

课题名称

中文： 人工智能专业课题：大数据算法模型与应用

英文： Algorithms for Big Data

课题关键词

人工智能|计算机科学|数据科学|数理统计|算法原理|机器学习|最优化|神经网络

课程描述

算法是指解题方案的准确而完整的描述，是一系列解决问题的清晰指令，算法代表着用系统的方法描述解决问题的策略机制。从技术上说，算法是一种中介，通过算法模型，将信息与用户进行匹配，本质是要解决信息和用户的精准匹配问题。无论是传统的机器学习算法，还是近年来兴起的深度学习算法，通过用户个人属性和网络应用使用过程中的数据记录，挖掘用户个人兴趣、需求，最终达成个人信息需求的精准匹配，这就是算法的使命。而算法和大数据相互依赖，算法能够从大数据中获得信息和洞察，而大数据则需要算法来进行有效的处理、分析和应用。这种相互关系在科技、商业和社会等领域都有广泛的影响。

本课程会介绍一些经典的算法设计和分析。我们将介绍算法技术,如动态程序设计、散列和数据结构,分治算法,网络流和线性规划。我们还将涵盖范围广泛的分析工具,如recurrences、概率分析,平摊分析和势函数。除了学习算法,我们还会涉及一些复杂性理论的研究——双重的算法设计（下界方法在这些模型中的显示和最优算法）。最后,我们将讨论新模型在现代大型数据集下的应用,比如在线算法、机器学习和数据流。

导师简介



卡内基梅隆大学终身教授

David Woodruff

- 卡内基梅隆大学计算机系终身教授
- UCB Simons Institute 数据科学项目创建者及主席
- IBM Almaden研究中心资深研究员
- STOC 2013、PODS 2010 最佳学术研究论文奖得主
- 曾获EATCS Presburger

研究方向

- Compressed Sensing 压缩传感
- Data Stream Algorithms and Lower Bounds
- 数据流算法和下限
- Dimensionality Reduction 降低维度
- Distributed Computation 分布式计算
- Machine Learning 机器学习
- Numerical Linear Algebra 数值线性代数
- Optimization 数据优化

课程大纲

1. 机器学习
2. 神经网络

3. 数值线性代数或优化
4. 稳健回归或Hessian草图
5. 实际问题与设计复杂模型
6. 理论（数据驱动）算法