Optimal path planning for mobile robot using Preference based Evolutionary Qlearning algorithm

Dr. G. Nirmala
Assistant Professor, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India. gnirmala.gnirmala@gmail.com

Dr. S. Selva Kumar
Professor, G.K.M. College of Engineering and Technology, Chennai, Tamil Nadu, India. sselvakumar@yahoo.com

Abstract

An efficient Preference based Evolutionary Qlearning algorithm (PEQL) for mobile robot path planning method is proposed that finds optimal path between source to destination using an improved policy evaluation with adaptive reward and policy switching process, the proposed mobile robot path planning algorithm is unique and novel since it finds the optimal path using policy switching and better success index than the conventional methods.

Keywords: *mobile robot, path planning, adaptive reward, policy switching*

Introduction:

Preference based Qlearning algorithm based mobile robot path planning finds an optimal path from a source and destination in a given environment and has become popular for emerging complex mobile robot applications such as robot soccer and planet exploration, surveillance [1, 2]. As a result of the stochastic property, Qlearning algorithm suffer from the convergence time of learning the environment; The path may not be optimal after termination [3, 4]. Since a Qlearning must run repeatedly until it finds an optimal path, low success rates increase the total execution time of the algorithm. To circumvent being surrounded into convergence time, both the evaluation and the natural selection step must be designed intelligently to obtain a well converged population [4]. However, conventional Qlearning methods focus mainly on the policy improvement steps, instead of on the evaluation and selection steps, by incorporating problem specific operations.

The key features of this proposed Evolutionary Qlearning method are,

1.PEQL is based on evolutionary markov decision process with policy switching process. This proposed method significantly increases the performance of mobile robot.
2.Conventional Qlearning (CQL) algorithm are complex which
provided high convergence cost than the proposed work.
3.Experiments are performing simulation in with six different grid world environments.

In this letter, we propose a preference based evolutionary Qlearning algorithm by improving the policy evaluation and policy switching steps. Moreover, the improved steps do not increase the time complexity of the evolutionary Qlearning algorithm. As a result of its simplicity and efficiency, the proposed method can be applied promptly to modern commercial mobile robot applications such as robotic vacuum cleaners, unmanned aerial vehicles and service robots for survival rescue [3, 4].

Existing Work: Path planning of mobile robot between source to destination had been done by various researchers on improving the performance of the learning environment by evolutionary Qlearning with various reinforcement learning with adaptive techniques like LSPI, Qlearning with eligibility trace algorithm, and conventional Qlearning (CQL) algorithm so far.

Proposed Work: The need for PEQL algorithm is to increase the performance of mobile robot path planning applications. PEQL is a policy switching based evolutionary markov decision process with policy evaluation step. This paper focuses on finding the optimal policy between source to destination with and without obstacles in the environment. Proposed PEQL is to obtain enhanced performance measures by improving its policy improvement step.

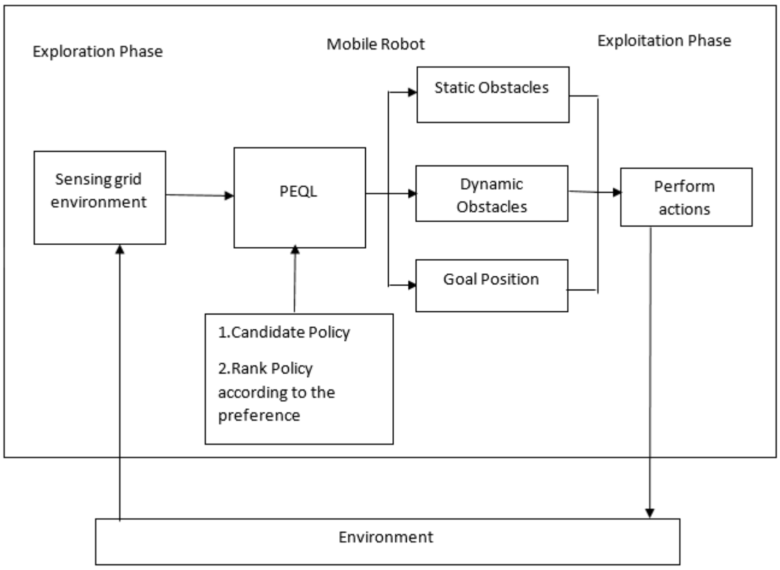

Figure.1. Block Diagram for mobile robot path planning using PEQL

PEQL Algorithm

Initialization

Choose population size n, t, $\pi(0)=\{\pi 1, \ldots\ldots \pi n\}$, where $\pi i \ \varepsilon \ \pi$. Set N= 0, and Pm, Pg and Pl in (0,1] and $\pi^* = \pi 1$

Repeat:

Policy switching: obtain $Q\pi$ for each $\pi \ \varepsilon \ \pi(k)$

Generate the elite policy of $\pi(k)$ defined as

$\pi^*(k) \ \varepsilon \ \{\text{argmax } Q\pi \ (x)\}$, $x \ \varepsilon \ X$

$\Pi \ \varepsilon \ \pi(k)$

Create n-1 random states Si, I = 1,………………n-1 of $\pi(k)$ and selecting m ε {2, …., n-1) with equal probability and selecting m policies in $\pi(k)$ with equal probability.

Generate n-1 policies $\pi(Si)$, defined as:

$\Pi(Si)(x) \ \varepsilon \ \{\text{argmax } Q\pi(x)\}$, $x \ \varepsilon \ X \ \Pi \ \varepsilon \ Si$

Stopping Rule:

If $Qt \ \pi^*(k) \neq N = 0, N = 0$

if $Qt \ \pi^*(k) = Qt \ \pi^*(k-1)$ and N = K, terminate PEQL

if $Qt \ \pi^*(k) = Qt \ \pi^*(k-1)$ and N < K, N= N+1

Policy mutation:

For each policy $\Pi(Si)$, i=1,…,n-1,

Generate a globally mutated policy $\pi m(Si)$ = with Pm

with 1-Pm using Pg and locally mutated policy $\pi m(Si)$ with

1- Pm using Pl and locally mutated policy.

Population generation:

$\pi(k+1) = \{\pi^*(k), \pi m(Si)\}$, i=1,….,n-1

$k \leftarrow k+1$

Now learning the environment of mobile robot is obtained at the learning phase and finding the optimal path between source and destination is obtained at path planning phase.

Result and Discussion:

The effectiveness of the proposed PEQL algorithm method evaluated, we demonstrated the performance of the LSPI, QL with eligibility trace, conventional CQL and the proposed method. In the experiments, we used six well-known grid environment, shown in Fig. 2. In Fig. 2 white and black color of environment represent free and obstacles, respectively. The source and destination are represented as small circle with and small star on the environment, respectively. we run each method repeatedly until it finds an optimal path between source to destination. The superiority of the proposed method is demonstrated by comparing the success rates of each method.

Success Index (SI) = No. of success episodes / Total number of episodes The grid map shows the mobile robot learning the environment taken for comparison of performance of learning measures.

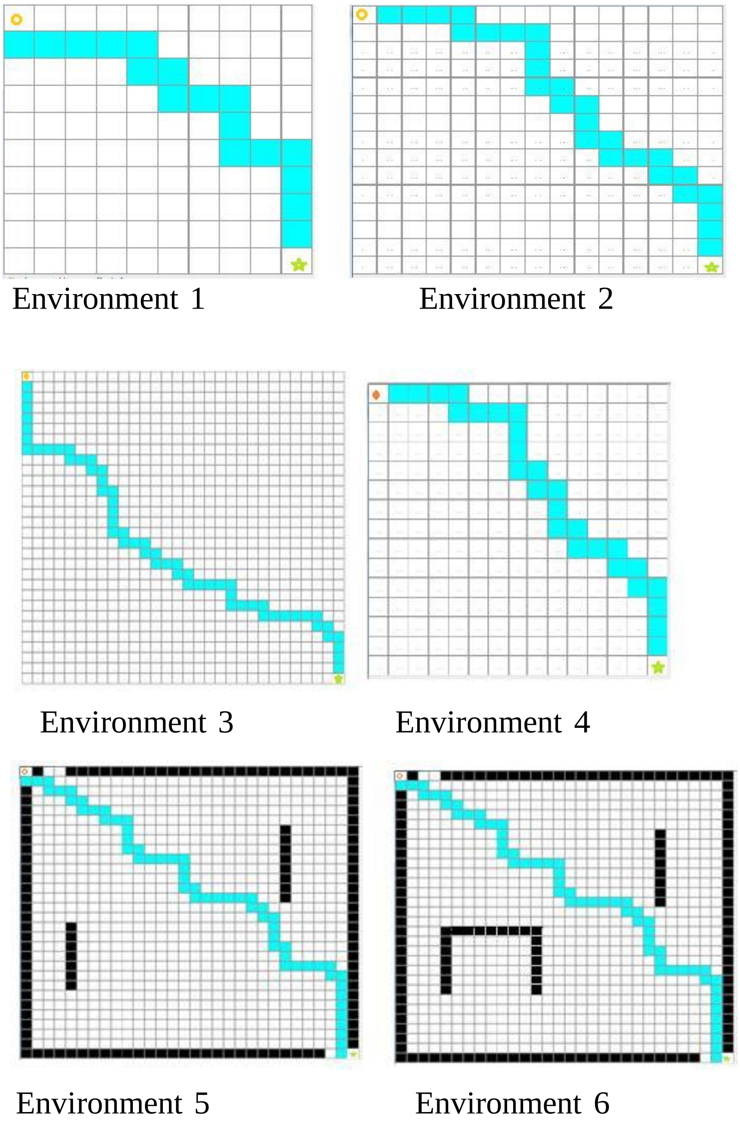

Figure.2. Environments used for comparison

Table 1. Success Index (SI) comparison of different environments

Performance based on Reliability Index				
Environment	LSPI	QL-eligibility trace	CQL	Proposed Work
Grid environment 1	87%	85%	86%	100%
Grid environment 2	86%	90%	60%	100%
Grid environment 3	76%	98%	86%	100%
Grid environment 4	78%	89%	82%	97%
Grid environment 5	79%	87%	84%	98%
Grid environment 6	87%	90%	79%	100%

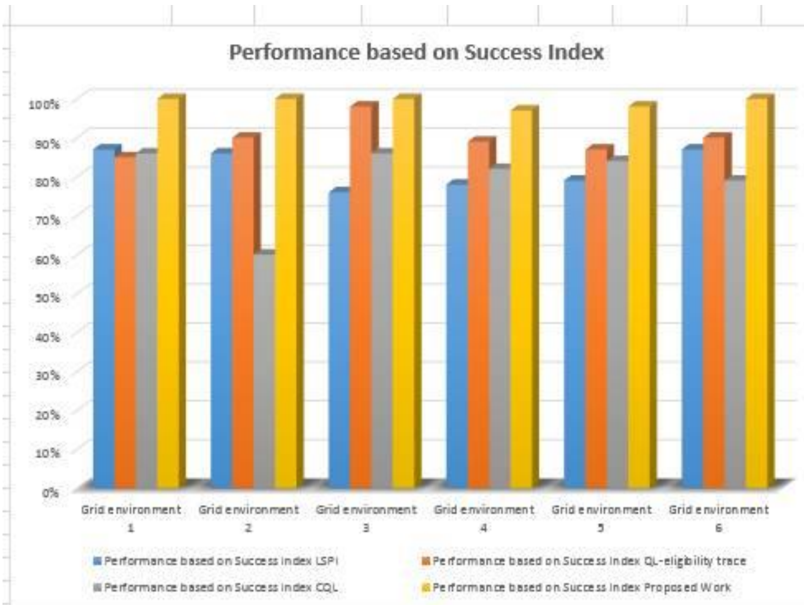

Figure 3. comparison of Success Index (SI) of different environments

The Success Index (SI) of the mobile robot path planning provided in table 1 are summarized by plotting the chart in figure 3.

In contrast, the method (existing method) may find a feasible path stably, but it suffers from very high execution time. Therefore, we can conclude that the proposed method increases the efficiency of the PEQL method based on its higher success rates without an increase of execution time from the evolutionary policy learning.

Conclusion:

In this work, a PEQL method on an improved evaluation and policy switching process is proposed to obtain better performance of mobile robot path planning between source to destination.

References

1. C. Watkins and P. Dayan, “Q-learning,” Mach. Learn., vol.
8, no. 3, pp. 279–292, May 1992.

2.Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, Ilya Sutskever, "Evolution Strategies as a Scalable Alternative to Reinforcement Learning", 2017, OpenAI, arXiv:1703.03864v2

3. Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, "Natural Evolution Strategies" Journal of Machine Learning Research, 2014.

4. Esref Bogar, Selami Beyhan, "A Hybrid Genetic Algorithm
for Mobile Robot Shortest Path Problem", International Journal of Intelligent Systems and Applications in Engineering", Advanced Technology and Science ISSN:2147-6799, pp.264-267, 2016.