

# 西南交大大数据高峰论坛聚焦 创新 联盟 应用

■本报见习记者 李勤 通讯员 蔡京君 马艳娟 朱正安

全球已经进入了数字化时代,数字化正深刻地影响着世界科技、全球教育和教育改革,高等教育不能错过数字化这个新的战略机遇,更不能错过数字化的时代。由西南交通大学与中国科学报社联合主办,信息数字化领域专家参与的“创新 联盟 应用大数据高峰论坛”于10月20日在西南交通大学举行。本论坛深度聚焦数字化革命浪潮下,创新、联盟、应用大数据的前瞻性思维、战略主旨、方略与举措,旨在对数字化战略实施作出积极贡献。本报刊登论坛上的专家发言,以飨读者。

西南交通大学校长徐飞:

## 数字化是世界和高等教育发展的战略机遇

数字化正深刻地影响着世界科技、人类生存方式和全球经济变革,强力实施数字化发展的战略机遇。第一,数字化引发了技术范式的革新。数字化的本质是数字化技术,是一种技术范式的创新革命,为通信、计算和存储等开辟了一条全新的科学问题解决路径,为信息化的实现提供了有力的技术支撑。第二,数字化带来了人类生存和生活方式的变革,带来了一种全新的生存方式——数字化生存(Being Digital)。数字化技术的广泛应用正在引发一场范围广泛的产品革命,数字产品正在主导着人类的生活。第三,数字化推进了经济的大转型,数字化使信息不仅是通信联系的纽带,而且已经成为生产力中最重要的因素,极大地推动了全球经济大转型。第四,数字化颠覆了传统教育的发展模式,以慕课(MOOCs)为代表的数字化教育模式的广泛应用和井喷式爆发,为高等教育带来了一场革命风暴,颠覆了传统的教育理念、教育方式、教育手段,实现了在线学习与课堂讨论的有效融合;数字化教育兼顾了班级制教学和个性化教学,最大限度地促进个性化学习;数字化技术使大学治理层级更加扁平,促进教育管理方式的再造和优化。

二是推进数字化战略是学校实现跨越式发展的战略举措。数字化时代所有高校站在同一起跑线上,学校要勇于抢抓数字化战略机遇,科学谋划,力争实现非线性、超常规的跨越式发展;同时,致力于大学的根本使命——人才培养,探索建立数字化育人模式,切实提高育人质量。数字化战略作为以知识生产、价值发现、文明融合展示学术竞争力、科技创新力和思想影响力的重要抓手,是学校实现弯道超车、跨越发展的重要举措。三是强力推动数字化战略。第一,正确认识学校推动数字化的战略意图和战略意志,数字化战略不是贴标签、赶时髦,是学校追赶世界科技前沿的必然选择,与师生学习、生活、工作、科研、教学息息相关,且人人可为;数字化除了投入,更重要的是谋划好、实施好,使投入能产生溢出效应。第二,实施数字化育人,切实肩负大学人才培养的根本使命。实施数字化育人,核心是探索建立数字化育人新模式。要改革教学方式,加大推进学校MOOCs课程建设力度,逐步试点推进“翻转课堂”及“混合式教学”模式;要注重人才培养过程,利用大数据分析,开展全过程、全方位、全员育人。建立学生从入学、学习、生活、就业与职业发展等数据档案,助力教育教学改革。第三,利用数字化实现学科交叉、融合与再造,持续提升学术竞争力。结合学校学科特色

科技部原党组成员、科技日报社原社长张景安:

## 创新驱动战略与大数据

第一,实施创新驱动战略是实现中国梦的关键。虽然我们创造了30年高速发展的奇迹,但目前我国高端核心技术和关键技术受制于人的局面还没有从根本上改变,与日本和韩国等邻国相比,世界级的、具有国际影响力的创新性大企业仍不够多,世界知名品牌也不够多。只有通过创新驱动战略才能改变这一现状,从而实现从中国制造到中国创造的提升,从大国向强国过渡。

目前我国创新中有三个难点,即垄断、融资难和知识产权的保护问题。在创新过程中还存在“四多四少”,即面向国内的多,面向国际的少;考虑眼前的多,考虑长远的少;模仿山寨多,独创少;创业多,创新少。

第二,自主创新不仅要尊重其规律和特点,还需要一个文化建设过程,进一步加强

国家“千人计划”学者、西南交通大学金融大数据研究院院长李维萍:

## 数据科学的利用

2009年,美国开放Data.gov网站,所有美国公民可以查到政府在财经、安全、能源、交通运输等各方面的数据。两年后达沃斯发布“大数据,大影响”,包括正反两方面,负面则是数据安全问题。数据安全是目前最受关注的问题,或许会成为以后十大重要专业之一。

2012年3月,奥巴马政府宣布投资2亿美元启动“大数据研究和发展计划”,提出大数据的研发对强化美国国土安全,转变教育学习方式,加快科学工程领域的创新和对社会的日常生活改变有着不可替代的重要作用。针对此计划,美国国家自然科学基金(NSF)、国立卫生研究院(NIH)、国防部(DOD)、能源部(DOE)、国防部高级研究计划局(DARPA)、地质勘探局(USGS)6个联邦机构共同提高收集、存储、保留、管理、分析和共享数据的核心技术,支撑大数据开发和应用所需的人才和经费。2012年,联合国发布了《大数据政策白皮书》,指出大数据对于联合国和各国政府来说是个历史性的机遇,人们可以使用极为丰富的数据资源,对社会经济进行前所未有的实时分析,帮助政府更好地响应社会和经济运行。2012年中国工信部发布《互联网“十二五”发展规划》,同年9月18日,中国科技部印发《中国云科技发展“十二五”专项规划》,都把大数据技术作为一项重点予以支持。2012年11

月,十八大提出将金融体制改革定为未来十年发展的重中之重,未来中国的金融业将依赖智慧数据分析,开发金融业务与金融服务创新。2013年12月1日,中国计算机学会发布《中国大数据技术与产业发展白皮书》,旨在推动学术界和业界的大数据发展。2014年5月1日,美国白宫发布《2014大数据白皮书》,大数据为美国经济、健康卫生、教育、国土安全、能源开发利用提供难得机遇。大数据的新特性在哪里?按结构上的特性,大数据可分为结构化数据(海量查找、统计与更新)、非结构化数据(图片、视频、符号、邮件、文档等)、半结构化数据(变为结构化数据,依照非结构化数据存储)。大数据不是随机抽样而是全部样本,从统计走向概率;不问因果而看相关。大数据还具有“4V”特性,即:量大(Volume,存储大)、计算量大、繁杂(Variety,来源多、格式多)、高速(Velocity,增长速度快、处理速度要求快)、价值(Value,海量捞针、全新方法)。大数据是资产,最早来源于电话、网络、信用卡,汇集数据对科学发展与进步至关重要。资本市场的tick数据的急剧增长,使得原有数据管理、统计方法和金融机制不能适用。在寻找新方法和待解之谜的新发现时,数据扮演着重要角色。即使是解释乱作一团的数据失败,也对科学发现有重要意义。不受规划的数据带来巨大的好处,连企业也常常

青睐。有耐心有远见,不为大数据而大数据,要切记大数据不是包治百病的灵丹妙药。这里不得不提到一本书《计算方法简述》,Naur在里面首次提到数据科学,他提出从不可能处获得数据是未来科学发展的一个部分。

数据科学与金融有什么关系?中国有6.32亿互联网用户,全球有20亿人生活在数字世界。扎克伯格,也就是Facebook的CEO,他今年推出一个全球合作计划,寻求更新更廉价的电脑、更高效的数据软件和全新的商业模式。Facebook市值为1900亿美元,只有6000名员工,他们没有固定资产,有的只是拥有的用户数据。怎么给数据估值?这是对现代数据金融一个巨大的挑战。大数据市场今年已创造51亿美元,估计到2015年将达到321亿美元,2017年将达到534亿美元。国际发展高级委员会呼吁进行“数据革命”,增加数据的数量与质量的可获得性和可使用性,倡导政府、统计部门、全球和地方非政府组织、学术和研究机构、私企通过新形式合作,使数据更易管理和使用。西南交通大学金融与大数据研究院的成立是在各方努力下促成的,希望能够整合资源、协同科研、分享方法、多方位合作。在此也呼吁政府成立大数据研发协调中心,建立大数据风险管理机制,每年发布与大数据相关的年报(白皮书)。

无论从哪个角度看,大数据仍处在初始阶段。尽管企业和媒体对大数据时代的到来表现了极大的热情,但科技和社会学者应当对此保持清醒,并对已经面临的和将会出现的问题有所准备。

今天数据积累的速度远远超过数据能被处理和能被利用的速度。2013年产生的数据中只有约22%有可能被用作分析,但结果只有5%实际被分析。照目前数据产生的速度,预计到2017年,全球又会积累超过四倍于今天的数据量。结果导致存储设备、数据中心快速膨胀。纵向累积的、缺乏多维度关联的数据的确越来越庞大,但这并不是“大数据”,而只不过是“数据大”而已。

在我们刚刚步入大数据社会时,应当对大数据作一些前瞻性思考。这些思考应当以数据为源头,以人类为终点。

当人类让数字讲话、把决策权更多地移交给“大数据”时,第一个重要问题就是数据的质量。“垃圾进,垃圾出”这句话早在50年前就被用来描述自动化处理数据时的质量问题,今天此话依然有效。研究表明,知识工作者平均花近一半的时间在寻找数据、验证数据、修正数据、剔除不靠谱的数据。数据质量问题可能发生在收集、存储、处理、传输和分享整个过程中的任何一个环节,但第一个关口仍是数据录入端(记录或采集口)。大多数情况下,数据输入端很少知道数据使用端为什么需要这种数据,用这些数据做什么。当使用端发现问题时,除非是系统性问题,大多被随手解决,少有追根溯源,从源头纠正问题。

那么数据的质量如何保证,谁应对质量负责呢?解决数据质量的问题,更多的是在管理不在技术。行之有效的方法是把数据质量的管控更多地转交给使用端(业务条线),建立数据“原料方”和“生产方”之间的直接的客服关系,形成制度和从数据源头和元数据层面控制质量。

大数据社会要求人们学会用一种全新的方式打量这个世界,工具、技术、技能和人才缺一不可。要及早部署和投资在以下方面:

数据分析工具和软件平台:人工智能(AI)技术,自然语言处理、模式识别、机器学习、预测分析、数据挖掘、信号处理和元数据管理等;大数据要求在数十、数百、甚至数千台服务器中进行大规模并行运算的软件,目前使用的大多数关系数据库管理系统、桌面数据库和可视化软件包已很难满足需求,因此仍需关注MPP数据库、分布式文件及分布式数据库的发展。

数据分析人才和算法模型:培养训练会使用大数据分析语言工具,如ECL、Cassandra、Hadoop、Hive、MongoDB等的人才只是整个需求的表层。再深入一层,要让大数据发挥作用,跨行业顾问、分析师和有行业经验的编程人员缺一不可。图形及视觉表现在人机对话和数据表达层作用极大,但融合技术和艺术,能将抽象形象化的语言、工具及人才尚不多见。没有数学模型很难想象如何“大”数据,数据只有通过算法模型才能被电脑解读,但数学模型只能逼近现实,无法百分之百与现实重合。对于解读数据关联的隐喻而言,这些模式非常有用,但必有局限性。所以如何建立能精准模拟世间万物的数学模型,是集理论研究与应用开发的重要地带。

进入大数据时代,一个“大”字很容易掩盖一切。现实情况是,我们可用的数据越来越多样化,但其中大部分的数据是人们在只知其来源不知其如何产生、质量被如何管控的情况下而使用。因此,要想让人们信赖基于大数据的决策,对所有数据源提前测试和试验必不可少。不管数据是大是小,真实可靠最重要。只有逼真的模型、精准的分析,才能体现大数据的价值。过去没有数据是瞎子摸象,现在数据太多是大海捞针,瞎子摸象和大海捞针的结果相差无几。大数据容易使人们的关注点从因果关系移动到相关关系。它的结果常常能帮助人们回答“是什么”而不是“为什么”。肤浅的使用大数据,有可能引导人们止步于探究事件背后的深层原因,满足于了解现象之间的联系,并利用这种联系得出是对非的解读。综观世界所有政府,他们既是最大的数据收集使用者,又是保护隐私的最高管理者。实现真正的数据共享是大数据腾飞的基础,主管部门应该多关注数据开放和共享的规则、公民信息安全和隐私的保护。

在可预见的未来,我们面临诸多挑战:技术挑战会出现在从信息搜索、数据捕捉、存储、传输、共享、分析直到可视化全过程。另外,必须面对大数据对社会人文的挑战。法律层面,更开放的网络会带来更多的数据窃用、滥用和非法监控。一旦强大的新型数学数据工具出现,如何控制它不会被用在恶意方面。人文伦理层面,在数据越来越多地获得否决权的数字社会,如何对待人的经验积累和直觉判断?大数据技术日新月异,我很高兴看到西南交大已不失时机地建立中心、开设课程,联合院校精英迎接大数据时代的到来。

## 大数据社会的前瞻性思考

IBM 大中国区政府与公共事业部技术总监文金言

**一图读懂 交大数字化战略**  
顺应时代发展潮流 抓住数字化发展机遇 推动跨越式发展

2014年10月20日西南交通大学举办数字化战略高峰论坛和大数据高峰论坛,正式发布《西南交通大学关于数字化战略的实施意见》

**人才培养**  
以教学资源数字化、教学支持网络化、管理信息一体化、人才培养信息数字化为基础,实现精细化的教学过程管理。

**数字化**  
课程改革: 收集学生作业、出勤、考试、构建深度交流的“课堂—在线”一体化教学环境  
个性化学习: 个性化学习方案  
新专业: 力争2015年开始在学院进行“大数据+专业(方向)试点本科招生”  
课程: 构建以MOOCs、微课程为主体的全方位立体化课程资源体系  
档案: 学生、教师、校友信息数据化,建立“学生成长档案系统”

**科学研究**  
重点面向轨道交通和前沿科技等领域,开展数字化技术等科学研究,努力取得对未来有重大影响的原创性成果。

**建立大数据科技创新专家委员会**  
重点开展云服务平台技术研究,启动面向生产制造领域的云工程研究中心建设

**结合学校“人才强校战略”和“国际化战略”的实质,大力引进大数据领域相关的领军人才**

**加强基于“北斗”空间信息技术的位置大数据处理与服务体系建设**

**开展多种研究 巩固轨道交通优势地位**  
结合学校生命医学行动计划,积极开展医学领域大数据、医学影像学等的研究

**高速列车综合安全技术**  
系统可信性自动验证  
新型智能能效控制及传动技术  
智慧城市与城市安全 大数据与智能  
**列车运行在途检测技术** 北斗导航系统  
轨道交通基础设施全寿命周期信息物理技术  
高速铁路网条件下铁路组织关键技术  
轨道交通牵引供电系统主动运维技术  
监测综合运输大数据处理技术

**学科建设**  
系统推进大数据相关的学科建设,通过实施数字化战略,促进大数据与各个学科的高度融合。

**相关学科**  
分析学科大数据需求,结合学校特色,形成5-6个大数据相关学科方向。

**交叉学科**  
建设10个左右的交叉交叉学科,形成在大数据研究开发等方面的合力。

**基础学科**  
加强数学、统计学、信息科学与技术等基础学科的建设,提升对大数据开发的基础技术。

**教学变革**  
建立新的课程体系、教材体系和授课方式。

**文科构建**  
大数据将促使文科研究范式发生改变。

**社会服务**  
与大数据相关龙头企业合作,加入与大数据相关的国家产业技术创新战略联盟,形成完整的大数据开发应用链。

**企业合作**  
大力推进与大数据龙头企业合作,资源共享,共同研究,开展大数据展示平台、应用平台相关的软件研发。

**科研基地**  
与行业、企业等共同申报国家级科研基地,加快国家大数据创新体系的建立。

**医学应用**  
与校内相关单位和附属医院应用云计算技术和智能技术,共同建设安全电子化的“数字化医院”。

**共享平台**  
加快产业园区与研究院的建设,搭建成果转化共享平台,实时将科研成果和产品推向市场发布。

**数字校园**  
“三网两平台一数据”(有线宽带网、无线网和园区物联网,管理服务平台和教育教学平台,学校公共数据)

**三网**  
建设全校有线无线网,完善校园网络的移动应用。  
有线网升级换代,提高对互联网的访问速度。  
包括一卡通门禁、图书档案、智能楼宇、安全监控等内容的园区物联网。

**二平台**  
管理服务平台: 师生、资产设备、财务、档案信息一体化管理。  
教育教学平台: 互动式的数字化学习平台和虚拟实验平台。

**一数据**  
学科发展: EVALUATING SHARED ANALYTICS  
基本建设: ASSETS  
设备购置: SUPERVISE

对资产的配置和使用进行科学的使用评价和使用监管提供数据支持