2024 大数据时代下统计理论和应用研讨会

会议手册

会议日期：2024 年 7 月 26 日-7 月 27 日

会议地点：惠州金海湾喜来登度假酒店

主办单位：南方科技大学统计与数据科学系

目 录

会议介绍

大数据分析在人工智能、生命科学、生物医学，信息科学、资源环境、认知科学、经济金融和人文科学等众多学科中发挥着关键作用。基于对海量数据资源的深度分析挖掘，人工智能在众多领域取得了革命性的发展，但人工智能系统的可解释性、可靠性等根本性问题，需要从理论上解决。这离不开统计学理论的支撑为大数据分析结论的准确性、可解释性、可靠性提供理论基础。为深入探讨大数据时代下统计理论的最新动态和发展趋势，以及探讨“脑动态影像构建和分析中的关键数学问题及其应用”及“基于信息融合和迁移的多元异质数据统计推断”等前沿统计理论和应用，“2024 大数据时代下统计理论和应用研讨会”将于 2024 年 7 月 26 日-27 日在惠州举行，主办单位为南方科技大学统计与数据科学系。本次研讨会特邀相关领域国内外杰出专家学者进行学术报告，深入探讨统计与数据科学及其相关应用领域的科研成果、前沿动态、应用推广，为统计与数据科学领域的专家学者提供学术交流平台。

主办单位

南方科技大学统计与数据科学系成立于 2019 年 4 月，以建设国际一流的教育培养和研究基地为目的。本系志在为国家培养出具有扎实的科学基础，思想活跃，创新意识和能力强，有国际视野，脚踏实地，有朝气、有理想的拔尖人才。本系已经建立起本硕博人才培养体系，拥有统计学和数据科学与大数据技术两个本科专业，及数学学科下概率论和数理统计硕博学位授予权，主要研究领域包括数理统计、生物医学统计、金融统计和数据科学。

目前共有 19 位教研序列教师和 4 位双聘教师，其中有讲席教授 3 人，教授 4 人，副教授 7 人，助理教授 9 人。本系拥有国际化、高水平的师资队伍，包括 1 名国际数学家大会邀请报告人，2 名国家自然科学奖二等奖获得者，1 名长江讲座教授，2 名国际数理统计学会（IMS）会士，1 名 IMS 常务理事，1 名美国统计学会（ASA）会士，1 名 IMS Medallion 讲座演讲者，1 名英国皇家统计学会会士、1 名英国计算机学会会士，1 名广东省特支科技创新青年拔尖人才，1 名深圳市杰出人才培养对象，1 名深圳市高层次国家级领军人才和 2 名深圳市优秀教师。 本系目前正处于一个高速发展时期，师资力量在未来的几年内将大幅度增长。本系诚邀国内外优秀的统计和数据科学学者加入我系的教师行列，同时热忱欢迎优秀的本科生、研究生、博士后到我系学习、研究。

欢迎浏览统计与数据科学系官方网站(https://stat-ds.sustech.edu.en/)以及关注南方科技大学统计与数据科学系“公众号”，获取更多关于本系的最新资讯。

南科大统计系官网二维码　　南科大统计系微信公众号

主要参会人员

大会报告嘉宾

姓名	职称	单位
唐年胜	教授	云南大学
史建清	教授	南方科技大学

邀请报告嘉宾（以演讲时间排序）

姓名	职称	单位
言方荣	教授	中国药科大学
张艳青	教授	云南大学
唐炎林	教授	华东师范大学
赵普映	教授	云南大学
蒋学军	副教授	南方科技大学
李婷	副教授	上海财经大学
康凯	教授	中山大学
周帆	副教授	上海财经大学
贺磊	副教授	安徽师范大学
陈志勇	副教授	福建师范大学
朱慧晨	研究助理教授	香港中文大学
曾鹏程	助理教授	上海科技大学
付金玉	讲师	南京审计大学
陈奇秀	研究生	南京林业大学

学术委员会

唐年胜 教授 云南大学　　朱仲义 教授 复旦大学

组委会

组委会主席

史建清 教　授 南方科技大学

蒋学军 副教授 南方科技大学

组委会秘书

王达奇 南方科技大学（wangdq3@mail.sustech.edu.cn）

邱　童 南方科技大学（qiut@mail.sustech.edu.cn）

天择教育

曹阜兴 18682092796（住宿、用车）

黄　丽 15814784425（缴费，开票）

会议日程

7 月 26 日（周五）14：00-18：20

时间	会议内容	演讲人	主持人
14:00-14:15	开幕式&合影	朱仲义	蒋学军
14:15-15:00	大会报告：TBA	唐年胜	朱仲义
15:00-15:25	SUDO: A Bayesian Subgroup-Speciffc Utility-Based Dose Testing– Optimization Design for Multi-Dose Randomized Basket Trials	言方荣	林金官
15:25-15:55	Variational Bayesian Logistic Tensor Regression with Application to Image Recognition	张艳青	林金官
15:55-16:20	Conformal Prediction in Non-Exchangeable Data Contexts	唐炎林	夏业茂
16:20-16:40	茶 歇		
16:40-17:05	Orthogonality Specification Testing With Complex Survey Data	赵普映	宗序平
17:05-17:30	Decorrelated forward regression for high-dimensional data analysis	蒋学军	宗序平
17:30-17:55	Combining Experimental and Historical Data for Policy Evaluation	李婷	刘应安
17:55-18:20	Blockwise Mixed Membership Model for Discovering the Clinical Heterogeneity of Parkinson’s Disease	康凯	刘应安
18:20-20:20	晚 宴		

7 月 27 日（周六）8：00-12：00

时间	会议内容	演讲人	主持人
08:00-08:45	大会报告： Analysis of Functional Real-world Data	史建清	蒋建成
08:45-09:10	Directional diffusion models	周帆	岳荣先
09:10-09:35	Optimal designs for active controlled does-response models with asymmetric errors	贺磊	岳荣先
09:35-10:00	Bayesian quantile regression for semiparametric spatial autoregressive models via free-knot splines	陈志勇	秦国友
10:00-10:20	茶 歇		
10:20-10:45	Causal Fairness with Optimal Transport	朱慧晨	柏杨
10:45-11:10	Information-theoretic co-clustering-based models for the integrative analysis of single-cell multi-omics data	曾鹏程	柏杨
11:10-11:35	Estimation of integrated volatility of volatility by the range-based volatility measure	付金玉	曹春正
11:35-12:00	Linguistic-Driven Partial Semantic Relevance Learning for Skeleton-Based Action Recognition	陈奇秀	曹春正
12:00-13:00	午 餐（酒店一楼盛宴西餐厅）		

报告摘要

1.TBA 唐年胜

2. **SUDO: A Bayesian Subgroup-Speciffc Utility-Based Dose Testing–Optimization Design for Multi-Dose Randomized Basket Trials**

言方荣 教授 中国药科大学理学院生物统计系主任，中国南京

Abstract: Phase II basket trials are increasingly adopted, as they enable the concurrent evaluation of multiple tumor types, thereby expediting the drug development process, especially for oncology treatments. With growing evidence supporting the promising efffcacy of lower doses for many novel agents, and in line with the FDA's Project Optimus, a trending practice in oncology phase IIA studies is to combine routine preliminary testing of treatment effects and dose optimization with investigations of multiple doses in a single trial. This paper proposes a novel Bayesian adaptive design for testing and optimizing subgroup-speciffc doses in multi-dose randomized basket trials. To address potential heterogeneity between subgroups, the Bayesian model averaging approach is utilized to adaptively cluster predeffned patient subgroups. A Bayesian hierarchical dynamic linear model is developed to facilitate efffcient information sharing across multiple doses and within speciffc subgroup clusters. Under the Bayesian inference framework, proof of concept for the treatment effect in each subgroup can be established, and the subgroup-speciffc optimal dose can be identiffed based on a utility 1function that quantiffes the trade-off between toxicity and efffcacy. Extensive simulation studies are conducted to evaluate the operating characteristics of the proposed design. The results demonstrate its favorable performance across various scenarios.

嘉宾简介：言方荣，中国药科大学理学院教授，博士生导师，生物统计系主任，生物统计与计算药学研究中心主任。主要研究领域包括：临床试验中的生物统计问题，自适应试验设计，生存分析与肿瘤精准治疗，癌症基因组学分析，生物医药大数据及医疗大数据分析理论及应用等。近年来以第一作者或通讯作者发表学术论文 120 多篇，单篇 SCI 影响因子最高 74.69，代表论文包括医学及肿瘤学顶级刊物 NEJM，Annals of Oncology , Clin Cancer Re (discussion paper), Genome Biology，AJRCCM，生物统计权威刊物 Journal of Statistical Software, The Annals of Applied Statistics，SMMR，JRSSC，Bioinformatics 等，入选江苏省"青蓝工程"中青年学术带头人。现主持国家自然科学基金面上项目 2 项，省部级及横向课题多项，出版肿瘤临床试验方法学专著《肿瘤临床试验贝叶斯设计方法》，主编《Advanced Statistics in Regulatory Critical Clinical Initiatives》（Chapmam & Hall/CRC）。

3. **Variational Bayesian Logistic Tensor Regression with Application to Image Recognition**

张艳青，云南大学数学与统计学院统计系教授

Abstract: Image recognition is an important research topic and has attracted considerable attention in various fields including video surveillance, human-computer interaction, and medical image recognition. Existing recognition techniques often ignore structural information of image data or depend heavily on the sample size of image data. To address the issue, we propose a logistic tensor regression model, and develop a novel variational Bayesian approach to make estimation and classification prediction based on the CANDECOMP/PARAFAC decomposition of tensor coefficients. To incorporate the sparsity of tensor coefficients, we consider the multiway shrinkage priors for marginal factor vectors of tensor coefficients. The key idea of the proposed method is to efficiently combine the prior structural information of tensor and utilize the matricization of tensor decomposition to simplify the complexity of tensor coefficient estimation. The ξ-transformation technique is presented to approximate the lower bound of the joint density function of marginal factors and latent variables. The coordinate ascent algorithm is employed to optimize variational lower bound. Based on variational posteriors for classification prediction, we obtain a predictive density approximation. Simulation studies and two real examples including flower image and chest X-ray image recognition show high classification metrics of the proposed method.

嘉宾简介：张艳青，云南大学数学与统计学院统计系教授，博士生导师，国家自然科学基金海外优青，云南省兴滇英才青年项目获得者，主要从事差分隐私、推荐系统、张量数据分析、缺失数据分析、贝叶斯分析等方面的研究，并在《Journal of Machine Learning Research》、《Journal of the American Statistical Association》等机器学习、统计学重要期刊杂志上发表多篇论文。

4. **Conformal Prediction in Non-Exchangeable Data Contexts**

唐炎林 教授 华东师范大学统计学院

Abstract: Conformal prediction is a distribution-free method for uncertainty quantification that ensures finite sample guarantee. However, its validity relies on the assumption of data exchangeability. In this talk, I will introduce several conformal prediction approaches tailored for non-exchangeable data settings, including clustered data with missing responses, nonignorable missing data, and label shift data. To provide an asymptotic conditional coverage guarantee for a given subject, we propose constructing prediction regions by establishing the highest posterior density region of the target. This method is more accurate under complex error distributions, such as asymmetric and multimodal distributions, making it beneficial for personalized and heterogeneous scenarios. I will present some numerical results to illustrate their effectiveness.

嘉宾简介：唐炎林，华东师范大学统计学院教授，博士生导师，统计学系主任；入选国家高层次青年人才计划（组织部）。2012 年 1 月博士毕业于复旦大学统计系，同年 5 月加入同济大学，2019 年 1 月加入华东师范大学。主要研究方向为分位数回归、高维统计推断、不完全数据统计建模，主持多项国家自然科学基金、上海市自然科学基金，担任 SCI 期刊 Statistica Sinica、Journal of the Korean Statistical Society 的编委。在 Biometrika、JRSSB、PNAS、Biometrics 等发表论文 30 余篇。

5. **Orthogonality Specification Testing With Complex Survey Data**

赵普映（云南大学统计系教授）and Mengmeng Xu

Abstract: In this paper, we consider specification analysis for linear quantile models in the context of complex surveys. An orthogonal projection-based nonparametric test is proposed for the correct specification of a linear conditional quantile function over a continuum of quantile levels. The proposed test statistic not only can be used to assess the validity of post design-based estimation inferences regarding the effect of conditional variable on the distribution of outcomes, but also can successfully incorporate design effects of complex survey data. We derive the limiting distribution of the proposed test statistic under the null and alternative hypothesis. In particular, we show that estimation of the unknown model parameters has no asymptotic impact on the proposed test statistic. To implement the test in practice, we propose a multiplier bootstrap procedure and establish its validity. The performance of the proposed method is evaluated through simulations, and the utility of the methodology is demonstrated by a real-world example.

Keywords: Complex survey design; Empirical processes; Multiplier bootstrap; Orthogonal projection; Quantile regression; Specification tests.

嘉宾简介：赵普映，云南大学数学与统计学院教授、博士生导师，兼任云南省应用统计学会理事长。主要从事缺失数据统计分析、复杂调查数据分析的研究工作。主持国家自然科学基金面上项目、重点项目子课题，云南省杰出青年科学基金项目。获第九届高等学校科学研究优秀成果奖（人文社会科学）青年成果奖、云南省自然科学奖一等奖，入选云南省“高层次人才引进计划”青年人才专项。

6. **Decorrelated forward regression for high-dimensional data analysis**

蒋学军　南方科技大学统计与数据科学系副教授

Abstract: Forward regression (FR) is a crucial methodology for automatically identifying important predictors from a large pool of potential covariates. While forward selection techniques achieve screening consistency in contexts with moderate predictor correlation, this property gradually becomes invalid when dealing with substantially correlated variables—especially in high-dimensional datasets where strong correlations exist among predictors. This challenge is not unique to forward selection methods and is encountered by other model selection approaches as well. To address these challenges, we introduce a novel decorrelated forward (DF) selection framework for generalized mean regression models, including prevalent models, such as linear, logistic, Poisson, and quasi likelihood. The DF selection framework stands out because of its ability to convert generalized mean regression models into linear ones, thus providing a clear interpretation of the forward selection process. It also offers a closed-form expression for forward iteration, to improve practical applicability and efficiency. Theoretically, we establish the screening consistency of DF selection and determine the upper bound of the selected submodel's size. To reduce computa-tional burden, we develop a thresholding DF algorithm that provides a stopping rule for the forward-searching process. Simulations and real data applications show the outstanding performance of our method compared with that of some existing model selection methods.

嘉宾简介: 蒋学军，南方科技大学统计与数据科学系副教授(长聘)，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，2010-2013 任中南财经政法大学副教授，2013 年 07 月加入南方科技大学，入选深圳市海外高层次人才孔雀计划（2016），曾获南方科技大学杰出教学奖（2018），深圳市优秀教师（2018），主持和完成国家（广东省）自然科学基金、深圳市基础研究面上项目等 10 余项。其主要研究方向包括分位数回归、变量选择、假设检验、高维数据降维与分析，金融统计与计量等，已在 Biometrika, Bernoulli, Statistics and Computing, Statistica Sinica, Econometrics Journal 等国际一流统计学及计量经济学期刊上发表 SCI&SSCI 论文 50 余篇并出版英文教材一部。

7. **Combining Experimental and Historical Data for Policy Evaluation**

李婷 上海财经大学统计与管理学院副教授

Abstract: This paper studies policy evaluation with multiple data sources, especially in scenarios that involve one experimental dataset with two arms, complemented by a historical dataset generated under a single control arm. We propose novel data integration methods that linearly integrate base policy value estimators constructed based on the experimental and historical data, with weights optimized to minimize the mean square error (MSE) of the resulting combined estimator. We further apply the pessimistic principle to obtain more robust estimators, and extend these developments to sequential decision making. Theoretically, we establish non-asymptotic error bounds for the MSEs of our proposed estimators, and derive their oracle, efficiency and robustness properties across a broad spectrum of reward shift scenarios. Numerical experiments and real-data-based analyses from a ridesharing company demonstrate the superior performance of the proposed estimators.

嘉宾简介：李婷，上海财经大学统计与管理学院副教授，博士毕业于复旦大学管理学院统计学系。曾赴美国德州大学 MD 安德森癌症中心、香港中文大学统计系，北卡罗来纳大学教堂山分校访问学习。研究方向包括函数型数据、医学基因影像数据分析、分位数回归和因果推断。在 Journal of the American Statistical Association, Annals of Applied Statistics, Statistica Sinica，Biometrics 等统计学期刊发表以及人工智能顶会 Neurips，ICML 上发表过多篇论文。

8. **Blockwise Mixed Membership Model for Discovering the Clinical Heterogeneity of Parkinson's Disease**

康凯 中山大学数学学院统计系教授

Abstract: Current diagnostic landscape for Parkinson's disease (PD) faces formidable challenges due to the heterogeneous nature of disease course, including that (i) the disease progression varies hugely between patients, (ii) various types of motor and nonmotor symptoms exist, and (iii) the time to develop those clinical symptoms differs significantly. To tackle these complexities, we propose a novel blockwise mixed membership model (BM3) to systematically unveil between-patient, between-symptom, and between-time clinical heterogeneity within PD. The key idea behind BM3, which is fundamentally different from conventional mixed membership models, is to partition multivariate longitudinal observed variables into distinct blocks, enabling variables within each block to share a common latent membership while allowing different latent memberships across blocks. Consequently, the heterogeneous PD-related symptoms across time are divided into clinically homogeneous blocks consisting of correlated symptoms and time-dependent visits. Moreover, BM3 assign each patient a subject specific vector characterizing partial membership across latent clusters, thus enabling the intricate description of individualized disease progression. We provide both theoretical and empirical justification for the identifiability and posterior consistency of the unknown blocking structures and model parameters. By applying BM3 to the Parkinson's Progression Markers Initiative (PPMI) data, we advance our comprehension of PD heterogeneity, paving the way for the development of more precise and targeted therapies to benefit patients.

嘉宾简介：康凯，博士，中山大学数学学院副教授。2015 年本科毕业于中山大学数学与计算科学学院，2020 年博士毕业于香港中文大学统计系，2020-2022 年于哥伦比亚大学从事博士后研究。2022 年 7 月加入中山大学数学学院，任副教授，硕士生导师。主要研究方向为潜变量模型，纵向数据与生存数据的联合建模，贝叶斯分析等。

9.**Analysis of Functional Real-world Data**

史建清， 南方科技大学教授、生物统计中心主任

Abstract: Data collected from wearable devices , EEG and various other types of data have been extensively utilized for health monitoring and disease diagnosis. However, analyzing such types of data poses a significant challenge. Take, for instance, the free living gait data typically collected over a 7-day period using AX6 devices. This type of data presents the following characteristics: 1) Data is in an arbitrary format, consisting of multivariate functional data of varying lengths and sizes, unique to each subject; 2) There is substantial heterogeneity, with features not only differing between subjects but also varying over time for the same subject; 3) The data exhibits a complex structure, with both mean and covariance structures potentially changing over time; 4) The dataset is of considerable size. In this talk, I will illustrate how our team has tackled the analysis of such data through examples we have worked on.

嘉宾简介：史建清，南方科技大学统计与数据科学系教授，理学院生物医学统计中心主任，英国皇家统计学会会士，科技部十四五重点项目主持、首席科学家。曾任英国国家艾伦图灵研究院图灵研究员，剑桥大学牛顿学院访问研究员，英国纽卡斯尔大学（Newcastle University）统计学教授，纽卡斯尔大学云计算和大数据研究中心副主任。主要研究方向包括函数型数据分析，生物医学统计，缺失数据分析，meta-analysis 等。在国际学术刊物上发表高水平学术论文 100 多篇，包括统计和医学顶级期刊 JRSSB, JASA, Biometrika, Nature Medicine 和 British Medical Journal。J. of Computational and Graphical Statistics 副主编, 曾任英国皇家统计协会《应用统计》(JRSSC)等国际期刊副主编，Guest AE for JRSS discussion paper。获 IEEE 康复游戏和健康国际年会最佳论文奖、美国统计协会非参数统计分会年度最佳论文奖。在 Chapman & Hall 出版专著：Gaussian Process Regression Analysis for Functional Data。

10. **Directional diffusion models**

周帆 上海财经大学统计与管理学院副教授

Abstract： Diffusion models have achieved remarkable success in diverse domains such as image synthesis, super-resolution, and 3D molecule generation. Surprisingly, the application of diffusion models in graph learning has garnered little attention. In this paper, we aim to bridge this gap by exploring the use of diffusion models for unsupervised graph representation learning. Our investigation commences with the identification of anisotropic structures within graphs and the recognition of a crucial limitation in the vanilla forward diffusion process when dealing with these anisotropic structures. The original forward diffusion process continually adds isotropic Gaussian noise to the data, which may excessively dilute anisotropic signals, leading to rapid signal-to-noise conversion. This rapid conversion poses challenges for training denoising neural networks and obstructs the acquisition of semantically meaningful representations during the reverse process. To overcome this challenge, we introduce a novel class of models termed directional diffusion models. These models adopt data-dependent, anisotropic, and directional noises in the forward diffusion process. In order to assess the effectiveness of our proposed models, we conduct extensive experiments on 12 publicly available datasets, with a particular focus on two distinct graph representation learning tasks. The experimental results unequivocally establish the superiority of our models over state-of-the-art baselines, underscoring their effectiveness in capturing meaningful graph representations. Our research not only sheds light on the intricacies of the forward process in diffusion models but also underscores the vast potential of these models in addressing a wide spectrum of graph-related tasks.

嘉宾简介：周帆，上海财经大学统计与管理学院副教授，博士毕业于美国北卡罗莱纳大学教堂山分校。主要研究方向包括强化学习，深度学习，因果推断。在 Journal of American Statistical Association，Journal of Machine Learning Research, NeurIPS, ICML, KDD 等统计学机器学习期刊和会议发表数十篇文章，曾获得泛华统计协会国际会议新研究者奖，北卡教堂山分校 Barry H. Margolin Award，并入选多个上海市青年人才计划。

11. **Optimal designs for active controlled does-response models with asymmetric errors**

贺磊 安徽师范大学副教授

Abstract: The optimal design of dose-finding studies with an active control has only been investigated in the literature for regression models that arise from an accurate error distribution. In this paper, we mainly determine optimal designs for estimating the smallest dose achieving the same treatment effect as the active control from the perspective of the second-order least squares estimation (SLSE). This estimation method has been demonstrated to be asymptotically more efficient than the ordinary least squares estimation (OLSE) when the error distribution is asymmetric. More precisely, we develop optimal design theory based on the SLSE, including equivalence theorems for \phi_p-optimal designs and a geometric characterization of \tilde{c}-optimal designs, which are then illustrated with examples. Furthermore, we investigate the finite sample properties of optimal designs and compare the SLSE with OLSE through simulation, numerical results show that the variance reduction of the SLSE is quite significant when optimal designs based on the SLSE are used for certain situations. Finally, the sensitivity of optimal designs to parameter mis-specifications has also been discussed.

嘉宾简介：贺磊，安徽师范大学副教授，硕士生导师。2019 年博士毕业于上海师范大学，师从上海师范大学岳荣先教授。主要从事贝叶斯分析、试验设计等方面的研究工作，主持国家自然科学基金青年项目和安徽省自然科学基金青年项目各 1 项，在 Bayesian Analysis、Journal of Statistical Planning and Inference、Statistical Papers 等国内外期刊上发表论文近 20 篇。兼任中国现场统计研究会试验设计分会理事，全国工业统计学教学研究会青年统计学家协会理事。

12. **Bayesian quantile regression for semiparametric spatial autoregressive models via free-knot splines**

陈志勇 福建师范大学数学与统计学院副教授

Abstract: We consider a Bayesian quantile regression (BQR) of the semiparametric spatial autoregressive models for spatial data to improve predicting performance. It can be used to capture the linear and nonlinear effects of covariates on the response at different quantile points. With the nonparametric function approximated by free-knot splines, we develop a Bayesian sampling-based method which can be performed via MCMC approach and design an efficient Metropolis-Hastings within Gibbs sampling algorithm to explore the joint posterior distributions. Monte Carlo simulations show that our estimator not only has robustness for different spatial weights matrices but also perform more better than the quantile regression (QR) and instrumental variable quantile regression (IVQR) estimators in the finite sample at different quantiles. The efficacy of the proposed model and estimation method is demonstrated on a real data application from Boston housing price.

嘉宾简介：陈志勇，福建师范大学副教授。近年来，致力于空间计量模型、非/半参数统计等方面的研究，已公开发表 20 余篇论文，主持国家自然科学青年基金和中国博士后科学基金面上一等资助，并作为骨干参与国家级和省部级项目等 6 项。

13. **Causal Fairness with Optimal Transport**

朱慧晨 香港中文大学统计系研究助理教授

Abstract: Algorithm fairness has been a hot topic in recent years. It is a big concern in many fields, especially healthcare, where machine learning algorithms are increasingly expected to help make clinical decisions. Biased data may lead to a biased decision-making algorithm. Consequently, the unfairness makes the algorithm unreliable. Unfortunately, it might make the whole decision-making system more biased. How to measure fairness has been discussed in many related works. However, the association-based fairness definitions are not always compatible with each other. Some causal fairness definitions are proposed and take the causal path into consideration. Most causal fairness definitions are consistent with other association-related fairness. One step further, principal fairness is a definition that is more applicable in healthcare since it concerns whether those who benefit from the treatment will receive it fairly. In this project, we will use principal fairness as the measurement. In order to handle the high-dimensional dataset in health record and possible biases/shift in the dataset, we intend to use autoencoder, which is a non-linear dimension reduction approach, and optimal transport, which helps us mitigate the biases/shift. Our proposed framework can make contributions to advance the fairness decision-making algorithm in healthcare and be able to handle extensive electronic health record data, such as Medical Information Mart for Intensive Care (MIMIC).

嘉宾简介：朱慧晨，2008-2012 年复旦大学统计系本科，2012-2014 年伊利诺伊大学香槟分校统计系硕士，2014-2019 年哥伦比亚大学生物统计系博士，主要研究领域在分位数回归，机器学习等，在 Biometrics, Biostatistics, Computational Statistics and Data Analysis, Journal of Multivariate Analysis, Statistics in Medicine 等杂志发表论文。

14. **Information-theoretic co-clustering-based models for the integrative analysis of single-cell multi-omics data**

曾鹏程，上海科技大学数学科学研究所助理教授

Abstract： Modern high-throughput sequencing technologies have enabled us to profile multiple molecular modalities from different cell subpopulation and even from the same single cells, providing unprecedented opportunities to assay cellular heterogeneity from multiple biological layers. In this talk, we will present two kinds of data integration models, one for multi-omics data profiled from different single cells and another for multi-omics data profiled from the same single cells, under the same framework of information-theoretic co-clustering. The first kind of models are based on the idea of transfer learning, utilizing the information in one dataset, the source data, to facilitate the analysis of another dataset, the target data. These models use the linked features in the two datasets for effective knowledge transfer. The second kind of model borrows the idea of multi-view learning, and takes into account the biological dependency across different types of genomic features. Our experiments on single-cell genomic datasets demonstrate that all of the proposed models improve the clustering performance, provide biological insights, and are computationally efficient.

嘉宾简介：曾鹏程，上海科技大学数学科学研究所助理教授。2018 年博士毕业于英国纽卡斯尔大学数学与统计学院，2019-2021 年在香港中文大学统计系从事博士后研究工作，之后加入上海科技大学。目前从事计算生物学、统计学和数据科学方向的研究工作，致力于开发新的统计模型和计算工具分析生物医学大数据（主要包括单细胞多组学数据和函数型数据）。已在相关领域的国际权威期刊（包括 Briefings in Bioinformatics, Bioinformatics 和 Journal of Computational and Graphical Statistics）发表论文近十篇。目前主持国家自然科学基金青年基金一项。

15. **Estimation of integrated volatility of volatility by the range-based volatility measure；**

付金玉，南京审计大学讲师

Abstract： The issue of market microstructure noise poses a significant challenge to the estimation of integrated volatility of volatility based on high-frequency data, owing to the accumulation of noise with increasing sampling frequency. In this paper, we provide new statistics to estimate the integrated volatility of volatility by the range-based volatility measure in the presence of microstructure noise as well as in the case where noise and price jumps are present. The associated unfeasible and feasible center limit theorems with convergence rate n^(1/8) for the new estimators are established in both cases. Simulation studies are conducted to check the finite sample performance of the proposed methodology and our findings suggest that if the jumps are large, the proposed method is prior to the integrated volatility of volatility estimator on the basis of returns. Additionally, a real study is used to illustrate the stylized facts of the proposed estimators later

嘉宾简介：付金玉博士，南京审计大学讲师。在《International journal of forecasting》、《Journal of Time Series Analysis》、《Journal of Systems Science & Complexity》、《应用概率统计》等发表数篇文章。主要从事高维高频数据、混频数据、高频数据、Meta 分析等方面的研究。

16. **Linguistic-Driven Partial Semantic Relevance Learning for Skeleton-Based Action Recognition**

陈奇秀, 南京林业大学

Abstract: Skeleton-based action recognition, renowned for its computational efficiency and indifference to lighting variations, has become a focal point in the realm of motion analysis. However, most current methods typically only extract global skeleton features, overlooking the potential semantic relationships among various partial limb motions. For instance, the subtle differences between actions such as "brush teeth" and "brush hair" are mainly distinguished by specific elements. Although combining limb movements provides a more holistic representation of an action, relying solely on skeleton points proves inadequate for capturing these nuances. Therefore, integrating detailed linguistic descriptions into the learning process of skeleton features is essential. This motivates us to explore integrating fine-grained language descriptions into the learning process of skeleton features to capture more discriminative skeleton behavior representations. To this end, we introduce a new Linguistic-Driven Partial Semantic Relevance Learning framework (LPSR) in this work. While using state-of-the-art large language models to generate linguistic descriptions of local limb motions and further constrain the learning of local motions, we also aggregate global skeleton point representations and textual representations (which generated from LLM) to obtain a more generalized cross-modal behavioral representation. On this basis, we propose a cyclic attentional interaction module to model the implicit correlations between partial limb motions. Numerous ablation experiments demonstrate the effectiveness of the method proposed in this paper, and our method also obtains sota results.

嘉宾简介：陈奇秀, 南京林业大学 2023 级硕士研究生。研究方向为骨骼动作识别。2023-2024 年参加研究生数学建模竞赛，获得国赛二、三等奖和省赛二等奖。

会议指南

一、会议时间、地点

1) 报到时间：7 月 25 日 15:30-18:30 惠州金海湾喜来登度假酒店

(地址：广东省惠州市惠东巽寮镇金海路 1 号 电话：0752-8328888)

2) 会议时间：2024 年 7 月 26 日-7 月 27 日

3) 会议地点：惠州金海湾喜来登度假酒店 二楼惠东厅

4) 26 号晚宴地点：惠州金海湾喜来登度假酒店 一楼宴会 3 厅

二、会议日程

7 月 26 日(周五)下午，大会报告，学术报告，详见会议手册 P7-P8

7 月 27 日(周六)上午，大会报告，学术报告，12:00 结束，13:10 离店

三、会议费用

参会代表会议注册费 2200 元，注册费由深圳市天择教育科技有限公司代为收取，并开具发票。

1. 转账汇款 （请备注姓名+学校（发票抬头））
   公司名称：深圳市天择教育科技有限公司
   开户行名称：中国银行股份有限公司深圳人民南路支行
   银行账号：767975397371

2. 请扫描下方的二维码，点击“付款附言”，填写【姓名+学校（发票抬头）】，该信息是开具发票的依据，请您在缴费时，务必填写清楚（发票将会发送至对应的邮箱）。

参会人员的交通、食宿费用自理。

四、会议回执

请在 7 月 22 日前扫码填写在线会议回执，以便准确统计及后续会议安排。

五、会务组联系人：

会议日程：王达奇 18600136066 邱 童 13510203716

住宿、车辆安排：曹阜兴 18682092796

缴费、发票事宜：黄 丽 15814784425

备注：

1）7 月 25 日 深圳宝安机场有接机服务，请在会议回执中填写航班号及落地时间，如有变动请致电 18682092796，25 日 18:30 晚宴。

2）惠东南站暂时未安排接站，建议自行打车前往酒店，距酒店约 33 公里，约 50 分钟；

3）7 月 27 日中午 12:00 休会，中午自助餐后预计 13:10 离开酒店，会务组统一安排大巴送至深圳宝安机场及惠东南站。

4）会议酒店至深圳宝安机场约 128 公里，驾车约 2 小时；至惠东南站 33 公里，打车约 50 分钟；至南方科技大学（深圳北站）需 2 小时 40 分。建议预定返程航班在 17:00 以后。

会议通讯录

姓 名	单 位	邮 箱
柏杨	上海财经大学	baiyang@msg.sufe.edu.cn
曹春正	南京信息工程大学	caochunzheng@nuist.edu.cn
曾鹏程	上海科技大学	zengpch@shanghaitech.edu.cn
陈奇秀	南京林业大学	1509328770@qq.com
陈志勇	福建师范大学	zychen1024@163.com
付金玉	南京审计大学	fujinyu@nau.edu.cn
傅博	复旦大学	"fu@fudan.edu.cn
郝红霞	南京审计大学	hxhao@nau.edu.cn
何海金	深圳大学	hehj@szu.edu.cn
贺磊	安徽师范大学	lhstat@163.com
蒋建成	北卡州立大学 夏洛特分校	jjiang1@charlotte.edu
蒋学军	南方科技大学	jiangxj@sustech.edu.cn
康凯	中山大学	kangk5@mail.sysu.edu.cn
李德旺	惠州学院	1787516277@qq.com
李婷	上海财经大学	tingli@mail.shufe.edu.cn
林金官	南京审计大学	jglin@nau.edu.cn
刘超	深圳大学	liuc36@szu.edu.cn
刘欣	东华大学	liuxin@dhu.edu.cn
刘应安	南京林业大学	"lyastat@163.com
潘灯	华中科技大学	pand@hust.edu.cn
秦国友	复旦大学	gyqin@fudan.edu.cn

邱世芳	重庆理工大学	sfqiu@cqut.edu.cn
邱童	南方科技大学	qiut@mail.sustech.edu.cn
史建清	南方科技大学	shijq@sustech.edu.cn
唐琳	云南大学	lintang@ynu.edu.cn
唐年胜	云南大学	nstang@ynu.edu.cn
唐炎林	华东师范大学	yltang@fem.ecnu.edu.cn
王达奇	南方科技大学	wangdq3@mail.sustech.edu.cn
王浩枫	香港浸会大学	hfwang@hkbu.edu.hk
王晓瑞	南方科技大学	sjxywxr@163.com
夏业茂	南京林业大学	ymxia@njfu.edu.cn
严晓东	山东大学	yanxiaodong@sdu.edu.cn
言方荣	中国药科大学	f.r.yan@163.com
殷鹏	深圳先进技术研究院	peng.yin@siat.ac.cn
岳荣先	上海师范大学	yue2@shnu.edu.cn
张艳青	云南大学	zhangyanqing@ynu.edu.cn
章迎莹	华东师范大学	yyzhang@fem.ecnu.edu.cn
赵慧	云南大学	zhaohuistat@ynu.edu.cn
赵普映	云南大学	pyzhao@you.edu.cn
赵彦勇	南京审计大学	yyzhao@nau.edu.cn
周帆	上海财经大学	zhoufan@mail.shufe.edu.cn
周晓东	上海对外经贸大学	xdzhou@suibe.edu.cn
朱慧晨	香港中文大学	huichenzhu@cuhk.edu.hk
朱仲义	复旦大学	zhuzy@fudan.edu.cn
宗序平	扬州大学	xpzhong@yzu.edu.cn

会议记录页