

Winter 2018



NEWSLETTER 第 8 期

二〇一九年三月



目录

访问人员 ----- 02

新闻动态 ----- 04

讲座报告 ----- 16



访问学者



蒋俊正 教授于 2005 年在桂林电子科技大学获学士学位，2011 年在西安电子科技大学雷达信号处理国家重点实验室获得博士学位。2011 年至今他在桂林电子科技大学信息与通信学院从事教学/科研工作，是教授和博士生导师。2016 年他主持完成国家自然科学基金一项和广西自然科学基金一项。目前，主持国家自然科学基金项目一项。在 IEEE Trans Signal Processing, Signal Processing, IET Signal Processing 等国际信号处理领域的重要期刊上发表 SCI 检索论文 18 篇，其中 SCI 二区 7 篇。他的研究领域包括：图信号处理理论与算法，分布式信号处理理论与算法，大规模传感器网络数据处理，大数据的半监督学习。

• 新入职成员 •



朱子沛

算法分析师

硕士（香港中文大学（深圳））

研究方向：自然语言处理

2019 年硕士毕业于香港中文大学(深圳)理工学院数据科学专业，研究方向主要围绕自然语言处理。曾在中国平安担任算法工程师一职，主要负责利用机器学习方法解决实际业务问题，为具体项目开发模型和算法支持风控、营销、投资决策等。



黄瀚铭

高级工程师

博士（新加坡南洋理工大学）

研究方向：定量金融，信用风险模型

黄瀚铭博士目前在深圳大数据研究所担任高级工程师。黄博士是一位金融分析师，拥有超过十年的海外和国内工作经验。他专注于宏观经济分析、数学建模、风险管理和量化金融。在加入 SRIB 之前，他曾在财务和数据分析领域担任过不同职位。



吴毓双

南京邮电大学

研究方向：目标检测



肖婕

香港中文大学（深圳）

专业：数据统计

研究方向：数据分析



徐宁

香港中文大学（深圳）

研究方向：强化学习与
多智能体强化学习



陈宇丝

香港中文大学（深圳）

专业：统计科学

研究方向：自然语言处
理、机器学习



郑玉健

哈尔滨工业大学（威海）

专业：软件工程

研究方向：

三维重建



郭俊甫

大连理工大学

专业：数学

研究方向：

三维点云的语义分割

新闻动态

我院成员访问深圳市城市交通规划设计研究中心

2018年12月11日，深圳市大数据研究院副院长蔡小强教授带领研究院成员访问深圳市城市交通规划设计研究中心，受到模型院院长邱建栋先生，技术质量部部长吕北岳先生和谭章智博士的接见。

访问期间，谭章智博士首先带领大家参观了城市交通规划设计研究中心布设在沙河西路的多功能智慧灯杆。据介绍，灯杆的设计是为深圳市智慧道路提出解决方案。以智慧灯杆为载体，集成多种传感设备，预留未来新增设备的挂载接口，实现对城市道路静态设施和智能设备的集约

化管理，实现多杆合一。紧接着，邱建栋院长带领大家参观了交通大数据实验展厅，并详细介绍了深圳市城市交通规划设计研究中心的发展现状和业务案例，给大家留下深刻印象。



深圳市大数据研究院 2018 年度战略研讨会

2018年12月15日，深圳市大数据研究院于深圳青青世界召开 2018 年度战略研讨会，香港中文大学（深圳）特邀教授及研究院全体教职工参加了本次会议。会议以分组讨论的形式进行，三个小组分别就研究院的科研发展、产业化及行政管理等相关重大事项进行了热烈讨论。

会议结束后，研究院师生们集体前往深圳青青世界景区进行游览。午后的青青世界阳光明媚，花草悠悠。景区内雨雾弥漫的热带雨林，新潮的打击乐表演，回归童年的趣味游戏，为每



参会嘉宾及教职工合影留念



研究院师生游览深圳青青世界

日埋头于科研工作中的师生们带来别样的乐趣。

晚餐期间，在研究院参与实习的香港中文大学（深圳）理工学院秦泽钰同学为大家带来了歌曲串烧、来自厦门大学的访问学生林志地用家乡话为大家高歌一曲《爱拼才会赢》；除了学生们的精彩表演，老师们也都纷纷上台展现自己的特长，现场气氛一片欢乐。

深圳市大数据研究院 2018 年度战略研讨会于 2018 年 12 月 15 日圆满结束。回顾 2018 年，深圳市大数据研究院发展迅速，这离不开研究院师生们的团结与拼搏，更离不开各界对我们的大力支持；展望 2019 年，新起点，新目标，深圳市大数据研究院将在科研道路上创造新的里程碑，奔向更加美好的未来！



研究员老师欢乐合唱



深圳市大数据研究院师生合影

大数据研究院“WattBalance”团队荣获“智能客服” 第二届数据应用创意节比赛冠军



2018 年 12 月 21 日，来自深圳市大数据研究院的“WattBalance”团队在成生辉和段东亮两位老师的带领下参加了中国南方电网广州供电局举办的“智能客服”第二届数据应用创意节活动决赛，以 90.14 的高分获得决赛的第一名。

据了解，本次活动以“数据赋能·智电未来”为口号，以“大数据与人工智能的电力服务运用”为创意方向，以“对内提质增效、对外合作创新”为比赛主题，吸引了来自全国 5 个省以及广州供电局内部共计 63 支优秀团队、65 个项目报名参赛。

经过 12 月 8 号初赛的激烈角逐,“WattBalance”团队从全国众多优秀企业团队和高校队伍中厮杀突围成功,顺利进入决赛,最终与其他五支团队在本周五上午进行冠亚季军之争。

决赛分为展厅成果展示和现场演讲答辩两个环节。尽管其他团队有类似大数据打假之类的新意,但“WattBalance”团队以别具一格的基于用户画像分析的精准营销、基于深度网络预测的精准供电及可视化一体化平台的成熟方案成功吸睛,并凭借着程嘉钰同学出色的演讲水平,获得专家评委和在场观众的一致好评,成功摘下第二届数据应用创意节的桂冠。



林志地同学在展厅给广州供电局董事长甘霖讲

香港中文大学（深圳）两大研究院获 深圳市基础研究机构建设授牌



1月6日,深圳市基础研究机构授牌仪式举行,广东省委副书记、深圳市委书记王伟中,深圳市委副书记、市长陈如桂和深圳市委常委、秘书长高自民为基础研究机构授牌。香港中文大学（深圳）牵头组建的深圳市人工智能与机器人研究院、深圳市大数据研究院获深圳市基础研究机构建设授牌。香港中文大学（深圳）校长、中国工程院院士徐扬生教授与副校长（学术）、加拿大皇家科学院院士罗智泉教授出席仪式并分别代表两大研究院接受牌匾,协理副校长兼科研处处长李学金教授也出席授牌仪式。

深圳市大数据研究院以建立国际一流、具有自主知识产权、核心技术优势突出的“大数据处理分析研发基地”为目标,最大限度发挥国家超级计算深圳中心在大数据产业研发中的核心平台作用,充分利用香港中文大学（深圳）国际化科研团队的科研力量,为打造集大数据分析处理技术创新研究、技术服务及交流、核心技术研究、人才引进与培养的综合性创新型研究机构,成为世界一流的大数据研发和公共服务平台,在智慧医疗、智能通信、智慧交通等城市管理等方面为深圳、全





省乃至国家的科学研究、产业发展、政府决策、公共服务做贡献。深圳市大数据研究院首任院长罗智泉教授是加拿大皇家科学院院士，IEEE Fellow，SIAM Fellow 以及香港中

文大学（深圳）副校长（学术）。

港中大（深圳）建校五年以来，学科建设已逐步完善，在机器人与人工智能、大数据与数据科学、未来智能网络、生物信息学、新材料和能源科学、国际金融与物流、制药和精准医学等领域逐渐组建世界级专家队伍，正在面向国家及珠三角地区战略性新兴产业发展需要，打造国际化科技创新平台。

深圳市大数据研究院联合主办 2018 年数据科学国际研讨会

12 月 20 日至 22 日，2018 年数据科学国际研讨会在香港中文大学（深圳）举行。本次研讨会由深圳市大数据研究院、香港中文大学（深圳）、数据与运筹科学研究院、西安电子科技大学，明尼苏达大学数学与应用研究院联合主办。研讨会聚集了来自国内外知名大学的专家和学者，共同交流讨论计算数学在信息科学和大数据分析领域应用的新思想、新技术和创新研究。



开幕式上，香港中文大学（深圳）副校长、深圳市大数据研究院院长罗智泉教授首先致欢迎辞，他简要介绍了学校近几年在办学规模、人才引进、学科建设上取得的成绩。

随后香港中文大学（深圳）理工学院执行院长、深圳市大数据研究院副院长崔曙光教授介绍了深圳市大数据研究院的发展现状、研究队伍、科研成果以及远景规划。



最后香港中文大学（深圳）数据与运筹科学研究院共同院长戴建岗教授介绍了数据与运筹科学研究院的发展状况。

本次研讨会邀请的演讲嘉宾包括：



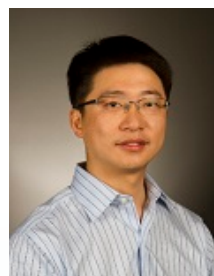
密西根大学
Alred Hero 教授



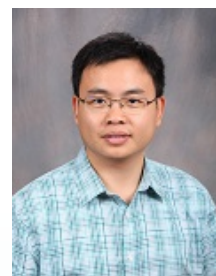
密西根大学
Brian D. Athey 教授



康奈尔大学
Lang Tong 教授



亚利桑那州立大学
Lei Ying 教授



德州农工大学
Le Xie 教授



麻省理工大学
David Simchi-
Levi 教授



香港科技大学
Daniel P. Palomar
教授



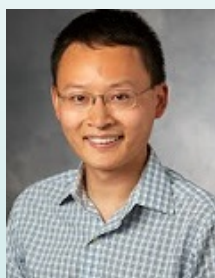
意大利 LUISS 大学
Marco Scarsini
教授



加州大学欧文分校
A. Lee Swindlehurst
教授



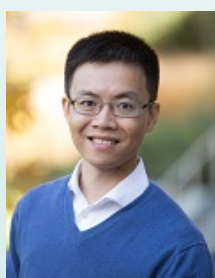
加州大学
洛杉矶分校
Wotao Yin 教授



斯坦福大学
James Zou 教授



香港中文大学
Wing-Kin Ma 教授



康奈尔大学
Yudong Chen 教授



亚利桑那州立大学
Junshan Zhang 教授



伊尔梅瑙工业大学
Martin Haardt 教授



伊利诺伊大学
香槟分校
Ruoyu Sun 教授



京东 AI 研究院
Bowen Zhou 博士



普林斯顿大学
Yuxin Chen 教授

我院老师指导两学生代表队于 AI-Trans 全球智能网络传输竞赛中获奖



1月12日，AI-Trans 全球智能网络传输竞赛总决赛在北京举行，深圳市大数据研究院和香港中文大学（深圳）两学生代表队“枯鱼肆队”与“WwW 队”遇强愈强，超水准发挥，以突出的算法设计和出色的成果演示获得评委的一致好评，分别夺得“最佳非机器学习奖”与“最佳演示答辩奖”。另外，我校深圳市大数据研

究院的李钰鹏老师荣获“最佳指导教师奖”。

AI-Trans 全球智能网络传输竞赛由清华大学计算机系牵头，联合中国计算机学会、中国人工智能学会智能信息网络专委会、中国通信学会信息通信网络专委会和 Power Info 协办，斯坦福大学计算机系、新加坡国立大学计算机系和华为未来网络理论实验室也作为合作伙伴积极参与了竞赛的组织。AI-Trans 将成为今年国际计算机多媒体领域顶级会议 ACM Multimedia 2019 的 Grand Challenge。

本次竞赛的初赛吸引了全球上百支队伍参加，其中，深圳市大数据研究院的李钰鹏老师带领我校“枯鱼肆”队（李漪静、石思源、王依然）、“WwW”队（郝森跃、谢国超、莫子恒）和“太阳神”队（何涛、李晓璇、冉晨宇）三支学生代表队伍参赛。经过两个多月的精心准备，三支代表队在初赛中表现出色，全部晋级决赛，成为全球 25 支决赛队伍之一。据悉，晋级决赛的 25 支队伍中，不仅包括来自中科院、清华大学、北京大学、上海交通大学、北京邮电大学、天津大学、武汉大学等国内一流高校和科研单位的代表队，还有来自快手科技、阿里巴巴的业界专业团队，以及来自美国、加拿大、德国、日本等国的参赛队伍。

近年来，数据规模的爆炸式增长、动态异质网络的个性化需求，都对网络传输提出了新的挑战。网络研究人员、设备供应商和互联网服务提供商一直在试图探究各种各样新兴科技以实现网络传输智能化。而最近的技术创新，则为实现“智能化”网络传输技术提供了新的机遇，例如，人工智能正在成为推动新工业革命的主流研究方向。我校大数据研究院的研究团队则希望通过机器学习、数据挖掘、控制论等方法实现自动化的高效、准确、公平的智能网络传输。

目前国内流行的直播平台如斗鱼、映客，都是默认选择较高码率，在频繁出现卡顿之后，会有相应的弹窗提示用户往下调为标清、流畅等，另外，也有平台默认选择低码率，如此一

来，用户就不能享有高清的观看体验。针对此问题，近年来，学界和工业界也一直在积极研究自适应调节码率算法，本次比赛旨在鼓励参赛者设计自己的自适应码率调节算法，用技术的力量解决这个现实问题。

参赛队伍的获奖感言



“WwW”代表队

队长：莫子恒 2016 级理工学院学生 新能源工程专业

这次是我第一次参加的正规编程比赛。这次比赛带给我两点很重要的收获。首先，通过与两位计算机科学专业的优秀队友合作，我得以接触及了解到许多计算机科学专业的先进技术以及新颖观念。例如这次的比赛需要我们设计一个算法来帮助视频直播 APP 在不同的网络环境下尽可能地为用户提供清晰流畅的视频播放。针对这个问题需要用到机器学习算法。此前我仅仅听闻机器学习是一个非常强大的计算机技术，但却一直没有深入了解。但得益于队友的帮助，我得以了解到机器学习的基本概念、算法思想及实现机器学习的一些数学理论。除此以外，在平时与他们的畅聊中，我还了解到了计算机科学领域中的一些技术突破以及一些技术的革新或淘汰。所有这些都大大打破了我对计算机科学的原有认识。

其次是这次竞赛高手云集，我们的对手不仅有在计算机及网络通信领域中的实例派高校如清华、北邮、武大和天大，还有非高校队伍如快手、阿里巴巴和中科院计算所。通过与他们同台竞技，我更加深刻的认识到自己现有的专业水平与实力与我身边的大部分人比起来还远远不足。在我们学校优秀的人之外，其他高校还有更优秀的人，而在他们之外则是计算机领域中更加优秀的人。而再在他们之外，更加优秀的人依然大有人在。这一次比赛结束后，带给我更多的是紧迫的危机感，而非欣喜。我明白，唯有加倍努力，方能跟上强者的步伐。我十分感谢我的队友郝森跃和谢国超，他们的帮助和知识让我们队伍能够一路披荆斩棘，勇往直前。同时刷新了我原有的旧知识，给我带来了许多新颖的思想。

特别感谢李钰鹏老师在这次竞赛中，对我们算法的设计思想方向进行了多次指导，让我们走了很多弯路。其次，他对我们算法的训练模型提出了宝贵的建议，使得我们的算法有很大的提升，这也使得我们得以在预赛中成功晋级。另外，老师还指出了我们答辩内容的很多错误和逻辑漏洞，并且对内容做出了适当调整，使我们的答辩更加严谨、完善。没有老师的指导，我们也无法获得“最佳演示答辩”奖。此外，也感谢深圳市大数据研究院和我们学校的大力支持。



队员：郝森跃 2016 级理工学院学生 计算机专业

这个比赛的初赛正值我校期末考试周，朋友拉上我说很想参加一次这种比赛，我虽然稍有犹豫，但还是报名参赛。之所以犹豫是因为之前我也报名过相关比赛，但最终都是不了了之。帮助我们克服这一困难的就是我们的指导老师李钰鹏，他督促我们根据比赛进度设置工作时间线并给出策略建议，正是因为拥有良好的团队合作和时间安排，我们才得以安全度过期末考试，又成功在上百支队伍中杀进决赛。

在决赛阶段，我们不断调整、训练我们的比赛模型，因为评分环境与规则和初赛有所区别，而准备时间又稍显紧张。顶着压力，我们一边重新收集数据进行训练，一边着手准备决赛答辩。作为主答辩选手，我下定决心要用英文描述我们的算法，不过这确实给演示文稿的制作以及对讲稿的熟练程度带来不小的压力，但是我的队友还是很积极配合。直到决赛现场，我都在拼命熟悉讲稿，并一直和队友还有指导教师交流，为了防止思路不清晰等一些错误。这应该是我第一次站在讲台上进行如此正式的答辩，虽然开始确实很紧张，但是通过 Intro 部分成功调解，我想这也是我校两年英语与 GE 课还有当 USTF 的经历所带来的经验吧。我校是现场使用英文答辩的唯一队伍，而我们更是全场本科生使用英文答辩的唯一队伍。

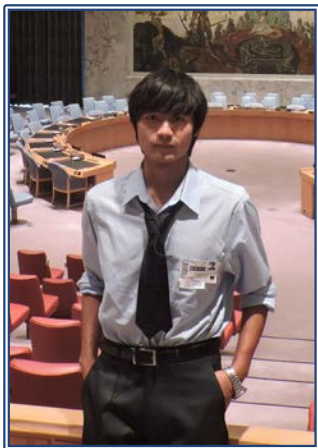


队员：谢国超 2017 级理工学院学生 计算机专业

这是我第一次参加 AI 比赛，是一次新鲜又收获颇丰的体验。

这次比赛的主要收获是 AI 比赛的经历。在这次比赛中，我主要负责模型搭建和编程任务。通过分析问题，我们认为这次比赛需要用到我们之前没有接触的强化学习。通过自学，我们提出了一个基于强化学习思想的监督学习模型。应对决赛环境的变化，我们又尝试了使用预赛环境训练决赛模型的方法。比赛提供了很多尝试自己奇思妙想的机会。

这次比赛的另一个收获是临危不惧的心态。因为预赛恰好是 final 周，在预赛之前，我们模型的表现并不太好。预赛第一天半天过去了，我们的前四次提交（每天只取五次提交的最高分）排名都在 50 名之外。这时候，我们没有急于通过调参把提交用完，而是重新思考这个问题。一个突然的想法进入我们的视野：在 reward_base 的基础上加上对 next_state 的预测，从而可以通过 BFS 得到更远的视野。于是，我们马上重新训练模型，并进行了最后一次提交。最终，我们进入了前 20，并且在之后的四天更进一步，最终得以进入决赛。



“枯鱼肆”代表队

队长：王依然 2018 级数据科学硕士生

一开始了解这次比赛是通过深圳市大数据研究院导师李钰鹏的介绍。在此之前，我并没有了解过自适应码率的问题，相比于主流算法，这是一个独特的算法领域，不同于传统机器学习或深度学习。自适应码率算法是基于实时的视频数据和特定传输环境来做出最优决策，本质上是一个最优化问题。但是传统的最优化建模方法在题目所给的环境中无法成立且很难应用于实际，因此解决问题的模型范围只有传统的 Based on Buffer and Throughput（基于缓冲和吞吐量）方法以及 Deep Reinforcement Learning（深度强化学习）方法。

经过两个多月不断的改进和尝试，我们从预赛一路杀入决赛，并最终斩获“最佳非机器学习奖”！这一路掉过很多坑，但是最终还是顺利将算法一步步完善。从一开始的论文阅读和技术积累，到代码环境搭建，到算法实现，再到测试提升，我们学到了很多，让我深知不断学习、和 learning by doing 的重要性。

除此之外，作为队长，既要平衡课业时间，还要调动大家的积极性，让我觉得颇有压力，对我也是一种锻炼，所幸队友相当给力，顺利大家一起坚持到了最后。尤其是在初赛的最后阶段，我们面临期末考试周，最后靠着挤出来的一些时间进行讨论和代码撰写，虽然累但是也有回报。



队员：石思源 2018 级数据科学硕士生

从去年 11 月的初赛到刚刚结束的决赛，一路以来枯鱼肆小分队从初期算法问题频出到预赛时起死回生再到最终决赛顺利通关，很感激深圳市大数据研究院的李钰鹏老师给予我们及时的指导，多次在我们算法遇到瓶颈的时候帮助我们找到了突破。也为我们小组能够默契配合集思广益最后取得一定成绩感到高兴。

在决赛中，我们与来自世界各地优秀的学界和工业界的队伍进行了切磋。他们的优势与风格各异，有的以紧跟前沿的算法模型出彩，有的以有效的变量设计出众，而我们的算法思路通透，简洁高效，得到了主办方的认可，非常荣幸能获奖。

队员：李漪静 2018 级数据科学硕士生

首先谢谢队友以及指导老师的大力 carry。经过这次比赛，我学到了很多关于网络传输和强化学习的知识，了解了一个项目从无到有，从简陋到丰富的过程，并且意识到了理解数据对于建立，改进模型的重要性。在决赛现场听了大牛和主办方对于人工智能在网络传输领域的应用，前景以及面临的问题的报告，这大大开阔了我的视野以及开拓了未来找工作的方向。

指导老师：深圳市大数据研究院李钰鹏

这次参加决赛的都是全球数一数二的高等院校和研究机构，不乏有业界直播和 OTT 行业的著名公司团队。为什么学界和业界都重视这次比赛？数据规模的爆炸式增长、动态异质网络的个性化需求，都对网络传输，特别是视频传输，提出了新的挑战。视频码率控制亟待解决，而且非常有研究价值。



在长达 3 个月的比赛中，参赛同学努力备战，从问题瓶颈的分析、码率控制算法的设计、论文深挖、文档阅读，到搭建算法框架、落地实现、迭代改进、鲁棒性和可扩展性测试，全程细致入微，不放过每一个可优化提高的环节。比赛中，三个代表队使用不同的算法架构，成功打入决赛，这十分不易，也展现了我校同学扎实的算法设计能力和逻辑分析素养，这与我校独具特色的高质量教学分不开。决赛中，我校同学设计的算法性能优异，架构和设计逻辑严谨细致、具有强鲁棒和高扩展性，获得来自学界和业界评委高度评价，并获颁两项大奖。这些奖十分不易，我为他们骄傲。得益于同学们的优秀表现，我有幸获评“最佳指导教师”。

我院两篇研究工作被录取为 CVPR 口头报告

近日，从 2019 年 IEEE 计算机视觉与模式识别会议（IEEE Conference on Computer Vision and Pattern Recognition, CVPR）传来喜讯，我院韩晓光博士团队的两份研究工作均被 CVPR(2019)录取为口头报告（Oral Presentation）。



第一份工作题为《A Skeleton-bridged Deep Learning Approach for Generating Meshes of Complex Topologies from Single RGB Images》为我院与华南理工大学和微软亚洲研究院合作共同完成，第一作者为我院暑期访问学生唐佳鹏，韩晓光博士为共同第一作者。该工作提出了一种全新的基于骨架表达的三维几何深度学习算法，非常有效地解决了计算机视觉中一个非常具有挑战性的问题—从单视角图像中重建完整的三维物体。尤其是对于具有复杂拓扑结构的物体，本文的算法较已有算法效果尤其显著。本工作受到会议评审者的一致好评，三位评审者均给出了强烈接受（Strong Accept）的意见。

第二份工作《Deep Reinforcement Learning of Volume-guided Progressive View Inpainting for 3D Point Scene Completion from a Single Depth Image》为我院与香港中文大学（深圳）、大连理工大学、中国科学技术大学以及阿里巴巴集团共同完成。韩晓光博士为第一作者，我院其他主要参与作者包括我院访问学生张肇轩、杜冬和杨明岱，崔曙光教授以及熊子祥访问教授。该工作针对计算机视觉中另一挑战性难题—基于单视角深度图恢复完整三维场景，提出了一种基于三维与二维卷积神经网络协同学习的多视角补全技术，并首次将深度强化学习用于引入该问题的求解过程。本文的方法在公开数据集上获得了世界领先水平。本工作受阿里巴巴创新研究计划资助。



CVPR 是由美国电气及电子工程师学会（IEEE）组织的计算机视觉领域最权威的国际会议之一，每年一届，与 ICCV（计算机视觉国际会议）和 ECCV（欧洲计算机视觉会议）并列为计算机视觉领域最顶级的三大国际会议。国际计算机视觉领域普遍认为，CVPR/ICCV/ECCV 三大会议代表了计算机视觉领域的旗舰和风向标，而其 Oral Paper 则基本代表当年度计算机视觉领域的最高水准。2019 年，CVPR 共收到有效投稿论文 5160 篇，最终会议论文的录用率为 25.2%，而大会口头报告（Oral）论文的录用率仅为 5.6%。

韩晓光博士，现任香港中文大学（深圳）理工学院与深圳市大数据研究院研究助理教授。他在 2017 年 9 月获得香港大学计算机专业博士学位。在此之前，他于 2009 年本科毕业于南京航空航天大学数学系，于 2011 年在浙江大学应用数学专业获得硕士学位，并于 2011 年至 2013 年间在香港城市大学担任研究助理。韩博士于 2017 年 9 月加入香港中文大学（深圳）

和深圳市大数据研究院，从事计算机科学领域的教学科研工作。他的主要研究方向包括计算机视觉，计算机图形学，人机交互以及医学图像处理。

韩晓光博士简介: <http://www.cuhk.edu.cn/academics/hanxiaoguang/index.html>

《A Skeleton-bridged Deep Learning Approach for Generating Meshes of Complex Topologies from Single RGB Images》 请详见网址:

<https://arxiv.org/pdf/1903.04704.pdf>

《Deep Reinforcement Learning of Volume-guided Progressive View Inpainting for 3D Point Scene Completion from a Single Depth Image》 请详见网址:

<https://arxiv.org/pdf/1903.04019.pdf>

苏州市人民政府办公室访问深圳市大数据研究院



2019年2月26日，苏州市政府副秘书长暨办公室主任张旗、苏州市财政局总会计师王强，苏州市行政审批局副局长庄朝鸣等市领导一行访问深圳市大数据研究院，并与研究院的成员举行了交流座谈会。出席此次座谈的研究院成员包括科研经理崔来中博士，研究员万翔博士，李镇

博士，周健君博士，成生辉博士及陆彦辉博士。

会上，崔来中博士首先介绍了深圳市大数据研究院的基本信息，发展现状，研究队伍及远景目标。随后，在研究院参会人员介绍了各自项目基本情况后，苏州市政府办一行参观了深圳市大数据研究院展示厅，并针对大数据领域的基础研究政府在大数据研究工作中提供的支持进行了热烈讨论。

会后，来访嘉宾参观了香港中文大学（深圳）校园。



讲座报告

一、纳米孔测序端到端映射中一种准确快速的连续小波动态时间规划算法

时间：12月22日星期六，下午 16:30–17:30

地点：诚道楼 207

演讲嘉宾：Sheng Wang 博士，阿卜杜拉国王科技大学

演讲摘要：

基于纳米孔的第三代测序带来了长读取、即时检测和无聚合酶链反应的前景。在纳米孔数据分析的步骤中，原始电流信号和参考预期信号（即，信号标记）之间的端到端映射用是关键部分。解决信号标记问题的经典算法之一是动态时间规整（DTW）。然而，超长纳米孔测序和采样速度的大小差异可以让问题变得复杂，并使得经典 DTW 不再有效。

此次讲座中，我们提出了一种新型的多级 DTW 算法：连续小波 DTW（cwDTW）。这种算法基于连续的小波变换，伴随着两个不同尺度的输入信号序列。我们的算法从两个序列的低分辨率小波变换开始，使得变换的序列短并且具有相似的采样率。然后，我们提取变换序列的峰值和最低点，用以形成具有相似长度的特征序列，它可以通过原始 DTW 容易地对齐。接下来，我们的算法通过构建依赖于上下文的边界以及启用对后者中变形路径的约束搜索，递归地将变形路径从较低分辨率级别投影到较高分辨率级别。

综合实验证明了 cwDTW 的效率和有效性，与原始 DTW 相比，cwDTW 具有加速度的显著提升（大约快 3000 倍）和对准精度的微小失误两个优势。我们还表明，cwDTW 可以作为精确医学中重要任务的关键模块，例如单核苷酸变体（SNV）的鉴定和表观遗传学中的甲基化检测。

嘉宾介绍：

Sheng Wang 博士是阿卜杜拉国王科技大学（KAUST）的研究科学家。他于 2005 年获得上海交通大学（SJTU）的生物化学学士学位；2010 年获得中国科学院大学（UCAS）生物信息学博士学位。随后，他前往芝加哥丰田技术学院（TTIC）和芝加哥大学与 Jinbo Xu 教授合

作，作为蛋白质结构预测的博士后。他目前的研究主要集中在基于纳米孔第三代测序的算法设计和数据分析。

Wang 博士已经在计算生物学领域发表了 17 篇会议论文，会议包括 ISMB, RECOMB, ECCB 等顶级会议。他还在知名期刊上发表了 42 篇文章，其中包括 1 篇于《PNAS》、1 篇于《Nature Protocols》、1 篇于《Cell Systems》、5 篇于《Nucleic Acids Research》、12 篇于《Bioinformatics》、2 篇于《PLoS Computational Biology》。他有三篇论文被 Web of Science 评为高被引论文，还有一篇论文在突破和创新类别中获得了 2018 年 PLoS 计算生物学研究奖。

二、无线功率和信息传输：非线性模型和波束形成优化

时间：1 月 9 日星期三，下午 13:30-14:30

地点：诚道楼 103

演讲嘉宾：王帅博士

演讲摘要：

无线电力和信息传输 (WPIT) 是一种有前景的技术，可应用于低功率设备例如 RFID 和传感器。然而，WPIT 的有效性取决于能量收集模型的精确度。虽然出于简单性，现有的工作大多使用线性能量收集模型，但真正的能量收集器会包含例如二极管等非线性元件。这导致模型和电路之间的不匹配，并由此导致 WPIT 性能的下降。为了解决上述问题，一种实用非线性能量收集模型被提出，这种模型能非常好地匹配能量收集产品的实验数据。利用非线性模型，两个 WPIT 系统中的波束形成设计得到了进一步的研究。第一个系统中研究了移动充电器的波束形成设计，目的是最大化受给定能量预算影响的数据速率。我们发现通过适当的移动路径，数据速率增益会比固定充电器的情况大。此外，一种用于规划移动充电器路径的分支定界方法也被提出。第二个系统中研究了发射机处的大规模天线阵列下的波束形成设计。为了在短时间内实现高性能，我们导出了仅需要计算梯度的加速一阶算法。理论和数值结果都表明，与现有技术方法相比，所提出的方法减少了计算时间的数量级，同时也保证了相同的性能。

嘉宾介绍：

王帅博士于 2018 年获得香港大学电子电气工程博士学位。目前，他在香港大学担任研究员。

三、为全局优化提供更好的平滑效果：受到罗尔的启发

时间：1月25日星期五，上午 11:00–12:00

地点：诚道楼 207

演讲嘉宾：Jon Lee 教授，密歇根大学

演讲摘要：

在混合整数非线性优化（MINLO）公式的全局优化背景下，我们考虑在零点处导数表现不佳的平滑单变量凹增函数（例如，根函数）。我们的问题是在于如何在不夸大 MINLO 模型和求解器的情况下对 NLP 求解器进行数值求解。在延伸 D'Ambrosio、Fampa、Lee 和 Vigerske，以及 Lee 和 Skipper 的早期工作后，我们给出了一些通用的可检查条件。在这些条件下，我们的平滑是凹陷和增加的、是一个低估者、并且主导一个简单的“移位平滑”。在此过程中，我将提到一些可追溯到 17 世纪后期的历史观点以及当时有争议的微积分主题。这是与 Luze Xu 和 Daphne Skipper 的联合工作。

嘉宾介绍：

Jon Lee 是密歇根大学的 G. Lawton 和 Louise G. Johnson 工程学教授。他从美国康奈尔大学获得博士学位。Jon Lee 是《组合优化的第课》（剑桥大学出版社）和开源书籍《线性优化第一课》（Reex Press）的作者。与此同时，他也是期刊《离散优化》的创始人，并且现任杂志《数学规划》的主编。Jon Lee 曾担任数学优化协会执行委员会主席，以及 INFORMS 优化协会主席。他曾被授予 INFORMS 计算机学会奖，是 INFORMS 院士。

四、概率张量模型中的推论：基础和创新

时间：3月15日星期五，下午 15:00–17:00

地点：诚道楼 207

演讲嘉宾：程磊博士，深圳市大数据研究院研究科学家

演讲摘要：

我们的世界充满了数据。这些数据通常出现在高维结构中，且在高维结构中，每个对象都可多个属性描绘而成。相应的例子可以涉及社会科学、药物、药理学和环境监测等多方面的数据，在此只是略举几例。为了理解多维数据，需要先进的计算工具来找出数据背后隐藏的模式，这就是张量模型发挥作用的地方。由于其卓越的代表能力，张量模型已经在许多领域实现

了最先进的性能，包括社交网络挖掘、图像处理、阵列信号处理和无线通信。然而，先前对张量数据分析的研究主要是从确定性的角度来看，因此不能量化模型的不确定性并控制模型的复杂性，但这些在模型批判和避免过度拟合中本应是非常重要的。幸运的是，贝叶斯建模和推理提供了一种自然的方法，并引发了近期人们对开发贝叶斯张量数据分析的兴趣增长。在本次演讲中，我们将介绍贝叶斯张量数据分析的基础和其最新进展，特别关注于张量规范多元分解模型。

嘉宾介绍：

程磊博士于 2017 年获得香港大学博士学位。2017 年至 2018 年，他在阿里巴巴集团担任算法工程师。2018 年 11 月，他作为研究科学家加入深圳市大数据研究院。他的研究兴趣在于可扩展的贝叶斯推理和张量数据分析。

五、城市交通自动驾驶的导航与预测

时间：3 月 22 日星期五，上午 11:00–12:00

地点：诚道楼 101

演讲嘉宾：Yuexin Ma，香港大学计算机科学博士

演讲摘要：

自动驾驶是一项重大而艰巨、且有可能对人们的日常生活产生深远影响的任务。它的目标是在无需人为干预的情况下，使车辆自动感知环境并在任何交通状况进行安全有效的导航。然后，这项技术在密集的城市环境中面临了一些挑战，因为同一城市道路中会出现不同类型的出行工具，包括汽车、自行车、公共汽车、行人等。自动驾驶系统按顺序依次分为四个模块：感知、预测、导航和控制。对于导航，我们首先为道路中有不同出行工具的情况设计一个碰撞避免算法，它可以为各种出行方式之间提供紧密的交流，并为它们进行实时的无冲突速度计算。其次，我们又为该情况考虑了特殊运动和动态约束，以使流量模拟更加真实。然后，基于 LiDAR 和相机扫描的街景图像，我们利用模拟交通流增强图片的现实感，以创建照片般逼真的模拟图像和渲染，为测试导航的算法提供了良好的模拟器。为了保证自动驾驶的安全性，系统应当具有分析其他交通工具的运动模式并预测其未来的轨迹的能力，以便自动驾驶的车辆能够做出恰当的导航决策。在预测方面，我们提出了一种基于 LSTM 的实时轨迹预测算法。相比现有的技术方法，这种算法的精度提高了 20%。接下来，我们研究了行人的行动模式，以进一步

改善自动驾驶技术的预测能力。为了更好地提升自动驾驶的研究，我们使用采集车来收集城市交通中的交通数据，包括基于 LiDAR 的 3D 点云和交通代理的轨迹。我们的新数据集具有约 150 分钟的连续数据，是一个用于城市街道的大型数据集，可以有助于 3D 检测、3D 跟踪、轨迹预测、运动规划和模拟任务的研究。

嘉宾介绍：

Yuexin Ma 目前是香港大学计算机科学系四年级博士生。他的研究兴趣涵盖人工智能、自动驾驶、机器人、计算机视觉和计算机图形学。他曾在《Science Robotics》、《AAAI》、《AAMAS》、《SIGGRAPH ASIA》、《SGP》、《PCM》、《CSCS》等多家学术期刊发表论文。他曾于 2015 年获得香港博士研究奖学金（HKPFS）、2016 年访问了法国计算机科学与自动化研究所（INRIA）、以及 2017 年访问了北卡罗来纳大学教堂山分校（UNC）。此外，他也曾在百度机器人和自动驾驶实验室进行了实习，并于 2018 年获得优秀个人奖。

六、因果建模与机器学习

时间：3 月 29 日星期五，上午 10:00–11:00

地点：诚道楼 207

演讲嘉宾：张坤教授，卡内基·梅隆大学

演讲摘要：

我们可以通过观察数据的统计分析找到因果结构吗？我们怎样才能找到两个变量之间的关联呢？如何在分布转移的情况下做出最佳预测？我们经常面临诸如此类的建模或预测问题。最近，随着大量数据的快速积累，因果发现——即学习来自纯观察数据的因果信息，以及机器学习，都令人们看到了诱人的机会以及巨大的挑战。

本演讲将集中讨论因果发现的最新进展，以及因果信息如何促进理解和解决从异构数据学习的某些问题。另外，我还将讨论因果发现的基本方法，并解决一些实际问题，包括数据异质性和测量误差的存在。最后，我将会以迁移学习作为例子，讨论在 i.i.d. 假设被取消的情况下，基础因果知识为何以及如何可以帮助我们学习从异构数据中学习。

嘉宾介绍：

张坤教授是卡内基梅隆大学（CMU）哲学系的助理教授、机器学习系的兼职教授、以及德国马克斯普朗克智能系统研究所的高级研究员。他的研究兴趣主要在于机器学习和人工智能，

特别是在因果发现和基于因果关系的学习中。他一直在积极从多种数据中开发开发自动因果发现的方法，并从因果的角度研究机器学习问题，其中包括迁移学习和深度学习。

他已经在解决因果关系中长期存在的问题方面做出了一系列贡献，其中包括如何从观察数据中的结果中辨别原因、如何处理因果分析中的测量误差、以及因果关系和数据异质性如何相互受益。他曾担任多个机器学习或人工智能会议的区域主席或高级计划委员会成员，包括 NIPS、UAI、ICML、AISTATS、AAAI 和 IJCAI。他组织过各种学术活动，以促进因果关系的跨学科研究。

七、大数据可视化：设计，艺术和其他

时间：3 月 27 日星期三，下午 14:00–15:00

地点：诚道楼 103

演讲嘉宾：郭文君，韩国亚洲大学工程学博士

演讲摘要：

在大数据时代，数据分析是增长最快的研究领域之一。数据科学家有必要通过将数据置于视觉环境中来帮助人们理解数据的重要性。如果您将数据视为食品材料，那么数据分析就是用具，而数据可视化则是烹饪方法。为数据可视化选择正确的设计方法对数据科学家来说意义重大。本次讲座将讨论改进工具的设计如何加速人们对数据的探索和理解，以及如何使数据表达巧妙。

嘉宾介绍：

Guo 博士于 2018 年获得韩国亚洲大学博士学位。她的研究主要集中在数据可视化和数据分析。此外，她的研究兴趣还包括创新数据艺术作品，旨在促进艺术、设计、技术和数据可视化之间重大交汇。



编辑：陈怿、王文琪

指导：罗智泉

深圳市大数据研究院

中国广东省深圳市龙岗区龙翔大道 2001 号道远楼 225 室

邮编：518172

电话：（86）755-84273615

<http://sribd.cn/index.php/cn/>