Path Planning Algorithm based on Partially POMDP Model

Zhang Yali

*Information Engineering Department，Henan Mechanical and Electrical
Vocational College*

Abstract

To address the problem that reactive navigation is prone to local optimality under uncertain and complex environments, a POMDP-based global path planning algorithm is proposed for mobile robots. A 6-tuple model is constructed for path planning under complex dynamic environments, and the global optimality is realizes by maximizing the accumulative reward function. State transition function and observation function are used to handle unknown obstacles and noisy perception by modeling the error probability. Belief state space is introduced, and a value iteration algorithm using point-based policy tree pruning is developed to solve for real time planning policy, which effectively reduces the computational complexity. Simulation results show that using this algorithm the robot can automatically adapt to different probing granularities, avoid obstacles under complex uncertain environments, and achieve the optimal paths.

Keywords: Mobile robot, path planning, grid method, POMDP, belief state, policy

1. Introduction

Path planning of autonomous mobile robot is one of the recent research hotspots at home and abroad. Currently, artificial potential field [1] and reactive navigation are the major path planning methods. In particular, artificial potential field takes environment as a virtual artificial force field. Target point generates gravitation while obstacle generates repulsion. Under the control of resultant force between gravitation and repulsion, the robot can avoid the obstacle to reach the target point [2,3]. Its method and principle are simple and easy to realize real-time control for bottom layer. However, it is difficult to adapt to the dynamic changes of environment. Moreover, there are such problems as unreachable target, local minimum and oscillation. Therefore, the method of building reactive navigation controller based on fuzzy control and neural network methods are emerging. Simulated annealing algorithm is adopted to conduct cost function search and realize robot path planning. However, reactive navigation lacks a global cognition for the environment, can hardly achieve cumulative global optimum and cause the robot to easily get into local trap in a complex environment.

To solve the partial trap problem of reactive navigation in an uncertain environment, this paper proposes a robot path planning algorithm based on Partial Observable Markov Decision Processes (POMDP), adopts six-tuple of POMDP for modeling of robot path planning in the dynamic and uncertain environment and applies cumulative reward function to maximally realize the global optimum of path planning. Aiming at the problems that the obstacles are unknown and there might be deviations in robot perception, belief state space is introduced, and value iteration algorithm based on point pruning policy tree is adopted to get real-time planning strategy and effectively reduce the algorithm complexity.

Information Engineering Department，Henan Mechanical and Electrical Vocational College, China
*Corresponding author: zhang20021026@163.com

2. Path Planning Model of MDP

2.1. Path Planning Model of POMDP

The path planning of autonomous mobile robot needs to face the dynamic and uncertain environment with unknown obstacles, conduct planning and avoid obstacles according to the sensor's perception for the environment, and find out an optimal and collision-free path from starting point to target point. The path planning of autonomous mobile robot in a complex environment needs to solve multiple problems:

(1) Mostly, the path planning of mobile robot adopts reactive navigation, lacks a global cognition for the environment and easily gets into local optimum.

(2) Being influenced by such factors as sensor precision and environmental noise, there are errors in robot's perception for the environment.

(3) In running process of the system, there are uncertainties in robot movement. That is, the robot cannot always complete the set actions accurately, and there are action failures. Therefore, the path planning of autonomous robot is actually about the optimal decision problem in a dynamic and uncertain environment [4]. POMDP is an ideal mathematical model to solve this problem[5].

According to the prior information of the system, POMDP conducts modeling for intelligent agent through environment observation, possibly executed action and the action's effect on environment. Globally optimal strategy can be obtained by getting the maximum value of cumulative reward. It can be described with a six-tuple $<S, A, Z, T, R, O>$ in which S is state set, A is action set, Z is observation set, and $T(s, a, s')$ is state transition function which means the probability of reaching the state s' by executing the action a from the state s. $R(s, a)$ is reward function which means the reward obtained by executing the action a in the state s. $O(s', a, z)$ is observation function which means the probability of reaching the state s' and obtaining the observation set z by executing the action a.

2.2. Modeling of State Transition Function

The positions of robot and obstacles are defined through 2D grid map. The state set S is used to describe the positions of robot and obstacles. At any moment, the state of each grid might be one of the following four types: being occupied by the robot, being occupied by the obstacle, being blank and terminal.

Definition 1 Suppose the environment is divided into N grids. The state set of robot path planning is defined as:

$$s_i = \begin{cases} s\_r & (grid\ is\ robot) \\ s\_o & (grid\ is\ obstacle) \\ s\_b & (grid\ is\ blank) \\ s\_t & (grid\ is\ terminal) \end{cases} \quad (1 \le i \le N) \tag{1}$$

Where, i is the number of grid where the robot is located, $S = \{s_1, s_2, \text{L } s_N\}$

Definition 2 Suppose the robot can move at four directions in the environment. The executable action set for the robot that is described by the action A is as below:

$$A = \{East, South, West, North\} \tag{2}$$

The state transition function $T(s, a, s')$ means the probability of reaching the state s' by executing the action a in the state s. That is, $T(s, a, s') = P(s' | s, a)$. When the robot executes a certain action, the expected effect might not be fully achieved, and there is a failure probability to some extent. In a practical environment, it can be obtained from statistical data. In a simulation environment, suppose the probability for the robot to successfully execute the action is δ, and the probability for the robot to fail to execute the action, move toward the opposite direction or remain still is $1 - \delta$.

2.3. Modeling of Observation Function

Use the observation set z to indicate robot's cognition for the environment. That is the result for detecting the positions of robot and obstacles by the sensor. Suppose the detection state for robot for the grids includes being occupied by the robot, being occupied by the obstacle, being blank, terminal and unknown state. In particular, the unknown state means the undetected grids.

Definition 3 Suppose the environment is divided into N grids. The observation set of robot for the environment is defined as:

$$z_i = \begin{cases} z\_r & \text{the detected grid is robot} \\ z\_o & \text{the detected grid is obstacle} \\ z\_b & \text{the detected grid is blank} \qquad (1 \le i \le N) \\ z\_t & \text{the detected grid is terminal} \\ z\_u & \text{the detected grid is unknown} \end{cases} \qquad (2)$$

Where, i is the number of grid where the robot is located, $Z = \{z_1, z_2, \text{L } z_N\}$

Being influenced by such factors as sensor precision and environmental noise, there might be errors between the grid state detected by the robot and the actual state. Such errors can be expressed by the observation function $O(s',a,z)$. $O(s',a,z) = P(z\,|\,s',a)$. In a simulation environment, suppose the probability for the robot to successfully detect the grid state is μ, and the probability for detection failure is $1-\mu$. To simplify the model complexity, the grid state is uniformly set as z_u when the detection fails.

2.4. Modeling of Reward Function

To prevent path planning from getting into local optimum, immediate $R(s,a)$ and cumulative reward $V_\pi(s)$ for path/action need to be concerned. $R(s,a)$ means immediate reward obtained from executing the action a in the state s. For a certain state s, the state after robot movement includes collision, being in blank grid and arriving at terminal. Set the reward obtained from robot movement in blank grid as a negative incentive, which can prevent the robot from wandering in the blank grid. Therefore, set the immediate rewards for reaching these three states as -100, -1 and 100.

To achieve the global optimum of path planning, obtain the shortest path by maximizing the cumulative reward. The cumulative reward $V_a(s)$ is obtained through Bellman equation:

$$V_t(s) = R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V_{t-1}(s') \qquad (3)$$

Where, γ is discount factor and its target is to have expected value converged.

$$V_t(s) = R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V_{t-1}(s') \qquad (4)$$

Due to the uncertainty of mobile robot for the environment observation, the environment for the robot is not always confirmed to be a certain state s but there are multiple states distributed in form of probability. Therefore, the belief state space B is introduced to indicate the probability distribution for the states observed by the robot in the practical environment state. The belief B at t is described by Formula (5) and calculated by Formula (6).

$$b_t = P(s_t \,|\, a_t, z_t, a_{t-1}, z_{t-1}, a_{t-2}, z_{t-2}, ..., a_0, z_0, s_0) \qquad (5)$$

$$b'(s') = P(s' \,|\, z, a, b)$$
$$= \frac{O(s',a,z) \sum_{s \in S} T(s,a,s') b(s)}{P(z,a,b)} \qquad (6)$$

The functions based on belief state and cumulative reward can be rewritten by Formula (4) as:

$$V_t^*(b) = \max_a [\sum_{s \in S} b(s)R(s,a) + \gamma \sum_{z \in Z} P(z|b,a)V_{t-1}^*(b)] \tag{7}$$

It can be seen that the process of getting maximal strategy is an iteration process of obtaining the maximum function. The iteration ending condition is shown in Formula (8):

$$\|V_t(b') - V_{t-1}(b)\| < \frac{\xi(1-\gamma)}{2\gamma} \tag{8}$$

The strategy of mobile robot's POMDP model decision is to map the state to the action. That is $\pi(s) \to a$. To obtain the globally optimal solution, the optimal strategy π^* is the action to reach the maximal expected value for discount cumulative reward and can be solved by Formula (9).

$$\pi_V^*(b') = \arg\max_a [\sum_{s \in S} b(s)R(s,a) + \gamma \sum_{z \in Z} P(z|b,a)V_{t-1}^*(b)] \tag{9}$$

3. Solution Algorithm Based on Point Pruning Policy Tree

Modeling of POMDPs is conducted by robot path planning. Robot path planning is transferred into the solving problem for maximal POMDP cumulative reward.

3.1. Analysis for Algorithm Complexity

It can be known from POMDP model of robot path planning that the scales of state space, action and observation space are $|S| = N^4$, $|A| = 4$ and $|Z| = N^5$ in which N is the number of grids. Suppose the number of belief state spaces at $t-1$ is $|\Gamma^{t-1}|$, the complexity of iteration process of the value is $O(|S|^2 |A||Z||\Gamma^{t-1}|^{|Z|})$. In $|\Gamma^{t-1}| = |A|^{\frac{|Z|^{T-1}-1}{|Z|-1}}$, T is the times span from 0 to t.

Therefore, to solve POMDP optimal strategy for robot path planning is a typical issue about "curse of dimensionality" [6]. To reduce the algorithm complexity, this paper proposes a value iteration solution algorithm based on point pruning policy tree.

3.2. Value Iteration Solution Algorithm based on Point Pruning Policy Tree

A policy tree is composed of actions and observations. Each node means the action executed by the robot, and each branch means the observation obtained after executing the action.

Each belief point is traversed according to the structure of policy tree. The optimal value function and optimal strategy are obtained by Formula (6) and (8). As belief points and the scale of policy tree increase in an exponential order according to the observation number, the policy tree needs to be pruned. The idea of algorithm based on point pruning policy tree lies in the following: In the process of policy tree generation, select the belief points with partial characteristics to replace the whole belief state space at $t-1$, obtain the policy tree q_{t-1} with the maximal value function in this belief state space, add the optimal sub-policy tree set and record it as Q_{t-1}. In the iteration at t, sub-policy tree generation will not traverse all actions and observations but traverses the optimal sub-policy trees of Q_{t-1} to achieve quick solution.

Table 1. Solution Algorithm

Algorithm 1: POMDP solution algorithm based on point pruning policy tree
1) for all b_t in B
2)   for all a_t in A
3)   for all z_t in Z
4)   get and choose b_t by z_t, a_t, b_{t-1}
5) $B' = B \cup b_t$
6) for all b_t in B'
7)   for all a_t in A
8)   for all q_t in Q_t
9) $V_{q_t}(b) = \sum_{s \in S} b(s) V_{q_t}(s)$
10) $V_t(b) = \max_{q_t \in Q} V_{q_t}(b)$
11) $\pi_t = \arg\max_{q_t \in Q} V_{q_t}(b)$
12) $t++$
13) return 1) until $\| V_t(b) - V_{t-1}(b) \| < \zeta$

4. Simulation and Result Analysis

4.1. Simulation Environment and Parameter Setting

When Matlab is used to set up the simulation environment and set robot's executing actions, the probability of successful execution δ is 0.90, and the probability of execution failure is 0.10. In case of action execution failure, it will remain at the original position. When the grid is non-obstacle, the probability for the robot to successfully detect the grid state is 1.0. When the grid is an obstacle, conduct three groups of experiments by adopting the probabilities for the robot to successfully detect the obstacles (1.0, 0.8 and 0.6) so as to verify the algorithm efficiency when the environment is partial observation for the robot. The discount factor γ is 0.85.

4.2. Environment Simulation of U-Shaped Obstacles

Aiming at the problems that some path planning algorithms usually get into local optimum in concave obstacles and the path planning cannot be conducted successfully, simulation is conducted in the environment of U-shaped obstacles in this paper. Simulate a 5×6 area and divide the environment into 30 1×1 grids. Starting point and terminal are fixed. The environment and experimental results are shown in Figure 1.

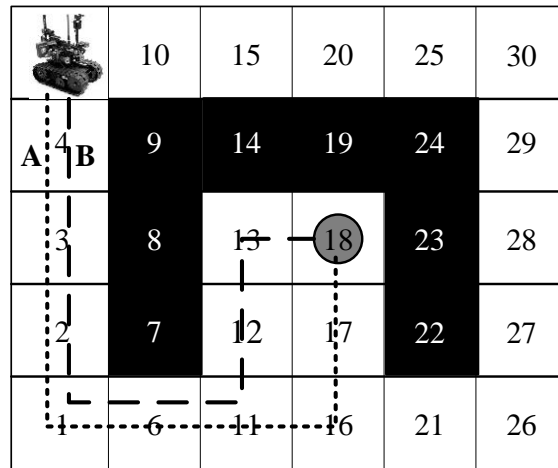**Figure 1. Path Planning Experiment of U-Shaped Obstacles**

In 30 simulation experiments, when μ is 1.0, 0.8 or 0.6, the robot can successfully plan path A or path B. The probabilities of path A and path B are 0.533 and 0.467. The results show that when the environment is partial observation for the robot, the robot can still plan an optimal and collision-free path from starting node to target node in U-shaped environment. This proves that the algorithm in this paper features strong robustness and optimizing capacity.

4.3. Environment Simulation of Random Obstacles

To prove the universality of the algorithm in this paper, a larger and more complex environment is generated at random for path planning. Simulate a 10×10 area and divide the environment into $100\ 1 \times 1$ grids. Starting point and terminal are fixed. Each grid is placed with obstacles at random, as shown in Figure 2.
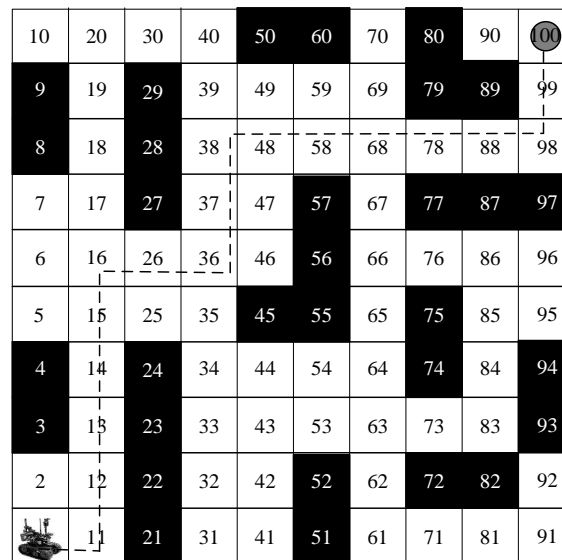

Figure 2. Experiment of Random Environment

μ is 1.0, 0.8 and 0.6, 10 simulation experiments are conducted. To simplify the complexity of path planning, when the reward values are the same, the paths without

swerving are firstly selected, and the experimental results are shown in Table 2.

Table 2. Results under Different Detection Accuracy Rates

μ	Path No.
1.0	2, 12, 13, 14, 15, 16, 26, 36, 37, 38, 48, 58, 68, 78, 88, 98, 99, 100
0.8	2, 12, 13, 14, 15, 16, 26, 36, 37, 38, 48, 58, 68, 78, 88, 98, 99, 100
0.6	2, 12, 13, 14, 15, 16, 26, 36, 37, 38, 39, 49, 59, 68, 78, 88, 98, 99, 100

It can be seen from the data of planned path algorithm that as the accuracy rate of robot detection reduces, the robot's planed paths are somewhat different but can reach the targets. When μ is 1.0 and 0.8, the globally optimal paths are obtained by the robot according to the planning. When it is reduced to 0.6, it can be seen that 39, 49, 59 and 69 are the grids with lower collision rates although the planed paths are relatively weaker than the optimal paths. This has ensured the collision-free path planning under relatively low detection accuracy rate, and the robot can adapt to a more complex environment.

5. Conclusion

Aiming at the problems that the calculated amount of path search of mobile robot is huge and it gets into the local optimum easily in the dynamic and complex environment, this paper cites the probability theory and proposes a global path planning algorithm based on POMDP model. Aiming at the inaccurate detection for unknown obstacles and uncertainty of robot movement, the prior information of the system is used to set up state transition model and observation model. Globally optimal solution is obtained by value iterative algorithm. The method based on belief point pruning policy tree is adopted to reduce the algorithm complexity. Simulation experiments are conducted in multiple complex environments to realize the optimal and collision-free path planning from starting node to target node and verify the algorithm efficiency, robustness and universality.

Referencea

[1] E.S.H. Hou and D. Zheng, "Mobile robot path planning based on hierarchical hexagonal decomposition and artificial potential fields", J of Robot. Sys., vol. 11, no. 7, **(1994)**, pp. 605-614.

[2] L. C. A. Pimenta, A. R. Fonseca and G. A. S. Pereira, "Robot Navigation Based on Electrostatic Field Computation", IEEE Tran. on Mag., vol. 42, no. 7, **(2006)**, pp.1459-1462.

[3] L. Xie, H. Chen and G. Xie, "Artificial Potential Field Based Path Planning for Mobile Robots Using Virtual Water-Flow Method", Comm. in Comp. and Infor. Sci., vol. 2, no. 2, **(2007)**, pp.588-595.

[4] M. A A, "Managing the Dynamics of a Harmonic Potential Field-Guided Robot in a Cluttered Environment", Indu. Ele. IEEE Tran. on, vol. 56, no. 2, **(2009)**, pp. 488-496.

[5] M. L. Littman, "A tutorial on partially observable Markov decision processes", J. of Math. Psy., vol. 53, no. 3, **(2009)**, pp. 119-125.

[6] S. Ross, J. P Pineau and S. Paquet, "Online Planning Algorithms for POMDPs", J. of Arti. Int. Res., vol. 32, no. 2, **(2008)**, pp. 663-704.

Author

Yali Zhang, she is a lecture in Henan Mechanical and Electrical Vocational College.Her major research interests includes software engineering and compilation technology,Intelligent information processing, *etc*.Research direction:Computer analysis.Data mining