

# 建人工智能“大厦” 夯数理基础为先

■本报见习记者 田瑞颖 记者 郑金武

当房子越建越高时,就需要回头看看地基是不是坚实。中国科学院院士、北京大学教授张平文团队最近在做一件事——给人工智能“大厦”夯基,即为人工智能机器学习常用的数学符号提供一个通用标准。

人工智能作为一个高度交叉而又发展迅猛的学科,吸引了来自数学、物理、化学、生物甚至人文社科等不同领域的众多研究人员参与其中。对于不同学科的研究人员来说,一个现实的问题就是如何消除信息鸿沟,统一数学符号等基础性工作,看似简单,却是人工智能长远发展的“地基”。

“目前,人工智能的数理基础还不成熟,但这也为广大青年学者提供了更多创新机会。”在日前举办的2020北京智源大会上,张平文等专家围绕人工智能下一个十年发展面临的重大问题进行交流探讨,呼吁加强人工智能数理基础等相关研究。

## 有望催生新的研究方向

中国科学院院士、清华大学教授张钹认为,第一代人工智能靠知识驱动,第二代人工智能靠数据驱动,目前进入第三代人工智能发展期,主要靠认知驱动。“认知驱动中,深度学习研究尤为重要,发展前景广阔,但仍需加强理论和方法等基础性工作。”

张平文也认为,机器学习等深度学习理论是当前人工智能数理基础研究的热门方向之一。虽然在本次人工智能浪潮中,深度学习理论表现良好,但研究人员还不能解决它的可解释性问题。

事实上,可解释性、可计算性、泛化性、稳定性等正是人工智能未来发展面临的重大理论挑战。例如可解释性,“分两个层面,一是从数学算法的角度解释,二是从数学模型的角度解释机器学习的问题。这些基础性工作都有待深入研究。”张平文说。

“机器学习是从数据中挖掘出有价值的信息,数据本身是无意识的,不能自动呈现有用的信息。”北京大学特聘教授张志华说,“怎样才能找出有价值的信息?首先要

# 早期智能生态需高度协作

“智能变革将带来巨大发展机遇,但挑战也不少。”近日,英特尔公司市场营销集团副总裁兼中国区总经理王锐在接受《中国科学报》采访时说。在她看来,推动智能变革需要重点解决三大挑战。一是传统基础设施已经不能满足智能化的新需求,要从云到端实现系统性升级。二是应用场景多元复杂,需要根据需求制定不同的解决方案。三是智能生态还处于发展早期,只有高度协作才能带来产业效益最大化。

“在互联网的‘上半场’,也就是消费互联网的发展历程中,中国已成为全球的领先者。现在,随着‘新基建’的推进,5G的发展,中国也将会成为互联网‘下半场’的领跑者。”王锐说。

她表示,人工智能(AI)、5G、边缘计算技术的“三方汇合”,创造了智能技术飞速迭代的需求和条件。结合当前正在爆发的云经济,将会促进诸多产业乃至整个社会加速向智能化转型,同时也会为新业务的增长带来巨大机遇。

例如,AI及数据分析将为金融、健康医疗、工业、通信及交通等多个行业开启新的机会。据市场研究公司IDC预测,到2021年,75%的商业企业各类应用将使用人工智能技术;到2025年,所有数据中将有约1/4是实时产生的,而在这一庞大的增量中有95%的数据将通过各种物联网设备产生。

英特尔亚洲人工智能销售技术总监伊红卫告诉《中国科学报》,人工智能和数据分析是未来10年起来决定作用的工作负载,也是关键的转折性技术。据她介绍,围绕AI,英特尔有三大战略,即从云到端的全栈解决方案、软件上的赋能与优化,以及智能生态建设。

为更好地支持数字化转型和智能化变革,英特尔近日在举办了数据创新峰会暨新品发布会。其宣布推出最新的数据平台产品组合,包括集成AI加速的英特尔第三代至强可扩展处理器、英特尔首个人工智能优化FPGA Stratix 10 NX、第二代英特尔傲腾持久内存、最新英特尔3D NAND SSD及相关软件解决方案,以在数据中心、云和智能边缘领域支持客户进一步加快人工智能和数据分析等工作负载的开发和部署,助力智能新基建建设。

在本次会议上,来自华大基因、金山云、国电南瑞集团、腾讯云、阿里云等行业企业的代表,分享了他们部署AI和数据分析解决方案、推动IT架构转型、提供创新云计算服务的实践。峰会还呈现了20余场应用专题深度解析和从学做到的AI实战课程,以及20余家英特尔合作伙伴的应用案例。



人工智能长远发展需夯实数理基础研究“地基”。  
图片来源:Unplash

给数据一个抽象的表示,然后进行建模,再进行计算。”

“机器学习想要实现的是从数据到决策到预测的过程,这意味着很多新问题会出现,而这些新问题都有赖于通过加强算法、模型等数理基础研究来解决。”张平文表示,数理基础研究无法在短时间内实现,需要开展大量工作。

反过来,“人工智能数理基础研究对数学发展也非常重要,研究期待未来可以产生类似于概率论的新的研究方向。虽然短时间内不太可能实现,但仍有这种可能性。”张平文说。

为此,张平文鼓励广大青年学者投身人工智能数理基础研究,“数理基础研究别具活力,希望通过大家的努力使它趋于成

熟,产生更多原创性成果”。

## 从应用出发定夯基计划

人工智能数理基础研究涉及数学、概率统计、计算理论等众多学科,不同学科、不同研究方向使用的数学记号和表述不尽相同。看起来相去甚远的两个记号,可能表达的是同一个概念;在一个学科已经公认的优美、简单的记号,在另一个学科或许还需被重新发现、重新定义,而新定义的记号很可能远不如已有的记号方便。

这给新进入该领域的科研人员和工程人员造成了不少麻烦。北京大学教授李铁军说:“机器学习的普及,迫切需要制定一套统一的数学记号。这将大大减轻研究人员在不同符号体系间反复切换的工作量。”

为人工智能研究提供一套统一的符号标准,这项工作“增强了不同领域研究人员的互相理解。研究人员在看其他领域成果的时候,也更容易看懂”。张平文呼吁研究人员在科研中尽量采用同一套符号标准来交流、写论文,并为完善符号标准而努力。

也就是说,基础研究要与应用结合起来。“开展人工智能研究不能凭空想象,必须了解实际问题,考虑实际应用需要。”张平文说。

北卡罗来纳大学教堂山分校终身教授朱宏图对此表示认同。他说:“人工智能领域的研究应该从应用角度出发。”

“人工智能的核心是如何利用收集的数据来为用户创造价值,提高用户满意度。”朱宏图认为,随着收集数据的方法越来越多,频率越来越高,为用户提供相应的方案也会越来越多。“这为人工智能研究带来了新机遇。”

朱宏图表示,从研究需要出发,为用户提供更好的方案和服务环境,也需要加强算法、模型等数理基础研究,“一方面要强调它与应用的连接,另一方面也应该基于

数学本身的基本问题开展研究。”李铁军认为,虽然数理基础研究当下可能跟应用联系并不紧密,但有可能提供一些最基本的数学语言,深化人类对人工智能数学结构的认识。

## 实践是检验合作实效的“金钥匙”

夯实人工智能“大厦”,加强数理基础研究,“主要有两条路,一是从低维空间走向高维空间,二是从低维空间走向高维空间,但无论哪条路都存在困难。”张平文说,不管走哪条路,都需要加强学术界与产业界及不同领域间的交流。

与会专家认为,一直以来,产业界与研究界的信息鸿沟制约着数理基础研究的进度。产业界拥有研究界缺乏的大数据,而产业界欠缺研究界对实际应用背后的科学认识。

朱宏图同时还担任滴滴出行公司的首席统计学家。他表示,人工智能领域相关研究,数据在其中起着越来越重要的作用。企业积累了丰富的数据,这是人工智能研究的重要基础。

目前,在张平文等业内人士的努力下,北京智源人工智能研究院已设立“人工智能数理基础”重大研究方向,联合全国相关科研院所和产业界人士,致力于打破基于计算机实验和神经科学的人工智能的惯用建模范式,以可解释性的新型人工智能模型、新型机器学习算法、深度学习的基础研究等为研究内容,建立以数学与统计理论为数理基础的新一代人工智能方法论。

香港浸会大学兼职教授李春霖认为,研究将会面临挑战。他举例说,在材料设计中,机器学习提供了新方法,但既有机器学习模型并不能直接拿来用,还需要把机器学习的模型和材料物理的模型结合起来。

在他看来,人工智能数理基础研究需要产业界和研究院所合作。这样的合作将为行业发展创造更大空间,而实践也将成为检验合作实效的一把“金钥匙”。

# 为智慧城市织一张立体网

■本报记者 李惠钰

突如其来的新冠肺炎疫情,不仅是对政府公共卫生应急管理能力的深度检验,也是对我国各地“智慧城市”建设水平的“突击大考”。

近日,在第四届世界智能大会上,中国科学院院士、深圳大学智慧城市研究院院长郭仁忠表示,“当今世界发展的两大主旋律就是城市化和信息化。对于生态城市建设,智慧之路是可行路径,也可能是必然路径。”

## “一网统管”成为可能

在渤海湾畔天津滨海新区内的中新生态城,智慧应用场景触手可及,多个项目也在第四届世界智能大会上亮相。

智慧小区内,高空抛物被全方位、无死角地实时监控,并自动显示抛物源头;在智慧图书馆,智能机器人每小时可自动盘点2万多册图书,智能分拣还书系统更是超过人工效率的10倍以上;交通领域,通过对信号灯优化配时,高峰时段道路拥堵时间减少13%……

疫情期间,中新生态城还上线了疫情防控小程序“新冠通”,基于云端采集技术,通过录入居民身份证号、家庭住址、户籍所在社区即可进行每日实时监测。

郭仁忠表示,智慧城市是运用物联网、云计算、大数据、空间地理信息集成等新一代信息技术,促进城市规划、建设、管理和服务智慧化的新理念和新模式。

可以说,智慧城市并不只是数字城管,一个挂在政府大厅里的巨大显示屏,也不只是人们手中出入社区的健康码,更是撬动社会治理专业化、智能化和法治化的支点,是切实提升城市管理效能、改善市民生活质量、缓解“大城市病”的金点子。

北京数字政通科技股份有限公司董事长吴强华表示,因疫情防控需要,各社区越来越重视“网格化”。以网格化管理为基础,以“一网统管”为引领,才能构建科学化、精细化、智能化的城市大脑运行体系。

给智慧城市建立“大脑”,就是让城市拥有类似人类大脑的功能,能够自我感知、自我学习、自我修正,让城市的各个“器官”协同工作。

“大数据汇聚”即多源数据接入,如同人的耳目;城市大脑负责智能分析研判,如同人的大脑;一网统管即运行处置体系,如同人的手脚。”吴强华表示,城市大脑驱动一网统管运行处置体系,才能实现智慧城市治理能力的提升。

## 时空大数据赋能城市建设

中国科学院空天信息创新研究院副



运用物联网、云计算、大数据、空间地理信息集成等新一代信息技术打造智慧城市  
图片来源:Unplash

研究员杨丽娜表示,全球大数据进入加速发展时期,数据总量每年增长50%,数字经济已成为全球经济增长的引擎,数据也成为城市的重要资产。

空间信息作为重要的时空大数据来源,能为智慧城市建设赋能添彩。“疫情期间,通过感知和定位技术,实现用户一段时间移动轨迹的数字化,进而分析出用户移动过程中的接触史和暴露史。”郭仁忠说。

“城市居民状态数据可视化,也为城市精细化管理提供了信息支撑。”杨丽娜表示,空间对地观测技术开始向天、空、地、人一体化纵深和全球化横向发展,空间对地观测手段持续向多平台、多传感器、多方向方向发展;遥感数据处理自动化、智能化和实时化、批量化,也向快速响应和准确翻译方向发展。这些技术都将赋能智慧城市建设。

此前,中国科学院在中天津生态城落地建设智慧城市数据汇聚平台,通过该平台可以实现数据实时共享,并进行三维可视化处理和应用。“就像是将泥沙碎石等原料,变成标准化、高质量、可复用的信息‘金砖’,为智慧城市建设提供原材料。”杨丽娜说。

她举例称,通过收集城市管理部件、案卷数据等,展现城市实时动态变化情况,实现城市管理动态监控;水环境监测数据、空气质量监测数据、噪声监测数据等,都将助力探索空气质量指数波动和水环境监测对比等。

而在江苏无锡,一系列的“时空大数据+”应用也让这座城市管理“耳聪目明”起来。例如,通过动态感知技术,可查看全市4000余辆出租车的实时位置和历史轨迹;还可对重点监控区域的土地利用开发状况进行监控等。

## 将安全纳入顶层设计

当前,“新基建”撬下“蓄力键”,为城

市发展赋能。与此同时,智慧城市也是5G、人工智能等技术落地应用的场景。市场研究公司IDC预测,到2023年,1/3的智慧城市市场将被5G影响,70%的一二线城市将使用5G技术。

但360政企安全集团副总裁刘霁提醒道,当整个城市运行在软件上时,“安全”就成为城市建设、管理、运营的基础。

“网络安全已成为城市大安全的重要入口,甚至关系到整个城市的社会稳定和运转。网络安全将成为未来城市关键基础设施的重要保障。”刘霁表示,当前,全球针对关键基础设施的攻击事件频繁发生。在美国、俄罗斯等发达国家,城市安全基础设施已经成为网络空间重点攻击的对象。

“如果将智慧城市的信息化架构展开来看,从应用层到最底层的网络通信和物联网感知层,都面临着不同程度的网络安全威胁。”刘霁说,“许多智慧城市都建设了门户网站,居民可以通过身份认证来办理线上业务。如果这些门户网站存在漏洞,公民身份信息就有被冒用的风险,甚至导致敏感数据的大规模泄露。”

对此,将安全能力纳入智慧城市的顶层设计已成为业界共识。天融信科技集团CEO李雪莹此前表示,定义安全达到的标准和级别,并细化到不同规模智慧城市的规划和设计中,是安全能够纳入智慧城市顶层设计的根本保障。此外,智慧城市中垂直行业众多,包括交通、水利、医疗等各个方面,对应到每个行业都应当通过规范化来指导行业的健康发展。

“只有先做到这两点,才能在智慧城市的顶层设计纳入安全建设。否则就会像过去多个行业的发展规律一样,‘建设先行,安全滞后’,导致安全成为保障性、补丁性的工作。由此看来,政策的保障和行业规范的要求对城市安全建设尤为重要。”李雪莹说。

# 一所一人一事



丁健指导研究进展。

## 丁健

中国科学院院士、肿瘤药理学家。他领导建立了符合国际规范的抗肿瘤药物筛选和药效学评价体系,重点围绕分子靶向抗肿瘤药物开展工作,取得了一批原创性科研成果,为创制具有自主知识产权的新药,尤其为我国抗肿瘤原创新药研制及创新能

已经年过花甲的中国工程院院士丁健,每天还忙碌在中国科学院上海药物研究所的实验室里,不仅为新药研发的每个环节把关,还关心着每项科研成果的转化落地。

国家自然科学奖、中国科学院杰出科技成就奖、谈家桢生命科学成就奖……这些荣誉在丁健看来并不值得过多关注。他认为:“只有我们研制的老百姓用得好的药真正投入市场,造福于患者,才值得庆祝和宣传。”

数十年来,丁健带领科研团队持之以恒地在抗肿瘤新药研发中不断探索,并取得令人瞩目的研究成果。他的目标十分明确和坚定:“做中国老百姓用得好的药,让中国的新药走向世界。”

## 做让老百姓用得好的药

一粒进口抗肿瘤药物,动辄数百元甚至数千元,比同等分量的黄金还贵。这对很多患者家庭来说,几乎是不可承受的。

自1992年从日本留学归来,丁健一直在中科院上海药物研究所潜心研制具有自主知识产权的抗肿瘤新药。“如果有100个患者,却只有两三个能负担得起这种药,那这项研究的应用价值就大打折扣。”丁健说,“要做药效好并且让老百姓用得好的好药。”

岁月染白了他的黑发,也见证着丁健辛勤浇灌得来的累累硕果。目前,他率领团队研发的具有自主知识产权的13种新药正在进行临床研究,其中4个新药获美国食品药品监督管理局(FDA)批准。其中,三靶激酶抑制剂AL3810、我国首个ALK二代抑制剂T二酸复瑞替尼、EGFR三代抑制剂120067等已在临床研究中观察疗效。他与团队共同申请国内外专利500余项,其中获授权150余项。在抗肿瘤药物的生物标志物和耐药机制研究方面,他带领团队发现了十余个生物标志物及多个增效减毒、克服耐药的联合用药方案;在《细胞》《癌细胞》发表论文300余篇,他引6000余次。上述研究成果对分子靶向抗肿瘤药物的精准用药、疗效监控、耐药克服等具有重要指导意义。

## 让中国原创新药迈向国际

“我们必须走自己的路,因为国外科学家做新药,首先考虑本国患者的特点,所以这些年来,我们做抗肿瘤药物主要瞄准中国人高发肿瘤,如肝癌、胃癌、食管癌等。首要目标就是要满足中国患者的迫切需求。”丁健曾强调。

2015年,丁健作为首席科学家牵头的中国科学院战略先导A类科技专项“个性化药物——基于疾病分子分型的普惠新药研发”正式启动,目标是对病人进行精准治疗。丁健表示,团队2012年就提出了要做个性化药物的想法。长期以来,在抗肿瘤药物研究领域,我国只能跟跑国际,现如今我国部分研究成果已经可以与国际比肩。

为此,丁健带领药物创新团队搭建了符合国际规范和水准的覆盖分子水平、细胞水平和整体动物水平的系统化三级抗肿瘤药物评价平台;建设了先进的药物转化研发平台;布局了蛋白激酶、表观遗传、肿瘤代谢、肿瘤免疫微环境四大研究领域,创建了一条具有国际竞争力的抗肿瘤药物研究管线。

## 加大科研成果的转化力度

丁健2005年至2013年担任中科院上海药物研究所所长期间,传承药物所创建之初“使中国药物科学化”的重要使命,坚持将“出新药”作为研究所的战略目标。研究所进一步提出了“以国家药物创新体系的建设为主线,实施合作共赢的国际化战略、阶梯递进的可持续战略、学科互补的集成创新战略和以人为本的人才开发战略”等四大战略,“做老百姓用得好的好药、新药”,带领全体药物所人取得了骄人成绩。

作为国家实施科技成果转化改革试点单位之一,中科院上海药物研究所自2015年以来共转化科技成果50项,其中丁健所带领的研究团队优先先行,先后有13个新药实现成果转化,合同金额逾20亿元,取得了显著的经济效益和社会效益,极大激发了科研工作者创新创业的积极性,也为中国的抗肿瘤领域储备了一批颇有应用前景的候选药物。

科技是国家强盛之基,创新是民族进步之魂。丁健非常庆幸自己回国投身科研事业,见证了祖国经济的腾飞,也亲身参与了中国科研事业的大发展。他说:“当前是我国科学研究大发展的盛世。”举国上下重视科技创新,政府积极鼓励科技成果转化。他将执着创新,为实现“做老百姓用得好的好药”的目标而不懈奋斗。(作者单位:中科院上海药物研究所)

# 做老百姓吃得好的好药

记中科院上海药物研究所研究员丁健  
■方婷 徐晓萍