**International Journal of Advanced Science and Research**
**ISSN: 2455-4227, Impact Factor: RJIF 5.12**
**www.allsciencejournal.com**
**Volume 2; Issue 1; January 2017; Page No. 01-05**

# Agent based optimization using reinforcement learning in maze environment

**[1] Savita Kumari Sheoran, [2] Poonam**

[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 is a complex environment where finding optimal path is always a challenge. Recently, traditional Q-Learning and Dyna-CA appear as an effective tool to solve such problems. 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 Rule Interpolation based Q-Learning (FRIQ) may be an effective tool to solve the maze problems. Further, MATLAB ® is a computational platform having effective environment to simulate static as well as dynamic maze environment. This research article theoretical analyze all these issues and based upon it decide a research direction for agent based optimization using reinforcement learning in maze environment.

**Keywords:** agent based optimization, reinforcement learning, maze environment, machine learning and learning models

## 1. Introduction

Reinforcement learning (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 (RL) is a common machine learning algorithm of many computational intelligence applications. RL is adapting the dynamic environment by trial and error style iterations [1-2]. This machine learning works in statistics, psychology, neuroscience and computer science. In the last ten year, it has attracted rapidly increasing interest in machine learning and artificial intelligence fields [3]. There are many 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 can't find in advance correlations of actions and states. Maze is a network of puzzle paths [4-6] and has a specific goal for finding 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 [7].

Recently many researches have been done on maze problem by using reinforcement learning but there is also a problem in reinforcement learning that it is ineffective in the case when the dimension count of the possible states and actions are relatively high. This drawback can be removed by using fuzzy inference systems [7]. Fuzzy Inference System is a system that maps input into output using fuzzy logic [8-9]. When Fuzzy Rule Interpolation (FRI) [10-12] is applied with discrete Q-Learning (Reinforcement Learning) [13-15] called Fuzzy Rule Interpolation based Q-Learning (FRIQ) [1, 16-17]. This paper presents all such research issues arise reinforcement learning in maze environment and a way for their possible solution. The rest of this paper is organized as follow: The remaining part of this section presents learning, its various types and maze environment. Section 2 gives a survey of available literature in the field. Section 3 ascribes the main research issues along with their auxiliary concerns in the proposed area. Section 4 summarizes the scope of this research and section 5 finally concluded the paper.

### 1.1 Learning

Learning is the act of acquiring new or modifying and reinforcing, existing knowledge, behaviors, skills, values or preferences and may be involve different types of information. The ability to learn is possessed by humans, animals, plants and machines. Learning means change in behavior occurs as a result of experience [18]. G. Murphy and H.P. Smith ascribed it as:

- ***Gardener Murphy***: "The term learning covers every modification in behavior to meet environmental requirement.
- ***Henry P. Smith***: "learning is the acquisition of new behavior or the strengthening or weakening of old behavior as the result of experience.

#### 1.1.1 Machine Learning

Machine learning means to train machines like a human or better than human to perform tasks. It is a subfield of computer science. It involves the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Author Samuel defined machine learning as a, "field of study that gives computers the ability to learn without being explicitly programmed [19].

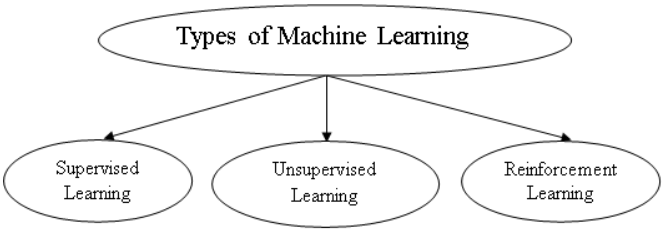


**Fig 1:** Types of Machine learning

Machine learning systems automatically learn programs from data. This is often a very attractive alternative to manually constructing them. Machine learning is used in Web search, spam filters, recommender systems, ad placement, credit scoring, fraud detection, stock trading, drug design, and many other applications [3].

There are two aspects of a given problem: One is, to call the programs that learn and improve on the basis of their experience, second is, directly program our computers to carry out the task. On the basis of learning, machine [22].

### 1.1.2 Supervised Learning

In this learning we make an algorithm according to problem. This algorithm needs both input and output data. Algorithm is generalized such that it is able to give correct result of all possible inputs. Here inputs are training data and output of this data is compared with actual output data [20]. The aim of supervised learning is to build a model that makes predictions based on evidence in the presence of uncertainty.

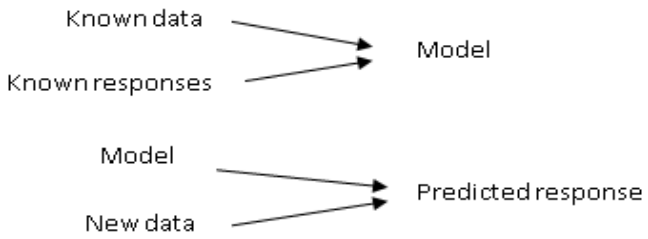


**Fig 2:** supervised learning process

In this learning, an algorithm takes a known set of input data and known responses to the data (output), and trains a model to generate reasonable predictions for the response to new data [20]. Agent of this learning is made for every input with goals and compare with actual responses and adjust its internal memory in such a way that it is more likely to produce the appropriate response, when next time it receives the same input [22].

### 1.1.3 Unsupervised Learning

In this learning, agent not gets any feedback from environment. This learning is used where we not know the output. In this learning, agent learns by finding hidden structure in unlabelled training data and is based on the similarities and differences among the input patterns [21]. Aim of this learning is to represent the inputs in more efficient and effective way. It categorizes data according to problem as by clustering or reducing set of dimensions. As there are no explicit target output associated with each input, these input representations can then be used by other parts of the system in a way that affects behavior [22].

### 1.1.4 Reinforcement Learning

Reinforcement learning is an area of machine learning inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward [22].

It is learning what to do-how to map situations to actions-so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics-trial-and-error search and delayed reward-are the two most important distinguishing features of reinforcement learning [22-23].

One of the challenges that arise in reinforcement learning and not in other kinds of learning is the trade-off between exploration and exploitation. To obtain a lot of reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in producing reward [24].

## 1.2 Basic Reinforcement Learning Models

The basic model of reinforcement learning is shown in figure 3. When intelligent agent interacts with the environment, it chooses an action to obtain the biggest reward. The interactive interface of intelligent Agent and environment includes action, reward and state. When each time intelligent agent interacts with the environment, first of all, the system accepts the input s (state), and then according to the internal inference mechanism, it gives output as a (action). Finally, the environment changes to new state 's' after accepting the action. The system accepts the input of the new state 's' and obtains the rewards and punishment signal 'r' of environment for the system [8,13,15]. It can be visualized through following steps:

1. A set of environment states;
2. A set of actions;
3. Rules of transitioning between states;
4. Rules that determine the scalar immediate reward of a transition;
5. Rules that describe what the agent observes.

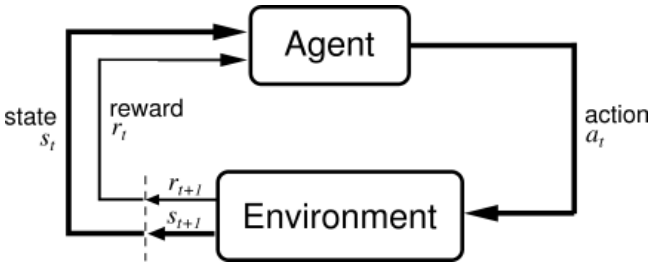


**Fig 3:** Basic Reinforcement Learning Model

## 1.3 Maze Environment

A Maze is a path or network of paths and has a goal. It has many obstacles in the path, so one has to cover these obstacles to find goal. Main motive is to find goal with shortest path with less time and more accuracy from these puzzle network of paths [25].

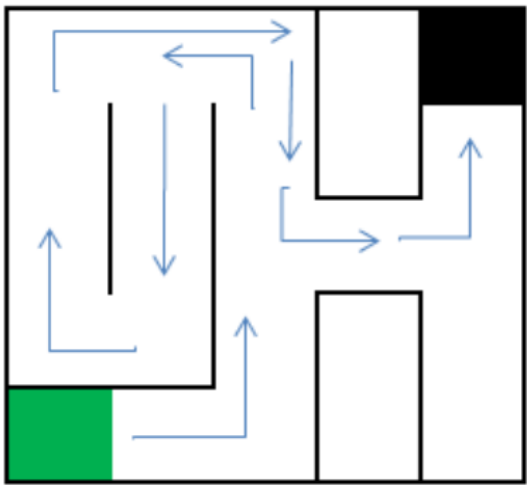

**Fig 5:** Not the most efficient short path in this maze

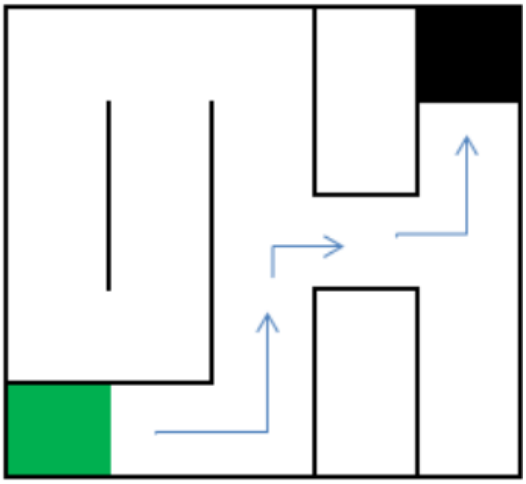

**Fig 6:** A better result of shortest path in this maze

In these figures, Green Square is an initial point and Black Square is goal point. There are two paths for find the goal. But the path in 5th figure is lengthy and time consuming [26]. There is a need of better algorithms and techniques for solving this puzzle network of paths (solving maze) for arriving at the goal. Maze solving means finding a route through the maze to arrive at the goal. This can solve manually or by computer. Here need is to solve these puzzle paths with minimum time, high speed and with high accuracy. So this is only possible by computer [27]. There many algorithms have been applied to solve maze as A*, Ant Colony Optimization (ACO), Neural Network, Hybrid Binary Particle Swarm Optimization (HBPSO) for multi-vehicle Search Area converge problem, Adaptive Critic Design for the generalized maze problem (path planning for general maze) [28].

Maze is available as game for entertainment, as by playing this game online, people release their stress, it also increases level of focusing time. Maze can also save life of people by making maze solving robot for detecting bombs places (during terrorism attack) and it can also save life of the Army, otherwise Army person is employed for detecting places of bombs [29]. Leonhard Euler was one of the first mathematicians who analysis plane maze mathematically. Maze is also used in psychology experiments to study spatial navigation and learning.

## 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 [1].

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 [30].

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 [31]. 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 [32].

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 [33].

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 [35].

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 [36]. 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 [37].

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 [38].

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 [1].

## 3. Research Issues

The maze problems are explored and optimized through different algorithms but still there are complexities and scope for further optimization as reasoned from above literature perusal. The following subsection addresses these issues in details.

### 3.1 Existing Research Gap

Reinforcement learning is used in many fields as game theory, control theory, robotics, operation research, statistics etc. This learning train machines through trial and error without need of specify how the task is to be achieved. In maze environment, problem of continuous state and action is solved by reinforcement learning. But as the number of states increases in the maze, reinforcement learning is ineffective.

So to remove this problem, fuzzy inference system is used. Fuzzy logic is a multi-valued logic, which gives solution in degrees of any problem not only in two values true and false.

### 3.2 Open Research Problem

We need a solution which can improve in reinforcement learning with increasing number of states in any problem. So the goal of this research is to give a demonstrative example for introducing the benefits of the Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning) over the traditional discrete Q-learning (reinforcement learning) and Dyna-CA learning.

The comparison of FRIQ-Learning, traditional Q-Learning and Dyna-CA learning can be compared using MATLAB ® tools. Such comparison will be based on convergence time and will be able to provide optimal solution in specific situation.

### 3.3 Research objective

Reinforcement learning has proven better in case when no idea of any problem "how do it" in advance than supervised learning. By combining the fuzzy inference system with reinforcement learning, it also removes the problem of high dimensionality of states.

There is a question for need of efficient solution according to time with increasing the states of any problem using Q-Learning (reinforcement learning) with fuzzy inference system. In order to bridge the existing research gap and address the open questions, following objectives may be purposeful:

1. To check the efficiency of FRIQ-Learning over traditional Q-Learning and Dyna-CA learning.
2. To check the efficiency of Dyna- CA learning over traditional Q-Learning.
3. To check the accuracy or result of FRIQ-Learning is better than traditional Q-Learning or not?
4. To check the accuracy or result of FRIQ- Learning is better than Dyna-CA learning or not.
5. To take advantages of both Q-Learning and fuzzy inference system to solve any problem.
6. Identify proposed machine learning approach to increase the time and speed to find goal of the problem.
7. To study and analyses the existing research work regarding of reinforcement learning.
8. Analyzing the accuracy measures for the proposed mechanism.

## 4. Scope of Research

As we know that every machine has need of better learning and more accuracy with high speed and less time. So by using reinforcement learning (Q-Learning) with fuzzy inference system, is a strong step to train machines with more accuracy, high speed and less time. It is very much helpful in problem of high state dimensionality.

It may be helpful in Game theory: (e.g. Backgammon, Chess, Solitaire, Checkers, maze, Control theory: (e.g. helicopter control) Operation research (e.g. Vehicle routing, Pricing, Targeted marketing), Robotics: (e.g. Robot Soccer, Air Hockey, Quadruped Gait Control).

## 5. Conclusion

This research paper theoretically analyzed the research issues pertaining to the agent based optimization in maze environment using reinforcement learning. It have ascribed the existing research gap and open challenge available before scientific and research community working in this regime. Based upon perusal of available literature it appears imperative that Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning) may be an effective strategy to find optimal solution in maze environment. Also MATLAB ® is suggested as effective platform to implement research hypothesis prescribed there in research. Since it is a theoretical conceptualization, the practical implementation of proposed objectives using synthetic or real time maze environment will depict the desired outcome.

## 6. References


1. Tamás Tompa, Szilveszter Kovács. Q-learning vs. FRIQ-learning in the Maze problem, IEEE Cognitive Infocommunications, 2015.
2. Lihong Li. A Unifying Framework for Computational Reinforcement Learning Theory Ph.D. Thesis, the State University of New Jersey, 2009.
3. Leslie PK. Reinforcement Learning: A Survey in Journal of Artificial Intelligence Research. 1996, 4.
4. Aurangzeb M, Lewis FL, Huber M. Efficient, Swarm-Based Path Finding in Unknown Graphs Using Reinforcement Learning, IEEE Control and Automation (ICCA), 2013.
5. 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.
6. Paul Werbos J. Generalization Maze Navigation: SRN Critics Solve What Feedforward or Hebbian Nets Cannot, IEEE Intelligent Control, 1996.
7. Janusz Starzyk A, Yinyin Liu, Sebastian Batog. A Novel Optimization Algorithm Based On Reinforcement Learning, Reinforcement Learning, 2006.
8. Swati Chaudhari1, Manoj Patil. Study and Review of Fuzzy Inference Systems for Decision Making and Control, International Journal of Advanced Computer

Research. 2014; 4(4).
9. Serge Guillaume. Designing Fuzzy Inference Systems from Data: An Interpretability-Oriented Review, IEEE Transactions on Fuzzy System, 2011; 19(3).
10. Peter Baranyi. A Generalized Concept for Fuzzy Rule Interpolation, IEEE Transactions on Fuzzy System, 2005; 12(6).
11. Chengyuan Chen. Rough-fuzzy rule interpolation" Information Science, 2016.
12. Szilveszter Kovács. Fuzzy Rule Interpolation in Practice, 2006.
13. Dávid Vincze, Szilveszter Kovács. Reduced Rule Base in Fuzzy Rule Interpolationbased Q-learning, 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, 2009.
14. Jan ˇSkach1, Ivo Punˇcoch´aˇr1, Frank Lewis2 L. Temporal-Difference Q-learning in Active Fault Diagnosis, IEEE 55th IEEE Conference on Decision and Control, 2016.
15. Christopher Watkins. Technical Note: Q-Learning, Machine Learning, 1992.
16. Dávid Vincze. Rule-base reduction in Fuzzy Rule Interpolation-based Q-learning, Debreceni Egyetemi Kiadó–Debrecen University Press, 2015.
17. Fu Bo, Chen Xin, HE Yong, Wu Min. An Efficient Reinforcement Learning Algorithm for Continuous Actions, IEEE 25th Chinese Control and Decision Conference, 2013.
18. Hal Daumé III. A Course in Machine Learning. 2012. Retrieved from: http://ciml.info/dl/v0_8/ciml-v0_8-all.pdf.
19. Nils Nilsson J. Introduction To Machine Learning, Stanford University Publication, 1998.
20. Shai Shalev-Shwartz, Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014.
21. Wang Qiang. Reinforcement Learning Model, Algorithms and Its Application, IEEE International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), 2011.
22. Yilin Kang, Student Member. Self-Organizing Agents for Reinforcement Learning in Virtual Worlds, IEEE World Congress on Computational Intelligence, 2010.
23. Lucian Bus oniu, Robert Babuˇska, Bart De Schutter. A Comprehensive Survey of Multiagent Reinforcement Learning", IEEE Transactions on Systems, Man, and Cybernetics, Part C, 2008; 38(2).
24. Jos´e Antonio Mart´ın H. An Effective Algorithm for Continuous Actions Reinforcement Learning Problems, 35th Annual Conference of IEEE Industrial Electronics, 2009.
25. Richard Sutton S, Andrew Barto G. Reinforcement Learning: An Introduction, The MIT Press, 2012.
26. Michael L. Littman, Model-Based Reinforcement Learning, NIPS, 2009.
27. Kao-Shing Hwang. Pheromone based Planning Strategies in Dyna-Q Learning, IEEE WCCI, 2016.
28. Abu Bakar Sayuti Saman. Solving a Reconfigurable Maze using Hybrid Wall Follower Algorithm, International Journal of Computer Applications, 2013; 82(3).
29. Mohit Ahuja, Baisravan Hom Chaudhuri, Kelly Cohen, Manish Kumar. Fuzzy Counter Ant Algorithm for Maze Problem, 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 2010.
30. Grant Gilbert Arthur Rivera. Path planning for general mazes, Mater Thesis, Missouri University Of Science And Technology, 2012.
31. Venkata Vara Prasad D. Knowledge based Reinforcement Learning Robot in Maze Environment, International Journal of Computer Applications, 2011; 14(7).