

破局数据“孤岛”难题：

健康大数据产业重构“方法论”

■本报见习记者 赵利利



经过一系列探索,健康大数据的应用之路开始逐渐变得清晰。

图片来源:百度图片

“2015年,‘互联网+’和大数据成为国家战略,当一门科学成为国家战略的时候,我们总有好奇心研究一下,大数据到底帮我们解决了哪些实际问题?”在近日召开的2017大数据发展促进委员会年会上,中软国际数据服务线咨询服务总监陈涛叩响了大数据从业者的心头之问。

在陈涛看来,互联网帮人们解决了连接的问题,而连接实现了数据的在线流动,“数据的在线流动则提供了很多不能直接感受到和看到的信息,它加快了我们对这个世界的速度。”

但在健康大数据领域,这种“流动性的魅力”似乎打了折扣。中国疾病预防控制中心慢病中心主任助理蒋伟坦言,“医疗健康这个行业一直都没有跟数据分开过,但是近几年大数据的发展,反而让医疗行业在大数据面前显得比较茫然。”

这是为什么?“大数据想要打破我们原来采用抽样手段通过整体数据来反映健康和医疗全貌的传统。”蒋伟道出了医疗大数据近年来发展中遇到的困惑,在这种情况下,旧的“数据方法论”难以为继,而新的“健康大数据应用体系”尚未成熟,健康大数据产业如何打破僵局?

严重的“孤岛”问题

所谓健康大数据,是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的健康数据的集合,蒋伟介绍道。

“健康大数据构架中数据信息的来源渠道、数据信息种类和大数据数据库的建设与常规大数据是相同的。从数据信息的来源渠道看,很多健康大数据来源于电子病例、居民健康档案、国家临床中心的相关数据、公共卫生及物联网数据库等。”蒋伟认为,正是这种数据种类的庞杂造成了健康大数据的复杂性,甚至成为目前整个健康大数据发展的瓶颈,而这种瓶颈在海量数据收集之后更加凸显。

蒋伟把问题归结为三个方面:数据“孤岛”现象;数据标准不统一;大数据技术未能有效利用,“从2010年左右医疗行业提出大数据至今,一直都没有很好地解决”。

目前医疗大数据的孤岛现象非常严重,“横向来讲,所有的医院都有自己的数据,我们力推医院数据打通,但是将近20年了一直没有达成成果”。蒋伟分析,纵向来讲,由于个人健康信息分为很多类型,从国家层面收集信息就意味着从上到下按不同的类型在地方做不同的数据平台,“各套纵向体系之间的信息没有形成很好的互联互通,大数据很难得以应用,而且造成很多数据资源的浪费”。

百度发布多款软硬件产品“把云带进生活”

本报讯(记者赵广立)11月16日,主题为“Bring AI to Life”的2017百度世界大会在北京举行。会上百度发布了手机百度10.0、Apollo小度车载系统、智能音箱“raven H”、共享Wi-Fi翻译机等多款人工智能(AI)软硬件产品。百度董事长兼首席执行官李彦宏在主题演讲中表示,随着AI时代的到来,百度希望用人工智能技术“把复杂的世界变得更简单”。

在大会现场,李彦宏演示了基于图像识别等技术打造的驾驶员疲劳监测系统,该系统可通过人脸(红外)识别判断,提醒司机避免疲劳驾驶。他透露,百度已与一汽集团签署战略合作协议,将共同推动该项技术产品落地。

疲劳监测系统是“Apollo小度车载系统”的众多功能之一。在智能驾驶论坛上,百度车联网事业部总经理徐勇明介绍说,Apollo小度车载系统提供了人车对话、人脸登录、疲劳监测、智能安全、AR导航、车家互联、个性化服务和内容等核心AI功能,致力于做到对驾驶人“听懂、看懂、关爱、守护”的贴心陪伴。

百度当天发布的产品中,另一款打出“贴心陪伴牌”的当属百度首款“标杆性”人工智能硬件产品“raven H”。在百度智能家居硬件总经理吕骋(前渡鸦科技创始人)的介绍中,基于“DuerOS 2.0”内核与百度数据库的“raven H”是一款“将科技和艺术融合的智能家居大脑”,它在5米左右唤醒的概率可达95%,而在1~3米的唤醒率则高达99%,平均唤醒时间1.4秒,在智能音箱市场上处于领先水平。

记者了解到,成立于2014年5月的渡鸦科技在被百度于今年2月收购之前,就专注于打造基于人工智能和新兴交互为基础的操作系统。被百度全资收购后,创始人吕骋携团队加盟百度,并出任百度智能家居硬件总经理。

吕骋在会上还公布了另外两款即将面世的产品——raven R、raven Q。其中raven R是raven H的“可灵活运动升级版”,不仅具备智能音箱所有功能,还是一款六轴全自动情感机器人;raven Q则被包装为“具备成熟人脸识别、计算机视觉和Apollo自动驾驶的所有百度人工智能时代技术之集成的终极硬件形态”的未来产品,现场并未予以实物或概念呈现。

百度在大会上还推出了“集搜索与信息流双引擎于一体”的手机百度10.0。使用手机百度10.0,用户可通过语音、AR等获取交互更自然、体验更真实的信息流。百度方面提供的数据显示,手机百度累计激活次数已突破20亿,信息流日均阅读总量月复合增长36%,短视频日均播放量月复合增长38%。

在技术平台方面,百度宣布即日起推出目前全球规模最大的“BROAD”百度AI公开数据集计划,包括室外场景理解数据集、视频精彩片段数据集、阅读理解数据集3个数据集。百度副总裁、AI技术平台体系(AIG)总负责人王海峰介绍说,这些数据是百度AI生态的真实数据,有些是首次发布,有些是截至目前行业内发布的最大的数据集。百度AI公开数据集计划将在2018年开放更多数据,并推出供在校师生免费使用的BROAD云计算平台,共同推进AI技术的发展和落地。

此外,蒋伟还表示,在一些信息化水平较高的地区,比如东部沿海地区,或由于企业自身原因,或受制于当时的基础条件,每个地区在建设信息化平台时的标准不统一,当多数据归类归集于同一数据库时,从清洗到达到可预算标准的过程需要消耗巨大的工作量。

业务导向的价值变现

针对大数据分散,总体价值难以变现的问题,陈涛给出了自己的看法。他表示,就目前而言,“大数据还没有厉害到无所不能的程度”。大数据解决的是相关性,并不能解决因果性的问题。他举了个例子,“如果来分析北京交通的拥堵情况,数据能够帮助我们找到很多相关的因素,比如道路建设情况、地铁路线、公交线路设计等。但是数据并没有告诉我们这些因素的成因,以及我们如何利用这些因素来解决交通拥堵的问题,最终数据还需要人来处理加工,从而支持决策的作出。”

陈涛认为,基于这种理解,可以梳理出数据的一些独有特性:业务附着性、真实性、价值不确定性及可复制性。

“数据从业务中来,最终又反馈给业务,支撑业务的改进和创新。大数据应用实际上要做的一件很基础的事情就是先要树立业务及相

关的业务流程。”陈涛表示,“其次,数据可以被当成一种实证方法,既然是实证方法,就必须要有真实,虚假数据毫无疑问。”但陈涛也不讳言具体实践中存在的掣肘,“靠数据本身以及数据所存在的IT信息系统并不能保证信息的真实性,因为数据和人的活动是紧密关联的,保证数据的真实性一定要人为制定很多流程制度”。

谈及数据资产时,很多人会关注,既然数据是资产,是不是可以交易、变现或者“卖钱”?陈涛表示,数据和一般性的商品之间还是有很多区别的,一是数据归属权问题,二是“数据的角度”。陈涛所谓的数据的角度,即“它是不是可以等价交换的方式来支撑”。陈涛解释道,“比如手机上的位置信息,它是数据。对于滴滴打车和外卖公司来讲,它们可以根据这个数据为我提供服务;对于公安部门来讲,这个信息可以帮他们快速找到想找的人,但同样的数据对于其他人可能毫无疑问。数据根据其角色不同,场景不同,目的不同,我们很难认为它的价值像商品一样具有等价交换价值的唯一性。”

此外,数据的可复制性和流动性也是其特性之一,陈涛说:“如果有一天我们发现数据真的成为资产了,真的可以变现的时候,安全非常重要。”

基于这四个特性,大数据应用其实涉及一到一整套完整的方法论。具体到医疗健康大

数据行业,其应用价值的体现也要遵循这样的路径。

建设规范化交互平台

事实上,经过一系列探索之后,健康大数据的应用之路开始逐渐变得清晰。在这个过程中,全球智能穿戴设备已经突破1亿件,我国目前拥有移动医疗用户2.93亿。正如蒋伟所言,“大家都迫切希望利用移动技术得到医疗健康服务”,移动医疗终端与健康大数据的采集和处理带来了新的思路。

可穿戴设备能实现诸多人体生命特征相关数据的采集和连续采集。通过大数据、云计算、物联网等技术应用,实时采集大量的用户健康数据信息和行为习惯,已经成为智慧医疗获取信息的重要途径。蒋伟表示,这些数据有望与电子病例、公共卫生大数据进行有效衔接,加以人工智能分析,从而推进覆盖全周期的预防、治疗、康复、保健的一体化健康服务发展模式,重构健康产业生态链。

尽管可穿戴设备发展蓬勃,但蒋伟认为目前依然存在数据采集单一导致的健康数据分析、健康服务提供不精确和行业数据标准不统一导致的数据难以共享的问题。“哪怕是同一智能穿戴设备,根据厂商自己的利益和价值取向,设备的数据标准也不一样,当数据真正拿来利用的时候,往往得不到医疗机构的认可。”蒋伟说,“对于健康数据来讲,它的保质期很短。”

“从整个行业来讲,需要一个规范化的平台。”蒋伟认为,这个平台要有两大功能,一个是汇总不同智能硬件的健康数据,集合成用户的整体运动健康数据电子档案,为医疗数据接入、慢病管理等提供可操作的健康数据依据;二是将同一用户多维度的排重数据,反向传输给智能硬件厂家,用来精确算法,优化产品,从而形成用户健康信息闭环。基于此,中国疾病预防控制中心和中国信息通信研究院标准所正在共同探索建立健康管理信息交互平台。

为什么要做这个平台?蒋伟表示,首先,针对数据来源的问题,力图使用平台去衡量、支持可穿戴健康数据采集数据的质量标准,开展健康服务评估。其次,融合各个采集设备的数据孤岛,实现数据共享,通过大数据算法提高健康医疗服务精准性。再次,面对大量的运动健康、身体指标、医疗信息等数据,提供技术、服务、用户等行业交流平台,推动我国智能健康医疗大数据的发展,实现移动医疗健康大数据的真正价值。

蒋伟最后讲述了平台建设的时间表,“2017年底一期试运行,2018年正式上线,预计2018年底实现300家移动健康设备平台接入”。

苹果被诉专利侵权 案中案初现端倪

■本报记者 李晨

11月14日,西电捷通公司诉苹果公司专利侵权案,由陕西省高级人民法院在西北政法大学公开开庭审理。经过此前多轮庭前质证之后,这起备受关注的中国标准专利侵权案正式进入关键的庭审阶段。

庭审陈述中,西电捷通方面认为苹果公司在其iPhone 6系列产品,iPhone 7系列产品及iPad平板电脑等产品中实施了其持有的一项专利,并根据专利侵权及侵权产品的实际销售情况,提出了1.5亿元人民币的索赔请求,同时申明保留根据被告侵权持续时间内的侵权产品实际销售数量,对所遭受的直接损失或间接损失增加索赔金额或另案索赔的权利。

该案在正式庭审之前便广受关注,这不仅在于它是继ICT领域“中国标准必要专利第一案”——西电捷通诉索尼专利侵权案后,西电捷通就同一涉案专利发起的第二起维权诉讼,且被告为目前全球市值近万亿美元的公司,还在于该案引发了系列“案中案”——这在外界看来,也预示着该系列案件的审理将旷日持久,耗资巨大。

纠纷源于一项标准必要专利

西电捷通和苹果公司的纠纷始于一起无线网络安全技术专利的维权诉讼。2016年4月,因认为其15年前研发并持有的名为“一种无线局域网移动设备安全接入及数据保密通信的方法”的专利(专利号为02139508.X)遭受侵权,西电捷通以专利侵权为由,将苹果电脑贸易(上海)有限公司、西安市国美电器有限公司诉至陕西省高级人民法院,要求法院判令被告立即停止专利侵权行为,停止销售侵犯西电捷通专利权的手续产品及赔偿经济损失。

该涉案专利系西电捷通研发推出的无线网络安全技术专利 WAPI(无线局域网鉴别与保密基础结构)的一项核心专利。

目前,全球WLAN已形成相对统一的技术架构,但安全技术分为两条发展

路线:一是美国主导的IEEE 802.11i标准,另一个就是我国主导的WAPI标准(WAPI技术于2003年被国家标准采纳)。基于上述技术路线形成的无线网络,业界分别称为Wi-Fi网络和WAPI网络。而该案涉案专利即为WAPI标准采纳的必要专利之一。就在今年的10月16日,Wi-Fi最高安全等级防护机制WPA2技术标准被曝出存在重大漏洞。

另据了解,为推动WAPI在全球市场的应用进程,2005年5月,西电捷通针对WAPI技术在美国提交了专利申请,于2013年10月获得美国授权。此外,2010年6月1日,WAPI核心支撑三元对等安全架构的核心技术实体鉴别(TePA-XX)成为国际标准,这是我国在网络安全基础技术领域第一个国际标准。

2010年,苹果曾与西电捷通签订了专利协议。双方遵照“公平、合理、无歧视”原则,明确2010~2014年的5年期间的专利许可费。合同到期后,原告方指出,“虽经西电捷通反复交涉,苹果公司拒绝与原告就专利许可进行实质性的磋商,反而以各种理由恶意拖延拒绝支付相关费用”。在经历了与苹果方面历时16个月的商讨但未取得任何实质性进展后,西电捷通方面认为苹果的种种举动纯属恶意磋商,遂发起专利侵权诉讼。

此次庭审过程中,双方就各自理由进行了举证和陈述,但未当庭展开辩论。双方争议焦点归纳为三点,即:涉案专利是否为标准必要专利;是否构成侵权;是否及如何实施侵权行为;如果侵权,侵权责任如何承担。对此,原告、被告双方均表示无异议。审判长随即宣布,接下来主要针对侵权判定问题展开审理。

随后,围绕涉案专利是否为标准必要专利、被告是否侵权问题,双方分别通过多媒体演示、专家辅助人证言及律师口头说明等方式,进行了举证和意见陈述。

案中有案

在西电捷通于2016年4月起诉苹

果公司专利侵权后,后者随即针对西电捷通涉案专利发起了专利无效请求。2016年5月,苹果方面以说明书公开不充分、缺少必要技术特征、权利要求得不到说明书的支持、不具备新颖性和创造性等理由,向国家知识产权局专利复审委员会提出涉案专利无效宣告请求。在获得专利复审委员会驳回后,苹果方面遂向北京知识产权法院提起诉讼,请求法院判决撤销被诉决定并判令专利复审委员会重新作出审查决定。

在发起专利无效请求后,苹果方面又接连针对西电捷通发起“专利使用费率”诉讼以及“滥用市场支配地位”诉讼。

在当日庭审中,原告代理律师指出,苹果方面在前述针对西电捷通的两个诉讼案中,均使用了“公平、合理、无歧视原则”以及“禁令”等标准必要专利才会牵涉的审理问题及术语,这也是原告认为被告苹果公司承认其产品中使用西电捷通的标准必要专利的理由之一。

“案中案”的复杂性还不止于此。在这次庭审中,原告代理律师还提到,在此前庭审及同一专利纠纷的索尼案审理过程中,苹果向索尼方面非法提供了重要帮助,向其非法泄露原告的商业信息,以期达到联合抵制支付专利许可费的目的。

2015年6月,西电捷通针对同一涉案专利侵权问题,向索尼移动通信产品(中国)股份有限公司(以下简称索尼)提起了专利侵权诉讼,并于2017年3月22日获得一审胜诉。

随后,索尼方面不服判决并提起上诉。据业内人士查证和披露,2017年7月14日,索尼向美国地区法院提起诉讼,要求从苹果公司调取苹果与西电捷通之间的许可协议,并最终及时调取了苹果应当对第三方承担保密义务的文件。

西电捷通有关负责人告诉《中国科学报》记者,针对此次专利侵权纠纷,苹果方面一直采取不正当甚至不合法的手段予以应对。针对这种做法,他们会坚决维护自己的权益。

简讯

“补贴退坡”开启新能源汽车新格局

本报讯11月16日,为期两天的2017第二届动力电池应用国际峰会暨第三届中国电池行业智能制造研讨会在北京拉开帷幕。中国工程院院士杨裕生表示,“补贴退坡”开启电动汽车新格局,主要表现为四个方面:开启企业主导培育市场的阶段;汽车企业要主动消化补贴退坡的“差额”;积分与纯电里程挂钩,副作用不可低估;发展电动汽车的初衷亟待重温。

本届峰会由中国化学与物理电源行业协会、电池中国网和天津力神电池股份有限公司联合主办。与会人员就新能源汽车及动力电池取得的成就、存在问题、解决方案、发展机遇等进行了探讨。(赵利利)

阿里巴巴宣布建立广东研发中心

本报讯11月22日,2017云栖大会·广东分会在广州举行。阿里云总裁胡晓明宣布,将设立阿里云广东研发中心,招募1000名云计算和人工智能工程师,推动前沿技术与广东产业融合。在大会上,阿里云还宣布将在广东建设其工业互联网云平台,将全国工业云总部定于广州,依据企业需求和场景,深度再造ET工业大脑,目标是服务全国10万家制造企业。

目前,广东一批制造业骨干企业已经开始使用阿里云的云计算和人工智能技术,来解决转型升级过程中的痛点问题。在大会现场,珠江啤酒宣布与阿里云达成战略合作,推动互联网、大数据、人工智能与啤酒产业的深度融合。

此外,OPPO、珠江钢琴、金立、美的等广东企业也出现在当日的发布会上,这批大型制造企业已经率先尝试使用前沿互联网科技与制造业融合,成为广东具有典型效应的工业互联网标杆项目。(彭科峰)

金山软件今年第三季度营收超13亿元

本报讯11月21日,金山软件有限公司发布2017年第三季度财报。财报数据显示,金山软件第三季度营收达人民币13.03亿元,同比增长16%。其中,网络游戏、云服务、办公软件及服务以及其他业务贡献的收入分别为7.47亿元、3.58亿元、1.98亿元,分别占总营收的57%、28%和15%。

金山软件董事长雷军表示:“2017年第三季度,金山软件各大业务板块均有优秀的表现。端游《剑网3》收入稳步提升,手游《剑网3》全新布局;金山云各垂直领域加足马力,培育核心竞争力领跑业界;此外,WPS办公软件实用用户数量再创新高,PC版全球月活用户量突破1亿。未来,我们将秉持创新、严谨的态度,力争保持网络游戏、云服务及办公软件业务的稳健增长。”(计红梅)

英特尔发布商用5G新空口调制解调器系列产品

本报讯11月17日,英特尔宣布其无线产品路线图取得重大进展,以加快5G的普及。

据介绍,其中亮点包括:推出英特尔首个支持5G新空口(5G NR)的多模商用调制解调器家族英特尔XMM 8000系列,以及英特尔最新LTE调制解调器英特尔XMM7660。英特尔还宣布已经成功实现了基于英特尔5G调制解调器的完整端到端5G连接,而这套早期5G芯片是英特尔发展史上的一座关键里程碑。当天的宣布还包括,在2017年世界移动通信大会上发布的英特尔XMM7560调制解调器已实现千兆级速度。

据悉,在2017年CES上宣布的英特尔早期5G芯片已经在28GHz频段上成功实现5G连接。连同英特尔移动试验平台,英特尔技术已经成为全球数十个5G测试的核心。未来英特尔将延续自己在Wi-Fi上的领先性,快速实现千兆级Wi-Fi的商业化,并计划从2018年开始推出下一代802.11ax Wi-Fi。(计红梅)

达能减排目标获国际倡议组织批准

本报讯11月15日,达能集团宣布公司的温室气体减排目标已经获得“科学碳目标”倡议组织(英文简称SBTi)的正式批准。同时,达能还宣布加入旨在改善农业、粮食安全和气候的“千分之四”碳封存项目。

SBTi是由“碳排放披露项目”、世界资源研究所、世界自然基金会和联合国全球契约组织联合发起的合作组织。SBTi认为,达能的减排目标与把全球气温升高控制在2℃以内的全球应对措施相符。

达能首席执行官范易谋对记者表示:“我们希望能与合作伙伴携手助力食品体系的变革,从达能供应链中的14万农户开始,打造以健康和可复原土壤为基础的可再生能源模式。”达能透露,在2008年至2016年期间,在除农业以外的包括运营、包装和物流环节在内的价值链上,达能的排放强度已经降低了50%。(赵广立)

2017PIMC未来领导力发展论坛举行

本报讯11月18日,由PIMC领导力发展研究院、咪咕文化、科大讯飞联合主办,长江商学院EE首席人才官学会(CHO学会)协办的2017PIMC未来领导力发展论坛在北京召开。

中国正和世界同步进入一个创业大繁荣的时代。然而,当前许多企业面临着借力科技创新发展的新趋势,都在不断寻找科技创新的切入点。该会议正是围绕这一主题展开各个议程的探讨。

论坛发言嘉宾围绕“团队崛起”将如何影响组织成长与领导力变革,如何挖掘“微数据”的管理价值,人工智能技术与组织与人才成长到底意味着什么等三方面与现场参与者进行了分享与探讨,发掘数据和物联网对企业管理的赋能。(赵广立)

第65届ISSCC峰会将在美国加州召开

本报讯11月17日,“2018芯片奥林匹克—IEEE国际固态电路峰会(ISSCC 2018)中国北京发布会”在京召开,该峰会将于2018年2月11日~2月15日在美国加州三藩市举行。

IEEE ISSCC被称为集成电路行业的芯片奥林匹克大会。峰会录用和发布了全球顶尖芯片及企业最新和最具研发趋势领先指标的芯片成果。ISSCC 2018录用的202篇论文来自18个国家的一流大学和研究机构及顶尖集成电路企业,主要领域包括了射频技术、电源管理和能量采集、无线发射器和有线通讯、模拟电路、前瞻技术等。(李惠钰)