控制与决策
Control and Decision

基于置信度上界的移动机器人信息路径规划方法

王轶强, 吴芝亮, 李群智

引用本文:
王轶强,吴芝亮,李群智. 基于置信度上界的移动机器人信息路径规划方法[J]. 控制与决策, 2023, 38(2): 395−402.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.1158

基于置信度上界的移动机器人信息路径规划方法

王轶强[1], 吴芝亮[1†], 李群智[2]

(1. 天津大学 机械工程学院，天津 300354；2. 中国空间技术研究院 北京空间飞行器总体设计部，北京 100094)

摘 要: 路径规划是移动机器人未知环境探索的关键问题,路径点的合理规划对提高环境探索的效率和环境场预测的准确性至关重要. 基于强化学习范式,提出一种适用于静态环境场探索的移动机器人在线信息路径规划方法. 针对基于模型训练算法计算成本高的问题,通过机器人与环境的交互作用,采用动作价值评估的方法来学习所获取的环境场历史信息,提高机器人实时规划能力. 为了提高环境预测准确性,引入基于置信度上界的动作选择方法来平衡探索未知区域与利用已有信息,鼓励机器人向更多未知区域进行全场特征探索,同时避免因探索区域有限而陷入局部极值. 仿真实验中,环境场分别采用高斯分布和Ackley函数模型. 结果表明,所提算法能够实现机器人环境探索路径点的在线决策,准确有效地捕捉全场和局部环境特征.

关键词: 移动机器人；未知环境探索；置信度上界；强化学习；信息路径规划

中图分类号: TP242 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.1158 **开放科学(资源服务)标识码(OSID):**

引用格式: 王轶强,吴芝亮,李群智. 基于置信度上界的移动机器人信息路径规划方法[J]. 控制与决策, 2023, 38(2): 395-402.

An informative path planning approach for mobile robots based on upper confidence bound algorithm

WANG Yi-qiang[1], *WU Zhi-liang*[1†], *LI Qun-zhi*[2]

(1. School of Mechanical Engineering，Tianjin University，Tianjin 300354，China；2. Beijing Institute of Spacecraft Systems Engineering，China Academy of Space Technology，Beijing 100094，China)

Abstract: Path planning for mobile robots is paramount in unknown environment exploration. Exploration efficiency and prediction accuracy largely depend on appropriate waypoint decision. In this paper, an informative path planning approach is proposed based on the reinforcement learning paradigm for static environment exploration. In contrast to model based algorithms, no assumption is presumed for the environmental features. The computational cost is reduced and the online planning capability is enhanced by evaluating the values of actions through robot's interaction with the environment. To improve the prediction accuracy, an action selection algorithm based on the Upper Confidence Bound (UCB) is utilized to balance exploration and utilization. Exploration in unknown areas is encouraged, which also potentially avoid stucking into local extrema. Numerical simulations have been performed on environments that are modeled with Gaussian distribution and Ackley function respectively. Results show that the characteristics of the entire environment field can be effectively captured using the proposed path planning approach.

Keywords: mobile robot; unknown environment exploration; upper confidence bound; reinforcement learning; informative path planning

0 引 言

移动机器人正越来越广泛地被人们用于异常环境数据的收集,为人们提供更好的视角来了解环境状况[1]. 具有自主能力的移动机器人可替代人类在复杂或极端环境下进行环境探索,获取环境信息[2],实施环境监测[3],协助灾难救援[4]等. 在这些任务场景下,通常要求机器人通过自主探索实现未知环境场的特征信息采集,搜寻局部特征极值[5]、表征全场物理特性[6]. 此类环境探索的核心是为机器人规划出合适的采样路径,从而提高机器人探索效率,同时有效捕捉环境场特征.

传统的路径规划方法旨在起点与终点之间寻得

收稿日期: 2021-07-03；**录用日期:** 2021-11-10.

基金项目: 国家自然科学基金项目(51975044)；国家重点研发计划项目子课题(2016YFC0301102).

责任编委: 訾斌.

†通讯作者. E-mail: zhlwu@tju.edu.cn.

满足一定优化目标的最优路径[7-8],但在探索异常环境时,通常面临环境先验知识欠缺、机器人探索终点位置不受限的情况. 因此,机器人需要采用信息路径规划方法[5, 9],根据实时探测信息,依据有效的探索策略进行在线决策和规划.

国内外学者以环境采样信息最大化[9]为目标开展了移动机器人信息路径规划方法的研究. 在处理环境信息方面,研究者通常借助环境模型来预测待采样位置的环境信息. 文献[3]设计了一种分层贝叶斯优化方法,帮助机器人找到最大化环境信息的路径;文献[5]提出一种基于距离的贝叶斯优化方法,解决了机器人在环境探索时勘探与开发的权衡问题,有效探测环境信息并减小极端区域的监测误差;文献[10]通过非参数贝叶斯回归引入不确定性建模,构建准确的三维环境模型并检测异常情况;文献[11]提出一种利用进化策略对连续空间中的路径进行优化的方法并引入重规划方案,获得环境信息的同时将探测重点放在有价值的区域. 上述信息路径规划方法的核心在于通过高斯过程回归模型估计机器人待采样位置的环境特征信息. 机器人对基于模型训练的预测信息进行评估,进而规划路径. 文献[12]结合人工势场法在环境数据值或者不确定性更大的区域选择路径点,提高机器人对环境特征及环境特征极值位置估计的准确性;文献[13]提出一种贝叶斯优化预测互信息的方法,通过选择视野内信息最丰富的位置进行采样来最大化环境信息;文献[14]通过预测环境数据变化梯度的极值及其不确定性引导机器人进行在线路径规划,获得环境场特征分布;文献[15]提出一种应用于标量场的长期自适应信息路径规划方法,并采用交叉熵作为目标函数对路径进行优化. 训练环境模型可以提供均值函数和协方差函数,量化表征环境数据及其不确定性[16],但样本数据维数和数量较大时,模型训练耗时长,计算成本高.

本文基于强化学习范式,提出一种适用于静态环境场探索的移动机器人信息路径规划方法. 探索过程中,机器人通过评估动作价值来学习所获取的环境场历史信息,而非通过训练环境模型估计待采样位置的环境信息,提高计算效率. 同时,机器人利用基于置信度上界(upper confidence bound, UCB)算法平衡探索与利用,实现路径点的在线决策. 仿真实验中,将本文方法分别应用于具有单极值和多极值的静态环境场探索,验证方法的可行性与有效性.

1 信息路径规划

假设机器人在二维平面内探索静态的未知环境场,机器人路径规划的目的在于选取合理的路径点,保证环境数据采样的有效性,提高环境探索效率. 因此,机器人需要根据获取的环境信息,实现自主在线决策,规划出后续采样路径点,完成环境特征分布和极值的探测.

本文采用无向图$G=\langle V,E\rangle$描述环境地图,如图1所示. 顶点集V表示机器人可行路径点的集合$\{v_{i,j}\}(i,j=1,2,\ldots,n,\ldots)$,边集$E$表示机器人可选择的路径集合$\{\langle v_{ij},v_{kl}\rangle\}(i,j,k,l=1,2,\ldots,n,\ldots)$. 采用直角坐标法对可行路径点进行编码,按由下至上、由左至右的顺序依次递增编号. 水平和竖直方向上,相邻顶点以等间距的方式设置. $v_{1,1}$为机器人的起始位置,对应直角坐标系XOY的原点.

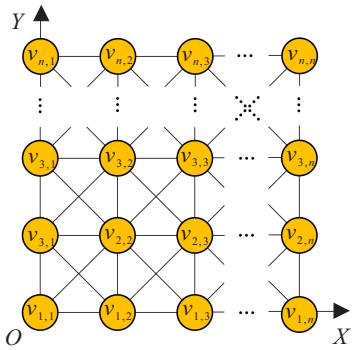


图1 环境地图的无向图表示

环境地图中路径点的坐标与节点编码之间的对应关系可以表示为

$$\begin{cases} x_m = x/\text{unit}+1, \\ y_m = y/\text{unit}+1. \end{cases} \tag{1}$$

其中:x_m和y_m分别表示两个维度上的节点编码,(x,y)表示路径点在直角坐标系中的坐标,unit表示竖直或水平方向相邻节点间的距离.

假设在每个路径点处机器人有若干种执行动作可选择,且机器人仅在其所到达的路径点处进行环境数据采集和后续路径点决策. 在环境探索过程中,机器人路径点决策过程如下:首先,在当前路径点获取位置信息与环境特征信息;然后,结合历史数据,评估在强化学习范式下定义的动作价值;最后,基于评估的不确定性和平衡探索与利用的理念,进行动作选择,完成路径决策.

2 动作价值计算

在未知环境探索过程中,环境信息是机器人路径点决策和规划的重要依据. 在强化学习范式下,机器人与环境发生交互作用,根据已获取的环境采样数据评估机器人动作价值. 本文基于环境特征值梯度定义机器人在路径点s_{t-1}采取动作a_{t-1}后所获得的即时奖励,如下所示:

$$R_{t-1}(s_{t-1},a_{t-1})=z_t-z_{t-1}. \tag{2}$$

其中: t 表示时间序列, z_t 表示当前路径点的环境采样数据, z_{t-1} 表示上一时刻路径点的环境采样数据, $a_{t-1}\in\{1,2,\dots,n,\dots\}$ 表示机器人在每个对应位置上向各个可选择方向的动作编码.

对于特征值变化单调的环境场, 可仅考虑当前路径点的即时奖励. 对于较为复杂的环境场, 机器人需要更好地学习环境历史数据, 更新对环境场的认知, 动作价值函数可根据即时奖励定义为

$$Q(a_t)=\sum_h \omega_h\cdot R_h(s_h,a_h)\Big/N_a. \tag{3}$$

其中: ω_h 为对应历史奖励分配的权值, N_a 为机器人采取过动作 a_t 的次数, $h\in\{0,1,2,\dots,t-1\}$ 为动作经验所在位置节点序列号.

权值 ω_h 的分配和历史路径点与当前路径点间的距离成反比, 归一化后可得

$$Q(a_t)=\eta\cdot\frac{1}{\sqrt{(x_h-x_t)^2+(y_h-y_t)^2}}. \tag{4}$$

其中: (x_h,y_h) 为动作经验对应的位置坐标, η 为归一化参数. 因为环境变化具有连续性, 历史动作奖励距当前位置越近, 环境特征相对越相似, 相应动作奖励获得的权重越大, 表示对距离当前位置较近的历史动作奖励更加重视. 通过上述权值分配方法, 机器人通过与环境进行交互并利用历史动作奖励计算动作价值的方式来评估环境场特征变化趋势.

3　动作选择策略

在动作价值计算的基础上, 采用非贪婪算法设计动作选择策略. 在动作选择过程中, 考虑了动作价值评估的不确定性, 从平衡探索与利用的理念出发, 采用基于置信上界算法的策略, 进行后续路径点决策. 算法需要机器人在探索初期具有一定的探索经验, 所以首先以初始随机策略指导机器人的移动, 之后采用UCB实现利用已有经验和探索未知区域的平衡, 指导机器人向更多区域进行探索, 继而预测出全场环境特征以及极值的位置和数值.

3.1　初始随机策略

在环境探索的初始阶段, 由于采样数据稀少, 动作价值评估存在较大不确定性, 机器人采取等概率随机策略选择背离起点方向的可行动作来进行环境探索, 向未知环境进行探索的同时尽可能为下一节中置信上界的动作选择提供多个方向上的动作经验. 如图2所示, 红色箭头表示可选择的动作方向, 绿色路径点表示历史采样位置, 蓝色线段表示路径. 根据式(3)和(4)进行动作价值的计算, 积累动作经验. 达到设定步数时, 初始探索阶段结束.

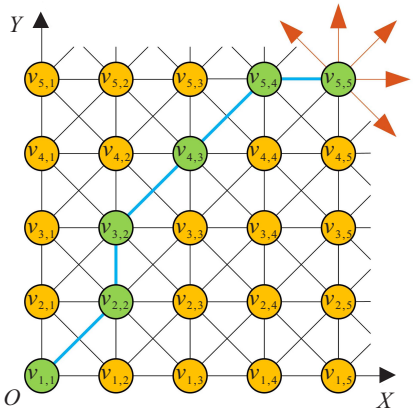


图 2　初始随机策略示意图

3.2　置信上界动作选择策略

利用即时奖励对动作价值的估计存在不确定性. 从长远的角度来看, 利用贪婪算法选择动作价值最大的动作可能并非最好的选择[17-18], 这可能会使机器人因受到不确定因素的干扰而选择非最优的动作. 为了鼓励机器人向更多未知区域进行全场特征探索, 同时避免因探索区域有限而陷入局部极值, 引入基于UCB的动作选择方法. 算法中, 采用下式评估动作价值计算的不确定性[19], 并根据动作价值及其不确定性计算每个动作的置信度上界:

$$U(a_t)=\sqrt{\frac{2\ln N_{\text{sum}}}{N_a}}, \tag{5}$$

其中 N_{sum} 表示机器人采取过的动作的总数. 之后, 在所有可选动作中选择置信度上界最大的动作, 如下所示:

$$A_t\doteq\arg\max_{a_t}\left(Q(a_t)+\sqrt{\frac{2\ln N_{\text{sum}}}{N_a}}\right). \tag{6}$$

对于小样本情况, 采用式(5)描述动作价值评估的不确定可能引起较大的偏差. 为了提高机器人探索未知区域的有效性, 在进行动作选择之前, 先对各个动作进行筛选. 根据机器人当前位置, 将环境地图划分为右上方、左上方、左下方、右下方4个区域, 如图3所示, 机器人位于 $v_{m,m}$ 处, 红色虚线将无向图分成4个区域.

分别计算4个区域中探索过的历史路径点数量与对应区域包括的所有路径点数量的比值 P_d, 对每个探索过的位置赋予与距当前所在状态的距离成反比的权重 τ, P_d 的计算方式如下:

$$P_d=\sum_{i=1}^{E_d}\tau_i\Big/N_d. \tag{7}$$

其中: $d\in\{1,2,3,4\}$ 为4个区域的编号, 分别与机器人右上方、左上方、左下方、右下方区域对应, E_d 为属

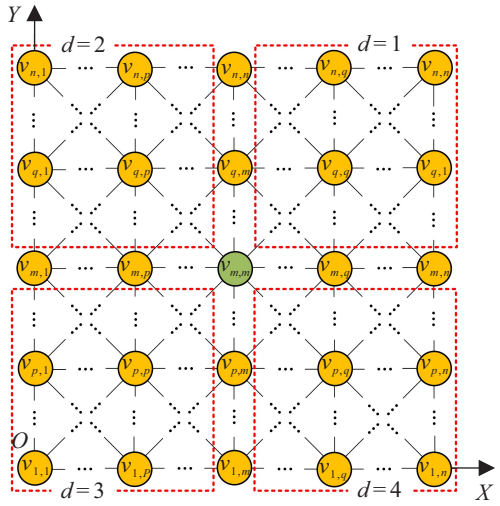


图 3 无向图分区

于在 d 区域的机器人探索过位置的数量, N_d 为机器人在对应区域包含的所有位置的数量. 当机器人处在环境地图边界上时, 根据机器人所处的位置, 将环境分为两个区域, $d \in \{1, 2\}$. 当机器人处在地图环境的4个角落位置时, 则无需划分区域.

计算得到的 P_d 值越大, 表示机器人对 d 区域的信息探索相对更充分. 为了引导机器人向未知的环境进行探索, 算法中排除了机器人向 P_d 值最大的区域探索. 之后, 根据式(6)在剩余的可选动作中进行选择.

4 仿真与分析

为了验证算法的有效性, 采用数值模拟方法分别进行具有单极值点和多极值点的二维环境场探索. 仿真实验中, 假设机器人在路径点可选择移动的方向共有8个, 即向上、向下、向左、向右、左上、右上、左下、右下.

通过在仿真环境中与传统六边形路径(hex-path)算法[20]、全覆盖(full coverage)算法[21]以及基于模型训练的不确定性采样(uncertain sampling, US)[22]方法进行对比, 验证本文算法的探索效率. 算法运行环境为: Window10 64 bits, Matlab R2016a, 处理器 Intel Core i5 9400, 主频2.9 GHz, 内存8 GB.

4.1 单极值点环境探索

环境模型采用高斯分布模型, 生成具有单极值点的环境场. 环境函数模型如下所示:

$$f(x,y) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}) \cdot \exp\Big(-\frac{1}{1-\rho^2}\cdot\Big(\frac{(x-\xi_1)^2}{\sigma_1^2} - \frac{2\rho(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2} + \frac{(y-\xi_2)^2}{\sigma_2}\Big)\Big). \tag{8}$$

其中: (ξ_1, ξ_2) 为极值点的位置坐标, 设置在(15,15)处; $\rho = 0$ 为相关系数; $\sigma_1 = 5$ 和 $\sigma_2 = 5$ 分别为横纵坐标的分布方差. 仿真实验中, 对应的环境场云图如图4所示. 实验参数设置: 环境场的尺寸为 20×20, 步长 unit $= 1$, 环境地图中共有441个路径点. 初始随机策略探索步数 $n_s = 10$.

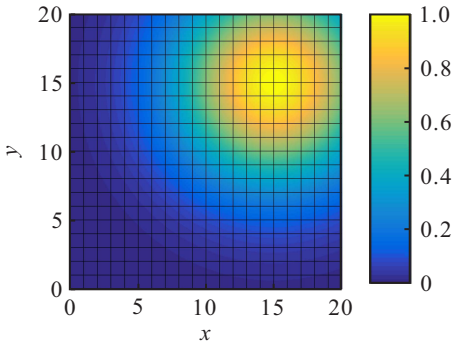


图 4 单极值环境场

移动机器人信息路径规划的实质是决策出合理的路径点, 提高全场环境特征预测的准确性. 因此, 采用全场环境特征预测的均方根误差、加权均方根误差评估算法对全场环境特征预测的准确性. 同时, 探索环境特征极值的位置和数值也是本文的目标, 采用对极值的定位误差和特征值预测误差来评估算法的性能.

均方根误差用于衡量整体环境的预测质量, 表示为

$$\text{rmse} = \sqrt{\frac{\sum_{i=1}^{M}[f(x_i,y_i) - \mu(x_i,y_i)]^2}{M}}. \tag{9}$$

加权均方根误差根据环境数据的大小分配权重, 对环境数值较高的区域给予相对更高的权重[5], 如下所示:

$$\text{wrmse} = \sqrt{\frac{\sum_{i=1}^{M}\Big[\frac{\lambda\cdot(f(x_i,y_i) - \mu(x_i,y_i))}{\max(\mu(x,y)) - \min(\mu(x,y))}\Big]^2}{M}}. \tag{10}$$

其中: $\mu(x,y)$ 表示高斯过程回归模型利用机器人在环境模型 $f(x,y)$ 中探索得到的采样数据训练出环境预测场, $\lambda = \mu(x_i,y_i) - \min(\mu(x,y))$ 表示对环境数据分配权重, M 表示环境中均布点的数量.

预测极值点的位置和特征值误差用于表征极值的定位误差和预测精度, 如下所示:

$$r = |\text{Location}(\max(f(x,y))) - \text{Location}(\max(\mu(x,y)))|, \tag{11}$$

$$l = |\max(f(x,y)) - \max(\mu(x,y))|. \tag{12}$$

图5为本文算法与hex-path算法、US算法以及full coverage算法的运行路径,所有算法起点为(0, 0)处,圆点表示对应路径的终点. 将4种算法分别进行上述4个方面的性能比较,结果如图6所示.

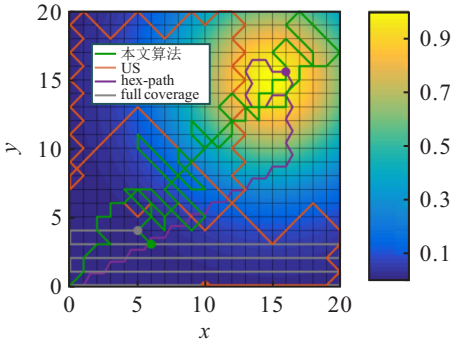
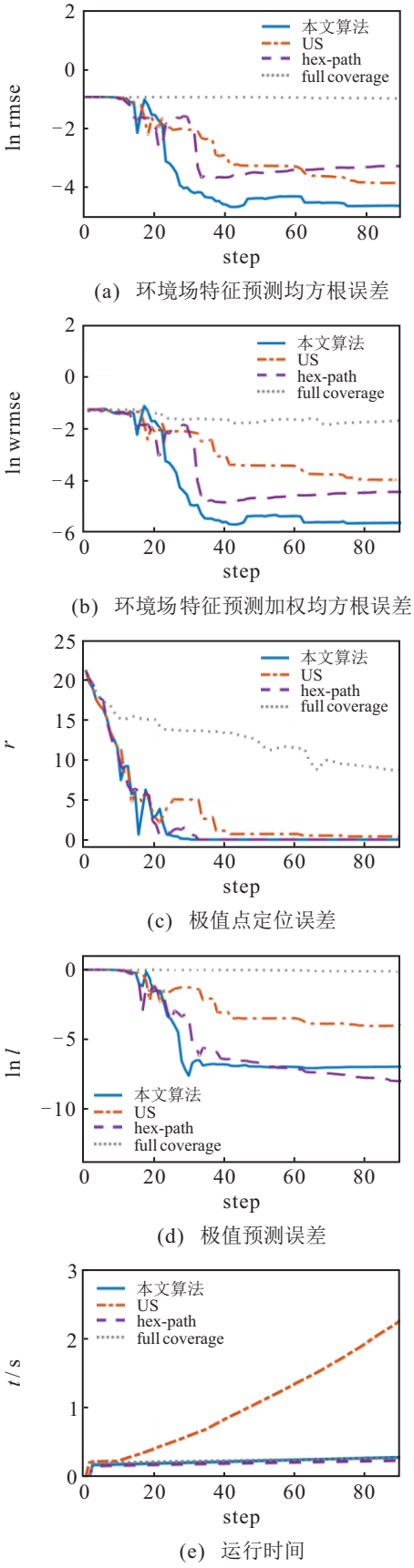


图 5　4种算法在单极值环境场中的路径对比

从图6(a)和图6(b)中的曲线变化可以看出,本文算法在单极值环境场中得到的均方根误差和加权均方根误差结果更小. 这说明本文算法对全场环境以及环境特征变化复杂的重点区域探索得更加充分,所以得到的预测结果误差较小. 同时从图6(c)和图6(d)中的曲线可以看出,本文算法和hex-path算法在环境特征极值的求解方面优于US算法及full coverage算法,收敛速度较快;对环境特征极值的数值预测,本文算法与hex-path算法相当. 这说明本文算法在单极值环境场中,在保证对全场特征取得较准确预测结果的同时,也能较为准确地获取环境场局部特征,实现利用已有经验和探索未知环境间的平衡.

采用hex-path算法时,机器人在单极值环境场中会朝着环境特征极值的方向前进并徘徊在其附近,所以对环境特征极值的位置及其数值的预测相对更加准确. 但该算法仅考虑了环境特征极值区域的数据采集,忽略了对环境中的其他区域进行探索,所以具有较高的误差. US算法注重指导机器人向着环境中信息不确定性更大的区域进行探索,但是没有考虑利用环境信息对环境场中的重点区域进行探索,在环境极值附近区域路径点较少,所以对环境特征极值的数值预测精度较低. 在full coverage算法下,机器人机械地进行环境采样,不能根据环境采样信息实时决策路径,所以各指标结果较差. 如图6(e),在单极值环境场中运行相同步数,本文算法耗时与hex-path算法及full coverage算法时间相当,明显低于基于模型训练的US算法. US计算量随样本数据规模呈正比增大,而本文算法采用动作价值评估的方法来学习所获取的环境场历史信息,实时规划路径,避免了基于模型训练算法所带来的计算成本高的问题.



(a)　环境场特征预测均方根误差

(b)　环境场特征预测加权均方根误差

(c)　极值点定位误差

(d)　极值预测误差

(e)　运行时间

图 6　4种算法在单极值环境场中的结果对比

4.2　多极值点环境探索

多极值点环境采用Ackley环境模型, Ackley函数可以模拟一种物质的扩散[16],模型函数为

$$f(x,y) = -k\cdot\exp\left[-b\sqrt{(x^2+y^2)/2}\right]- \exp\left[(\cos(cx)+\cos(cy))/2\right]+k+\exp(g). \tag{13}$$

参数设置为$k=-36, b=2.2, c=\pi, g=1$. 仿真实验中, 对应的环境场云图如图7所示. 图中环境特征极值位置为$(15,15)$, 在环境场中还存在多个局部极值, 环境地图尺寸为20×20, 步长unit $=1$, 共有441个路径点. 初始随机策略探索步数$n_s=10$.

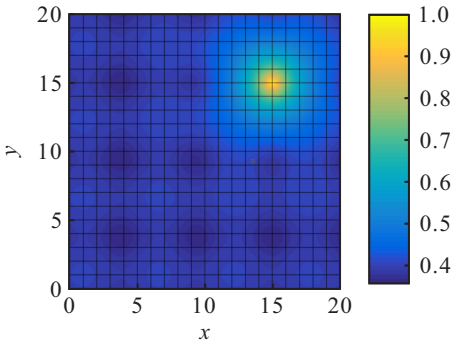


图 7 Ackley环境场

4种算法的运行路径如图8所示, 算法性能同样采用全场预测的均方根误差、加权均方根误差、对环境特征极值的定位误差以及数值预测误差来评估, 如图9所示. 图9也显示了本文算法与hex-path算法、US算法及full coverage算法的结果比较.

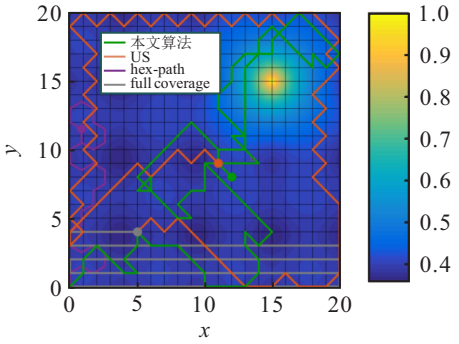
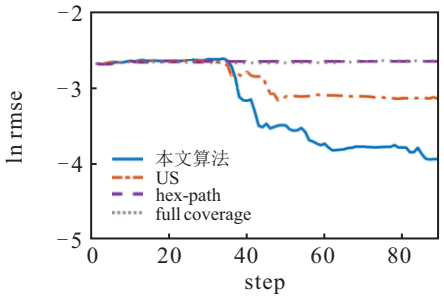


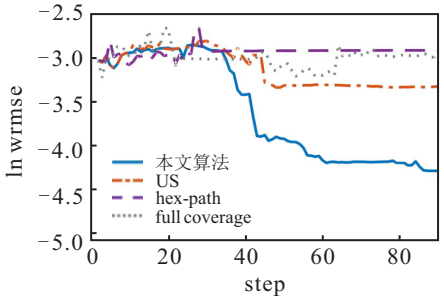
图 8 4种算法在Ackley环境场中的路径对比

从图9(a)和图9(b)可以看出, 本文算法在Ackley环境场中的均方根误差和加权均方根误差明显低于其他算法, 对全场环境以及重点区域的预测更加准确. 图9(c)和图9(d)说明本文算法在Ackley环境场中对环境特征极值的定位误差和数值预测误差更小. 本文算法基于环境特征变化梯度定义即时奖励, 着重考虑选择动作价值及不确定性大的动作, 所以在相对充分获得全场环境特征的同时, 对环境特征极值的位置及数值预测也更加准确.

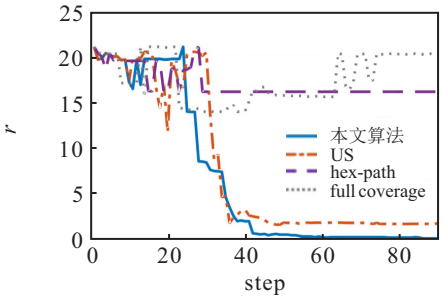
从图9(a)～图9(d)可以看出, hex-path算法在Ackley环境场中表现不佳, 对全场环境和局部特征预测误差大. US算法在复杂环境场中与在单极值环境



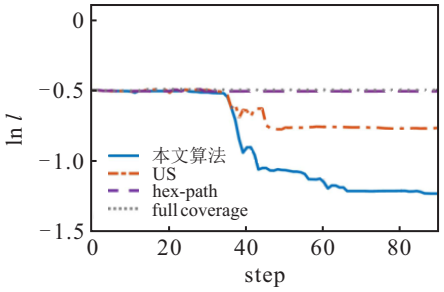
(a) 环境场特征预测均方根误差



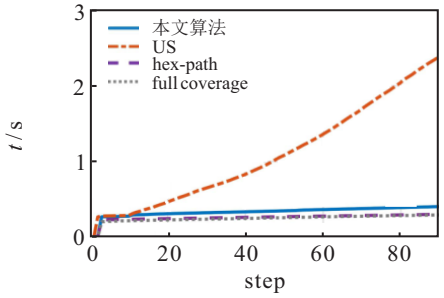
(b) 环境场特征预测加权均方根误差



(c) 极值点定位误差



(d) 极值预测误差



(e) 运行时间

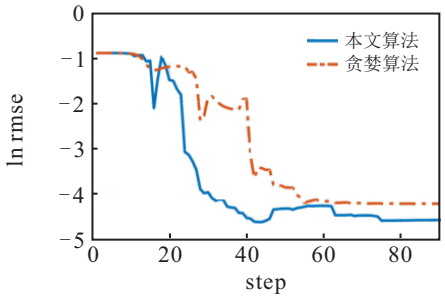
图 9 4种算法在Ackley环境场中的运行结果对比

场中表现相当, 能较好反映全场特征, 但预测准确度相对较差. full coverage各项指标最差, 在没有完成全场覆盖的时候不能反映全场特征. 如图9(e), 在多极值环境场中运行相同步数, 本文算法耗时与hex-path

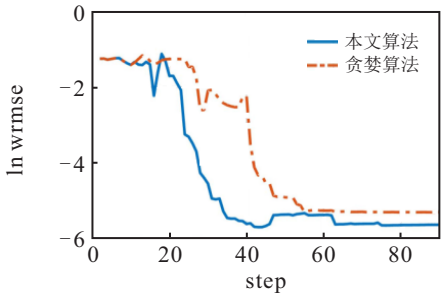
算法及full coverage算法相当,明显低于US算法.

4.3　UCB的作用

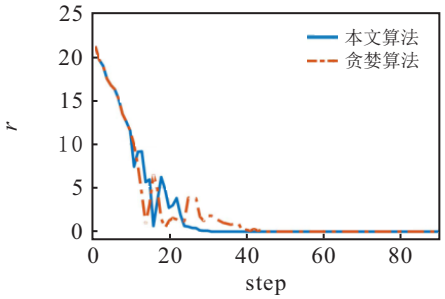
为了检验UCB算法在利用已有经验和探索未知环境之间的权衡作用,将本文算法与贪婪算法进行对比. 图10和图11分别为单极值高斯环境场和Ackley环境场中的结果.



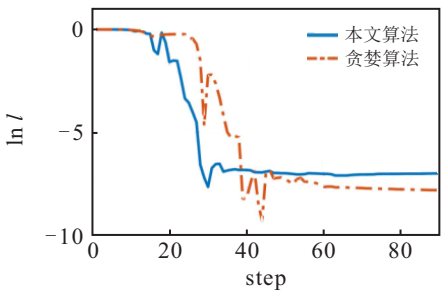
(a)　环境特征预测均方根误差



(b)　环境特征预测加权均方根误差



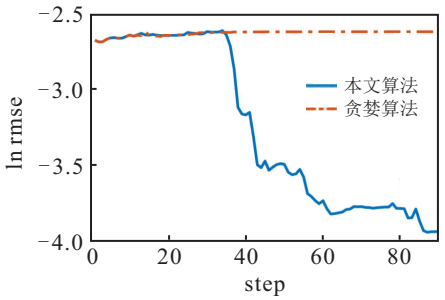
(c)　极值点定位误差



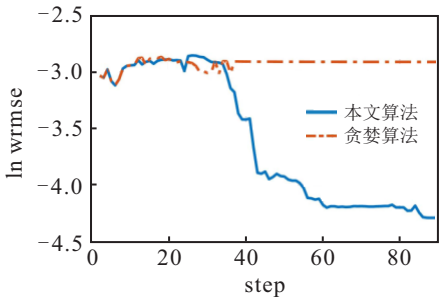
(d)　极值预测误差

图 10　两种算法在单极值环境场中的数据对比

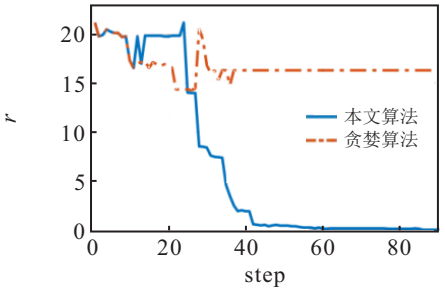
由图10(a)和图10(b)可以看出,本文算法和贪婪算法对单极值高斯环境场的探测效果相当,但本文算法收敛速度相对较快. 从图11对比的结果可以看出,本文算法在Ackley环境场中的效果明显优于贪婪算法,具有相对较低的均方根误差和加权均方根误差,且对环境特征极值搜索效率更高,对其数值预测更



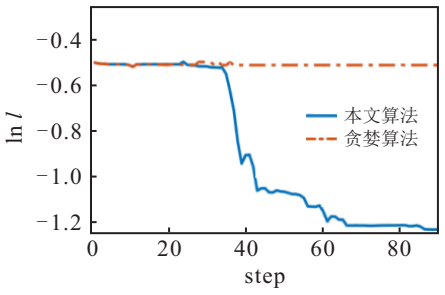
(a)　环境场特征预测均方根误差



(b)　环境场特征预测加权均方根误差



(c)　极值点定位误差



(d)　极值预测误差

图 11　两种算法在Ackley环境场中的数据对比

加准确. 利用贪婪算法,机器人容易陷入环境场中的局部极值. 两种算法的结果对比说明本文算法可以平衡利用已有经验和探索未知环境的关系,避免机器人陷入局部极值.

5　结　论

本文基于强化学习范式提出了一种面向静态环境场探索的机器人信息路径规划方法. 机器人根据实时采样数据,计算动作价值,并结合基于置信度上界的动作选择方法进行利用已有经验和探索未知区域间的平衡,实现路径点决策. 仿真实验结果表明,与传统方法相比,本文算法在环境场特征极值评估与全场环境预测方面均具有较好的准确度和较快的收敛

速度.

本文研究仅利用单机器人在静态无障碍环境中进行全场探索,未来的工作计划重点考虑多机器人在多极值的复杂环境场中协同信息路径规划问题,以完成对多环境特征极值环境的探测任务.

参考文献(References)

[1] Dunbabin M, Marques L. Robots for environmental monitoring: Significant advancements and applications[J]. IEEE Robotics & Automation Magazine, 2012, 19(1): 24-39.

[2] 张晓平, 阮晓钢, 肖尧, 等. 基于内发动机机制的移动机器人自主路径规划方法[J]. 控制与决策, 2018, 33(9): 1605-1611.
(Zhang X P, Ruan X G, Xiao Y, et al. Mobile robot autonomous path planning method based on intrinsic motivation mechanism[J]. Control and Decision, 2018, 33(9): 1605-1611.)

[3] Marchant R, Ramos F. Bayesian optimisation for informative continuous path planning[C]. 2014 IEEE International Conference on Robotics and Automation. Hong Kong, 2014: 6136-6143.

[4] 孙辉辉, 胡春鹤, 张军国. 移动机器人运动规划中的深度强化学习方法[J]. 控制与决策, 2021, 36(6): 1281-1292.
(Sun H H, Hu C H, Zhang J G. Deep reinforcement learning for motion planning of mobile robots[J]. Control and Decision, 2021, 36(6): 1281-1292.)

[5] Marchant R, Ramos F. Bayesian optimisation for intelligent environmental monitoring[C]. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura-Algarve, 2012: 2242-2249.

[6] Morere P, Marchant R, Ramos F. Sequential Bayesian optimization as a POMDP for environment monitoring with UAVs[C]. 2017 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2017: 6381-6388.

[7] 张玮, 马焱, 赵捍东, 等. 基于改进烟花-蚁群混合算法的智能移动体避障路径规划[J]. 控制与决策, 2019, 34(2): 335-343.
(Zhang W, Ma Y, Zhao H D, et al. Obstacle avoidance path planning of intelligent mobile based on improved fireworks-ant colony hybrid algorithm[J]. Control and Decision, 2019, 34(2): 335-343.)

[8] 黄鲁, 周非同. 基于路径优化D*Lite算法的移动机器人路径规划[J]. 控制与决策, 2020, 35(4): 877-884.
(Huang L, Zhou F T. Path planning of moving robot based on path optimization of D* Lite algorithm[J]. Control and Decision, 2020, 35(4): 877-884.)

[9] 阮晓钢, 郭威, 黄静, 等. 机器人信息增益RRT环境探索算法[J]. 控制与决策, 2021, 36(11): 2683-2689.
(Ruan X G, Guo W, Huang J, et al. Robot RRT based on information gain for environment exploration[J]. Control and Decision, 2021, 36(11): 2683-2689.)

[10] Hollinger G A, Englot B, Hover F, et al. Uncertainty-driven view planning for underwater inspection[C]. 2012 IEEE International Conference on Robotics and Automation. Saint Paul, 2012: 4884-4891.

[11] Hitz G, Galceran E, Garneau M È, et al. Adaptive continuous-space informative path planning for online environmental monitoring[J]. Journal of Field Robotics, 2017, 34(8): 1427-1449.

[12] Neumann P P, Asadi S, Lilienthal A J, et al. Autonomous gas-sensitive microdrone: Wind vector estimation and gas distribution mapping[J]. IEEE Robotics & Automation Magazine, 2012, 19(1): 50-61.

[13] Bai S, Wang J K, Chen F F, et al. Information-theoretic exploration with Bayesian optimization[C]. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, 2016: 1816-1822.

[14] Yan S X, Li Y P, Feng X S. An AUV adaptive sampling method based on Gaussian process regression[J]. Robot, 2019, 41(2): 232-241.

[15] Li Y, Cui R X, Yan W S, et al. Long-term adaptive informative path planning for scalar field monitoring using cross-entropy optimization[J]. Science China Information Sciences, 2019, 62(5): 1-3.

[16] Blanchard A, Sapsis T. Bayesian optimization with output-weighted optimal sampling[J]. Journal of Computational Physics, 2021, 425: 109901.

[17] Sutton R S, Barto A G. Reinforcement learning: An introduction[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 1054.

[18] Wei Y Y, Zheng R. Informative path planning for mobile sensing with reinforcement learning[C]. IEEE Conference on Computer Communications (IEEE INFOCOM 2020). Toronto, 2020: 864-873.

[19] Peter A, Nicolò C B, Paul F. Finite-time analysis of the multiarmed bandit problem[J]. Machine Learning, 2002, 47(2/3): 235-256.

[20] Russell R A. Chemical source location and the robomole project[C]. Proceedings Australian Conference on Robotics and Automation. Citeseer, 2003: 1-6.

[21] Li K, Chen Y F, Jin Z Y, et al. A full coverage path planning algorithm based on backtracking method[J]. Computer Engineering & Science, 2019, 41(7): 1227-1235.

[22] MacKay D J C. Information-based objective functions for active data selection[J]. Neural Computation, 1992, 4(4): 590-604.

作者简介

王轶强(1996−), 男, 硕士生, 从事移动机器人路径规划的研究, E-mail: wyq2019@tju.edu.cn;

吴芝亮(1979−), 女, 副教授, 博士, 从事移动机器人运动规划、多智能体协同运动规划等研究, E-mail: zhlwu@tju.edu.cn;

李群智(1978−), 女, 高级工程师, 博士, 从事空间探测器总体设计、空间智能系统总体设计等研究, E-mail: 13681332025@139.com.

(责任编辑: 齐 霁)