Engineering Science and Technology, an International Journal 19 (2016) 651–669

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: http://www.elsevier.com/locate/jestch

Full Length Article

Intelligent-based multi-robot path planning inspired by improved classical Q-learning and improved particle swarm optimization with perturbed velocity

P.K. Das [a,*], H.S. Behera [a], B.K. Panigrahi [b]

[a] Department of Computer Science & Engineering and Information Technology, VSSUT, Burla, Odisha, India
[b] Department of Electrical Engineering, IIT, Delhi, India

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form
9 September 2015
Accepted 22 September 2015
Available online 15 December 2015

Keywords:
Q-learning
Path planning
Mobile robots
Energy
IPSO-DV
Khepera II

ABSTRACT

Classical Q-learning takes huge computation to calculate the Q-value for all possible actions in a particular state and takes large space to store its Q-value for all actions, as a result of which its convergence rate is slow. This paper proposed a new methodology to determine the optimize trajectory of the path for multi-robots in clutter environment using hybridization of improving classical Q-learning based on four fundamental principles with improved particle swarm optimization (IPSO) by modifying parameters and differentially perturbed velocity (DV) algorithm for improving the convergence. The algorithms are used to minimize path length and arrival time of all the robots to their respective destination in the environment and reducing the turning angle of each robot to reduce the energy consumption of each robot. In this proposed scheme, the improve classical Q-learning stores the Q-value of the best action of the state and thus save the storage space, which is used to decide the Pbest and gbest of the improved PSO in each iteration, and the velocity of the IPSO is adjusted by the vector differential operator inherited from differential evolution (DE). The validation of the algorithm is studied in simulated and Khepera-II robot.

1. Introduction

The path planning problem in mobile robotics is considered a complex task. It [1] determines a path for the robot to reach in predefine goal location from a specified starting location without hitting various obstacles in the given environment. The path planning problem has been classified into different categories. One of the classifications is static and dynamic path planning based on the environmental information. In the static path planning, the obstacles and goals are motionless. But in the dynamic path planning, the obstacles and goals are moving in the environment each time, and also the environment is changing every time. Another classification is local and global path planning. Robot navigates through the obstacles by steps and determines its next position to reach the goal by satisfying constraints like path, time and energy optimality [2–8], with the help of the local path planning scheme. In global planning, the robot decides the entire collision free path before its movement toward the goal from a specified initial position. The above mentioned global planning is termed as *offline planning* [9]. Local path-planning, which includes navigation and online planning, is sometimes referred to as navigation only in the literature. The phrase motion planning, which includes the notion of time with the position of a robot on a planned trajectory, is often used in the context of path-planning. In path planning, we need to generate a collision free trajectory path in the world map by avoiding the obstacles, and path is optimized with respect to certain criteria. However, the environment may be vast, dynamic, imprecise, uncertain and partially non-structured. In such environment, the mobile robots often used the machine learning to become aware about its environment. In early, research was used the supervised learning to train the robots to determine its next position in the given world map based on the sensory data gained from the environment. But it has provided the best result for mobility management of robots in fixed maps. However, it is difficult to guide the robot to decide its next position, although the acquired knowledge to small changes in the robot's world map. So a complete training is required for the robot with both old and new sensory data–action pair to overcome the above problem.

Reinforcement learning is considered as an alternative learning policy, which is based on the principle reward and penalty. In this learning an agent performs an action on the environment and

* Corresponding author. Tel.: +919439005466; fax: 06632430573.
E-mail address: daspradipta78@gmail.com (P.K. Das).
Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2015.09.009

receives an immediate reward or penalty based on the action. The learner adapts its parameter based on the status of (reward/penalty) the feedback signal from its environment. Since the exact value of the futuristic reward is not known, it is guessed from the knowledge about the robot's world map. The primary advantage of reinforcement learning lies in its inherent power of automatic learning even in the presence of small changes in the world map. There exist extensive research on multi-robot navigation on reinforcement learning that has been tested in many simulated environments [3,10–18] but on a limited basis in real-world scenarios. A real-world environment poses more challenges than a simulated environment, such as enlarged state spaces [11], increased computational complexity, significant safety issues (a real robot can cause real damage), and longer turnaround times for results. This research measures how well reinforcement-learning technique, such as Q-learning, can apply to the real robot for navigational problem [19,20]. The author [21] has implemented the multi-robot navigation in the Khepera-II environment by designing an adaptive memetic algorithm (AMA) by utilizing the composite benefits of Q-learning for local refinement and differential evolution (DE) for global search. In the paper [22] multi-robot navigation is solved in the real world map by hybridization of the Artificial Bee Colony (ABC) for global search and Q-learning for local refinement; the performance is evaluated in terms of runtime, cost function and accuracy. The paper [23] used the Lyapunov design principle in the reinforcement learning to switch control policy instead of training the agent for control policy and combine PSO and Q-value-based reinforcement learning for neuro-fuzzy system design. The multi-goal Q-learning algorithm has been modeled to solve the multiple goal learning problems in the virtual team [24]. In this presented paper, we modified the classical Q-learning algorithm (CQL), hereafter called improved Q-learning (IQL), and is integrated with an improved particle swarm optimization (IPSO) hybridized with DV, called IQ value-based IPSO-DV, to improve its performance for path-planning problem of multi-robots.

The online trajectory path planning of multi-robot from specified initial position to a goal position without hitting obstacles and a teammate is presented in this work. In a multi-robot path planning problem, each robot has a specified initial and goal position in a given environment, and each robot has to plan its collision free path without hitting any of the colleagues or obstacles present in the map through offline or online approach. The obstacles present in the environment may be static or dynamic. However, in this paper, we have considered static obstacles in the given environment for the robots, and the robot is treated as a dynamic obstacle for other robots. The path planning problem for multi-robot can be solved by two different approaches, such as centralized or distributed approach. The cost or objective function and the constraints for computing the path for all the robots are considered together in the centralized approach [25,26], whereas in the distributed planning [27] each robot determined its collision free trajectory path independently without making collision with static obstacles or colleagues at the time of moving toward the destination. The multi-robot navigational problem has divided into two smaller problems: velocity planning and path planning. In the first phase, each robot constructs the individual path by satisfying the optimum path for each robot. In the velocity planning, each robot avoids the collision with obstacles and the teammates. Many researchers are using the multi-robot navigational problem as a meta-heuristic optimization problem, and different meta-heuristic optimization algorithms have been used to generate the optimum trajectory collision free path for each robot, such as genetic algorithm (GA), particle swarm optimization (PSO) [28,29], and differential evolution (DE) [26].

In our research, we have integrated reinforcement learning techniques with an improved particle swarm optimization (IPSO) hybridized with DV to compute the optimal trajectory path of all the robots from specified initial positions to fixed goal positions in the cluttered environment and with an objective to minimize the path distance for each robot. In this paper, we enhance our implementation of IQ value-based IPSO-DV algorithm to determine the trajectory of path for multiple robots from predefined initial positions to predefined target positions in the environment with an objective to minimize the path length of all the robots. The result shows that the algorithm can improve the solution quality in a reasonable amount of time. This paper contributed to improve the classical Q-learning algorithm for improving the global path planning problem of the multi-robots by integrating it with IPSO-DV and improve the convergence rate, and performance metrics are evaluated in terms of path deviation, path traveled, number of turns and total time required to reach the destination. Finally, the efficiency of the IQ value-based IPSO-DV will be proven through the simulation as well as the Khepera robot, and the results are compared with other evolutionary computing, such an IPSO-DV, IPSO and DE.

The remaining part of the paper is outlined as follows. Problem formulation for the multi-robot navigation has been elaborated in section 2. Classical Q-learning and its limitation is introduced in section 3. Classical Q-learning has improved based on the proposed properties called improved Q-learning, and overcomes the limitation of the classical Q-learning, as introduced in section 4. The algorithm for the improved Q-learning is presented in section 5. The classical particle swarm optimization and improved particle swarm optimization are described briefly in section 6. A differential evolution algorithm is presented in section 7. Theoretical description and its algorithm of the hybrid IPSO-DV for path planning of multi-robot is presented in section 8. The QIPSO-DV algorithm-based multi-robot path planning is given in section 9. Implementation of hybrid QPSO-DV and performance analysis is briefly described in section 10. Section 11 provides the experimental result with the Khepera II robot. Conclusions are listed in section 12.

2. Problem formulation for multi-robot navigation

The multi-robot navigation problem is formulated as to compute the next location for each robot from its current location in the environment by avoiding collision with teammates (which is dynamic in nature) and obstacles (which are static in nature) in its path to reach the goal. The set of principles is considered in formulating multi-robot path planning problem with the help of the following assumption:

Assumptions

1. Current position/initial position and goal positions/target position of all the robot is known in prior coordinate system.
2. At any instant of time, the robot can decide any action from a set of predefined actions for its motion.
3. Each robot is performing its action until reaching their respective target position in steps.

The following principles have been taken care of for satisfying the given assumptions.

1. For determining the next position from its current position, the robot tries to align its heading direction toward the goal position.
2. The alignment may cause a collision with the robots/obstacles (which are static in nature) in the environment. Hence, the robot turns its heading direction with a certain angle either to the left or right for determining its next position from its current position.
3. If a robot can align itself with a goal without collision, then it will move to that determined position.
4. If the heading direction is rotated to the left or right, then it is required for the robot to rotate the same angle about its z-axis; if it is the same for more than one, then decide randomly.