去中心化多智能体滤波

Dom Huh加州大学戴维斯分校，加利福尼亚州戴维斯 95616 dhuh@ucdavis.edu

Prasant Mohapatra加州大学戴维斯分校，加利福尼亚州戴维斯 95616
pmohapatra@ucdavis.edu

arXiv:2301.08864v1 [cs.MA] 21 Jan 2023

摘要

本文探讨了在离散状态空间中的多智能体定位应用采用去中心化通信时需要考虑的问题。在这一框架下，我们扩展了贝叶斯滤波器这一离散状态估计的基础概率工具，在其中增加了一个贪心信念共享步骤，以传播信息并改善局部估计的后验分布。我们将此方法应用于一个基于模型的多智能体网格世界，其中每个智能体都维护着所有智能体状态的信念分布。实验结果证实了我们提出的扩展在去中心化协作任务中的实用性。本研究的代码库可在以下仓库中获取[1]。

1 引言

多智能体系统的去中心化催生了模块化的联合行为，因为这类系统要求智能体在局部范围内进行观测、行动和通信。鉴于灵活性、实用性和生存能力的提升带来的吸引力， [10], 已有大量文献综述了面向多样且复杂的去中心化多智能体系统应用的技术 [3, 2, 1]。近年来，这类研究主要聚焦于将这些方法用于控制任务 [4, 8, 5, 15, 17]。本研究则主要致力于为此类控制应用构建丰富且定义明确的特征空间，同时遵守去中心化约束。

为实现去中心化多智能体系统（Dec-MAS）的效用，我们探讨了去中心化通信的特性所带来的一些问题。从本质上说，去中心化网络严格限制通信，使其仅以点对点的方式进行，因此只能建立基于物理条件（即感知范围）的连接。因此，智能体通常依赖某种消息传递方式，而非使用集中式通信设施在智能体之间传输信息。先前的研究 [11] 指出，除非采用全连接拓扑（这通常难以实现），否则这种限制会导致知识库之间出现不一致。这种不一致会增加实现适当协调的难度，并降低可预测性。我们提出的贝叶斯滤波算法扩展旨在缓解这些问题，同时不违反局部性约束。

本文研究如何将去中心化概念扩展到一种基础的非参数状态估计算法——贝叶斯滤波器 [12], ，并应用于部分可观测跟踪应用。本研究借鉴了许多先前工作的思路 [9, 16, 6], ，但我们着重研究协作跟踪任务，以评估所提出的贪心信念共享方法在概率状态估计中的效果。

[1] https://github.com/domhuh/ma-gym

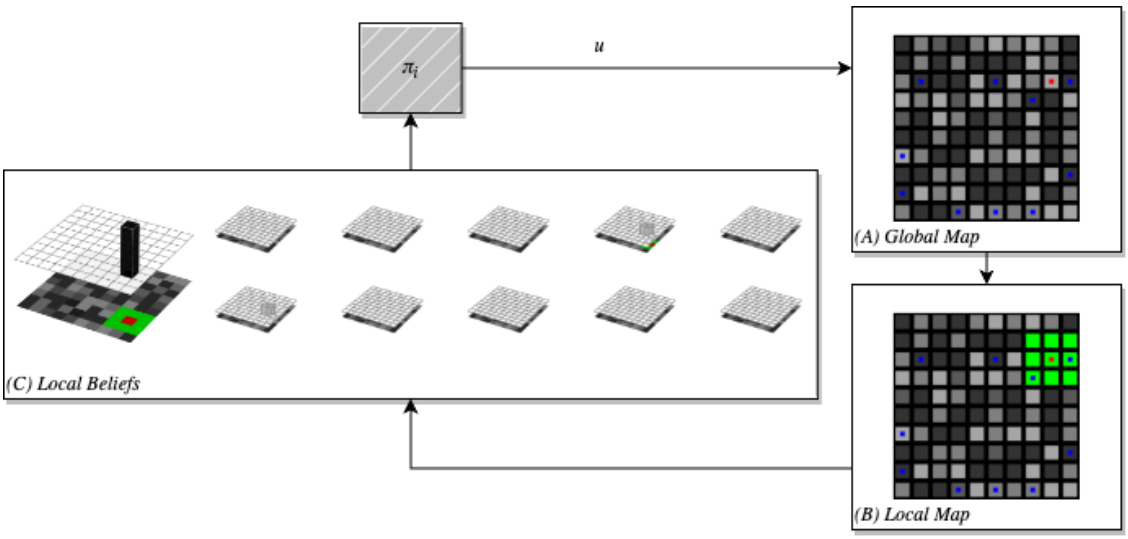


图 1：去中心化多智能体系统过滤概述：(A) 环境的动态和潜在状态在全局地图层面进行处理。(B) 对全局信息的部分内容进行屏蔽，以保留去中心化属性并构建局部信息。(C) 每个智能体利用 (B) 提供的局部信息更新并处理其内部信念，估计每个智能体的状态，从而做出明智的决策。

2 系统概述

图 1 展示了我们用于定位的去中心化多智能体系统滤波系统示意图。由于我们选择要处理的任务会对各系统的构建方式产生重大影响，因此本节将介绍该任务。在我们的应用中，每个智能体 $a_i \in A$（其中 A 为所有智能体的集合）利用来自局部地图层级 L_i的信息，维护自身的局部信念 B_i。重申一下，该系统的目标并非设计一种新颖的控制策略 π，而是开发一种动态特征空间，供控制策略用于各种去中心化多智能体系统控制任务。

2.1 任务表述。

我们专注于去中心化多智能体系统，并遵循去中心化部分可观测马尔可夫决策过程框架。具体而言，我们在一个二维 $H \times W$ 网格世界中开展研究，其中有多个智能体。每个智能体只能访问自身的局部观测，该观测取自其所在的网格单元。此外，智能体还可以访问其感知范围内协作智能体的观测和控制量。如果感知范围内存在外部实体，则会在该智能体的观测中附加一个指示位，但不提供其他信息，因为我们假设这些实体不具备协作性。还可以附加更详细的感知观测，但本研究并未采用。

每个智能体都知道环境模型（即状态转移动力学和发射概率），并且该模型被定义为平稳、确定性且满射（即具有别名化单元特征）。状态转移动力学 T 遵循王图，该图基于智能体的控制空间 U 构建，其中包含切比雪夫移动。感知范围由同一距离度量的阈值限定。系统采用排序规则，以避免多个智能体试图占据同一单元时可能出现的平局。发射概率定义了从局部观测到潜在状态空间中一个候选子集的映射。我们将这一去中心化多智能体系统的形式化框架扩展至以下任务：聚集和捕食者-猎物。这两项任务的有限时域均为离散的，且上限为 $t_{done} = 30$。

聚集。 该任务的目标是让尽可能多的智能体彼此处于对方的感知范围内，因此奖励函数定义为每个智能体感知范围内智能体总数之和

。对于此任务，所有智能体都遵循相同的控制策略 $\pi_i : B_i \to U$，该策略将局部信念空间映射到控制空间，并使每个智能体向其他智能体靠近。因此，该策略利用各智能体自身的信念计算所有智能体的可能位置，并通过随机采样打破平局。所选动作会使智能体朝所有智能体平均可能位置的方向移动。

捕食者-猎物。该任务的目标是让一组称为捕食者的智能体找到并捕获另一组称为猎物的智能体。奖励函数为感知范围内有猎物的捕食者总数。对于捕食者，我们会让智能体朝着其认为猎物所在的位置移动，同时彼此保持距离。猎物采用与聚集任务相同的策略，但猎物不是朝捕食者的平均位置移动，而是远离该位置。在我们的实验中，猎物使用全局信息。

对于这两项任务，我们建立一种平稳的联合行为，不再进一步考虑如何设计能够提升规划、稳定性或适应性的方法。

2.2 分布式贝叶斯滤波器

每个智能体都会在本地为每个智能体维护状态估计。为计算智能体的信念，我们使用标准的贝叶斯滤波算法，根据本地控制量和观测进行更新。为纳入附近智能体发送的信息，我们利用邻近智能体的信念，贪心地更新该智能体的信念。因此，我们执行以下递归更新：控制、观测和信念共享。对于外部实体（即猎物），我们会加入其相应的状态估计，并遵循相同的更新规则，只做细微调整。所有信念先验均初始化为在潜在状态空间上均匀分布。

控制更新。 在这一步中，我们使用智能体的控制量，结合状态转移动力学和先验信念，更新局部后验信念。对于离散情形，我们可以将状态转移动力学$p(x'|x,u)$表示为线性张量算子 $\mathcal{T}_t$ ，以纳入时间步 t所采取的控制量。

$$\mathcal{T}_t = \begin{bmatrix} p(X|X,u_{0,t}) \\ \cdots \\ p(X|X,u_{n,t}) \end{bmatrix} \tag{1}$$

其中， n 是智能体的数量， X 表示潜在状态空间，因此， $p(X|X,u_{i,t})$表示智能体 i 在采取动作 $u_{i,t}$时， X 上的转移矩阵。据此，我们可以按以下方式表述更新过程。

$$\bar{bel}_{t+1} = \mathcal{T}_t bel_t \tag{2}$$

对于彼此不在感知范围内的智能体，我们会在状态转移动力学中屏蔽该智能体被观测到的控制量，为屏蔽动作 u_m分配相应处理，并将其转移概率设为所有可能控制量 $U' = U \backslash u_m$的期望。

$$p(x'|x,u_m) = \frac{1}{|U'|} \sum_{u \in U'} p(x'|x,u) \tag{3}$$

观测更新。 然后，我们处理每个局部智能体的观测 z_i ，根据发射概率 $p(z_i|x)$对每个信念的后验分布进行筛选。

$$bel_{i,t+1} = \eta(p(z_i|x) \odot \bar{bel}_{i,t+1}) \tag{4}$$

其中， η 是一个归一化常数，使滤波后的后验分布积分为 1。同样，对于感知范围之外的智能体，我们在发射概率中分配一个掩蔽观测，并将其定义为在整个网格世界中均匀分布。

信念共享更新。 后验分布经过滤波后，智能体通过固定轮次的消息传递协作改进其局部估计。通过这些轮次，我们引入了传递性和更广的隐式通信范围。具体而言，每个智能体 i 与邻近智能体集合 N_i 交流各自的信念，并基于熵度量贪心地更新其后验分布。

$$bel_{(i,j),t} = \underset{j \in A}{\operatorname{argmin}} \mathcal{H}(bel_{(n,j),t}), n \in N_i \tag{5}$$

我们断言，这种贪心更新只会单调提升估计准确度，不会有陷入次优收敛的风险，因为所有其他针对后验分布的同质更新，都是基于真实信息的平稳同分布并由其推导而来。

证明： 假设 T 是由狄拉克 δ 函数 $\delta(z)$ 表示的遍历状态分布，其微分熵 T 趋近于 $-\infty$。令 X, Y 为关于 T同分布的两个独立分布。因此，若关系熵定义为 $\mathcal{H}(X) \leq \mathcal{H}(Y)$，则我们断言，相对于 T的相对交叉熵为 $H(T, X) \leq H(T, Y)$；等价地，在仅允许进行我们所定义的更新这一假设下，我们可以表述为 $P_X(z) \geq P_Y(z)$。换言之， X 包含于 Y中，而 T 包含于 X中。

令 $D(P, Q) \propto \mathcal{H}(Q, P)$ 为与交叉熵成正比的散度度量。根据此定义，若 $\mathcal{H}(X) \leq \mathcal{H}(Y)$，则在没有其他冲突项主导该度量的前提下，我们可以断言 $D(T, X) \leq D(T, Y)$。满足此性质的散度度量示例包括反向 KL 散度和 Wasserstein 距离 [14]；这种性质可以由定义得出（参见公式 6），也可以通过下界得到。

$$D_{KL}(T|X) - D_{KL}(T|X) = H(T, X) - H(T, Y) \leq 0 \tag{6}$$

或者，我们可以直观地推断，熵较低的估计反映了非屏蔽更新中更近期的信息，因此具有更大的信息增益。这会使其与目标分布的重叠度更高。

外部信念。 对于外部实体的状态估计，我们遵循上述相同的流程，但确保非合作智能体之间不会发生信息泄漏。因此，我们对控制和观测更新都采用相同的屏蔽方式，并根据当前任务对两者进行相应处理。

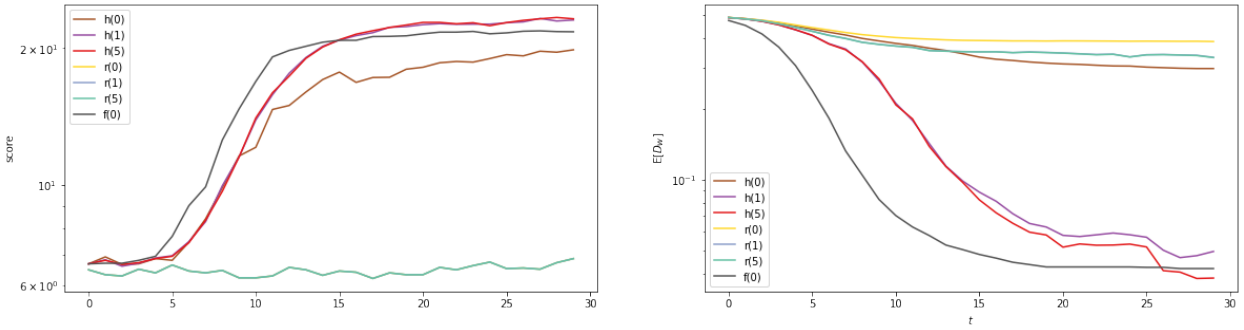
此外，我们通过使用局部层级提供的信念和信息，确保不会发生全局信息泄漏。因此，对于捕食者-猎物任务中的观测更新，我们会判断指示位 I 是否已置位，以确定猎物是否在感知范围内。我们使用智能体估计来尝试定位自身，并使用张量算子 $\mathcal{T}_s$，根据感知范围传播猎物的估计。该算子的定义与状态转移动力学类似，但依据的是感知范围。随后，我们处理相应外部智能体的信念更新 e。

$$bel_{(i,e),t} = \eta(\mathcal{T}_s bel_{(i,i),t} \odot bel_{(i,e),t}) \tag{7}$$

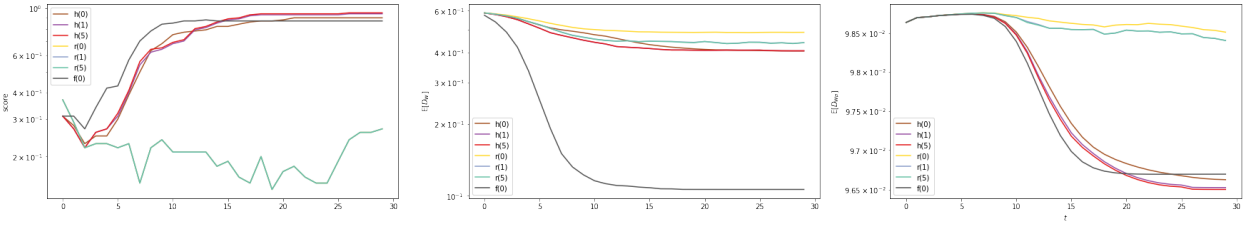
最后，我们指出，信念分布的不确定性与决策过程密切相关。我们将论述范围限定在纯状态估计内，同时承认，控制在全面优化其性能（即平衡探索）方面确实发挥着重要作用。

3 实验

我们的实验揭示了我们的改动是否充分适应了去中心化通信网络的局限性，以及我们在滤波系统中引入的不同设计考量对两项任务的影响。



(a) 聚集任务。



(b) 捕食者-猎物任务。

图 2：去中心化多智能体系统滤波系统的定量评估：图中显示累积奖励（得分）的平均值，该值取自 100 次仿真运行；同时还显示总期望 1–Wasserstein 距离 $\mathrm{E}[D_w]$，以及图 2b 中猎物与目标分布之间的期望 1–Wasserstein 距离 $\mathrm{E}[D_{wp}]$。对去中心化系统 $h(\cdot)$、预言机系统 $f(\cdot)$ 和随机系统 $r(\cdot)$ 进行了仿真，并针对随机策略和启发式策略测试了 0、1、5 轮消息传递。

3.1 设置

为了评估我们的系统，我们使用第 2 节中针对两项任务所述的相同配置，以及各自的控制策略。我们将网格世界的大小设为 $H = W = 10$，智能体数量设为 5，每个智能体的感知范围为 1。作为评估指标，我们跟踪累积奖励，它可作为智能体之间信念信息一致性的替代衡量指标；同时跟踪局部信念分布与真实分布（用狄拉克 δ 函数表示）之间的期望 1–Wasserstein 距离，后者可直接衡量真实散度。所有实验均在基于 [7, 13] 的环境中使用 100 个不同的随机种子运行。结果见图 2a 和图 2b。

3.2 使用随机控制策略。

我们首先测量控制策略选择的影响，并使用去中心化滤波系统比较随机策略和启发式策略。由于随机策略并非依据其知识库采取行动，实际表现并未揭示太多信息，只说明这些任务无法通过随机控制量解决。真实分布之间的总散度确实显示，启发式策略与随机策略在估计准确度方面存在显著差异。这一结果再次证实，控制策略对状态估计的真实散度具有重要影响，因为每个潜在状态都包含一定的信息增益，而控制量在很大程度上决定了能够观测到哪些状态。

3.3 与预言机系统进行比较。

预言机系统是一个全连接的通信网络，其中每个智能体的感知范围都覆盖整张地图。我们定义预言机系统遵循启发式策略，并使用

经典贝叶斯滤波，而不进行任何消息传递和信念共享。对于这两项任务，奖励函数仍要求智能体处于与去中心化系统相同的感知范围内。结果表明，预言机系统在两项任务初期确实优于我们的去中心化系统，这正如预期，因为每个智能体都能接收到信息增益最高的更新。在衡量捕食者-猎物实验中的总真实散度时，这一点尤为明显。然而，有趣的是，我们的系统最终表现超过了预言机系统，收敛至更高的得分和更低的真实散度。因此，我们证明，即使去中心化通信受到限制，只要进行适当的调整，局部信念分布仍可收敛至接近真实分布。

关于我们的系统优于预言机系统的原因，我们认为，所采用的控制策略未必能平衡这些任务取得良好表现所需的探索与利用。有限的信息增益带来了更大的不确定性，从而提供了所需的探索，这可能使真实散度更低。

3.4 增加消息传递轮次的影响。

观察消息传递轮次更多的系统时，我们发现，其改进主要体现在真实散度上，而性能表现则呈现相似的模式。即使在随机策略下，采用消息传递的系统真实散度也更低。结果确实表明，增加消息传递轮次存在相当显著的边际收益递减；从 ≥ 1 个消息传递轮次开始，改进幅度急剧下降。或许在规模更大、更加密集的环境中，更广泛的传递通信会更加重要。

4 结论

在这项工作中，我们提出了对贝叶斯滤波算法的扩展，使其更适用于采用去中心化通信网络的多智能体系统。我们所做的调整包括一个基于贪心熵的信念共享步骤，以及通过定义掩码更新来修改环境模型。结果表明，我们改进后的贝叶斯滤波算法应用于去中心化多智能体系统，能够充分适应去中心化通信的有限感知范围。未来的工作可以着重于利用本研究开发的特征空间来设计控制策略。

参考文献

[1] Mohammad Afshari 和 Aditya Mahajan。“用于最小化团队均方误差的多智能体估计与滤波”。载于：*IEEE* 信号处理汇刊69 (2021年)，第 5206–5221 页。[2] Michael Beard、Ba Tuong Vo 和 Ba-Ngu Vo。“一种大规模多目标跟踪解决方案”。载于：*IEEE* 信号处理汇刊68 (2020年)，第 2754–2769 页。[3] V. Berge-Cherfaoui 和 B. Vachon。“一种多传感器融合的多智能体方法”。载于：第五届先进机器人国际会议“非结构化环境中的机器人”。1991年，第 1264–1269 页，第 2 卷。 DOI: 10.1109/ICAR.1991.240381。[4] Daniel S.Bernstein、Eric A. Hansen 和 Shlomo Zilberstein。“去中心化 POMDP 的有界策略迭代”。载于：第十九届国际人工智能联合会议论文集。IJCAI’05。苏格兰爱丁堡：Morgan Kaufmann 出版社，2005年，第 1287–1292 页。[5] Per Bostr¨om-Rost、Daniel Axehill 和 Gustaf Hendeby。“使用泊松多伯努利混合滤波器进行搜索与跟踪的传感器管理”。载于：*IEEE* 航空航天与电子系统汇刊57.5 (2021年)，第 2771–2783 页。 DOI: 10.1109/TAES.2021.3061802。

[6] Joseph R Bourne 等。“用于目标估计与定位的去中心化多智能体信息论控制：寻找燃气泄漏”。载于：国际机器人研究期刊39.13 (2020年)，第 1525–1548 页。 DOI: 10. 1177/ 0278364920957090。电子版：https: // doi. org/ 10.1177 / 0278364920957090。 URL: https://doi.org/10.1177/0278364920957090。[7] Greg Brockman 等。*OpenAI Gym*。2016年。电子版：arXiv:1606.01540。[8] Philip M. Dames、Pratap Tokekar 和 Vijay R. Kumar。“使用一组移动机器人检测、定位和跟踪数量未知的移动目标”。载于：国际机器人研究期刊36 (2017年)，第 1540–1553 页。[9] Prashant Doshi 和 Piotr J. Gmytrasiewicz。“使用粒子滤波器近似多智能体环境中的状态估计”。载于：第四届自主智能体与多智能体系统国际联合会议论文集。AAMAS ’05。荷兰：计算机协会，2005年，第 320–327 页。 ISBN: 1595930930。 DOI: 10.1145/1082473.1082522。 URL: https://doi.org/10.1145/1082473.1082522。[10] Hugh F. Durrant-Whyte。“分布式数据融合入门指南”。载于：2006年。[11] S. Grime 和 H.F. Durrant-Whyte。“分布式传感器网络中的数据融合”。载于：*Control En-gineering Practice* 2.5 (1994年)，第 849–863 页。 ISSN: 0967-0661。 DOI: https : / / doi . org / 10 . 1016 / 0967 - 0661(94) 90349 - 2。 URL: https : / / www . sciencedirect . com / science / article / pii / 0967066194903492。[12] Leslie Pack Kaelbling、Michael L. Littman 和 Anthony R. Cassandra。“在部分可观测随机领域中的规划与行动”。载于：《人工智能》101.1 (1998年)，第 99–134 页。 ISSN: 0004-3702。 DOI: https : / / doi . org / 10 . 1016 / S0004 - 3702(98) 00023 - X。 URL: https : / / www.sciencedirect.com/science/article/pii/S000437029800023X。[13] Anurag Koul。*ma-gym*：基于 *OpenAI Gym* 的多智能体环境集合。https : / / github.com/koulanurag/ma-gym。2019年。[14] Muka Nakazato 和 Sousuke Ito。“基于 Wasserstein 距离的随机热力学中熵产生的几何方面”。载于：(2021年2月)。[15] Zengyi Qin 等。“使用去中心化神经屏障证书学习安全多智能体控制”。载于：*CoRR* abs/2101.05436 (2021年)。arXiv: 2101.05436。 URL: https://arxiv.org/abs/2101.05436。[16] Luke Zettlemoyer、Brian Milch 和 Leslie Kaelbling。“使用无限嵌套信念的多智能体滤波”。载于：神经信息处理系统进展。D. Koller 等编。第 21 卷。Curran Associates, Inc.，2008年。 URL: https : / / proceedings . neurips . cc / paper / 2008 / file / 6c3cf77d52820cd0fe646d38bc2145ca-Paper.pdf。[17] Qingrui Zhang、Hao Dong 和 Wei Pan。“用于去中心化多智能体控制的基于李雅普诺夫的强化学习”。载于：分布式人工智能国际会议。施普林格。2020年，第 55–68 页。