

# 中科院虚拟经济与数据中心主任石勇： 数据挖掘好比“大海捞针”

■本报见习记者 任芳言

做数据挖掘的人，会和医学期刊扯上什么关系？

2020年4月，一项关于新冠疫情变化与复工复产的研究登上了《柳叶刀》子刊 *EClinical Medicine*。该研究将不同年龄人群划为7类，刻画了他们在家庭、学校、工作场合等情景下的接触模式，并据此分析新冠病毒如何传播，用数量方法给出了疫情期间复工复产的多种模型。

领衔这项研究的作者之一，是中国科学院虚拟经济与数据科学研究中心(下称中心)主任石勇。在他看来，“数据模型是有生命力的。数据挖掘的结果是粗糙知识，进一步过滤叫智能知识，智能知识的发现比数据挖掘的结果更重要”。

## 迅速出成果的秘诀

石勇认为，传统传染病模型预测相对固定，将之与大数据结合、演算，就能发现病毒传播的端倪，并据此预测潜在风险。2020年2月7日完成武汉的城市模型后，团队又选取了北京、深圳等5个城市，根据各个城市经济发展的预期值进行实证分析，模拟出不同城市疫情防控强度与复工复产方案间的利弊关系。

## 简报

### 北京强化科技支撑限塑行动

本报讯 近日，《北京市塑料污染治理行动计划(2020—2025年)》(以下简称“限塑10条”)由北京市发改委、生态环境局正式发布实施，从10个方面提出了北京治理塑料污染的明确要求。

“限塑10条”聚焦餐饮、外卖平台、批发零售、电商快递、住宿会展、农业生产等六大重点行业强化减塑力度，严控一次性塑料制品向自然环境泄漏；同时提出，充分发挥科技资源丰富的优势，聚焦重点领域和关键环节，强化塑料污染治理的科技支撑。根据“限塑10条”，北京将加强塑料领域科技研发，塑料污染治理技术推广，还将建立完善塑料污染治理标准体系。(郑金武)

### 国际磁生物学前沿研究中心 落户合肥“科学岛”

本报讯 近日，中国科学院合肥物质科学研究院发布消息，国际磁生物学前沿研究中心(以下简称iMFR)落户合肥“科学岛”。

iMFR依托稳态强磁场实验装置，以“科学岛”为起点，联合中科院电工研究所和中科院地质与地球物理研究所，与国际相关研究团队紧密合作，聚焦磁效应、磁感应和磁技术三大方向，面向世界科技前沿和人民生命健康，探索磁生物学的未知领域。

目前，iMFR来自中国、美国、英国、法国、俄罗斯、捷克、巴西7个国家的18位成员组成，研究单元包含中国科学院、哈佛大学、约翰斯·霍普金斯大学和牛津大学等国际顶级大学和科研机构，未来将进一步吸纳更多相关科学家加盟。(桂运安)

### 华南理工大学—华大集团 签订战略合作协议

本报讯 近日，“十年再出发”华工—华大基因科学创新班毕业典礼暨战略合作签约仪式在深圳国家基因库举行。华南理工大学副校长朱敏、华大集团联合创始人、董事长汪建、华大集团党委书记、华大基因常务副总裁杜玉涛出席活动。

朱敏与杜玉涛签订了华南理工大学—华大集团战略合作协议。根据协议，双方将发挥各自优势，共同开展人才培养、科技研发、产业转化等方面的合作，共同促进基因组、生命大数据、智能制造与产业转化的深度融合。(朱汉斌 卢庆雷)

## 中科院助力江西建筑陶瓷产业清洁发展

本报讯(记者陈欢欢)近日，江西省建筑陶瓷产业基地内36家陶瓷企业获得江西济民可信(高安)清洁工业燃气项目供气，全面进入试运行期。据悉，高安项目采用中科院工程热物理研究所两大核心技术——循环流化床煤气化和气化飞灰焚烧发电，企业制气成本仅为天然气价格的1/3~1/2，不产生焦油、酚类等污染物，也无废水产生，且实现了固废资源化利用，经济和环保效益显著。

结果显示，利用不同模型，决策者可根据当地防控工作和复工计划预测出相应疾病传播风险，得到城市长期经济发展的基本判断，为疫情防控与科学决策提供了有力支撑。

回看这项研究，石勇等人之所以能在疫情早期迅速发布相关研究模型，有赖于长期与合作伙伴保持的密切联络。“我们可以把很多学科在短时间内集合到一起。”石勇表示。他所在的中心作为交叉学科单位，招生范围包括计算机、管理、数学、生物医学等专业，“只要把这些学生匹配起来，可以发挥很大的科研能动性”。

## 沉淀自己 也沉淀数据

团队研究生张林姿告诉《中国科学报》，由于疫情，在无法与国内合作者面对面交流的情况下，石勇每天会专门留出几小时用于交流工作、推动进展。

“据说牛顿就是在瘟疫封闭期间发现了万有引力，所以不要焦虑，这是很好的沉淀自己的时间。”石勇对学生如是说。

那段日子里，石勇也在思考中心的过去和下一个十年。2004年正式成立以来，中心一直在为国家宏观经济发展提供决策依据。

2010年，全国个人信用评分系统问世，该项工作由石勇团队与中国人民银行征信局和征信中心联合完成，这项系统对中国人的日常商业经济活动产生了重大影响。最近，石勇正带领团队与中国金融期货交易所合作，对大宗级别体量的交易数据进行分析，预测金融期货的变化及可能产生的影响。

这些年来，石勇一直强调异构数据、非结构化数据的概念。当下，让大量“有噪声”的随机数据沉淀下来，实现数据挖掘最优化，仍然是他的工作目标：“当针掉到海里，你要知道针可能掉到哪里去。”

“只有高质量的数据才能产生更好的结果。”石勇再次强调。因为在这方面的努力，他成为首届成思危基金会优秀科研成果奖获得者。

## 人与算法皆纯粹

石勇的团队每周举行一次讨论，大家围桌而坐，讨论与数据挖掘有关的国际最前沿研究。团队成员李彪告诉《中国科学报》，讨论的目的很简单：让大家一直跟着领域的前沿走。因为讨论班的学术氛围太好，以至于已经毕业甚至毕业多年的学生也愿意专门抽出

时间参加。

这样的氛围与石勇本人的行事风格不无关联。在李彪等年轻后辈的眼中，石勇始终是一位简单、纯粹的学者。已过耳顺之年的他仍在不断学习，对新鲜事物保持关注，但他永远衣着朴素——深蓝色的夹克和登山鞋穿了多年，公文包磨破了却想不起来去换。

“做科研不为搞多大名堂、创造多大利益。”李彪表示，这样的言传身教影响了每一位加入团队的后辈。石勇尊重每个人的想法，“但你必须踏踏实实去做”。李彪说。

为了让学生更早接触数据科学，石勇面向中国科学院大学的研究开设了《多元统计分析与机器学习》这门课程，介绍数据科学研究中用到的各种方法。

曾任课程助教的李彪还记得，这堂从不点名的课有40多人选修，但每次都有60多人坐在教室，直到课程结束。“一学期上完，每个人都相当于完成了一个大规模的机器学习项目。”李彪说。

在石勇看来，算法“都是人写出来的”，“一点都不神秘”。高精度的算法只有配上准确的数据，得到的结果才会更理想。他坚信，“把大数据弄清楚，可以把各个方面的工作往前推进”。



近日，在高黎贡山国家级自然保护区，巡护队员李加华在安装红外相机。

在怒江西岸，被誉为“世界物种基因库”的云南高黎贡山深处，有一支队伍常年穿行在入迹罕至的原始森林间，他们自称是手握“物种仓库”钥匙的保管员。他们是高黎贡山国家级自然保护区泸水专职巡护队。

如今，安装红外相机、协助科研单位野外物种成了巡护队最主要的工作。他们协助科研人员摸清了高黎贡山西坡片马辖区的珍稀濒危物种怒江金丝猴的种群数量，发现珍稀濒危植物滇桐野生居群，观测到了数十种珍稀野生动物，成为科研人员可靠的帮手。

新华社记者江文耀摄

## 智源研究院发布 2020 全球人工智能十大进展

本报讯(记者郑金武)2020年12月31日，北京智源人工智能研究院发布了《2020人工智能十大进展报告》。报告由该研究院全体智源学者共同参与，对2020年全球范围内人工智能领域的科学系统、算法等方面进行了评价分析，最终形成了十大进展成果。其中，中国学者取得的进展有3项。

这十大进展分别是：OpenAI发布全球规模最大的预训练语言模型GPT-3；DeepMind的AlphaFold2破解蛋白质结构预测难题；深度势能分子动力学研究获得戈登·贝尔奖；DeepMind等用深度神经网络求解薛定谔方程促进量子化学发展；美国贝勒医学

院通过动态颅内电刺激实现高效率“视皮层打印机”功能；清华大学首次提出类脑计算完备性概念及计算系统层次结构；北京大学首次实现基于相变存储器的神经网络高速训练系统；麻省理工学院仅用19个类脑神经元实现控制自动驾驶汽车；Google与Facebook团队分别提出全新无监督表征学习算法；康奈尔大学提出无偏公平排序模型可缓解检索排名的马太效应问题。

在中国学者取得的进展中，北京应用物理与计算数学研究院王涵所在的深度势能团队将第一性原理精度分子动力学模拟规模扩展到1亿原子，同时计算效率相比此前人类

最好水平提升1000倍以上，极大提升了人类使用计算机模拟客观物理世界的能力。

清华大学张悠慧、李国齐、宋森团队首次提出类脑计算完备性概念以及软硬件耦合的类脑计算系统层次结构，通过理论论证与原型实验证明该系统的硬件完备性与编译可行性，扩展了类脑计算系统应用范围，使之能支持通用计算。

北京大学杨玉超团队提出并实现了一种基于相变存储器电导随机性的神经网络高速训练系统，有效缓解了人工神经网络训练过程中时间、能量开销巨大且难以在片上实现的问题。

## 深圳三大基础研究机构通过筹建期验收

本报讯(见习记者刁雯蕙 通讯员李志鑫)近日，中国科学院深圳先进技术研究院(以下简称深圳先进院)牵头建设的深圳合成生物学创新研究院、深港脑科学创新研究院、深圳先进电子材料国际创新研究院三大深圳基础研究机构通过筹建期验收评估，标志着上述基础研究机构将正式进入建设期。

深圳合成生物学创新研究院成立于2019年1月，共建单位包括深圳市第二人民医院和深圳华大生命科学研究院。筹建期

内，研究院组建了一支年轻有活力、多学科交叉的创新队伍，人员规模超500人，引进了36位全职PI，超过60%的团队骨干来自国外知名高校。

深港脑科学创新研究院成立于2019年1月，共同牵头单位为香港科技大学，共建单位包括南方科技大学、深圳大学和北京大学深圳研究生院。筹建期内，研究院遴选并组建了69个高水平科研团队，人员规模近700人，依托深圳先进院设立了7个研究中心，

建成了非人灵长类脑疾病模型构建及转化研究等6个科研支撑平台。

深圳先进电子材料国际创新研究院于2019年6月19日注册成立，是由深圳先进院和宝安区人民政府共同建设的市级二类事业单位。筹建期内，研究院组建了一支超200人的全职人员队伍，围绕电子封装材料领域关键科学问题和“卡脖子”技术问题，布局了八个研究方向，建立了从基础研究到中试放大、产线验证的全链条研发平台。

煤炭转化为清洁燃气、电能、热能和建筑原料，真正实现了清洁生产。该项目解决了制约高安建陶产业发展的环保瓶颈问题，为当地陶瓷企业淘汰落后燃气产能、清洁化生产提供了经济环保的解决方案。

据悉，在中科院科技成果转化转移转化重点专项(弘光专项)的支持下，循环流化床煤气化技术已经在工业燃气、焦化、建材和有色冶金等行业大规模应用，在合成氨行业实现多个项目投产，并迈入现代煤化工领域。气化飞灰焚烧发电技术是在中科院战略性先导科技专项课题支持下研发成功的煤炭清洁高效梯级利用关键技术，在高安项目中的大规模成功应用，将加速该产业化应用。

## 发现·进展

华东理工大学

## 提出乙炔选择性加氢 催化剂设计新策略



乙炔选择性加氢催化剂设计概念图

本报讯 近日，华东理工大学化工学院、化学工程联合国家重点实验室催化反应工程团队在乙炔选择性加氢Pd催化剂表面结构调控和Ni催化剂活性位点设计方面取得研究进展，相关成果分别发表于中国工程院院刊《工程》和《德国应用化学》。

研究人员从金属催化剂表面结构调控入手，采用XAS、原位XRD、程序升温技术和DFT计算相结合的研究方法，发现了以纳米炭材料为载体制备的Pd催化剂表面存在C原子。通过向炭载体中引入石墨氮物种，进一步调控Pd活性位点电子结构，将乙炔选择性进一步提高约30%。同时，从金属催化剂活性位点设计入手，创新提出通过构筑结构稳定、长程有序Ni-Ga金属间化合物实现Ni活性位点原子级精准调控策略。

理论计算结果表明，相比于Ni和Ni<sub>3</sub>Ga催化剂，NiGa催化剂中Ni活性位点被Ga完全隔离，使得乙炔和乙烯分子以π键模式择优吸附，表现出显著提高的乙炔选择性；实现了粒径相似Ni<sub>3</sub>Ga和NiGa纳米催化剂的可控制备，发现NiGa金属间化合物催化剂中Ni活性位点处于完全孤立状态，且乙炔和乙烯分子通过π键构型吸附。性能考评结果表明，Ni位点完全隔离的NiGa催化剂乙炔选择性显著高于对比催化剂，证实了理论预测结果。(潘聪聪 黄辛)

相关论文信息：

<https://doi.org/10.1016/j.eng.2020.06.023>

<https://doi.org/10.1002/ange.202004966>

中科院长春应用化学所

## 吡啶类天然产物 全合成研究获进展

本报讯(记者沈春蕾)在报道了Dippinine-Chippiine同系天然产物(+)-Troncarpine的首次不对称全合成及其绝对构型确定后，中科院长春应用化学研究所研究员韩福社课题组在该领域研究中再次取得重要进展。相关研究成果近日发表于《德国应用化学》。

Dippinine-Chippiine类天然产物是依波加天然产物次级代谢物中的一个亚家族，具有独特的强刚性稠环骨架，尤其是所含的手性氮杂桥环骨架，在吡啶类天然产物中十分罕见。此外，Dippinine类分子中含有包括桥头手性季碳在内的多个手性中心，并且受C20位绝对构型的影响，分子中的全碳六元环为椅式构象，而Chippiine分子中的六元环为船式构象。在活性方面，Dippinine-Chippiine类天然产物对逆转肿瘤细胞耐药性具有潜在的良好活性。

基于其独特、新颖、复杂的结构特征带来合成上的挑战性以及潜在的生物活性，开展此类天然产物的全合成研究，无论是对推动有机合成新方法、新策略发展，还是对深入开展相关天然产物及其新型类似物生物活性研究，都有重要科学意义。

相关论文信息：

<https://doi.org/10.1016/j.tet.2020.131641>

江南大学

## 评估油酸 对炎症因子影响

本报讯(记者王方)近日，江南大学食用油营养与安全科技创新团队在《食品科学与营养评论》发表综述文章，通过Meta分析论述了油酸与炎症因子的相关性。

油酸是地中海饮食模式和橄榄油的主要脂肪酸成分(占70%以上)，也是膳食脂肪和血浆甘油三酯中的主要脂肪酸。该文筛选719项研究，通过成人的随机对照试验揭示了增加油酸剂量对炎症因子水平的影响。在719项已确定的研究中，有31项随机对照试验符合条件，涉及1634名受试者。结果表明，补充油酸显著降低了C反应蛋白。然而，饮食摄入油酸对肿瘤坏死因子- $\alpha$ 、白介素-6、纤维蛋白原、纤溶酶原激活物抑制物-1活性、可溶性细胞间粘附分子-1和可溶性血管细胞黏附分子-1没有显著影响。

该研究系统分析和评估了油酸对炎症因子的影响，为预防和控制以心血管疾病为主的慢性疾病、合理调整膳食结构，以及制定公共营养卫生政策提供理论依据，对进一步阐明膳食脂肪结构对炎症的影响及作用机理具有重要意义。

相关论文信息：

<https://doi.org/10.1080/10408398.2020.1854673>