摘要

近年来，强化学习在机器学习领域中用于解决决策问题，并取得了显著进展。但同时，该领域也面临着许多挑战，如采样效率低、信用分配问题、探索与利用的平衡问题和泛化问题等。其中，弱泛化能力是阻碍强化学习在实际场景中广泛应用的主要障碍。此外，强化学习算法还可用于处理包含多个智能体（机器人、仪器、汽车等）在同一环境下交互的系统，即多智能体系统。在大多数实际多智能体场景中，智能体无法直接接触完整的状态信息，只能依赖于自己所观察到的信息。这种情况下存在一个非常具有挑战性的问题，即非平稳问题。每个智能体不仅要面对可变的环境，还会受到其他智能体因不断学习而改变的策略所影响。

本文研究如何基于智能体通信来设计强化学习算法，从而提升智能体的性能。本文期望智能体在通信的帮助下，能够形成更优秀的策略，并获得一些良好的特性，例如良好的泛化性和更好的协调能力。当智能体允许与外部通信后，按照通信对象可以分为两种类型。其一，智能体与智能体之间进行通信。智能体与智能体进行通信可以帮助智能体获取到更多的有关其他智能体的信息; 其二，人类也可以传递消息给智能体。与机器不同，人类可以通过另一种通信协议 (自然语言) 来传递信息。自然语言具有很强的抗干扰性，能够传递抽象的概念进行概括，并含有大量描述世界的先验知识。

本文从智能体-智能体通信入手，通过解决系统非平稳性，保证了智能体在各种复杂环境下训练的稳定性和收敛性。此外，通信让智能体之间能够互相协同，形成更好的合作策略。值得注意的是，仅仅提高智能体在某种特定任务的性能是不够的，本文期望能提高智能体策略的泛化能力，即智能体可以在不同的环境或不同的任务中都能保持良好的性能。因此，本文还研究如何通过人类-智能体通信，传递额外的人类先验消息，帮助智能体提高泛化能力。本文主要创新点包括以下四个方面：

首先，为了让智能体在智能体间通信过程中具有处理消息冗余的能力，本文提出了一个基于因果推理的个体推理通信模型，这是领域中第一个采用单对单智能体通信机制的模型。该模型仅仅通过智能体自身观察来评估附近智能体的重要性，进而选择与合适的智能体进行通信，从而减少通信量。模型主要的通信控制模块（先验网络）是通过因果推理在学习过程中动态地标记有影响力的智能体，再通过神经网络去学习该先验知识。目标通信框架能在减少通信的情况下，进一步提升模型性能。

其次，为了进一步提高智能体的性能，智能体需要加强和其他智能体的协调能力。本文提出了一种基于环境模型的多层级顺序通信模型，可以有效地基于智能体间通信解决误协调问题。顺序通信把智能体分为不同决策层级 (上层智能体在下层智能体之前

做出决策)，下层智能体可以基于上层智能体的动作对自身决策进行更好的调整，从而解决误协调问题。在顺序通信的框架下，智能体相对于其他方法能够获得额外的正确的动作信息，从而获得更好的性能。本文还证明了顺序通信学习到的策略可以保证单调改进和收敛。

再次，为了保证智能体在不同场景的策略泛化能力，智能体需要更好的语言理解能力，以便利用人类对环境的先验知识来理解不同的环境。本文提出了一种多智能体语言定位框架–实体划分器。其通过关联环境实体来理解给定的描述当前环境运行规则文本手册的含义。基于所理解的文本内容，智能体将笼统的指令分解为多个子目标并仅基于合理的“自身”子目标来作出决策。此外，智能体通过“其他”子目标来进行智能体建模，从而评估其他智能体的策略。本文提出的框架是在多智能体环境中通过语言定位将策略泛化到未见过环境的首次尝试。

最后，为了提高智能体的学习能力，使其能够在开放世界中像人类一样不断学习并积累环境知识，加深对开放世界的理解，本文提出了一种对强化学习友好的视觉语言模型 CLIP4MC。该模型通过度量当前视频帧输入和语言提示语的相关性，生成相应的奖励信号，使智能体能够在开放式世界不断学习。因为不再需要针对特定任务进行奖励设计，所以奖励生成器为智能体在开放世界中的学习提供了可能性。

本文系统研究了多种类型的智能体通信方法，并验证了它们对智能体性能的各种提升。这为该领域未来的研究提供了强有力的技术支持。未来的研究可以进一步探索如何将不同类型的通信方法整合到一个统一的框架中，以便让智能体更高效地完成更复杂的任务。

关键词：多智能体强化学习，多智能体通信，基于语言的强化学习。