超越任务性能：异构多智能体破坏性觅食中序贯协作的基于指标的分析

Alejandro Mendoza Barrionuevo*, SamuelYanes Luis, Daniel GutiérrezReina, Sergio L.Toral Marín电子工程系，塞维利亚大学，*Av. deLos Descubrimientos s/n*，塞维利亚，*41003*，西班牙

arXiv:2602.10685v1 [cs.MA] 11 Feb 2026

摘要

本文研究在部分可观测且存在时间角色依赖的条件下运行的异构多智能体系统中的协作分析问题，并将其置于多智能体破坏性觅食场景中。与大多数以往研究主要关注任务完成方面的算法性能不同，本文提出了一套系统化的通用协作指标，旨在刻画效率，以及团队与智能体之间的协调和依赖、公平性与敏感性。这些指标旨在适用于与觅食类似的不同多智能体序贯领域。所提出的指标体系分为三个主要类别，共同提供对协作的多层次刻画：主要指标、团队间指标和团队内指标。这些指标已在一个现实的破坏性觅食场景中得到验证，该场景以异构自主车辆执行动态水面清洁为背景。场景涉及两个存在序贯依赖关系的专业团队：一个负责搜索资源，另一个负责摧毁资源。研究评估了若干具有代表性的方法，涵盖基于学习的算法和经典启发式范式。

关键词：自主移动机器人、协作多智能体系统、协作分析、评估指标、异构觅食任务、部分可观测性、搜索与摧毁

1. 引言

在若干现实世界环境中，多智能体系统能够通过机器人或自主智能体的合作来完成复杂任务，例如搜救任务、清理过程或环境探索等。当智能体具有不同的感知或行动能力，且必须协调以实现共同目标时，这种范式尤为有用。

其中许多场景都可以用多智能体觅食（MAF）这一通用框架来描述。在这一问题中，智能体必须探索空间、定位并取回资源，并对其采取行动，同时平衡搜索效率与共享信息的协同利用 [1]。它通常被定义为一项称为搜索与归巢的两步任务 [2]，，但以此为起点可以衍生出许多变体 [1]。本文聚焦于所谓的破坏性觅食，即资源在被访问时会被移除或消耗 [3]。更具体地说，本文研究的是两个具有不同角色的异构智能体团队之间的序贯协作场景：一组搜索者负责觅食搜索任务，另一组清除者负责访问并移除物品。超越概念性表述——

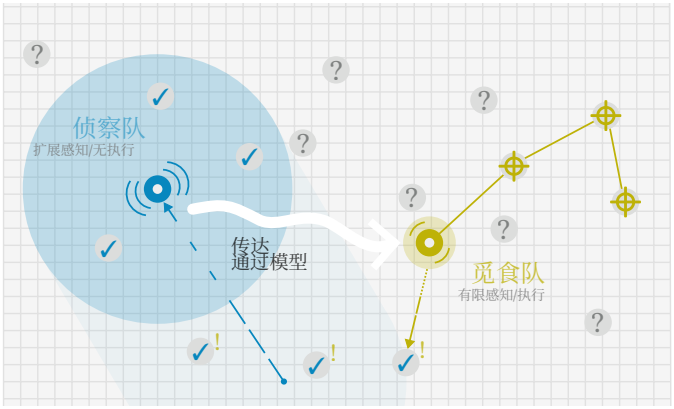

图 1：一支具备扩展感知能力的智能体团队探索环境并发现资源（侦察者），另一支感知能力有限的团队则专注于清除已发现的元素（觅食者）。

问题的……在现实世界的机器人部署中，异质性自然会出现，因为团队很少由完全相同的智能体组成。相反，由于设计和运行约束，机器人往往具有不同的形态、感知能力、移动限制和能耗特征，并且通常会采用针对互补角色进行优化的专用机器人。这导致智能体之间的信息获取和行动能力不对称，使合作不仅有益，而且必不可少——

*通讯作者电子邮箱： amendoza1@us.es; syanes@us.es; dgutierrezreina@us.es; storal@us.es()

这也进一步凸显了研究部分可观测条件下角色专业化合作的重要性。

与经典的多智能体觅食方法不同，这一场景结合了动态且具有破坏性的资源、部分可观测性，以及角色之间的时间依赖关系；这些特征在现有文献中鲜有探讨。尽管多智能体觅食问题已得到广泛研究 [1], 但现有方法大多侧重于开发针对特定任务相关指标优化性能的算法。然而，据作者所知，目前尚无可在不同问题变体中一致应用的通用指标，也没有研究专门探讨合作的定量度量，无论是异构团队之间的合作，还是单个智能体之间的合作。

在这项工作中，破坏性觅食被应用于一个现实问题：利用自主水面航行器（ASV）清理漂浮在水生环境中的塑料废弃物，这一问题也称为搜索与清理合作问题。在这一场景中，智能体必须在动态且不确定的条件下运行，废弃物会受风和洋流等环境因素影响而随时间移动。为使本研究具有普适性，我们将环境建模为一张图，其中节点表示可能包含一个或多个废弃物的空间位置。如图1所示，侦察者是轻型、快速且感知范围较广的智能体；它们每步可移动多个节点，并负责更新废弃物分布的共享模型。相比之下，觅食者速度较慢、重量较大且感知能力有限，只需访问某个节点即可移除其中的所有物品。这两种角色之间的合作具有顺序性和互补性：侦察者通过发现并传达新的废弃物位置，提升对整体情况的认知；觅食者则利用这些信息规划移动路线并清除废弃物。两类智能体均以同步方式行动，并且受能源限制，完成任务的移动步数有限。因此，总体目标有两个：在固定时间范围内尽可能多地清除废弃物，并在整个过程中维持对环境的准确建模。本研究旨在分析信息有限且存在时间依赖关系的异构智能体之间的合作如何产生并演变。为此，我们提出一组通用的协作指标，用于量化智能体之间的合作程度和协同效应，突破了传统性能指标仅关注任务完成情况的局限。因此，本研究的三项主要贡献如下：

- 基于两支团队（侦察者和觅食者）的合作，对异构破坏性序贯觅食问题进行形式化描述。该描述明确刻画了部分可观测条件下，角色分工不同的团队之间的时间依赖关系，以及基于网格的场景中的有限动态资源。
- 提出一组新颖且具有描述性的合作指标，旨在评估异构多智能体觅食系统中团队间和团队内交互的效率。这些指标针对现有多智能体觅食评估中一个众所周知的局限：不再局限于最终任务表现，并有助于更好地理解驱动合作行为的机制。
- 通过所提出的指标开展实验分析，研究不同配置和算法如何影响系统的整体表现。这些指标用于分析协调模式的涌现、依赖关系、抗噪鲁棒性以及努力分配。该分析验证了这些指标作为诊断工具的实用价值，表明它们能够帮助深入理解某些方法表现更优的原因、合作在何种条件下会减弱，以及设计选择如何影响集体行为。

2. 相关工作

过去几十年来，关于MAF和协作探索的研究显著增长，这尤其得益于机器人学和算法的进步，如 [1]所示。随着机器人系统不断发展，觅食原理逐步应用于工程化的多智能体团队，由此产生了多种用于协同搜索、资源收集和任务分配的方法。近期研究更加关注异构团队，其中智能体具有不同的感知、移动或执行能力，例如 [4]。觅食有许多变体，具体取决于不同特征，例如收集的物品类型相同还是不同（多目标觅食） [9], 、归巢是前往单一地点还是有多个地点[19], ，或者发现物品后是否会立即将其摧毁，如破坏性觅食 [3]。表1汇总了本节提到的算法。

文献 [10] 提出了一种集中式强化学习（RL）策略，旨在应对竞争性多智能体觅食场景。在其模型中，目标是在与追求相同目标的对抗团队竞争的同时，尽可能多地收集可用资源。与我们的研究不同，团队之间存在利益冲突，但他们提出的收集任务可归类为破坏性觅食。在 [11] 中，作者将觅食者建模为非破坏性搜索环境中的 RL 单智能体，目标固定不动且分散分布。智能体根据感知状态决定继续前进还是随机转向。这两项研究都仅根据收集到的资源来制定评估指标。在 [17] 中，一种异构机器人群体

表1：觅食相关工作及其所采用指标的总结。

Ref	觅食变体	算法	所采用的指标
[4]	异构 集中式 地点 MAF	遗传算法	环境平均覆盖率；资源收集率 收集率
[5]	同构搜索 MAF	莱维飞行与信息素协同方法	未探索区域，探索时间
[3]	同质 破坏性与 非破坏性多智能体觅食	Lévy 调制 相关 随机 Walk	搜索效率（访问的目标地点数与总 行进距离之比）；均方位移
[6]	同质破坏性多智能体觅食	采用莱维飞行的线性奖励-无动作 分布	覆盖区域
[7]	同质破坏性多智能体觅食	柯西分布的莱维游走	发现目标总数
[8]	异质破坏性多智能体觅食	最优联盟形成；随机 搜索；基于车道的搜索；基于网格的 搜索。	平均任务完成时间
[9]	均质与异质 多目标觅食	基于运动模式的反应式控制	交付物品数量；多样性指标
[10]	异构多智能体觅食	深度 Q 学习	平均奖励；平均资源收集量
[11]	单智能体觅食	强化学习（Projective Sim- ulation)	平均搜索效率
[12]	同质中心地 MAF	异步优势演员-评论家	食物沉积总量，消融研究
[13]	同质中心地 MAF	合作 切换 算法 for 觅食	平均觅食时间、食物返回总量、平均—— 路径长度、可扩展性研究
[14]	同质破坏性多智能体觅食	双环境群体觅食式 覆盖路径规划	时间步数、缺失点数量, 已覆盖目标点数量、平均距离
[15]	异质多重觅食	受生物觅食启发的通信 tion	功耗、端到端延迟
[16]	异构多智能体觅食	仿蚂蚁导航	成功返回次数
[17]	异构多智能体觅食	任务专用群体（S-群体）	资源收集, 探索覆盖率, Task 完成时间
[18]	单智能体多重觅食	基于先验知识的方法	平均路径长度、耗时、成功率

提出了一种基于角色专业化的方案，以优化 MAF。与我们的研究类似，该任务被划分为两个相互依赖的子任务，分别由不同团队执行：搜索和运输。尽管他们使用的性能评估指标多于以往研究，但作者仍未使用指标分析合作。

在 [12] 中，采用异步优势演员-评论家（A3C）算法解决了中心地觅食问题。该系统基于信息素协同，允许智能体使用虚拟信息素进行通信。尽管研究进行了消融研究，但对于智能体觅食行为的关注仅限于食物沉积总量这一指标，并未研究其他协作指标。在 [13], 中也探讨了类似案例，该研究提出采用一种名为觅食协作切换算法的算法，基于群体智能解决大规模系统（包含数百或数千个简单智能体）中的问题。所采用的指标评估了完成任务所需的平均觅食时间、返回巢穴的食物总量、优化路线的平均长度以及系统的可扩展性。文献 [5] 研究了仿生群体机器人学在危险放射性环境（如核设施）探索和建图中的应用。因此，这是一种仅搜索觅食策略。系统的成功主要通过探索时间和环境覆盖率来衡量按探索时间和环境覆盖百分比。

文献 [16]介绍了一种机器人导航算法，该算法基于 Formica rufa 蚂蚁觅食时的生物行为。其突出特点是采用无地图策略：机器人利用视觉地标、罗盘数据和经过的时间来记忆并重复路线。与我们的工作类似，智能体分为侦察者和觅食者两种角色。然而，其评估指标十分有限，几乎没有分析任务是否成功完成并返回。在 [14], 中，智能体的主要任务是为需要覆盖的区域进行路径规划，这些区域内设有必须访问的目标点。目标点一经访问便会消失，因此这可归类为破坏性觅食问题。所用指标仅关注智能体访问目标点所需的步数，或其访问的目标点数量。在 [15], 中，研究采用最优觅食理论中的猎物模型，将节点视为觅食者，使其根据能量和时间效率自主决定要“捕获”哪些通信信道。其他文章虽未明确将该问题称为觅食，但与之类似，可将智能体的行为建模为对资源的策略性搜索。例如, [18], 中所述的服务机器人技术，将在家中搜索动态对象视为一种觅食行为。

在 [7], 中，提出了一种基于莱维游走的搜索策略，并将其

应用于群体机器人学领域的破坏性觅食场景。该算法的新颖之处在于，每当发现一个目标时，便增加用于利用（密集搜索）的耗时。其他研究（如 [6], ）也探索了破坏性觅食，提出了用于搜索和消除随机分布目标的去中心化的基于智能体策略。在 [8], 中，研究人员考察了多无人机系统中的搜索与摧毁任务，分析了协同探索策略和依赖资源的目标消除联盟形成机制。与我们的研究不同，该研究未涉及专业化团队之间的序贯协作，其评估仅关注任务完成时间这一主要性能指标，没有考虑智能体的协作相关指标。

相关工作反映了文献中的一个更广泛趋势：尽管许多研究探讨了密切相关的多智能体觅食或搜索与清除问题，但评估方法却缺乏一致性，如表1所示。各项研究往往会定义各自特定于任务的性能指标，通常侧重于最终结果（例如，资源回收总量或任务完成所需时间），这使得不同方法难以比较。此外，即使采用了等价的指标，这些指标往往也不足以揭示促成观测到的性能表现的合作机制和交互模式。当应用黑箱优化技术（如深度强化学习或元启发式方法）时，这一局限尤为突出。在这些情况下，对所产生行为的解读很大程度上取决于单一效用函数（例如奖励或适应度），使我们无法了解合作、冗余或角色专业化究竟是如何形成的。文献中仍缺乏适用于评估异质协作觅食场景的通用指标统一框架，因为现有研究往往提出的是领域特定度量。为弥补这一空白，本文提出探讨一个互补的研究方向，重点是为协作多智能体系统建立评估指标和基准测试方法的形式化体系。所提出的指标旨在刻画覆盖效率、冗余、公平性、协作强度和故障鲁棒性等方面，从而能够超越原始任务性能，对策略进行更具可解释性和系统性的分析。

3. 问题表述与表示

本节介绍一种多智能体场景，后文将用它来阐释和评估所提出的合作指标。该形式化描述旨在提供一个通用且足够抽象的框架，用于分析部分可观测性和动态资源条件下异构团队的序贯协作动态。对环境、智能体及其能力的定义，将作为实验基础，用于研究不同协调策略如何影响集体表现，以及这种影响如何超越传统的最终结果指标。

3.1. 觅食环境

本文研究的问题由一个序贯合作破坏性觅食场景构成，其中两个异构智能体团队在动态且部分可观测的环境中运行。智能体必须共同探索并清理一片大小为 $H \times W$、表示为图 $G = (V,\ E)$ 的区域，其中每个节点 $v_{i,\ j} \in V$可包含一个或多个废弃物。边集$E\ \subseteq\ V \times V$ 定义了相邻节点之间的连通性，指定智能体在相邻位置之间可执行的所有可行转移，也就是两个相邻节点之间可能的移动方式。在本研究中，每个节点最多可与周围八个节点相连。因此，边数少于八条的节点对应于环境中邻近障碍物或边界的位置，这些障碍物或边界被假定为已知。智能体可以采取动作 $a \in A$ ，沿着一条可用的边移动。

侦察者团队由轻型、快速且感知范围较广的智能体组成，每个时间步可移动两个节点，并更新有关废弃物分布的共享模型。觅食者团队则由较重、速度较慢的智能体组成（每个时间步移动一个节点），且感知范围有限。当觅食智能体访问某个节点时，该节点上的所有废弃物元素都会被移除，从而形成破坏性的资源交互。

两类智能体均同步移动，且在每个回合中，智能体可执行的时间步数不得超过上限 T。每个智能体对环境的了解都不完整，所获信息来自共享模型。总体目标有两方面：首先，在时间范围T内尽可能多地移除废弃物；其次，尽可能减小估计的废弃物分布与实际分布之间的差异，并将其表示为建模误差 $E(t)$，通常通过计算环境真实状态与感知状态之间的均方根误差（RMSE）来衡量。这一表述与多智能体觅食领域的近期观点一致：资源收集效率与信息获取之间的联系日益受到重视 [20, 21]。这些研究指出，任务的集体表现取决于智能体感知并共享环境信息的准确程度。

车队中 N 辆车辆的位置表示为集合 $\mathcal{P}\ =\ \{p_n\ \mid\ n\ =\ 1, 2, \ldots, N\}$，其中 p_n 表示车辆 n的位置。车辆可能处于的每个位置都位于一个节点内，例如$p_n\ =\ v_{i,\ j} \in V$。图表示为大小为 $H \times W$的矩阵 $M[i, j] \in \{0, 1\}$，其中 $M[i, j] = 1$对应一个可通行节点$v_{i,\ j}$，否则为 0。

3.2. 物品动态

图 2 所示的垃圾物品集合表示为 $B\ =\ \{b_k\ =(x_k,\ \ y_k)\ \mid\ k\ =\ 1,\ 2,\ \ldots\ldots,\ \ K\}$，其中

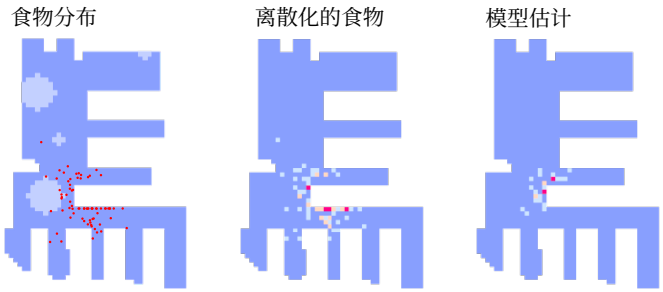

图 2：觅食环境场景，包括食物分布以及侦察者和觅食智能体不同的视野（左）、离散化的食物分布（中），以及基于智能体感知的模型估计（右）。

每个元素 b_k 表示连续参考坐标系中单个物品的空间坐标。按照觅食领域的术语，垃圾物品将称为食物。每个回合开始时，物品总数 K取自正态分布。从可访问位置中随机选取一个节点作为污染源，并围绕该位置按照多元正态分布生成 B 中的元素。因此，每个回合开始时，食物都具有独特的空间配置，且觅食过程中不会引入额外物品。物品的位置会在整个回合中动态变化，以模拟环境影响。具体而言，考虑两个组成部分：风和随机扰动。风建模为恒定速度向量 v_{wind} ，对所有物品的作用相同，并在回合开始时从均匀分布中抽样得到。通过各物品独立的速度变化 v_{rand}引入随机波动，这些变化在每个时间步分别从均匀分布中抽样。因此，每个物品的位置按如下方式更新：

$$b_k^{t+1} = b_k^t + \Delta t \left(w_{\text{wind}} \cdot v_{\text{wind}} + w_{\text{rand}} \cdot v_{\text{rand}} \right),$$

其中， w_{wind} 和 w_{rand} 为加权系数，在本研究中均设为 1。

食物的空间分布可表示为一个维度为 $H \times W$的矩阵 Y，其中每个元素$Y[i,j]$ 对应于给定时间步中集合 B内落在与节点$v_{i,\ j}$相关联的物理区域 Area(i, j) 内的物品数量。形式上，可表示为

$$Y[i,j] = \left| \{ b_k \in B \mid (x_k, y_k) \in \text{Area}(i,j) \} \right|,$$

因此， $Y[i,j] = 0$ 表示该单元格中没有资源，而 $Y[i,j]$ 的值为正则表示其中存在一个或多个食物。

每当觅食智能体访问一个单元格时，该单元格内的所有食物资源都会被视为已收集，并从集合 B中移除。假设觅食智能体的存储容量无限，且拾取过程由智能体的局部控制系统在内部管理，不会影响觅食的决策逻辑。因此，该觅食环境被视为破坏性环境。由此可将位于 $p_n \in \mathcal{P}$ 的觅食者在时间 t 收集的食物数量定义为

$$C(p_n, Y) = Y[i,j] \quad \text{such that} \quad v_{i,j} = p_n.$$

视觉覆盖范围 $\Theta(p_n)$ 表示位于位置 p_n的车辆视野内的节点集合，定义为所有位于感知半径 ρ内的节点：

$$\Theta(p_n) = \{ v \in V \mid \|v - p_n\| < \rho \}.$$

在本研究中，侦察者的感知半径设为（$\rho = 4$），而觅食者的感知半径则限制为单个节点，因为它们的视野有限。如图 2 左侧所示，智能体的视野展示了它们能够观察到的地图区域，其中侦察者的视野范围要广得多。所有智能体共享估计的食物分布，并将其表示为矩阵 $\hat{} Y[i,j]$，初始值设为零。基于这些限制，智能体在每个时间步根据观察结果（如图 2 右侧所示）更新模型，即：只要相应节点落入任一智能体的视野内， Y 中的每个单元格值$\hat{}$都会更新为与 $Y[i,j]$ 的真实值一致：

$$\hat{Y}[i,j] \leftarrow Y[i,j] \quad \text{if} \quad v_{i,j} \in \Theta(p_n).$$

4. 指标定义

如第2节所述，尽管多智能体觅食已得到广泛研究 [1], ，但大多数研究采用的评估指标难以推广，因为这些指标与特定任务或环境密切相关。此外，其中大多数指标仅关注基于性能的评估标准。据作者所知，目前仍缺少通用指标，无法持续一致地评估多智能体觅食问题不同变体之间的合作。为解决这些局限，本节提出一组广义指标，旨在评估整体系统性能，以及智能体和团队在序贯觅食任务中产生的合作交互。这些指标分为几个互补类别：

1. **主要指标**，用于评估多智能体系统在资源清除或信息准确性方面的整体性能。2. **团队间指标**，用于衡量异构团队在依赖关系和协调方面的合作程度。3. **团队内指标**，用于量化同一团队中各智能体的贡献和公平性。

4.1. 主要指标

这些指标属于最常用于评估多智能体算法的性能指标类别，因为它们概括了系统层面的整体任务完成度和效率。虽然这些指标并未明确衡量合作行为或智能体之间的互动，但它们为理解后续更详细的团队间和团队内分析提供了必要背景。在本研究中，我们扩展并系统化了这组主要指标，以建立共同基线，支持对不同算法进行一致的比较。

目标完成百分比（*PTA*）。该指标衡量系统完成主要目标的整体成效。设 U 表示与任务完成相关的效用变量（例如发现或移除的资源数量），U^{UB} 表示其对应的上界。在所述破坏性觅食情境中，对于觅食者，该指标表示在一个回合中成功从环境中移除的资源总量占最初存在的物品总数的比例；对于侦察者，则表示发现的资源总量所占的比例。其定义如下：

$$PTA = \frac{U}{U^{UB}} \cdot 100 \in [0, 100], \tag{1}$$

当 PTA 的值等于 100 时，表示智能体完全完成了任务；较低的值则表示任务仅部分完成或完成效率较低。该指标提供了一个简单的全局指标，用于衡量系统性能，且不受智能体内部协调或个体贡献的影响。

正如第2节所讨论的，这是一种以往研究中常用的指标，例如 [4, 7, 10], ，尽管这些研究采用了不同的名称或略有差异的表述。在本研究中，PTA 被引入为一种统一指标，可用于量化各种觅食场景中的任务完成程度，前提是目标或任务总数已知或可估算。该指标可从外部观测，因为它依赖于有关任务完成情况的真实信息，且无需获取智能体的内部状态或策略。因此，既可以在执行过程中在线计算，也可以在回合结束后离线计算。

均方根误差 *(RMSE)*。在觅食场景中，该指标用于量化食物分布估计地图 Y 与真实底层分布 Y之间的差异。它衡量模型的准确性，并反映智能体维护环境最新表示的能力。由于在本研究中，Y 和 Y 均为稀疏矩阵，逐元素误差会导致误差变化较大，因此采用高斯滤波器G（$\sigma = 1$）来比较食物分布的密度：

$$RMSE(t) = \sqrt{\frac{1}{|V|} \sum_{v \in V} \left(G \circledast Y_{v,t} - G \circledast \hat{Y}_{v,t} \right)^2} \tag{2}$$

数值越低，表示模型估计在各节点上越准确、越一致；数值越高，则意味着存在因信息过时、不完整或不正确而导致的不匹配。

尽管第2节讨论的相关工作中未明确采用均方根误差，但在涉及真实状态估计的问题中（例如 [22, 23]），只要有真实状态可用，均方根误差就是一种广泛采用的指标。其公式与具体任务语义无关，因此适用于任何涉及发现、跟踪或估计静态或动态潜在真实状态的觅食场景，并且可以在线或离线计算。

达到 X（NT_X）所需的归一化时间。该指标用于评估时间效率，即系统达到总目标一定比例 X 所需的速度，并以归一化尺度表示，以便比较不同领域或不同回合时长。它衡量达到目标所需时间占总可用时间的比例，定义如下：

$$NT_X = \frac{t_X}{T} \in [0, 1], \tag{3}$$

其中，t_X 表示系统首次达到总目标百分比 X 时的时间步（例如，$PTA_t \geq X$），而 T 表示每个回合可能的最大时间步数。

$T_X = 0$ 表示目标比例 X在回合开始时立即达成，而 $T_X = 1$ 则表示在规定时间内始终未能达成。这种归一化表述提供了一个无量纲的时间效率指标，可直接比较时间范围或环境动态各不相同的实验。

尽管第2节所述的先前研究并未采用这种确切形式的指标，但文献中常使用类似的完成时间指标来评估效率，例如 [8] 或 [13], ，不过这些指标通常未经归一化，或针对特定任务而设计。与 PTA 指标相同，它是一种外部可观测指标，因为它依赖于有关任务完成情况的真实信息；既可以在达到阈值 X时在线计算，也可以离线计算。

吞吐量。它提供了一个归一化指标，可用于比较智能体数量或回合时长不同的配置。形式化地说，它可以定义为：

$$Throughput = \frac{Resources\ Collected}{|N_f| \cdot T}, \tag{4}$$

其中，$Resources\ Collected$ 表示觅食智能体在回合中处理的资源总数，$|N_f|$ 表示觅食者总数，T 表示时间步数。此指标也可用于

计算侦察队搜索物品时的吞吐量，将收集的资源改为已发现的资源。

虽然以往的觅食和搜索与服务研究中也隐含地使用了类似的效率概念 [3, 11], ，但通常以回合级形式报告，而非采用归一化的每步每智能体指标。尽管该指标既可在线计算，也可离线计算，但最好在回合结束时进行评估，因为中间值可能会受到瞬态效应的显著影响。由于吞吐量采用归一化形式，因此可应用于随时间演变定义的觅食场景，无论其以连续时间还是离散步表示；只要涉及资源收集、服务或消除，且不受具体任务语义或环境规模的限制，均可适用。

空闲度指标。 该指标用于评估智能体在一段时间内对降低环境空闲度的贡献效果，即自某个区域或节点上次被智能体访问以来经过的时间。

形式化地说，节点的空闲度$v_{i,\ j}$ 在时间 t的值，记为 $I_{i,\ j}(t) \in [0, 1]$，表示节点持续未被访问时长的归一化度量。每经过一个时间步，空闲度都会根据遗忘因子 $f \in (0,\ 1)$逐步增加；该因子控制有关先前访问过的节点的信息变得过时的速度。每当节点被访问，其空闲度值便重置为零。因此，频繁被重新访问的节点保持较低的空闲度值，而被忽略时间较长的节点，其空闲度值则会逐渐趋近于 1。

环境的平均空闲度（MI）可定义为所有可访问节点的平均空闲度$|V|$:

$$MI(t) = \frac{1}{|V|} \sum_{v_{i,j} \in V} I_{i,j}(t), \tag{5}$$

空闲度降低率（IRR）衡量平均空闲度随时间下降的速率，反映智能体为保持环境得到充分探索和更新而付出的集体努力:

$$IRR(t) = -\frac{dI(t)}{dt} \approx \frac{I(t) - I(t+1)}{\Delta t}. \tag{6}$$

$IRR(t)$的值越高，表明智能体正在高效访问未探索或过时的区域，从而降低全局空闲度、提升环境感知，并维持空间新鲜度。尽管在第2节所综述的觅食文献中尚未明确采用该指标，但它在机器人探索和巡逻问题等相关领域中已被广泛使用 [24, 25]。只要可获得访问时间戳，该指标便完全可观测，并且既能在线计算以实时监测探索动态，也能离线计算，即通过对已完成回合的统计数据进行汇总。空闲度指标在需要食物搜索阶段的觅食场景中尤为有用，尤其是当食物在整个回合中持续出现并需要均衡重访时。

4.2. 团队间指标

团队间指标旨在评估具有不同角色的异构团队如何有效地相互协调并相互依赖。在顺序合作设定中，各团队的表现与其他团队所执行的动作紧密耦合，因此孤立评估并不足够。本节提出的指标刻画了这种交互的关键方面，包括时间对齐、协作依赖或退化鲁棒性等。

发现到服务延迟（*DSL*）。该指标刻画协同觅食系统中各序贯角色之间的时间依赖关系，用于衡量团队间协调的效率。形式上，DSL 定义为侦察智能体发现资源的时刻 t_d 与觅食智能体实际处理或移除该资源的时刻 t_c之间的时间间隔，并以时间步总数 T进行归一化:

$$DSL = \frac{t_c - t_d}{T}. \tag{7}$$

DSL 值越低，表示协同响应越迅速、越高效，也反映任务之间具有更强的时间耦合。

团队间时间滞后（*ITL*）。该指标量化两个协作团队在活动峰值或性能峰值方面的时间错位，反映它们在整个回合中的贡献是否同步。设 $R_1(t)$ 和 $R_2(t)$ 分别表示团队 1 和团队 2 在总计 T 个时间步内某项性能指标（例如 PTA）的时间序列。定义$t_1^* = \arg\max_t R_1(t)$ 和$t_2^* = \arg\max_t R_2(t)$ 为各团队达到峰值性能时对应的时间步。

团队间的时间滞后量计算如下:

$$ITL = \frac{|t_1^* - t_2^*|}{T} \in [0, 1], \tag{8}$$

其中 T 为该回合的总时长。

因此， $ITL = 0$ 表示团队间实现了完美的时间同步，即双方的活动峰值同时出现。 ITL 值越高，表示时间不同步程度越高，也就是说，一个团队达到性能峰值的时间显著早于或晚于另一个团队。需要注意的是，在顺序执行或角色专业化的任务中，例如先搜索再摧毁，团队间存在一定的时间偏移是意料之中、甚至是理想的。在这些情况下，ITL反映的是任务的协调结构，而非效率低下。解读指标时应始终结合具体情境，而不能孤立看待。

协作成功率（*CSR*）。 该指标量化觅食任务的成功完成对具有互补功能的团队之间合作的依赖比例。在当前环境中，它是

由侦察队首先发现、并由觅食队收集的物品数量。它可以表示为：

$$CSR = \frac{\text{Tasks performed through cooperation}}{\text{Total tasks}}. \quad (9)$$

CSR 的值较高表明，大多数觅食任务都受益于团队之间的相互依赖的协作。

随机扰动下的协作敏感性（*CSSC*）

我们引入了一项受控扰动实验，用于量化目标团队（记为 T_A）的表现如何取决于合作团队（记为 T_B）的行为质量。团队 T_B 执行随机化策略 $\pi^{\varepsilon}_{T_B}$：在每个决策步，以 $1-\varepsilon$的概率选择其名义策略 π_{T_B} 规定的行动 a，并以 ε 的概率从其动作集中均匀随机选择一个行动。扰动参数 $\varepsilon \in [0,1]$ 在完美行为（$\varepsilon = 0$）与完全随机行为（$\varepsilon = 1$）之间进行插值。

形式化地，对于每个 ε，智能体 $t_b \in T_B$ 的混合策略为

$$\pi^{\varepsilon}_{t_b}(a \mid s) = (1-\varepsilon)\,\pi_{t_b}(a \mid s) + \varepsilon \frac{1}{|A_{t_b}|}, \quad (10)$$

其中，A_{t_b} 是智能体 t_b的离散动作集。

退化曲线通过扰动参数 $\varepsilon_{T_B} \in [0,1]$逐步向团队 T_B 的动作选择策略中注入随机噪声，并测量团队 T_A由此产生的表现来获得。对于ε_{T_B}的每个取值，都有一个相应的表现值 $R_{T_A}(\varepsilon_{T_B})$ 记录下来，得到退化曲线 $\varepsilon_{T_B} \mapsto R_{T_A}(\varepsilon_{T_B})$。该曲线刻画了团队 T_A 的性能对团队 T_B 可靠性的敏感程度。

随后，采用线性回归对 $R_{T_A}(\varepsilon_{T_B})$ 与退化因子 ε_{T_B} 之间的关系进行建模。通过计算最佳拟合直线的斜率，可得到一个标量敏感性指标，称为敏感性斜率，用于量化团队 T_B 的行为受到的扰乱逐渐加剧时，团队 T_A 的性能会在多大程度上退化。由此得到的指标反映了团队间依赖程度以及协作过程的鲁棒性，其定义如下：

$$SS_{T_A \leftarrow T_B} = \text{slope}\Big(\text{linreg}\big(\varepsilon_{T_B}, R_{T_A}(\varepsilon_{T_B})\big)\Big), \quad (11)$$

SS_{T_A} 的值越高，意味着随着合作团队的可靠性降低，表现退化得越严重，表明合作的依赖性更高、鲁棒性更低。需要注意的是，退化曲线不一定总能用单一的线性近似来表示。实际上，曲线在某些区段可能呈现不同的最佳拟合斜率，对应协作过程中的不同敏感阶段。例如，当 ϵ值较小时，可能会观察到轻微的初始退化；而当其达到关键腐化阈值时，表现则会更加明显地下滑。

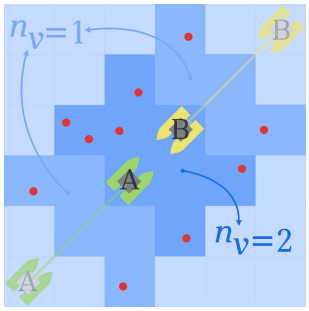

图3：两个智能体的视野中包含相同节点时，它们之间覆盖重叠的示例。

4.3. 团队内指标

团队内指标侧重于考察各团队内部的协调动态，评估承担相同角色的智能体之间如何分配职责、努力和空间覆盖。这些指标反映公平性、冗余和集体效率等方面，揭示团队表现究竟源于均衡合作，还是不均衡的贡献。

贡献基尼指数。该指标通过衡量个体对集体目标的贡献不平等程度，评估智能体之间工作量分配的公平性。它基于基尼系数 [26], 这一广为人知的不平等统计量，其取值范围为0（完全平等）到1（最大不平等）。在多智能体觅食的情境中，该指标量化了各团队发现或收集资源的均匀程度。令c_n 表示智能体 n 处理的资源数量，该智能体属于团队 T，其中 $n = 1,\ 2,\ \ldots,\ |N_T|$。贡献基尼指数定义如下：

$$G = \frac{\sum_{i=1}^{|N_T|}\sum_{j=1}^{|N_T|}|c_i - c_j|}{2\,|N_T|^2\,\bar{c}}, \quad (12)$$

其中，$\bar{c}$ 表示每个智能体的平均贡献。

$G = 0$ 表示智能体之间完全平等，即所有智能体对集体目标的贡献相等。可能的最大值取决于团队中的智能体数量 $|N^T|$。在极端情况下，若由单个智能体完成全部工作，而其他智能体毫无贡献，基尼系数将达到 $G_{\max} = (|N^T| - 1)/|N^T|$。该指标提供了一个以公平性为导向的视角，可补充效率指标，从而分析合作表现是通过均衡还是不均衡的贡献实现的。

覆盖重叠（*CO*）。该指标量化每个时间步中当前覆盖区域由多个智能体冗余覆盖的比例。它反映了空间协调的效率：较低的值表示覆盖分布良好，而较高的值则表明存在显著的冗余。

令 $\Theta(p_n,\ t)$ 表示在时间 t，位于位置 p_n 的车辆视野内的节点子集。对于

对于每个节点 $v \in V$，定义时间 t 覆盖该节点的不同智能体数量：

$$n_v(t) = \sum_{n \in N} \mathbf{1}\{v \in \Theta(p_n, t)\}, \tag{13}$$

那么，重叠度可以表示为：

$$CO(t) = \frac{\sum_{v \in V} \mathbf{1}\{n_v(t) \geq 2\}}{\sum_{v \in V} \mathbf{1}\{n_v(t) \geq 1\}}. \tag{14}$$

图3对此进行了直观展示。因此，$CO(t)$ =0 表示覆盖完全不重叠，而较高的 $CO(t)$值则表示智能体之间存在冗余探索

每个智能体的边际贡献 *(MC)*。该指标通过衡量移除某个智能体后团队整体性能的下降程度，量化该智能体在各个角色中的个体合作影响。

设 R_{full} 表示整个团队取得的某项性能指标（例如总奖励、收集的资源或期望服务成本的降低），R_{-i} 表示将智能体 i 从团队中移除后获得的性能。则智能体 i 的边际贡献定义为：

$$MC_i = \begin{cases} \dfrac{R_{\text{full}} - R_{-i}}{R_{\text{full}}} & \text{if greater values of } R \text{ are better,} \\ \dfrac{R_{-i} - R_{\text{full}}}{R_{-i}} & \text{if lower values of } R \text{ are better.} \end{cases} \tag{15}$$

接近 0 的值表明该智能体对任务结果影响很小（冗余度高或用途有限），而接近 1 的值则说明该智能体发挥着关键作用。如果该值为负，则表明移除该智能体后，系统的实际表现更好（该智能体产生了负面影响）。该指标还反映了其余智能体的韧性，因为移除智能体后性能下降幅度较小，说明团队内部具有较强的补偿能力。

5. 实验研究

本节旨在分析不同路径规划算法如何影响所研究觅食环境中的性能、信息交换和合作效率。该环境中的异构团队在部分可观测且存在时间依赖性的条件下运行。为此，本研究考察所提出的合作指标和性能指标的变化，观察各算法如何在整个觅食过程中影响协作模式、信息流和集体效率的涌现。研究特别关注团队间和团队内指标的演变，这些指标不仅揭示了整体性能，也揭示了推动该性能的底层合作动态。

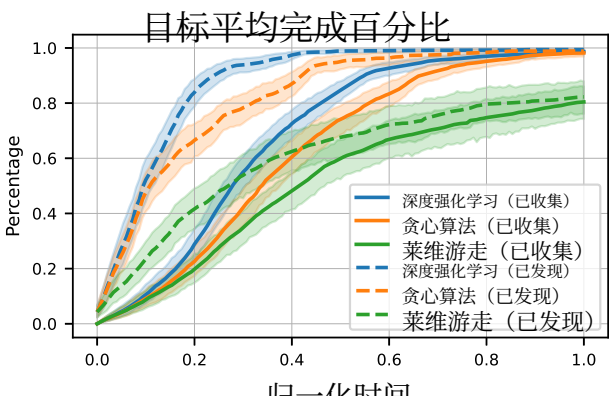

图 4：3 种算法在归一化时间内各团队目标的目标完成百分比（PTA）指标，数据为 100 个回合的平均值。置信区间以阴影区域表示。

比较分析采用了多种算法，包括现代学习方法和经典启发式方法。首先，考虑一种双重深度Q学习（DDQL）策略，其架构和训练流程遵循文献[23]，中的描述，并采用类似的奖励函数。该深度强化学习（DRL）方法按角色采用共享神经策略：同一团队类型（例如侦察者或觅食者）中的所有智能体都调用同一个神经网络，由该网络根据各自的局部观测输出动作。这对应于一种分散式执行的参数共享方案。此外，还评估了两种非学习基线方法。第一种是贪心策略，智能体选择即时奖励最大的行动，而不进一步考虑未来后果。第二种是基于莱维游走的方法：侦察智能体采用受生物觅食效率启发的莱维游走分布进行随机探索，而觅食者则使用最短路径导航（Dijkstra算法）前往最近的已知资源位置。当没有已知食物时，觅食者采用与侦察者相同的探索模式。所有算法都在相同的觅食环境中评估，均运行相同的100个回合并取平均值，且使用相同的随机种子，以确保食物动态和出现情况具有完全相同的变化性。所用代码可在Github（https://github.com/amendb/CollaborationMetrics）获取，因此结果可

n

可复现。

用于进行比较的评估场景取自一个典型的体育码头，如图 2 所示；该场景中的走廊狭窄，点与点之间通行困难，详见 [23]。场景由一个包含 $H = 62 \times W = 46$ 的矩阵构成，其中有 1297 个可访问节点。每个回合都会生成独特的食物分布。系统从所有可访问节点中随机选取一个初始热点，并在其周围按照多元正态分布随机散布物品，从而确保这些元素从排放源均匀分布。物品数量则在每个回合开始时从正态分布中抽取。

表2：主要指标：回合结束时 PTA 和 RMSE 的平均值及置信区间（CI）。

算法	$\mathbf{PTA_D}$ （%）		$\mathbf{PTA_C}$ （%）		RMSE	
	Mean	置信区间（95%）	Mean	置信区间（95%）	Mean	置信区间 (95%)
DRL	99.41	0.46	98.68	0.56	0.0010	0.0005
贪心算法	98.85	1.04	98.20	1.15	0.0017	0.0012
莱维游走	82.28	5.93	80.43	5.98	0.0165	0.0051

表 3：主要指标：达到目标收集量所需的归一化时间及最终吞吐量。

算法	$\mathbf{NT}_{50\%}^{\mathbf{PTA_C}}$	$\mathbf{NT}_{90\%}^{\mathbf{PTA_C}}$	吞吐量
DRL	0.281	0.512	0.245
贪心算法	0.349	0.583	0.244
莱维游走	0.477	0.735	0.199

每个回合开始时，物品数量也从正态分布中抽取。物品是动态的，即会在每个时间步沿两个分量移动。首先是环境洋流，以回合期间保持不变的风向表示；该风向在回合开始时从均匀分布中抽取。其次，每个时间步都会为每件物品添加一个随机运动分量，以增加运动的不可预测性。每个回合的最长持续时间为 150 步；如果智能体在时限前清除所有元素，回合也可能提前结束。这个上界为完成任务留出了合理余量，并通过统一可用交互预算，确保算法之间的公平比较，同时也能反映任务提前结束时的效率差异。环境中部署了两个团队，每个团队包含两个智能体，因此构成了一个由专业化角色组成的四智能体异构多智能体系统。保持团队规模固定有助于进行受控评估，尤其是在分析移除一个团队中的智能体以降低其行为表现所产生的影响时。

下文将详细分析所有提出的评估指标，说明它们在不同算法配置下的表现。分析将从指标的解释、比较结果，以及其对异构多智能体觅食环境中的协调动态、劳动分工和效率所提供的洞见等方面展开。

5.1. 主要指标

图4展示了不同受评算法和团队的平均 PTA 随时间的变化及其置信区间。在此背景下，该指标反映了不同团队各自的目标：对于侦察队，它衡量已发现资源（食物）的百分比 PTA_D；对于觅食队，它表示成功收集的资源百分比 PTA_C。结合表2（列出了最终均方根误差和 PTA 指标及其置信区间）对曲线进行综合解读，可以看出各算法在整体任务效率方面存在明显差异，从而能够直接比较其收敛速度和最终性能水平。

基于深度强化学习的方法在两支团队中都表现出明显更快的初期增长速度，表明其早期决策和探索与利用的平衡更为有效。表 3 也通过数值体现了这一点，其中展示了觅食队达到 $NT_{50\%}$ 和 $NT_{90\%}$ 的 PTA_C 指标所需的归一化时间，即消灭场景中相应比例的全部元素所需的时间（在回合期间）。相比之下，贪心算法和莱维游走等启发式策略的初期进展较慢。不过，贪心策略在回合中期取得了显著进展，使其表现暂时接近基于深度强化学习的方法。这一行为表明，尽管贪心算法缺乏长期规划，但当局部机会密集聚集时，它能够有效加以利用。至于莱维游走策略，由于其随机性很强，既不考虑环境的形状，也不利用从环境中收集的信息，因此在 PTA_D（已发现物品的比例）方面明显落后。不过，尽管该算法较为基础，其检测到的大多数物品最终都能被收集。莱维游走能够在非常开放的场景中实现广泛的全局覆盖，利用其探索特性进行长距离移动；但其集中搜索并利用资源丰富区域的能力有限，尤其是在元素的动态变化要求进行信息与方向更新时，因为其移动方式并不具备适应性。

图5所示的估计食物分布平均均方根误差，也呈现出与 PTA 指标相似的趋势。在这种情况下，均方根误差反映了资源时空估计的准确性。数值越低，表明智能体越能可靠地预测环境的演变。与 PTA 结果一致，基于深度强化学习的方法能够以最快且最稳定的速度降低预测误差，而贪心算法的改进较慢，但仍较为明显。莱维游走策略的均方根误差显著较高，且斜率几乎保持不变，这与其不采用基于信息的决策、而是进行探索的特性相符。这进一步支持了这一观察：尽管 Lévy 能够广泛覆盖环境，却无法改进对高价值区域的认知。

图6所示的 MI 和 IRR 指标进一步支持了这一解释。这些指标衡量地图各区域被重复访问的频率。在这种情况下，数值越低，表示对环境的覆盖越均衡、越持久。贪心算法使 MI 降低得最快且最持久，这表明它会持续将强化搜索集中在最近最少访问的区域。这种行为源于其纯粹的短视特性：智能体总是选择能够带来最高即时奖励的行动，而不考虑长期收益。因此，贪心算法倾向于积极前往当前有望提升覆盖率的位置（如 IRR 所示），而不是转而占据未来可能拥有更高资源密度的区域。不过，如图4和图5所示，这种策略对于此类问题确实表现出很强的有效性。

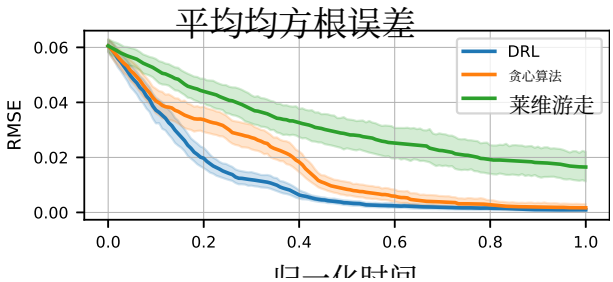

图5：物品分布随归一化时间变化的均方根误差（RMSE），为3种算法在100个回合上的平均值。置信区间以阴影区域表示。

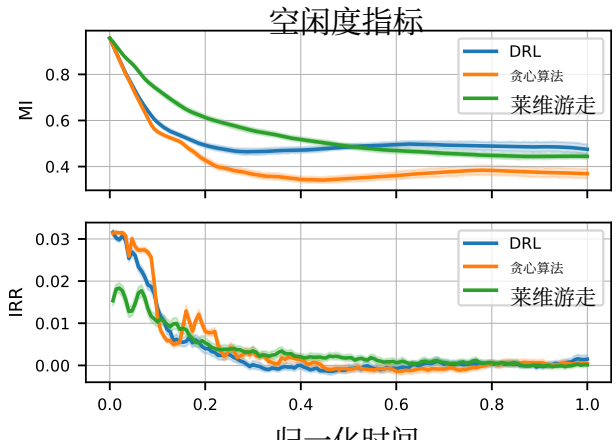

图6：归一化时间内的空闲度指标，取3种算法在100个回合中的平均值。置信区间以阴影区域表示。

能够提升覆盖率的位置（如 IRR 所示），而不是转而占据未来可能拥有更高资源密度的区域。不过，如图4和图5所示，这种策略对于此类问题确实表现出很强的有效性。

图6显示，深度强化学习（DRL）的互信息（MI）最初急剧下降，反映出其在初始阶段迅速覆盖环境，但随后稳定在高于贪心算法的水平。值得注意的是，最终莱维游走策略的互信息超过了它。这种表现似乎表明，深度强化学习侦察者在形成对环境的整体表征后，会将注意力转向更新已知含有资源的区域，而非继续广泛探索。因此，图5中的均方根误差（RMSE）持续下降，而互信息保持稳定。相比之下，莱维游走的互信息值保持适中并逐步改善，这与其固有的广泛随机探索模式相符：它会反复扫过环境中的大片区域，而不会有意优先探索特定区域。

表3还展示了吞吐量（Throughput），其衡量指标为每步每个智能体收集的平均物品数量。所得数值表明，DRL与贪心算法的生产力几乎相同，而Lévy则明显较低。这一结果与PTA和空闲度指标中观察到的趋势一致：尽管贪心算法缺乏长期战略推理，但其强烈的利用行为能够将附近的发现迅速转化为已处理物品，从而维持与DRL相当且具有竞争力的收集率。相比之下，DRL通过更加均衡的努力分配实现了相近的吞吐量，优先依据累积知识进行有依据的移动，这一点也体现在其更优的均方根误差表现上。与此同时，Lévy方法表现出的较低吞吐量与其探索性特征相符——它能够实现广泛覆盖并具备较高的发现潜力，但不太擅长将智能体集中到高价值区域进行资源移除。主要指标进一步强化了这样一种观点：相近的性能可能源自本质不同的算法，凸显了分析多个互补指标的重要性。

5.2. 团队间指标评估

图7展示了发现到服务延迟 指标，该指标衡量侦察者发现物品的时刻与觅食者成功处理该物品的时刻之间的归一化时间。分布表示显示，莱维游走策略实现了最低的平均延迟，其次是DRL和贪心算法。Lévy的这一优异性能可归因于其觅食者的确定性行为：该行为经过设计，使得当发现某个元素时，觅食者会沿着通往最近已知目标的最短Dijkstra路径移动，从而确保物品被检测到后能够快速完成资源服务。相比之下，DRL表现出略高的延迟，这可能是由于策略驱动决策在利用与更广泛的情境推理之间进行权衡，偶尔会为了长期收益而延迟即时服务。最后，对小提琴图形状进行详细检查可以发现，贪心算法较高的平均延迟似乎是由一个明显较长的尾部造成的。这表明，尽管贪心算法能够实现快速的发现到服务周期，但也经常经历极端延迟。当智能体陷入局部最优决策且无法快速到达场景中存在的物品时，就很可能发生这些长延迟事件。总体而言，这些结果表明，延迟揭示了协调效率方面的运行差异，而这些差异无法从PTA或吞吐量等单项性能指标中看出。

图8中 ITL指标的分布呈现出与算法时间协调能力相关的类似见解。它衡量侦察者（PTA_D）与觅食者（PTA_C）的峰值PTA达成之间的时间偏差，反映两个团队在完成各自子目标时的同步程度。结果表明，贪心策略的表现再次略差。对于深度强化学习，分布更集中在均值附近，表明发现峰值与服务峰值之间的同步性更加稳定。这种稳定

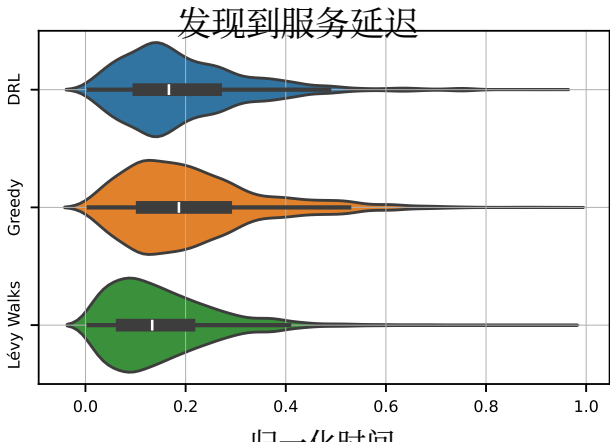

图7：3种算法在100个回合中发现至服务延迟（DSL）指标的分布情况。

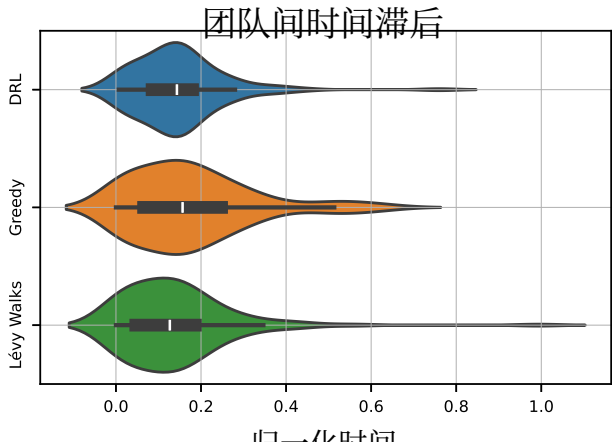

图8：3种算法在100个回合期间的团队间时间滞后（ITL）指标分布表示。

性可归因于觅食者相对于侦察者的策略性位置安排：通过学习，觅食者会隐式利用资源更可能在其附近被发现这一预期，从而减少不必要的移动，使 ITL值紧密聚集。相比之下，Lévy和贪心算法的小提琴图形状更圆润、分布更分散，反映出该指标具有更大的变异性。因此，这是一个结果与DSL指标相似、但方法略有不同的指标，因为它关注的是性能峰值。

另一个有参考价值的指标是协作成功率，它量化了通过两个团队明确协作完成的任务比例，也就是最初由侦察者发现、随后由觅食者处理的物品所占比例。图9展示了该指标随时间的变化，表 4则报告了最终数值。起初，三种算法的协作表现相近；但在回合进行约20%时，深度强化学习（DRL）和贪心算法的协作表现明显提升，且最终数值相近。这可以用算法中设计的操作依赖关系来解释：深度强化学习（DRL）和贪心算法中的觅食者只专注于服务已知物品，因此高度依赖侦察者进行发现。于是，随着侦察者发现有关食物位置的信息，协作率自然上升。相比之下，当已知物品模型为空时，莱维游走觅食者也会进行探索性移动，遵循随机的莱维模式。虽然这能扩大环境覆盖范围，但也降低了对侦察者的严格依赖，因此整体协作成功率较低。对该指标的分析反映了设计选择如何直接影响团队间协作的形成及其效率。

该随机扰动下的合作敏感性分析评估了当一支团队的行为逐渐退化时，另一支团队的表现会受到怎样的影响。为此，我们逐步提高侦察者策略中随机动作所占的比例，比例由

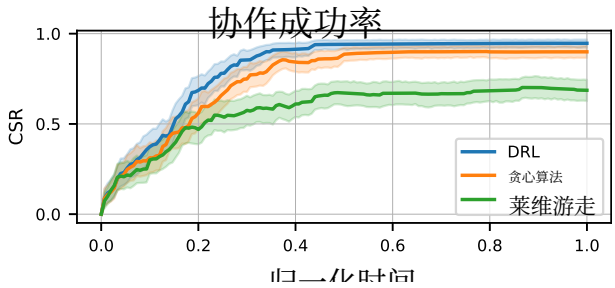

图9：3种算法在100个回合中的协作成功率（CSR）指标随归一化时间的变化，取平均值。置信区间以阴影区域表示。

参数 $\varepsilon \in [0,1]$ 控制，增幅为 0。05；与此同时，觅食者遵循其正常的决策策略，随后再以相反的方式进行评估。为确保公平比较，所有算法均使用相同的随机种子生成随机动作。图10展示了随着 ε_{scouts} 增加，评估回合结束时 PTA_C指标的变化情况。正如预期，随着侦察者的可靠性降低，所有算法的表现均有所下降。不过，最佳拟合直线的斜率（SS）揭示了不同的敏感程度，表 4 对此进行了数值展示。

这种行为可以从算法特性来解释。对于 Lévy 而言，其敏感性较低，部分原因在于其基线性能本身较低——它起步时的 PTA 显著低于 DRL 和贪心算法所达到的近乎最大值，因此性能下降的空间较小。此外，Lévy 的探索策略本就高度依赖随机移动模式，因此进一步增加随机性所造成的扰动相对较轻。相比之下，DRL 和贪心算法都高度依赖侦察者：它们的觅食者将努力全部用于服务已知物品，因此更容易受到侦察者行为遭到扰动的影响。在这两种方法中，DRL 表现出明显更强的鲁棒性，这反映在其较低的绝对 $SS_{scouts}^{PTA_C}$ 上。这可归因于其学习策略能够在

噪声条件下进行泛化；而贪心算法严格追求即时奖励，不具备任何前瞻性结构，因此在随机扰动下性能下降得更为迅速。

图11报告了预测模型误差（均方根误差 ）在参数 $\varepsilon_{foragers}$增大时，随机扰动如何影响误差的变化。所有算法的均方根误差都会随着 Epsilon 增大而上升，但退化的幅度和速度各不相同：Lévy 的退化最为显著；贪心算法的斜率略高于深度强化学习，而深度强化学习对引入的噪声最为稳健。这种模式由若干相互关联的原因所致。当觅食者受到扰动，未能移除已发现的资源时，会有两种影响共同导致侦察者维持的均方根误差升高。首先，已检测到的物品会在环境中停留更久并继续移动，因此侦察者先前的观测很快就会过时；预测模型因而会累积系统误差。其次，未能移除物品会增加环境中的元素数量，因此更难及时更新所有元素的位置，进而降低模型质量。发现到服务的循环中断得越久，这两种影响就越强，而这正是参数 ε 增大时发生的情况。算法特有的因素也会调节这种敏感性。深度强化学习侦察者（以及整体的深度强化学习算法）能够学习有关食物空间分布和动态的行为，从而进行更有信息量的更新，并采用隐式恢复策略。这种适应性使其在觅食队受到扰动时的 SS_{forag}^{RMSE}更小。贪心侦察者虽由局部信息和即时奖励驱动，但仍会依据观测到的物品位置采用确定性规则，因此其退化速度快于深度强化学习，但慢于Lévy。相比之下，Lévy侦察者基本遵循随机探索轨迹，且不会优先进行模型驱动的更新；因此，当觅食者未能移除物品时，侦察者既不会通过对高价值位置重新采样来进行补偿，导致均方根误差增幅最大。此外，根据先前的指标也可以推断，Lévy将部分探索任务交由觅食者完成。系统越依赖觅食者来维持对环境真实情况的掌握，当这些觅食者受到扰动时，预测模型的退化就越严重。

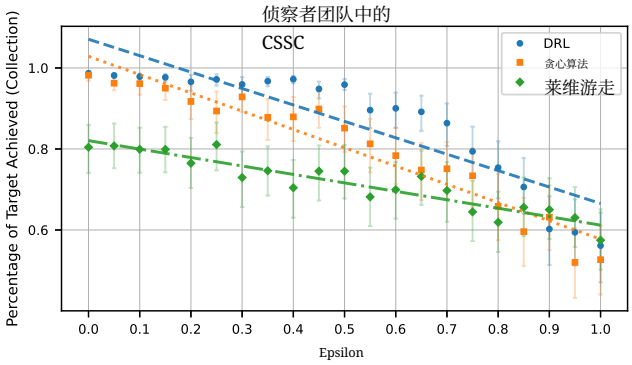

图10：侦察队在采集性能指标（PTA_C）上的随机扰动下的协作敏感性（$CSSC$）。最佳拟合直线展示了各算法下采集性能的下降速率。图中包含95%的置信区间。

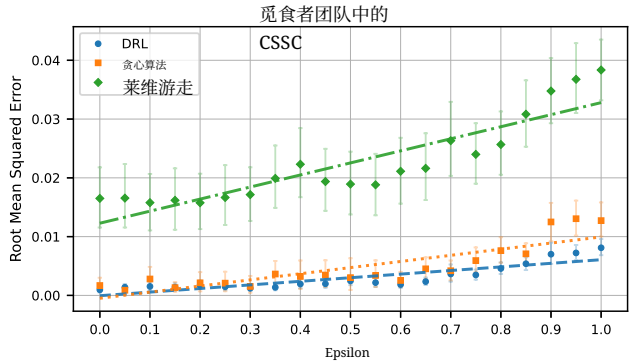

图11：觅食队在模型均方根误差上的随机扰动下的协作敏感性（$CSSC$）。最佳拟合直线展示了各算法下模型的退化速率。图中包含95%的置信区间。

最后，正如指标定义中所述，图10和图11显示，退化曲线具有
噪声引入后，不同程度的性能退化会呈现出明显不同的阶段。特别是在图10中，基于深度强化学习的方法表现出两阶段特征：随着 ϵ增加，性能在初始区域内缓慢下降；超过某个破坏阈值（约为 $\epsilon = 0.5$）后，性能则会显著下滑。这表明存在两个截然不同的敏感性区间：在初始的鲁棒区间内，学习策略能够容忍适度的随机扰动；而在第二个区间中，一旦噪声超出已学习的协作模式，合作便会骤然恶化。相比之下，贪心算法的性能下降更为渐进。这一现象在实际应用中尤其具有重要意义，因为这表明，在轻度策略退化（例如通信噪声、模型误差或部分故障）下，基于深度强化学习的合作比贪心算法更具韧性，尽管两者最终都会在严重破坏下失效。

5.3. 团队内指标评估

图12展示了应用于每个团队内觅食者工作负荷的基尼系数随时间的平均变化。由于每个团队中只有两名觅食者，该系数的最大可能值为0.5（其中一名完全工作，而另一名处于不活动状态），最小值为0（工作完全均衡）。在回合开始时，所有算法的数值都较为相近，约为0.2–0.3，这反映了由于探索初始阶段以及元素发现不规律所导致的中度不平等但仍不稳定的分配。然而，大约在回合长度的20%至后续阶段之间，可以观察到明显的分化：DRL和贪心算法都开始持续降低其内部不平等程度，并在回合约60%处稳定在明显低于初始值的水平。

表 4：团队间指标：最终平均合作成功率和敏感度斜率指标。

算法	CSR	$SS^{PTA_C}_{侦察者}$	$SS^{RMSE}_{觅食者}$
DRL	0.946	-40.55	0.0061
贪心算法	0.899	-45.14	0.0104
莱维游走	0.686	-20.86	0.0205

在DRL中，最终的平衡更加明显，表明其觅食者之间具有更好的运作协调和更高效的工作分配。相比之下，在莱维游走中，曲线呈现出明显更平滑、几乎平坦的下降趋势，意味着智能体之间服务负载的不平等持续存在。这一结果更有趣之处在于，DRL和贪心算法所表现出的更高公平性，并非源于明确施加的合作分配机制，而是源于它们在地图上的运作方式或位置选择所产生的涌现效应。在DRL中，这一效应尤其明显，因为其学习策略融合了环境状态和其他智能体位置的信息，这表明它们学会了选择能够避免干扰和冗余的动作，否则其奖励将会降低。尽管DRL具备这一优势，但这似乎并不是其在基尼系数方面具有良好公平性的唯一原因。贪心算法虽然缺乏对队友的感知，并且仅针对即时局部奖励进行优化，但同样实现了相对均衡的工作负荷。这种平衡表明，它并非源于合作推理，而是由目标变得可用的速度所诱导的性质所造成：如上文第5.1节中PTA指标的变化所示，DRL和贪心算法都表现出快速且持续的新元素发现，从而使目标被频繁纳入环境模型。在这些条件下，单个采集者无法独自利用所有已发现的资源，因此两个智能体不可避免地会承担原本会仅由其中一方累积的任务，进而产生涌现的公平性。相比之下，莱维游走产生的发现模式更慢且更不规律，资源以更小的批次、间隔更长的时间出现。当每次只有少量目标处于已知状态时，距离这些目标最近的采集者往往会垄断可用任务，而另一名采集者到达该位置时已经太晚。因此，工作负荷的不平衡会持续更长时间，这反映在显著更高的基尼系数值上。

图13展示了侦察队覆盖重叠指标的演变，揭示了各算法截然不同的行为特征。正如预期，所有方法都出现初始峰值，因为所有智能体均从同一区域出发，因而观察到相互重叠的区域。回合开始后不久，重叠急剧下降：深度强化学习和贪心算法在回合时长约10%时达到接近零的数值，而莱维游走则需要更长时间才会下降。Lévy表现出更长的初始重叠，因为其

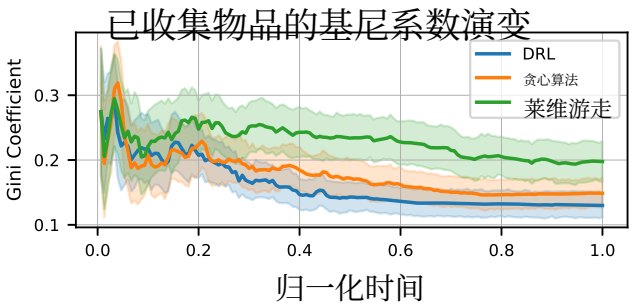

图12：觅食队随归一化时间变化的基尼系数指标，取3种算法在100个回合中的平均值。置信区间以阴影区域表示。

智能体遵循包含许多短步的伪随机轨迹，在空间上分离所需时间更长；这与深度强化学习和贪心算法不同，后两者引入更多方向性决策，从而迫使更快分离。一旦分离，Lévy在回合剩余时间内几乎保持零重叠，且置信带极窄。这反映了其高度探索性与随机性：智能体持续在全局范围内分散，覆盖大体上彼此分散的区域。

相比之下，深度强化学习和贪心算法在初始下降后均呈现逐步上升的趋势，分别稳定在约0.15和0.10。这表明，尽管两种策略都优先考虑覆盖，但它们往往会在出现高价值资源簇的区域汇聚。在贪心算法中，这种行为自然产生：当发现高密度区域时，由于接近最近资源会获得奖励，两个采集者都会被吸引到相同的目标。在深度强化学习中也会出现类似现象，不过在这种情况下，智能体学会了即便重叠区域中的奖励需要共享，向高价值区域汇聚仍然有益，因为这能提高共享模型的准确性，而这正是奖励函数本身的设计所致。由于深度强化学习会优化长期奖励，侦察者会策略性地调整位置，靠近预期回报较高且更有可能通过新发现获得额外奖励的区域。适度的重叠可能有益，使侦察者能够强化对动态变化的资源位置所进行的局部更新；但过度重叠则会浪费覆盖。

如第 5.1 节所分析，表2报告了完整系统——两名侦察者和两名觅食者——所达到的绝对性能，为所有算法提供了共同的参照。深度强化学习和贪心算法在物品发现和清除方面均达到近乎完美的性能，基线均方根误差极低，约为 10^{-3}；相比之下，莱维游走的模型误差显著更高，任务完成度更低。表5列出了用于计算边际贡献指标的相应消融设置，其中移除了一名侦察者或一名觅食者。

在所有算法中，移除一名侦察者似乎会

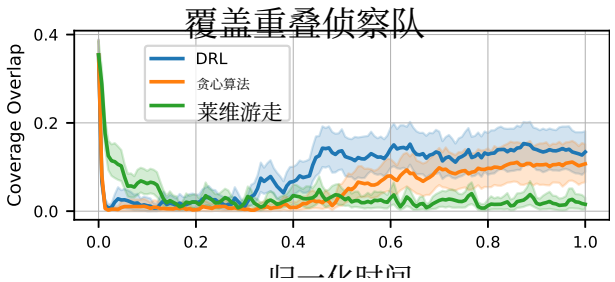

图13：侦察队的覆盖重叠（CO）指标随归一化时间的变化，针对3种算法在100个回合上取平均。置信区间以阴影区域表示。

表5：团队间指标：用于边际贡献分析的消融团队配置与完整配置下的性能对比。

算法	配置	PTA_D (%)	PTA_C (%)	RMSE
DRL	完成	99.41	98.68	0.0010
	1S–2F	87.92	93.55	0.0077
	2S–1F	95.14	99.32	0.0017
贪心算法	完全	98.85	98.20	0.0017
	1S–2F	82.26	88.08	0.0130
	2S–1F	94.89	98.70	0.0029
莱维游走	完成	82.28	80.43	0.0165
	1S–2F	72.50	76.17	0.0223
	2S–1F	73.85	78.60	0.0202

比移除一名觅食者对性能指标造成更大的负面影响。这表明，在所提出的场景中，探索和信息收集是执行收集任务和保证模型准确度的最关键因素。即使在莱维游走等方法中（当没有已知物品时，该方法会让觅食者进行探索），减少专职侦察者也会显著降低发现物品的比例，最终限制后续收集行动的成效。这些发现表明，尽管侦察者在收集任务执行中扮演间接角色，但他们是智能体协调的基础推动者。RMSE 中也观察到类似效应：移除侦察者时，RMSE 的增幅显著大于移除觅食者时。在深度强化学习和贪心算法中，移除觅食者只会使 RMSE 适度上升，但这主要归因于环境中剩余的活跃元素数量更多，而非探索效率下降，因为发现率（PTA_D）几乎保持不变。相比之下，当没有已知目标时，莱维游走依赖觅食者进行探索，因此移除觅食者后，PTA_D的下降比其他方法更为明显。这些行为模式与第 5.1 节和第 5.2 节基于所研究指标得出的先前推断一致。

边际贡献表6所报告的边际贡献，以标准化形式量化了团队中智能体数量的个体影响，但务必结合表5中的数值来理解这些信息。边际贡献指标

表 6：团队间指标：基于角色的边际贡献（MC）。

算法	MC（1 名侦察者）			边际贡献（1 名觅食者）		
	PTA_D	PTA_C	均方根误差	PTA_D	PTA_C	均方根误差
DRL	0.0589	0.1090	0.8701	0.0009	0.0359	0.4118
贪心算法	0.1090	0.1623	0.8692	0.0015	0.0337	0.4138
莱维游走	0.0743	0.0986	0.2601	0.0447	0.0818	0.1832

表明，就均方根误差而言，深度强化学习和贪心算法中侦察智能体的边际贡献非常高。然而，这种效应很大程度上是由尺度造成的：由于这两种算法在完整系统中的基线误差极低（见表2），因此消融设置中的绝对误差即使只有小幅增加（表5），也会转化为较大的相对贡献。相比之下，莱维游走的参考均方根误差要高得多，因此侦察智能体对均方根误差的边际贡献较低。此外，觅食者较少的配置会带来更多食物积累，无论采用何种协作策略，这都会进一步增加独立维持模型准确性的难度，这一点也体现在边际贡献指标中。

与 PTA_D和PTA_C 相关的边际贡献进一步阐明了两个团队的不对称角色。对于 PTA_D，移除一名侦察者会导致所有算法的表现持续且显著下降，在贪心算法和深度强化学习中尤为明显，这证实了侦察者在信息获取中的核心作用。相比之下，在 PTA_D 中，对于深度强化学习和贪心算法，移除一名觅食者带来的边际贡献几乎可以忽略不计，这表明在这些方法中，觅食者在发现过程中的作用有限。莱维游走则呈现出不同的模式：在 PTA_D中，觅食者的边际贡献不可忽视，这反映出在没有已知目标时，该方法依赖觅食者进行探索行为。

对于收集指标 PTA_C，边际贡献分析显示，在所有算法中，移除一名侦察者造成的性能下降都大于移除一名觅食者。这表明，尽管觅食者是直接负责收集的智能体，但侦察者提供的准确、及时的信息是有效收集的更大制约因素。在深度强化学习和贪心算法中，这一效应尤为显著，反映出它们高度依赖由侦察者驱动的发现过程，以维持高效的收集策略。在莱维游走中，无论移除侦察者还是觅食者，PTA_C中的性能下降程度都非常相近，这是由上述算法的特性所致。

6. 结论

本研究探讨了在部分可观测和动态任务分布条件下运行的异构多智能体系统中，如何表征并量化序贯协作。为此，我们开发、形式化并通过实验

分析了一套全面的协作指标，专门用于序贯觅食环境中探索团队与服务团队之间的交互。所提出的指标分为三大类：主要性能指标、团队间合作指标和团队内指标。结果表明，仅凭传统性能指标不足以全面描述合作行为，因为孤立地使用指标可能导致片面甚至误导性的解读。实验研究证实，联合分析这些指标，能够对合作动态进行显著更丰富、可解释性更强的分析，为未来合作机制研究奠定坚实基础。这些指标还能作为实用工具，指导新算法的设计，因为它们能够清晰揭示常用方法无法呈现的协调结构、敏感性和低效之处。

此外，还必须承认本研究存在若干局限。本研究聚焦于一种特定的双重角色结构（侦察者和觅食者），这可能会限制某些指标直接推广到更复杂的团队层级。不过，值得注意的是，本研究采用的简化假设，例如智能体具有移动能力、资源分布固定且初始时未知并且不会再生、承载能力无限，或环境拓扑，不会限制所提指标的普适性或有效性。这些指标独立于上述设计选择，可有助于设计其他更具合作性的场景。

未来工作可以从多个方向扩展这一框架，例如将其应用于更贴近现实的觅食平台，并纳入明确的能量约束、通信延迟、丢包和有限带宽等因素，从而引入新的限制。此外，还可以将该框架应用于具有两个以上功能角色和团队组织的场景。

参考文献

[1] O. Zedadra、N. Jouandeau、H. Seridi 等，《多智能体觅食：现状与研究挑战》，Complex Adaptive Systems Modeling 5 (1) (2017) 3。数字对象唯一标识符：10.1186/s40294-016-0041-8。[2]E. H. Ostergaard、G. S. Sukhatme、M. J. Matari，《涌现式接力搬运：一种提升多机器人受限空间觅食任务性能的简单机制》，载于：第五届自主智能体国际会议论文集，AGENTS '01，计算机协会，纽约，美国纽约州，2001，页码 29–30。数字对象唯一标识符：10.1145/375735.375825。

[3]F. Bartumeus, M. Luz, G. Viswanathan, J. Cata- lan，《动物搜索策略：随机游走的定量分析》，Ecology 86 (2005) 3078–3087。数字对象唯一标识符：10.1890/04-1806。

[4]N. Bezzo, J. P. Hecker, K. Stolleis, M. E. Moses, R. Fierro，《在协作任务中利用异构机器人系统》(2015)。arXiv:1509.00948。[5]H. Ardiny, A. M. Beigzadeh，《利用仿生群体机器人学增强放射性环境探索：莱维飞行与信息素协同方法的比较分析》，Robotics and Autonomous Systems 181 (2024) 104794。数字对象唯一标识符：10.1016/j.robot.2024.104794。[6] D. Calitoiu, D. Milici，《使用非合作智能体建模：对随机分布物体的破坏性搜索》，载于：2010年夏季计算机仿真会议论文集，SCSC '10，国际计算机仿真学会，圣迭戈，美国加利福尼亚州，2010年，第102–109页。数字对象唯一标识符：10.5772/14264。[7]O. Zedadra, M. Idiri, N. Jouandeau, H. Seridi, G. Fortino，《基于莱维行走的搜索策略：在破坏性觅食中的应用》，载于：2018年国际编程与系统研讨会（ISPS），2018年，第1–5页。数字对象唯一标识符：10.1109/ISPS.2018.8379010。[8] J. George, S. PB, J. B. Sousa，《多无人机搜索与摧毁任务的搜索策略》，Journal of Intelligent & Robotic Systems 61 (1) (2011) 355–367。数字对象唯一标识符：10.1007/s10846-010-9486-8。[9]T. Balch，《多机器人觅食中多样性对性能的影响》，载于：第三届自主智能体年会论文集，AGENTS '99，计算机协会，纽约，美国纽约州，1999年，第92–99页。数字对象唯一标识符：10.1145/301136.301170。

[10] I. O'Flynn, H. Šiljak，模型共享与有限通信下机器人团队中角色的涌现（2025）。arXiv:2505.00540。

[11] G. Muñoz-Gil, A. López-Incera, L. J. Fiderer, H. J. Briegel，《最优觅食策略可以通过学习获得》，《新物理学杂志》26 (1) (2024) 013010。数字对象唯一标识符：10.1088/1367-2630/ad19a8。

[12] S. Shaw, E. Wenzel, A. Walker, G. Sartoretti，Formic：通过隐式通信实现多智能体强化学习觅食，《IEEE机器人与自动化快报》7 (2) (2022) 4877–4884。数字对象唯一标识符：10.1109/LRA.2022.3152688。

[13] O. Zedadra, H. Seridi, N. Jouandeau, G. Fortino，《一种用于多智能体觅食的协作切换算法》，《人工智能工程应用》50 (2016年) 302–319。数字对象唯一标识符：10.1016/j.engappai.2016.01.025。

[14] J. Zhang, P. Zu, K. Liu, M. Zhou，基于群体觅食的双重环境自适应覆盖路径规划方法，《IEEE 控制论汇刊》54 (3) (2024) 1882–1893。数字对象唯一标识符：10.1109/TCYB. 2023.3268844。

[15] B. Atakan, O. B. Akan，间歇连接移动认知无线电自组织网络中受生物觅食启发的通信，《IEEE 车辆技术汇刊》61 (6) (2012) 2651–2658。数字对象唯一标识符:10.1109/TVT.2012.2198928。

[16] I. Karpova，基于视觉地标的受蚂蚁启发的导航算法，Robotics and Autonomous Systems 193 (2025) 105082。数字对象唯一标识符：10.1016/j.robot.2025.105082。

[17]E. Debie、K. Kasmarik、M. Garratt，任务导向的机器人群集：通过专业化角色最大限度地提高效能，Robotics and Autonomous Systems 194 (2025) 105133。数字对象唯一标识符：10.1016/j.robot.2025.105133。

[18] Y. Zhang、G. Tian、J. Lu、M. Zhang、S. Zhang，移动机器人在家庭环境中的高效动态目标搜索：一种基于先验知识的方法，IEEE车辆技术汇刊 68 (10) (2019) 9466–9477。数字对象唯一标识符：10.1109/TVT.2019.2934509。

[19]G. Debout、B. Schatz、M. Elias、D. Mckey，《蚂蚁的多巢制：我们知道什么、自以为知道什么，以及还有哪些工作有待完成》，《林奈学会生物学杂志》90 (2) (2007) 319–348。数字对象唯一标识符:10.1111/j.1095-8312.2007.00728.x。[20] L. Pitonakova、R. Crowder、S. Bullock，《觅食机器人群可塑性的信息流原则》，《群体智能》10 (1) (2016年) 33–63。数字对象唯一标识符:10.1007/s11721-016-0118-1。[21]L. Pitonakova、R. Crowder、S. Bullock，《用于理解机器人群觅食的信息—成本—收益框架》，《群体智能》12 (1) (2018) 71–96。数字对象唯一标识符:10.1007/s11721-017-0148-3。[22] M. Popović、T. Vidal-Calleja、G. Hitz、J. J. Chung、I. Sa、R. Siegwart、J. Nieto，《一种用于无人机地形监测的信息丰富路径规划框架》，《自主机器人》44 (6) (2020) 889–911。数字对象唯一标识符:10.1007/s10514-020-09903-2。[23] A. M. Barrionuevo、S. Y. Luis、D. G. Reina、S. L. T. Marín，《利用深度强化学习，通过异构自主水面车辆优化水体中的塑料废弃物收集》，IEEE《机器人与自动化快报》10 (5) (2025) 4930–4937。数字对象唯一标识符:10.1109/LRA.2025.3555940。[24]S. Y. Luis、D. G. Reina、S. L. T. Marín，《一种通过自主水面车辆解决水资源巡逻问题的深度强化学习方法：以亚帕卡莱湖为例》，IEEE Access 8 (2020) 204076–204093。数字对象唯一标识符:10.1109/ACCESS.2020.3036938。[25]C. Yan、T. Zhang，《多机器人巡逻：一种基于预期空闲时间的分布式算法》，《国际先进机器人系统期刊》13 (6) (2016) 1729881416663666。数字对象唯一标识符:10.1177/1729881416663666。[26] J. L. Gastwirth，《洛伦兹曲线与基尼指数的估计》，《经济学与统计学评论》(1972年) 306–316。数字对象唯一标识符:10.2307/1937992。