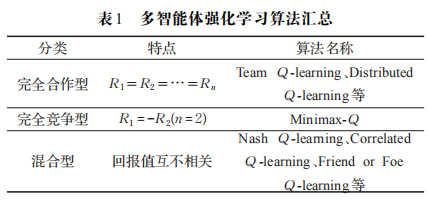
<1 多智能体深度强化学习研究综述_孙彧.pdf>

MDRL按照无关联型、通信规则型、互相合作型和建模学习型4种分类方式阐述了现有的经典算法；DRL算法的实际应用，现有的测试平台；multiagent DRL在理论、算法、应用方面面临的挑战和未来的发展方向。



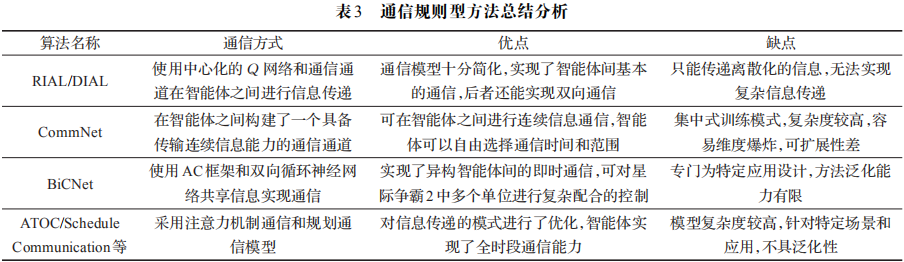
MARL 的算法根据其回报函数的不同可以分为完 全 合 作 型(Fully Cooperative)、完 全 竞 争 型(Fully Competitive)和混合型(Mixed)三种任务类型，完全 合作型算法中智能体的回报函数是相同的，即 *R*1 = *R*2 = ⋯ = *Rn* ，表示所有智能体都在为实现共同的目标而努力；完全竞争型算法中智能体的回报函数是相反的，环境通常存在两个完全敌对的智能体，它们遵循SG(随机博弈)原则，即 *R*1 = -*R*2，智能体的目标是最大化自身的回报，同时尽可能最小化对方回报；混合型任务中智能体的回报函数并无确定性正负关系，该模型适合自利型（Self-interested）智能体，一般来说此类任务的求解大都与博弈论中均衡解的概念相关，即当环境中的一个状态存在多个均衡时，智能体需要一致选择同一个均衡。该类算法主要面向静态任务。

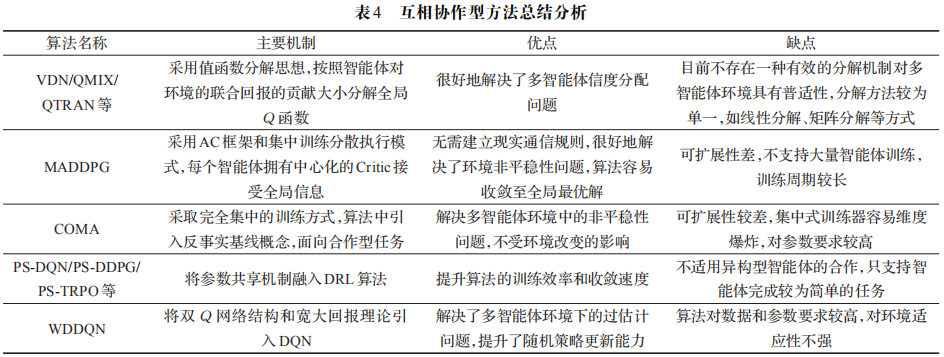
MARL遵循随机博弈(SG)过程，SG具体定义。图二MARL基本框架。

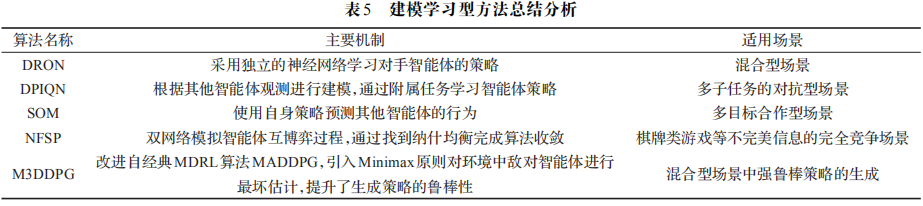
MARLA缺点：探索利用矛盾（Explore and Exploit）和维度灾难（Curse of Dimensionality）；多智能体环境非平稳性（None stationary）问题；多智能体信度分配（Multiagent Credit Assignment）问题；最优均衡解问题；学习目标选择问题等

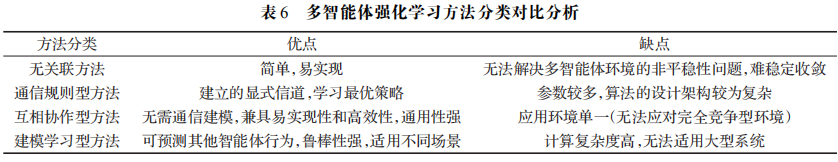
MARL和DL结合产生MDRL，按照智能体之间的通联方式，大致将当前的 MDRL 方法分为：无关联型、通信规则型、互相协作型和建模学习型4大类。

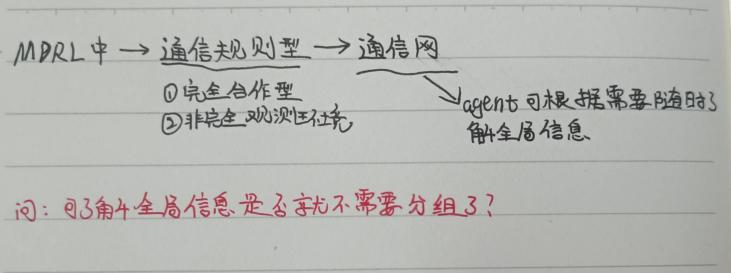




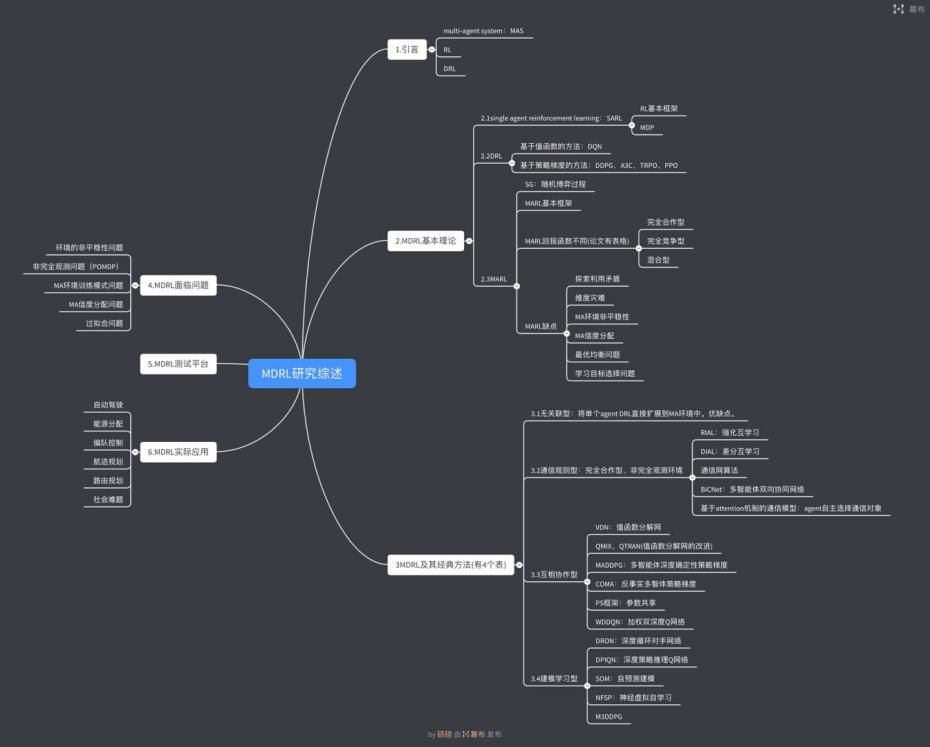








POMDP在这篇中有具体定义。



Multi-Agent Reinforcement Learning: a critical survey<2 MARL a critical survey.pdf>

博弈论，SGs

