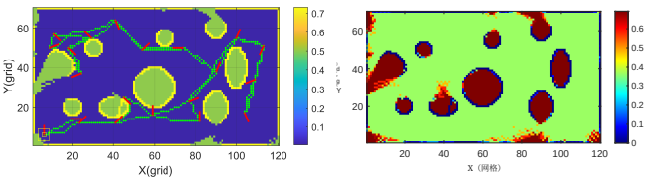
arXiv:2301.00523v1 [cs.RO] 2 Jan 2023

用于自主机器人探索的贝叶斯广义核推断

Yang Xu，*IEEE* 学生会员，Ronghao Zheng†，*IEEE* 会员，Senlin Zhang，*IEEE* 会员，以及 Meiqin Liu，*IEEE* 高级会员

摘要—本文旨在实现在复杂场景中兼具高效率与理想性能的信息论机器人探索。我们构建了一种连续轻量级推断模型，用于预测机器人尚未经过显式评估的候选动作的互信息（MI）及其相关预测置信度。这使机器人探索中的决策阶段能够以近似对数级复杂度运行，也有利于在大型、无结构且杂乱的环境中进行在线探索，因为评估和决策需要更多空间采样。我们还设计了一个目标函数，用于平衡互信息值最高的局部最优动作与预测方差较高的全局选择。大量数值仿真和数据集仿真表明，所提出的方法在不同环境中能够达到理想的效率，同时不损失探索性能。我们还为机器人领域提供了在 GitHub 上发布的开源实现代码。



(a) (b)

图1. 基于互信息的未知**非结构化**环境中的机器人主动探索。(a) 信息丰富的轨迹及其生成的占据栅格地图；(b) 生成的互信息曲面。请注意，重合的黄色方块表示起点和终点。设置了最低信息阈值，以选择信息量更丰富的探索动作，例如，左侧中部和右上方区域的信息量低于该阈值，因此未进行探索。请注意，本文中互信息的单位为 [0,1]比特。

I. 引言

近年来，机器人探索在地下环境、海洋环境和行星任务等先验未知环境中的应用日益普遍 [1]–[3]。在现有文献中，最先进的探索方法倾向于在每次迭代中使用信息论度量，例如香农互信息（MI） [4]及其衍生指标 [5]–[8], ，以准确评估候选控制动作带来的信息增益，并选择和执行信息量最大的动作。因此，探索问题自然地转化为一个序贯最优决策问题。图1展示了一个典型的探索示例。

直观而言，解决这一问题的方法是采用贪心策略，并增加候选动作，包括离散动作空间中的采样节点 [9], [10],、可用视点 [11], [12],或特殊运动基元 [13], [14], 。然而，贪心选择的探索性能与地图网格上动作空间的离散采样分辨率/方法密切相关：分辨率较粗可能导致次优动作/路径；分辨率较细则会生成更多样本，更有可能选出最优动作，但计算

在这种情况下，评估所有候选动作的信息增益的成本将变得高昂，因为前向

[1]Yang Xu、Ronghao Zheng 和 Senlin Zhang 就职于中国杭州 310027 浙江大学电气工程学院。{xuyang94,rzheng,slzhang}@zju.edu.cn[2]Meiqin Liu 就职于中国西安 710049 西安交通大学人工智能与机器人研究所。liumeiqin@zju.edu.cn[3]所有作者还就职于中国杭州 310027 浙江大学工业控制技术国家重点实验室。

†通讯作者

评估过程中的模拟需要大量的射线投射和互信息计算。值得注意的是，在三维环境中，这些影响会更加明显，因为维度增加需要更多的样本。

本文旨在实现一种更高效、准确的方法，在机器人探索中无需穷举且高成本地评估所有候选动作，即可找到信息量最大的动作。具体而言，我们的主要贡献有三点：

1）我们提出一种贝叶斯核空间互信息推断方法，仅使用部分经显式评估的样本，即可构建机器人动作与互信息值之间的连续替代评估模型，并以对数时间复杂度高效预测控制动作的互信息；

2）我们设计了一种由预测互信息值和不确定性构成的奖励函数，用于找到最佳动作，在探索与利用之间实现权衡；该方法已通过数值仿真和数据集仿真得到验证；

3）此外，我们在此[1] 公开了所提方法的开源实现，供机器人领域使用。

本文的组织结构如下。第二节介绍近期基于学习的机器人探索方法相关工作。第三节对问题进行建模，第四节介绍我们基于贝叶斯核的互信息推断方法。第五节给出基于合成数据和真实世界数据集的仿真结果及讨论，最后第六节总结全文。

II. 相关工作

在机器人探索中，监督学习技术提供了一种强大的工具，可通过以下方式近似找到全局最

[1]https://github.com/Shepherd-Gregory/BKI-exploration

优值：利用连续动作空间中的少量动作训练预测模型，无需昂贵地评估目标函数；此外，这种方法在黑箱推断中也具有更好的可解释性 [15]–[17]。

在 [18], Bai 等人中，使用高斯过程（GP）对机器人探索先验未知区域时控制动作与显式评估的互信息（MI）之间的关系进行建模。在 [19], 中，他们进一步将贝叶斯优化（BO）引入信息论机器人探索，通过多次迭代优化高斯过程预测，从而快速降低地图熵并确保计算效率。通常，贝叶斯优化假设目标函数具有一个先验分布，并构建预测模型来描述机器人动作与其互信息之间的潜在关系。它还会评估基于高斯过程先验和样本导出的采集函数，随后选择使采集函数最大化的下一个查询点，并平衡探索（全局）与利用（局部）之间的权衡。随着观测（训练样本）增多，贝叶斯优化通过迭代不断提高后验分布的精确度。Francis 等人 [17]没有评估离散视点，而是将自主探索与建图任务建模为一个约束贝叶斯优化问题，旨在寻找最优连续路径。

然而，上述基于贝叶斯优化的机器人探索方法的主要瓶颈在于，训练动作的数量N 会直接影响最终的预测精度以及计算成本。这意味着，要获得更高的探索性能，就必须付出高昂的计算代价。通常，更新和查询高斯过程模型（贝叶斯优化背后的核心）的总体时间复杂度为 $\mathcal{O}(N^3)$。这不可避免地损害了机器人探索任务的推断效率和实时性能，尤其是在大规模三维场景中。

近年来，深度神经网络（DNNs）被引入，以更高效地预测最优感知动作。Bai 等人 [20] 使用大量随机生成的二维地图训练深度神经网络，从而生成建议动作，并确保推理时间为常数。图神经网络（GNNs）也与强化学习方法相结合，以从探索图而非度量地图或视觉图像中学习最佳动作 [21], [22]。然而，基于神经网络的机器人探索方法需要预先使用大量训练样本，而且在不同环境中的适应性和泛化能力也受到限制，未来可能需要对此开展进一步研究。

令人鼓舞的是， [23] 提出的贝叶斯核推断（BKI）技术，使我们能够在简化模型上进行高效的精确推断，而不必在精确的生成模型（例如高斯过程）上进行代价高昂的近似推断。BKI 将局部核估计扩展到针对指数族似然函数的贝叶斯推断，使推理的运行时间仅为 $\mathcal{O}(\log N_q)$（N_q：查询样本的数量）。这些显著优势拓展了 BKI 在机器人学中的应用，包括传感器不确定性估计 [24], 高速导航 [25],，以及利用稀疏传感器测量进行环境建图，例如地形可通行性建图 [26], 3D 占据状态建图 [27],语义映射 [28]。

受 [19] 的启发， [23], 使用 BKI，以高效且闭式的方式推断空间互信息，从而确定那些尚未通过高计算开销（例如 [4]）明确评估互信息值的控制动作。与 [18] 和 [19], 等现有工作相比，我们的方法保持了与先前方法相近的准确性，同时在需要大量明确评估样本的复杂场景中表现得更加高效且适用。

III. 预备知识与概念

为便于讨论，本文主要考虑在二维环境中，利用配备视场（FOV）有限的基于波束的测距传感器的移动机器人进行信息论探索。本文的结果也可方便地扩展到三维情形。

A. 信息论探索

通常，机器人会在其可行配置空间 $\mathcal{X} \subseteq SE(2)$ 内生成一组候选动作$\mathcal{X}_{action}$ 。我们还假设，该配置空间已按固定分辨率在二维静态网格地图上离散化。数值集合 $m \in [0, 1]$ 表示独立栅格单元的占据水平，并可通过经典的对数几率法 [29]进行更新和查询。假设未观测地图单元 ξ 的占据值服从均匀分布，即$p(m_\xi) = 0.5$。

这里，我们采用经典的香农互信息 [4] 作为候选配置 $x_i = [p_i^x, p_i^y, \psi_i] \in \mathcal{X}_{action}$的信息度量，其中 p_i^x和 p_i^y 表示机器人在地图上的位置， ψ_i 表示机器人的航向角。从信息论的角度看， x_i 的期望信息增益可根据当前地图熵以及给定在 x_i处获得的新测量时的条件熵进行评估：

$$I(m; x_i) = H(m) - H(m|x_i). \tag{1}$$

信息论机器人探索的目标是选择使期望互信息 x_{best} 最大的最佳动作：

$$x_{best} = \underset{x \in \mathcal{X}_{action}}{\operatorname{argmax}} \ I(m; x_i). \tag{2}$$

值得注意的是，每种配置的互信息都可以分解为独立波束上的互信息，再通过射线投射映射到栅格单元，随后对各栅格单元上的互信息进行累加以求近似，其时间复杂度最坏为地图分辨率的平方级 λ_m [7]。这也增加了机器人探索的评估成本。

B. 贝叶斯广义核推断

考虑一个基于监督学习的预测随机模型推理问题 $p(\mathbf{y}|\mathbf{x})$，给定一系列观测 N observations $\mathcal{D} = \{(\mathbf{x} = \{x_i\}, \mathbf{y} = \{y_i\})\}_{i=1}^N$，其中$\mathbf{x}$和$\mathbf{y}$ 分别表示已评估配置的集合及其产生的互信息值 $I(m;\mathbf{x})$。主要目标是推理待评估目标输入$\mathbf{x}^*$ 的后验分布 $p(\mathbf{y}^*|\mathbf{x}^*, \mathcal{D})$。通过在潜在空间 $\mathbf{\Theta}$中为输入$\mathbf{x}$关联潜在参数 $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^N \in \mathbf{\Theta}$ ，即可解决该问题，其中似然 $p(\mathbf{y}|\boldsymbol{\theta})$已知。

因此，对$\mathbf{y}^*$ 的推理可以表述为对与$\mathbf{x}^*$相关的目标参数 $\boldsymbol{\theta}^*$ 进行推理:

$$p(\mathbf{y}^*|\mathbf{x}^*,\mathcal{D}) = \int_{\boldsymbol{\Theta}} p(\mathbf{y}^*|\boldsymbol{\theta}^*)p(\boldsymbol{\theta}^*|\mathbf{x}^*,\mathcal{D})d\theta^*, \tag{3}$$

其中，潜变量的后验分布可以使用贝叶斯定理表示为: $\propto \int_{\boldsymbol{\Theta}} \prod_{i=1}^{N} p\ y_i|\theta_i\ p\ \theta_{1:N}\ \boldsymbol{\theta}^*|x_{1:N}\quad {}^*\ d\theta_{1:N}.$ $(\boldsymbol{\theta}^*|\mathbf{x}^*,\ \mathcal{D})$ ( ) (, , x)

通过强假设潜在参数 $\theta_{1:N}$ 在给定目标参数 θ^*的条件下相互独立：$p(\theta_{1:N},\ \boldsymbol{\theta}^*|a_{1:N},\ \mathbf{x}^*) = \prod_{i=1}^{N} p(\theta_i|\boldsymbol{\theta}^*,\ x_i,\ \mathbf{x}^*)p(\boldsymbol{\theta}^*|\mathbf{x}^*)$，可以边缘化潜变量 $\theta_{1:N}$，进而得到 $p(\boldsymbol{\theta}^*|\mathbf{x}^*,\ \mathcal{D}) \propto \prod_{i=1}^{N} p(y_i|\boldsymbol{\theta}^*,\ x_i,\ \mathbf{x}^*)p(\boldsymbol{\theta}^*|\mathbf{x}^*)$。

BKI 进一步定义了一种具有特殊平滑性约束的分布，其扩展似然$p(y_i|\boldsymbol{\theta}^*, x_i,\mathbf{x}^*)$ 与由 f表示的似然 $p(y_i|\theta_i)$ 之间的 Kullback-Leibler 散度（KLD） $D_{KL}(g||f)$ 有界；也就是说，满足 $D_{KL}(g||f) \leq \rho(\mathbf{x}^*,\mathbf{x})$ 的最大熵分布g 具有如下形式： $g(\mathbf{y}) \propto f(\mathbf{y})^{k(\mathbf{x}^*,\mathbf{x})}$，其中 $\rho(\cdot,\cdot)$: $\mathcal{X}\times\mathcal{X} \to \mathbb{R}^+$是平滑性界，而 $k(\cdot,\cdot)$: $\mathcal{X}\times\mathcal{X} \to [0,1]$ 是可由 ρ唯一确定的核函数。代入式（3），可得:

$$p(\boldsymbol{\theta}^*|\mathbf{x}^*,\mathcal{D}) \propto \prod_{i=1}^{N} p(y_i|\boldsymbol{\theta}^*)^{k(\mathbf{x}^*,\mathbf{x})}p(\boldsymbol{\theta}^*|\mathbf{x}^*) \tag{4}$$

因此，使用指数族中的似然并假设相应的共轭先验，即可精确推断后验分布。

IV. 用于机器人探索的贝叶斯核推断

为了高效评估在空间动作空间中采样的未知机器人配置的精确互信息，我们通过贝叶斯核推断来解决这一问题。

A. 贝叶斯核空间互信息推断

如第三节.B 所述，我们假设互信息值$\mathbf{y}$ 与潜在参数 $\boldsymbol{\theta}$ 之间的底层似然模型服从均值未知、协方差 $\boldsymbol{\Sigma}$固定且已知的高斯分布 $\boldsymbol{\mu} \in \mathbb{R}^N$:

$$p(\mathbf{y}|\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}), \boldsymbol{\Sigma} = diag(\sigma^2) \in \mathbb{R}^{N\times N}, \tag{5}$$

因此，其共轭先验也可以用高斯分布来描述，使用超参数 ζ 和目标样本输入$\mathbf{x}^*$:

$$p(\boldsymbol{\mu}|\mathbf{x}^*) = \mathcal{N}\left(\boldsymbol{\mu}_0(\mathbf{x}^*), \frac{1}{\zeta(\mathbf{x}^*)}\boldsymbol{\Sigma}(\mathbf{x}^*)\right), \tag{6}$$

其中， $\boldsymbol{\mu}_0$ 和 ζ 分别是均值的初始信念及该信念的不确定性。 $\zeta = 0$表示没有置信度，而 $\zeta \to \infty$ 表示拥有完整的先验知识。在探索未知区域时，由于我们对该信念没有太多先验信息，因此假设 ζ 是一个相当小的正数。

因此，给定观测 $\mathcal{D}$，我们可以将公式（6）和公式（5）代入公式（4）：

$$p(\boldsymbol{\mu}^*|\mathbf{x}^*,\mathcal{D}) \propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma^2}k(\mathbf{x}^*,x_i)\right) \cdot \exp\left(-\frac{1}{2}\frac{(\mu_i-\mu_0)^2}{\sigma^2}\zeta\right), \tag{7}$$

由此可以推导出互信息的均值和协方差的后验分布，如下所示:

$$\begin{aligned} I(\mathbf{x}^*) = \mathbb{E}[\mathbf{y}^*|\mathbf{x}^*,\mathcal{D}] = \mathbb{E}[\boldsymbol{\mu}^*|\mathbf{x}^*,\mathcal{D}] = \frac{\overline{\mathbf{y}}+\zeta\mu_0}{\zeta+\overline{k}} \simeq \frac{\overline{\mathbf{y}}}{\overline{k}}, \\ \sigma_I(\mathbf{x}^*) = \mathbb{V}[\boldsymbol{\mu}^*|\mathbf{x}^*,\mathcal{D}] = \frac{\boldsymbol{\Sigma}}{\zeta+\overline{k}} \simeq \frac{\boldsymbol{\Sigma}}{\overline{k}}, \end{aligned} \tag{8}$$

其中$\mathbf{y}$ 和 k 可通过核函数计算:

$$\overline{k} = \Sigma_{i=1}^{N} k(\mathbf{x}^*,\mathbf{x}),\ \overline{\mathbf{y}} = \Sigma_{i=1}^{N} k(\mathbf{x}^*,\mathbf{x})y_i. \tag{9}$$

给定一组显式评估的观测 $\mathcal{D}$ 作为输入，然后我们便可使用公式（8）和公式（9）轻松计算互信息及对应的置信度，针对测试空间配置 $\mathbf{x}^*$ 。

B. 核函数选择

BKI 方法中的核函数会直接影响计算效率和精度，因此选择合适的核函数至关重要。在 [26]–[28],中，所选的稀疏核会移除远离查询点的训练点，从而利用 k-d 树，以对数级运行时间高效且精确地评估观测结果（例如占据状态、可通行性和语义类别）。

与建图任务中通过机载传感器获取充足训练数据不同，机器人探索在有限空间内的每个时刻总是生成并评估相对较少的候选配置，因此在机器人探索任务中无需剔除罕见的训练样本。

在指数核函数中，我们更倾向于使用 Mat´ern 核，因为它能够处理地形 [30], [31],的突变；这是因为应用场景中的潜在障碍物和未知结构此前从未被观察到，它们会使互信息值发生显著变化。典型的 Mat´ern 核函数如下:

$$k(\mathbf{x}^*,\mathbf{x}) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}r}{\ell}\right), r = ||\mathbf{x}^*-\mathbf{x}||, \tag{10}$$

其中，正参数 ν 和 ℓ 分别为平滑度常数和特征长度尺度， $\Gamma(\cdot)$ 和K_ν 分别为伽马函数和修正贝塞尔函数。实际应用中，我们选择形式为 $k(\mathbf{x}^*,\mathbf{x}) = (1+\frac{\sqrt{3}r}{\ell})\exp(-\frac{\sqrt{3}r}{\ell})$ 的 Mat´ern 3/2 核（$\nu = 3/2$）。

C. 基于 *BKI* 的机器人探索

在机器人探索中，我们希望机器人朝着预测互信息值较高的地点移动，以最大化局部信息增益；但这种贪心式“探索”可能会导致不理想的路径，甚至可能更糟，比如被困在杂乱区域。因此，预测不确定性较高的未探索地点也值得探索，因为它们

可能有助于机器人在先验未知区域中规划全局最优路径，这也被称为“利用”。

因此，我们将互信息值的预测置信度与预测的互信息相结合，以实现探索与利用之间的权衡；随后，我们可以根据公式（2）和公式（8），获得使信息目标函数最大化的建议动作：

$$x_s = \underset{x \in \mathcal{X}_{action}}{\operatorname{argmax}} \ \alpha I(m; x_i) + (1-\alpha)\sigma_I(x_i), \qquad (11)$$

其中，$\alpha \in [0,1]$ 是权衡因子。

基于我们提出的 BKI互信息推断方法的自主探索框架如算法1所示，其中算法2是BKI优化模块。

算法1 BKI探索( )

要求： 第 k个时间步的占据栅格地图 m_k，先前的机器人位姿 $x_{hist} = x_{0:k-1}$ 和当前位姿 x_k，显式评估样本数量 N，信息阈值 I_{th}，查询样本数量 N_q，while循环计数上限 N_{loop}

迭代 = 0
当 $x_{hist} \neq \emptyset$ 且迭代 $< N_{loo}$ $_p$ **do**
  迭代 = 迭代 + 1
  //示例 N 训练动作
  $\mathbf{x} \leftarrow Sampling(x_k, m_k, N)$;
  // 显式评估这些动作式（1）
  对每个 $x_i \in$**x 执行**
    $m_{virtual} \leftarrow Raycasting(x_i, \ m_k)$;
    $I_i \leftarrow ComputeMI(m_{virtual})$;
    $\mathbf{y} \leftarrow \mathbf{y} \cup I_i$;
  结束循环
  $\mathbf{x}^* \leftarrow Sampling(x_k, \ m_k, \ N_q)$;
  //使用算法*2*查找建议动作
  $\{x_{best}, I_{best}\} \leftarrow BKIOptimization(\{\mathbf{x},\mathbf{y}\}, \mathbf{x}^*)$;
  如果 $max(I_{best}) > I_{th}$ **则**
    $x_{k+1} \leftarrow x_{best}$(最大信息索引);
    $x_{hist} \leftarrow x_{hist} \cup x_{k+1}$;
  else
    $x_{k+1} \leftarrow x_{k-1}$; // 返回上一个动作
    从 x_{hist} 中移除 x_{k-1} ;
  结束 if
  //执行该动作并更新地图
  $P_{local} \leftarrow Astar(x_k, x_{k+1})$ // 通过 *A** 规划局部路径
  $m_{k+1} \leftarrow OccupancyGridMapping(P_{local})$;
结束 while 循环

命题1 我们提出的方法在算法*1*的每个 *while* 循环步骤中的时间复杂度为：

$$\underbrace{\mathcal{O}(NN_zN_c^2)}_{\textit{explicit MI evaluation}} + \underbrace{\mathcal{O}(N_{epoch}N\log N_q)}_{\textit{BKI MI inference}} \qquad (12)$$

其中，N_{epoch} 是训练轮次的数量，N_z 和 N_c分别是每个传感器扫描的波束数量，以及最坏情况下一个波束与栅格地图相交的栅格单元数量。

算法2 BKI优化( )

要求： 训练集 $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$、当前动作集待评估 $\mathbf{x}^*$，训练轮次 N_{epoch}，因子 α

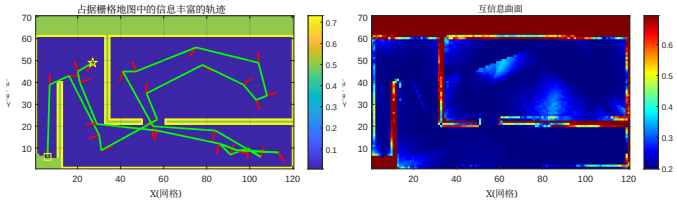
$x_{best} \leftarrow \{\}$, $I_{best} \leftarrow \{\}$;
对每个轮次 **执行**
  //使用公式（11）计算核函数
  $k \leftarrow KernelFunction(\mathbf{x}^*, \ \mathbf{x})$;
  // 使用公式（8）计算互信息和不确定性
  $\overline{k} \leftarrow \Sigma k, \ \overline{y} \leftarrow k \cdot \mathbf{y}$;
  $I^* \leftarrow \overline{y}/\overline{k}, \ \sigma_I^* \leftarrow \Sigma/\overline{k}$;
  $ObjFunc \leftarrow \alpha I^* + (1-\alpha)\sigma_I^*$;
  $x_s = max(ObjFunc)$;
  if $x_s \in$**x 然后**
    $x_{best} \leftarrow x_{best} \cup x_s, I_{best} \leftarrow I_{best} \cup y_s$
  else
    // 使用式（1）显式计算互信息
    $I_s = CalculateMI(x_s)$;
    //添加到 $\mathcal{D}$
    $x_{best} \leftarrow x_{best} \cup x_s$,$\mathbf{x} \leftarrow \mathbf{x} \cup x_s$;
    $I_{best} \leftarrow I_{best} \cup I_s$, $\mathbf{y} \leftarrow \mathbf{y} \cup I_s$;
  结束 if
结束循环
返回 x_{best}, I_{best}

值得注意的是，文献 [18] 中的基于 GP 的机器人探索方法和文献 [19] 中的基于 BO 的方法，在显式互信息评估方面的时间成本与我们的方法相同，但这两种方法分别需要以 $\mathcal{O}(N^3 + N^2N_q)$和 $\mathcal{O}(N_{epoch}(N^3 + N^2N_q))$的计算复杂度执行昂贵的 GP 推断以计算互信息。这一理论比较结果表明，我们的基于 BKI 的探索方法在时间效率上优于基于 GP 的方法，尤其是在需要更多样本 N 和 N_q 来快速评估的大规模、杂乱环境中。

V. 结果与讨论

在本节中，我们使用配备 3.6 GHz Intel i3-9100F CPU 和 32G 内存的台式电脑进行数值仿真和数据集实验，以验证所提出的基于 BKI 的机器人探索方法的有效性。信息阈值为 $I_{th} = 0.05$ 比特，权衡因子为$\alpha = 0.5$。在所有仿真中，我们采用 Mat´ern 核函数作为高斯过程的核函数，核参数为 $\ell = 1$ 和 $\nu = 3/2$ 。我们还为 BKI 方法选择参数 $\zeta = 0.001$ 和 $\sigma = 0.01$ 。假设机器人位姿已知，并在测距传感器的视场内均匀采样机器人的候选动作。我们针对所有地图进行了 20 次蒙特卡洛试验。

我们采用基于贪心的优化（在仿真中称为“NBO”）、仅使用 1 个轮次进行优化的批量 GP（“批量 GP”） [18], 以及使用多个轮次的基于 GP 的贝叶斯优化（“GP-BO”） [19] 与我们的方法进行比较：一种是仅使用 1 个优化轮次的“批量BKI”，另一种是使用多个轮次的“BKI-BO”。同时，为了验证时间效率，我们为每种方法采用了 $N = 30$ 和 $N = 60$ 个样本两种设置，其中 GP-BO 30



(a) 信息丰富的轨迹　　(b) 互信息曲面

图1。未知**结构化**环境中基于 BKI 的机器人探索示例。黄色方块：起点；黄色星形：终点；红线：机器人每次动作时的方向。

以及 BKI-BO 30 使用 $N_{epoch} = 15$ 次迭代；GP-BO 60 和 BKI-BO 60 在 BKI 优化中使用 30 个轮次。我们还在所有仿真中设置了$N_q = 8N$ 。

A. 合成环境结果

为模拟室内和野外场景，我们生成了两张 24 m × 14 m 的合成地图：一张是由若干墙壁围绕的结构化迷宫地图（如图2, N_{loop} = 50所示），另一张由圆形和椭圆组成的非结构化地图（如图1, N_{loop} = 150所示）。两张地图的分辨率均为 0.2 m。模拟测距传感器的视场为 ±1.5 rad，分辨率为 0.05 rad，最大感知范围为 6 m。机器人初始位置为 [1.2 m, 1.2 m] ，航向角为 0 rad，并尝试探索先验未知地图。最大化信息目标函数所得到的代表性路径如图1和图2所示。

结构化地图和非结构化地图的定性结果分别如图 3 和图 4 所示。为了直观地比较不同方法的探索性能，我们在图中展示了各方法地图熵和覆盖率的变化过程，其中实线和虚线表示各方法在蒙特卡洛试验中的均值，阴影区域表示标准差。

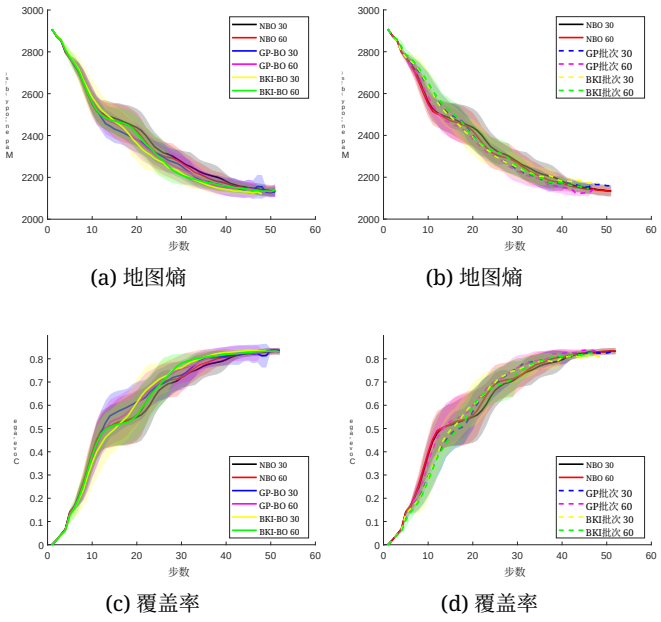
图 3 显示，BKI 和 GP 方法的性能与 NBO 方法相近，因为这个结构化场景相对较小且简单，尤其是在初始阶段，只有一条走廊可供前进。相比之下，图 4 表明，NBO 方法需要花费更多时间（约 50∼70 步）才能收敛并结束探索，而 BKI 和 GP 方法在熵减和覆盖率方面与 NBO 方法相当，也能完成探索。

此外，如图 5 所示，我们将显式计算的互信息作为真实值，并在随机选定的一个步骤中，使用少量训练样本通过 BKI-BO 和 GP-BO 方法计算互信息预测误差。这表明，在面对具有挑战性的情况时，基于 BKI 的方法在互信息推断精度方面可以与基于 GP 的方法相媲美。

简而言之，这些结果验证了我们的 BKI 方法在典型的结构化场景和非结构化场景中具有与基于 GP 的探索方法相媲美的性能。

B. 数据集结果

为了在更复杂的环境中测试我们的方法，我们选择包含狭长走廊的西雅图地图 [32]



(a) 地图熵　　(b) 地图熵

(c) 覆盖率　　(d) 覆盖率

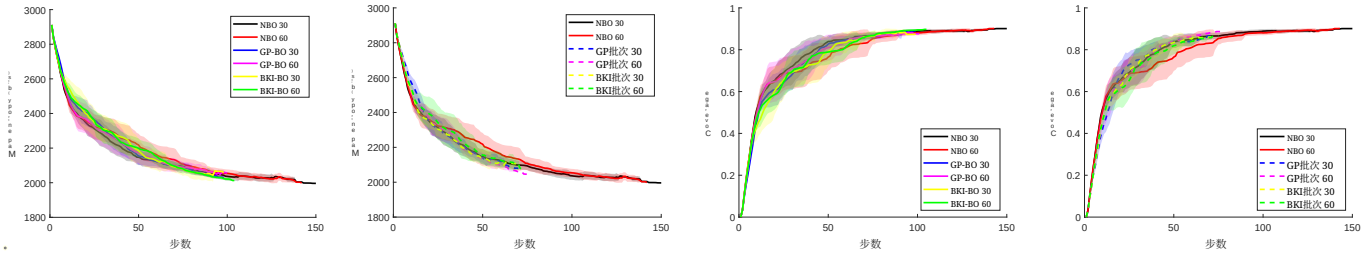
图 3. 合成结构化地图的地图熵和覆盖率结果。

以及杂乱房间，如图 6 所示。地图尺寸为 24m×14m，分辨率为 0.2 m。我们使用一台模拟激光扫描仪，在视场为 $\pm\pi/3$ rad、最大量程为 4 m 的范围内均匀发射 20 个波束。机器人从 [13, 57] m 处出发，初始航向角为 $-\pi/2$ 。 N_{loop} 设为 100。

图 7 展示了地图熵和覆盖率的对比曲线。图 7(a) 显示，探索开始后，BKI-BO 方法的地图熵下降得更快，且最终达到的水平低于其他方法，其中 BKI-BO 60 表现最佳。在这个典型的杂乱地图中，GP-BO 方法略逊于我们的 BKI-BO 方法，但几乎迎头赶上；它们也远胜于 NBO 方法。图 7(b) 中的曲线表明，批量 GP 和批量 BKI 的表现相近。我们还可以从图 7(c) 和 (d) 中看出，BKI-BO 方法的覆盖率曲线略早于 GP-BO 方法收敛，且达到的数值更高；所有基于 BO 的方法探索未知区域的速度都远快于 NBO 方法。这一结果证明，我们的 BKI 方法更适用于大型杂乱环境。

C. 时间效率

我们已经展示了在前述典型场景仿真中的探索结果。与最先进的方法相比，我们基于BKI的方法在效率和准确性方面均表现出理想的探索性能。为了进行更直观、具体的比较，我们进一步分析了所有地图中各方法每个探索步骤的时间成本。如表I所示，结果以均值和标准差的形式展示了整个探索过程中每一步的时间成本，以及不同方法在每一步中用于评估和决策的时间平均占比。



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

图4. 合成非结构化地图的熵和覆盖率结果。

表I 不同探索方法的时间成本比较

方法	合成结构化地图	合成非结构化地图	西雅图地图 [32]
NBO 30	95.29% / 10.4455 ± 0.9409	96% / 12.1683 ± 1.3856	96.95% / 4.8434 ± 0.7311
NBO 60	95.38% / 10.9967 ± 1.0676	95.93% / 12.4971 ± 2.1583	96.93% / 5.3502 ± 0.9009
GP批次 30	5.15% / 0.4387 ± 0.0246	4.66% / 0.2805 ± 0.0232	12.68% / 0.2134 ± 0.0169
GP批次 60	6.44% / 0.4444 ± 0.0487	5.89% / 0.3021 ± 0.0362	14.94% / 0.2291 ± 0.0226
BKI批次 30（我们的方法）	3.05 / 0.4324 ± 0.0276	2.93% / 0.2485 ± 0.0346	7.67% / 0.2036 ± 0.0254
BKI批次 60（我们的方法）	3.87% / 0.4407 ± 0.0384	3.56% / 0.2731 ± 0.0356	9.05% / 0.2065 ± 0.0229
GP-BO 30	49.71% / 0.9435 ± 0.0609	48.09% / 0.6083 ± 0.1121	67.97% / 0.5203 ± 0.0554
GP-BO 60	**74.03% / 1.8265 ± 0.1189**	**72.99% / 1.3558 ± 0.1190**	**84.26% / 1.0528 ± 0.1124**
BKI-BO 30（我们的方法）	39% / 0.7518 ± 0.0683	39.14% / 0.514 ± 0.0966 30	54.03% / 0.3903 ± 0.1175
BKI-BO 60（我们的方法）	**53.74%/ 0.9952 ± 0.1061**	**54.31% / 0.7363 ± 0.1186**	**62.45%/ 0.4955 ± 0.1775**

注：每步推理的时间成本（百分比）/各方法每步探索的总时间成本（秒）

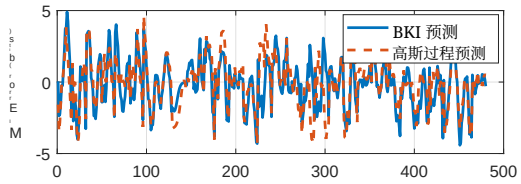
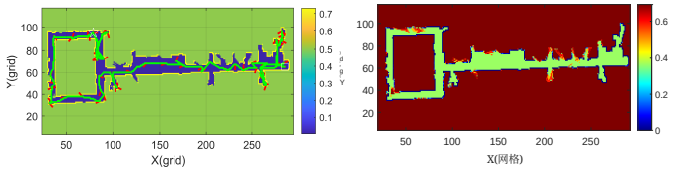


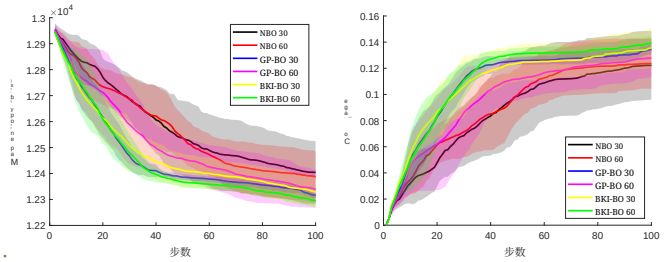
图 5. 在随机选取的探索步骤中，使用较少样本训练的 BKI 和 GP 方法进行互信息预测误差比较的一个具有挑战性的示例。



（a）探索轨迹 （b）互信息曲面

图6. 在大型杂乱西雅图地图中进行基于 BKI 的机器人探索示例 [32]。白色方块：起点；白色星形：终点。

在这10种方法中，基础NBO方法每一步的时间开销最大（约为BKI和GP方法的 8~50倍以上），而其他基于GP和BKI的方法耗时则少得多，体现了基于贝叶斯优化的方法的效率。我们可以从两个方面进一步分析这些结果。从上排到下排，在所有地图中，我们基于BKI的方法在决策和推理方面的时间效率均优于相应的基于GP的方法



（a）地图熵率 （b）覆盖率

图7. 西雅图地图的地图熵和覆盖率结果（省略批量方法）。

当样本数量增加时，例如 BKI批次 30/60 与 GP批次 30/60，以及 BKI-BO 30/60 与 GP-BO 30/60 的对比。我们还可以观察到，在使用更多训练轮次时，BKI 方法的运行速度快于 GP 方法。在探索过程中，BKI 方法还可显著节省时间；例如，在结构化地图中，与 GP-BO 30 和 GP-BO 60 相比，所需时间分别减少约 20% 和 45%。

从左栏到右栏，在非结构化且规模较大的杂乱地图中，上述差异愈发明显。例如，在西雅图地图中，GP-BO 30 和 GP-BO 60 的单步时间成本分别降低了约 25% 和 53%。这也验证了，与其他方法相比，我们提出的基于 BKI 的机器人探索方法能够显著提高时间效率，同时不损失整体探索性能。

VI. 结论

本文主要贡献在于提出一种面向未知环境中信息论机器人探索的高效新型学习方法。具体而言，我们通过引入贝叶斯核推断方法，构建了一个连续的信息增益评估模型，用于预测大量采样机器人动作的互信息。与最先进的方法相比，该模型将互信息预测的时间复杂度降低至对数级。我们还设计了一个结合预测互信息与不确定性的目标函数，以平衡探索与利用。此外，我们在自主探索框架下，通过对不同场景进行大量仿真验证了所提方法。结果表明，总体而言，该方法在效率上优于基于贪心策略和基于高斯过程的探索方法，同时不损失探索性能；这一优势在非结构化且规模较大的杂乱场景中尤为明显。未来的工作主要包括研究使用不同的 α 值和核函数时的探索性能，并将我们的方法扩展至三维场景。

参考文献

[1]H. Azp´urua、M. F. M. Campos 和 D. G. Macharet，“面向地下环境和受限空间的三维地形感知自主探索”，载于*2021 IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，2021年，第2443–2449页。[2] J. Strader、K. Otsu 和 A.-a. Agha-mohammadi，“面向行星探测车的感知感知自主桅杆运动规划”，《现场机器人学杂志》，第37卷，第5期，第812–829页，2020年。[3] P. Stankiewicz、Y. T. Tan 和 M. Kobilarov，“静态海洋环境中自主水下航行器的自适应采样”，《现场机器人学杂志》，第38卷，第4期，第572–597页，2021年。[4] B. J. Julian、S. Karaman 和 D. Rus，“基于互信息的测距机器人建图应用控制”，*The International Journal of Robotics Research*，第33卷，第10期，第1375–1392页，2014年。[5] B. Charrow、S. Liu、V. Kumar 和 N. Michael，“利用 Cauchy-Schwarz二次互信息进行信息论建图”，载于*2015 IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，2015年，第4791–4798页。[6] Z. Zhang、T. Henderson、S. Karaman 和 V. Sze，“FSMI：用于信息论建图的香农互信息快速计算”，*The International Journal of Robotics Research*，第39卷，第9期，第1155–1177页，2020年。[7] Y. Xu、R. Zheng、M. Liu 和 S. Zhang，“CRMI：用于信息论建图的高置信度互信息”，*IEEE*机器人与自动化快报，第6卷，第4期，第6434–6441页，2021年。[8] Y. Xu、R. Zheng、S. Zhang 和 M. Liu，“基于粒子滤波器的机器人探索高置信度定位与建图”，载于*2022 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2022年，第1–7页。[9] G. A. Hollinger 和 G. S. Sukhatme，“基于采样的机器人信息收集算法”，*The International Journal of Robotics Research*，第33卷，第9期，第1271–1287页，2014年。[10] M. G. Jadidi、J. V. Mir´o 和 G. Dissanayake，“基于采样的增量式信息收集及其在机器人探索和环境监测中的应用”，*The International Journal of Robotics Research*，第38卷，第6期，第658–685页，2019年。[11] A. Bircher、M. Kamel、K. Alexis、H. Oleynikova 和 R. Siegwart，“用于三维探索的滚动时域‘下一最佳视点’规划器”，载于*2016 IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，2016年，第1462–1468页。[12] B. Charrow、V. Kumar 和 N. Michael，“用于最大化互信息的多机器人控制策略近似表示”，自主机器人，第37卷，第4期，第383–400页，2014年。[13] K. Yang、S. Keat Gan 和 S. Sukkarieh，“用于无人机在未知且杂乱环境中探索的基于高斯过程的RRT规划器”，《先进机器人学》，第27卷，第6期，第431–443页，2013年。

[14] B. Charrow、G. Kahn、S. Patil、S. Liu、K. Goldberg、P. Abbeel、N. Michael 和 V. Kumar，“采用轨迹优化进行稠密三维建图的信息论规划”。载于*Robotics: Science andSystems*，第11卷，2015年，第3–12页。[15] R. Marchant 和 F. Ramos，“面向信息丰富的连续路径规划的贝叶斯优化”，载于*2014 IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，2014年，第6136–6143页。[16] R. Oliveira、L. Ott、V. Guizilini 和 F. Ramos，“定位不确定性下安全导航的贝叶斯优化”，载于机器人研究国际研讨会。Springer，2020年，第489–504页。[17] G. Francis、L. Ott、R. Marchant 和 F. Ramos，“通过贝叶斯探索构建占据地图”，*The International Journal of RoboticsResearch*，第38卷，第7期，第769–792页，2019年。[18] S. Bai、J. Wang、K. Doherty 和 B. Englot，“连续动作空间的推理赋能信息论探索”，载于机器人研究国际研讨会。Springer，2015年，第419–433页。[19] S. Bai、J. Wang、F. Chen 和 B. Englot，“基于贝叶斯优化的信息论探索”，载于*2016 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2016年，第1816–1822页。[20] S. Bai、F. Chen 和 B. Englot，“通过深度监督学习实现移动机器人的自主建图与探索”，载于*2017 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2017年，第2379–2384页。[21] F. Chen、J. D. Martin、Y. Huang、J. Wang 和 B. Englot，“通过图上的深度强化学习实现不确定环境下的自主探索”，载于*2020 IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2020年，第6140–6147页。[22] T. Wang、R. Liao、J. Ba 和 S. Fidler，“NerveNet：利用图神经网络学习结构化策略”，载于*2018*国际学习表征会议（*ICLR*），2018年。[23] W. R. Vega-Brown、M. Doniec 和 N. G. Roy，“多元指数族上的非参数贝叶斯推断”，神经信息处理系统进展，第27卷，2014年。[24] V. Peretroukhin、W. Vega-Brown、N. Roy 和 J. Kelly，“PROBE-GK：使用广义核的预测稳健估计”，载于*2016 IEEE*机器人与自动化国际会议（*ICRA*），2016年，第817–824页。[25] C. Richter、W. Vega-Brown 和 N. Roy，“面向未知环境安全高速导航的贝叶斯学习”，载于机器人研究国际研讨会。Springer，2015年，第325–341页。[26] T. Shan、J. Wang、B. Englot 和 K. Doherty，“用于地形可通行性建图的贝叶斯广义核推断”，载于机器人学习会议。PMLR，2018年，第829–838页。[27] K. Doherty、T. Shan、J. Wang 和 B. Englot，“利用贝叶斯广义核推断实现学习辅助的三维占据建图”，*IEEE*机器人学汇刊，第35卷，第4期，第953–966页，2019年。[28] L. Gan、R. Zhang、J. W. Grizzle、R. M. Eustice 和 M. Ghaffari，“用于可扩展稠密语义建图的贝叶斯空间核平滑”，*IEEE*机器人与自动化快报，第5卷，第2期，第790–797页，2020年。[29] S. Thrun、W. Burgard 和 D. Fox，概率机器人学。MIT出版社，2005年。[30] C. E. Rasmussen 和 C. K. I. Williams，机器学习中的高斯过程。MIT出版社，2005年。[31] S. T. O’Callaghan 和 F. T. Ramos，“高斯过程占据地图”，*The I nternational Journalof Robotics Research*，第31卷，第1期，第42–62页，2012年。[32] A. Howard 和 N. Roy，“机器人数据集存储库（RADISH）”，2003年。 [在线]。网址：http://radish.sourceforge.net/