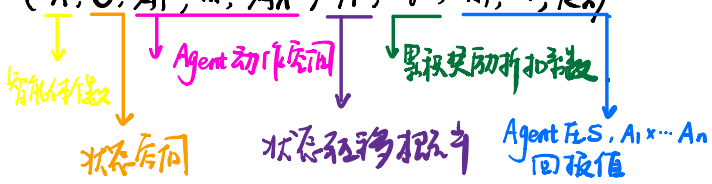
MARL 随机博弈过程描述：

分解为 State Game 策略组合

与环境交互更新每一个状态阶段博弈中 Q 值

多环境状态过程：

$(n, S, A_1, \cdots, A_n, Tr, \gamma, R_1, \cdots, R_n)$



$Tr: S \times A_1 \times \cdots \times A_n \times S' \to [0,1]$

$R_i: S \times A_1 \times \cdots \times A_n \times S' \to \mathbb{R}$

MARL 目标：

每个状态下的纳什均衡策略（最优策略）。

↓

$V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i^*, \cdots, \pi_n^*) \geq V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i \cdots, \pi_n^*)$

$\forall \pi_i \in \Pi_i \quad i = 1, \cdots, n$

所有智能体的联合策略

在纳什均衡处，对于 All Agent 都不能改变自己策略

从而获得更大的奖励。

最优策略数学描述：

$(\pi_1^*, \pi_2^*, \cdots, \pi_i^*, \cdots, \pi_n^*)$ $i=1,\cdots,n$

对于 $\forall s \in S$ 满足

① 对状态价值函数 $V_i(s)$

$$V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i^*, \cdots, \pi_n^*) \geq V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i, \cdots, \pi_n^*)$$

$\forall \pi_i \in \Pi_i \quad i=1,\cdots,n$

② 对状态-动作价值函数 $Q_i(s, a_1, \cdots, a_n)$

$$\sum_{a_1 \cdots a_n \in A_1 \times \cdots \times A_n} Q_i^*(s, a_1, \cdots, a_n) \pi_1^*(s, a_1) \cdots \pi_i^*(s, a_i) \cdots \pi_n^*(s, a_n) \geq \sum_{a_1 \cdots a_n \in A_1 \times \cdots \times A_n} Q_i^*(s, a_1, \cdots, a_n) \pi_1^*(s, a_1) \cdots \pi_i(s, a_i) \cdots \pi_n^*(s, a_n)$$

Bellman 公式

$$Q_i^*(s, a_1, \cdots, a_n) = \sum_{s' \in S} Tr(s, a_1, \cdots, a_n, s')\left[R_i(s, a_1, \cdots, a_n, s') + \gamma V_i^*(s')\right]$$

$$V_i^*(s) = \sum_{a_1 \cdots a_n \in A_1 \times \cdots \times A_n} Q_i^*(s, a_1, \cdots, a_n) \pi_1^*(s, a_1) \cdots \pi_n^*(s, a_n)$$

如何得到最优策略呢？

MARL 算法层面

MARL 随机博弈过程

从 Reward 角度进一步分类：

① fully cooperative $(R_1 = \cdots = R_N)$

② fully competitive $(n=2 \quad R_1 = -R_2)$

③ Mixed

针对不同博弈过程，

设计 MARL 回报函数：

① 共同目标，可以共同最大化。

$\Longrightarrow$ i) 自身学习动态行为的稳定性

ii) 对其他代理动态行为的适应

{ Stability：趋于平稳的策略。 必收性

Adaptation：其他 Agent 改变策略时 保持或改善性能。 合理性

↓

自我play，最优策略， Nash equilibrium

所收敛到的什么均衡 与 动态SG的性能之间的联系 尚不清晰。

聚焦稳定性的算法，一般认为 Agents 都正时。

适应能力算法，考虑其它 Agents 的行为。

又考虑稳定性，不考虑收敛性，对其他智能体的跟踪。

算法：

Task type → ↓ Agent awareness	Cooperative	Competitive	Mixed
Independent	coordination-free	opponent-independent	agent-independent
Tracking	coordination-based	—	agent-tracking
Aware	indirect coordination	opponent-aware	agent-aware

事件类型，　　对其它 Agent感知.

Task type / Agent awareness	Cooperative	Competitive	Mixed
Independent Agent 之间相互独立 考虑其它Agent 行为			