2018 年中南大学生物信息学前沿研讨会

湖南，长沙

2018-10-26

2018 年中南大学生物信息学前沿研讨会

(2018 年 10 月 26-28 日,中南大学)

程序安排

研讨会场所 ： 中南大学民主楼小礼堂

外地专家入住宾馆 ： 福盛源大酒店（湖南省长沙市岳麓区丰顺路与黄鹤路交汇）

10 月 26 日（星期五）下午: 与会专家报到

报到地点：福盛源大酒店一楼大厅(14:30-18:00)

晚餐：福盛源大酒店（18:00-19:30）

10 月 27 日(星期六)上午：

开幕式（8:30-8:40）	
第一场学术报告	
主持人：李敏 (中南大学)	
8:40-9:25	Methods for detecting topological domains from 3D genomic maps 高琳 教授（西安电子科技大学计算机学院）
9:25-10:10	Two complementary strategies for biclustering algorithm 李国君 教授（山东大学数学学院）
全体与会专家照相	
茶歇（10:10-10:30）	
第二场学术报告	
主持人：高琳 (西安电子科技大学)	
10:30-11:15	Integration and network analysis of multi-dimensional data of complex diseases 邹秀芬 教授（武汉大学数学与统计学院）
11:15-12:00	最长保守子序列问题建模预算法 朱大铭 教授（山东大学计算机与技术学院）
中餐（12:00-13:30）: 福盛源大酒店	

10 月 27 日(星期六)下午：

第三场学术报告	
主持人：朱大铭　(山东大学)	
14:30-15:15	Deep Learning for Mining Protein-Binding Motifs in DNA Sequences 黄德双 教授（同济大学电子与信息工程学院）
15:15-16:00	iTongue: An Intelligent Tongue Image Classification System Based on Traditional Chinese Medicine Constitution 梁艳春 教授（吉林大学计算机科学与技术学院）
16:00-16:45	Core-genome scaffold comparison reveals the prevalence that inversion events are associated with pairs of inverted repeats 王鲁生 教授（香港城市大学）
茶歇（16:45-17:00）	
重点项目启动汇报与咨询	
17:00-18:00	高通量蛋白质组学计算的基础理论与算法 李敏 教授（中南大学）
晚餐（18:00-19:30）: 福盛源大酒店	

学术报告摘要及报告专家介绍

2018 年 10 月 27 日 8:40-9:25

高琳 （西安电子科技大学计算机学院 教授，博士生导师）

个人简介：高琳，女，博士，西安电子科技大学计算机学院二级教授，省级重点学科学术带头人，西安电子科技大学学术委员会委员。担任计算机学会“生物信息专业组”副主任，运筹学会“计算生物信息学分会”常务理事，人工智能学会“生物信息学与人工生命专业委员会”副主任,细胞生物学学会“生物信息学与系统生物学分会”理事。陕西省大数据与云计算产业技术创新战略联盟常务理事。在生物数据分析与挖掘、模式识别与机器学习、图论与组合优化等方面进行了长期研究，承担了国家自然科学基金重点、重大研究计划和面上等项目，在 Nucleic Acids Research、Bioinformatics、Human Molecular Genetics、Briefings in Bioinformatics、BMC Bioinformatics、BMC System Biology、IEEE/ACM TCBB 等期刊发表论文 100 余篇。

报告题目：Methods for detecting topological domains from 3D genomic maps

报告摘要：The chromosome conformation capture (3C) technique and its variants have been employed to reveal the existence of a hierarchy of structures in three-dimensional (3D) chromosomal architecture, including compartments, topologically associating domains (TADs), sub-TADs and chromatin loops. We proposed a generic and efficient method to identify multi-scale topological domains (MSTD). We first applied MSTD to detect promoter-anchored interaction domains (PADs) from promoter capture Hi-C datasets across 17 primary blood cell types. The boundaries of PADs are significantly enriched with one or the combination of multiple epigenetic factors. Moreover, PADs between functionally similar cell types are significantly conserved in terms of domain regions and expression states. Cell type-specific PADs involve in distinct cell type-specific activities and regulatory events by dynamic interactions within them. We also employed MSTD to define multi-scale domains from typical symmetric Hi-C datasets and illustrated its distinct superiority to the-state-of-art methods in terms of accuracy, flexibility and efficiency.

2018 年 10 月 27 日 9:25-10:10

李国君 （山东大学数学学院 教授，博士生导师）

个人简介：山东大学二级教授。1996 年获中科院数学与系统科学研究院博士学位。长期从事图与组合优化、计算机科学和生物信息学研究。证明了以 Chvátal 猜想为代表的四个图论难题；结束了数个可近似性问题的长期争疑；刷新了生物信息学领域多个经典的算法和软件。

报告题目：Two complementary strategies for biclustering algorithm

报告摘要：Recognizing complicated biclusters submerged in large scale datasets (matrix) has been being a highly challenging problem. We introduce a biclustering algorithm BicGO consisting of two separate strategies which can be selectively used by users. The BicGO which was developed based on global optimization can be implemented by iteratively answering if a real number belongs to a given interval. Tested on various simulated datasets in which most complicated and most general trend-preserved biclusters were submerged, BicGO always extracted all the actual bicluters with accuracy 100%, while on real datasets, it also achieved an incredible superiority over all the salient tools compared in this article. To our best knowledge, the BicGO is the first tool capable of accurately identifying any complicated (e.g., constant, shift, scale, shift-scale, order-preserved, trend-preserved, etc), any shapes (narrow or broad) of biclusters with overlaps allowed. In addition, it is also highly parsimonious in the usage of computing resources.

2018 年 10 月 27 日 10:30-11:15

邹秀芬 （武汉大学数学与统计学院 教授，博士生导师）

个人简介：邹秀芬 女，教授，博士生导师。武汉大学数学与统计学院教授，中国工业与应用数学学会数学生命科学专业委员会副主任，中国运筹学会计算系统生物学常务理事, 国际学术期刊International Journal of Data Mining and Bioinformatics 编委。长期从事智能计算与计算系统生物学的研究工作,主持一项国家自然科学基金重点项目和多项国家自然科学基金面上项目及重大研究计划培育项目。已在国际重要学术期刊“PLOS Computational biology”, “Information Sciences”, “IEEE Transactions On Systems, Man, and Cybernetics”, “Soft Computing”, “SIAM on Applied Mathematics”, “Applied Mathematical Modeling”, “IEEE/ACM Transactions on Computational Biology and bioinformatics” 等国际重要学术刊物上发表六十余篇相关的SCI学术论文。

报告题目：Integration and network analysis of multi-dimensional data of complex diseases

报告摘要：Integration analysis of high dimensional data is a challenging problem in computational systems biology and bioinformatics. In this talk, based on the tensor framework, I present a new centrality measure SVT to identify the key nodes of multilayer networks and its application to two multi-dimensional datasets related to cancers. In addition, a low rank tensor decomposition method is proposed to impute missing values of multi-dimensional single cell data.

2018 年 10 月 27 日 11:15-12:00

朱大铭 （山东大学计算机与技术学院 教授，博士生导师）

个人简介：山东大学教授，博士生导师。中国计算机学会理论计算机科学专委会委员，生物信息学会委员。长期从事生物信息学/计算生物学，算法与计算复杂性研究，发表学术论文 150 余篇。擅长生物信息学组合优化问题的算法和近似算法设计。

报告题目：最长保守子序列问题建模预算法

报告摘要：基因组保守子序列被认为是保持生命特征的关键序列。将寻找基因组保守子序列的计算需求形式化为在多个基因组中寻找最长公共样本子序列的组合优化问题。设计该问题的多项式时间求解算法，和限制固定常数个样本的多项式时间求解算法。利用该算法计算出人类和大猩猩基因组草图的最长公共样本子序列。

2018 年 10 月 27 日 14:30-15:15

黄德双 （同济大学电子与信息工学院 教授，博士生导师）

个人简介：黄德双教授，工学博士，博士生导师，英国利物浦约翰摩尔大学教授，同济大学特聘教授，中国科技大学兼职教授，2000 年度中科院“百人计划”入选者，国际模式识别学会会士，国际神经网络学会理事，IEEE 高级会员，2015 年国际神经网络联合会议大会主席，国际智能计算系列学术会议发起人与主席，IEEE/ACM Transactions on Computational Biology and Bioinformatics, Neural Networks 等国际杂志编委。

报告题目：Deep Learning for Mining Protein-Binding Motifs in DNA Sequences

报告摘要：准确识别转录因子与 DNA 序列的绑定位点对于理解转录调控很有帮助，而通过实验识别 motif 以及传统的 motif 预测方法均存在一定的缺陷，因此研究人员开始研究将卷积神经网络用于 DNA 序列中 motif 的识别并取得良好的效果。但是这些方法也存在如下问题：1-序列中核苷酸之间的依赖性；2- DNA 序列中不同的结合位点的长度不同；3-DNA 序列中蕴含的弱监督信息；4-DNA 序列中转录因子上下文信息对自身的影响。为了解决上述问题，我们做了以下工作：针对问题 1 和 2，我们提出高阶卷积神经网络模型（HOCNN）；针对问题 3，我们提出将多示例学习机制融入现有的卷积神经网络（WSCNN）；针对问题 4，我们使用双向递归神经网络来捕获转录因子绑定位点上下文信息对位点自身的影响(WSCNNLSTM, KEGRU)。在不同数据集上的实验表明我们提出的上述方法取得了良好的预测效果。

2018 年 10 月 27 日 15:15-16:00

梁艳春 （吉林大学计算机科学与技术学院 教授，博士生导师）

个人简介：梁艳春，博士，国家二级教授，吉林大学计算机科学与技术学院博士生导师。1997 年于吉林大学数学系博士毕业，获理学博士学位。1990-1991 年在英国曼彻斯特大学做访问学者，2000 年-2004 年分别在新加坡国立大学、新加坡高性能计算研究所做访问教授，2006 年-2010 年期间到意大利特伦托大学做访问教授，2011 年-2018 期间到美国密苏里大学做访问教授。多年来一直从事计算智能和生物信息学等领域的研究。主持国家自然科学基金项目、863 项目、欧盟项目和省部级科研项目 20 余项，获省部级科技进步奖特等奖 1 项、一等奖 4 项，二等奖 3 项。发表学术论文 400 余篇，其中 SCI 收录 180 余篇，EI 收录 200 余篇。出版学术著作 4 部。发表的论著被国内外同行他人引用 4,000 余篇次。连续四年（2014-2017 年）入选《中国高被引学者榜单》。

报告题目：iTongue: An Intelligent Tongue Image Classification System Based on Traditional Chinese Medicine Constitution

报告摘要：Tongue image classification is a key component in traditional Chinese medicine (TCM). For thousands of years, TCM practitioners have judged the patient's health status by examining the tongue's color, shape, and texture. Tongue images can also indicate the overall health status and pre-disease indications without any significant symptoms, which provides a basis for preventive medicine and lifestyle adjustment. With the improvement in imaging and pattern recognition methods, computer aided tongue diagnoses have a great potential to play an important role in TCM by providing more accurate, consistent and objective health assessment.

Constitution theory is a very important part of TCM theory, which plays an important role in disease prevention. The theory categorizes the physical status into nine constitutions, including Normal, Qi-deficiency, Yin-deficiency, Yang-deficiency, Phlegm-dampness, Dampness-heat, Blood-stasis, Qi-stagnation, and Special-Constitution type. We have developed an efficient system to identify the constitution of Chinese medicine from an image by using convolution neural network (CNN). The CNN architecture is composed of residual networks. The system is trained on Pytorch using thousands of tongue images with constitutions labeled by TCM experts. The nine-class classification accuracy is 61.3%.

The system iTongue is currently available as an App at the iOS and Android App stores and also at the Wechat public platform, free of charge. It can automatically determine whether a picture uploaded by users is a tongue picture, and identify TCM constitution classification based on the tongue pictures. The system gives the user's related lifestyle and diet recommendations. Meanwhile the system records the tongue images of the same person over time so that abnormal changes can be detected sensitively and timely. Hence, the system may potentially benefit a broad population and reduce healthcare costs by preventive medicine.

2018 年 10 月 27 日 16:00-16:45

王鲁生 （香港城市大学 教授，博士生导师）

个人简介：Lusheng Wang is a professor of Department of computer science, City University of Hong Kong. His research interests include algorithms and bioinformatics.

报告题目：Core-genome scaffold comparison reveals the prevalence that inversion events are associated with pairs of inverted repeats

报告摘要：Genome rearrangement plays an important role in evolutionary biology and has profound impacts on phenotype in organisms ranging from microbes to humans. The mechanisms for genome rearrangement events remain unclear. Lots of comparisons have been conducted among different species. To reveal the mechanisms for rearrangement events, comparison of different individuals/strains within the same species or genus (pan-genomes) is more helpful since they are much closer to each other. Results: We study the mechanism for inversion events via core-genome scaffold comparison of different strains within the same species. We focus on two kinds of bacteria, Pseudomonas aeruginosa and Escherichia coli, and investigate the inversion events among different strains of the same specie. We find an interesting phenomenon that long (larger than 10,000 bp) inversion regions are flanked by a pair of Inverted Repeats (IRs) (with lengths ranging from 385 bp to 27476 bp) which are often Insertion Sequences (ISs). This mechanism can also explain why the breakpoint reuses for inversion events happen. We study the prevalence of the phenomenon and find that it is a major mechanism for inversions. The other observation is that for different rearrangement events such as transposition and inverted block interchange, the two ends of the swapped regions are also associated with repeats so that after the rearrangement operations the two ends of the swapped regions remain unchanged. To our knowledge, this is the first time such a phenomenon is reported for transposition event.