Yang CHEN, Dianxi SHI, Huanhuan YANG, Tongyue LI, Zhen WANG, 2024. An anti-collision algorithm for robotic search-and-rescue tasks in unknown dynamic environments. *Frontiers of Information Technology & Electronic Engineering,* 25(4):569-584. https://doi.org/10.1631/FITEE.2300151

An anti-collision algorithm for robotic search-and-rescue tasks in unknown dynamic environments

Key words: Search and rescue; Reinforcement learning; Game theory; Collision avoidance; Decision-making

Corresponding author: Dianxi SHI
E-mail: dxshi@nudt.edu.cn
ORCID: https://orcid.org/0000-0002-8112-371X

Motivation

1. Online autonomous exploration in unknown dynamic environments attracts significant interest, particularly in the context of search-and-rescue planning. When multiple objectives need to be collected in a dynamic environment, the execution results would be poor.

2. When reinforcement learning is used to prevent collisions, the effectiveness and efficiency of motion planning heavily rely on the design of the sample selection policy.

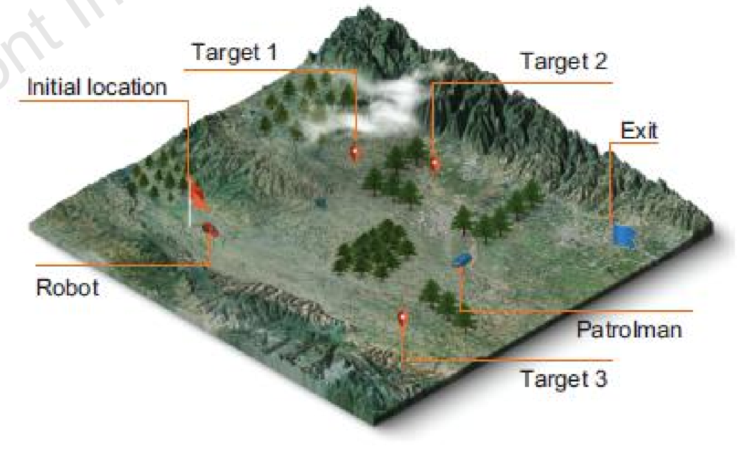


Fig. 1 Robot search-and-rescue task

Main idea

1. We develop a multi-objective layered structure with the aim of simplifying the complexity of the multi-objective problem. This approach allows us to reduce the computational complexity of the overall problem.

2. We propose a risk-monitoring mechanism that relies on the relative position of dynamic risks to aid in decision-making. This mechanism assists in creating an intuitive environment model.

3. To maximize rewards, we incorporate the principles of game theory into our approach. We introduce a method called MNDQ, which combines reinforcement learning (RL) with Dyna-Q and mixed-strategy Nash equilibrium (MNE).

Method

The method integrates the Dyna architecture with three key technologies: multi-objective layered structure, risk-monitoring mechanism, and mixed-strategy Nash equilibrium (MNE) policy

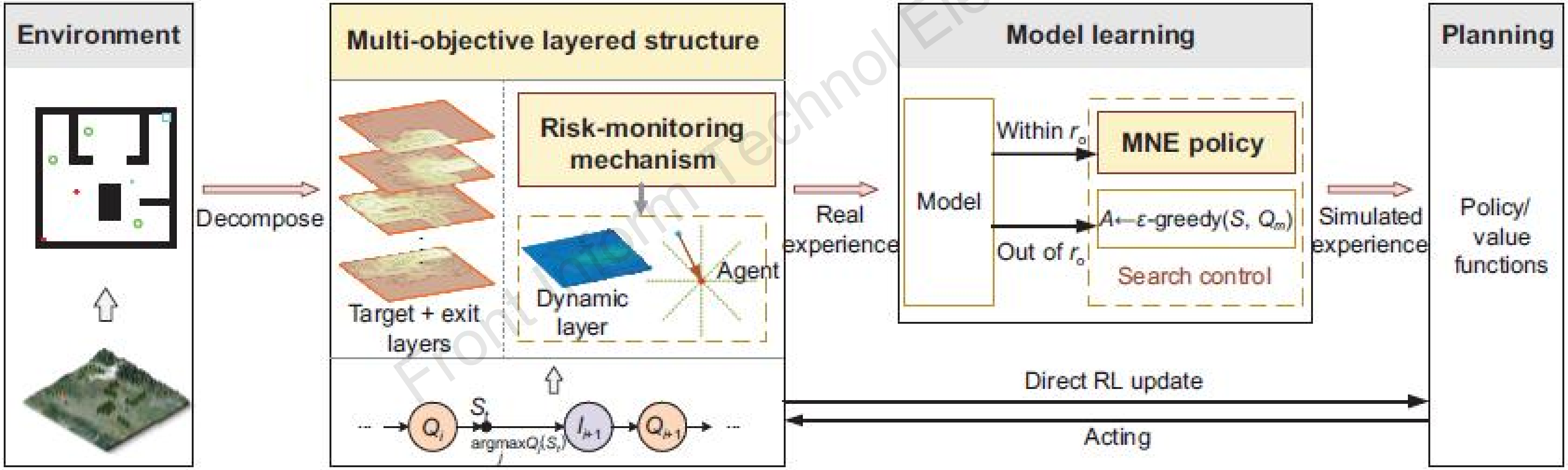


Fig. 4 The general framework of the mixed-strategy Nash equilibrium based Dyna-Q (MNDQ) algorithm. The framework integrates the Dyna architecture with three key technologies: multi-objective layered structure, risk-monitoring mechanism, and mixed-strategy Nash equilibrium (MNE) policy (RL: reinforcement learning)

Method (Cont’d)

1. Multi-objective layered structure

We design a multi-objective layered structure to simplify these multi-stage tasks. The task model is layered to improve learning accuracy and efficiency. The overall task is decomposed into a group of discrete subtasks associated with each other through the multi-objective layered structure.

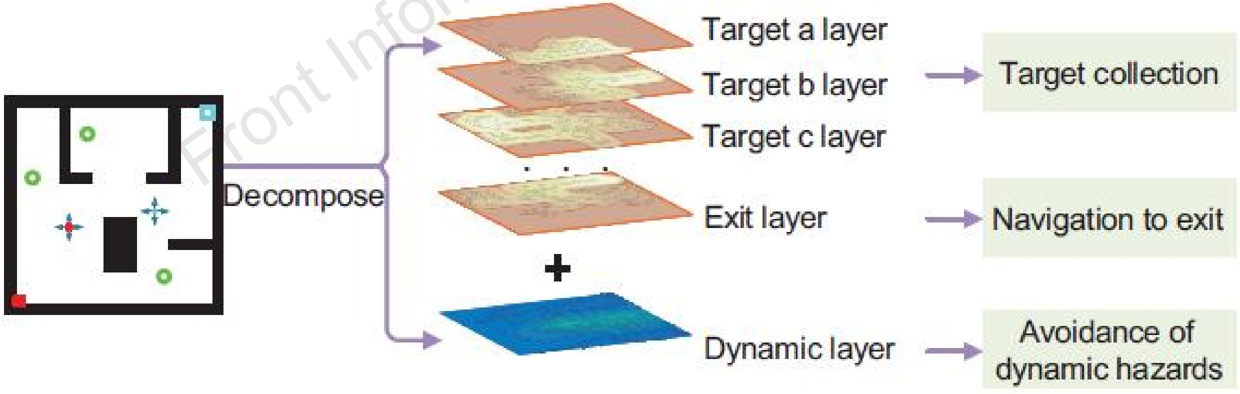


Fig. 7 Multi-objective layered structure

Method (Cont’d)

2. Risk-monitoring mechanism

A risk-monitoring mechanism is proposed to provide a clearer representation of the impact of a patrolman. This mechanism is based on the assessment of the relative positions of the dynamic risks. The state of a dynamic risk is expressed using the Manhattan distance and relative orientation.

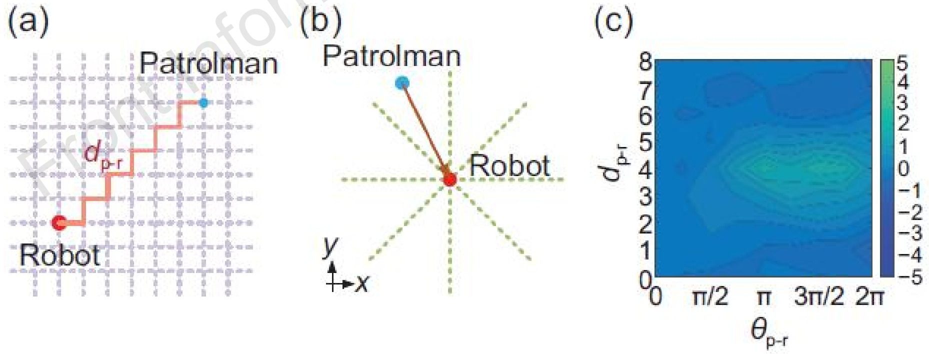
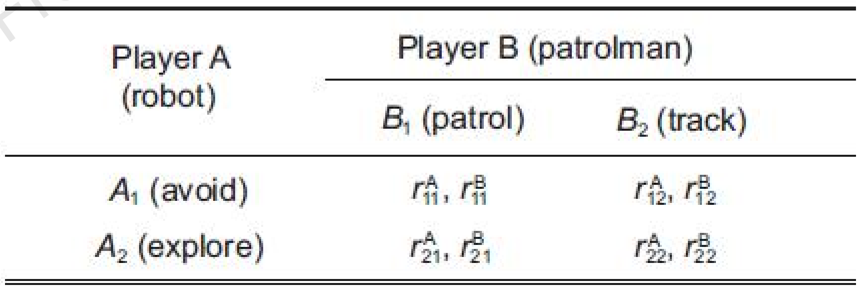


Fig. 8 Manhattan distance $d_{\mathrm{p-r}}$ (a), relative orientation $\theta_{\mathrm{p-r}}$ (b), and the contour plot of $\max_{A} Q_{\mathrm{e}}(S_{\mathrm{e}})$ (c)

Method (Cont’d)

3. The mixed-strategy Nash equilibrium policy

Simulated transitions in Dyna-Q are started in state–action pairs selected uniformly and randomly from all previously experienced pairs. The random selection policy may slow down the convergence rate. Experience and samples should be focused on specific state–action pairs. To prevent the only action from being predicted by the patrolman and to encourage the robot to explore potential solutions, we propose an MNE policy to select an action in the form of probability distribution.

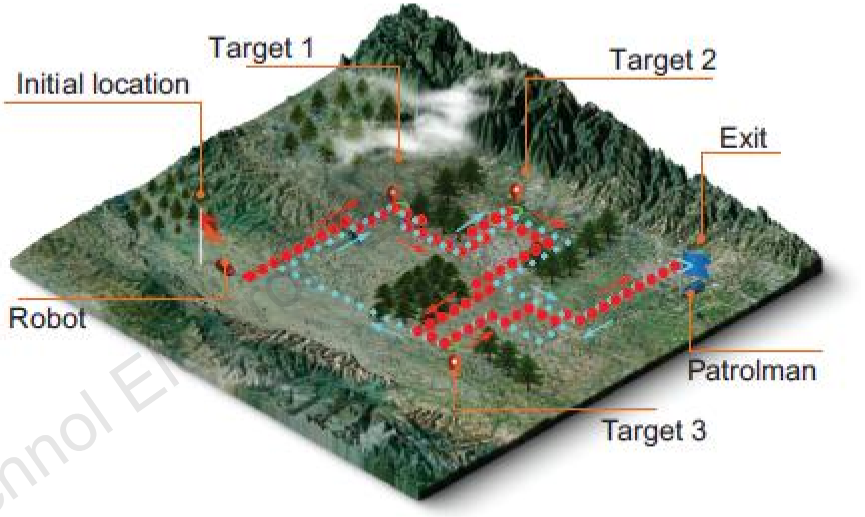


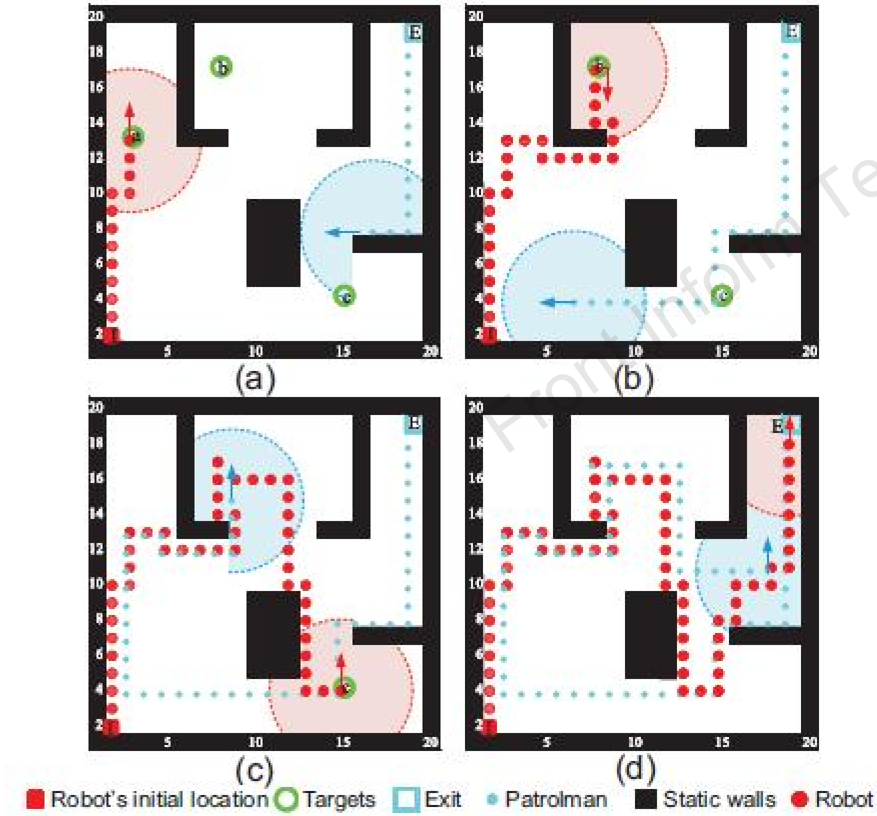
Player A (robot)	Player B (patrolman)	
	B_1 (patrol)	B_2 (track)
A_1 (avoid)	r^A_{11}, r^B_{11}	r^A_{12}, r^B_{12}
A_2 (explore)	r^A_{21}, r^B_{21}	r^A_{22}, r^B_{22}

Fig. 9 Structure of the game played between players A and B

Major results

The complete trajectory in an unknown environment map with obstacles and a patrolman:

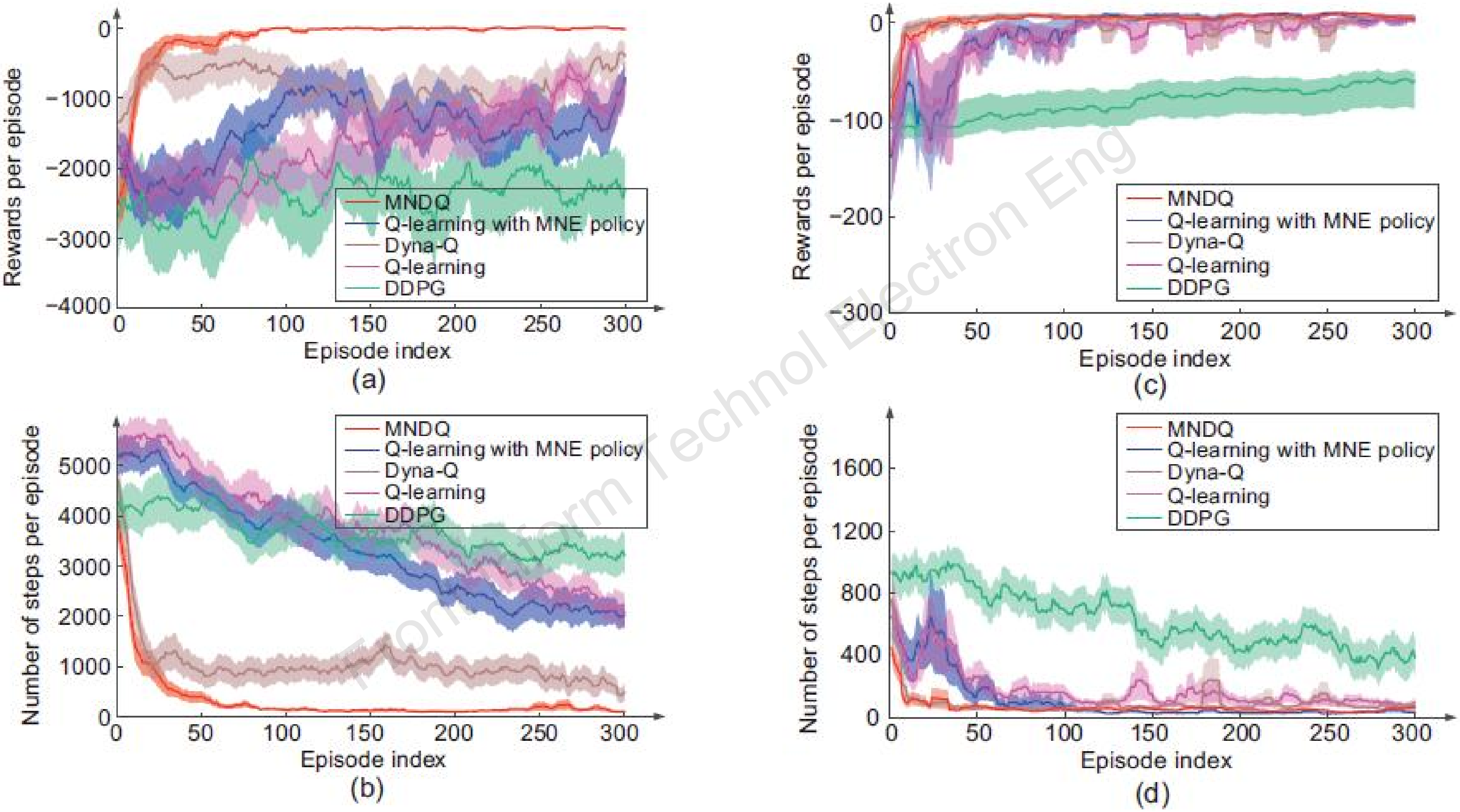




(a) (b) (c) (d)

Planned trajectory with static obstacles: (a–c) current positions of the robot (solid red dot) and the patrolman (solid blue dot) when the robot collects the targets; (d) the robot reaching the exit.

Major results (Cont’d)



Average learning curves in the unknown dynamic environments with static obstacles: (a) rewards under the map size 20×20; (b) number of steps under the map size 20×20; (c) rewards under the map size 10×10; (d) number of steps under the map size 10×10

Conclusions

- In this paper, we investigated the robot search-and-rescue problem in unknown dynamic environments and proposed an MNDQ algorithm. The algorithm adopted a multi-objective layered structure to decompose the task into smaller learning subtasks.

- We proposed a risk-monitoring mechanism to help the robot generate a collision-free static trajectory. These mechanisms were designed to express and simplify the tasks.

- Furthermore, we combined the knowledge of the game theory and designed the MNE policy to choose better specific state–action pairs and enhance the sample efficiency. The policy enabled the agent to make decisions in the form of probability to maximize the expected rewards and improved the performance of Dyna-Q.

Yang CHEN received her M.S. degree from University of Chinese Academy of Sciences, Beijing, China. She is currently pursuing her Ph.D. degree with the School of Computer Science, Peking University, Beijing, China. Her current research interests include reinforcement learning and artificial intelligence.

Dianxi SHI received the B.Sc., M.Sc., and Ph.D. degrees in computer science from the National University of Defense Technology, Changsha, China, in 1989, 1996, and 2000, respectively. He is currently a Professor with the Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, Beijing, China. He has presided over and participated in the National 863 Project, the National Key Research and Development Plan, the National Natural Science Foundation, the Major Projects of Core Electronic Devices, High-end Generic Chips, and Basic Software more than 20 times. His research interests are in distributed object middleware technology, adaptive software technology, artificial intelligence, and robot operation systems. Dr. SHI was a recipient of second in the National Science and Technology Progress Awards twice and first in provincial-level scientific and technological progress awards three times.