Analysis of Reinforcement Learning in Maze Environment

Savita Kumari Sheoran[1], Poonam[2]

[1]Associate Professor & Chairperson
Department of Computer Science & Applications
Chaudhary Ranbir Singh University, Jind (Haryana) – INDIA

[2]M.Phil. (CS) Scholar
Department of Computer Science & Applications
Chaudhary Ranbir Singh University, Jind (Haryana) – INDIA

Abstract

Maze environment presents complex path grids made of an arbitrary number of squares of varying width and length with restricted movements. Since its geometry incorporate varying level of knowledge, hence presents interesting challenge before Artificial Intelligence (AI) community. Reinforcement Learning is used to trace optimal solution in maze learning environment, where agent learns its behavior through trial and error. The discrete Q-Learning, Dyna-CA and Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning) are commonly used and proven Reinforcement Learning methods in Machine Learning to solve such puzzles. This research paper aim to finding out such method which converges in minimum time. We he simulated the maze environment with and without obstacles configurations over MATLAB computational platform to compare the real time parameter of convergence time. The performance results were analyzed and presented. The final results reveals that FRIQ-learning outperform the others under all conditions.

Keywords: Reinforcement learning, discrete Q-learning, DYNA-CA learning, FRIQ-learning, maze problem.

1. INTRODUCTION

Reinforcement learning (RL) is a learning theory that came from animal theory and now applied on machines to work like a human being. This learning is used to train machines by trial and error. There are many reinforcement techniques as Q-Learning, SARSA, Temporal difference, function approximation etc. RL is a sub-area of Artificial Intelligence. In this learning an autonomous agent is situated in an unknown environment to find goal with positive and negative reinforcement. Here agent is any software or hardware entity which faces the state of the environment and takes an action according to positive or negative reinforcement. Agent learns from trial and error. Reinforcement Learning is a common machine learning algorithm of many computational intelligence applications. It adapts the Dynamic environment by trial and- error style iterations. This machine learning works in statistics, psychology, neuroscience and computer science. In the last decade, the field of machine learning and Artificial Intelligence has accrued a greater interest from all sides' fields. Reinforcement learning work in that cases when exact principles of operation are unknown and either only goal or some expected results are known. Some problems as maze problem, in which we cannot find in advance correlations of actions and states. Maze is a network of puzzle paths and has a specific goal that needs to find through these puzzle paths. Through reinforcement learning, where the system learns to achieve the goal from scratch without initial knowledge, based only on rewards and punishments given by the environment in a trial-and-error style, can solve better than other machine learning techniques.

The rest of this paper is organized as follow: Section 2 gives a survey of available literature. Section 3, recapitulates various Reinforcement learning methods of discrete Q-Learning, Dyna-CA and Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning). Section 4, presents the simulation results and finally paper is concluded in section 5.

2. LITERATURE REVIEW

There are a lot of research has been done by researchers using the reinforcement learning technique with modification and combination of other existing techniques such as fuzzy logic, neural network etc. Fuzzy Q-Learning is an extension of Q-Learning into fuzzy environment. My research is focus on improvement in Q-Learning method using fuzzy interpolation rules in maze environment. In 1996, Hamid and R. Berenji introduce a GARIC-Q method for improving the speed and applicability of fuzzy Q-Learning through generalization of input space by using fuzzy rules. Generalization is to generalize between similar situations and actions. Generalization is very necessary in large problems for automatic learning. This method provides the first step toward a true intelligent system where agents can explore the environment and learn from their experience.

In 1999, Chris Gaskett, David Wettergreen, and Alexander Zelinsky proposed wire-fitted neural network for continuous action on continuous states, but it had still problem of smoothly varying control action. In 2008, Alessandro Lazaric, Marcello Restelli and Andrea Bonarini propose a sequential Monte Carlo approach with reinforcement learning, it specially focused on continuous actions. It was called SMC- learning. In 2008, Hado van Hasselt and Marco A. Wiering developed a new algorithm called calca (Continuous Actor Critic Learning Automaton), which uses a continuous actor has clear advantages over Q-Learning and SARSA algorithm. It also works even when some actions are removed from action space after some time of learning. But it takes small number of actions at one time.

In 2009, Jos´e Antonio Mart´ın H. and Javier de Lope presented a new Reinforcement Learning algorithm was presented by called Ex a Reinforcement Learning Algorithm. It was applied on continuous actions. H. Wang and X. Guo proposed an algorithm called HRLPLA. This is a Dynamic web service composition algorithm. It considers both function and QoS. It solves the problem of low composition efficiency in RL for service composition when encounters large scale services. The algorithm also has the advantages of high effectiveness and strong adaptability.

In 2011, Edwards and W. M. Pottenger proposed a new technique "Higher Order Q-Learning. This technique drastically reduces the amount of exploration required in the initial stages of learning. It combines reinforcement learning with Higher Order Learning. This is especially important for online learning mechanisms, where learned is rapidly required. Edwards and W. M. Pottenger presented a new method Motivated learning, is a combination of reinforcement learning and the goal creation system (GCS). ML based agent, which has the ability to set its internal goals autonomously, is able to fulfill the designer's goals more effectively than RL based agent. Motivated learning has better performance than Reinforcement learning, especially in Dynamic changing environment. The Motivated Learning Agent`s essential aim is to survive in a hostile, Dynamic changing environment.

In 2013, FU Bo, Chen Xin, HE Yong, Wu Min proposed a fast and effective Reinforcement Learning algorithm called Dyna CA (Continuous Action). This algorithm gets continuous actions over states, but not focuses directly on state space. In 2015 Tamás Tompa, Szilveszter Kovács shown the benefits of FRIQ-learning (Fuzzy Rule Interpolation-based Q-learning) over the traditional Q-learning. They proved this by applying these methods on maze problem. They compared maze problem based on the convergence speeds in iteration steps. There is also need to know the time of convergence of these methods with different configuration of maze.

3. Learning Models for Maze Problems

Reinforced Learning is well know computational intelligence techniques which solve complex problems by learning from unknown environment. Recently, traditional Q-Learning and Dyna-CA appear as an effective tool in such environment. But major shortcoming with these reinforced learning techniques is that they are not effective when the dimension count of the possible states and actions are relatively high. In such scenario FRIQ-learning (Fuzzy Rule Interpolation-based Q-learning) is much effective method. This research aims to experimentally explore these models in detail. The subsequent sub-sections of this section will theoretically elucidate them in detail.

3.1. Q-Learning:

Watkins and others proposed Q-learning algorithm in 1989 as a model of free reinforcement learning method. It is an iterative algorithm that assumes an initial condition before the first update occurs to find fixed point solution of a complex puzzle. The Q- Learning has discrete state, action and Q-function representation and hence provide sluggish response towards decision making.

3.2. Dyna-CA Learning

Dyna-CA is machine learning paradigm where agent learn through hit and trial during interaction with the environment. It has superiority over Q-Learning model due to its continuous learning behavour. It improve its policies dynamically to feed the agent during the action and hence more adaptive in comparison to former. In Dyna-CA all the state around current state makes best selection and improve speed by gaining experience from immediate stages.

3.3 FRIQ-Learning

The FRIQ-learning is an extension of the discrete Q-learning by augmentation of Fuzzy Rule Interpolation-based Q-function representation. The knowledge representation in FRIQ-learning is a fuzzy rule-base and hence it can handle continuous state and action space more effectively. In this model agent learn through automatic rule-base construction and rule-base reduction strategies based upon principle of reward and punishment. At the onset the system constructs rules and in subsequent steps modifies the rules taking care the rewards and punishments. Ever time the prediction goes far away from set rule, the system auto insert a new rule based upon ever gathered learning experiences. This model have natural benefit over the former two models due to its adaptive and automotive learning style.

4. PERFORMANCE RESULTS

The performance of three commonly deployed Reinforcement Learning methods viz. traditional Q-Learning, DYNA-CA Learning and FRIQ-Learning is tested for parameter of convergence time using MATLAB platform. To check performance in varied conditions following four maze configurations are considered:

- ✓9x6 sized maze without obstacles.
- ✓9x6 sized maze with obstacles.
- ✓18x12 sized maze without obstacles.
- ✓18x12 sized maze with obstacles.

The convergence time performance for these four situations in case of discrete Q-Learning, Dyna-CA Learning and FRIQ-Learning is represented pictorially in subsequent part of this section. The perusal to pictorial staging depicts that convergence time increase with increase in maze size as well as introduction of obstacle.

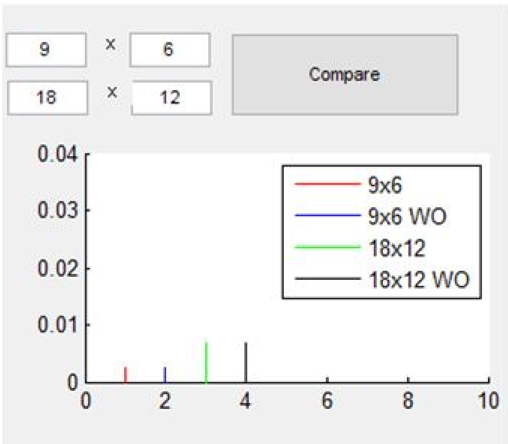

Figure 1: Maze without obstacle

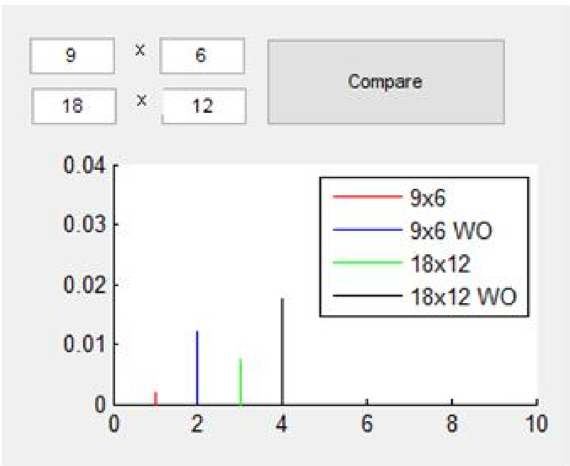

Figure 2: Maze with obstacle = .01

The convergence time calculated by implementing maze configurations without obstacle using MATLAB in iteration steps of discrete Q- Learning are presented below. Figure 1, show results without obstacle, while figure2, shows it with obstacle of 0.01.

The convergence time calculated for Dyna-CA learning without obstacle is presented below. Figure 3, show results without obstacle, while figure 4, shows it with obstacle of 0.01.

The convergence time calculated for FRIQ-Learning without obstacle is presented below. Figure 5, show results without obstacle, while figure 6, shows it with obstacle of 0.01.

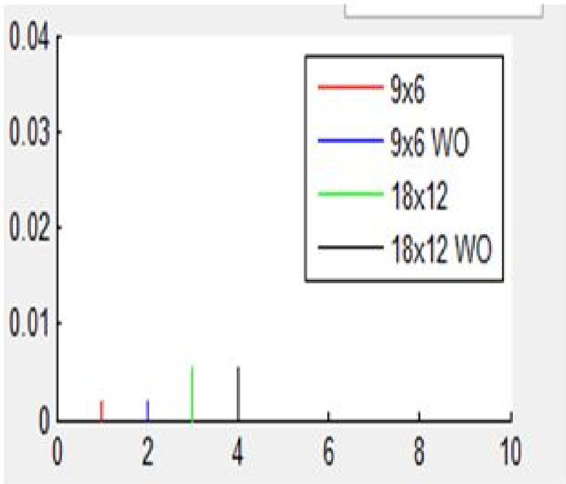

Figure3: Maze without obstacle

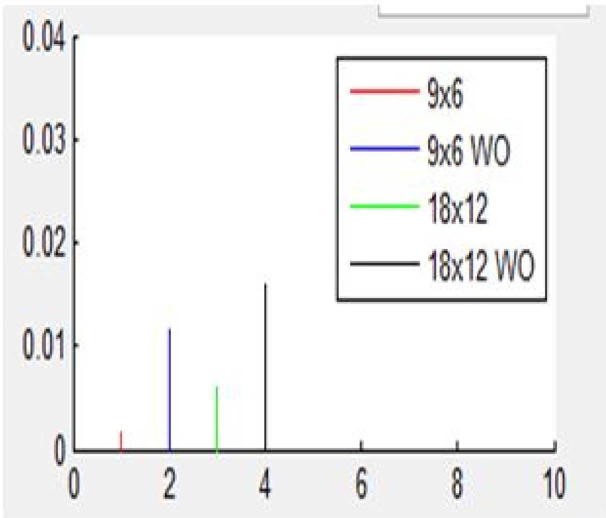

Figure 4: Maze with obstacle = .01

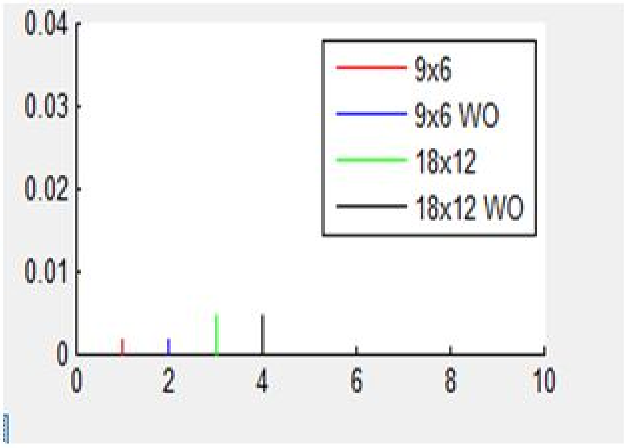

Figure 5: Maze without obstacle

The cohesive intra comparison of convergence time for three learning methods is represented below. Figure 7, represents the kit for maze problem without obstacle while figure 8, depicts the same with an obstacle of value 0.01.

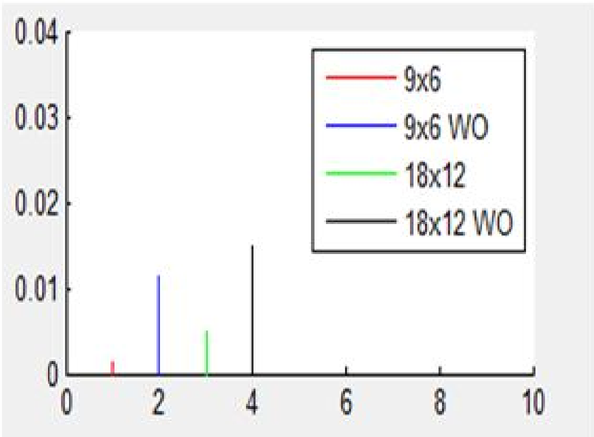

Figure 6: Maze with obstacle = .01

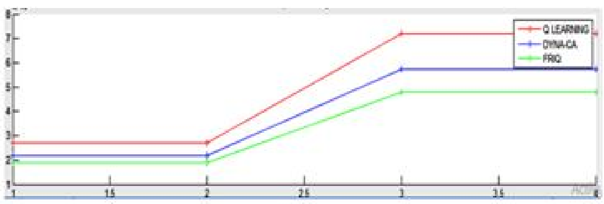

Figure 7: Composition of Convergence Time without Obstacle

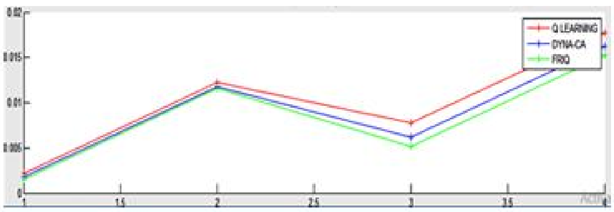

Figure 8: Composition of Convergence Time with Obstacle = 0.01

The perusal to figure 7 and 8 reveals that FRIQ-Learning outperform the traditional Q-Learning and Dyna-CA Learning with and without obstacle.

5. CONCLUSION

In this research article, we have analyzed state of art machine learning models for agent based learning to solve maze puzzles. A simulation study to analyze convergence time is constituted to compare the performance of three state of art leaning models viz. discrete Q-Learning, Dyna-CA and Fuzzy Rule Interpolation-based Q-learning in varied condition of obstacle. The numerical simulation results generated through MATLAB tool reveals that Fuzzy Rule Interpolation-based Q-learning has require minimum time overhead and hence considered best to find optimal path in maze puzzles. Since the whole study is carried out in 2-dimensions only. The 3-dimensional extension of study may be much fruitful to solve intricate geometrical problems in Euclidean and Non-Euclidian space.

References

[1]. Tamás Tompa, Szilveszter Kovács, "Q-learning vs. FRIQ-learning in the Maze problem", IEEE Cognitive Infocommunications, 2015.
[2]. Leslie P.K." "Reinforcement Learning: A Survey" In Journal of Artificial Intelligence Research, volume 4, 1996.
[3]. M. Aurangzeb, F. L. Lewis, and M. Huber, "Efficient, Swarm-Based Path Finding in Unknown Graphs Using Reinforcement Learning*", IEEE Control and Automation (ICCA), 2013.
[4]. Donald Wunsch, "The Cellular Simultaneous Recurrent Network Adaptive Critic Design for the Generalized Maze Problem Has a Simple Closed-Form Solution", IEEE INNS-ENNS, 2000.Paul J. Werbos, "Generalization Maze Navigation: SRN Critics Solve What Feedforward or Hebbian Nets Cannot", IEEE Intelligent Control, 1996.
[5]. Swati Chaudhari1, Manoj Patil, "Study and Review of Fuzzy Inference Systems for Decision Making and Control", International Journal of Advanced Computer Research 4(4), 2014.
[6]. Serge Guillaume, "Designing Fuzzy Inference Systems from Data: An Interpretability-Oriented Review", IEEE Transactions on Fuzzy System, 19(3), 2011.
[7]. Peter Baranyi, "A Generalized Concept for Fuzzy Rule Interpolation", IEEE Transactions on Fuzzy System, 12(6), 2005.
[8]. Chengyuan Chen, "Rough-fuzzy rule interpolation" Information Science, 2016.
[9]. Szilveszter Kovács, "Fuzzy Rule Interpolation in Practice" 2006

[10]. Fu Bo, Chen Xin, HE Yong, Wu Min, "An Efficient Reinforcement Learning Algorithm for Continuous Actions", IEEE 25th Chinese Control and Decision Conference, 2013.
[11]. Lucian Bus¸oniu, Robert Babuˇska, and Bart De Schutter, "A Comprehensive Survey of Multiagent Reinforcement Learning", IEEE Transactions on Systems, Man, and Cybernetics, Part C, 38 (2), 2008.
[12]. Richard S. Sutton and Andrew G. Barto, "Reinforcement Learning: An Introduction", The MIT Press, 2012.
[13]. Abu Bakar Sayuti Saman, "Solving a Reconfigurable Maze using Hybrid Wall Follower Algorithm", International Journal of Computer Applications , 82 (3), 2013.
[14]. Mohit Ahuja, Baisravan HomChaudhuri, Kelly Cohen and Manish Kumar, "Fuzzy Counter Ant Algorithm for Maze Problem", 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 2010.
[15]. D. Venkata Vara Prasad, "Knowledge based Reinforcement Learning Robot in Maze Environment", International Journal of Computer Applications, 14 (7), 2011.