WDS

DataScience@Web

稀疏奖励——好奇心驱动

汇报人：陈海燕

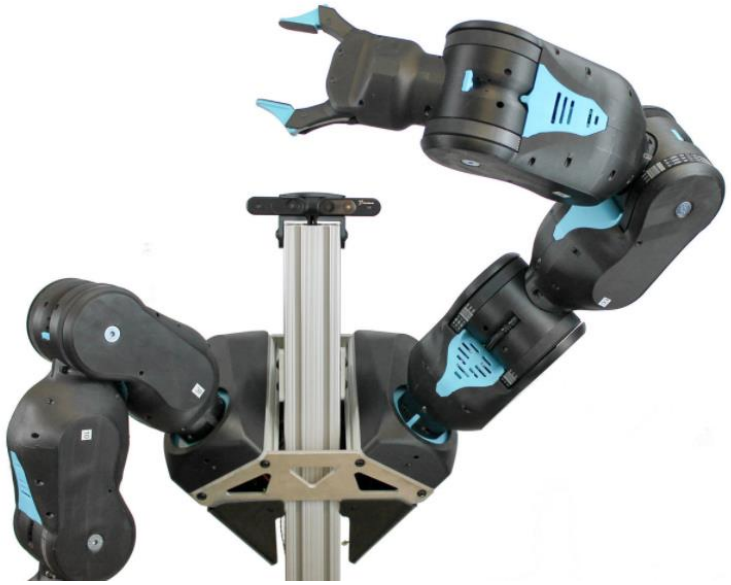
2021.07.27

2021WDS暑期讨论班

目录

- 稀疏奖励问题(Sparse Reward)
- 人为设计奖励(Reward Shaping)
- 好奇心驱动(Curiosity Driven)
- 总结

稀疏奖励



- 拿起螺丝刀—对准螺丝钉—拧螺丝刀

人为设计奖励(Reward Shaping)

Take “Play”,
$r_{t+1} = 1, r_{t+100} = -100$

Take “Study”,
$r_{t+1} = -1, r_{t+100} = 100$

人为设计奖励(Reward Shaping)

Take “Play”,

$r_{t+1} = 1, r_{t+100} = -100$

Take “Study”,

$\cancel{r_{t+1} = 1}, r_{t+100} = 100$

$r_{t+1} = 1$

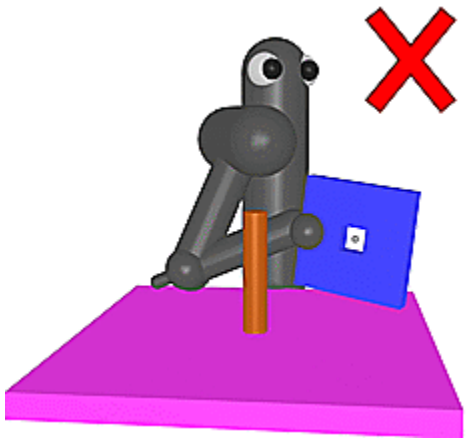
人为设计奖励(Reward Shaping)

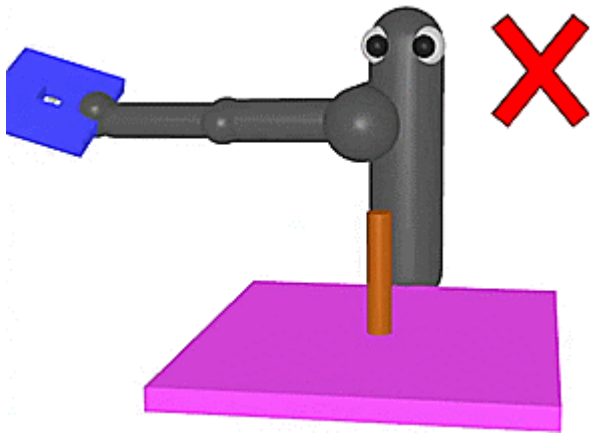
VizDoom

Parameters	Description	FlatMap	CIGTrack1
living	Penalize agent who just lives	-0.008 / action	
health_loss	Penalize health decrement	-0.05 / unit	
ammo_loss	Penalize ammunition decrement	-0.04 / unit	
health_pickup	Reward for medkit pickup	0.04 / unit	
ammo_pickup	Reward for ammunition pickup	0.15 / unit	
dist_penalty	Penalize the agent when it stays	-0.03 / action	
dist_reward	Reward the agent when it moves	9e-5 / unit distance	

- 10:1

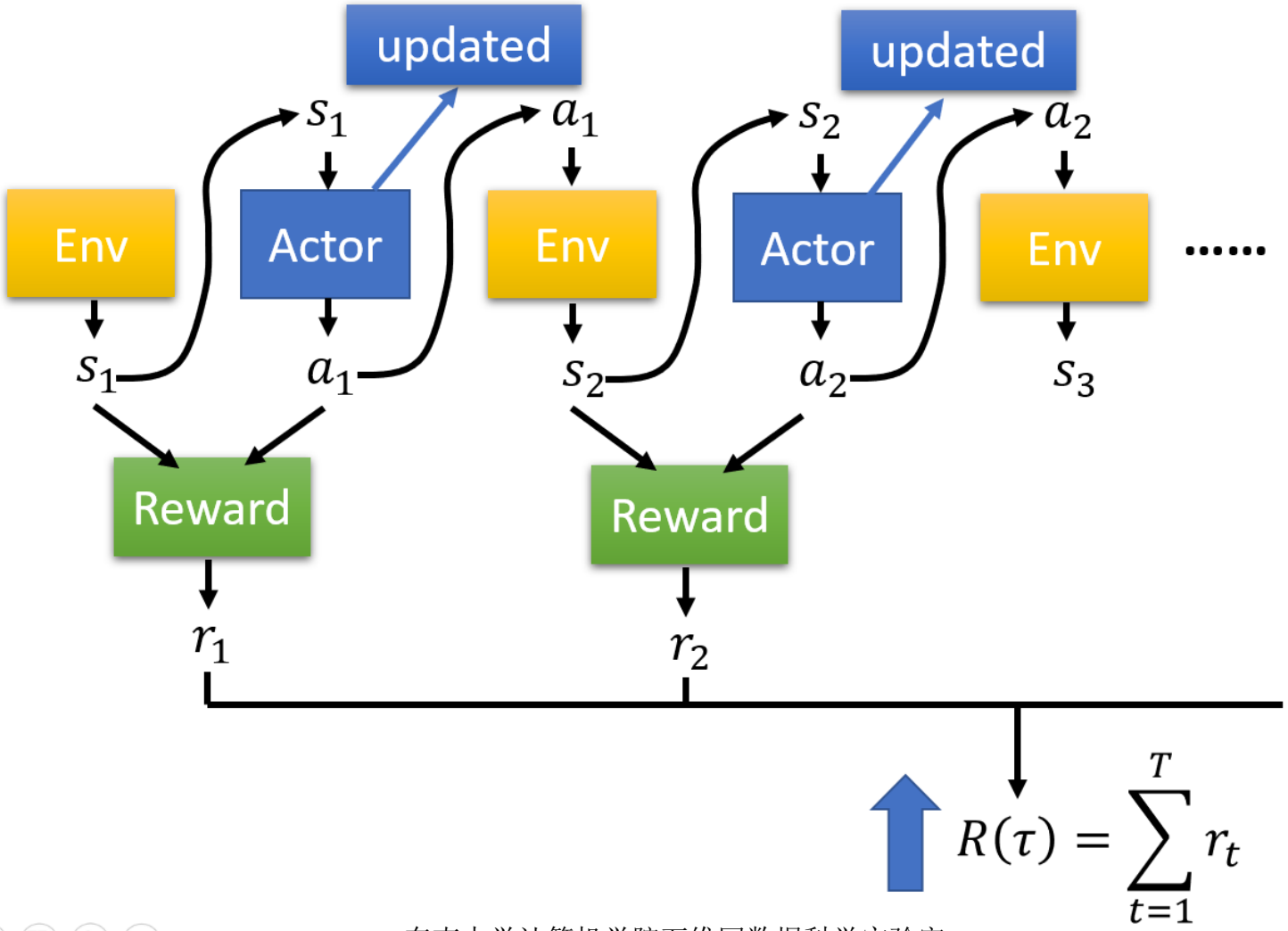
人为设计奖励(Reward Shaping)



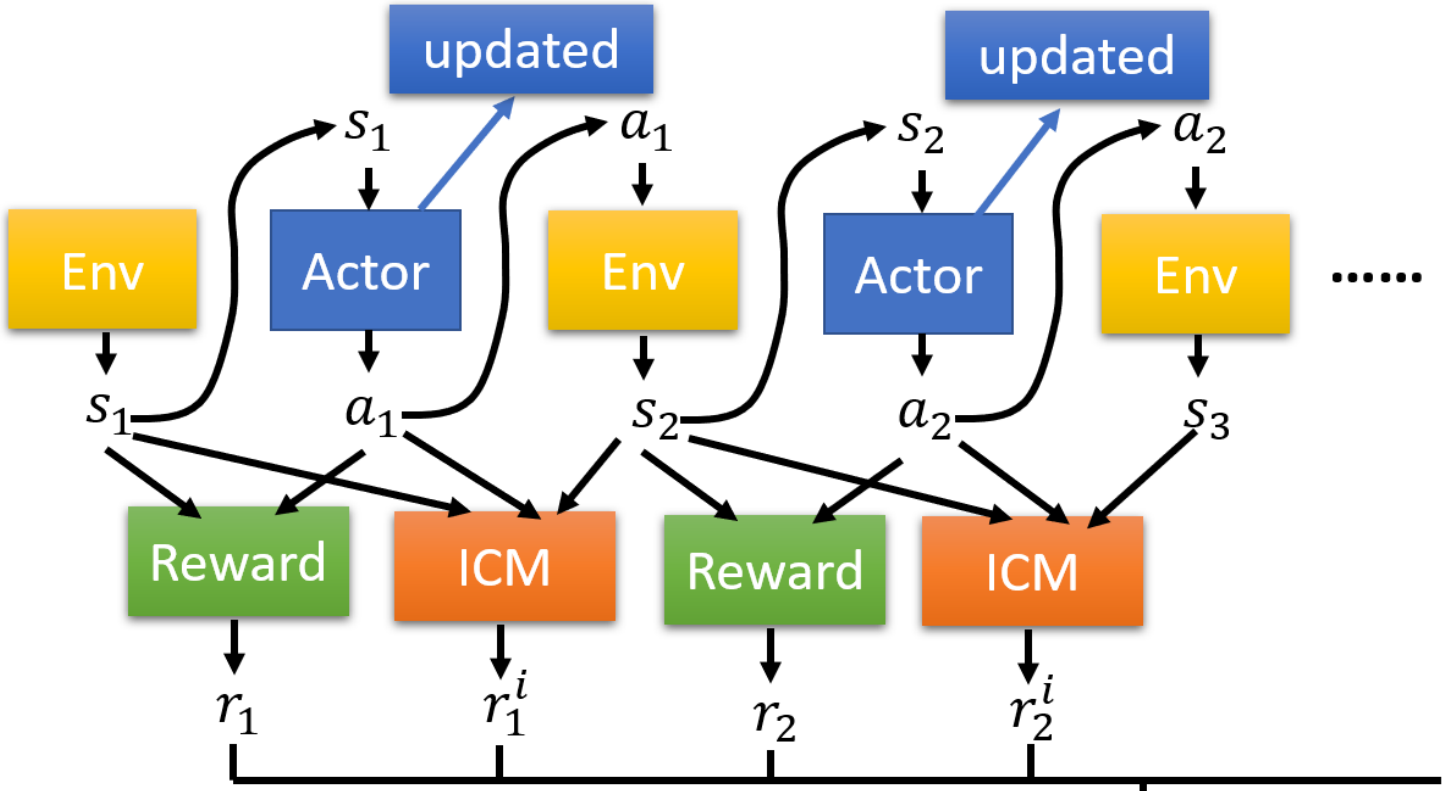


靠近柱子就获得奖励

好奇心驱动(Curiosity Driven)



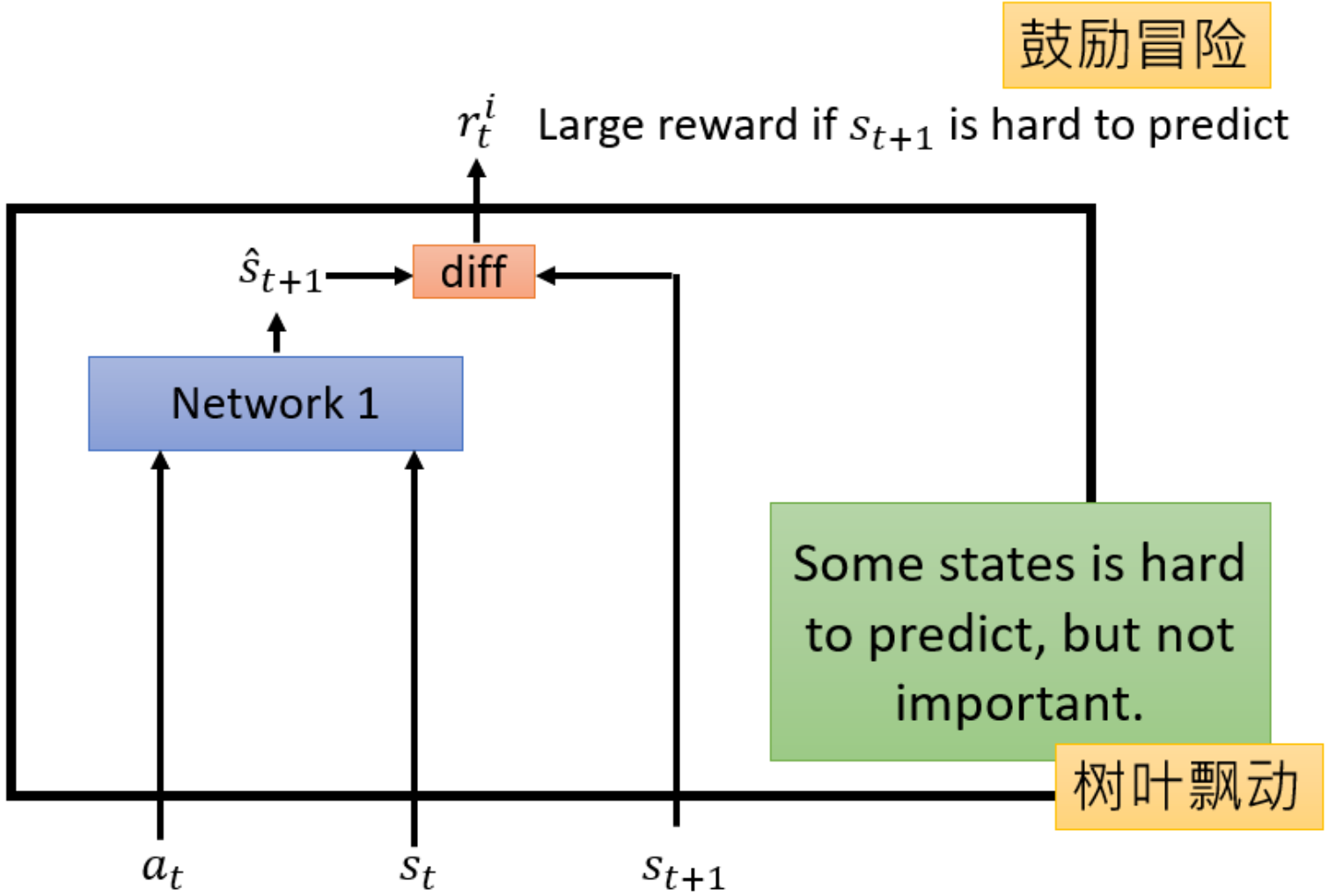
好奇心驱动(Curiosity Driven)



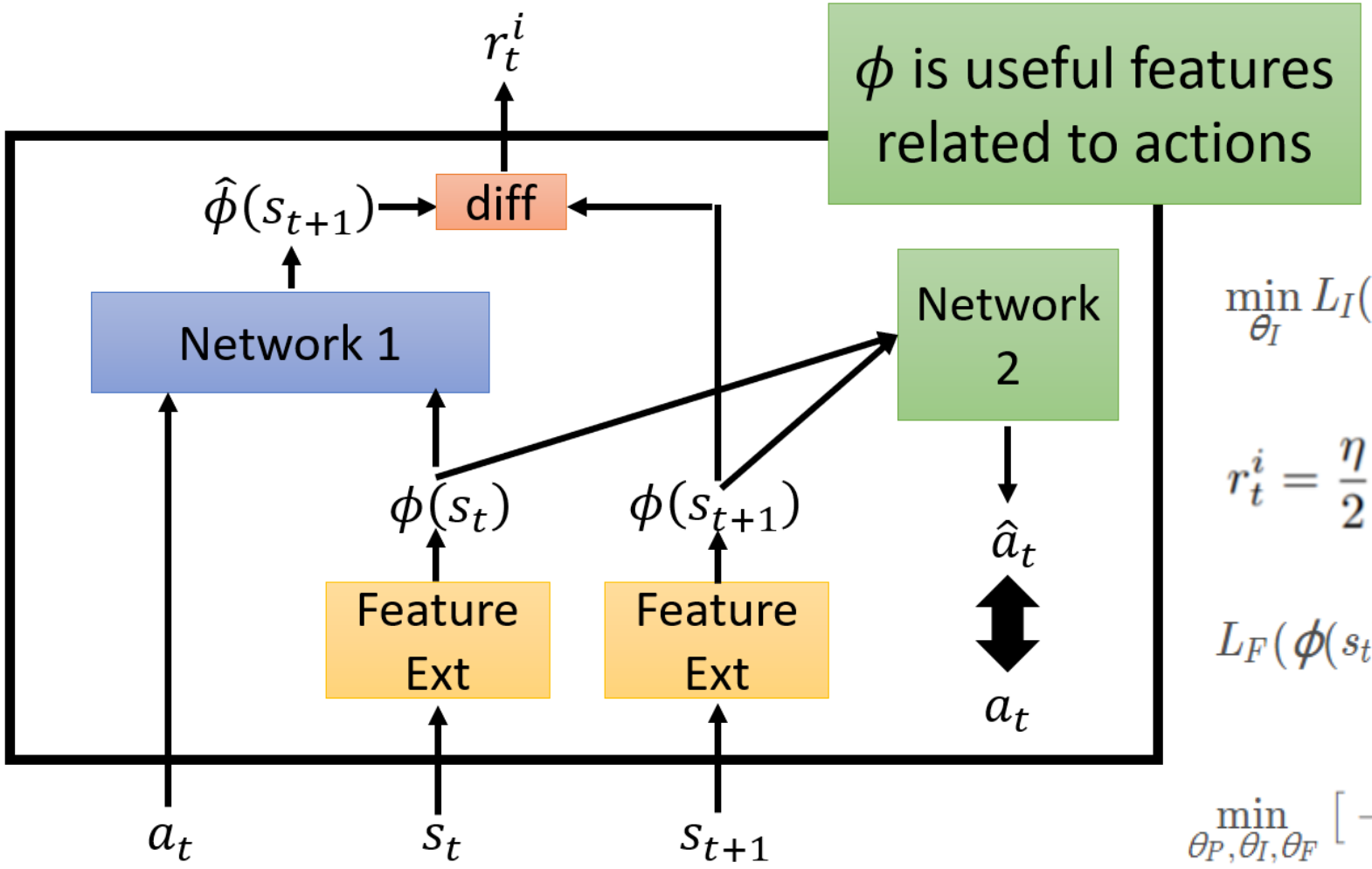
ICM = intrinsic curiosity module

$$R(\tau) = \sum_{t=1}^{T} r_t + r_t^i$$

好奇心驱动(Curiosity Driven)



好奇心驱动(Curiosity Driven)



$$\hat{\phi}(s_{t+1}) = \infty \quad ?$$

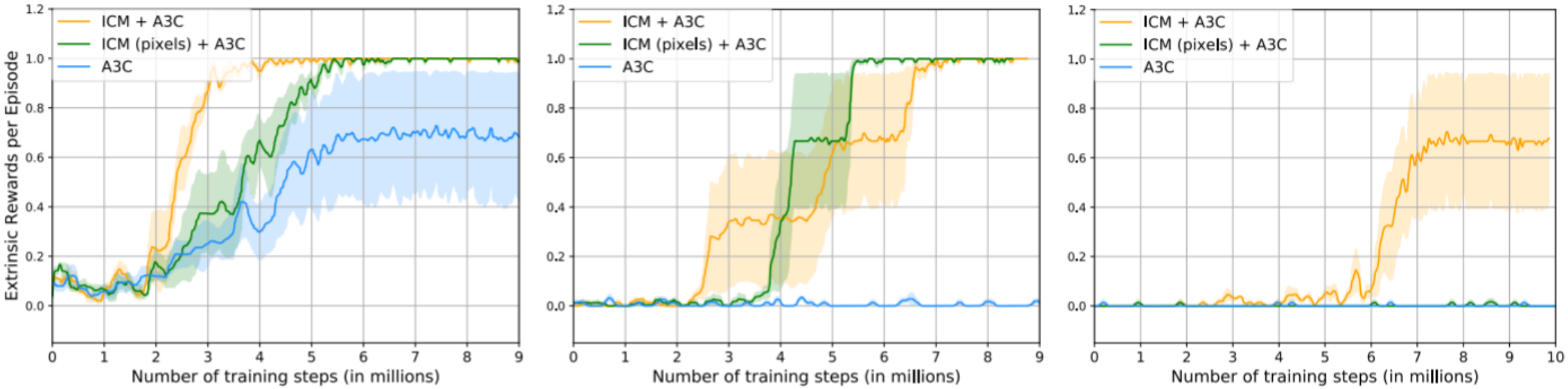
$$\min_{\theta_I} L_I(\hat{a}_t, a_t)$$

$$r_t^i = \frac{\eta}{2}\|\hat{\phi}(s_{t+1}) - \phi(s_{t+1})\|_2^2$$

$$L_F(\phi(s_t), \hat{\phi}(s_{t+1})) = \frac{1}{2}\|\hat{\phi}(s_{t+1}) - \phi(s_{t+1})\|_2^2$$

$$\min_{\theta_P, \theta_I, \theta_F} \left[-\lambda E_{\pi(s_t;\theta_P)}\left[\sum_t r_t\right] + (1-\beta)L_I + \beta L_F \right]$$

好奇心驱动(Curiosity Driven)



(a) “dense reward” setting (b) “sparse reward” setting (c) “very sparse reward” setting

好奇心驱动(Curiosity Driven)

- 适用情况
  - 密集奖励 ×
  - 稀疏奖励 √
    - 简单奖励（胜/负）
- 使用方法
  - 配合其他强化学习模型

总结

- 人为设计奖励(Reward Shaping)
  - 需要专业知识
- 好奇心驱动(Curiosity Driven)
  - 更加通用的方法
- 课程式学习(Curriculum Learning)
- 层次强化学习(Hierarchical RL)

Thanks!