一种基于心智理论方法的测试时缓解策略，用于应对涌现式对抗通信

扩展摘要

Nancirose Piazza SAIL实验室 - 新海文大学 美国康涅狄格州西黑文 npiazza@newhaven.edu

Vahid Behzadan SAIL实验室 - 新海文大学 美国康涅狄格州西黑文 vbehzadan@newhaven.edu

arXiv:2302.07176v1 [cs.MA] 14 Feb 2023

摘要

多智能体系统（MAS）研究的是共享环境中的多智能体交互。合作通信是部分可观测环境中共享信息的一项基本构成。合作式多智能体强化学习（CoMARL）是一种学习框架，在该框架中，我们学习智能体策略，或通过合作机制实现学习，或学习表现出合作行为的策略。具体而言，已有研究探讨如何让 CoMARL 智能体学会传递消息；然而，当非合作型智能体能够接入合作团队的通信信道时，研究发现它们会学会发送对抗性通信消息，破坏合作团队的表现，尤其是在目标依赖有限资源时。为解决这一问题，我们提出一种技术，利用心智理论（ToM）的局部表述，在接受任何智能体的消息之前，区分其表现出的合作行为与非合作行为。我们在一个采用集中式训练、分散式执行（CTDE）的 CoMARL 基准中进行了实证评估，证明了所提技术的有效性和可行性。此外，尽管我们提出的心智理论防御明确用于测试时，但我们强调，心智理论是一种用于设计认知防御的构想，而非防御本身的目标。

关键词

对抗性通信；心智理论；多智能体强化学习；测试时防御

ACM参考文献格式：

Nancirose Piazza 和 Vahid Behzadan。2023年。作为应对涌现式对抗通信的测试时缓解措施的心智理论方法：扩展摘要。载于*第22届国际自主智能体与多智能体系统会议（AAMAS 2023）*论文集，英国伦敦，*2023*年*5*月*29*日至*6*月*2*日，IFAAMAS，共7页。

1 引言

多智能体系统存在于许多对社会运转至关重要的重要领域，将自动化的近期部署扩展至自主移动交通工具[27],智能交通系统[16], 金融交易投资组合[8],以及云边缘系统中的联合用户调度，以最小化用户设备的延迟和即时服务质量（QoS）[23]。针对这类问题的多智能体方法能够将这些复杂且计算上不可行的任务转化为可分布式、并行且可在本地管理的工作负载。在涉及学习序列策略和行为的任务中，多智能体强化学习（MARL）框架涌现出了前景广阔的方法。 [4] 对多智能体强化学习开展的综合综述，印证了人们对多智能体强化学习方法长期以来的关注，以及对相关方法的多元视角，例如价值分解、策略梯度和通信。 [25] 近期发表的论文则提供了一系列与当前研究相关的理论和算法。

对智能体间关系的研究通常将其描述为合作型或非合作型（例如竞争型）。完全合作的智能体通常被建模为团队博弈，并可获得共享奖励。一些针对完全合作型多智能体强化学习任务（例如 [1], [20],[21], [14], [11]）的基于价值的研究，会考虑估计智能体的个体价值函数、共享奖励对应的全局价值函数，和/或通过某种机制进行奖励塑形。另一方面，智能体也可以被描述为完全竞争型：一个智能体对其局部奖励的最大化，会导致另一个智能体的奖励收益受损。介于完全竞争与完全合作之间的是混合型，此时智能体可能同时表现出竞争行为与合作行为。

通信被视为合作智能的一大支柱，是实现协商（例如通过协商促成通信的涌现 [5]）以及共享或传递信息的重要渠道。通信多智能体强化学习通常采用消息传递的形式，以处理部分可观测信息和/或为协调提供信号提示。一些通信多智能体强化学习的设计架构会整合图模型来聚合各种输入，以进行决策，通常遵循集中训练、分散执行（CTDE）范式。然而，这类机制和设计可能遭到利用。 [3] 的研究表明，自利智能体能够学会向合作团队的通信信道发送对抗性消息，从而损害合作团队的表现。将这一情况扩展到现实世界场景，成功传递对抗性信息可能导致级联故障和灾难性事件（例如，复杂网络中级联故障的影响 [22]。）

本文的贡献如下：

- 我们提出将心智理论（ToM）作为防御对抗性通信的一种认知机制；具体而言，我们的心智理论设计基于历史上可观察到的邻居智能体行动。
- 我们提出一种基于信念的信任防御机制，供同策略智能体或已安全获取其他合作型智能体策略的智能体在测试时使用。这种防御无需额外训练，并能让智能体根据可观察到的邻近智能体行为，对其他智能体形成分散式心智理论信任信念

第*22*届国际自主智能体与多智能体系统会议论文集（*AAMAS 2023*），*A. Ricci*、*W. Yeoh*、*N. Agmon*、*B. An*（编），*2023*年*5*月*29*日至*6*月*2*日，英国伦敦。 © 2023年国际自主智能体与多智能体系统基金会（www.ifaamas.org）。保留所有权利。

，以此选择采纳哪些智能体传达的消息。

- 我们展示了来自对抗性通信环境 CoverageEnv 的实证结果。
- 我们讨论了这一实现如何归入现有分类体系，并与其他类似文献进行比较，以区分彼此的差异。

2 问题陈述

我们用一个元组来定义多智能体强化学习设置:

$$< V, A, S, P, \{R^i\}_{i\in V}, \{O^i\}_{i\in V}, O, \gamma >$$

其中 V 是智能体集合 $\{1, 2...|V|\}$， A 是动作空间， S 是状态空间， P 是转移概率矩阵， $R^i : (i_a) \to \mathbb{R}$ 是智能体 $i \in V$ 采取动作 a 时的局部奖励函数， O^i 是智能体 i 在状态 s 下传输的观测， O 是智能体 i 在状态 s 下的真实观测，而 $\gamma \in [0, 1)$ 是折扣因子。

此外，我们寻找一个联合策略 $\pi : (\pi_1, \pi_2...\pi_{|V|})$，该策略生成联合动作 $(a_1, a_2, ..., a_{|V|})$，使每个合作的同质智能体的局部奖励最大化。

智能体的图网络 $G =< V, E >$ 其中 V 是智能体集合 $\{1, 2...|V|\}$， E 是任意一对通信智能体之间现有边的集合 $V \times V$，该集合可能会随时间动态变化。边集合 E 表示智能体对 (i, j) 之间的双向去中心化通信信道集合，其中智能体 i 将 O^i（智能体 i 的观测）传输给智能体 j。该观测或传输的消息可以是任何信息，例如局部观测/状态、经过设计的信息（如信任分数）、局部奖励等。在去中心化执行中，通信中共享的是编码信息，而不是局部观测；但我们将局部观测纳入通信，使智能体能够自行评估其他智能体的观测。换言之，我们不再让一个智能体直接信任其他智能体的编码信息，而是要求传输设计者能够理解的状态观测。此外，还可以对状态观测采用传统的、非机器学习的过滤防御，通过让设计者理解有效扰动来最大限度地减少具有迷惑性的对抗性消息。

我们通过信任历史机制来缓解对抗性通信：根据先前观察到的其他智能体的自承诺行为，调整对其的信任概率及调整方向和幅度，并以心智理论为目标。与人类推理类似，智能体会根据其他智能体的观测行为进行推理，并基于过去的证据更新信念，例如信用记录。在本研究中，我们考虑同质智能体（依照环境而定），但只要评估者可用且对队友公开，这种方法也可以扩展到任何异质智能体团队。打个比方，个人用于电子邮件的公钥可以上传到共享密钥服务器，供他人使用。

3 相关工作

在多智能体强化学习环境中学习通信[7]，可以将由于部分观测而导致的困难任务转化为信息更完整的任务，从而最大限度地提高任务竞争的成功率。然而，当通信信道能够直接影响一部分智能体时，攻击面会扩大，这凸显了意识到对抗性通信的重要性。此外，由于这些智能体采用半监督学习，若没有适当的约束来避免学习不良行为，多智能体强化学习[3] 中就可能出现对抗性通信等行为，而且若无先验知识便难以应对。

可信度建模是防御的一种常见方向。例如，有一篇名为《使用隐马尔可夫模型对上下文感知动态信任进行建模》的论文，由 [10] 撰写，该论文对智能体交互进行建模，理想情况下还会对智能体意图进行建模。我们的工作也属于可信度建模，但我们采取了不同的方向，选择受可信性启发的模型。Schillo 等人 [18] 的一项相关研究将其他智能体过去的观测动作作为初始信任模型。该研究也利用了其他智能体的观测动作。尽管基于共识的防御是可行的（我们也曾考虑过），我们还是没有采用这种方案，而是倾向于更偏向自我评估的可信度模型，以尽量减少对外部输入的需求。

近期一项由 [13] 开展的调查旨在推动分类体系标准化并建立多智能体强化学习智能体的鲁棒性标准，提出了多智能体强化学习面临的开放问题，并依据其分类体系整合了现有研究。我们认为，我们的工作应归入后端感知与过滤方法。该调查根据应激因素（如噪声等干扰）将方法划分为操作前、操作中和操作后三类。我们的方法会在执行期间更新信念，因此很可能属于针对分布外应激因素的操作中方法；不过，由于对手仍然关注其自身在环境中的表现，我们的防御也试图应对分布内应激因素。

[12] 提出了一种针对对抗性通信的防御方法：将变分自编码器贝叶斯模型用作消息过滤方法，根据基于互信息和信息一致性得出的真实性置信度，对其他智能体的消息进行加权。其工作与我们通过概率信念处理问题的方式相近；不过，我们不需要构建信念状态并训练另一个模型，而是让信念状态依赖于分回合执行期间可观察到的行为。 [6]的另一项研究则利用基于共识的决策，在群集问题中检测并消除对抗性机器人。相比之下，他们的方法更倾向于在检测到对抗行为后作出反应，并依赖未受损的多数方；而他们通过集中式方法评估信任的方案，与我们的去中心化方法形成对照且互为补充。

或者，完整的对抗型智能体可能并不现实。相比之下，合作型智能体遭到劫持、在一段时间内受到攻击，或因不确定性等非蓄意恶意因素而受影响，可能更为常见。在这种情况下，[26] 的研究阐述了风险感知智能体以及面对不确定性时状态韧性的当前研究方向。许多应对不确定性的方法也可用于增强对非合作行为（包括对抗行为）的韧性。例如， [17] 提出了一种通过分布式韧性控制器实现韧性编队控制的方法，使移动机器人智能体组成的群体能够更好地应对非合作或有缺陷的智能体。检测对抗型智能体也可以采用更偏向身份验证的方法。 [15] 的研究利用物理指纹，并分析合法邻近智能体的这些指纹，以识别实施欺骗的对抗型智能体。

4 自揭示行为

假设我们有一组双人博弈，已知这些博弈的分布，但玩家的估值函数是私有的。玩家可以进行“廉价磋商”，即通过口头方式交流消息；在此情况下，消息中会陈述他们将采取的动作。此外，当消息中陈述的动作可以被观测到时，用于交流的消息可称为自揭示或自承诺。来自 p_1 的自揭示消息会向 p_2 告知当前的博弈 G_i。在即时收益情形下， p_2的消息会向 p_1 揭示双方的关系可能是合作还是竞争，从而传递有关当前博弈的信息。当来自 p_1 的消息与 p_2对 p_1动作的观测 a_1一致时，该消息就是自承诺的。这种承诺对智能体冲突解决等研究领域十分重要，例如谈判[2], ；不过，我们的工作并不试图解决冲突，而是根据行为预测智能体角色。为使本文的术语与其他研究保持一致，如果后续观测到的动作与根据先前消息推断出的预期动作相符，我们就称该动作是一致的。通过将消息描述为一致的，我们可以像人类一样建立其他智能体的信任历史。

虽然学习通信的理想用途是鼓励合作学习行为，但信号不必是自承诺的；换句话说，信号可能是胡言乱语，此时无法从消息中推断出任何信息，也可能具有欺骗性，如 [3] 所示。虽然在合作任务中，人们通常不会预期出现非自承诺信号，但面对目标不一致的情况时，在合作环境中训练的策略可能会在处理非合作行为时出现适应不良且脆弱的问题。此外，传输中的噪声也可能导致沟通失误，因此应将其纳入韧性考量。一种有助于团队智能体在存在非合作行为时保持正常行动的方法，是根据它们的历史建立信念，其中既包括他人发送的消息，也包括未被直接观测到的其他智能体动作。类似于 [12] 中由某种经验信念决定可信度的情况，我们认为，如果其他智能体的消息不是自承诺的，那么通信消息就不应影响智能体的决策。

5 测试时的心智理论

心智理论（ToM）是将其他智能体合理化为处于某种信念状态的过程，通常通过贝叶斯建模实现。贝叶斯模型具有信念状态，用于根据离散或连续的时间度量以及一些公共信息，近似估计一系列结果的概率频率。控制理论中一种广为人知的贝叶斯估计算法是卡尔曼滤波器，参见 [9] 关于其应用的综述；该算法能够根据一系列观测结果估计随时间变化的未知变量和不确定性。虽然我们不使用卡尔曼滤波器，但它当然可以作为目标跟踪系统（例如自动驾驶机动车）的心智理论防御。此外，当智能体学会消息编码后，卡尔曼滤波器的扩展可能更难建模。心智理论在人工智能领域已有充分讨论；然而，对其他智能体的信念状态进行显式建模，可能需要额外且繁重的计算负载，在大规模智能体系统（例如群体机器人学）中未必可行[19] ，而理想情况下每个智能体的计算开销本来并不高。然而，均值场多智能体强化学习（MF-MARL）[24]可以作为大规模 N 智能体的一种替代框架，在该框架中，我们只记录定义半径范围内智能体的可观测动作。

我们强调，心智理论应影响防御的设计，而不应将某一种明确的心智理论解释——即直接建模其他智能体的信念状态——视为其实现方式的唯一表征。

我们从合作智能的角度看待心智理论：智能体遵循这样的推理：‘如果我是他们，我会怎么做？我会发送什么消息来合作？他们观测到的行为是否与我根据其发送的消息所做的预期一致？’

理想情况下，如果智能体 i 观察到智能体 j的行为与自己的预期一致，那么智能体 i 可能会倾向于信任智能体 j发送的通信消息并加以利用。如果行为与预期不一致，则可以丢弃该消息。

许多防御依赖分布外检测，包括用于消息过滤的 VAEB 防御 [12]。然而，这类防御需要在测试时之外进行额外训练，部分原因是其重点在于转换高维数据。相反，我们将注意力放在维度较低的可观测智能体动作上。我们的心智理论信念防御类似于人类的信用记录，利用观测到的动作历史来更新信任概率。这种防御针对测试时，因为智能体的信念只与当前回合及其历史相关，这与传统的跨回合学习方法不同。此外，我们认为应采用分层防御，以尽量减少对抗性通信可能造成的损害；这意味着我们的防御与传统防御相辅相成。

6 近端策略优化

在多智能体和单智能体强化学习中，推导策略有两种方法：基于价值的方法和基于策略的方法。基于价值的方法通过优化参数来估计最优的 Q^* 函数；在多智能体场景中，这可能是一个联合 Q 函数，用于提取联合策略 π。策略梯度方法通过梯度下降优化参数，以最大化动作的期望对数似然，直接 π。演员-评论家是一种方法，它在评论家/基于价值的函数（Q、V或 A/优势函数）的辅助下优化演员/策略 π。合作策略梯度的形式如下：

$$g_k^i = \mathbb{E}_\pi\left[\sum_{j\in V}\nabla_\theta \pi_\theta^j(a^j|O)A^i(s)\right]$$

该方法已被证明以概率 1 收敛到所有智能体期望回报总和的局部最大值[3] ，并且联合策略使累积奖励总和最大化。

7 一致行动的公共价值评估

鉴于采用演员-评论家架构这一假设，我们假设每个智能体都可以访问价值函数 Q、 V或 A。尽管在实践中，智能体的部署仅涉及演员/策略本身，但我们也允许智能体包含价值函数 $\hat{}V$。我们定义

价值差异 $\hat{}\rho$ 如下：

$$\hat{\rho} = \hat{V}(s_j, \pi_i(s_j)) - \hat{V}(s_j, \pi_j(s_j)) \quad (1)$$

其中，$\pi_j(s_j)$表示智能体 i在给定s_j时对智能体 j所采取动作的观察。由于假设智能体 i和 j的策略相同，因此可以依据智能体 i 的策略进行评估。我们引入阈值 ρ：若在时间步 $t+1$ 针对消息 m_{ij}^t进行评估时，$\hat{}\rho > \rho$，则认为智能体 j的动作不一致。同样，若评估结果处于阈值范围内，则认为该动作一致。

随机策略的扩展可以使用样本的经验集合来估计训练期间的 $V(s_j, \pi_i(s_j))$，从而为样本创建计算所得的 Kullback–Leibler 距离分布，并将其均值用作阈值：

$$KL(\hat{V}(s_j, \pi_j(s_j)) || \hat{V}(s_j, \pi_i(s_j))) \quad (2)$$

算法 1： 更新_一致性_计数

结果： 更新一致性计数
对于智能体 ∈ 所有 _智能体 **执行**
  智能体.信任 = {0 : 0<style id="2">, 1 : 0}∗ 长度(所有_智能体)
end
A_c ←评论家的优势函数
a ←观测到的动作
$o\prime$ ← 目标智能体通信
已评估_索引 ←目标智能体索引
对于合作_智能体 ∈ 所有_智能体 **执行**
  评估者 ← 合作_智能体
  如果 $|\max_{a'} A_c(o') - A_c(o'a)| <= \rho$ **那么**
    评估者.信任[已评估_索引][1] +=1
  else
    评估者.信任[已评估_索引][0] +=1
  end
end

算法 2： 更新_信念

结果： 使用.....调整一个智能体对另一个智能体的信念
  先前观察到的动作的方向以及
  与一致性计数类别成正比的幅度。
对于智能体 ∈ *Coop*_智能体 **执行**
  对于智能体_索引 ∈ 所有¯智能体 **执行** 智能体.信念[智能体_索引] = 1.0 // 概率 1.0
  end
end
s ←学习率
对于 $t \in \{2,...N\}$ **执行**
  j →询问智能体
  i →被询问智能体
  如果 $\pi^j(o_{t-1}^i) = a_{t-1}^i$ **那么**j.belief[i] + s × (j.trust[i][1] / t) 提高信任概率。
  else
    j.belief[i] - s × (j.trust[i][0] / t) 降低信任概率。
  end
end

算法 1 和算法 2 中概述的理想实现如下：每个智能体都会持有关于所有其他智能体的概率信念，并根据各自的概率决定是否信任其他智能体。由于没有不信任的证据，默认将所有智能体的概率设为 1.0。在每个环境步之后，所有智能体都会根据从该环境步中观测到的动作重新评估对其他智能体的信任，并依据前一个动作是否一致来调整其信念；调整幅度与观测到的一致动作或不一致动作的累积和成正比。

8 实验结果

我们使用了由 [3],training 提供的对抗通信代码库，并采用 PPO 行动者-评论家等默认设置；不过，我们将策略设为始终选择分布中概率最高的动作。覆盖任务所处的环境要求一组智能体尽可能覆盖网格世界地图。智能体可以访问中央通信信道，其提取的特征图会由图移位算子（GSO）加权；该算子通过聚合图神经网络（AGNN）确定相邻节点的权重。在去中心化执行和去中心化通信中，GSO 在本地计算；我们假设第三方（类似于注册机构（RA））可以验证智能体位置的有效性。

我们使用 $N = 5$ 个智能体训练合作策略，共进行 600 万个时间步。在固定合作策略的条件下，自利策略训练了 600 万个时间步。随后，在固定自利策略的条件下，又进行了 600 万个时间步的重新适应训练。评估设置包含一个自利型智能体（Agent0）和三个合作型智能体（Agent1、Agent2、Agent3）。

我们在图 1 中展示了基线比较：变分自编码器贝叶斯消息过滤器（VAEB）在训练环境中使用提取的样本／单个智能体编码后的特征图进行训练，其理想性能与我们的心智理论防御（ToM）（所选学习步长乘数为 3.7）、次级基线无防御（NoDef），以及合作型智能体的理想性能（理想合作）进行比较。使用生成的样本 s，我们设定：如果$||s||_2 > c$ 对于某个选定的 c 常数成立，则拒绝该消息；否则接受该消息。就 VAEB 的理想性能而言，心智理论防御在图 1 中的覆盖率表现与之相当。接着，我们在图 2 中展示，当 VAEB 应用于已在自利型消息影响下重新适应其通信消息的合作型智能体时，它并不起作用。这表明新学习到的通信内容很可能偏离 VAEB 的分布，导致较少真实且有用的信息影响智能体决策。这说明，VAEB 等消息过滤方法，以及 [12]提出的更复杂的高斯过滤器 VAEB，都难以应对非平稳学习表征，并且需要针对分布偏移进行额外训练。消息过滤通常依赖消息的平稳性来保持一致性，而非依赖序列动作期望的一致性。

为支持我们的假设，我们发现心智理论的表现不受适应性训练的影响。此外，我们发现，自利型智能体在能够访问共享

通信时表现更好，这也在预料之中。我们还发现，性能下降主要归因于虚假通信消息，而非环境中存在对手：允许自利型智能体向合作团队发送正确消息后，其表现与自利型智能体不进行通信时相近。我们纳入了自利型智能体在进行通信时的基线（Adv NoDef），因为这表明其性能提升直接导致合作团队的性能损失，这是共享通信遭到操纵所致。

我们按智能体提供了图4和图3，分别显示假阴性和假阳性的均值计数。可以看到，由于假阳性多于假阴性，且图6和图5分别显示了合作型智能体的真阳性均值计数，以及非合作/自利型智能体的真阳性均值计数，因此，我们的防御在隐式设计上更倾向于不信任智能体，而非信任智能体。此外，虽然 VAEB 防御在合作消息分布固定时能够有效拒绝分布外消息，但如果重新适应训练没有试图将消息学习限制在当前 VAEB 的消息分布范围内，就会导致更多假阳性。我们还提供了图7，展示各智能体在适应性训练后每一步的 F1分数。结果显示，随着步数增加，我们的防御的 F1分数也随之提高，而 VAEB 则趋于停滞，这很可能是由于假阳性增加所致。重新适应训练后，我们的防御性能没有下降，这表明这种额外的防御不仅比持续重新训练成本更低，而且在叠加于其他防御之上时也具有灵活性。

我们尝试的另一个值得关注的方向是构造对抗性消息；然而，相较于 VAEB，我们未能成功利用 AGNN架构生成有效的定向攻击。多次尝试使我们认为，规避攻击——即寻找能够改变所有其他智能体动作的对抗样本，尤其是但不限于定向攻击——可能比寻找能够欺骗采用标准深度学习架构的单个智能体的对抗样本更困难。

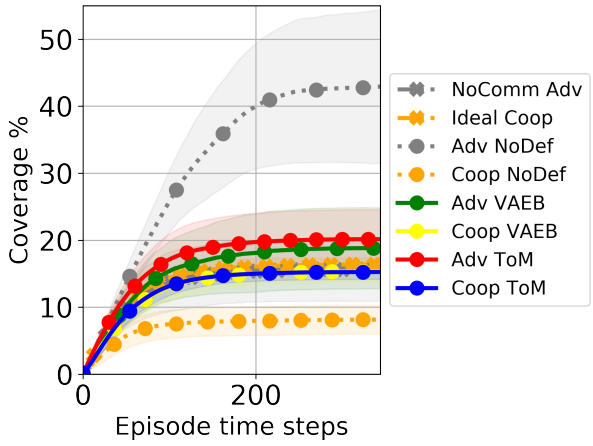

图 1：CoverageEnv——心智理论性能与理想防御基线 VAEB 及无防御情形的对比

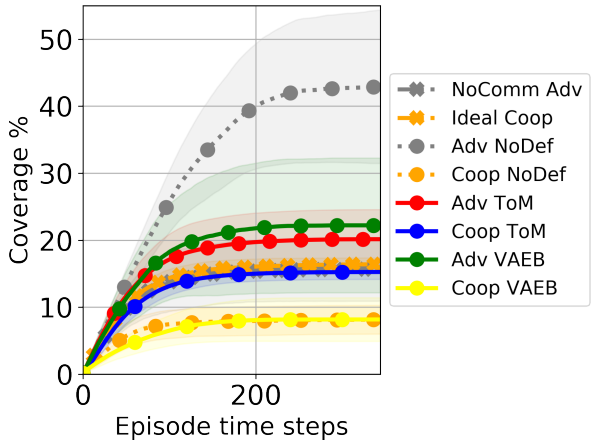

图 2：CoverageEnv——重新适应训练后，心智理论性能与防御基线 VAEB 的对比

9 结论

本文提出了心智理论（ToM），通过观测到的动作来合理解释其他智能体的行为，从而确定对其他智能体所传递消息的信任信念。我们的防御机制适用于测试时，无需额外训练或重新训练；在允许合作型智能体进行适应的情况下，该机制也易于与其他防御机制结合。我们研究了一个会出现对抗性通信的多智能体环境，并展示了心智理论信任机制防御的用法，将采用该防御机制时合作团队的平均表现，与不设防时合作团队的平均表现以及采用变分自编码器贝叶斯防御基线时合作团队的表现进行了比较。

10 引文与参考文献

[1] Adrian K. Agogino 和 Kagan Tumer。2004年。统一时间与结构信用分配问题。载于第三届自主智能体与多智能体系统国际联合会议论文集——第2卷（纽约，纽约州）（*AAMAS '04*）。IEEE计算机学会，美国，980–987。

[2] Martin Beer、Mark D'inverno、Michael Luck、Nick Jennings、Chris Preist 和 Michael Schroeder。1999年。《多智能体系统中的协商》。知识工程评论 14, 3 1999 285 289 10 1017 0269888999003021 （sep），–。

[3] https://doi.org/ 。/S Jan Blumenkamp 2020 和 Amanda Prorok。。多智能体强化学习中的对抗性通信涌现。arXiv: 2008 02616 [ ]。cs.RO

[4] 2008 Lucian Busoniu、Robert Babuska 和 Bart De Schutter。。多智能体强化学习综述。《系统、人和控制论》*C* 部分：应用与评论，*IEEE* 汇刊38（04 2008年），156–172。https://doi.org/10.1109/TSMCC.2007.913919[5] Kris Cao、Angeliki Lazaridou、Marc Lanctot、Joel Z Leibo、Karl Tuyls 和 Stephen Clark。2018年。通过协商实现的涌现通信。https://doi.org/10.48550/ARXIV.1804.03980[6] Mingxi Cheng、Chenzhong Yin、Junyao Zhang、Shahin Nazarian、Jyotirmoy Deshmukh 和 Paul Bogdan。2021年。多智能体系统通用信任框架。载于第 *20* 届国际自主智能体与多智能体系统会议论文集（英国，虚拟会议）（*AAMAS '21*）。国际自主智能体与多智能体系统基金会，里奇兰，南卡罗来纳州，332–340。

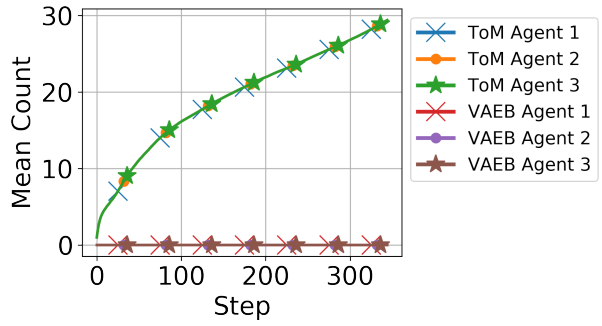

图3：经过重新适应训练后，假阴性（可信对手）智能体的均值计数

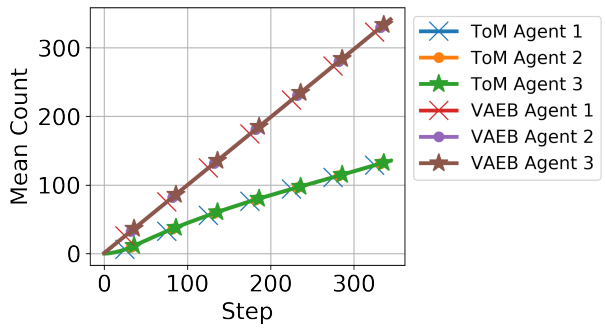

图4：经过重新适应训练后，假阳性（不可信协作者）的均值计数

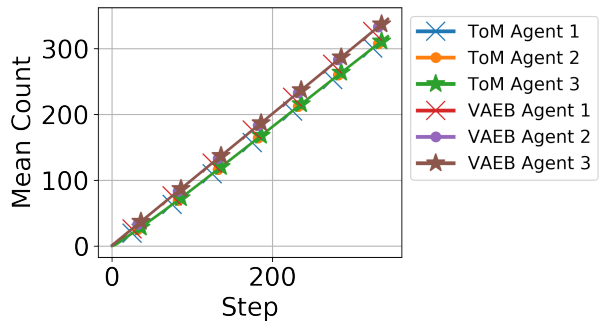

图5：经过重新适应训练后，真阳性对手检测的均值计数

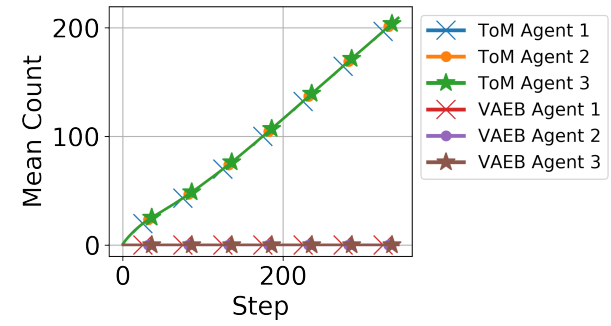

图6：经过重新适应训练后，真阳性协作者检测的均值计数

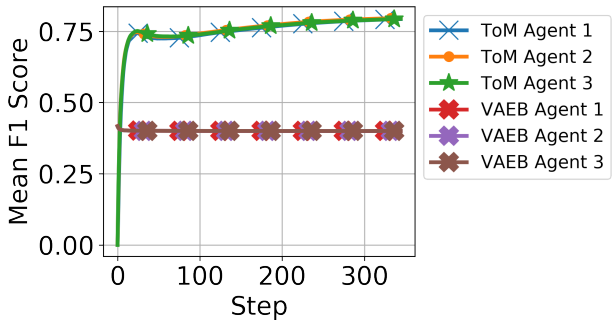

图7：采用重新适应训练时各步骤的F1分数均值

[7] Jakob N. Foerster、Yannis M. Assael、Nando de Freitas 和 Shimon Whiteson。2016年。利用深度多智能体强化学习进行通信。arXiv: 1605 06676 [ ]。cs.AI

[8] Jinho Lee、Raehyun Kim、Seok-Won Yi 和 Jaewoo Kang。2020年。MAPS：基于多智能体强化学习的投资组合管理系统。第二十九届国际人工智能联合会议（2020年7月）。https://doi.org/10.24963/ijcai.2020/623

[9] Qiang Li、Ranyang Li、Kaifan Ji 和 Wei Dai。2015年。卡尔曼滤波器及其应用。载于*2015*年第*8*届智能网络与智能系统国际会议（*ICINIS*）。74 77 10 1109 2015 35 – 。https://doi.org/ 。/

[10] ICINIS。。Xin Liu 和 Anwitaman Datta。。使用隐马尔可夫模型对上下文感知动态信任进行建模。载于第二十六届*AAAI*人工智能会议论文集（多伦多，安大略省，加拿大）（*AAAI'12*）。AAAI出版社，1938–1944。[11] David Mguni、Joel Jennings、Emilio Sison、Sergio Valcarcel Macua、Sofia Ceppi 和 Enrique Munoz de Cote。2019年。协调群体：在非合作系统中诱导理想均衡。载于第*18*届自主智能体与多智能体系统国际会议论文集（蒙特利尔，魁北克省，加拿大）（*AAMAS '19*）。国际自主智能体与多智能体系统基金会，南卡罗来纳州里奇兰，386–394。[12] Rupert Mitchell、Jan Blumenkamp 和 Amanda Prorok。2020年。对抗性通信环境下基于高斯过程的消息过滤，实现鲁棒多智能体协作。arXiv: 2012 00508 [ ]。cs.RO Amanda Prorok、

[13] Matthew Malencia、Luca Carlone、Gaurav S. Sukhatme、Brian M. Sadler 和 Vijay Kumar。2021年。超越鲁棒性：弹性多机器人系统方法的分类体系。arXiv: 2109 12343 [ ]。cs.RO Tabish Rashid、

[14] Mikayel Samvelyan、Christian Schroeder de Witt、Gregory Farquhar、Jakob Foerster 和 Shimon Whiteson。2018年。QMIX：用于深度多智能体强化学习的单调价值函数分解。https: //doi.org/ 10 48550 1803 11485 。/ARXIV。。

[15] Venkatraman Renganathan 和TylerSummers。2017年。分布式多机器人系统的抗欺骗协调。载于*2017*年多机器人与多智能体系统国际研讨会（*MRS*）。135–141。https://doi.org/10.1109/MRS。2017.8250942[16] Sergey Satunin 和 Eduard Babkin。2014年。采用组合拍卖的智能交通系统建模的多智能体方法。*Expert Syst. Appl.* 41, 15 2014 6622 6633 10 1016 2014 05 015 （11月），– 。

[17] https://doi.org/ 。/j.eswa。。Kelsey Saulnier、David Saldaña、Amanda Prorok、George J. Pappas 和 Vijay R. Kumar。2017年。移动机器人团队的弹性群集。*IEEE*《机器人学与自动化快报》2（2017年），1039–1046。

[18] Michael Schillo、Petra Funk 和 Michael Rovatsos。2000年。利用信任检测人工社会中的欺骗性智能体。应用人工智能14, 8 (2000年), 825–848. https://doi.org/10.1080/08839510050127579 arXiv:https://doi.org/10 1080 08839510050127579 . / Melanie Schranz、Martina Umlauft、[19] Micha Sende 和 Wilfried Elmenreich。2020年。群体机器人行为及当前应用。《机器人与人工智能前沿》7 (2020年)。[20] Kyunghwan Son、Daewoo Kim、Wan Ju Kang、David Earl Hostallero 和 Yung Yi。2019年。QTRAN：通过变换学习分解，以实现协作式多智能体强化学习。https://doi.org/10 48550 1905 05408 . / ARXIV. . Peter Sunehag、Guy Lever、Audrunas Gruslys、Wojciech Marian Czarnecki、[21] Vini- cius Zambaldi、Max Jaderberg、Marc Lanctot、Nicolas Sonnerat、Joel Z. Leibo、Karl Tuyls 和 Thore Graepel。2017年。面向协作式多智能体学习的价值分解网络。https://doi.org/10.48550/ARXIV.1706.05296[22] Lucas D Valdez、Louis Shekhtman、Cristian E La Rocca、Xin Zhang、Sergey V Buldyrev、Paul A Trunfio、Lidia A Braunstein 和 Shlomo Havlin。2020年。复杂网络中的级联故障。《复杂网络期刊》8, 2 (05 2020年)。https:// doi.org/10.1093/comnet/cnaa013 arXiv:https://academic.oup.com/comnet/article-pdf/8/2/cnaa013/33582729/cnaa013.pdf cnaa013。[23] Chunmei Xu、Shengheng Liu、Cheng Zhang、Yongming Huang、Zhaohua Lu 和 Luxi Yang。2021年。基于多智能体强化学习的协作式云边缘系统分布式传输。*IEEE*车辆技术汇刊 技术 70, 2 2021 1658 1672 10 1109 2021 3055511 ( ), – 。[24] https://doi.org/ 。 /TVT. 。 Yaodong Yang 、 Rui Luo 、 Minne Li 、 Ming Zhou 、 Weinan Zhang 和 Jun Wang。2018年。平均场多智能体强化学习。https://doi.org/10.48550/ ARXIV.1802.05438[25] Kaiqing Zhang、Zhuoran Yang 和 Tamer Başar。2019年。多智能体强化学习：理论与算法精选综述。https: //doi.org/10 48550 1911 10635 。/ARXIV. 。Lifeng Zhou 和 Pratap Tokekar。 。[26] 2021 不确定与对抗环境中的多机器人协同与规划。《机器人学最新报告》2, 2（2021年4月）, 147–157。https://doi.org/10.1007/s43154-021-00046-5[27] Ming Zhou、Jun Luo、J ulian Villella、Yaodong Yang、David Rusu、Jiayu Miao、Weinan Zhang、Montgomery Alban、Iman Fadakar、Zheng Chen、Aurora Chongxi Huang、Ying Wen、Kimia Hassanzadeh、Daniel Graves、Dong Chen、Zheng- bang Zhu、Nhat Nguyen、Mohamed Elsayed、Kun Shao、Sanjeevan Ahilan、Baokuan Zhang、Jiannan Wu、Zhengang Fu、Kasra Rezaee、Peyman Yadmel- lat、Mohsen Rohani、Nicolas Perez Nieves、Yihan Ni、Seyedershad Banijamali、Alexander Cowen Rivers、Zheng Tian、Daniel Palenicek、Haitham bou Ammar、Hongbo Zhang、Wulong Liu、Jianye Hao 和 Jun Wang。2020年。SMARTS：面向自动驾驶的可扩展多智能体强化学习训练平台。arXiv:2010.09776 [cs.MA]