

统计与大模型学术研讨会会议日程安排

第一天（2025 年 10 月 31 日，星期五，南方科技大学专家公寓大堂）

14:00—18:00	签到，晚餐
-------------	-------

第二天（2025 年 11 月 01 日，星期六，商学院 343 会议室）

时间	报告人	主持人	议程
上午			
08:20 - 08:30	荆炳义		致开幕词及参会嘉宾介绍
08:30 - 09:00	严晓东教授 西安交通大学	胡湘红（深圳大学）	Robust Safety Guarantee for Large Language Models via Preference-Augmented Distributional Alignment
09:00 - 09:30	蒋建成讲席教授 大湾区大学		Semiparametric estimation with DNN for unknown nonlinear effects in double sampling designs
09:30 - 10:00	荆炳义讲席教授 港中深&南科大 @河套学院	姜丹丹（西安交大）	Rethinking Intelligence through Embodiment and Autonomous Agents
10:00—10:30	蒋学军研究员 南方科技大学		Cross-Semantic Transfer Learning for High-Dimensional Linear Regression.
10:30 - 10:50	茶歇		休息
10:50 - 12:20	全体参会人员	荆炳义	国自然专题研讨
12:20 - 14:30			午餐 & 休息

参会嘉宾简介

荆炳义 南方科技大学

报告题目: Rethinking Intelligence through Embodiment and Autonomous Agents (通过具身性与自主智能体重新审视智能本质)

报告摘要: 人工智能正改变药物发现流程, 使分子相互作用的预测更加准确与可扩展。本次报告将介绍一种利用扩散模型进行分子对接的新方法, 通过生成并精化配体在蛋白结合位点的构象来实现对接。在标准基准测试上, 实验证明所提出的方法在构象质量与排序指标上均优于当前最先进的对接技术。

嘉宾简介: 荆炳义教授, 香港中文大学(深圳)人工智能学院, 校长永平讲座教授。南方科技大学统计与数据科学系讲席教授, 大数据创新中心主任。国家自然科学奖二等奖, 教育部长江学者讲座教授, 教育部高等学校自然科学奖二等奖, 美国统计学会会士 (ASA Fellow), 数理统计学会会士 (IMS Fellow), 国际统计学会当选会士 (ISI Elected Member)。中国现场统计学会多元分析委员会理事长, 先后分别担任 Ann Appl Probab, JBES, 《中国科学》等国际学术期刊副主编。研究兴趣包括: 概率统计、计量经济、网络数据、机器学习、生物信息, 大模型等领域。在 Ann of Stat, Ann of Prob, JASA, JRSSB, Biometrika, JoE, JBES, Bioinformatics, JMLR, NeurIPS 等顶级期刊及顶会发表论文 130 余篇。与业界有丰富的合作经验, 曾获华为火花奖、华为优秀合作成果奖。

蒋建成 大湾区大学

报告题目: Semiparametric estimation with DNN for unknown nonlinear effects in double sampling designs (基于深度神经网络的双重抽样设计中未知非线性效应的半参数估计)

报告摘要: 本研究针对目标变量的直接观测稀少且获取成本高昂, 而大规模代理数据易于获得的典型问题, 提出了一种高效的回归分析方法。此类“双重抽样 (double-sampling) 设计”在临床研究等领域广泛存在。传统参数方法易受模型设定错误影响, 非参数方法又易陷入维度灾难。为克服上述局限, 我们提出一种结合深度神经网络 (DNN) 与半参数效率理论的统一框架。该方法利用 DNN 从代理数据中学习高维协变量的低维表示, 有效捕捉复杂的非线性关系并缓解维度问题; 在此基础上构建的高效估计量引入稳健的零均值修正项, 保证即使代理模型被误设, 估计仍保持无偏且方差降低。理论分析与实证研究均表明, 该方法能够充分利用大规模代理数据, 在有限的高质量验证样本下显著提升推断的精确性与稳健性。

嘉宾简介: 蒋建成博士毕业于南开大学数学系。2024 年 7 月回国受聘于大湾区大学为讲席教授。回国前是北卡罗来纳大学夏洛特分校数学与统计系和数据科学学院双聘教授, 并兼任该校夏洛特可信人工智能研究中心联合负责人。他聚焦计量经济学与统计学的核心领域, 包括分布式计算、金融时间序列、高维统计学习、非参数平滑、分位数回归和人工智能等方向。迄今已在 Annals of Statistics, JASA, JRSSB, Biometrika, The Econometrics Journal 等顶尖期刊发表 70 余篇同行评审论文。

姜丹丹 西安交通大学

嘉宾简介：姜丹丹现任西安交通大学数学与统计学院教授、博士生导师，陕西省基础科学（数学、物理学）研究院副院长，国家级青年人才计划入选者、陕西省高层次人才引进计划青年项目入选者，主要从事随机矩阵、高维统计分析等理论研究及其在通信领域中的应用研究。研究成果发表在 *Annals of Statistics*、*Biometrika*、*Bernoulli*、*Statistica Sinica* 等统计学权威期刊，含 ESI 高被引论文 1 篇。著有《大维统计分析》、《大维随机矩阵谱理论在多元统计分析中的应用》2 本学术专著。主持国家重点研发计划课题 1 项，主持国家自然科学基金项目 3 项，其中面上项目 2 项、青年项目 1 项，主持省部级科研项目 6 项，主持横向项目 2 项。

严晓东 西安交通大学

报告题目：Robust Safety Guarantee for Large Language Models via Preference-Augmented Distributional Alignment (基于偏好增强分布对齐的大型语言模型稳健安全性保障)

报告人简介：严晓东，西安交通大学数学与统计学院教授，博士生导师，入选国家级青年人才项目和校内青拔 A 类支持计划，荣获“华为火花奖”，“滴滴盖亚学者”，研究方为统计决策、统计推断和统计计算等。学术成果发表在著名期刊 JRSSB, AOS, JASA, JOE 以及人工智能顶级会议 NeurIPS, ICML, AAAI 等 50 余篇。在“高等教育出版社出版”以独立主编出版了《机器学习》、《数据科学实践基础-基于 R》两部教材。

蒋学军 南方科技大学

报告题目：Cross-Semantic Transfer Learning for High-Dimensional Linear Regression (高维回归的跨语义迁移学习)。

报告摘要：现有的高维线性回归迁移学习方法通常假设源域与目标域之间的特征一一对应，从而将适用范围局限于语义匹配的特征情形。然而，在许多实际应用中，源域与目标域的特征虽不相同，却可能在预测作用上具有相似性，形成一种跨语义相似。为充分利用这种更广义的可迁移性，我们提出了跨语义迁移学习 (Cross-Semantic Transfer Learning, CSTL) 框架。该方法通过加权融合惩罚 (weighted fusion penalty) 将目标域的每个回归系数与源域的所有系数进行比较，从而捕捉潜在的跨语义关联。权重由 SCAD 惩罚函数的导数确定，可有效近似理想加权方案，在保留可迁移信号的同时滤除源域特有噪声。在计算实现方面，我们基于交替方向乘子法 (ADMM) 对 CSTL 进行了高效求解。理论上，我们证明在较弱条件下，CSTL 估计量可在高概率下达到理想估计 (oracle estimator) 的性能。仿真研究与真实数据分析结果均表明，CSTL 在存在跨语义或部分信号相似的情形下，均显著优于现有方法。

报告人简介：蒋学军，南方科技大学统计与数据科学系副系主任(教学与本科事务)、长聘副教授、研究员、博士生导师，于 2009 年博士毕业于香港中文大学统计系，2009-2010 在港中文从事博士后研究，2013 年 07 月加入南方科技大学系中南财经政法大学副教授，入选深圳市海外高层次人才孔雀计划 (2016)，深圳市优秀教师 (2018)，主持和完成国家自然科学基金、深圳市基础研究面上项目等 10 余项。其研究方向和兴趣涉及大规模复杂数据统计分析、高维统计推断，金融/应用统计，迁移学习及表征学习，辅助学习等，已在统计学顶级期刊 *Biometrika* 及 *Bernoulli*, *Statistica Sinica*, *Statistics and Computing*, *The Econometrics Journal*，中英文版的中国科学等国际知名学术期刊上发表 SCI&SSCI 论文近 60 篇，授权专利 2 项，及出版英文教材一部。

胡湘红 深圳大学

嘉宾简介：胡湘红，深圳大学教授，本科毕业于华南农大理学院，硕士毕业于中国人民大学统计学院，19 年博士毕业于南方科技大学-香港浸会大学联合博士培养项目，并获 2 校数学系博士学位。2019 年 10 月至 2024 年 5 月在香港科技大学数学系从事博士后工作，2024 年 7 月加入深圳大学任助理教授、教授，24 年入选海外优青。主要研究领域：生物医疗统计、统计机器学习算法、数据科学、人工智能与生命科学交叉研究。目前从事的主要研究包括基于遗传学数据的因果推断，跨种群大规模基因组学数据分析等。近五来在 PNAS, American Journal of Human Genetics (AJHG), Nature Communications 以及 Bioinformatics 等国际重要期刊上发表多篇文章。

王浩枫 香港浸会大学

嘉宾简介：王浩枫，2022 年博士毕业于南方科技大学统计与数据科学系，23 年截至目前于香港浸会大学数学系从事博士后研究、即将以百人计划入职深圳大学数学科学学院，任助理教授。