强化学习与深度强化学习

【课程时长】

3 天（6 小时/天）

【课程简介】

强化学习是当前最热门的研究方向之一，广泛应用于机器人学、电子竞技等领域。本课程系统性的介绍了强化学习（深度强化学习）的基本理论和关键算法，包括：马尔科夫决策过程、动态规划法、蒙特卡罗法、时间差分法、值函数逼近法，策略梯度法等；以及该领域的最新前沿发展，包括：DQN 及其变种、信赖域系方法、Actor-Critic 类方法、多 Agent 深度强化学习等；同时也介绍大量的实际案例，包括深度强化学习中最著名的工程应用：Alpha Go。

【课程对象】

计算机相关专业本科；或理工科本科，具备初步的 IT 基础知识的人员

第一天 强化学习

第一课 强化学习综述

1.强化学习要解决的问题
2.强化学习方法的分类
3.强化学习方法的发展趋势
4.环境搭建实验（Gym，TensorFlow 等）
5.Gym 环境的基本使用方法

第二课 马尔科夫决策过程

1.基本概念：马尔科夫性、马尔科夫过程、马尔科夫决策过程
2.MDP 基本元素:策略、回报、值函数、状态行为值函数
3.贝尔曼方程
4.最优策略
案例：构建机器人找金币和迷宫的环境

第三课 基于模型的动态规划方法

1.动态规划概念介绍
2.策略评估过程介绍
3.策略改进方法介绍
4.策略迭代和值迭代
案例：实现基于模型的强化学习算法

第四课 蒙特卡罗方法

1.蒙特卡罗策略评估
2.蒙特卡罗策略改进
3.基于蒙特卡罗的强化学习
4.同策略和异策略
案例：利用蒙特卡罗方法实现机器人找金币和迷宫

第五课 时序差分方法

1.DP，MC 和 TD 方法比较
2.MC 和 TD 方法偏差与方差平衡
3.同策略 TD 方法：Sarsa 方法
4.异策略 TD 方法：Q-learning 方法
案例：Q-learning 和 Sarsa 的实现

第二天 从强化学习到深度强化学习

第一课 基于值函数逼近方法（强化学习）

1.维数灾难与表格型强化学习
2.值函数的参数化表示
3.值函数的估计过程
4.常用的基函数

第二课 基于值函数逼近方法（深度学习与强化学习的结合）

1.简单提一下深度学习
2.深度学习与强化学习的结合
3.DQN 方法介绍
4.DQN 变种：Double DQN, Prioritized Replay, Dueling Network
案例：用 DQN 玩游戏——flappy bird

第三课 策略梯度方法（强化学习）

1.策略梯度方法介绍
2.常见的策略表示

3.常见的减小方差的方法:引入基函数法，修改估计值函数法
案例：利用 gym 和 tensorflow 实现小车倒立摆系统等

第四课 Alpha Go（深度学习与强化学习的结合）

1.MCTS
2.策略网络与价值网络
3.Alpha Go 的完整架构

第五课 GAN（深度学习）

1.VAE 与基本 GAN
2.DCGAN
3.WGAN
案例：生成手写数字的 GAN

第三天 深度强化学习进阶

第一课 AC 类方法-1

1. PG 的问题与 AC 的思路
2. AC 类方法的发展历程
3. Actor-Critic 基本原理

第二课 AC 类方法-2

1. DPG 方法
2. DDPG 方法
3. A3C 方法
案例：AC 类方法的案例

第三课 信赖域系方法-1

1.信赖域系方法背景
2.信赖域系方法发展路线图
3.TRPO 方法
案例：TRPO 方法的案例

第四课 信赖域系方法-2

1.PPO 方法
2.DPPO 方法简介
3.ACER 方法
案例：PPO 方法的案例

第五课 多 Agent 强化学习

1.矩阵博弈
2.纳什均衡
3.多人随机博弈学习
4.完全合作、完全竞争与混合任务
5.MADDPG
案例：MADDPG 的案例等