

多智能体深度确定性策略梯度算法研究与改进*

劳天成, 刘义, 范文慧

(清华大学 自动化系, 北京 100084)

摘要: 针对多智能体深度确定性策略梯度算法(MADDPG)在某些场景下,尤其是在部分可观察环境与稀疏奖励条件下,不一定能学习到最优策略的问题,采用观察叠加法和在深度网络中加入长短期记忆网络(Long Short-Term Memory, LSTM)层的方法对MADDPG算法进行了改进,通过含遮蔽区的捕食者-猎物场景验证了改进的算法在智能体决策上的有效性;引入后验经验回放(Hindsight Experience Replay, HER)方法对MADDPG算法进行了改进,通过合作通讯场景和合作导航场景的对比实验验证了改进的算法能够使智能体获得的高价值经验大大增加,可以提高MADDPG算法收敛速度,有助于智能体学习到最优策略。

关键词: 多智能体; 深度强化学习; 部分可观察; 稀疏奖励

DOI: 10.13568/j.cnki.651094.651316.2023.05.08.0001

中图分类号: TP301.6 **文献标识码:** A **文章编号:** 2096-7675(2023)06-0717-07

引文格式: 劳天成, 刘义, 范文慧. 多智能体深度确定性策略梯度算法研究与改进[J]. 新疆大学学报(自然科学版)(中英文), 2023, 40(6): 717-723.

英文引文格式: LAO Tiancheng, LIU Yi, FAN Wenhui. Exploration and improvement of multi-agent deep deterministic policy gradient algorithm[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2023, 40(6): 717-723.

Exploration and Improvement of Multi-Agent Deep Deterministic Policy Gradient Algorithm

LAO Tiancheng, LIU Yi, FAN Wenhui

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: In order to solve the problem that the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm may not be able to achieve the optimal policy in some scenarios, especially in partially observable environments and sparse reward conditions, this paper adopts the observation stacking method and introduces LSTM layers in deep networks to improve the algorithm and the effectiveness in decision-making of the improved algorithm is verified on the predator-prey scenario with sheltered area. Besides, HER method is introduced to improve the algorithm. The experiment on the cooperative communication scenario and the cooperative navigation scenario validates that the improved algorithm can actually enable the agents to obtain more high-value experience, which speeds up the convergence and makes it easier to learn the optimal policy.

Key words: multi-agent; deep reinforcement learning; partial observability; sparse reward

0 引言

多智能体深度强化学习是人工智能领域机器学习方向的研究热点,广泛应用于工业智能、自主智能、群体智能等实际领域,具有极其重要的研究意义和实际应用价值. 在强化学习的应用中,可以与环境交互,并在交互过程中学习策略的个体被称为智能体. 它以自身观察(在完全可观察条件下,等同于环境状态)作为输入,基于策略做出动作,然后得到环境的反馈(称之为奖励或回报)以及新的观察. 强化学习算法可以分为三类,分别是策略梯度法、基于值函数的方法和基于actor-critic结构的方法^[1].

策略梯度法以累积回报为目标函数,通过梯度上升法来逼近最优策略. 其中最经典的方法是REINFORCE算法^[2]. 此外,为了避免在梯度上升过程中,策略参数变化过大导致的振荡、难以收敛的问题,TRPO算法^[3]和P-

* 收稿日期: 2023-05-08

基金项目: 国家重点研发计划“基于人工智能与区块链技术的生态环境新型治理体系研究与示范应用”(2021YFC1809000).

作者简介: 劳天成(1998-),男,硕士生,从事多智能体强化学习的研究, E-mail: ltc2207@163.com.

PO算法^[4]相继被提出,它们都对每次迭代时策略参数的变化作了限制.基于值函数的方法则是估计在任意观察下各个动作的价值,通过选择价值最大的动作就可以导出策略.DQN算法^[5]在Q-learning方法^[6]的基础上通过引入深度网络解决了状态空间无穷的问题.而当动作空间无限时,DDPG方法^[7]作为actor-critic结构算法的代表,通过引入用于选择动作的新网络解决了动作选取问题,其中actor负责动作的选取,critic负责动作价值的评估.

而在多智能体的设定下,情况会更加复杂,依据奖励函数设置不同,多智能体场景可以分为完全合作型、完全竞争型和混合型^[8].本文将要讨论的MADDPG算法^[9]全称为multi-agent deep deterministic policy gradient,中文名为多智能体深度确定性策略梯度算法,是美国OpenAI公司在2017年提出的,是通用于各种多智能体场景的算法,是DDPG算法在多智能体情景下的拓展.

MADDPG算法在多种场景下都有不错的表现,但在部分可观察的环境下,由于智能体获得信息的不足,往往只能学习到次优的策略.本文通过对历史信息利用来扩展智能体所获得的信息,从而令智能体可以习得更优的策略.面对稀疏奖励场景时,MADDPG算法下智能体往往由于成功经验不足,难以探索到最优策略,本文引入HER方法^[10],相较原算法在合作通讯场景和合作导航场景都取得了明显的提升.已有学者将HER方法与MADDPG算法结合^[11]并验证了有效性,本文将更进一步地说明HER方法如何应用于具体场景,并对它的价值作了更多分析.

1 MADDPG算法

MADDPG算法采用了actor-critic结构,每个智能体都有一个独立的中心化的critic和一个分布式的actor.此外,MADDPG不假设环境是完全可观察的,记智能体 i 的观察为 o_i ,动作为 a_i .以智能体 i 为例,为了有利于收敛,其actor和critic各自有两套网络:actor当前网络,记为 $\mu_i(o_i)$,参数为 θ_i ;actor目标网络,记为 $\mu'_i(o_i)$,参数为 θ'_i ;critic当前网络,记为 $Q_i^\mu(\mathbf{x}, a_1, \dots, a_N)$;critic目标网络,记为 $Q_i^{\mu'}(\mathbf{x}, a_1, \dots, a_N)$;其中 \mathbf{x} 代表状态信息,如果获得全局的状态信息比较困难,可令 $\mathbf{x} = (o_1, \dots, o_N)$.算法框架如图1所示.

MADDPG算法的策略网络(actor)的优化目标是最大化 $Q_i^\mu(\mathbf{x}, a_1, \dots, a_N)|_{a_i = \mu_i(o_i)}$,其梯度为:

$$\nabla_{\theta_i} J(\mu_i) = E_{\mathbf{x}, a \sim D} [\nabla_{\theta_i} \mu_i(o_i) \nabla_{a_i} Q_i^\mu(\mathbf{x}, a_1, \dots, a_N)|_{a_i = \mu_i(o_i)}] \quad (1)$$

其中 D 表示经验回放池,经验元组记为 $(\mathbf{x}, \mathbf{x}', a_1, \dots, a_N, r_1, \dots, r_N)$.注意到 \mathbf{x}, a 是一并从经验回放池中采样的,在代入 Q 函数时,只有 a_i 是重新计算的,这样足以使得梯度流动到智能体 i 的策略网络;当然,如果能获知当前所有智能体最新的策略网络,也可以尝试将所有智能体动作都重新计算,但这也增大训练的开销.

而critic网络的更新与DQN类似,只是其中 a' 的选取由actor目标网络完成,其优化目标是减小下述的损失函数值:

$$L = E_{\mathbf{x}, a, r, \mathbf{x}'} [(Q_i^\mu(\mathbf{x}, a_1, \dots, a_N) - y)^2], y = r_i + \gamma Q_i^{\mu'}(\mathbf{x}', a'_1, \dots, a'_N)|_{a'_k = \mu'_k(o_k)} \quad (2)$$

目标网络的更新采用软更新方法,即令 $\theta'_{\text{new}} = \tau\theta + (1-\tau)\theta'$,其中 τ 为超参数.

实际中深度网络使用的是多层感知机(MLP).在多智能体强化学习中,所有智能体的策略都在同时迭代,如果每个智能体单独训练,在使用经验回放时, (o, a, r, o') 与 (o, a, r', o'') 可能会在经验回放池并存,即面对相同的观察做出了相同的动作却有不同奖励和新的观察,这种不稳定性对智能体的训练是有害的;而且critic只利用本地观察(即智能体自身的观察)进行训练也容易因为没有考虑其它智能体的动作,习得的策略不够好.采用中心化的critic,不仅解决了经验回放的冲突问题,也解决了训练时critic网络无法直接关注到其它智能体的问题;而在训练完毕之后,则因actor网络只需要本地观察,仍便于实际应用.

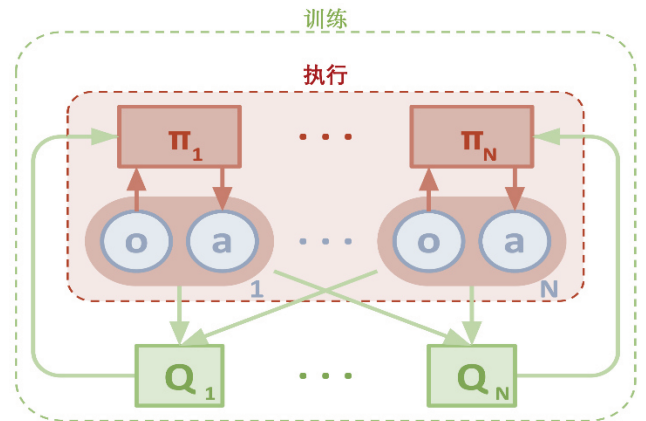


图1 MADDPG算法框架概览^[9]

2 算法改进

2.1 观察叠加法和引入LSTM网络法

在实际中,智能体往往很难获知环境的全部状态,而只能拥有对环境的观察,此时问题可以建模为部分可观察马尔科夫决策过程(POMDP),用六元组 (S, A, R, P, Ω, O) 表示,其中: S 为观察空间, A 为动作空间, R 为回报, P 为状态转移方程, Ω 为观察空间, $o \in \Omega$,且 $o \sim O(s)$.图2是对POMDP的示意.

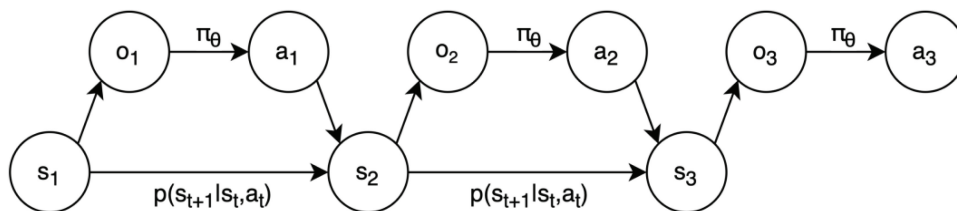


图2 部分可观察马尔科夫决策过程示意

在部分可观察的情况下,智能体需要在不知道真实状态的情况下进行决策;这也容易使得算法的收敛更为困难.

本文的思路是通过挖掘历史信息来增强智能体对当前状态的了解,最简单的方法是观察叠加,即 t 时刻智能体的观察输入由 (o_{t-k}, \dots, o_t) 叠加而成,记单次观察的维度为 m ,则拼接后的观察维度为 $m(k+1)$.事实上在利用DQN算法玩Atari游戏时同样使用了帧叠加的技巧^[5].为了更好地编码历史信息,本文也使用了LSTM网络.这个做法借鉴了DRQN算法^[12],它在DQN的基础上引入了LSTM层,在对每帧图片做随机遮挡的Pong游戏上取得了不错的表现.LSTM是循环神经网络(RNN)的一种,能够利用序列数据中存在的前后依赖性.LSTM通过精心设计的门机制较好地解决了长依赖的问题^[13].

不同于DRQN中将一层全连接替换为LSTM层的做法,本文在现有网络中增加了一层LSTM.此外,在DRQN中,只有一个深度网络,即Q网络;而MADDPG算法中包括critic网络和actor网络,均为多层感知机,本文的做法是将actor网络和critic网络做相同的架构改动.具体而言,原先的深度网络结构为:输入-全连接-ReLU^[14]-全连接-ReLU-全连接-输出.引入LSTM层之后的结构为:输入-全连接-ReLU-LSTM-全连接-ReLU-全连接-输出.后续将做如此修改的MADDPG算法称为MADRDPG.训练时使用随机更新的方法:每批次训练,从经验回放池随机采样批次大小的固定时间步(超参数 T)的经验.

2.2 引入HER方法

本文通过引入后验经验回放(HER)方法,增强了MADDPG算法在稀疏奖励场景下的表现.稀疏奖励场景是指智能体在其中长时间只能获得0的回报.解决这个问题有两个典型的思路,分别是依赖模型的方法和依赖数据的方法.前者主要指分层强化学习方法,通过分层弱化了对于持续奖励的依赖;而依赖数据的方法常常能够嵌入多种模型,适合作为对已有模型的加强,这里提到的HER方法就是其中的典型代表.

鉴于在稀疏奖励的场景下,智能体很难获得有价值的经验.HER的想法是,是否有可能将失败的经验转为成功的经验,从而加速智能体学习呢?HER的提出借鉴了UVFA^[15]的做法(两者强调的内容有所差异,UVFA的提出是针对多目标情形,HER强调对稀疏奖励问题的有效性),将智能体的观察拆解为目标和状态,在特定状态下,目标达成.

具体的,用 o 表示智能体的观察, s 表示智能体的状态, g 表示智能体的目标,则 o 可表示为 $g||s$, $||$ 表示拼接.以导航问题为例, g 为每回合给定的目标地点,自身坐标则包含在状态中.回合结束后,将该回合放入经验回放池.HER所做的关键是,重新给定一个目标 g' ,并假设该回合的目标为 g' ,重新计算该回合中智能体的观察($o' = g' || s$)与奖励,得到新的一个回合的经验,并放入经验回放池.当重新给定多个目标 g'_i 时,则新增多个回合的经验.这么做的前提是使用异策略方法,此时原理上对生成经验的策略没有限制.特别的,若将回合末智能体状态对应的目标作为重新给定的目标,比如在导航问题中以智能体最后到达的位置作为给定的目标点,那么智能体就获得了高价值的成功经验.

当智能体处于某一状态时,目标达成,这样的对应关系记为 $g_s = f(s)$.表现不错的新目标选取方法^[10]包括:1)以回合末对应的智能体状态对应的目标作为 g' ;2)以回合将来的 k 个时刻的智能体状态对应的目标作

为 g' , 此时 g' 有 k 个; 3) 从当前回合中随机抽取 k 个时刻的智能体状态, 取其对应目标. 第一种方法最为直观, 也是本文所采用的.

本文将HER方法的应用扩展到多智能体情形, 其扩展的关键在于不同场景下如何确定状态与目标的对应关系; 后续的实验中也对HER方法的价值做了更深入的分析.

3 实验

3.1 实验场景

带遮蔽区的捕食者-猎物场景^[9]: 包括3个速度较慢的捕食者(红方)和1个速度较快的猎物(绿方)以及2个障碍物. 此外本实验引入了两个遮蔽区, 用于构建部分可观察属性. 图3为各场景示意图, 对应本场景, 其中较小的圆为智能体, 较大的绿色圆为遮蔽区, 灰色圆为障碍物. 奖励设置上, 红方采用合作性奖励. 红绿双方碰撞, 则红方正奖励, 绿方负奖励, 红方有较小的引导性的基于与绿方的最短距离的奖励; 且绿方会因为离开场景边界受到惩罚.

合作通讯场景^[9]: 包括3个不同颜色的地标, 一位说者, 一位听者. 当听者到达目标地标, 任务完成. 说者与听者是合作关系, 说者的观察为需要抵达的地标的颜色, 听者只能知道3个地标的相对位置. 说者不能移动, 但可以对听者发出通讯信息, 听者和说者的奖励相同. 图3(b)是任务完成时的示意图, 小圆为地标, 灰色为说者, 浅蓝色为听者. 这里设置为稀疏奖励: 听者到达目标地标指定范围内得到正奖励, 否则奖励为0.

合作导航场景^[9]: 包括3个合作者和3个地标, 合作者们需在避免互相碰撞的情况下到达这3个地标, 我们希望地标均被占据(与智能体相交即判为被占据). 图3(c)为奖励最大化时的场景, 小圆为地标, 大圆为智能体. 奖励设置上, 若有 n 个地标被占据, 则给予奖励 n . 同时, 若当前智能体与其它合作者碰撞, 则会受到惩罚.

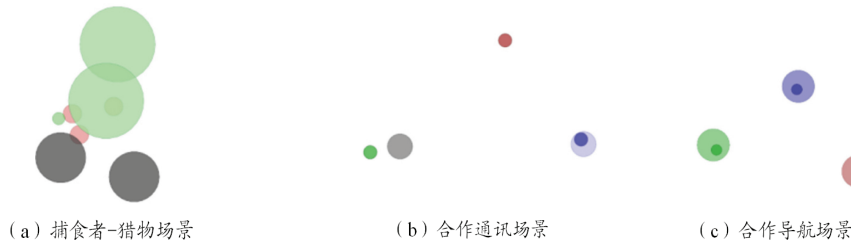


图 3 实验场景示意图

3.2 实验结果

在捕食者-猎物场景对观察叠加法和引入LSTM层方法进行了效果验证. 场景中, 遮蔽区内外的智能体不能互相观察到, 这对于捕食者构成了更大的挑战, 在追逐猎物时可能会因为猎物跑入或跑出遮蔽区而丧失追踪目标. 实验中也发现, 加入遮蔽区后, 捕食者的回合奖励大幅下降. 实验中, 仅对捕食者应用这两个方法, 并比较了应用前后捕食者奖励的变化.

观察叠加法的效果如图4所示, 横轴为训练时的回合数, 纵轴是每1 000回合智能体的平均回合奖励(下同). t 时刻智能体的观察由 $k+1$ 帧组成, 图中 $k=0$ 即表示没有做观察叠加. 由图4左图可知, 采用观察叠加的方法, 确实可以使得捕食者奖励增加, 对缓解部分可观察问题是有效的. 从捕食者动作观察来看, 在追逐过程中, 当猎物进入遮蔽区而丢失视野时, 捕食者仍能大致向着猎物的方向前进. 图4右图对不同 k 的情形作了对比. k 从1增加到4的过程中, 捕食者学习效果先升后降, $k=3$ 时最佳. 后续的下落可能是因为观察输入维度过高, critic和actor都需要从更高维度的输入中获取信息(相对而言冗余信息也会更多), 增大了学习难度.

MADRDGP的效果如图5所示, T 为选定的时间步. 由图5左图可知, 在当前场景, 选择时间步为8与观察叠加法中 $k=3$ 的结果是较为接近的. 相比之下, MADRDGP方法训练会更为耗时; 优势在于, 训练完成后, 测试时MADRDGP算法下智能体每个时间步只需要当前帧的观察作为输入(历史信息存储在LSTM层的细胞状态和隐层状态中). 图5右图展示了选取不同时间步对奖励值的影响, 就训练60 000回合的数据曲线而言, 末尾时间步为8和12的曲线均值接近, 高于时间步为4的情形; 考虑到训练时增加时间步会增加训练时间, 若在训练效果与训练时间中取平衡, 这里可取时间步为8.

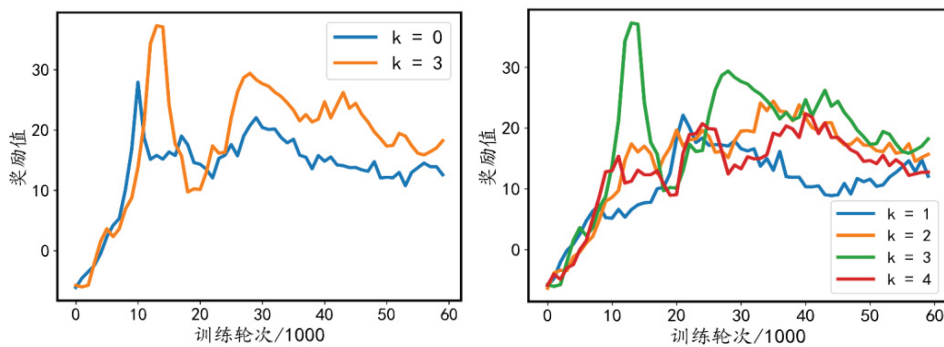


图 4 观察叠加法对于捕食者奖励值的影响

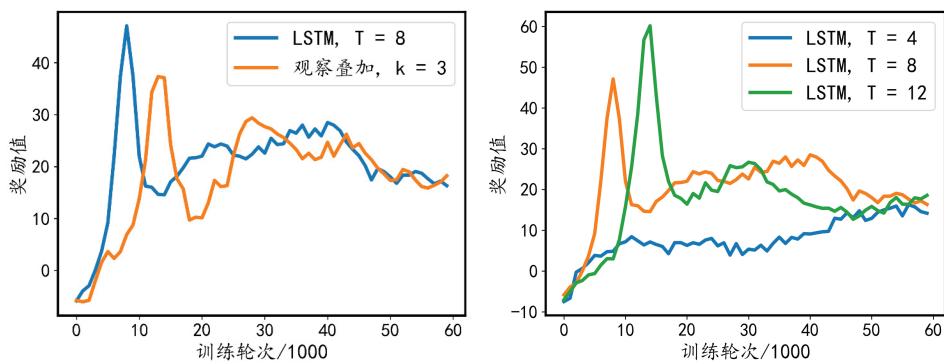


图 5 引入LSTM网络对于捕食者奖励值的影响

之后是对引入HER方法效果的验证, 在合作通讯场景下, 效果如图6所示. 新目标地标的选取方法是: 回合末说者通讯动作(3维)中第*i*维最大, 则选择第*i*个地标作为新的目标地标, 且其位置更改为回合末听者的位置. 图6 (a) 展示了方法的有效性. 此外, 观察发现, 不使用HER方法, 测试时每次听者都会奔向同一个颜色的目标. 这是由于探索的偶然性, 听者在训练初期较多地成功碰到该颜色地标并获得奖励, 便习得了到达该颜色地标的策略, 奖励的稀疏性使得很难跳出这样的局部最优. 而引入HER方法之后, 成功经验增加, 引导智能体习得了正确的策略.

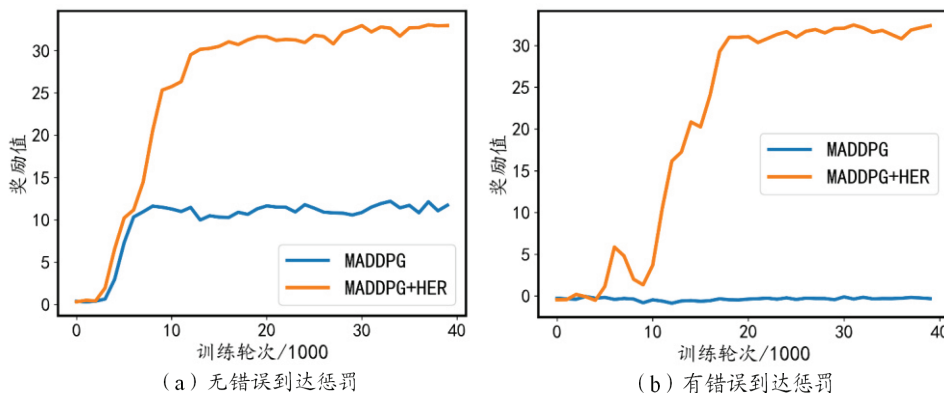


图 6 引入HER方法在合作通讯场景下的效果

上述新目标选取方法记为方法1. 同时实验中还对对比了另一种新目标选取方法的效果, 记为方法2, 其中目标地标仍为该回合原定地标, 只将其位置更改为回合末听者的位置. 事实上图6 (a) 展示的是典型情况, 由于通讯合作任务的困难和实验的随机性, 每次训练可能会收敛到不同的策略. 对每种方法(包括原MADDPG算法)重复进行了12次实验, 每次训练40 000回合; 结果是在方法1下, 智能体有10次学到了正确策略(平均回合奖励可达30), 方法2为两次, 而原方法为0次. 相比方法2, 方法1更好地引导了通讯动作的学习(相当于给出了通讯规则), 从而使得整个合作任务的学习更为顺利.

考虑在学习到的错误策略中,智能体会到达错误的地标,于是在奖励中加入了碰撞非目标地标的惩罚项,结果如图6(b)所示.在这样的奖励设置下,不使用HER方法,听者习得的策略是不碰撞任何地标;而使用HER方法后仍然能习得正确策略.同样重复进行了12次实验,HER方法下(使用方法1选取新目标),智能体有11次学习到了正确策略,原方法下为0次(偶尔能学习到在某1或2个特定目标下成功的次优解).

图7则展示了引入HER方法在合作导航场景下的效果.新目标的选取方法是:对于1号合作者,找到离它最近的地标,将该地标的位置改为1号合作者回合末的位置;对于2号合作者则在剩下的地标中找离它最近的,以此类推.由图7(a)可知,在稀疏奖励设置下,HER方法能够加速收敛,最终收敛时的奖励值也稍高.观察发现,若不使用HER方法,则合作者的策略不够灵活,比如每个合作者每次都会去向固定颜色的地标.相比之下,使用HER方法时每个合作者每次占据的地标不是固定的,能够根据每次初始化后的实际情况选择去占据的地标,更为灵活的策略使得合作者们的碰撞减少,以及更快地占据3个地标.此外,基于对HER方法的认识,在非稀疏奖励下增加成功经验也应当是有效的.图7(b)对应的实验中,将占据地标的奖励改为了基于每个智能体与最近的地标的距离持续给出的奖励,为负值.结果验证了在非稀疏奖励下的有效性,但效果不如在稀疏奖励情形下显著.

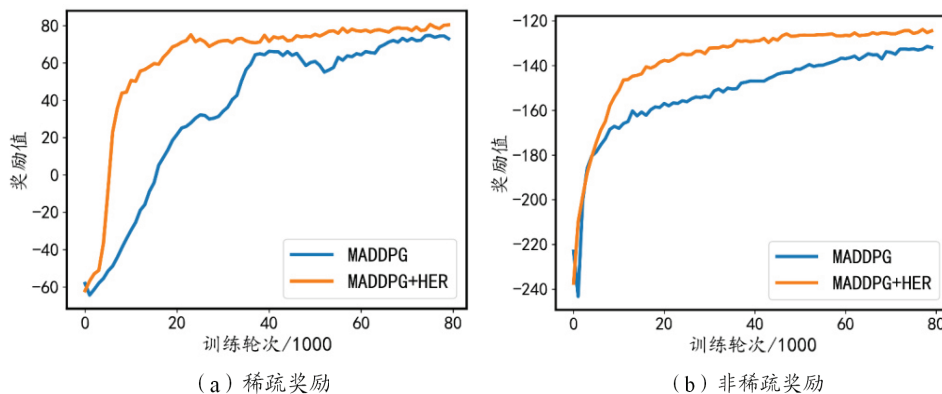


图7 引入HER方法在合作导航场景下的效果

4 结论

本文研究了多智能体深度确定性策略梯度算法(MADDPG),采用了观察叠加和引入LSTM层的方法增强了MADDPG在部分可观察场景下的表现.观察叠加法实现简单,但会因叠加后智能体输入的观察维度的过度提升而增加其决策的困难;引入LSTM层的方法同样有效,但同时会大幅增加训练时间;两种方法可根据实际情况选用.引入后验经验回放方法增强了MADDPG在稀疏奖励下的表现,在多智能体合作通讯场景中原算法下智能体难以达成任务目标,引入HER方法缓解了这一问题,使得智能体能够较快地习得正确策略;在多智能体合作导航场景下原算法也能完成任务,而引入HER方法后智能体可以学习到更优的策略,同时进一步验证了在非稀疏奖励场景下引入HER方法是有效的.

本文虽然在MADDPG改进研究方面取得了一些有效的成果,但是也存在一些不足.在引入LSTM网络的方法上,可以考虑别的形式,比如增加一个单独的LSTM网络,以actor为例,可将智能体自身历史观察、历史动作作为LSTM网络的输入,而其输出则与智能体当前观察拼接,这样也能做到对历史信息的利用.此外,在HER方法的使用上也存在一些限制,比如对于追逐场景,猎物(也是智能体)的位置一直在变化,很难在不更改智能体状态的情况下使用HER方法.后续可以借鉴分层的思想,在团队协作任务中,以指挥者为上层、小组成员为下层,尝试解决稀疏奖励和长期奖励的问题.

参考文献:

- [1] 李茹杨, 彭慧民, 李仁刚, 等. 强化学习算法与应用综述[J]. 计算机系统应用, 2020, 29(12): 13-25.
- [2] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992, 8(3): 229-256.
- [3] SCHULMAN J, LEVINE S, ABBEEL P, et al. Trust region policy optimization[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille: PMLR, 2015.

- [4] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB/OL]. (2017-08-28)[2022-12-20]. <https://doi.org/10.48550/arXiv.1707.06347>.
- [5] MNIEH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518: 529-533.
- [6] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3): 279-292.
- [7] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. (2019-07-05)[2022-12-20]. <https://doi.org/10.48550/arXiv.1509.02971>.
- [8] 孙彧, 曹雷, 陈希亮, 等. 多智能体深度强化学习研究综述[J]. 计算机工程与应用, 2020, 56(5): 13-24.
- [9] LOWE R, WU Y, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2017.
- [10] ANDRYCHOWICZ M, WOLSKI F, RAY A, et al. Hindsight experience replay[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2017.
- [11] 孙彧, 徐越, 潘宣宏, 等. 基于后验经验回放的MADDPG算法[J]. 指挥信息系统与技术, 2021, 12(6): 78-84.
- [12] HAUSKNECHT M J, STONE P. Deep recurrent Q-learning for partially observable MDPs[C]//2015 AAAI Fall Symposium. Palo Alto: AAAI, 2015.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [14] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2011.
- [15] SCHAUL T, HORGAN D, GREGOR K, et al. Universal value function approximators[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille: PMLR, 2015.

责任编辑: 张自强

(上接第708页)

- [11] 李春霞, 徐泽进, 乔曼, 等. 催化裂化油浆超临界萃取组分热缩聚生成中间相沥青的定量研究[J]. 石油学报(石油加工), 2015, 31(1): 145-152.
- [12] 查庆芳, 张玉贞, 巩春伟, 等. 催化裂化油浆溶剂抽提及其中间相沥青[J]. 炭素技术, 2001(2): 1-5.
- [13] 刘银全, 王秀娥, 李安邦, 等. 针入法沥青软化点和可纺性的简易测定[J]. 新型碳材料, 1994(3): 30-31.
- [14] 卢超. 催化裂化油浆预处理及制备针状焦的研究[D]. 上海: 华东理工大学, 2013.
- [15] 梁文杰. 重质油化学[M]. 东营: 石油大学出版社, 2000.
- [16] ESER S, JENKINS R G. Carbonization of petroleum feedstocks I: relationships between chemical constitution of the feedstocks and mesophase development[J]. Carbon, 1989, 27(6): 877-887.
- [17] ESER S, JENKINS R G. Carbonization of petroleum feedstocks II: chemical constitution of feedstock asphaltenes and mesophase development[J]. Carbon, 1989, 27(6): 889-897.

责任编辑: 张自强 刘敏