



软件学院20周年
软件学科50周年

山东大学软件学院“强院兴校、特软启航”

院庆系列学术活动

智能生物信息处理与分析论坛

特邀专家 潘毅 深圳理工大学

高琳 西安电子科技大学

黄德双 同济大学

蔡禄 内蒙古科技大学

厉力华 杭州电子科技大学

王吉华 德州学院

郭茂祖 北京建筑大学

承办单位：山东大学软件学院

山东大学人工智能国际联合研究院

中国计算机学会济南分部

时间：2021年9月25日

地点：山东大学软件园校区软件学院学术报告厅



山东大学软件学院“强院兴校、特软启航” 院庆系列学术活动

智能生物信息处理与分析论坛

随着生物技术的迅猛发展，海量的生物信息大数据快速积累，为研究生命的产生和发展规律提供了良好的契机。但是，生物信息大数据通常具有样本少、维度高、数据结构复杂、数据种类多、数据缺失等鲜明特点，导致了对数据的解读速度远远落后于数据的产出速度。因此，如何有效的融合智能信息处理、生物学理论以及统计学方法，建立强大的理论方法和实用工具，智能高效地处理复杂生物数据，已经成为以数据为驱动的智能生物信息处理与分析学科领域的发展重点和研究热点。

本次活动，我们邀请了来自国内外多所高校的领域知名学者，分享智能生物信息处理最新研究成果，共议智能生物信息处理与分析的未来发展方向，必将促进人工智能+生物医学多学科之间的交叉融合。

2021年，适逢山东大学建校120周年、山东大学软件学科发展50周年、国家示范性软件学院建院20周年华诞。本次学术研讨会围绕“强院兴校、特软启航”的院庆主题举办，是庆祝山东大学软件学院院庆系列学术活动之一。在此，谨向长期以来关心支持软件学院发展的各界朋友致以崇高敬意，向海内外校友致以亲切问候！

日程

时间	主题	主讲人
8:30-8:35	开幕词	刘士军(山东大学)
8:35-8:40	欢迎词	仇道滨(山东大学)
8:40-9:20	人工智能在疾病预测中的应用	潘毅 (深圳理工大学)
9:20-10:00	Computational problem in single-cell omics data analysis	高琳(西安电子科技大学)
10:00-10:20	茶歇	
10:20-11:00	DNN based computational analyses for motifs in biological binding sequences	黄德双 (同济大学)
11:00-11:40	Nucleosome assembly and disassembly in vitro are governed by chemical kinetic principles	蔡禄(内蒙古科技大学)
14:10-14:50	Radiogenomics and its clinical implications	厉力华(杭州电子科技大学)
14:50-15:30	固有无序蛋白 (IDPs) 与功能性长非编码 RNA	王吉华 (德州学院)



报告人简介：潘毅教授，美国医学与生物工程院院士、乌克兰工程院院士、英国皇家公共卫生学院院士、英国工程技术学会会士、日本学术振兴会会士和国家特聘专家，目前担任中国科学院深圳理工大学（筹）计算机科学与控制工程学院院长。他以江苏省理科状元考入清华大学计算机科学与工程系，1982年和1984年分别获得清华大学工学学士学位和硕士学位，1991年获得美国匹兹堡大学计算机科学博士学位。他曾是美国乔治亚州立大学计算机科学系主任、生物系主任、文理学院副院长、州校董教授、校级杰出教授，并担任过清华大学、北京大学、浙江大学等高校访问讲席教授或客座教授。

潘毅教授的主要研究领域是以云计算、大数据分析、人工智能、深度学习等为工具，进行生物信息和医疗信息的研究。在此领域已发表250多篇SCI期刊学术论文，其中100多篇发表在顶尖的IEEE/ACM Transactions/Journals学术期刊上；另在国际学术会议录上发表150多篇学术论文，出版编著了40多本书。他应邀在60多个国际大会上作了大会主题演讲，并在美国和许多世界著名大学作了近百个学术报告。潘毅教授目前是《Big Data Mining and Analytics》的主编，生物信息学著名杂志《IEEE/ACM TCBB》和中国计算机英文杂志《JCST》的副总主编，曾是John-Wiley《生物信息学系列丛书》与John-Wiley《无线网络和移动计算系列丛书》的创办人兼主编。他担任或担任过七种IEEE Transactions期刊副编辑以及十多种国际期刊的编委，已在几十个重大国际大会上任大会总主席和程序委员会主席。

报告题目：人工智能在疾病预测中的应用

报告摘要： AI has been successfully used in many applications such as computer vision, automatic speech recognition, natural language processing, audio recognition, and medical imaging processing and disease diagnosis. Recently, our group has designed a method to detect a type of epilepsy-benign epilepsy with centrotemporal spikes, which is the most popular epilepsy with children. In our method, we use three sources of data: hand-crafted features from MRI images based on doctors' knowledge, 3D MRI images and 4D functional MRI images. The final prediction decision is obtained by fusing the three prediction results through another neural network. Our idea is to take advantages of all three data sources which have different strengths and important features to achieve the best prediction results. We have done many experiments which show that the proposed method is truly better than any existing prediction method. Currently, we are working on predicting Autism disease through video data (including facial expression, movement style, and voice), genetic features, and MRI images through effective fusion methods. Future improvements including how to design better AI systems, and how to use more data sources and how to effectively fuse the data will also be outlined in this talk.



报告人简介：高琳教授，西安电子科技大学计算机科学与技术学院二级教授，博士生导师，中国计算机学会“生物信息专委会”副主任，人工智能学会理事，人工智能学会“生物信息学与人工生命专业委员会”副主任，运筹学会“计算生物信息学分会”副理事长。在计算生物信息学、数据挖掘与机器学习、图论与组合优化算法方面进行了长期研究，承担国家自然科学基金重点项目、重大研究计划等项目，参与科技部精准医疗重点专项。研究成果获陕西省自然科学技术一等奖，电子学会自然科学技术二等奖，单细胞环状伪轨迹推断算法入选“2019年中国生物信息学十大算法与工具”，在《Nature Communications》，《Science Advances》，《Advanced Science》，《Nucleic Acids Research》，《PLoS Computational Biology》，《Bioinformatics》等期刊发表论文 200 余篇。

报告题目： Computational Problem in Single-Cell Omics Data Analysis

报告摘要： The rapid development of single-cell multi-omics sequencing technology has made it possible to explore cells in multiple dimensions (genes, transcriptions, epigenetics, and even spatial), which has been highly valued by life science research and has posed new challenges to bioinformatics research. How to deeply understand the cellular function described by multi-omics data is a difficult problem and how to achieve feature complementarity of multi-omics data poses a computational challenge. In this talk, I will discuss the related computational problem including cell type, cell communication and the multi-omics data integration.



报告人简介：黄德双教授，同济大学电子与信息工程学院特聘教授、博士生导师。长期从事神经网络、模式识别与生物信息学方面的研究，在国内外等学术期刊上发表了超过 230 篇 SCI 论文；曾荣获教育部和安徽省自然科学一等奖各 1 项、人工智能学会科技进步一等奖 1 项；担任国家科技创新 2030—新一代人工智能重大项目“面向复杂数据处理的新型神经网络模型研究”项目首席专家；担任期刊 IEEE/ACM Transactions on Computational Biology & Bioinformatics 等杂志编委。他也是 IEEE Fellow，国际模式识别学会 (IAPR) Fellow，2000 年度中科院“百人计划”入选者，中国计算机学会生物信息学专业委员会副主任委员。

报告题目：DNN based Computational Analyses for Motifs in Biological Binding Sequences

报告摘要： Transcription factors (TFs) are proteins that bind to specific DNA sequences to regulate the expression of the genome. The identification of transcription factor binding sites (TFBSs) is extremely important for understanding the process of gene expression. On the one hand, classic deep learning methods for TFBSs prediction usually fail to capture the dependencies between genomic sequences since their commonly used one-hot codes are mutually orthogonal. On the other hand, these methods usually perform poorly when samples are inadequate. To address these two challenges, we developed a novel language model for mining TFBSs using human genomic data and ChIP-seq datasets, named Transcription Factors Fine-tune Language Model (TFFLM). First, we compared TFFLM with the existing widely used methods on 69 datasets and we achieved the state-of-the-art performance. Moreover, we conducted comparative experiments on complex TFs such as POLR2A and small data set such as KMD5A and the results show that TFFLM still achieved a significant improvement. Finally, through visualization analysis of one-hot encoding and TFFLM, we found that one-hot encoding completely cut off the dependencies of DNA sequences themselves, while TFFLM using language models can well represent the dependency of DNA sequence.



报告人简介: 蔡禄教授, 内蒙古科技大学生命科学与技术学院教授, 博士生导师。内蒙古自治区“功能基因组生物信息学重点实验室”主任, “表观遗传学与生物信息学科技创新团队”负责人。开展生物信息学、表观遗传学等领域研究。曾从事研究方向有: 基因序列的信息学分析、DNA 结构、蛋白质-蛋白质相互作用等。

目前主要关注基因表达在表观遗传学层面的调控机制, 开展核小体定位、组蛋白修饰、RNA 可变剪接等问题研究。获国家自然科学基金三等奖 1 项, 内蒙古自治区自然科学奖一等奖 1 项, 自治区科技进步奖一等奖、二等奖各 1 项, 获自治区优秀教学成果奖一等奖 2 项。主持国家自然科学基金项目 6 项, 发表 SCI 收录论文 50 余篇。出版《表观遗传学前沿》学术专著 1 部, 《生物信息学》教材 3 部。中国细胞生物学学会功能基因组信息学与系统生物学分会委员、中国生物工程学会生物信息学与计算生物学分会委员。

报告题目: Nucleosome assembly and disassembly in vitro are governed by chemical kinetic principles

报告摘要: As the elementary unit of eukaryotic chromatin, nucleosomes in vivo are highly dynamic in many biological processes, such as DNA replication, repair, recombination or transcription in order to allow the necessary factors to gain access to their substrate. The dynamic mechanism of nucleosome assembly and disassembly has not been well described so far. We proposed a chemical kinetic model of nucleosome assembly and disassembly in vitro. In the model, efficiency of nucleosome assembly is positively correlated with total concentration of histone octamer, reaction rate constant and reaction time. All the corollaries of the model were well verified for Widom 601 sequence and the six artificially synthesized DNA sequences, named as CS1-CS6, by using salt dialysis method in vitro. The reaction rate constant in the model may be used as a new parameter to evaluate the nucleosome reconstitution ability with DNAs. Nucleosome disassembly experiments for Widom 601 sequence detected by Förster resonance energy transfer and fluorescence thermal shift assays demonstrated that nucleosome disassembly is the inverse process of assembly and can be described as three distinct stages including opening phase of the (H2A-H2B) dimer/(H3-H4)₂ tetramer interface, release phase of the H2A - H2B dimers from (H3-H4)₂ tetramer/DNA and removal phase of the (H3-H4)₂ tetramer from DNA. The present work confirms that nucleosome assembly and disassembly in vitro are governed by chemical kinetic principles.



报告人简介: 厉力华教授, 国家杰出青年基金获得者、“新世纪百千万人才工程”国家级人选。浙江省科技发展咨询委员会委员、之江实验室学术咨询委员会委员、中国生物医学工程学会图像信息与控制专业委员会副主任委员(候任主任委员)、中国图象图形学会医学影像专业委员会副主任委员。主要从事医学影像和智能生物医学研究, 2002 年获中国侨界贡献奖(创新人才)。发表包括 Nature 子刊在内的科研论文和国外著作章节 200 多篇(章), 获发明专利 20 多项, 其中包括两项美国专利。现主持国家自然科学基金重点项目、国家重点研发计划战略性科技创新合作专项、国家重点研发计划数字诊疗专项课题, 合作开展国家自然科学基金委员会-浙江省人民政府联合基金重点项目和国家重点研发计划变革性技术专项等研究工作。

报告题目: Radiogenomics and its Clinical Implications

报告摘要: 近年来, 随着基因测序和医学影像技术的发展, 作为影像学与生物组学相交叉所衍生出来的影像基因组学成为一个新的热点研究方向。本报告以肿瘤精准诊疗为背景, 从临床的视角解析影像基因组学研究的生物学意义和临床价值。希望为从事该领域研究的同行在课题设计和研究开展上提供一点参考。



报告人简介：王吉华教授，德州学院二级教授，山东师范大学联合博士生导师，山东省生物物理重点实验室/山东省猪群健康大数据及智能监测工程实验室主任，德州学院生物物理研究院名誉院长；国务院特殊津贴专家，山东省教学名师；中国生物物理学会理事，中国生物物理学会生物相分离与相变分会理事，山东省物理学会生物物理专业委员会主任，科技部、国家基金委项目评审专家。主要从事计算生物和纳米生物传感技术研究，已在 *Nature Communication*, *Nucleic Acids Research*, *Biophys J* 等发表论文 80 余篇，多次应邀在国内外学术会议做报告。主持国家自然科学基金面上课题 4 项，联合承担完成国家自然科学基金委重点课题 1 项、重大研究计划 1 项；主持完成省部级课题 6 项。获得山东省自然科学二等奖、三等奖各 1 次。山东省黄大年式生物物理交叉创新团队负责人、齐鲁劳动奖章获得者等。

报告题目：固有无序蛋白（IDPs）与功能性长非编码 RNA

报告摘要：无序蛋白（IDPs）和长非编码 RNA（lncRNA）均是柔性很大的生物大分子，具有重要生物功能。无序蛋白（IDPs）是一类在天然状态下没有稳定的三维结构，却具有重要生物功能的柔性很大的蛋白质。我们分析了 IDPs 的某些序列特征及其相关复合体的构象特征，探索了蛋白质无序程度与物种最适生长温度的关联性，建立了无序蛋白与其它成分相互作用数据库等，加强了对 IDPs 的序列和构象特征、及其功能的认识。长非编码 RNA（lncRNA）发挥着越来越重要的生物功能，与人类重大疾病密切相关。但是，目前发现具有生物功能的 lncRNA 只是冰山一角，而深入系统认识这些实验证实的功能性 lncRNA 的特征性质，对发现新的 lncRNA 及其新功能具有十分重要的意义。因此，经过多年努力我们建立了目前国际上实验验证最全综合性功能性长非编码 RNA 数据库 EVLncRNAs（EVLncRNAs1.0 和 EVLncRNAs2.0（<https://www.sdklab-biophysics-dzu.net/EVLncRNAs2/>）），并基于此数据库发展了潜在功能性长非编码 RNA 预测算法 EVlncRNA-pred 和 EVlncRNA-deep，同时，基于纳米生物传感等技术探索非编码 RNA 与其他成分互作特征，这为认识长非编码 RNA 的特征和功能、鉴定发现新的长非编码提供了有力工具。

特邀嘉宾



郭茂祖教授，北京建筑大学电气与信息工程学院院长，计算机科学与技术系全职讲席教授、博导，“建筑大数据智能处理方法研究”北京市重点实验室主任；省杰出青年科学基金获得者、宝钢优秀教师奖获得者，曾留学瑞典、英国。研究领域：生物信息学、机器学习与人工智能、智能建造与智慧城市等。1997年博士毕业于哈工大计算机应用专业，2002年破格评为教授，2003年评为“计算机科学与技术”、“软件工程”、“生物医学工程”学科博导，目前为北建大“智能建造工程”、“建筑技术科学”学科方向博士生导师。任中国计算机学会生物信息学专委会副主任、中国人工智能学会机器学习专委会常委，国家自然科学基金委重大研究计划指导专家组成员。曾任国家科学技术奖会评专家、中国机器学习会议大会主席。是国家自然科学基金重点项目、国家重点研发课题负责人，曾获教育部自然科学奖、吴文俊人工智能自然科学奖（均排名第1），发表论文300余篇。指导毕业博士22人、硕士80余人。