智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

场景感知的分布式多智能体目标搜索方法

马成宇, 刘华平, 葛泉波

引用本文:
马成宇,刘华平,葛泉波. 场景感知的分布式多智能体目标搜索方法[J]. 智能系统学报, 2022, 17(6): 1244–1253.
MA Chengyu,LIU Huaping,GE Quanbo. Scene-aware decentralized Monte Carlo Tree Search of target discovery[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(6): 1244–1253.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202110012

DOI: 10.11992/tis.202110012
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20221008.1008.004.html

场景感知的分布式多智能体目标搜索方法

马成宇[1], 刘华平[2], 葛泉波[3,4]
（1. 南通大学 电气工程学院, 江苏 南通 226019; 2. 清华大学 计算机科学与技术系, 北京 100084; 3. 同济大学 电子与信息工程学院, 上海 201804; 4. 南京信息工程大学 自动化学院, 江苏 南京 210044）

摘 要: 在视觉语义导航任务中, 智能体通过视觉信息, 寻找并导航到给定对象类别的目标处。然而, 大部分现有的研究都是使用基于学习的框架来完成任务, 这些研究在现实世界中应用的训练成本非常高, 可移植性很低, 并且它们只适用于单智能体, 效率低下、容错能力差。为解决上述问题, 本文提出一种基于场景感知的分布式多目标优化蒙特卡洛树搜索模型, 该模型中多智能体实时在线规划并且不需要预先训练, 利用场景感知先验知识结合观测信息实时对环境进行估计, 并且利用改进的蒙特卡洛树搜索进行路径规划以此搜索目标。在 Mtterport3D 数据集中进行的实验表明, 该模型在效率方面比单智能体有着显著的提高。

关键词: 场景图谱; 分布式; 目标搜索; 蒙特卡洛树搜索; 多目标优化; 动作规划; 多智能体系统; 视觉语义导航
中图分类号: TP391 **文献标志码**: A **文章编号**: 1673−4785(2022)06−1244−10

中文引用格式: 马成宇, 刘华平, 葛泉波. 场景感知的分布式多智能体目标搜索方法 [J]. 智能系统学报, 2022, 17(6): 1244–1253.
英文引用格式: MA Chengyu, LIU Huaping, GE Quanbo. Scene-aware decentralized Monte Carlo Tree Search of target discovery[J]. CAAI transactions on intelligent systems, 2022, 17(6): 1244–1253.

Scene-aware decentralized Monte Carlo Tree Search of target discovery

MA Chengyu[1], LIU Huaping[2], GE Quanbo[3,4]
(1. School of Instrumentation and Electrical Engineering, Nantong University, Nantong 226019, China; 2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; 3. School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China; 4. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In visual semantic navigation, agents find and navigate to the target of a given object category through visual information. However, the majority of existing studies complete the task using a learning-based framework. These studies have a high training cost in the real world, low portability, and are only suitable for single-agent systems with low efficiency and poor fault tolerance. To address the above issues, this paper suggests a decentralized multi-objective optimization Monte Carlo Tree Search model based on scene awareness. In this model, multi-agent plans online in real time and does not need to train in advance. The improved Monte Carlo Tree Search is used for path planning to search targets, and the environment is estimated in real-time by using scene awareness prior knowledge combined with observation information. Experiments in the Matterport3D dataset demonstrate that the effectiveness of the model is significantly higher than that of a single agent.

Keywords: scene graph; decentralization; target search; Monte Carlo Tree Search; multi-objective optimization; action planning; multi-agent system; visual semantic navigation

在视觉语义导航任务中, 智能体需要规划自己的动作与环境产生交互, 根据观测到的视觉信息来导航到目标位置。根据目标可见还是不可见, 该任务可分为可见场景和不可见场景。在不可见场景下, 本文将任务称作目标搜索任务。它在人工智能领域有着非常重要的地位, 也有着广泛的应用前景, 包括室内搜索、灾难救援、仓库搬运、战场探索等。同时, 视觉语义导航任务也有利于其他具有挑战性的研究, 比如: 具身知识问答[1]、

收稿日期: 2021−10−13. **网络出版日期**: 2022−10−08.
基金项目: 国家自然科学基金项目 (U1613212).
通信作者: 刘华平. E-mail: hpliu@tsinghua.edu.cn.

视觉语言导航[2]、视觉音频导航等。

近年来, 许多学者开始进行视觉语义导航任务的相关研究, 文献 [3] 提出了一种利用强化学习使智能体仅通过视觉信息导航到目标位置的方法。文献 [4] 提出了一种“面向目标的语义探索”系统来提高智能体的搜索表现。还有一些研究利用强化学习方法作为智能体动作决策模块, 包括 DQN[5]、A3C[6]、PPO[7] 和分层强化学习方法[8] 等。同时也有一些方法致力于利用监督学习[9]、迁移学习[10] 解决视觉语义导航任务。但是, 这些研究基本上都是使用基于学习的方法使智能体找到目标, 当搜索场景发生变化, 智能体还需要重新训练, 而且在真实世界中获取数据集的成本很高, 这使得方法的可移植性较差, 无法在现实场景中应用。同时这些方法都只适用于单智能体, 导致算法效率低, 容错性能差, 一个智能体执行了错误动作将导致整个任务失败。

多智能体协同技术有着很大的研究潜力, 它相比于单智能体有着更高的效率和容错能力, 使得系统可以完成单智能体无法完成的更复杂的任务。多智能体协同技术有着广阔的应用场景, 比如: 工业生产、军事对抗、家庭服务、复杂环境中的搜索救援等。但是一个错误的协作策略反而会降低智能体工作的效率, 因此智能体之间的协作策略是多智能体系统研究的重点。早期的一些研究致力于使用约束条件下寻找最优解的方法来规划最优路径[11]。随后, 越来越多的研究利用启发式算法[12-14] 规划多智能体动作来最大化对整个环境的覆盖率, 以此完成搜索[15-16]。但是这些启发式算法的收敛速度非常慢, 复杂场景下的搜索效率很低。另外, 一些学者研究了多智能体的具身任务, 比如使两个智能体在场景中找到目标重物并合作举起重物[17]、使两个机器人协作找到目标并将其搬到另一个位置[18], 但是这些研究仍然使用了基于学习的方法, 而且只考虑了两个智能体的场景, 无法适用于更多智能体的任务场景。当智能体在环境中执行目标搜索任务时, 关于关联性物体的先验知识可以有效地帮助智能体搜索目标[19-20]。这一类利用场景图谱增加搜索效率的方法同样适用于本文的多智能体任务。

蒙特卡洛树搜索算法 (Monte Carlo tree search, MCTS) 是一种通过随机采样在动作空间建立搜索树来寻找最优决策的方法。它在可以被描述为连续决策树的场景中有着十分重要的应用。在早期研究中, 蒙特卡洛树搜索通常作为 AI 游戏中的动作决策器, 特别是棋类游戏[21]。最近, 一些研究开始使用蒙特卡洛树搜索算法来解决优化问题[22-23]、作为智能体的动作规划器, 文献 [24] 提出了一种帕累托蒙特卡洛树搜索算法来进行多目标优化, 但是该算法只适用于单智能体。文献 [25] 提出了一种分布式多智能体蒙特卡洛树算法, 文献 [26] 利用分布式蒙特卡洛树搜索算法来规划机械臂动作完成三维场景重建任务。

本文提出了一种分布式多目标优化蒙特卡洛树搜索算法框架, 不需要提前训练, 智能体执行一个动作后, 利用视觉观测信息结合场景认知实时更新对场景的估计生成奖励地图, 随后通过奖励地图驱动蒙特卡洛树搜索算法重新规划智能体的动作, 如此反复直到任务成功或到达限制条件。

本文结合了多智能体系统和场景先验知识的优点来解决未知环境中的目标搜索任务。主要贡献如下:

1) 本文将单智能体视觉语义导航任务扩展到了多智能体系统, 并且利用物体间的空间位置关系作为场景先验知识, 使智能体在发现关联性物体时进行有倾向的规划, 从而提高智能体完成任务的效率。

2) 本文提出了一种分布式多目标蒙特卡洛树搜索算法, 在现有的分布式蒙特卡洛树算法的基础上结合帕累托最优原理, 实现分布式多目标优化, 使得智能体群体在探索关联性物体区域同时兼顾未被探索的区域。在不需要提前训练的情况下实时进行规划, 快速地完成目标搜索任务。

3) 本文将算法在 Matterport3D 仿真环境中实现并进行实验验证, 实验结果表明本文提出的方法相较于单智能体, 效率有着显著的提升。

1　问题描述

在多智能体视觉语义导航任务中, 智能体的目标是通过以自我为中心的视觉观测信息和目标的类别, 与其他智能体协同发现并导航到环境中目标的位置。本文旨在使多智能体系统在环境中执行动作的过程中, 根据观测到的信息, 实时更新对环境的估计, 并且重新规划动作以尽可能少的动作数寻找并导航到目标附近。

本文设定在环境S中, 存在多个不同种类的物体, 表示为集合$\boldsymbol{O}=\{O_1,O_2,\cdots,O_K\}$, 其中每一个元素$O_k(k=1,2,\cdots,K)$表示一个确定的物体种类, K表示环境$\mathcal{S}$中存在的物体种类个数。假设有N个智能体, 在每一个步骤t内智能体r在执行动作后访问过的位置以及朝向为p_r, 除了智能体r之外的所有其他智能体访问过的位置和朝向为$p_{(r)}$。本文采用 Matterport3D 中的栅格地图作为导航地图,

所以智能体的动作视为在栅格上的向前一格、向右一格和向左一格 3 个动作，因此智能体的控制指令即动作空间为$\mathcal{A}$，其共有 3 个动作，分别为前进 25 cm、左转 90°后前进 25 cm、右转 90°后前进 25 cm。在不同的场景下，也可以采用其他导航方法，如 Navmesh 方法，此情况下智能体动作空间中的动作分别为发送指令移动到与所有该智能体所在导航点相邻的导航点坐标处。所有智能体的动作集合为$a=\{a_1,a_2,\cdots,a_N\}$，$a_{(r)}$表示除了智能体r之外的所有智能体的动作，即$a_{(r)}=a\backslash a_r$。集合Q_r^t表示智能体在步骤t观测到的关联性物体信息，集合$Q_{(r)}^t$表示其他智能体观测到的关联性物体信息。任务中目标的种类记为$O_j\in O$。多智能体的目标是在尽可能少的动作数下，在环境中找到种类为O_j的物体。

2　算法框架

本文提出的模型由 3 个主要模块组成：关联物体匹配模块，奖励地图更新模块，分布式多目标蒙特卡洛树搜索动作规划模块。如图 1 所示，智能体在接收到输入目标种类后，根据场景图谱得到与目标存在关联性的物体种类。智能体执行动作后得到视觉观测，并判断是否检测到关联性物体，之后将观测信息Q_r^t以及自身的位置p_r^t发送给其他智能体，同时接收其他智能体的相关信息$Q_{(r)}^t$和$p_{(r)}^t$，并根据$Q^t=Q_r^t+Q_{(r)}^t$和$p^t=p_r^t+p_{(r)}^t$更新奖励地图。最后分布式多目标优化蒙特卡洛树搜索动作规划模块根据奖励地图重新对智能体进行动作规划。智能体之间的协作分为两个阶段：第一阶段智能体在生成奖励地图时考虑其他机器人的动作信息来更新对环境奖励的估计，第二阶段智能体在多目标优化蒙特卡洛树搜索过程中 MCTS 模拟阶段的奖励计算时考虑到了其他机器人的动作。

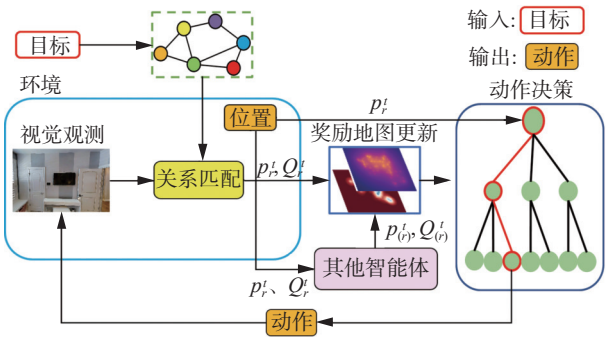


图 1　本文算法框架

Fig. 1　Algorithm framework of this paper

2.1　场景图谱

本文将场景先验知识以无向图的形式表现，场景图谱$G=\{B,E\}$，B中的节点表示不同的物体种类，边E表示两个类别物体之间特殊的位置关系。本文从视觉基因组数据集[27]中抽取场景中存在的物体种类之间的关系。该数据集包括了 100 000 多张图片，每一张图片中平均有 21 种物体，18 种属性，18 对物体之间的关系。本文从中抽取在 Matterport3D 数据集中存在的所有的物体种类之间的关系来表示场景先验知识图谱，并记录各个物体之间出现关系的频率。图 2 是场景图谱示意图，图中圆表示物体，线表示物体之间的空间关系。此外，由于本文的场景图谱是从数据集中直接提取的物体之间普适关系，所以不用提前进行预训练。

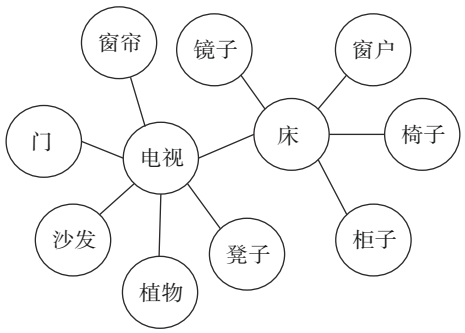


图 2　场景图谱示意

Fig. 2　Scene-aware decentralized multi-agent system for target discovery

2.2　奖励地图更新

2.2.1　奖励地图

本文根据 Matterport3D 数据集生成搜索场景的占用地图，并且利用每一个步骤智能体执行动作后得到的观测信息来产生对环境的估计。本文根据两种不同的机制对场景进行估计进而产生两种奖励地图，假设奖励地图M_{reward}为一个$2\times L\times W$的矩阵，其中L和W表示场景的长度和宽度，矩阵中的每一个元素表示环境中 25 cm×25 cm 的方格。

2.2.2　奖励地图更新机制

本节将介绍奖励地图的更新机制，在大部分搜索任务中，目标之间都会存在一定的关联性。比如椅子通常在桌子旁边。本文将这些先验知识输入到多智能体系统中，用以对场景进行估计并更新奖励地图，奖励地图更新流程见图 3。

智能体在搜索的过程中，首先根据现有的信息对 Matterport3D 中场景栅格地图上的每一个可导航点即对评估每一个栅格的奖励值更新奖励，其中单个栅格大小为 25 cm×25 cm，所有场景大小都在 150×150 栅格之内。

在更新第一张奖励地图的过程中，如果某个智能体发现了与目标O_j相关联的物体O_i，那么智

能体认为O_i的附近可能出现目标, 并推测每一个可导航点出现目标的概率。智能体假设目标O_j在关联物体O_i周围出现的概率呈高斯分布。本文将每一个导航点出现目标O_i的概率分布值作为奖励累加到奖励地图中:

$$\mathcal{F}(p_{\text{nav}}) = \frac{1}{2\pi\sigma_i^2}\exp\left[-\left(\frac{\|p_{\text{nav}} - p_{o_j}\|}{2\sigma_i^2}\right)\right]$$

式中: p_{nav}表示导航点在地图中的坐标; σ_i表示物体O_i与目标O_j之间关联性的强弱, 该参数由两个物体关系在视觉基因组数据集中出现的频率决定。智能体在与其他智能体通信交换信息后, 利用观测信息Q^t获取被发现的关联性物体的种类以及坐标, 并根据上述方法更新奖励地图 1 中各个导航点的奖励。

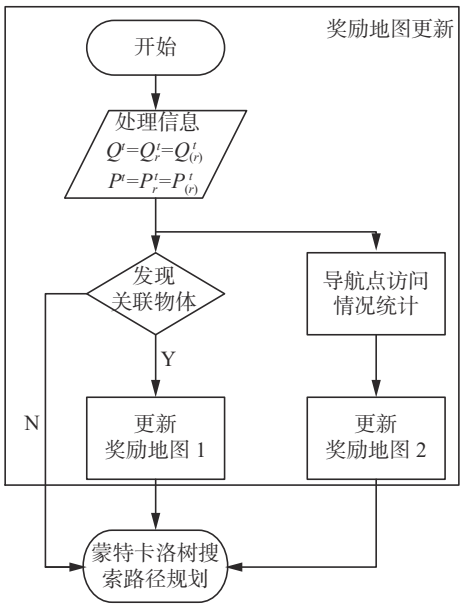


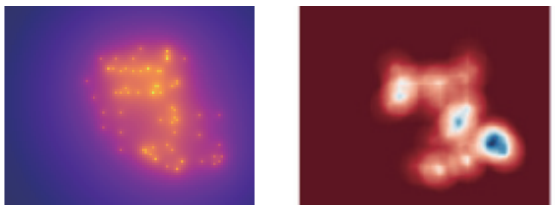
图 3　奖励地图更新过程

Fig. 3　Chart of reward map update process

同时对于第二张奖励地图, 本文设定在任务开始时, 所有导航点奖励均为 1, 随着步骤的增加, 智能体探索过的区域奖励会逐渐降低, 从而驱动智能体偏向于探索未探索过的区域, 具体更新方式为: 统计各个导航点以自身为中心周围9×9栅格范围内所有导航点的被访问情况, 其自身奖励为周围9×9栅格范围内未被访问的导航点个数和范围内所有导航点个数的比值。智能体在与其他智能体通信交换信息后, 利用所有智能体的历史位置坐标集合p, 更新各个可导航点的访问情况, 并根据上述方法更新奖励地图 2 中各个导航点的奖励。如图 4 所示, 两张地图即为根据上述方法更新后的奖励地图。

2.3　分布式多目标蒙特卡洛树搜索算法

本节将介绍分布式多目标蒙特卡洛树搜索算法规划智能体动作的具体步骤, 主要分为两个部分: 分布式智能体系统, 多目标优化蒙特卡洛树搜索算法。



(a) 奖励地图 1　　(b) 奖励地图 2

图 4　奖励地图

Fig. 4　Reward map

2.3.1　分布式智能体系统

在搜索过程中, 智能体异步、循环执行以下 3 个步骤: 1)使用多目标蒙特卡洛树搜索算法, 该算法在考虑到其他智能体动作的情况下增长搜索树; 2)通过决策树选择下一步动作a_r; 3)发送位置信息p_r和观测信息Q_r^t, 接收其他智能体的位置信息$p_{(r)}$和观测信息$Q_{(r)}^t$。无论通信是否成功, 智能体都会继续重复上述 3 个步骤直到任务成功或者达到最大步骤数。其中多智能体系统算法的全局目标函数f是所有智能体在场景中移动, 从奖励图中沿路径采集奖励向量总和。

本文中单个智能体通过优化自己的局部目标函数f_r来优化全局目标函数f达到分布式控制的效果[25], 定义f_r为智能体r在环境中执行动作a_r的全局目标函数值与执行无奖励动作$a_r^{\varnothing}$之间的差值:

$$f_r(a) = f(a_r \cup a_{(r)}) - f(a_r^{\varnothing} \cup a_{(r)})$$

式中: $a_r^{\varnothing}$为空集, 即不执行任何动作。首先, 智能体r根据奖励地图扩展搜索树$\mathcal{T}_r$; 其次, 智能体选择搜索树中访问次数最多的分支作为下一步的动作并且执行, 检测下一个坐标位置p_r^t, 获得观测信息Q_r^t; 然后, 智能体将p_r^t、Q_r^t信息发送给其他智能体, 同时接收其他智能体的相应信息$p_{(r)}^t$、$Q_{(r)}^t$; 最后, 利用这些信息更新奖励地图$\mathcal{M}_{\text{reward}}$。

2.3.2　多目标优化蒙特卡洛树搜索算法

智能体使用多目标优化蒙特卡洛树搜索算法[24]规划最优动作。在智能体每一个步骤扩展的搜索树中, 搜索树的节点代表了智能体在地图中的位置, 边代表了智能体的动作, 由于智能体有 3 个动作, 所以每一个节点可以扩展 3 个子节点。在模拟的过程中, 奖励向量通过采集两张奖励地图中的奖励获得, 并且根据帕累托最优原理选择节点。

帕累托原理: 假设$\boldsymbol{X}_m$是与选择m相关联的D维向量, 表示第m个选择, $X_{m,d}$是它的第d个元素。当且仅当满足以下条件可以称第m个选择比其他选

择n更好，即m支配n，表示为$m \succ n$：

1) $\boldsymbol{X}_m$中的任何元素都不小于$\boldsymbol{X}_n$中相应位置的元素，即$\forall d = 1,2,\cdots,D, X_{m,d} \geqslant X_{n,d}$；

2) $\boldsymbol{X}_m$中至少有一个元素大于$\boldsymbol{X}_n$中相应位置的元素，即$\exists d \in \{1,2,\cdots,D\}, X_{m,d} > X_{n,d}$。

如果只满足第一个条件，则称m弱支配n，表示为$m \succcurlyeq n$。而如果存在一个元素d_1使得$X_{m,d_1} > X_{n,d_1}$，同时存在另一个元素d_2使得$X_{m,d_2} < X_{n,d_2}$，则称m和n不可比较，即$m \| n$。

本文用两种不同的策略更新两张奖励地图，即$D=2$，则多目标优化需要解决以下最大化问题：

$$a_r^* = \arg\max_{a \in \mathcal{A}} \{f_r^1(a_r), f_r^2(a_r)\}$$

式中：a_r表示智能体r的动作；$\mathcal{A}$表示智能体的动作空间；$f_r^1(a)$、$f_r^2(a)$分别表示智能体r在执行动作a_r后，根据奖励地图 1 和奖励地图 2 获得的局部目标函数值。

如图 5 所示，在搜索过程中，智能体根据当前对环境的估计所生成的奖励地图，利用多目标优化蒙特卡洛树搜索算法规划动作，当智能体沿着路径行进时，不断利用深度相机对场景进行观测，并且与其他智能体通信交换信息。之后根据观测信息重新对环境进行估计并更新奖励地图。图 5 中，c是最大模拟次数，a_{sim}为模拟过程中随机选择的动作。

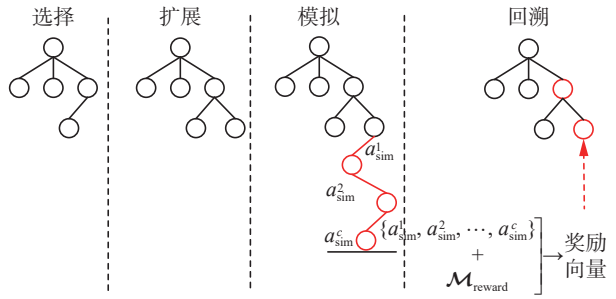


图 5　多目标优化蒙特卡洛树搜索示意

Fig. 5　Schematic diagram of multi-objective optimization Monte Carlo tree search

智能体将当前位置作为根节点扩展搜索树，多目标优化蒙特卡洛树搜索算法扩展搜索树的步骤如下：

选择：算法从根节点开始访问，逐步利用每一个子节点的帕累托置信上限向量，选出其中的帕累托最优点集，再根据智能体的偏向选择当前节点的某一个子节点作为下一个访问的节点直到访问到某个可以被扩展的节点处。

扩展：在可扩展节点处选择未被扩展的动作，得到智能体执行动作后的位置作为子节点。

模拟：根据预先定义好的策略，从新扩展的该节点进行模拟。在本文的目标搜索任务中，策略被设定为随机执行动作。智能体收集在奖励地图$\mathcal{M}_{\text{reward}}$中采集到的相应路径的奖励向量值，并且该奖励值通过局部目标函数f_r计算获得。

回溯：在进行模拟之后，得到的奖励被回溯累加到访问该节点之前，选择和扩展步骤访问过的节点上。在本文算法中，搜索树的每一个节点都保存了它被访问的次数以及累积的通过模拟获得的奖励向量值。并且利用式 (1)[24] 计算该节点的帕累托置信上限相量：

$$U(\mathcal{Z}_m) = \frac{\boldsymbol{R}(\mathcal{Z}_m)}{V(\mathcal{Z}_m)} + \sqrt{\frac{4\ln V(\mathcal{Z}) + 2}{2V(\mathcal{Z}_m)}} \tag{1}$$

式中：$\mathcal{Z}$表示当前节点；$\mathcal{Z}_m$表示$\mathcal{Z}$节点的第m个子节点；$\boldsymbol{R}()$ 表示该节点累积的奖励；$V()$ 表示该节点被访问的次数。

智能体循环执行上述 4 个步骤直到达到最大迭代次数，扩展完毕后算法选择访问次数最多的节点的边作为动作。

在蒙特卡洛树搜索算法中，每一个节点的置信上限都是标量，选择节点时都是选择置信上限最大的节点作为下一个访问节点，而本文方法所形成的置信上限是一个向量，代表了对场景的两种不同的估计：将关联性物体周围的区域作为感兴趣区域，将未被智能体探索过的区域作为感兴趣区域。这导致了智能体在动作规划中需要考虑两种不同的策略：探索和利用。本文利用帕累托最优原理，从所有子节点的置信上限向量中选择出帕累托最优点集，帕累托最优点集中的节点不可比较，不会被另一个节点支配。之后根据智能体不同的偏好来选择下一个访问的节点。这样可以达到在搜索过程中，完成智能体之间的合作，确保探索和利用策略之间的平衡。

帕累托最优点集：假设一个由点构成的集合$\mathcal{V}$，当且仅当满足以下条件时，$\boldsymbol{P}^* \subset \mathcal{V}$称为帕累托最优点集：

$$\begin{cases} \forall v_m^* \in \boldsymbol{P}^* \text{且} \forall v_n \in \mathcal{V},\ v_m^* \nprec v_n \\ \forall v_m^*,\ v_n^* \in \boldsymbol{P}^*, v_m^* \| v_n^* \end{cases}$$

智能体在扩展搜索树时的节点选择过程中，首先计算每个子节点的帕累托置信上限向量，然后使用生成的帕累托置信上限向量构建近似帕累托最优集，最后根据智能体的偏好选择最佳的节点访问。在本文实验中，智能体被设置为在任务初期，偏向于访问奖励地图 1 中奖励较高的区域来探索关联性物体周围的区域，随着动作数的增长逐渐偏向于访问奖励地图 2 中奖励较高的区域来探索之前没有探索过的区域，以获得最大化

场景覆盖率。

3 仿真实验

3.1 实验设置

本文将算法在 Matterport3D 数据集[28] 中进行实验验证, Matterport3D 数据集包括了 90 个不同的由真实房间 3 维重建生成的场景。本文从中选择 10 个场景来验证算法。对于目标种类, 本文从 Matterport3D 数据集和视觉基因组数据集中出现次数都比较多的物体种类中选择 6 个物体种类作为目标, 其包括椅子、桌子、床、马桶、电视机、水池。

3.2 实验细节

在实验中, 智能体配备了深度摄像机, 可以执行前进 25 cm, 左转 90°后前进 25 cm, 和右转 90°后前进 25 cm 三个动作。如果目标和关联性物体在智能体的视觉观测范围内、并且与智能体之间的距离小于 1.5 m, 视为智能体发现了该物体。

综上所述, 对于不同目标种类, 本文通过两种不同的实验设置来验证算法的有效性。在这两种实验中, 本文都在 10 个场景中各运行 100 个验证轮次共 1000 次实验。实验 1 在每一个轮次开始前, 从床、马桶、电视机、水池中随机选择一类作为目标。当每一个智能体在环境中进行 200 个动作后, 如果还没找到对应种类的目标, 则视为该轮次实验失败。如果过程中发现了目标, 则视为该轮次实验成功, 进入下一轮次。最后实验通过成功率以及成功轮次智能体平均路径长度作为衡量标准。通常来说, 在一个场景里会存在许多个某些种类的物体。所以第 2 个实验分别以椅子、桌子作为目标, 智能体在执行 200 个动作后, 实验通过统计所发现的目标数量占场景中该目标的总数量的比重作为算法的衡量标准。同时, 本文在定性实验中展示了在场景中寻找一个电视机和寻找所有电视机的过程。

实验中, 本文设置蒙特卡洛树搜索算法每次的最大迭代次数为 1 000 次, 最大模拟次数 c 为 5 次。

3.3 评估函数

在视觉语义导航任务中, 一般通过衡量智能体完成任务的成功率以及完成任务时的效率来评价算法的优劣。因此本文设置成功率和平均路径长度作为综合评价标准。但对于本文的多智能体视觉语义导航任务来说, 智能体之间互相协作在探索和利用之间进行平衡, 使得智能体在搜索关联性物体附近区域的同时也会逐渐重视未探索的区域来增加对场景的覆盖率。为了更加直观地体现本文提出的多目标优化的效果, 本文除综合评价标准外还设置了第二个评价标准, 从而设计了两个不同的实验。在实验 1 中, 本文使用成功率 (ω), 平均路径长度 (θ) 两种评估函数检验算法的效果, 成功率被定义为

$$\omega = \frac{1}{N_{\text{task}}} \sum_{i=1}^{N_{\text{task}}} R_i$$

式中: 第i个轮次实验成功时$R_i = 1$, 否则$R_i = 0$; N_{task}是实验总轮次数量。成功率越高多智能体搜索效果越好。

平均路径长度被定义为

$$\theta = \frac{1}{N_{\text{success}}} \sum_{i=1}^{N_{\text{success}}} \text{PL}_i$$

式中: PL_i表示智能体在第i个成功的轮次中, 移动的平均路径长度; N_{success}表示成功轮次的总数。平均路径长度越低多智能体搜索目标的效率越高。

在实验 2 中, 本文使用搜索覆盖率 (ρ) 作为评估函数来检验算法的效果, 搜索覆盖率被定义为

$$\rho = \frac{1}{N_{\text{task}}} \sum_{i=1}^{N_{\text{task}}} \rho_i$$

式中: ρ_i是智能体在第i个轮次实验中执行 200 次动作之后, 所发现的目标数量占场景中该目标总数量的比重。ρ越高说明多智能体系统能够搜索到的目标越多, 性能越好。

3.4 实验结果

实验 1 将本文算法与文献 [25] 中的算法、无场景感知先验知识的算法、随机算法分别进行比较, 实验结果见表 1。实验 1 结果表明, 随着智能体数量的增加, 本文算法与文献 [25] 中算法相比较, 在智能体数量为 2、3、4 时, 本文算法成功率和平均路径长度都优于文献 [25]。随机方法的平均路径长度会比其他方法的平均路径长度短, 因为随机算法使得智能体容易在初始点附近徘徊, 只能搜索到初始点附近的目标, 较远的目标基本无法搜索到, 成功率低, 因此平均路径长度反而会缩短。而智能体数量从 1 个增加为 2 个时, 平均路径长度不缩反长也是因为一个智能体时很难搜索远处的目标, 导致成功率很低。由实验结果可以看出, 本文算法更适用于视觉语义导航任务, 本文提出的多目标优化蒙特卡洛树搜索可以有效地提高搜索成功率, 而且多智能体系统可以大幅度地提高效率。在相同的条件下, 有场景感知先验知识模块的模型优于无场景感知先验知识模块的模型, 这说明即使没有提前训练的最普适

的物体关联性场景先验知识仍然可以有效帮助多智能体系统搜索目标。此外，虽然本文的场景先验知识总体提高了机器人搜索的成功率以及效率，但是在特别场景下反而会起到减少成功率的反作用，比如在场景中某一与目标相关联的物体周围并没有目标，而智能体仍然会在其附近搜索。因此随着场景先验准确率的提高，智能体整体性能也会有所提高。

表 1　实验 1 中 4 类算法效果对比

Table 1　Comparison of four kinds of algorithms in experiment 1

方法	N=1		N=2		N=3		N=4	
	ω/%	θ/m	ω/%	θ/m	ω/%	θ/m	ω/%	θ/m
本文方法	33.35	16.68	54.68	20.58	63.63	17.70	75.72	15.83
文献[25]方法	30.01	16.46	52.47	21.06	62.56	18.10	70.39	16.66
无场景图谱方法	28.23	17.53	45.67	24.80	59.38	21.47	63.61	18.54
随机方法	10.69	12.72	16.68	13.06	28.37	13.32	37.24	11.81

实验 2 将本文算法与文献 [25] 算法、无场景感知先验知识的本文算法相比较，结果见表 2。可以看出，随着智能体数量的增加，算法可以在有限的步骤内找到更多的桌子或椅子，表现了多智能体系统在多目标搜索任务中的优越性。同时文献 [25] 中算法效果低于本文算法，这说明多目标优化蒙特卡洛树搜索算法能够有效地平衡智能体探索和利用策略，增加了对场景的覆盖率。图 6 为 3 类算法发现目标时智能体执行动作数分布的比较图。其中，本文所提方法使得智能体发现目标的步骤数相对于无场景感知先验知识的方法来说，更加集中在前 100 步，这说明了场景图谱可以有效地帮助智能体根据物体关联性快速地找到目标。文献 [25] 方法步骤分布比起本文方法虽然相差不大，但是无法保证搜索覆盖率。

表 2　实验 2 中 3 类算法效果对比

Table 2　Comparison of three kinds of algorithms in experiment 2　　%

方法	N=1		N=2		N=3		N=4	
	桌子	椅子	桌子	椅子	桌子	椅子	桌子	椅子
本文方法	28.13	35.43	43.86	43.52	72.91	60.55	83.14	67.91
文献[25]方法	22.04	30.19	39.55	40.87	55.36	50.28	68.65	55.69
无场景图谱方法	27.63	34.25	41.48	41.74	67.73	58.39	78.35	64.14

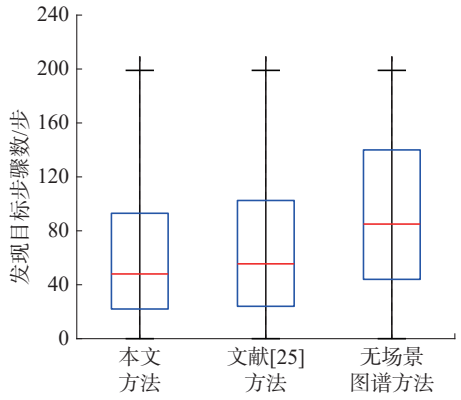


图 6　3 类算法发现目标步骤数分布

Fig. 6　Distribution of target-discovery steps of three kinds of algorithms

为了更加具体地展现算法的实现过程，本文在 Matterport3D 中的“8WUmh-Lawc2A”场景进行实验定性结果的展示。如图 7 所示，在每一个步骤中，智能体根据观测到的信息结合先验知识推测目标可能出现的区域并更新奖励地图 1，同时逐渐减少探索过区域对应的奖励地图 2 的奖励，之后智能体利用蒙特卡洛树搜索算法规划动作探索兴趣区域。可以看到，在搜索过程中智能体 1 已经很接近目标，但是视野中没有发现目标，当智能体 3 发现了关联性物体沙发后，在奖励地图 1 上更新了奖励，客厅变成了兴趣区域，使得智能体 1 探索该区域并找到目标。如图 8 所示，在场景中存在 5 个电视机作为目标，实验目的是在每个智能体移动 500 步内找到所有的电视机。在搜索过程中，智能体根据找到的关联性物体实时更新奖励地图 1，同时逐渐减少智能体已经探索过的地区的奖励地图 2 中的奖励。随着步骤的增加，智能体趋向于探索之前没有探索过的区域来最大化对场景的覆盖率，以此尽可能多地找到目标。最终，4 个智能体协作找到了场景中的所有目标。

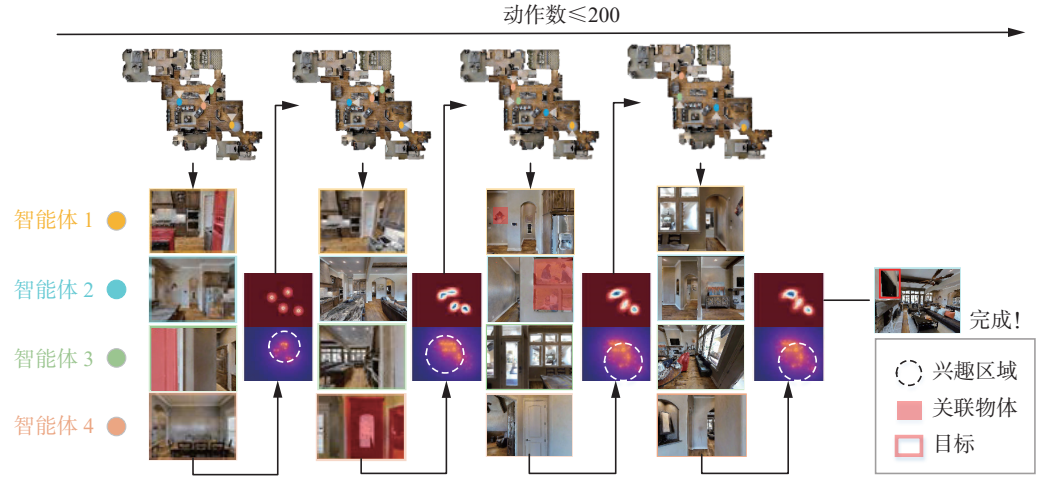


图 7　单目标搜索定性结果

Fig. 7　Qualitative results of a single target search

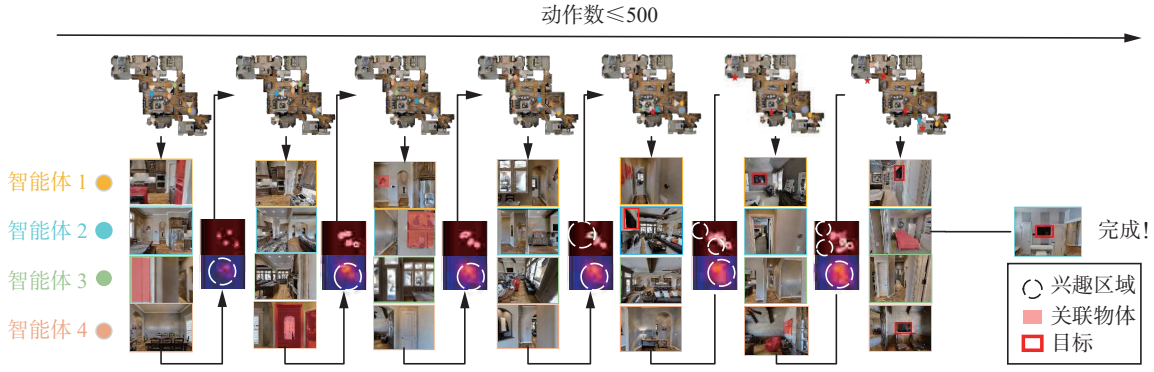


图 8　多目标搜索定性结果

Fig. 8　Qualitative results of a multi-target search

4　结束语

仿真实验结果可得: 本文提出的分布式多目标蒙特卡洛树搜索算法框架适用于室内场景的多智能体视觉语义导航任务, 无需提前训练, 智能体在场景中实时更新、实时规划。相比较于其他文献的分布式算法、无场景感知先验知识本文算法、随机算法, 本文提出的主动感知方法对环境探索的效率更高, 覆盖率更大。同时, 由于算法分布式运行, 智能体在搜索过程中即使通信没有成功也可以继续进行规划, 因此有着很强的容错能力。本文算法仅在 Matterport3D 仿真环境中进行验证, 因此以后可以在真实环境中实现算法的深入研究, 并且可以根据不同的任务需求, 通过其他算法加入对场景先验的实时更新模块, 从而更好地完成任务。

参考文献:

[1] ANDERSON P, WU Qi, TENEY D, et al. Vision-and-language navigation: interpreting visually-grounded navigation instructions in real environments[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018 : 3674−3683.

[2] DAS A, DATTA S, GKIOXARI G, et al. Embodied question answering[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018 : 1−10.

[3] ZHU Yuke, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]//2017 IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017: 3357−3364.

[4] CHAPLOT D S, GANDHI D, GUPTA A, et al. Object goal navigation using goal-oriented semantic exploration [EB/OL]. (2020−07−02)[2021−10−13]. https://arxiv.org/abs/2007.00643v2.

[5] LIANG Yiqing, CHEN Boyuan, SONG Shuran. SSCNav: confidence-aware semantic scene completion for visual semantic navigation[C]//2021 IEEE International Conference on Robotics and Automation. Xi'an: IEEE, 2021: 13194−13200.

[6] DRUON R, YOSHIYASU Y, KANEZAKI A, et al. Visual object search by learning spatial context[J]. IEEE robotics and automation letters, 2020, 5(2): 1279–1286.

[7] CAMPARI T, ECCHER P, SERAFINI L, et al. Exploiting scene-specific features for object goal navigation [C]//European Conference on Computer Vision. Cham: Springer, 2020: 406−421.

[8] YE Xin, YANG Yezhou. Efficient robotic object search *via* HIEM: hierarchical policy learning with intrinsic-extrinsic modeling[J]. IEEE robotics and automation letters, 2021, 6(3): 4425–4432.

[9] MOUSAVIAN A, TOSHEV A, FIŠER M, et al. Visual representations for semantic target driven navigation [C]//2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE, 2019 : 8846−8852.

[10] WU Qiaoyun, GONG Xiaoxi, XU Kai, et al. Towards target-driven visual navigation in indoor scenes *via* generative imitation learning[J]. IEEE robotics and automation letters, 2021, 6(1): 175–182.

[11] 王贺彬, 葛泉波, 刘华平, 等. 面向观测融合和吸引因子的多机器人主动 SLAM[J]. 智能系统学报, 2021, 16(2): 371–377.
WANG Hebin, GE Quanbo, LIU Huaping, et al. Multi-robot active SLAM for observation fusion and attractor[J]. CAAI transactions on intelligent systems, 2021, 16(2): 371–377.

[12] LIN Qiuzhen, LIU Songbai, ZHU Qingling, et al. Particle swarm optimization with a balanceable fitness estimation for many-objective optimization problems[J]. IEEE transactions on evolutionary computation, 2018, 22(1): 32–46.

[13] 楼传炜, 葛泉波, 刘华平, 等. 无人机群目标搜索的主动感知方法 [J]. 智能系统学报, 2021, 16(3): 575–583.
LOU Chuanwei, GE Quanbo, LIU Huaping, et al. Active perception method for UAV group target search[J]. CAAI transactions on intelligent systems, 2021, 16(3): 575–583.

[14] 吴莹莹, 丁肇红, 刘华平, 等. 面向环境探测的多智能体自组织目标搜索算法 [J]. 智能系统学报, 2020, 15(2): 289–295.
WU Yingying, DING Zhaohong, LIU Huaping, et al. Self-organizing target search algorithm of multi-agent system for envi-ronment detection[J]. CAAI transactions on intelligent systems, 2020, 15(2): 289–295.

[15] LUO Tianze, SUBAGDJA B, WANG Di, et al. Multi-agent collaborative exploration through graph-based deep reinforcement learning[C]//2019 IEEE International Conference on Agents. Jinan: IEEE, 2019 : 2−7.

[16] HU Jinwen, XIE Lihua, XU Jun, et al. Multi-agent cooperative target search[J]. Sensors (Basel, Switzerland), 2014, 14(6): 9408–9428.

[17] JAIN U, WEIHS L, KOLVE E, et al. Two body problem: collaborative visual task completion[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019 : 6682−6692.

[18] JAIN U, WEIHS L, KOLVE E, et al. A cordial sync: going beyond marginal policies for multi-agent embodied tasks[M]//Computer Vision–ECCV 2020. Cham: Springer International Publishing, 2020: 471−490.

[19] YANG WEI, WANG XIAOLONG, FARHADI A, et al. Visual semantic navigation using scene priors[EB/OL]. (2018−10−15) [2021−10−13]. https://arxiv.org/abs/1810.06543.

[20] QIU Yiding, PAL A, CHRISTENSEN H. Learning hierarchical relationships for object-goal navigation[EB/OL]. (2020−11−18) [2021−10−13]. https://arxiv.org/abs/2003.06749v2.

[21] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676): 354–359.

[22] 孙润稼, 刘玉田. 基于深度学习和蒙特卡洛树搜索的机组恢复在线决策 [J]. 电力系统自动化, 2018, 42(14): 40–47.
SUN Runjia, LIU Yutian. Online decision-making for generator start-up based on deep learning and Monte Carlo tree search[J]. Automation of electric power systems, 2018, 42(14): 40–47.

[23] 邱云飞, 于智龙, 郭羽含, 等. 蒙特卡洛树搜索下的整合多目标可持续闭环供应链网络优化 [J]. 计算机集成制造系统, 2022, 28(1): 269–293.
QIU Yunfei, YU Zhilong, GUO Yuhan, et al. Integrated multi-objective sustainable closed-loop supply chain network optimization under MCTS[J]. Computer integrated manufacturing systems, 2022, 28(1): 269–293.

[24] CHEN Weizhe, LIU Lantao. Pareto Monte Carlo tree search for multi-objective informative planning[EB/OL]. (2021−11−02) [2021−10−13]. https://arxiv.org/abs/2111.01825.

[25] BEST G, CLIFF O M, PATTEN T, et al. Dec-MCTS: Decentralized planning for multi-robot active perception[J]. The international journal of robotics research, 2019, 38(2/3): 316–337.

[26] SUKKAR F, BEST G, YOO C, et al. Multi-robot region-of-interest reconstruction with dec-MCTS[C]//2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE, 2019 : 9101−9107.

[27] KRISHNA R, ZHU Yuke, GROTH O, et al. Visual genome: connecting language and vision using crowd-

sourced dense image annotations[J]. International journal of computer vision, 2017, 123(1): 32–73.

[28] CHANG A, DAI A, FUNKHOUSER T, et al. Matterport3D: learning from RGB-D data in indoor environments[C]//2017 International Conference on 3D Vision. Qingdao: IEEE, 2017: 667−676.

作者简介:

马成宇, 硕士研究生, 主要研究方向为多智能体系统。

刘华平, 副教授, 博士生导师, 中国人工智能学会理事、中国人工智能学会认知系统与信息处理专业委员会秘书长, 主要研究方向为机器人感知、学习与控制、多模态信息融合。主持国家自然科学基金重点项目 2 项。吴文俊人工智能科学技术奖获得者。发表学术论文 300 余篇。

葛泉波, 教授, 博士生导师, 主要研究方向为工程信息融合方法及应用、人机混合系统智能评估。主持国家自然科学基金青年基金项目 1 项。发表学术论文 100 余篇。

第四届国际高性能大数据暨智能系统会议
The 4th international conference on high performance big data and intelligent systems

第四届国际高性能大数据暨智能系统会议(The 4th International Conference on High Performance Big Data and Intelligent Systems, HDIS 2022)拟于 2022 年 12 月 9 日至 12 月 11 日在中国天津举办。 会议旨在搭建高性能计算、大数据及人工智能领域高端前沿交流平台, 促进海内外专家学者的交流与合 作, 推动智能技术进步和智能产业发展。本次会议将汇聚全球顶级专家、学者和产业界优秀人才, 共同围绕 国际热点话题、核心关键技术、产业发展及挑战等进行开放式研讨。 会议由中国计算机学会 (CCF)、中国人工智能学会 (CAAI) 联合主办, IEEE Computer Society 技术支持, 天津理工大学、澳门大学、中国科学院半导体研究所、中国科学院深圳先进技术研究院、CCF 高性能计算专 业委员会、CAAI 神经网络与计算智能专业委员会、CAA 模式识别与机器智能专业委员会、中国智能计算产 业联盟共同承办。会议论文集将由 IEEE Xplore®出版, EI 收录, 优秀论文将会推荐至 SCI/EI 期刊发表。热 忱欢迎广大同仁踊跃投稿并莅临本届会议!

投稿要求:

1.论文未曾在国内外杂志或会议上发表。

2.稿件写作必须使用英文, 并严格按照模板要求排版。

3.所有论文采用网上投稿, 请访问会议官网进行投稿。https://www.hdis.world/public/portal/list/index/id/9.html

会议报名: 请登录会议官网 http://www.hdis.world/, 报名注册。

联系方式:

李老师, 010-82304554, hpbdis@semi.ac.cn

薛老师, 13920254011, xuewanli@email.tjut.edu.cn