



数学、计算机与生命科学交叉研究
青年学者论坛

2021年5月15日 云论坛
腾讯会议 ID: 460813776

会议手册

青年学者论坛组委会

2021年05月15日

会议简介

“数学、计算机与生命科学交叉研究”青年学者论坛旨在加强青年学者之间的联系，交流生命科学与计算生物学研究领域的最新成果，了解相关国内外发展动态和研究热点，促进我国数学、计算机与生命科学交叉研究与应用实践的更好发展。前七届论坛已分别于2013年至2019年的每年5月成功举办，获得了与会青年学者的普遍好评。2020年新冠疫情肆虐全球，影响至今。为继续促进数学、计算机与生命科学交叉研究领域青年学者的交流，我们将以线上会议的形式，于2021年5月15日举办第八届“数学、计算机与生命科学交叉研究”青年学者特别云论坛。

此次云论坛将围绕“AI与生命科学”、“单细胞多组学与空间转录组学”两个专题进行。会议将以国内外在数学、计算机与生命科学交叉研究领域取得突出成果的青年科学家的学术报告为主，并邀请演讲嘉宾针对“AI与生命科学”和“单细胞多组学与空间转录组学”领域的研究现状与未来发展趋势进行讨论和展望。

本届论坛由中国科学院数学与系统科学研究院/国家数学与交叉科学中心、中国科学院遗传与发育生物学研究所、中国生物信息学会（筹）主办，中国运筹学学会计算系统生物学会、中国细胞生物学学会功能基因组信息学与系统生物学会、中国科学院青年创新促进会、北京生物信息学研究会协办。会议组织者诚挚邀请从事相关研究的广大青年科学家与研究生参会。

会议主席：

王秀杰 中国科学院遗传与发育生物学研究所

张世华 中国科学院数学与系统科学研究院

指导委员会（按照姓氏汉语拼音排序）：

主席： 陈润生 郭 雷 马志明

委员： 敖 平 卜东波 陈洛南 邓明华 冯建峰 高 琳 巩馥洲 韩敬东 胡晓东
黄德双 李 雷 李 梢 李 霞 李国君 李亦学 林 魁 刘 娟 吕金虎 沈百荣
孙 啸 孙之荣 王亚东 魏冬青 吴家睿 闫桂英 于 军 张成岗 张德兴 张立新
张奇伟 章祥荪 张学工 郑伟谋 周水庚 周天寿 邹秀芬

论坛会务组：

史 庆 党大昌 董康宁 刘星言

时间和地点

2021年5月15日 9:00-16:40

腾讯云会议（会议号：**460813776**）

腾讯直播地址：<https://meeting.tencent.com/l/fPbPDnqxPiov>

会议日程

2021年5月15日 (9:00 am - 4:40 pm)

时间	报告人	报告题目
9:00 - 12:00 AI 与生命科学专场 主持人: 王秀杰		
9:00 - 9:05	孙之荣教授 开幕致辞	
9:05 - 9:35	高欣	Towards accurate biomedical genomics anywhere anytime
9:35 - 10:05	张平	人工智能与老药新用
10:05-10:35	曾坚阳	基于机器智能的分子识别模式解析
10:35 - 10:40 休息、交流		
10:40 - 11:20	林伟	人工智能中的若干理论问题
11:20 - 11:50	张强锋	单细胞基因组学数据整合的人工智能方法
11:50 - 12:00	焦玉霞	中国期刊推介 GPB
12:00 - 1:30 中场休息		
1:30 - 4:10 单细胞多组学与空间转录组学专场 主持人: 张世华		
1:30 - 2:00	徐讯	基于 DNA 纳米球测序芯片的高分辨率大视场空间转录组技术
2:00 - 2:30	李婧翌	Applications of generalized additive models and copulas to single-cell RNA-seq computational method development: PseudotimeDE and scDesign2
2:30 - 3:00	谢志	基于深度学习的单细胞转录组数据缺失值填补方法
3:00 - 3:10 休息、交流		
3:10 - 3:40	蔡军	基于深度学习模型的单细胞组学数据分析
3:40 - 4:10	任仙文	单细胞转录组测序与细胞空间关系重构揭示新冠肺炎免疫特征
4:10 - 4:40	演讲嘉宾专题讨论	
4:40 会议结束		

报告简介

1. Towards Accurate Biomedical Genomics Anywhere Anytime

高欣 沙特阿拉伯阿卜杜拉国王科技大学

Abstract:

Current genetic diagnosis by next-generation sequencing requires a large investment of resources and offers little point-of-care portability. Furthermore, it is unable to detect many types of genetic variations including large deletions, duplications, and balanced translocations that are relevant to human diseases and health. Comparing to other sequencing technologies, Nanopore sequencing owns the advantages of point-of-care (i.e., sequencing anywhere anytime), long reads (i.e., assembly-free to detect various genetic variations), and PCR free (i.e., sample preparation is easy). However, its application is severely limited by a number of challenges, including low base-calling accuracy, lack of training data for AI-based methods, and computational burden on reads mapping. In this talk, I will focus on our efforts on developing computational methods to tackle key open problems in Nanopore sequencing. In particular, I will introduce our recent works on developing a collection of computational methods to decode raw electrical current signal sequences into DNA sequences, to simulate raw signals of Nanopore, and to efficiently and accurately align electrical current signal sequences with DNA sequences. I will further introduce their applications in biomedicine and healthcare.

Biography: Dr. Xin Gao is a professor of computer science in Computer, Electrical and Mathematical Sciences and Engineering Division at King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is also the Associate Director of the Computational Bioscience Research Center (CBRC), Deputy Director of the Smart Health Initiative (SHI), and the Lead of the Structural and Functional Bioinformatics Group at KAUST. Prior to joining KAUST, he was a Lane Fellow at Lane Center for Computational Biology in School of Computer Science at Carnegie Mellon University. He earned his bachelor degree in Computer Science in 2004 from Tsinghua University and his Ph.D. degree in Computer Science in 2009 from University of Waterloo. Dr. Gao's research interest lies at the intersection between computer science and biology. In the field of computer science, he is interested in developing machine learning theories and methodologies related to deep learning, probabilistic graphical models, kernel methods and matrix factorization. In the field of bioinformatics, his group works on building computational models, developing machine learning techniques, and designing efficient and effective algorithms to tackle key open problems along the path from biological sequence analysis, to 3D structure determination, to function annotation, to understanding and controlling molecular behaviors in complex biological networks, and, recently, to biomedicine and healthcare. He has published more than 250 papers in the fields of bioinformatics and machine learning. He is the associate editor of GPB, BMC Bioinformatics, and so on.

2. 人工智能与老药新用

张平 美国俄亥俄州立大学

摘要:

基于传统流程的新药开发是一项高投入、长周期且高风险的研发工作。随着大数据和人工智能技术的不断发展，药物重定位及其相关的创新开发方法逐渐成为一种药物开发的主流战略。本报告将讨论近年来药物重定位的一些新工作，特别是基于药物扰动基因差异表达数据和真实世界数据上的深度学习方法，以及其在冠心病和新冠肺炎上的一些应用案例。

个人简介:

张平博士是俄亥俄州立大学（The Ohio State University）的助理教授，创建并领导医学人工智能（Artificial Intelligence in Medicine）实验室。他是美国医学信息学学会（AMIA）会士，ACM 杰出发言人和 IEEE 高级会员，担任多个医疗信息学权威期刊的编委。他的研究重点是机器学习，数据挖掘及其在生物医学信息学和计算医学中的应用。张博士发表了 60+ 篇论文，并申请了 25+ 项专利。他于 2018 年获得 IBM 发明大师（Master Inventor）称号，并于 2014 年获得 AMIA Marco Ramoni 杰出论文提名。他的个人主页是：<http://pingzhang.net/>

3. 基于机器智能的分子识别模式解析

曾坚阳 清华大学

摘要:

鉴定和预测分子之间的识别模式是生物学和药学领域的重要科学问题。DNA、RNA、蛋白质等生物大分子之间的相互作用是生命活动的基础，而大分子靶点与小分子化合物的相互作用是小分子药物研发的关键。近年来，高通量实验技术和海量生物数据的涌现以及先进人工智能技术的兴起，为深入探索分子间的识别机制提供了一个崭新的机会，但同时也对计算模型提出了新的挑战。围绕从海量生物数据中解析分子间识别模式这一目标，我们从信息整合、特征提取、异构网络等多维度出发，开发了多个先进的机器学习方法用以解析大分子间、小分子-蛋白间的识别模式，理解全基因组尺度的基因表达调控机理，准确预测药物-靶点互作，进而提高药物发现效率。

个人简介:

曾坚阳现任职于清华大学交叉信息研究院，长聘副教授，博士生导师。于 1999 年和 2002 年分别获得浙江大学的学士和硕士学位。2011 年，在美国杜克大学(Duke University)获得计算机科学博士学位。2011 年至 2012 年期间，在杜克大学计算机科学系和杜克医学院从事博士后研究。2012 年作为海外人才被引进清华大学交叉信息研究院。长期致力于人工智能/机器学习和生命科学领域的交叉学科研究。共发表学术论文 70 余篇，其中通讯作者论文包括 Nature Machine Intelligence、Nature Communications、PNAS、Cell Systems、Nucleic Acids Research 等，合作作者论文包括 Nature 期刊等。研究成果获得 ESI 高引论文、2019“吴文俊人工智能自然科学”三等奖、2018 年度和 2019 年度“中国生物信息学十大进展”、2019 年度“中国生物信息学十大算法和工具”、2020 年度世界人工智能大会青年优秀论文、国际会议 ICIBM 2019、PDCAT2005 最佳论文等。担任国际期刊 IEEE/ACM Transactions on Computational Biology and Bioinformatics 的编委。多次在计算生物学领域的国际一流会议 ISMB、RECOMB 担任程序委员会委员。

4. 人工智能中的若干理论问题

林伟 复旦大学

摘要:

面向未来,新一代人工智能发展愿景多瞄向结构拟人、功能仿人的智能个体、群体的构建与应用。这个报告主要向大家介绍,我们正在和准备在这三方面开展的研究工作:(1)新型智能算法创建,以挖掘和预测时空数据的基本特性;(2)高级机器学习算法设计,以创建具有可解释性的学习训练新理论;(3)仿人智能与群智智能的实现,以构建多空间智能体以任务实现为目标的协同演进和调控策略。

个人简介:

林伟博士近年来主要开展:有限与无限维动力系统、复杂和随机网络调控动力学、参数识别与数据同化、因果关系以及这些理论、模型与算法在计算系统生物学、社会科学以及人工智能领域的深度应用等交叉研究工作。在 PRL, PNAS, Nature Comm., Nature Phys., IEEE 汇刊, SIAM 汇刊等上发表学术论文 90 余篇。2011 年获教育部霍英东基金会高校青年教师二等奖,2019 年获世界华人数学家联盟最佳论文奖,2019 年获上海市五一劳动奖章(科技创新类)。于 2019 年获国家杰出青年基金资助,2018 年获选国家重点研发计划重点项目负责人,2013 年获得国家自然科学基金优秀青年基金资助。担任 Research、IJBC 的 AE, Chaos 编辑咨询委员会委员,CSF 的 Editor。

5. 单细胞基因组学数据整合的人工智能方法

张强锋 清华大学

摘要:

Single-cell RNA-seq and ATAC-seq analyses have been widely applied to decipher cell-type and regulation complexities. However, experimental conditions often confound biological variations when comparing data from different samples. For integrative single-cell data analysis, we have developed SCALEX, a deep generative framework that maps cells into a generalized, batch-invariant cell-embedding space. We demonstrate that SCALEX accurately and efficiently integrates heterogenous single-cell data using multiple benchmarks. It outperforms other integration methods, especially for datasets with partial overlaps, accurately aligning similar cell populations while retaining true biological differences. We demonstrate the advantages of SCALEX by constructing continuously expandable single-cell atlases for human, mouse, and COVID-19, which were assembled from multiple data sources and can keep growing through the inclusion of new incoming data. Analyses based on these atlases revealed the complex cellular landscapes of human and mouse tissues and identified multiple peripheral immune subtypes associated with COVID-19 disease severity.

个人简介:

张强锋博士 2006 年在中国科大获得计算机博士学位，主要从事计算复杂性和算法研究。于 2012 年在哥伦比亚大学获得生物物理的第二个博士学位，研究领域为计算结构生物学。随后在斯坦福大学从事基因组学博士后研究。2015 年加入清华大学；实验室致力于结构生物学、基因组学、人工智能和大数据交叉领域研究。在 RNA 结构研究方面，开发了细胞内 RNA 结构高通量解析新技术，并应用于解析新冠病毒等 RNA 病毒基因组结构图谱，发现并验证了病毒 RNA 保守结构对其传播的作用。实验室还致力于开发结构生物学人工智能新算法。以通讯作者身份发表 Cell 等杂志学术文章多篇。

6. 基于 DNA 纳米球测序芯片的高分辨率大视场空间转录组技术

徐讯 深圳华大基因研究院

摘要:

从分子表达精度解析组织的空间结构对于理解生命的复杂性非常重要。当前的技术在分辨率和视场方面都存在比较大的局限,限制了空间组技术的广泛应用。我们结合 DNA 纳米球 (DNB) 微阵列芯片和原位 RNA 捕获技术,开发了空间转录组测序技术 Stereo-seq。这种方法可以以前所未有的纳米级分辨率对组织切片进行高通量转录组分析,视场可扩展至厘米级。我们基于 Stereo-seq 技术构建了脑图谱和小鼠胚胎发育时空图谱,并在组织边界、细胞类型鉴定、细胞互作、基因互作、细胞命运等方面进行了系统解析,同时我们提出时空组学未来在海量数据计算和存储、细胞分辨率解析、空间组 3D 重构、细胞互作和命运调控解析等方向的算法挑战。

个人简介:

徐讯,博士,研究员。现任华大集团首席执行官、深圳华大生命科学研究院院长,国际标准化组织/生物技术委员会副主席 (ISO/TC276), ISBER (国际生物及环境样本库协会)中国区主席、世界经济论坛未来理事会委员、全国生物样本标准化委员会 (SAC/TC559) 专家委员。具体研究方向包括测序仪及相关技术开发,单细胞测序技术,时空组测序技术(stereo-seq),合成仪及相关技术,以及测序合成技术在合成生物学、疾病诊疗和农业等方向的应用转化研究。目前已发表在包括《自然》、《科学》、《细胞》等国际顶级科学杂志在内的 SCI 收录论文 246 篇,其中,第一作者或并列第一作者 15 篇,通讯作者 41 篇,近五年引用次数超过三万余次。主持和参与包括国家 863 计划国产测序仪项目课题和发改委产业集聚等项目 18 项。获得专利 27 项,另有 75 多项正在申请中。曾荣获“科技部大挑战青年科学家”、“鹏城杰出人才奖”、“广东省科学技术奖”一等奖、“教育部自然科学奖”二等奖。入选国家百千万人才工程“有突出贡献中青年专家”荣誉称号。

7. Applications of generalized additive models and copulas to single-cell RNA-seq computational method development: PseudotimeDE and scDesign2

李婧翌 美国加州大学洛杉矶分校

Abstract:

Part 1: PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. To investigate molecular mechanisms underlying cell state changes, a crucial analysis is to identify differentially expressed (DE) genes along the pseudotime inferred from single-cell RNA-sequencing data. However, existing methods do not account for pseudotime inference uncertainty, and they have either ill-posed p-values or restrictive models. Here we propose PseudotimeDE, a DE gene identification method that adapts to various pseudotime inference methods, accounts for pseudotime inference uncertainty, and outputs well-calibrated p-values. Comprehensive simulations and real-data applications verify that PseudotimeDE outperforms existing methods in false discovery rate control and power.

Part 2: scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. A pressing challenge in single-cell transcriptomics is to benchmark experimental protocols and computational methods. A solution is to use computational simulators, but existing simulators cannot simultaneously achieve three goals: preserving genes, capturing gene correlations, and generating any number of cells with varying sequencing depths. To fill this gap, we propose scDesign2, a transparent simulator that achieves all three goals and generates high-fidelity synthetic data for multiple single-cell gene expression count-based technologies. In particular, scDesign2 is advantageous in its transparent use of probabilistic models and its ability to capture gene correlations via copulas.

Biography: Jingyi Jessica Li is an Associate Professor in the Department of Statistics (primary) and the Departments of Biostatistics, Computational Medicine, and Human Genetics (secondary) at University of California, Los Angeles (UCLA). She is also a faculty member in the Interdepartmental Ph.D. Program in Bioinformatics. Prior to joining UCLA in 2013, Jessica obtained her Ph.D. degree from the Interdepartmental Group in Biostatistics at University of

California, Berkeley, where she worked with Profs. Peter J. Bickel and Haiyan Huang. Jessica received her B.S. (summa cum laude) from the Department of Biological Sciences and Technology at Tsinghua University, China in 2007. Jessica and her students focus on developing statistical and computational methods motivated by important questions in biomedical sciences and abundant information in big genomic and health related data. On the statistical methodology side, her research interests include association measures, asymmetric classification, and high-dimensional variable selection. On the biomedical application side, her research interests include bulk and single-cell RNA sequencing, comparative genomics, and information flow in the central dogma. Jessica is the recipient of the Hellman Fellowship (2015), the PhRMA Foundation Research Starter Grant in Informatics (2017), the Alfred P. Sloan Research Fellowship (2018), the Johnson & Johnson WiSTEM2D Math Scholar Award (2018), the NSF CAREER Award (2019), the UCLA DGSOM Keck W. M. Keck Foundation Junior Faculty Award (2020), and the MIT Technology Review 35 Innovators Under 35 China (2020).

8. 基于深度学习的单细胞转录组数据缺失值填补方法

谢志 中山大学

摘要:

单细胞测序技术实现了在单个细胞层面上对基因表达的量化研究, 广泛用于细胞异质性、肿瘤细胞进化等研究中。随着实验技术的迅速发展, 一次实验的细胞数可以达到 1 万甚至超过 10 万量级。但是由于单细胞的 RNA 含量低, 只有少量基因表达能检测到, 导致生物信号模糊, 后续分析困难。现有的基于传统统计模型的单细胞转录组缺失数据填补工具例如 SAVER, MAGIC 等无法处理大数据量, 而基于深度学习的填补工具 scScope 等虽然可以处理几十万甚至百万级别的细胞数, 但是效果往往不及传统方法。我们提出一个不假设基因表达分布的混合深度学习模型结构(DISC), 对填补后的表达值进行约束, 并借助深度神经网络对复杂函数的拟合能力来自适应单细胞表达分布。DISC 能够恢复多套数据的基因结构。进一步实验发现, DISC 能够有效找回降采样所造成的细胞类型丢失。最后, 在 130 万细胞数的小鼠大脑数据集上, DISC 也能够准确的找到主要的细胞类型。DISC 能够为处理和分析百万量级的单细胞数据提供重要的技术支撑。

个人简介:

谢志、教授, 中山大学健康医疗大数据国家研究院副院长、中山大学中山眼科中心眼科研究所副所长、中山大学学术委员会委员、广东省生物信息学会理事长。重庆医科大学临床医学本科, 新西兰林肯大学应用计算专业一级荣誉学士以及计算系统生物学博士学位, 美国约翰霍普金斯大学医学院博士后, 美国国家癌症研究院研究员。2013 年入职中山大学。研究方向为生物医学大数据和人工智能。主持国家自然科学基金海外合作基金、科技部重大平台等项目, 在《Cell》, 《Nature Methods》等国际期刊发表学术论文 47 篇, 总影响因子超过 450 分。

9. 基于深度学习模型的单细胞组学数据分析

蔡军 中国科学院北京基因组研究所

摘要:

单细胞转录组作为单个细胞的特征，能更加精确地定义细胞的类型。常规的基于单细胞转录组的分类方法首先是进行无监督的聚类，然后根据每个集群（Cluster）特异表达的细胞标记基因来对集群进行标注。虽然基于无监督的分类方法更容易发现新细胞类型，但人工标注的过程费时费力。目前已有的基于监督学习的自动分类方法，大部分无法兼顾到方法的解释性以及新细胞类型的发现。我们构建了决策过程可解释的深度学习网络模型，单细胞胶囊网络(single cell Capsule Network, scCapsNet)，模型使深度学习网络的决策黑箱透明化、可解释；实现了单细胞组学数据中细胞亚型的准确推断和基因调控模块的深度挖掘。相对于其他单细胞转录组自动分析工具，单细胞胶囊网络能更稳定更高效地分辨出属于新细胞类型的细胞。并且，单细胞胶囊网络能通过模型的内部参数找出细胞类型相关基因。通过细胞类型相关基因，单细胞胶囊网络能将基因与细胞类型直接联系起来，极大地提高了深度学习模型的可解释性。本质上，单细胞胶囊网络将基因的表达特征和细胞类型特征进行低维编码，这样的编码富含生物学意义。

个人简介:

蔡军，中国科学院北京基因组研究所（国家生物信息中心），中国科学院精准基因组医学重点实验室，研究员；中国科学院大学，存济医学院，岗位教授；博士生导师。致力于将计算生物学与基因组学相结合，发展和利用包括单细胞组学在内的新型组学技术和数据分析方法，揭示肿瘤基因组和早期胚胎细胞基因组中的遗传和表观遗传变异，研究肿瘤发生和胚胎发育过程中的细胞群体异质性和命运决定。在 Nature Machine Intelligence、PNAS、Nature Communications、Stem Cell Reports、Nucleic Acids Research 和 Bioinformatics 等国际期刊发表多篇通讯作者论文。承担科技部、自然科学基金多项科研项目，科技部国家精准医学重点研发计划“基因组学与表观遗传信息深度挖掘”课题负责人。

10. 单细胞转录组测序与细胞空间关系重构揭示新冠肺炎免疫特征

任仙文 北京大学

摘要:

新冠病毒给全球带来了巨大灾难,揭示其感染与免疫机制对于治疗与防控都有重要意义。基于单细胞转录组测序(scRNA-seq)和细胞自组织的空间关系重构算法CSOmap,本研究揭示了新冠病毒感染的新宿主细胞类型、不同上皮细胞引发的不同免疫响应,并解析了外周血免疫组成与疾病严重程度、发病阶段、年龄、性别等临床指标的关联,鉴定了介导新冠肺炎过度免疫反应的关键配体-受体分子和细胞因子风暴的潜在源头,为理解新冠病毒如何启动宿主免疫系统和治疗新冠肺炎提供了新机制和新靶点。

个人简介:

任仙文,副研究员,国家自然科学基金优秀青年项目获得者。2004年毕业于南开大学,获生物学、信息与计算科学双学位;2007年毕业于军事医学科学院,获遗传学硕士学位;2010年毕业于中国科学院数学与系统科学研究院,获得运筹学与控制论博士学位。2010年至2013年任职于中国医学科学院病原生物学研究所,致力于基于宏基因组测序的病原体检测分析。2016年任职于北京大学生物医学前沿创新中心(BIOPIC),致力于单细胞转录组测序的生物信息学分析和肿瘤免疫的研究,针对大规模单细胞测序数据的聚类、注释、时空重建等重要前沿生物信息学问题提出了关键技术创新,在Cell、Nature、Nature Medicine、Cell Research、Nature Communications等国际知名杂志上发表论文50多篇。