文章编号:1671-4598(2022)11-0192-07 DOI:10.16526/j.cnki.11—4762/tp.2022.11.028 中图分类号:TP391.9 文献标识码:A

基于Q学习的水下滑翔机路径规划方法

张鑫菠，李 乐，冀海军，彭星光
（西北工业大学 航海学院，西安 710072）

摘要：针对水下滑翔机路径规划问题，提出了一种基于Q学习的水下滑翔机路径规划方法；考虑到水下滑翔机在执行一些特定任务时会提前给定俯仰角及深度参数，且航向角选择范围通常是几个离散角度值，文章针对典型的几种俯仰角情况分别设计了航向动作选择集，这避免了Q学习方法“维数爆炸”问题；根据水下滑翔机航程最短的目标和障碍物外部约束条件，设计了奖励函数与动作选择策略；相较于传统路径规划方法，文章提出的方法不需要提前知道环境信息，而是在学习过程中根据环境的反馈选择最优动作，因此该方法在不同的环境条件下有优良的迁移能力；仿真结果表明，该方法能在未知环境中为水下滑翔机规划出规避障碍且航程短的路径。

关键词：水下滑翔机；路径规划；Q学习方法

Method of Underwater Glider Path Planning Based on Q-Learning

ZHANG Xinbo, LI Le, JI Haijun, PENG Xingguang
(Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Aiming at the problem of underwater glider path planning, an underwater glider path planning method based on Q-Learning is proposed. Considering that underwater gliders will give the parameters of pitch angle and depth in advance as performing some specific tasks, and the selected range of azimuth angel is usually several discrete angle values, the selection sets of azimuth action for several typical pitch angles are designed respectively, which avoids the problem of "dimension explosion" of Q-learning method. According to the shortest path target of underwater glider and the external constraints of obstacles, the reward function and action selection strategy are designed. Compared with the traditional path planning methods, the proposed method does not need to know the environmental information in advance, but the optimal action is selected by the environmental feedback in the learning process. Therefore, this method has the excellent migration ability under the different environmental conditions. The simulation results show that this method can plan the obstacle avoidance and short route for underwater glider in unknown environment.

Keywords: underwater glider; path planning; Q-learning method

0 引言

水下滑翔机是一种通过调节自身重心、浮力实现下潜与上浮的新型水下航行器，其凭借低成本、低功耗、长续航等特点，已经成为了一种重要的海洋观测平台。水下滑翔机路径规划是指在任务区域中规划一条满足给定目标函数（例如路径长度最短等）并满足避开障碍物等约束条件的，从起点到目标点的最优路径。路径规划主要包括环境建模方法和路径搜索算法。水下滑翔机在执行海洋环境监测和海洋安全保障等任务时，都需要进行路径规划。因此对水下滑翔机路径规划方法进行研究具有重大意义。

在水下滑翔机路径规划方面，国内外学者做了大量研究。当前水下滑翔机路径规划方法分为两类：传统路径规划方法和基于仿生学的路径规划方法[1]。传统路径规划方法如人工势场法、快速步进法和A*算法等；基于仿生学的路径规划方法如粒子群优化算法、蚁群优化算法、狼群算法和人工神经网络算法等。人工势场法[2]借鉴了物理学中势场理论的概念，将障碍物的排斥力与导航目标的引力模型结合起来，虽然简单高效，但存在目标不可达、振荡等固有缺陷。快速步进法[3]使用一阶数值近似来求解方程，具有可靠性强、易执行等优点，但是建模过于简单，未充分考虑障碍物形状和滑翔机下潜深度等约束条件。传统A*算法[4]以当前所在点为中心，对其各方向上限定范围内的点作为拓展备选点进行遍历，搜索速度快，但实时性差、运行效率低。粒子群优化算法[5]是一种基于迭代的优化方法，通过迭代的方法在搜索域中寻找最优值，参数少、收敛速度快，但是粒子间信息交互少，容易陷入局部最优。蚁群优化算法[6]是基于模拟蚁群觅食行为寻找优化路径的一种自然估算算法，具有鲁棒性、优良的分布式计算以及

收稿日期:2022-03-28； **修回日期**:2022-05-13。
基金项目:国家自然科学基金项目(61903304);中央高校基本科研业务费项目(3102020HHZY030010);“111”引智计划项目(B18041.0)。
作者简介:张鑫菠(1998-),男,陕西汉中人,硕士研究生,主要从事水下机器人规划与控制等方向的研究。
彭星光(1981-),男,贵州贵阳人,教授,主要从事群体智能、演化计算、机器学习及其在无人系统的应用等方向的研究。
通讯作者:李 乐(1986-),男,陕西渭南人,讲师,主要从事水下机器人协同规划与控制等方向的研究。
引用格式:张鑫菠,李 乐,冀海军,等.基于Q学习的水下滑翔机路径规划方法[J].计算机测量与控制,2022,30(11):192-198.

易于与其他算法结合等优点，但是在进行全局路径规划时收敛速度慢，易陷入局部最优。狼群算法[7]是仿生自然界狼群的捕猎行为提出的一种群体智能优化算法，不易受参数影响，简单易实现，但是也存在收敛速度慢、易陷入局部极值、稳定性差及后期局部搜索能力差等缺点。人工神经网络算法[8]是一种受生物神经网络功能的运作启发而产生的算法数学模型，容错性强、自学能力强，但存在着路径错判、计算量大、学习滞后等问题。这些方法都需要提前建立环境模型，当环境信息不能提前获得或者问题模型过于复杂时，使用这些方法往往效果不佳。针对上述不足，强化学习方法成为一种用于水下滑翔机路径规划的研究方向。

强化学习方法不需要提前得知环境信息，也不需要预先给定任何数据，而是通过不断与环境交互，从环境中获得累积行为奖励的方式来更新模型参数，进而寻求最优解[9]，并且可以根据优化目标来设计奖励函数。这种方法自适应性强，运行效率高，能避免陷入局部最优解。因此适合于水下未知复杂环境中的滑翔机路径规划问题。

本文提出基于 Q 学习的水下滑翔机路径规划方法。首先，在水下滑翔机工作深度一定的条件下，对典型的几种俯仰角情况分别设计了航向动作选择集；其次，根据水下滑翔机路径最短的目标和障碍物外部约束条件，设计了奖励函数与动作选择策略，给出了 Q 值更新函数。最后，通过仿真实验验证了算法的有效性。

1　问题描述

传统的路径规划方法都需要提前建立环境模型，当环境信息不能提前获得或者问题模型过于复杂时，这些方法往往不能使用。目前 Q 学习路径规划方法多仅考虑二维水平动作变化。

综合考虑水下滑翔机的自身运动特性和滑翔过程水下环境的影响，水下滑翔机的路径规划需要满足以下约束条件。

1.1　路径总长度约束

水下滑翔机在水中能滑翔的总航程有限，规划出的从起点到终点的路径总长度不能超过水下滑翔机能滑翔的最大航程。

1.2　障碍物约束

水中环境未知复杂，会有各种岛礁、船舶等障碍物阻挡水下滑翔机的运动，规划路径要规避障碍物威胁。

1.3　滑翔参数约束

水下滑翔机在俯冲与上浮时，要满足自身运动的俯仰角、横滚角、偏航角约束以及航行深度等参数约束。

2　Q 学习方法概述

Q 学习是一种典型的无模型强化学习方法，利用迭代方法直接优化一个可迭代计算的 Q 函数，也是一种增量式的在线学习[10-12]，它迭代时采用状态－动作对的 $Q(S,a)$ 函数，其中 S 表示状态，a 表示动作。在 Q 学习方法中，智能体每一次学习迭代时需要考察每一个行为，确保学习过程收敛。

Q 学习首先初始化 $Q(S,a)$ 的值，基于当前状态 S_t，使用动作选择策略确定下一步动作 a_t，然后环境立刻反馈一个奖励值 r_t 与新的状态值 S_{t+1}，根据 Q 值更新函数更新对应状态 S_t 和动作 a_t 的值，然后基于新状态值 S_{t+1} 进行同样的步骤[13-15]。当新的状态值为目标状态或者满足结束条件时，完成一次循环迭代，继续从起始状态开始新的循环迭代，直到学习结束[16]。Q 学习本质上就是不断地优化可迭代计算的 $Q(S,a)$ 函数值，直到其收敛到最优值[17-19]。Q 值更新函数如下：

$$Q(S_t,a_t) \leftarrow (1-\alpha)Q(S_t,a_t)+\alpha[r_t+\gamma \max_{a_t} Q(S_{t+1},a_t)] \tag{1}$$

其中：S_t 是当前状态，a_t 是当前动作，r_t 是在 S_t 状态下选择动作 a_t 得到的奖励值，S_{t+1} 是执行动作 a_t 后的新状态，$0<\alpha<1$ 表示学习率，$0<\gamma<1$ 表示衰减率。

Q 学习过程包括多次迭代循环，每次迭代循环又包括多次动作 a 的选择与执行，动作 a 选择与执行步骤如下，假定当前为第 t 次动作选择：

1）已知当前状态 S_t；

2）根据动作选择策略，得到并执行动作 a_t；

3）得到奖励值 r_t 和新状态值 S_{t+1}；

4）更新 $Q(S_t,a_t)$ 的值；

5）进入第 $t+1$ 次动作选择。

3　Q 学习水下滑翔机路径规划方法

基于 Q 学习的水下滑翔机路径规划方法包括 Q 学习要素设计与算法流程两部分。Q 学习要素是实现基于 Q 学习水下滑翔机路径规划的基础。算法流程是实现路径规划的方法与步骤，具体内容在本节给出。

3.1　水下滑翔机 Q 学习要素设计

水下滑翔机 Q 学习要素主要由水下滑翔机状态、动作集合、奖励函数、动作选择策略、Q 值表五部分组成，如图 1 所示。水下滑翔机根据当前状态选择要执行动作的依据是动作选择策略。具体 Q 学习要素的设计在下文给出。

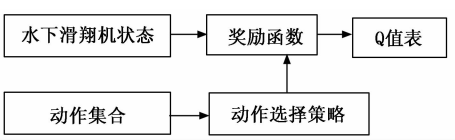


图 1　Q 学习要素表示

3.1.1　水下滑翔机状态表示

在解决路径规划问题中，对环境空间进行栅格化表示是一种常用的方法。本文用二维栅格表示水下任务环境的水平面。对环境空间的栅格化表示是为了在仿真实验中模拟环境对水下滑翔机动作的反馈，算法本身不需要提前建立环境模型。水下滑翔机在环境中的状态用该时刻滑翔机在栅格中的坐标点位置 S（x，y）表示。

3.1.2 动作集合表示

考虑到水下滑翔机在执行一些特定任务时会提前给定俯仰角 θ 及深度值 h，且航向角 ψ 的选择范围通常是几个离散角度值。因此假设给定深度值为 h，对 m 种典型的俯仰角 $\{\theta_1, \theta_2, \cdots, \theta_m\}$ 分别设计 n 种典型的航向动作选择集 $\{\psi_1, \psi_2, \cdots, \psi_n\}$，如图 2 所示。这些动作的步长由公式（2）～（4）得到：

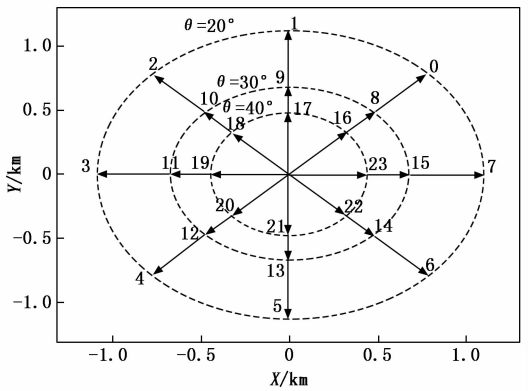


图 2 动作选择集合

$$x = \frac{2h}{\tan\theta} \cdot \cos\psi \tag{2}$$

$$y = \frac{2h}{\tan\theta} \cdot \sin\psi \tag{3}$$

$$step = \sqrt{x^2 + y^2} \tag{4}$$

其中：θ、ψ、h 分别表示俯仰角、航向角、深度，x、y 表示横、纵坐标，$step$ 表示步长大小。

3.1.3 奖励函数设计

Q 学习的目的是得到从起始点到目标点的最大累积奖励，此奖励是环境反馈得到，直接影响到学习的速度和效果，因此设计合理高效的奖励函数非常重要。本文设计的奖励函数如下。

当到达目标点：

$$\mathrm{Reward} = 500 \tag{5}$$

当碰撞到任何一个障碍物：

$$\mathrm{Reward} = -500 \tag{6}$$

当在自由区域运动时：

$$\mathrm{Reward} = \Omega[(x_{\mathrm{Aug}} - x_{\mathrm{Target}})^2 + (y_{\mathrm{Aug}} - y_{\mathrm{Target}})^2] \tag{7}$$

式中，x_{Aug}，y_{Aug} 是水下滑翔机在栅格环境中当前位置的横、纵坐标；x_{Target}，y_{Target} 是目标点的横、纵坐标；Ω 为缩放系数，用来调节自由区域运动时的奖励值范围，使其奖励值的绝对值远小于到达目标点或碰撞到障碍物时奖励值的绝对值，通常任务区域面积越大 Ω 取值越小。

3.1.4 动作选择策略

本文在 ε-greedy 方法基础上结合水下滑翔机运动特点，设计了一种新的动作选择策略。算法 1 给出了水下滑翔机 Q 学习动作选择策略算法流程。首先判断俯仰角 θ 的大小，θ 的选择范围为设定的 m 种典型俯仰角 $\{\theta_1, \theta_2, \cdots, \theta_m\}$。根据俯仰角大小，选择对应的 n 种典型的航向动作选择集 $\{\psi_{i1}, \psi_{i2} \cdots \psi_{in}\}$。然后设定一个贪婪值 ε，在每次选择动作时会随机产生一个 p，当 p 小于 ε 时随机选择动作值。当 p 大于 ε 时，则选择动作集合当中 Q 值最大者作为此次选择的动作。

算法 1：

IF $\theta=\theta_i, i\in(1,\mathrm{m})$

$a_i = \{\psi_{i1}, \psi_{i2}, \cdots \psi_{in}\}$

End IF

Generate p randomly, $\mathrm{p}\in(0,1)$

IF $p > \varepsilon$

$a = argmax(Q(s, a_i))$

Else

$a = Random(a_i)$

End IF

3.1.5 Q 值表初始化方法

在 Q 学习的初始学习阶段，系统对于环境状态基本一无所知[20]，对动作的选择都是随机盲目的，因此使用先验知识初始化 Q 值表对提高 Q 学习算法的收敛速度非常重要。

在水下滑翔机的路径规划问题中，本文使用栅格环境中坐标点与目标点的直线距离进行 Q 值的初始化，即式（5）所示，距离目标点越远的点初始 Q 值越小。这种根据与目标点距离初始化 Q 值表的方法，使得水下滑翔机具有向目标点移动的趋势，减小了搜索训练的盲目性，提高了 Q 学习方法的学习速度。

$$Q(S, a_i) = \frac{1}{(x_s - x_{\mathrm{Target}})^2 + (y_s - y_{\mathrm{Target}})^2} \tag{8}$$

3.2 基于 Q 学习的水下滑翔机路径规划方法流程

基于 3.1 节水下滑翔机 Q 学习要素设计，水下滑翔机的状态由位置坐标 S（x，y）表示。给定深度值 h，对典型的 m 种俯仰角 $\{\theta_1, \theta_2, \cdots, \theta_m\}$ 分别设计 n 种典型的航向动作选择集 $\{\psi_1, \psi_2, \cdots, \psi_n\}$，当水下滑翔机碰到障碍物时奖励函数值为 −500，到达目标点时奖励函数值为 500，在自由区域时奖励函数如公式（7）所示。算法 1 给出了动作选择策略。按照如图 3 所示具体流程图，实现基于 Q 学习的水下滑翔机路径规划。

具体步骤如下。

步骤 1：给定水下滑翔机初始位置 S，给定深度 h，俯仰角 $\theta=\theta_1$，航向动作集合 $\{\psi_1, \psi_2, \cdots, \psi_n\}$

步骤 2：根据 Q 值表初始化方法，给 Q 值表赋初值；

步骤 3：根据动作选择策略，从俯仰角对应的动作选择集中选择动作 a；

步骤 4：根据 3.1.3 节给出的奖励函数得到立即的奖励值 r 和执行动作 a 后的新的滑翔机位置 S'；

步骤 5：更新对应于状态 S 与动作 a 的 Q 值：

$$Q(S,a) \leftarrow (1-\alpha)Q(S,a) + \alpha[r + \gamma \max Q(S', a)] \tag{9}$$

步骤 6：判断水下滑翔机是否碰到障碍物，是则结束此

轮学习，转到步骤 1，否则继续进行下一步。

步骤 7：判断水下滑翔机是否到达目标区域，未到达则转到步骤 3，选择下一动作。若已到达目标区域，继续进行下一步。

步骤 8：判断是否完成 m 种不同俯仰角设定下的路径规划，是则输出 m 条路径中最短路径的规划路径点。否则重新给定俯仰角，进行新一轮学习。

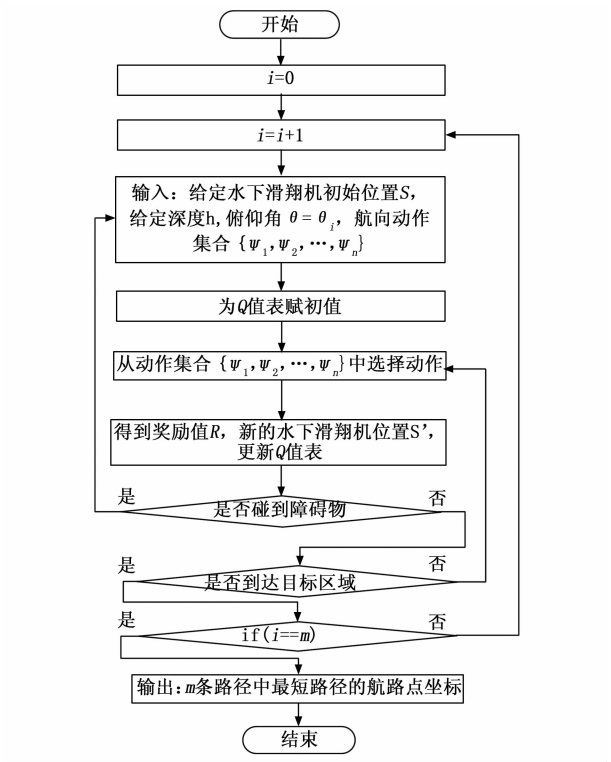


图 3　基于 Q 学习的水下滑翔机路径规划方法流程图

4　仿真实验及结果分析

4.1　仿真实验环境

水下滑翔机路径规划区域为 12 km×12 km 的二维栅格区域，表示水下任务环境，设置栅格粒度大小 Ngrid = 1 km。假设水下滑翔机的单周期最小滑翔距离为 0.3 km，滑翔深度为 0.2 km。在圆形障碍物环境、矩形障碍物环境与复杂障碍物环境 3 种规划区域中进行验证。在 Python 中建立水下滑翔机路径规划的初始环境空间，圆形障碍物、矩形障碍物与复杂障碍物信息在下文给出。

圆形障碍物环境中，起点与终点坐标分别为（0.5，0.5）km 和（10，10）km，4 个圆形障碍物半径均为 0.6 km，几何中心坐标分别为（1，3）km、（5，4）km、（5.5，6.5）km、（8.5，7.5）km；矩形障碍物环境中，起点与终点坐标分别为（0.5，0.5）km 和（10，9）km，3 个矩形障碍物边长均为 1.5 km，几何中心坐标分别为（1，3）km、（4，1）km、（5.75，5.75）km；复杂障碍物环境中，起点与终点坐标分别为（0.5，0.5）km 和（9.6，8.5）km，共有 6 个障碍物，其中 1 个圆形障碍物半径为 1 km，其余 5 个障碍物是由边长为 1 km 的矩形组合而成，几何中心坐标分别为（1，5）km、（2，10）km、（4，1）km、（5.5，4.5）km、（5.5，8.5）km、（9.5，3.5）km。

4.2　Q 学习水下滑翔机路径规划仿真

基于 Q 学习的水下滑翔机路径规划方法，根据滑翔机的当前位置坐标 S，在 Q 值表中找出对应当前位置坐标的最大 Q 值，进而执行此 Q 值所对应的动作 a，然后得到一个立即的环境奖励值 R 与新的位置坐标 S'，根据奖励值 R 更新对应于状态 S 与动作 a 的 Q 值，Q 值更新的方法如 3.1.3 节公式（5）（6）（7）所示。仿真参数设定为：$\alpha=0.1$，$\gamma=0.9$，$\varepsilon=0.1$，$\Omega=-0.000\,1$，学习最大次数为 3 000 次。

4.2.1　Q 学习水下滑翔机路径规划方法验证

水下滑翔机在特定任务中的路径规划需要提前设定俯仰角，俯仰角可选范围并非连续值，而是根据水下滑翔机自身性能、任务参数以及海洋状况确定的几个典型离散数值。因此有必要对多种俯仰角下所规划路径进行比较，以保证规划的路径最短。

给定深度值 $h=0.2$ km。设俯仰角个数 $m=3$，取值为 $\{\theta_1=20°,\ \theta_2=30°,\ \theta_3=40°\}$。设航向动作选择集包含动作个数 $n=8$，动作选择集合 {0，1，2，3，4，5，6，7} 表示当俯仰角为时，航向角 {315°，270°，225°，180°，135°，90°，45°，360°} 分别对应的动作。动作选择集合 {8，9，10，11，12，13，14，15} 表示当俯仰角时，航向角 {315°，270°，225°，180°，135°，90°，45°，360°} 分别对应的动作。动作选择集合 {16，17，18，19，20，21，22，23} 表示当俯仰角时，航向角 {315°，270°，225°，180°，135°，90°，45°，360°} 分别对应的动作。针对 20°、30°、40°3 种典型俯仰角取值，使用 3.1.2 中提出的动作集合，在复杂障碍物环境下分别规划。图 4 为 3 种俯仰角设定下的路径规划结果图，在不同的俯仰角设定下，均能规划出避免碰撞的安全路径。

图 5 为 3 种俯仰角设定下，每一轮学习所规划出的从起点到目标点的路径长度变化。在俯仰角 $\theta=20°$ 的路径规划中，经过 400 轮学习可以寻得最短路径，路径长度为 14.282 km。在俯仰角 $\theta=30°$ 的路径规划中，经过 700 轮学习可以寻得最短路径，路径长度为 15.218 km。在俯仰角 $\theta=40°$ 的路径规划中，经过 2 300 轮学习可以寻得最短路径，路径长度为 15.241 km。由于基于 Q 学习的水下滑翔机路径规划方法依靠不断探索学习来寻求最优解，因此在寻找最短路径后会继续学习，在图 5 中体现为最短路径与非最短路径接替出现，形成震荡。

表 1 记录了 3 种俯仰角设定下，最终规划路径的路径点个数、最短路径长度、与理论最短距离之差。理论最短距离指起点与目标点之间的直线距离。由表中数据知，在复杂障碍物环境中，设定俯仰角 θ 为 20°时，相较于 30°、40°，可以规划出长度最短、与理论最短距离差最小的路径。此时路径长度 14.282 km、与理论最短距离差为 2.165 km。

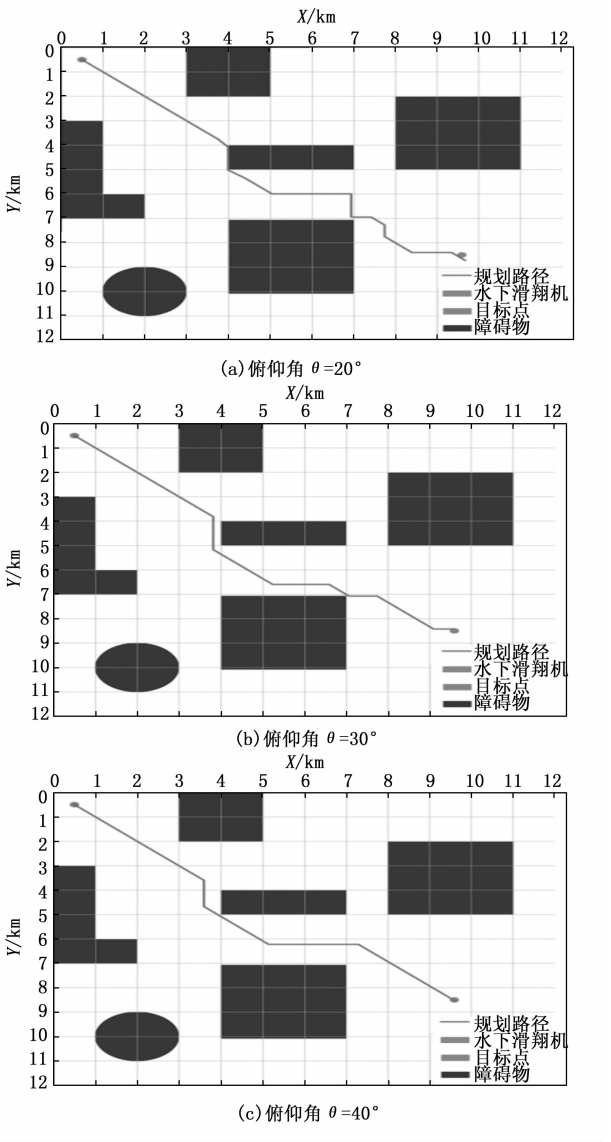


图 4 不同俯仰角设定下的路径规划结果

表 1 不同俯仰角设定下的路径规划结果

俯仰角 θ/°	路径点个数/个	最短路径长度/km	与理论最短距离之差/km
20	13	14.282	2.165
30	20	15.218	3.101
40	31	15.241	3.124

4.2.2 不同障碍物环境下的路径规划

由于水下环境复杂多样，针对圆形障碍物与矩形障碍物环境分别给出了 3 种不同场景，验证基于 Q 学习的水下滑翔机路径规划方法的完备性。图 6 是水下滑翔机在圆形障碍物初始环境中的路径规划仿真，图 6 中 a)、b)、c) 分别表示圆形动态障碍物几何中心在 (5，4) km、(3，3) km，(7.5，6.5) km 三个不同位置时的规划结果。图 7 是矩形障碍物初始环境中的路径规划仿真，图 7 中 a)、b)、c) 分别表示矩形动态障碍物几何中心在 (5.75，5.75) km、(8.55，6.75) km，(8.55，5.75) km 三个不同位置时的规划结果。

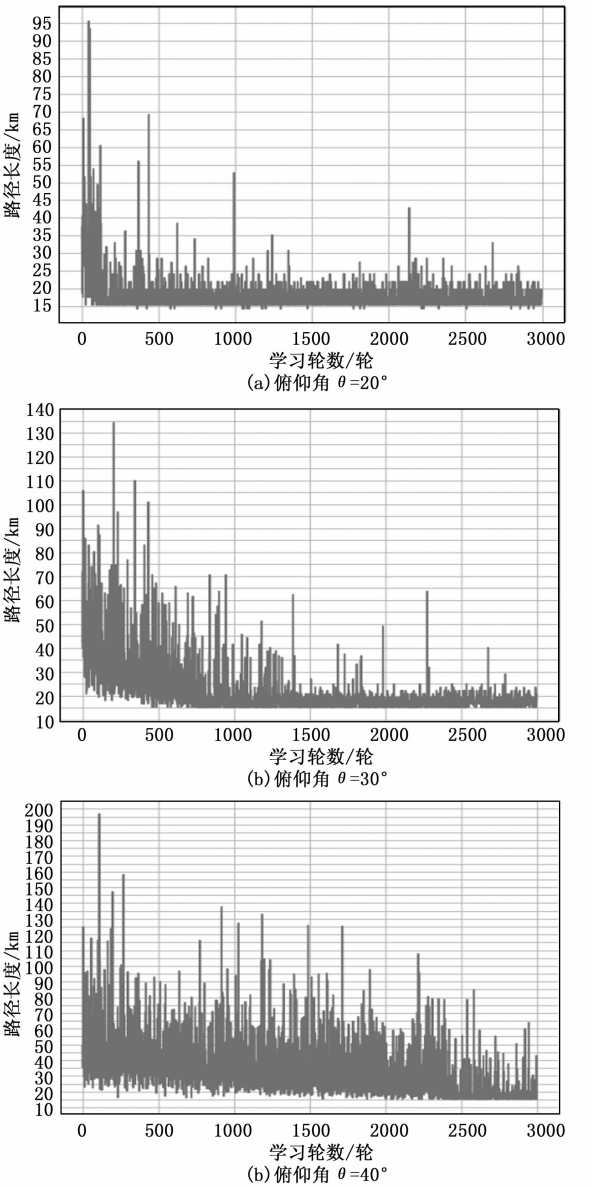
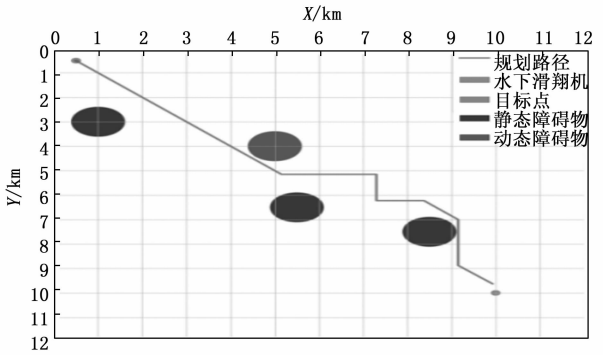


图 5 不同俯仰角设定下的路径长度变化

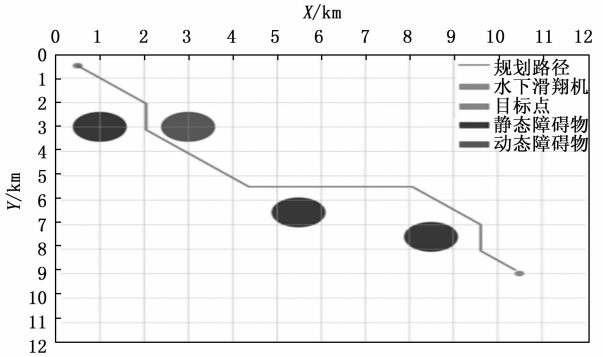
表 2 与表 3 记录了圆形与矩形两种障碍物环境中，不同位置动态障碍物下的规划路径的路径点个数、最短路径长度、与理论最短距离之差。理论最短距离指起点与目标点之间的直线距离。

表 2 圆形障碍物环境路径规划结果

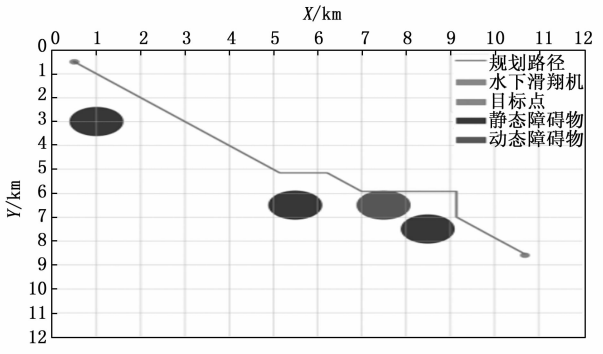
动态障碍物位置	路径点个数/个	最短路径长度/km	与理论最短距离之差/km
(5,4)	15	14.927	1.492
(3,3)	15	14.618	1.494
(7.5,6.5)	14	14.164	1.139



(a)圆形动态障碍物在（5，4）

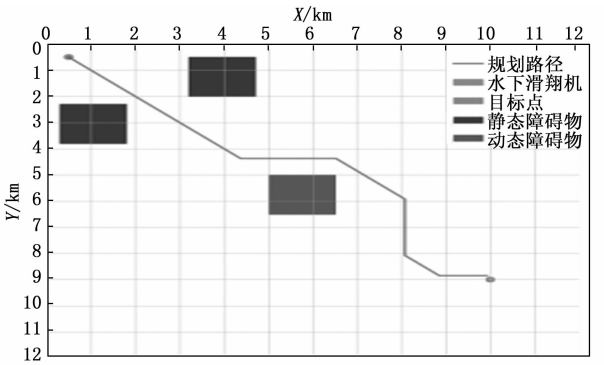


(b)圆形动态障碍物在（3，3）

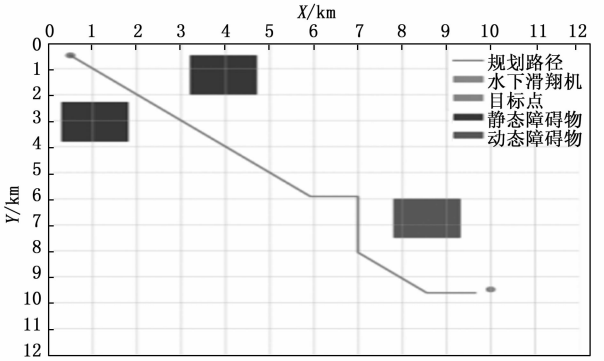


(c)圆形动态障碍物在（7.5，6.5）

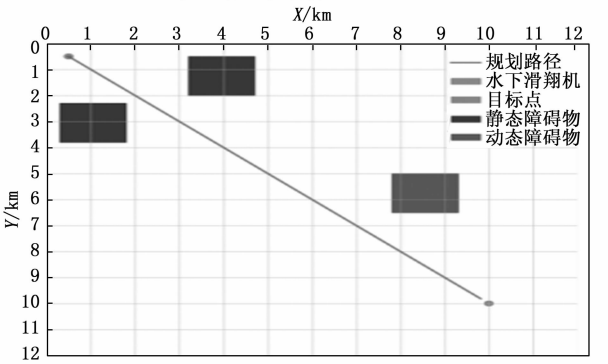
图 6　圆形障碍物环境下路径规划结果



(a)矩形动态障碍物在（5.75，5.75）



(b)矩形动态障碍物在（8.55，6.75）



(c)矩形动态障碍物在（8.55，5.75）

图 7　矩形障碍物环境下路径规划结果

表 3　矩形障碍物环境路径规划结果

动态障碍物位置	路径点个数/个	最短路径长度/km	与理论最短距离之差/km
(5.75,5.75)	14	14.143	1.395
(8.55,6.75)	14	13.118	0.381
(8.55,5.75)	13	13.172	0.263

由表 2 数据知，在圆形障碍物环境中，当动态障碍物位置在（5，4）km 时，规划最短路径长度为 14.927 km，与理论最短距离差为 1.492 km；当动态障碍物位置在（3，3）km 时，规划最短路径长度为 14.618 km，与理论最短距离差为 1.494 km；当动态障碍物位置在（7.5，6.5）km 时，规划最短路径长度为 14.164 km，与理论最短距离差为 1.139 km。在圆形障碍物环境中，以上 3 种情景中，当动态障碍物位置在（7.5，6.5）km 时所规划出的路径长度与理论最短距离差最小。

由表 3 数据知，在矩形障碍物环境中，当动态障碍物位置在（5.75，5.75）km 时，规划最短路径长度为 14.143 km，与理论最短距离差为 1.395 km；当动态障碍物位置在（8.55，6.75）km 时，规划最短路径长度为 13.118 km，与理论最短距离差为 0.381 km；当动态障碍物位置在（8.55，5.75）km 时，规划最短路径长度为 13.172 km，与理论最短距离差为 0.263 km。在矩形障碍物环境中，以上 3 种情景中，当动态障碍物位置在（8.55，5.75）km 时所规划出的路径长度与理论最短距离差最小。

仿真实验数据表明：基于 Q 学习的水下滑翔机路径规划方法能有效解决滑翔机在不同环境下路径规划问题且效率较高。能够在多个俯仰选择下规划出最优路径，路径长度与理论最短路径长度差距较小。当环境发生变化时，依然能得到最优路径。

5 结束语

本文提出基于Q学习的水下滑翔机路径规划方法，不需要提前得知环境信息，而是通过环境反馈的奖励值，使水下滑翔机学习对障碍物的有效规避，得到较短路径。仿真实验数据表明，本文提出的方法能满足水下滑翔机在复杂环境中的路径规划任务需求，并且在不同的环境条件下能够进行迁移，具有良好的通用性。

参考文献：

[1] LI D, WANG P, DU L. Path planning technologies for autonomous underwater vehicles - A Review [J]. IEEE Access, 2018, 7 (1): 9745 - 9768.

[2] 丛玉华，赵宗豪，邢长达，等. 基于改进人工势场的无人机动态避障路径规划 [J]. 兵器装备工程学报，2021，42 (9)：170 - 176.

[3] SONG R, LIU Y, BUCKNALL R. A multi-layered fast marching method for unmanned surface vehicle path planning in a time-variant maritime environment [J]. Ocean Engineering, 2017, 129 (5): 301 - 317.

[4] 刘雅丽，高立娥，李 乐. 通信距离约束下异构水下爬游机器人任务分配 [J]. 计算机测量与控制，2021，29 (9)：204 - 209.

[5] ZHANG B, DUAN Y, ZHANG Y, et al. Particle swarm optimization algorithm based on beetle antennae search algorithm to solve path planning problem [C] //2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, IEEE, 2020: 1586 - 1589.

[6] 袁梦婷，时宏伟. 基于 ADS-B 的无人机感知与规避蚁群算法模型 [J]. 西北工业大学学报，2021，39 (4)：761 - 769.

[7] 陈 暄，孟凡光，吴吉义. 求解大规模优化问题的改进狼群算法 [J]. 系统工程理论与实践，2021，41 (3)：790 - 808.

[8] 陈南凯，王耀南，贾 林. 基于改进生物激励神经网络算法的多移动机器人协同变电站巡检作业 [J]. 控制与决策，2022，37 (6)：1453 - 1459.

[9] GUO S, ZHANG X, ZHENG Y, et al. An autonomous path planning model for unmanned ships based on deep reinforcement learning [J]. Sensors, 2020, 20 (2): 426 - 432.

[10] CASTAG A, GUÉRIAU M, VIZZARI G. Demand-responsive rebalancing zone generation for reinforcement learning-based on-demand mobility [J]. AI Communications, 2021, 34 (1): 73 - 88.

[11] YANG Y, JUNTAO L, LINGLING P. Multi-robot path planning based on a deep reinforcement learning DQN algorithm [J]. CAAI Transactions on Intelligence Technology, 2020, 5 (3): 177 - 183.

[12] LIU X, ZHANG D, ZHANG T, et al. Novel best path selection approach based on hybrid improved A * al gorithm and reinforcement learning [J]. Applied Intel ligence, 2021, 51 (12): 9015 - 9029.

[13] WANG B, LIU Z, LI Q, et al. Mobile robot path planning in dynamic environments through globally guided reinforcement learning [J]. IEEE Robotics andAutomation Letters, 2020, 5 (4): 6932 - 6939.

[14] 江其洲，曾 碧. 基于深度强化学习的移动机器人导航策略研究 [J]. 计算机测量与控制，2019，27 (8)：217 - 226.

[15] 畅 鑫，李艳斌，田 淼，等. 基于一维卷积循环神经网络的深度强化学习算法 [J]. 计算机测量与控制，2022，30 (1)：258 - 265.

[16] HU H, ZHOU Y, WANG T, et al. A multi-task algorithm for autonomous underwater vehicles 3D path planning [C] // 2020 3rd International Conference on Unmanned Systems (ICUS), Beijing, China, IEEE, 2020: 972 - 977.

[17] 张子迎，陈云飞，王宇华，等. 基于启发式深度 Q 学习的多机器人任务分配算法 [J]. 哈尔滨工程大学学报，2022，22 (6)：1 - 7.

[18] 马 彬，陈海波，张 超. 基于改进深度 Q 学习的网络选择算法 [J]. 电子与信息学报，2022，44 (1)：346 - 353.

[19] 周 彬，郭 艳，李 宁，等. 基于导向强化 Q 学习的无人机路径规划 [J]. 航空学报，2021，42 (9)：506 - 513.

[20] 肖 浩，廖祝华，刘毅志，等. 实际环境中基于深度 Q 学习的无人车路径规划 [J]. 山东大学学报 (工学版)，2021，51 (1)：100 - 107.

(上接第 191 页)

[17] CUI Y, GAO S, ZHENG Y. Application of ZigBee Location Fingerprint Method in Positioning of Railway Tunnel Staff [C] // 2018 Chinese Automation Congress (CAC), 2018, pp. 3283 - 3287.

[18] LING J, WANG L, JI H, et al. UWB-Based Real-Time Continuous Positioning System in NLOS Tunnel Enviro-nment [C] // 2018 International Conference on Cyber Enabled Distributed Computing and Knowledge Discovery (CyberC), 2018, 142 - 1424.

[19] NAKAMURA M, et al. Development of a simple multiple-position identifying system with a long range multiband leaky coaxial cable for rescue operations in tunnels or passages in underground facilities [C] // 2010 Asia-Pacific Microwave Conference, 2010, 163 - 166.

[20] 黄文摄. 北斗卫星导航技术在隧道内定位的应用 [J]. 卫星应用，2022 (2)：66 - 69.

[21] 杨 柯，蔡成林，等. 一种高精度的 GNSS 伪距单点定位加权算法 [J]. 计算机仿真，2019，36 (4)：229 - 233，239.

[22] 王奉帅，刘聪锋. 迭代最小二乘卫星定位算法 [J]. 无线电通信技术，2018，44 (4)：339 - 342.