

5月第二次周报

时间：2019.5.2 - 2019.5.8

内容：帮学姐跑 NYT 和 SKE数据集、学习强化学习MDP有关知识、了解了一下pytorch

1. MDP：马尔可夫决策过程

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

其中的 R_{t+1} 是t时刻采取action之后立刻获得的reward

- Bellman方程

$$v(s) = R_s + \gamma \sum_{s' \in S} P_{ss'} v(s')$$

- Policy

$$\pi(a|s) = P[A_t = a | S_t = s]$$

- Value Function

$$v_{\pi}(s) = E_{\pi}[G_t | S_t = s]$$

- Action-Value Function

$$q_{\pi}(s, a) = E_{\pi}[G_t | S_t = s, A_t = a]$$

- V函数和Q函数的关系

$$\begin{aligned} v_{\pi}(s) &= E_{\pi}[R_t + \gamma v_{\pi}(S_{t+1}) | S_t = s] \\ q_{\pi}(s, a) &= E_{\pi}[R_{t+1} + \gamma q_{\pi}(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \end{aligned}$$

通过Q函数来求V函数：

$$v_{\pi}(s) = \sum_{a \in A} \pi(a|s) q_{\pi}(s, a)$$

通过V函数来求Q函数：

$$q_{\pi}(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_{\pi}(s')$$

2. 在学习了RL有关的知识后又重新读了学姐代码的HRL论文

3. 学习有关pytorch的内容

- 有关Tensor的内容
- Autograd：自动求导机制
- 学习用torch.nn构建神经网络