes an intelligent agent’s in core technologies. With the advancement of more robust and efficient algorithms, there is still a requirement for more work to be done. Thus the main aim of this study is to provide the review of reinforcement learning and its applications by utilizing various algorithms from machine learning perspective.

Index Terms— Algorithms, Applications, Artificial Intelligence, Machine Learning, Reinforcement Learning, Technology

1 INTRODUCTION

Technology plays a dynamic part in our day to day life since last few years and someway we as a whole are relying upon it in order to get maximum benefit and comfort. We are living in an era of Information Technology i.e. the era of smart innovations where every single individual related with this innovation either intentionally or unintentionally. Artificial Intelligence (AI) is one of the greatest innovation that came into existence in the recent years. As the need of Artificial intelligence is growing day by day. According to Marriot, 2011 the world’s leading research and advisory company Gartner Group predicted that, “By 2020, consumers will manage 85% of their association with the enterprise without interrelating with humans.” Artificial intelligence is the field of study that portray the capability of machine learning just like humans and the ability to react to specific practices. Fuji and Managi, 2018 highlighted that with the rising popularity of Artificial Intelligence, it is frequently utilized reciprocally with machine learning. Machine learning is the developed approach in the field of Artificial Intelligence. It is the science in where computers acts like humans without being explicitly customized. Machines are instructed by machine learning techniques to deal with the information precisely. To interpret the data machine learning, datasets or machine learning algorithms are used for viewing that data pattern and information. Machine Learning algorithms are extensively divided into four categories namely Supervised Learning, Unsupervised Learning, Reinforcement learning and Recommender System Learning.

Supervised learning is the assignment of machine learning that deciphers the capacities from information. This procedure includes the preparation information that is separated into the preparation information and the testing dataset model. The prepared information has the yield factors that are to be anticipated or characterized and the testing model uses the concealed tried information to get to the exactness of the model. In other words, an agent is used to estimate the values of the objective for each information and then stores all into the memory for further reference. This method is additionally used to figure out the bugs and to amend these mistakes for the desired output. Commonly used supervised algorithms are: Decision Tree, Native Bayes and Support Vector Machines.

In Unsupervised learning perceptions are not received by the agent. The agent need to learn at its own by receiving and delivering the specific contribution as the set of instructions. Here clusters are utilized to represents the information in an effective manner since groups diminished the arrangement of measurements. This is the reason this learning procedure is mainly used for bunching and highlights the reduction process. The two fundamentals algorithms for clustering and dimensionality reduction are: K-Means Clustering and Principal Component Analysis.

In Reinforcement learning the choices depends upon the moves to make as the result. So Reinforcement Learning depends on the correspondence between an agent that executes an activity and its environment that gives positive or negative reaction. The objective are accomplished by experimentation collaborations with condition. For becoming powerful machine learning system Reinforcement learning joins the fields of supervised learning and dynamic programming. Reinforcement Learning is successfully applied in numerous disciplines like Game theory, Operation research, Robotics, Economics, Information theory, Control Theory, Simulation based optimization, Statistics and Genetic algorithms.

In Recommender system learning the clients especially the online users that can customize their sites as per the client determinations and prerequisite. There are two manners by which the users can mine their information are Content based recommendation and Collaborative suggestion. It enables the users to recreate their information proficient, insightful and novel recommendations.

As discussed above Reinforcement learning is reflected as one of the type of machine learning algorithms along with unsupervised learning supervised learning and recommender systems learning. Sutton and Barto, 2017 indicated in his study that there is an open option for other paradigms too. The present study reveals that Reinforcement Learning is an agent based Artificial Intelligence machine learning algorithm and can be used in various applications. In addition of that, this paper also highlights the Reinforcement Learning by utilizing few algorithms from Machine learning perception.
2 LITERATURE REVIEW

Arthur Samuel characterizes that the Machine learning is the field of study that enables computers to learn without being explicitly programmed. According to Anderson, 1986 Machine learning is related to the frameworks that consequently improve their performance. Marsland, 2015 briefs that in machine learning the machines learn themselves to tackle explicit issues. As highlighted by Musumeci et al, 2018 Machine Learning is one of the most suitable way to handle the approach that performs the analysis of network data and fault management. As indicated by Lewis et al, 2008 that in the field of artificial intelligence reinforcement learning can take care of issues ideally by cooperating with its condition and furthermore by changing its control arrangements. Busoniu et al, 2009 clarifies that reinforcement learning is utilized to locate the best arrangement to expand the reward. Primarily it got from the instrument of regular learning and furthermore dependent on incentives and penalties on the environment. Flore, 2015 stated that Reinforcement learning is the problem that is faced by an agent and that has to be learned by trial and error interfaces through the dynamic environment. According to Sutton, 1992 Reinforcement learning always depend upon delayed result and experimentation. Tiwana et al., 2014 highlighted that reinforcement learning based framework for Quality of Services (QoS) used for the enhancement in Fourth generation (4G) Networks. Hou et al, 2017 proposed a technique that is powerful for taking care of choice issues which are consecutively enhanced. Reinforcement learning is identified with programming that depends on Markov decision process. Olafati, 2006 states that reinforcement learning algorithms are used to represents the unavoidably and social procedure. It uses state activity learning parameters that builds the components of the factors exponentially. Vidhate et al, 2016 stated that reinforcement learning is a methodology that is utilized for the improvement of multi-agent learning and furthermore a structure that have new strategies which demonstrates that simulated results and also acquire more results. Carlucho et al, 2017 proposed that on steady Q-learning procedure for portable robots a versatile PID control. This procedure doesn’t require and earlier knowledge but it can understand the procedure that is differ from the conventional procedure. Hung et al, 2017 also recommended that by applying Q-learning algorithms to a small flocking fixed-wing UAV’s to figure out how to run and it is also simulated that it can fly in the non-stationary stochastic environment.

3 REINFORCEMENT LEARNING

Reinforcement Learning is an auxiliary learning technique an agent looks at the space of possible procedures and gets an input on the results of the alternatives made. These alternatives are found out through experimental communications with a dynamic situation. It is likewise characterized by contrasting the problem with the different controls of study in machine learning. There are two principle approaches for solving reinforcement learning problems. The first methodology is to look at the space of practices in order to discover one that performs well in the environment. Sutton and Barto, 1998 stated in his study that this methodology has been taken to work in the genetic calculations and programming as well as some more novel search strategies. The second methodology is to utilize the factual strategies and dynamic programming techniques to compute the utility of taking activities in conditions of the world.

3.1 Structure of Reinforcement Learning Framework

![Fig 1: Structure of Reinforcement Learning](image)

The basic structure of Reinforcement Learning is as shown in Fig1 where the agent communicates with a situation with the help of sensor data, changing the environment and acquiring a reward for its activity. The states are parameters or highlights that depict the environment. The sensor information dependable states. Being in a given state S, the value function assesses the various moves that can be made. Hence some sort of reward is to be predicted. Reinforcement learning agent detects the environment and learns the ideal approach by taking activities in each condition. It is straightforward strategy of conduct and learning calculations. The agent will figure out how to streamline the new states. It relies upon the work whether the explanation relates to the ideal conduct. In every emphasis the agent identify its present state i.e. (s ∈ S), select an activity (a ∈ A), potentially changing its state and receives a reward signal (r ∈ R). In this procedure the agent needs to acquire valuable encounters related to states, activities, state developments and rewards to act ideally and the assessment of the framework happen simultaneously with learning process. The fundamental goal of Reinforcement learning is to figure out how to guide states of actions while augmenting a reward signal. In Reinforcement learning a self-ruling agent pursues experimentation procedure to become familiar with the ideal activity to perform in each state so as to achieve its objective.

Reinforcement Learning is generally demonstrated as a Markov Decision Process (MDP). It is the scientific establishment model utilized for single agent Reinforcement Learning. In this model complications are of successive decision-making in which the activities have to be chosen at each state by visiting the concerned framework. Such issues are across the board in stochastic control hypothesis and their foundations can be followed. Reinforcement Learning issues can be demonstrated as a Markov Decision Process (MDP) categorized by a quadruple (S, A, P, R) where S is the set of states, A(s) is the set of activities accessible in state S, P : S × A × S → [0, 1] is a progress dispersion that determines the probability of penetrating the specific state in the wake of making a given move in a given state, R : S × A → r is the action that indicates the quick reward while making a given move in a given state. There is an objective of figuring out how select activities that boost the aggregated total of remunerations over time r0 + yr1 + yr2 + ....... where γ ∈ (0, 1) is the markdown factor and decides how emphatically prompt the rewards that are weighted contrasted with remunerations later on. A markdown factor γ < 1 ensures that the future limited return R_t is dependably a limited number if the prompt reward...
is limited. If $\gamma < 1$ then the reward is not exponentially exactly same as the reward received at the initial stage then MDP is known as the limited reward MDP and when $\gamma = 1$ then the value of MDP is known as undiscounted.

4 STEPS FOR REINFORCEMENT LEARNING PROBLEM
These steps are applied to Reinforcement Learning problem before it came into existence

4.1 Understanding your Problem
Reinforcement learning isn’t essentially required every problem. There should be check for each and every problem before applying reinforcement learning algorithm the following attributes are taken into consideration a) trial-and-error an experimental technique b) delayed rewards c) can be displayed as MDP d) to check that the problem is a controlled problem or not.

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

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

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

5 APPLICATIONS OF REINFORCEMENT LEARNING
Coming up next there are some application zones in which reinforcement learning is utilized as activities in a domain to amplify some thought of combined reward.

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

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

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

5.4 Machinery Applications
Reinforcement learning is a kind of Machine Learning calculation that permits programming agents and machines to expressly decide the ideal conduct inside a particular setting and to expand its performance. These applications can’t be modified. It comprise of Manufacturing, Inventory Management, Delivery Management, Power Systems and Finance Sector.

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

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

6 REINFORCEMENT LEARNING ALGORITHMS
An Agent is a significant perspective in Reinforcement learning. It is also known as decision maker and learner. Whatever that is outside the agent is considered as a domain of that agent. By using trial-and-error interactions this domain of Reinforcement Learning framework becomes a powerful agent for mapping all the situations to activities. These activities are done by using the single-agent and multi-agent framework that are in contrast regarding their properties. In multi-agent framework other adjusting agents are also used that makes the environment unstable, violating the Markov property on which the conventional single agent depends upon. The following are the different Reinforcement Algorithms that are used with multi-agent systems

6.1 Minimax-Q Learning Algorithm
Minimax-Q learning algorithm for zero-sum games situations in
which the learning player expands its payoffs in any circumstance. The player’s passion towards the game is opposite. Initially, it is of worth capacity that support learning calculation. In this algorithm the player attempts to augment its normal incentive even with the most exceedingly awful conceivable activity decision of the adversary.

6.2 Nash-Q Learning Algorithm
Hu and Wellman, 2003 proposed a zero-sum game framework of Minimax-Q learning algorithm to general-aggregate games and build up a Nash-Q learning calculating algorithm for multi-agent reinforcement learning techniques. To spread out Q-learning to the various multi-agent learning domains, there are several joint activities of participating agents instead of purely individual actions are required to be taken into consideration. Due to this significant variance among single-agent and multi-agent Reinforcement learning agents this algorithm needs to continue with the Q values for both i.e. the learner itself and other players too. The main reason to discover Nash equilibria at each state to get Nash equilibrium approaches for updating the Q values. Nash-Q learning algorithm is applied only after the Nash Q-value is defined. This defined value is characterized as the estimated total of limited rewards where all the agents are required to follow indicated Nash equilibrium policies onwards. Further Hu and Wellman, 2003 again highlighted that this learning algorithm in multi-player situation combines with Nash equilibrium strategies under certain conditions and extra expectations to the payoff structures.

6.3 Friend-or-Foe Q-Learning (FFQ) Algorithm
In FFQ algorithm each and every agent that is in the framework is known as either “friend” or “foe.” Here the equilibria can be categorized either in the form of coordination or in the form of adversarial. FFQ-learning algorithm can provide stronger convergence assurance when compared with Nash-Q learning algorithm.

6.4 rQ-Learning Algorithm
The rQ - learning algorithm is developed to deal with large search space problems. In this algorithm r-state and an r-action set is always to be well-defined at initial stage. An r-state is initialized by the set of first-order relations like goal in front, team robot to the left, etc. whereas an r-action is initialized by the collection of pre-conditions and post-conditions with a generalized action. To define an r-action appropriately, there is a condition that is to be satisfied, the condition is “if an r-action is appropriate to the specific occurrence of an r-state, then it must be appropriate to all the occurrences of that r-state. This algorithm is convenient with large search space problems that might be very hard to describe an r-state and an r-action set appropriately, particularly for the situation with inadequate information. Moreover, Morales, 2003 states that in the r-state space there is no assurance that the number of well-defined r-actions are acceptable to discover the best arrangement of primitive activities and problematic approaches can be made.

6.5 Fictitious Play Algorithm
In Nash-equilibrium-based learning there is trouble in finding the outcomes of Nash equilibria then fictitious play algorithm is used and it provide another technique to manage with multi-agent framework. Cao, 1997 and Suematu, 2002 stated that in this algorithm the other strategies that are represented by experimental dissemination and also the players are required to keep their own Q values that are associated with the combined activities and are weighted by their conviction appropriation. The fictitious play algorithm converts the variant of individual Q-learning for the stationary approaches of different players and for the non-stationary approaches of different players this fictitious play algorithm have been utilized in either modest games where the players can demonstrate their adversarial opponents called rival modeling or combined games where the players can learn their Q values of their joint activities known as Joint Action Learner.

6.6 Multi-Agent SARSA Learning Algorithm
Nash-Q and Minimax-Q learning algorithms are actually known as the off-strategy Reinforcement Learning algorithms because these algorithms changes the max operator of specific Q-learning algorithm by their preeminent reaction and that reaction is known as the Nash equilibrium policy. In Reinforcement Learning an off-approach learning algorithm continuously attempts to combine the best Q values of optimal strategy irrespective of what strategy is to be executed. According to Sutton, 1998 SARSA algorithm is one of the on policy Reinforcement Learning algorithm that attempts to converge to ideal Q values of the strategy that are currently executed. Suematu, 2002 indicated in his study that a SARSA based multi-agent algorithm knowns as EXORL (Extended Optimal Response Learning) was developed to overcome that drawbacks of the Minimax-Q and Nash-Q learning algorithms.

6.7 Policy Hill Climbing (PHC) Algorithm
This algorithm updates the Q values in the same manner as the fictitious play algorithm does, but it maintains the mixed policy that is also known as the stochastic policy by executing the hill-climbing in the space of these policies. Bowling and Velso, 2002 proposed a PHC algorithm WoLF (Win or Learn Fast) by embracing an idea of Win or Learn Fast and also by utilizing the variable learning rate. While using this algorithm it results immediately to the agent learning if they are not doing effectively and cautiously. This change in the learning rates will be useful for the convergence by not over fitting to the different agents evolving strategies.

7 CONCLUSION
In today’s world people are constantly looking for an agreeable life that is the major reason that we have always relied on machines to accomplish our work more effortlessly in a faster and well-organized way. In the past years, machines have been utilized to reduce the physical work, but in the present era with the advancement of Artificial Intelligence humans seek to make machines that are not only strong but also intelligent therefore the concept of machine learning has come into existence and that become an area of study that is gaining momentum day by day. In this paper categories of machine learning i.e. supervised learning, unsupervised learning, recommender system and reinforcement learning are discussed that also displays the various applications under machine learning. Reinforcement learning is one of the type of machine learning that expounds its application and research to the wider region of control and various decision problems that are not usually handled by supervised or unsupervised learning techniques. To deal with such type of situations Reinforcement Learning has come into existence and become one of the most intelligent agent’s due to its various
characteristics like online learning, self-improving and very less programming effort. This paper reviews the various reinforcement learning algorithms that can decrease the number of state space, improves the learning productivity at the initial state of the testing and then accelerate the convergence. Multi-Agent Q-learning algorithm are also discussed that are used to construct the probable field of artificial intelligence to set the Q values according to the previous information. Here the reinforcement learning applications are also discussed that illuminates how reinforcement learning has become an impressive technology and able to attained remarkable outcomes over the various set of challenging problems. In addition with the development of more vigorous and effective algorithms, efforts have been made to provide some resolutions to these challenges. Hence, there is necessity for resolving problems related to learning techniques, decomposition, methods for approximation and combination of partiality related real life problems by using latest practices in reinforcement learning and new methods.

REFERENCES


[26] R. Ollfati-Saber, Flocking for Multi-agent Dynamics


