人工智能原理1

2024年6月18日　　22:32

----------决策与强化学习----------

1.简单决策

（1）基本概念：

- 效用函数U，期望效用EU，最大期望效用MEU。
- 理性偏好

（2）决策网络：

- 机会节点，效用节点，决策节点
- 信息价值，完美信息价值VPI（不具有可加性）

2.复杂决策

（1）马尔科夫

- 从s到s’的概率只取决于s，而不取决于以前状态的历史。

（2）序贯决策：效用函数依赖一系列状态和动作

（3）MDP：具有马尔科夫转移模型和加性奖励的序贯决策问题为MDP

（4）贝尔曼方程

（5）区分状态效用与动作效用，动作效用函数又被称为Q函数

（6）MDP的表示

（7）求解MDP

- 价值迭代
- 策略迭代
- 线性规划
- 蒙特卡洛规划(在线算法)

（8）老虎机问题

- 基廷斯指数
- 重启MDP

3.强化学习

（1）一些分类

- 基于模型的强化学习：学习转移模型和奖励函数
- 无模型的强化学习：Q学习，策略探索
- 被动强化学习：策略固定
- 主动强化学习：策略不固定，主要问题是探索

（2）被动强化学习

- 直接效用估计：对多组实验结果，直接计算状态效用的平均值
- ADP：求模型的P，R，用来更新效用
- TD学习：关于误差的学习，是对ADP的近似，每次更新当前状态效用

值，是一种对ADP的近似，其本质是对误差的学习。

（3）主动强化学习

- 无固定策略，智能体可以自主决定采取什么行动。
- 从ADP入手，先学习完整的转移模型，再通过探索函数，决定当前应该执行最佳动作还是随机探索。

（4）强化泛化

- 用线性函数或非线性函数近似效用函数，其中每个变量值是状态抽象出来的特征值
- 学习方式可采用近似直接效用估计，类似神经网络朝误差最小的方向对权重进行参数更新
- 也可采用近似TD进行学习，或用深度强化学习。