一、问题

1、现有 Multiagent 架构缺点

1）、分布式独立结构

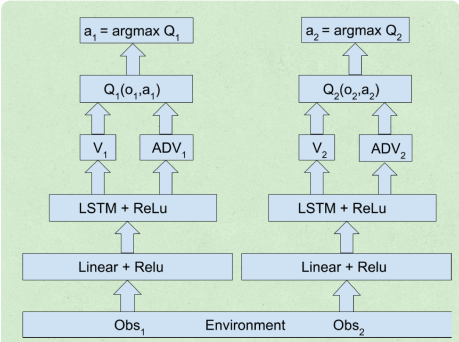


Figure 14: Independent Agents Architecture

① Observation 部分可观 PO-MDP

Policy π_i 对每个是单独Policy

合作，部分场景都完成.

② non-Stationary Env

2）结合中心式架构

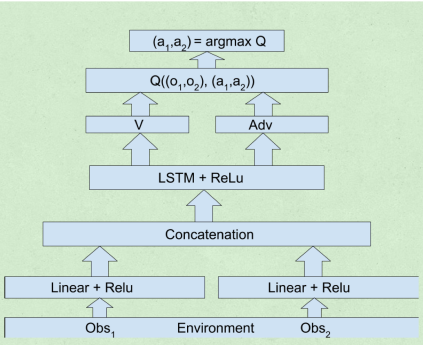


Figure 20: Combinatorially Centralized Architecture

① credit assignment

Team reward

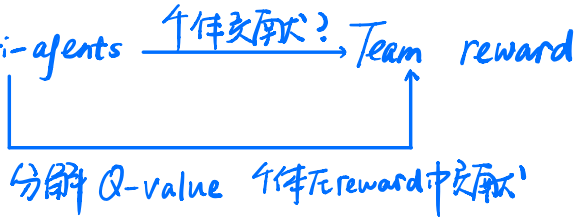
谁的贡献大？

② Joint action space

还有 agent 指数增加

Paper：每个 Agent 在 Team reward 中贡献大小？

Idea



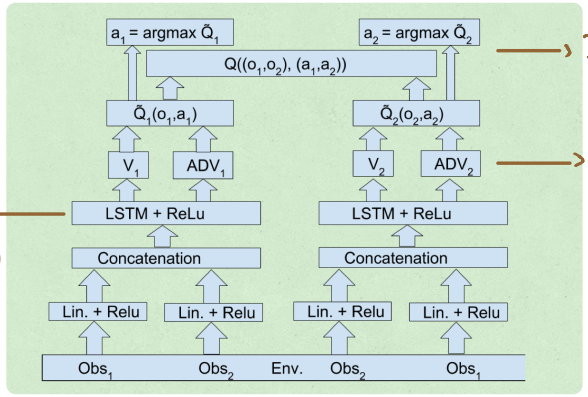
二、A Deep-RL Architecture for Coop-MARL

the Identified issues ⟵ value-decomposition

$Q((h^1, h^2, \dots, h^d), (a^1, a^2 \dots, a^d)) \approx \sum_{i=1}^{d} \tilde{Q}_i(h^i, a^i)$ 线性分解，合作场景

局部observation 用 history代替 $h_t = a_1 o_1 r_1, \dots, a_{t-1} o_{t-1} r_{t-1}$ 得到更优策略

论文中没有线性分解证明，大量实验验证，假设。



DRN (Target & ER)

dueling DQN 加速学习速度

LSTM：history 减轻局部观测的影响

三、实验：3 Task

Paper 没有证明 Value Decomposed 分解，只是用大量实验证明。

1、场景

① Switch：长廊环境到达各自的目的地。

② Fetch：取放货，另一个 Agent 接力（类似栅格环境吗）

③ Checkers：

	Apple	lemon
Agent1	10	-10
Agent2	1	-1

让 Agent1 拿 Apple，Agent2 拿 Lemon 得分最高。

a、算法比较

Agent	V.	S.	Id	L.	H.	C.
1						
2	✓					
3	✓	✓				
4	✓	✓	✓			
5	✓	✓	✓	✓		
6	✓	✓	✓		✓	
7	✓	✓	✓	✓	✓	
8	✓					✓
9						✓

Table 1: Agent architectures. V is value decomposition, S means shared weights and an invariant network, Id means role info was provided, L stands for lower-level communication, H for higher-level communication and C for centralization. These architectures were selected to show the advantages of the independent agent with value-decomposition and to study the benefits of additional enhancements added in a logical sequence.

不同的方法技巧

VD方法好 Individual Agent方法

各有VD好于 Centralized方法

应用到多种架构分析。

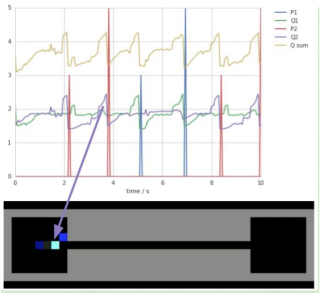


Figure 6: The learned Q-decomposition in Fetch. The plot shows the total Q-function (yellow), the value of agent 1 (green), the value of agent 2 (purple), rewards from agent 1 (blue) events and agent 2 (red). Highlighted is a situation in which agent 2's Q-function spikes (purple line), anticipating reward for an imminent drop-off. The other agent's Q-function (green) remains relatively flat.

Fetch场景，说明 Value-decomposed 的效，有效高 Q-value 的 Agent 对 Team Reward 贡献大，

Paper 思路清晰，以实验分析为主导，对比了所有子的的架构，综合分析 Value-Decomposed 的作用效果。

论文不足之处：① Team 数模 分解特性有效性。

② Value-decomposed 特性局限。

—— 场景

特性 Decomposed 效果不如非特性 Decomposed