彪哥带你学强化学习

12.A2C算法

DEEPLY UNDERSTAND REINFORCEMENT LEARNING

讲师：韩路彪

A2C算法

A2C含义：Advantage actor critic，优势演员评论家

思路：在价值基础上减去基线

小王用1万块钱进货，1个月盈利100，小张用2万进货，1个月盈利120，谁盈利多？

经济增加值（EVA）：EVA = 净利润 - 资本 × 资本成本率

小王：EVA = 100 - 10000*3%/12 = 75
小张：EVA = 120 - 20000*3%/12 = 70

小李用1万块钱进货，1个月盈利100，小赵用1万进货，1个月盈利120

用1万块钱进货，1个月平均盈利110
小李：-10，小赵：10

A2C算法

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}[(\sum_{t=0}^{\infty} \nabla_\theta \pi(a_t|s_t,\theta) q_\pi(s_t,a_t))]$$

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}[(\sum_{t=0}^{\infty} \nabla_\theta \pi(a_t|s_t,\theta) A_\pi(s_t,a_t))]$$

优势函数 $A_\pi(s_t,a_t) = Q_\pi(s_t,a_t) - V_\pi(s_t)$