《强化学习》教学大纲

《强化学习01》

一、课程基本信息

课程名称/英文名称	强化学习/Reinforcement Learning	课程代码	SI252
课程层次	研究生课程	学分/学时	4/64
先修课程	无	授课语言	双语
开课单位	信息科学与技术学院	课程负责人	邵子瑜

二、课程简介

“强化学习”是人工智能相关课程群中的一门重要的学科基础课，是研究强化学习的数学基础，算法设计以及应用的学科。本课程通过讲授强化学习的基础知识以及各类常用的强化学习算法， 使学生获得强化学习的基本理论、基本方法和基本技能;了解强化学习发展的概况以及前沿研究方向，初步掌握强化学习算法的分析、设计方法，为未来从事人工智能相关领域工作打下扎实基础。

为了保持教学效果，选课前会开展背景知识测试，未达到基础标准的同学建议学好预修课程再来选课。

三、课程教学目标

知识认知能力培养：能掌握强化学习相关的基本知识，包括多臂老虎机，马尔科夫决策过程，动态规划以及经典的强化学习算法(如蒙特卡洛算法，时序差分算法等)，并对各类深度强化学习算法有所了解。

综合素质能力培养：能够流利的表达，思路清晰富有逻辑;具备科学精神和工程师的基本素养，具有批判性思维和创新思维；能进行团队协作，具备合作精神和人际沟通能力。

四、课程教学方法

课堂讲授与讨论:强化学习知识点基本以课堂讲授为主，在讲解基本知识点的基础上，关注课程重点难点内容的讲授，采用启发式教学方法，引入丰富多彩的实际应用案例，激发学生的兴趣热情，引导学生对经典案例问题展开思考和讨论，使学生能够从工程思想出发，阐释相应的物理图像，

运用数学工具进行建模和分析，并使用计算机仿真和数值计算来验证解决方案的有效性，从而解决实际领域的相关问题。

五、课程教学内容与安排

每周四个学时的教学，详细教学大纲如下：

第一章 强化学习数学基础 （2周8个学时）

1.1 第一节 条件概率与贝叶斯

1.2 第二节 概率论不等式

1.3 第三节 条件期望

1.4 第四节 贝叶斯统计推断初步

第二章 多臂老虎机（2周8个学时）

2.1 第一节 随机老虎机基础

2.2 第二节 带上下文信息的老虎机

2.3 第三节 汤普逊抽样

2.4 第四节 多臂老虎机的各种应用

第三章 马尔科夫决策过程（3周12个学时）

3.1 第一节 马尔科夫链

3.2 第二节 马尔科夫链蒙特卡洛算法

3.3 第三节 马尔科夫奖励过程

3.4 第四节 马尔科夫决策过程

3.5 第五节 贝尔曼最优方程

3.6 第六节 强化学习介绍

第四章 动态规划 （2周8个学时）

4.1 第一节 策略评估

4.2 第二节 策略提高

4.3 第三节 策略迭代

4.4 第四节 价值迭代

第五章 蒙特卡洛算法 （2周8个学时）

5.1 第一节 On-policy下的蒙特卡洛预测与估值

5.2 第二节 On-policy下的蒙特卡洛控制

5.3 第三节 Off-policy下的预测

5.4 第四节 Off-policy下的蒙特卡洛控制

第六章 时序差分学习 （2周8个学时）

6.1 第一节 时序差分预测

6.2 第二节 单步时序差分算法

6.3 第三节 On-policy下的时序差分控制算法 (Sarsa)

6.4 第四节 Off-policy下的时序差分控制算法 (Q-Learning)

第七章 深度强化学习 （3周12个学时）

5.1 第一节 神经网络初步

5.2 第二节 深度学习算法简介

5.3 第三节 策略梯度算法及其各种变种

5.4 第四节 Actor-Critic 算法及其各种变种

5.5 第五节 其他深度强化学习算法

5.6 第六节 深度强化学习算法的各种应用： 围棋，机器人，无人车系统，网络等

六、考核方式和成绩评定方法

本课程重视过程考核，围绕重要教学目标开展考核，评定细节如下:
(1)作业:占比 60%。 六次算法作业，包括理论分析和编程两个部分;每次作业满分占总成绩比例是10%， 总共6次，占比60%。

(2)期末考试:占比40%。以课程研究项目 (Course Research Project)的形式开展，要求每位同学在任课老师的指导下阅读文献提出新的想法，并通过课上学习到的知识对研究问题

进行理论分析，算法设计和编程实现。最终形成项目报告文件，并要求做课程演讲汇报(Course Project Presentation)。

七、教材和参考书目

(一)、推荐教材

无

(二)、参考书目

无

八、学术诚信教育

本课程高度重视学术诚信，严禁抄袭、作弊等行为。“在学习、科研、实习实践等活动中，学生应恪守学术道德，坚守学术诚信，保护知识产权，坚持勇于创新、求真务实的科学精神，努力培养自己严谨求实、诚实自律、真诚协作的科学态度，成为良好学术风气的维护者、严谨治学的力行者、优良学术道德的传承者。” (具体请参

见《上海科技大学学生学术诚信规范与管理办法（试行）》文件要求，如果教师有更具体的要求，请详细列出。）

九、其他说明(可选)

无

《Reinforcement Learning》Syllabus

1.Basic course information

Course name	Reinforcement Learning	Course code	SI252
Course Level		Credit/Contact Hour	4/64
Prerequisite	Null	Teaching Language	
School/Institute	School of Information Science and Technology	Instructor	邵子瑜

2.Course Introduction

Null

3.Learning Goal

Null

4.Instructional Pedagogy

Null

5. Course Content and Schedule

Null

6.Grading Policy

Null

7. Textbook & Recommended Reading

(1) Textbook

Null

(2) Recommended Reading

Null

8.Academic Integrity

This course highly values academic integrity. Behaviors such as plagiarism and cheating are strictly prohibited. Please list more if you have more specific requirements.

9.Other Information (Optional)

Null