深圳大学计算机与软件学院
College of Computer Science & Software Engineering,
Shenzhen University

大数据技术与应用研究所 Online学术讲座

对抗机器学习的一些进展

报告人： 朱军 教授

主持人： 王熙照 教授

日　　期： 26th Oct, 2020 （星期一）

时　　间： 14:30-15:30 PM

地　　点： 会议链接：https://meeting.tencent.com/s/eWBTQvEFjF90

会议号：880 377 231 会议密码：444555

演讲嘉宾简介

朱军，清华大学计算机系教授、北京市智源学者、人工智能研究院基础研究中心主任。2001到2009年获清华大学计算机学士和博士学位，之后在CMU做博士后，2011年回清华任教，2015到2018年任卡内基梅隆大学兼职教授。主要从事机器学习基础理论、高效算法及应用研究，在国际重要期刊与会议发表论文百余篇。担任顶级期刊IEEE TPAMI的副主编、AI编委，担任机器学习国际大会ICML2014地区联合主席， ICML、NeurIPS、IJCAI、AAAI、ICLR等国际著名会议的领域主席，ICML2020最佳论文评选委员会委员。获科学探索奖、中国计算机学会（CCF）自然科学一等奖、CCF青年科学家奖等，入选科技部中青年创新领军人才、MIT TR35中国先锋者以及IEEE Intelligent Systems评选的“AI’s 10 to Watch”等。带领团队研制“珠算”深度概率编程库、“天授”强化学习库和RealSafe对抗攻防平台，获得首届“对抗样本攻防竞赛”国际竞赛所有三个任务的冠军、ViZDoom对抗决策国际竞赛2018年冠军等7项、部分算法成为主流开源软件FoolBox、CleverHans的标准算法。

摘要

在开放动态的应用环境中，人工智能技术面临着对抗噪声的干扰。大量工作显示，性能良好的深度学习模型（或一般的机器学习方法）通常容易被对抗噪声攻击，这给实际应用带来了很多风险。如何有效进行攻击以及如何进行防守受到了学术界和工业界的广泛关注。谷歌在NIPS 2017会议上举办了首届对抗攻击与防御的国际竞赛，促进了相关研究。在这个报告中，将介绍深度学习对抗攻击与防守方面的一些最新进展，包括对抗攻防的全面评测、赢得竞赛的解决方案以及一些近期的工作。

诚挚邀请全院感兴趣的老师同学参加