教学大纲

课程基本情况
强化学习基础，08617190，2023-2024 学年第二学期
3 学分，研究生选修课
开课专业：力学（工业与系统工程）
前置基础课程：概率论、最优化理论与算法或对等课程
时间：周一第 7-9 节（3:10-6:00 pm）
地点：三教 305

授课教师
尤鹏程，助理教授，pcyou@pku.edu.cn
办公室：王克桢 1003
答疑：周五下午（邮件预约），王克桢 1003

助教
都檬阁，dumengge@stu.pku.edu.cn
答疑：周二下午（邮件预约），王克桢 1008

课程资源
- 参考教材
  - Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction* (2nd Edition) ISBN-13: 978-0262039246.
    Online（http://www.incompleteideas.net/book/the-book-2nd.html）
- 课外推荐阅读
  - Csaba Szepesvári, *Algorithms for Reinforcement Learning*
    Online（https://www.ualberta.ca/~szepesva/papers/RLAlgsInMDPs.pdf）
  - Sean Meyn, *Control Systems and Reinforcement Learning*
    Online（https://meyn.ece.ufl.edu/control-systems-and-reinforcement-learning/）
  - Dimitri P. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. I (4th Ed.) and Vol. II (4th Ed. Approx. Dynamic Programming)
  - Dimitri P. Bertsekas, *Reinforcement Learning and Optimal Control* (1st Edition)
  - Torr Lattimore and Csaba Szepesvári, *Bandits Algorithms*
    Online（https://tor-lattimore.com/downloads/book/book.pdf）
- 北大教学网 https://course.pku.edu.cn/
  - 课程录像
  - 答疑讨论
  - 作业发布与提交
- 实验计算资源：咨询助教

课程信息

- 简介

  强化学习广泛应用于跨学科的研究和工作中，然而现成的算法和代码往往容易让人忽视其严谨的定义和性质。该课程将重点探究强化学习方法的底层数学基础，讨论马尔科夫决策过程模型的有效性，推导与证明经典算法的基本原理，并结合作业与编程应用加深学生对相关理论知识的理解。主要的内容分为基于模型的方法，例如确定性与随机性的动态规划，以及无模型的方法，即广义的强化学习方法，包括 Monte Carlo、Temporal Differences、n-step bootstrapping 等 on-policy 和 off-policy 表格型法，及其拓展至广泛实际应用的近似方法：值函数近似、策略近似等。课程可能会在学期间穿插安排专题讲座，邀请业界的专家学者分享强化学习研究和应用的前沿进展；期末将以团队项目形式考核，旨在鼓励学生利用课堂上所学的知识和技巧对具体问题进行建模和求解。

- 主要内容
  - 强化学习简介（3 学时）

    背景、应用、问题定义、问题建模
  - 马尔科夫决策过程（6 学时）

    状态、动作、环境模型、马尔科夫性质、有限/无限时域、策略、值函数、动态规划建模与算法
  - 模型驱动方法（12 学时）

    a. 基础：片段型/连续型任务、马尔科夫策略与静态策略的充分性、算子理论

    b. 策略评估（预测）：Bellman 方程、迭代策略评估、收缩理论

    c. 控制：最优性原理、贪婪策略、策略改善定理、策略迭代、价值迭代
  - 无模型强化学习方法（18 学时）

    a. 基础:多臂老虎机问题与算法、随机近似算法原理

    b. 表格型方法

        i. 蒙特卡洛预测、TD(0)预测、n-step bootstrapping

        ii. 蒙特卡洛控制、TD 控制——SARSA 和 Q-learning

    c. 近似方法

        i. 值函数近似——线性模型和深度神经网络模型

        ii. 资格迹

        iii. 策略近似——策略梯度方法
  - 期末项目展示（3 学时）
- 课程考核方式（不低于 65%的 A）
  - 作业（20%）：四次
  - 编程实验（30%）：六次
  - 期中考试（15%）
  - 期末项目（35%）
    - 项目计划报告（10%）
    - 展示（15%）
    - 项目结题报告（10%）

    - 课堂参与（不超过 5%的加分）
- 关键日期
    - 期中考试：第九周，4 月 15 日当堂
    - 期末项目
        - 项目计划报告：第十周，4 月 22 日 11:59 pm
        - 展示：第十六周，6 月 3 日当堂
        - 项目结题报告：6 月 10 日 11:59 pm
- 作业、编程实验政策
    - 发布、提交均在北大教学网
    - 一般周一发布，下个周一晚上 11:59 pm 截止提交
    - 迟交即无效，但每位学生有两次推迟提交的机会（多一周时间，截止次周周一晚上 11:59 pm），自动计算，不需汇报
    - 允许讨论，但需独立完成，如发现抄袭迹象将取消对应成绩，屡教不改将记录上报
- 其它
    - 病假、事假：欢迎提前邮件授课教师告知情况
    - 课程反馈：如对课程进度、强度有任何疑问或建议，欢迎邮件授课教师