

## 课题七：强化学习

强化学习是智能体 (Agent) 以“试错”的方式进行学习, 通过与环境进行交互获得的奖赏指导行为, 目标是使智能体获得最大的奖赏, 强化学习不同于连接主义学习中的监督学习, 主要表现在强化信号上, 强化学习中由环境提供的强化信号是对产生动作的好坏作一种评价(通常为标量信号), 而不是告诉强化学习系统 RLS(reinforcement learning system)如何去产生正确的动作。由于外部环境提供的信息很少, RLS 必须靠自身的经历进行学习。通过这种方式, RLS 在行动-评价的环境中获得知识, 改进行动方案以适应环境。

请详细介绍强化学习算法, 并给出 2-3 个强化学习的应用实例, 实现比较。

### 课题来源:

- 1) 动态偏好多目标进化优化方法研究, 64 万, 2019-2022
- 2) 基于动态环境的多目标进化方法研究, 国家自然科学基金, 27 万, 2015-2018;
- 3) 用高维多目标进化算法解决网络文本信息特征值提取的问题研究, 省科技计划项目, 4 万, 2014-2015;
- 4) 基于偏好的多目标进化优化方法, 省教育厅重点项目, 8 万, 2018 年-2020 年;
- 5) 算法设计与分析, 教育部共同育人项目, 3 万, 2017-2018 年;
- 6) 偏好多目标进化算法关键问题研究, 排名第 2, 国家自然科学基金, 67 万, 2018-2021;
- 7) 高维多目标进化算法关键问题研究, 排名第 2, 国家自然科学基金, 77 万, 2014-2017。

### **支持的课程目标**

**课程目标 1:** 掌握各种常用算法的基本思想、适用条件和设计方法，能运用算法设计的基本策略解决相应的具体问题。

该目标支持毕业要求 1.2 掌握计算机科学与技术专业所需的算法分析与程序设计等知识，具有计算思维能力，能将其用于复杂工程问题模型的实现。

**课程目标 2:** 能够对软件工程相关的复杂工程问题进行分析，对问题抽象和建模，设计并实现解决该问题的算法，并能对该算法的复杂度进行理论分析，对算法和程序进行优化。

该目标支持毕业要求 2.3 能够应用数学、自然科学和计算机科学与技术的基本原理对其相关的复杂工程问题进行提炼、定义、建模、分析和评价。

**课程目标 3:** 对软件工程问题结合领域技术规范、政策和法规，设计和选择合适的算法，并能，并能分析算法适用范围及影响。

该目标包含课程思政内容，支撑毕业要求 3.2 能够在系统方案设计环节中体现创新意识，同时考虑多方面、多层次因素的影响，如社会、健康、安全、法律、文化以及环境等因素。

**课程目标 4:** 能够对软件工程相关的复杂工程问题中的新型算法进行跟踪学习。

该目标包含课程思政内容，支持毕业要求 12.1 认识到自主学习的必要性，具有自主学习和终身学习的意识，掌握跟踪软件工程学科前沿、发展趋势的方法和途径。

**要求：**1. 对目前常用的强化学习算法进行调查；

2. 请实验比较分析至少 2 种现有的强化学习算法（计算机科学与技术专业要求至少 3 种）；

3. 强化学习采用“试错”的方式进行学习，请调查人类或生物的学习行为中还有哪些场景采用“试错”的方式进行学习，并分析这种学习行为的优缺点。