

大数据的开放式创新

Open Innovation of Big Data

吴甘沙/WU Gansha

(驭势科技有限公司, 北京 100080)
(UISEE Co.Ltd., Beijing 100080, China)

大数据创新的最高境界是用构建数据生态来改变竞争格局——数据源解决数据供给,数据创意者从数据中创造价值,而这又有赖于大数据处理和分析技术。在开放式创新的体系中,5种元素扮演3种角色。

- 数据源: 开放数据, 基于数据安全流通和定价的数据市场;
- 大数据分析和处理技术: 开放的基础设施, 以及开放的社会化分析服务;
- 数据创业者/应用服务: 跨越领域界限的开放数据思维。

它们五行相生, 互相作用, 形成价值的涌现。

1 开放数据的发展及问题

数据开放的主体首先是政府和科研机构, 即把非涉密的政府数据, 以及纳税人支持的一些科研数据开放出来。越来越多国家推出了统一的政府开放数据门户。中国在2015年也推出了《促进大数据发展行动纲要》, 将开放数据作为工作重点。在开放数据运动的风起云涌之下, 现在更多的企业也开始开放数据, 实现数据的价值化, 并建构生态系统和护城河。

收稿时间: 2016-01-12
网络出版时间: 2016-02-29

中图分类号: TP393 文献标志码: A 文章编号: 1009-6868 (2016) 02-0002-05

摘要: 大数据是社会从网络化演进到智能化的技术基础, 更是未来数字经济的基础资产和货币。认为目前大数据的创新主要局限在技术栈和组织内部, 数据的可获得性、处理和分析技术的缺乏以及封闭系统的数据思维成为制约创新的因素。提出大数据开放式创新的要素: 通过开放数据及基于数据安全流通和定价的数据市场解决数据供给, 开放基础设施及社会化分析服务实现技术共享, 最后通过跨领域的开放数据思维获得数据创意。认为开放式创新重构了数据生态, 将改变大数据的竞争格局。

关键词: 大数据; 开放创新; 匿名化; 数据定价

Abstract: Big data is the technical foundation of an evolving society, from the networking to intelligent age, and plays the role of critical assets and currencies of future data economy. However, today big data innovation is limited to technical stacks and within the organizations, and suffers from unavailability of data, lack of processing and analytics technologies, and closed-world thinking. This paper discusses key factors of open innovation for big data: unleash the data supply via open data and data marketplaces with secure exchange and pricing, democratize the technologies through open data infrastructure and socialized analytics services, and finally harvest innovative data ideas by “crossover” thinking. Open innovation restructures the data ecosystem and will reshape the competitive landscape of big data.

Key words: big data; open innovation; anonymization; data valuation

万维网之父 Tim Berners Lee 提出了数据开放的五星标准^[1], 以保证数据质量: 一星是开放授权的格式, 比如说 PDF; 二星是结构化, 把数据从文件变成了像 Excel 这样的表; 三星是开放格式, 如 CSV; 四星是能够通过统一资源标识符 (URI) 定位每一个数据项; 五星是能够跟其他数据链接, 形成一个开放的数据图谱。

数据开放与开源软件也形成了共振。主流的数据开放门户, 像 data.gov, 都基于开源软件。Data.gov 用 WordPress 做数据内容呈现, 用 CKAN 做数据目录, 甚至 data.gov 的整个架构也在 GitHub 开源了。英特尔在麻

省理工学院的大数据研究中心研发了开源的 DataHub 系统, 支持对开放数据的多人协作分析, 具有数据版本管理和多编程语言交互的能力。

数据开放中会遇到很多问题。

(1) 数据权属的问题。数据属于谁? 属于采集人? 还是属于生产者? 抑或是属于被观察的客体? 在特定情况下, 拥有权如何分割 (比如离婚) 或者转移 (比如继承)?

(2) 敏感数据的界定。比如位置信息数据在欧洲属于敏感数据, 而在日本不属于敏感数据。另外各个不同行业有进一步规定, 比如美国的《健康保险便利和责任法案》对个人

健康信息的隐私性、机密性和完整性做了规定；而在征信领域则有《公平信用报告法》对个人信用方面的信息做了规定。敏感数据需要法律和行业法规的界定。

(3)敏感数据的脱敏。如果开放数据中具有敏感数据,就要做数据的脱敏。脱敏最简单的做法是去标识,但是去标识未必能够彻底脱敏。美国研究显示:即使把姓名、地址等标识信息拿掉,只要有邮政编码、性别、生日等3项信息,就有60%~90%的可能性锁定个人。即使去标识很彻底,仍有“阿喀琉斯之踵(致命弱点)”。一种攻击的方法是通过多数数据源的比对来缩小搜索范围,重新标识;另一种方法是基于统计的攻击,比如根据两个打分再加上一定的时间范围约束,还是有接近70%的可能性锁定个人。

(4)防止隐私攻击的匿名化技术。比较典型的如k-anonymity和L-diversity等,但在敏感属性不够多样化,或攻击者具有背景知识时,这两种技术仍不够鲁棒。目前最好的一种技术叫差分隐私,即把噪声加入到数据集中,但仍保持它的一些统计属性,支持特定的机器学习算法。

这些困难和挑战都不能阻挡开放数据运动的深入人心。在数据(尤其是商业数据)仍然无法充分流通的今天,开放数据无疑能够让具有数据思维和分析能力的创意者点石成金,把死的、消耗成本的数据变活、创造利润。

2 基于数据安全流通和定价的数据市场

数据之于数据社会,就如同水之于城市或血液之于身体——城市因河流而诞生,也受其滋养;血液一旦流动停滞,身体就有危险。所以,在数据化生存的今天,一定要让数据流动起来。数据开放更多适用于政府公共数据和纳税人资助的科研数据,而更多涉及个人隐私或企业机密的

数据无法通过简单的开放获得。如果把数据看作一座冰山,公开的只是露出海面的一点点,绝大多数藏在暗黑的海面以下。

数据拥有者不愿意把数据拿出来,有两个原因:担心数据被偷窃;对自己并无好处。所以,解决时该问题时需要把握两点:保障数据的安全流通;对数据的使用进行定价,而实现这两个关键的载体是数据市场。

数据市场并非是新概念。早年的综合数据市场多进行原始数据集的下载交易,由于数据容易复制,版权保护困难,这种形态逐渐被几种新的形态取代:

(1)为特定用户定向采集或加工数据,如某公司从事人脸分析技术,委托第3方采集各类、各种姿态和光照条件的人脸数据,或某公司具有大型数据集,需要特定的服务来做标注。

(2)专业领域的数据服务,如交通领域的Inrix或金融领域美国三大征信公司。

(3)不给出整个数据集,只能基于查询或应用程序接口(API)提供数据的受控访问,中国出现的数据交易市场多为此类型。

(4)不给出原始数据,只交易加工信息,这是之前大数据时代的主流,有些公司(如彭博社)甚至提供专门的终端保证信息服务。

随着数据生态的完善,数据市场的形态将更为丰富。首先,上述形态多为数据提供者与数据请求者的简单交易关系,而未来市场的参与者可能同时是提供者与请求者。其次,交易将不仅是简单的“给”和“得”,而是融合、使用从而产生新的衍生价值。因此,数据的定价不是那些比特的固有价值,而是在这一次“使用”中产生的当前价值。数据市场应该是使用和买卖一站式服务,并且是先使用再买卖。

Steven Johnson的TED演讲《伟大创意的诞生》是从咖啡馆说起,它创

造了一个安全的空间,让不同的人做思想碰撞,创造新的想法。数据何尝不需要这样一个咖啡馆,让各方的数据能够产生“化学作用”。“数据咖啡馆”项目^[2]基于多方安全计算,试图解决3个问题:安全可控的开放;数据市场和云计算的一体化;数据定价的问题。

然而,绝大多数数据的价值是不确定的,这正是数据的外部性。这种属性决定了数据与石油本质上的区别:石油的价值在燃烧的一瞬间实现并消失了,但数据能够反复使用,产生不可预期的新价值。基于Moody的信息估值七律,可以衍生出数据估值七律:

(1)数据可以被无限次共享,可以产生更大的总体价值,但多次复制会使所有权复杂化,增加成本;

(2)数据用得越多,价值越大;

(3)数据价值会随时间衰变;

(4)数据越精确,价值越大;

(5)多个独立数据源的融合为1+1>2;

(6)更多的数据不见得能带来更多的价值;

(7)数据不会损耗,反而会越多越多。

这些基本原则对数据的定价具有指导意义——数据的使用频度、新鲜度、质量、外部性等都是重要变量。Glue Reply公司据此提出了基于使用的估值模型。

另一方面,Gartner分析师Doug Laney——大数据3V的提出者,把信息和数据的估值模型分成非金融模型和金融模型。

我们期待未来的数据市场有灵活的数据定价模型,该模型既考虑数据的使用历史和时间嬗变所形成的基础价值,又能计量当前的这次租用中可量化的价值,计算出这次交易的数据定价。同时,如果这次使用有多方数据参与,根据各方在计算中贡献的大小,对其数据分别进行定价。

数据的安全流通和定价将鼓励

数据拥有者将其数据参与流通,对其数据价值化、货币化和资产化,从而形成“收集-使用-价值化-更多收集-更多使用”的正向反馈,为开放式创新提供更广泛的原材料供给。

3 开放的基础设施

笔者的同事 Eric Dishman 罹患肾病 23 年,尝试了各种治疗方案,甚至换肾,一直没有进展,直到他选择了基于基因分析的精准治疗。整个测序和锁定致病基因片段的过程花了 3 个月;接着,数 TB 的基因数据被拷到硬盘里,在美国东西岸传来递去,颠簸了 4 个月以后方才形成了治疗方案。虽然他现在已经恢复健康,但 7 个月的等待对于任何一个病人来说都是煎熬。

原因很简单,对于专业的医疗健康和生命科学机构来说,计算和存储的基础设施并不是他们所擅长。要知道,就连大数据领域内部也是隔行如隔山,做数据分析的人很难理解分布式的存储和处理系统。事实上系统部署的困难已经成为目前拦在大数据产业前面的一座大山。

要致富,先修路(基础设施)。在现实生活中的这个朴素道理也适用于大数据。基于云计算的公共基础设施,特别是大数据系统作为平台服务,是搬走这座大山的希望所在。在其他的一些国家,很多以数据思维见长的小型创新企业已经开始受益于这一趋势。

Decide.com 是笔者一直关注的一家创业公司(后被 Ebay 收购)。它每天吸入几十万条商品价格数据以及相关的新闻(这也是开放数据),分析后告诉顾客买什么牌子、型号以及预测何时买最划算。在其神奇的背后,只有 4 个博士精心调制算法,他们不用担心基础设施的问题,因为亚马逊已经把计算和存储能力作为基础设施开放出来了。

Prismatic 是另一家创造神奇的公司,它能读懂用户关心什么,发掘用

户新的兴趣,实时地、个性化地推荐阅读。这家公司在很长一段时间内只有 4 个员工,3 个是学生,然而估值已经达到好几亿美金。之所以能够把神奇的数据思维变成现实,同样要感谢亚马逊的云计算把脏活累活都干了。

把大数据系统装在云上第 1 代大数据奋斗者的梦想。早在 2007 年, Hadoop 解决方案的领导者 Cloudera 成立伊始,就已经在憧憬这一愿景(从 Cloudera 这个名字可以看出)。然而,这条道路并不顺利。

首先,把 Hadoop 这样的重型系统跑在虚拟机里是很大的挑战,大数据这样的输入输出(I/O)密集型应用与虚拟化技术有点“水土不服”,性能下降严重。经过业界和社区多年的努力,这如今已经不是问题。而像 Spark 这样的新贵是生在云里,长在云里,与云相得益彰。

其次,对于大数据的早期用户来说,把数据放在云里是有疑虑的,一来大数据的搬动太过困难;二来数据安全没有保障。这些年来,云计算的积累效应悄然间改变了数据生态,越来越多的数据一生下来就在云里。而对于初尝云滋味的客户,亚马逊甚至专门设计了容量达 50 TB 的、可托运小型存储设备帮助他们把数据搬到云里。Spark 的商业化推动者 Databricks 也顺势与亚马逊结盟,在其 AWS 云服务上部署 Databricks 云,可以利用大量已经存在于亚马逊云的数据,这真是一个妙招。

而数据安全的保障有赖法律法规、行业自律和技术推动三箭齐发。目前关于大数据权利的立法已在酝酿之中,行业规范更是走在前列(如第 1 节所述)。在行业自律上,我们看到了阿里云发起的《数据保护倡议》。然而,没有技术推动,法律法规和行业自律会制约大数据的云部署。本小节开始讲的基因数据在磁盘里周游世界的故事,还是会一再重演,因为美国的《美国健康保险便利

和责任法案》对数据在网络上的传输施加了很多限制。

Eric Dishman 的癌症经历引起了计算机科学家的深思。男性有一半的几率罹患癌症,女性的几率也达到 1/3。相比之下,过去 50 年癌症的治愈率只提升了 8%,在各种疑难重症中进步最小。究其原因,癌症作为一种长尾病症,需要足够多的数据样本才能有所突破,而《美国健康保险便利和责任法案》等法规对于数据共享的限制使得各大科研机构只能各自为战,相对较少的数据样本制约了生命科学技术的发展。

想象一下,如果第 2 节中所谈的多方安全计算技术能够使数据在法规允许的范围内共享和互通,癌症研究将大不一样。鉴于此,英特尔和俄勒冈健康科学大学等科研机构开始陆续推动基于安全多方计算的协作癌症云。

我们预计:随着云观念越来越深入人心,大数据和高性能计算在云中的部署将呈现加速之势。这时候,云作为一种开放基础设施的优势将得到充分展现。

还是回到 Eric Dishman 的案例。历时 7 个月的诊断过程固然有数据磁盘在路上的延误,另一个重要原因是计算基础设施的缺乏。在生命科学领域中(尤其是生命信息学),非常罕见地呈现了高性能计算和大数据分析齐头并进的态势,寻常的科研院所无法维护完美支持两种运算的基础设施。

可以想见:未来的几年中融合高性能计算和大数据分析能力的云基础设施将变得普及。我们有一个雄心勃勃的愿景:到 2020 年,像 Eric Dishman 这样的患者,一天之内就能完成全基因组测序,锁定致病基因,且形成个性化用药和修复方案。相比起他 7 个月的经验来说,计算能力与时俱进的开放基础设施能缩短数百倍的等待时间。另一个非常热门的领域——脑科学研究如今也面临

计算力有不逮的局面,一次功能性核磁共振对大脑的完整数据采集将获得 500~600 GB 左右的数据,而对其进行完整的分析耗时 6 h。我们期待 2020 年这个工作将在 1 s 内完成,也就是说,能够对脑部活跃成像做一些实时的分析,这对脑科学和类脑计算的研究来说将打开一扇前所未有的大门。

4 开放的社会化分析服务

《哈佛商业评论》说数据科学家是 21 世纪最性感的职业。而麦肯锡认为:2018 年前美国这类人才的缺口达到数十万,特别是能够做深度分析的分析师有 50%~60% 的缺口。也难怪,一个合格的数据科学家必须精通数理统计和计算机科学,对数据敏感,对业务理解。现有的计算机科学或数学的教育体系,无法批量生产这样的人才。我们看到基于慕课(MOOC)的数据科学课程获得了数百万学生的参与,很多大学开始推出在线数据科学课程和学位,相信基于互联网的新型教育体系将在人才供给中扮演更重要的角色。但是,短期内人才饥渴是非常现实的问题,这对于矢志立于大数据潮流之巅的企业来说,不免英雄气短。

与此同时,一股轰轰烈烈的资源革命在互联网卷过,共享经济充分利用互联网将闲散资源与需求对接,解决了供需失衡的问题。设想数据科学家的技能和时间也是一种资源(克莱·舍基将其称作“认知盈余”),应该也能够在这一框架下提高使用效率。这就是所谓的开放的社会化分析服务。

这种服务对我们的社会来说并不陌生。某种意义上,这是一种古老智慧“悬赏”和现代“众包”思维的合体。1714 年,英国议会悬赏 20 000 英镑的“经度”大奖促使一个钟表匠发明了航海天文钟,完全改变了航海史和征服史。18 世纪,拿破仑悬赏 12 000 法郎征集储存食物的方法,促使

一个商人之子发明了罐头。近现代史上这样的悬赏还有很多,比如跨大西洋飞行、月球车、宇航员手套等。另一方面,众包完全改变了当代知识的生成和解决问题的方式,比如维基百科。

那么,开放的社会化分析服务该如何工作呢?下面我讲几个故事。

Netflix 在 2006—2009 年之间向大众发起数据分析挑战赛,希望能够通过预测用户星级评分来提升推荐引擎的效率,目标是提升 10%,为此设了百万美金大奖,吸引了全世界 180 多个国家 4 万多支团队来参加。非常可惜的是 Netflix 没有采用第 1 名的算法。那这个比赛是否没有价值呢?不然,大数据生态系统中最受关注的 Spark 平台正是因为这个比赛形成了灵感和最早的原型。大赛的价值往往不在赛场里。

第 2 个故事关于休利特基金会。它征集一个对学生的短论文进行自动化评分的算法,因此设立了 10 万美元奖金的 Automated Student Assessment Prize。第 1 轮大赛先向十多家专业的教育科研机构开放,而第 2 轮则是在 Kaggle 平台上向社会开放。Kaggle 坐拥数十万具有专业知识和自由时间的分析师,而具有数据分析需求的企业只要把数据和挑战赛规则放到网上,分析师们就可以八仙过海、各显神通、一较高低。结果出人意料,这些业余爱好者搞出来的算法,远胜于专业机构的算法。更让人大跌眼镜的是前 3 名获得者分别是美国一位机械工程专业的本科生,斯洛文尼亚一位计算机系的博士生,和新加坡一位 39 岁的保险精算师。第 1、3 名获奖者刚刚从 Coursera 慕课平台上学完了斯坦福机器学习的课程,刚刚学完去参赛,就摘得桂冠,这是非常颠覆的。Netflix 大赛的获奖团队都是高大上的科研人员,包括两个 AT&T 的研究主管,而这次竟然让几个初通机器学习门径的学生拿到了大奖。竞赛改变了学生的命运,第 1

名转向了数据科学专业,而斯洛文尼亚和新加坡的两位优胜者在美国找到了职业发展的巨大空间。

第 3 个故事是关于一家很小的初创公司 Jetpac,它在 IPAD 上做一个关于旅游的应用。这个公司非常小,做技术的两个人,一个 CTO,另一个是程序员,他们希望有一个自动化的算法在很多照片中筛选出最好的照片。但两个人学识有限,于是他们在 Kaggle 平台上搞了一个比赛,因为资金有限,就出了 5 000 美金,没想到还是吸引到了 400 多支团队参赛,最终他们确实选到了一个合适的算法,让这个应用脱胎换骨。Jetpac 马上就拿到了 240 万美金的风险投资,他们的精明之处在于:利用社会的资源为其贡献才智,换来资本的青睐。

对于当前的“大众创业、万众创新”,数据科学的专业性门槛必然导致洛阳纸贵;而这样的思想众包平台将解决数据智慧的短缺,提升众创的成功率。

大家试想,Kaggle 这个平台,也就数十万注册用户,咱们中国毕业生每年都是千万,学科学工程专业的也有好几百万,在中国可资利用的社会化分析力量一定更为强大。

鉴于此,中国计算机学会大数据专家委员会主办了“中国好创意”全国青年大数据创新大赛。首先,它是学生学习数据科学,切磋数据分析技术的平台;第二,像中国好声音一样,它一定是年轻人展现自己的平台,就像吴晓波所言,这个时代是无名山丘崛起为峰的时代,这个时代需要这么一个平台;第三,操作系统 BSD 的发明人 Bill Joy 提出了 Joy 定律:在这个时代,无论公司再牛,世界上最聪明的绝大多数人都是为其他人工作的。那么最好的办法就是打开组织的边界,让组织虚拟化,让世界上成千上万的人帮忙你解决难题。同时,对于数据科学家/工程师来说,数据分析能力将成为其行走江湖的独特品牌,纵横于不同企业之间,最大化

其价值。

5 跨领域数据思维

2013年,一种病毒在上海和安徽爆发,国家派出了很多工作组,前往各个现场采样,对10 000个样本进行分析。他们寻找的是H7N9禽流感病毒。笔者当时在想,我们的生物科技人员要是大数据思维多好!早在2005年,Craig Venter——这位被称为“科学界Lady Gaga”的奇人,已经在对纽约的空气做全体的基因组测序。如果对源头菜市场的空气做全体的检测,不正是大数据全集思维相对于采样的优势吗? Venter的跨界思维并不止于此。2014年,他的创业公司“人类长寿”从Google挖走了顶级计算机科学家,谷歌翻译首席科学家Franz Och。在这里,Franz将运用大数据去解密人类基因组的奥秘^[3-5]。

同样,生物科学的思维也能帮助大数据。百度首席科学家吴恩达,曾经一度迷惘人工智能走进了死胡同:识别杯子需要一种算法,识别人脸又是一种算法,识别汽车还要一种算法,似乎永远无法穷尽人的智能。直到有一天,神经科学方面的最新进展让他大开眼界:科学家把大脑皮层负责听力的区域与听力器官的神经连接剪断,连到视网膜,过了一段时间,这部分区域竟然能够形成视觉理解了;同样,负责触觉的区域也可以被训练成具有视觉功能。吴恩达获得了顿悟:原来人脑只有一套算法实现各种认知功能,从此他走上了深度神经网络的研究之路。

Farecast.com是人工智能学者Oren Etzioni开的一个创业公司(后被微软的Bing收购),他携数据思维切入到了航空公司白热化的价格竞争之中。通过洞悉机票随季节、燃油价格、天气状况甚至特定事件的变化趋势,他推出了机票价格预测服务。如果到此为止,这不失为一个精彩的跨界数据思维案例,但真正使其成为经典的是:在预测服务后Farecast.com增

加了10美元的“Fareguard”保险服务,如果购买后一周内价格下跌,公司将补足差价。

前文的另一个案例Decide.com帮助顾客预测某个商品何时买最划算。同样,Decide.com对于某些商品提供价格保险,如果消费者购买后一段时间内商品降价,那么公司会补偿差价。

The Climate Corporation把气候学和农艺学揉在一起,告诉农民播种的时机,或为恶劣天气做好准备。真正天才的创意在于:他们把保险业引入到三角关系中——通过微气象建模预测异常气候的发生,帮农民办理保险,并在气象灾害发生后,自动理赔、打款。当气候学、农艺学和金融学以一种全新的方式组合在一起,造就了一家10亿美元的公司。

读者从上述的3个例子能够读出什么?

大数据的预测分析和保险是完美搭档,创造了新的商业模式。推而广之,大数据的预测分析与金融也能产生很多新的商业机会,因为金融本质上就是跨越时空的价值交换,而大数据则能够发现时空之间的价值剪刀差。这毫无疑问也要拜跨界思维之赐。

相比信息,数据的价值有很高的外延空间。信息的意义是明确的,价值也是确定的。而数据有外部性,它因为某种目的被采集,又可以无限服务于新的目的。克强指数采用的3个数据——耗电量、铁路货运量和贷款发放量,都不是为衡量经济运行状况而设计的,然而总理跨界的数据思维使其能够反映中国的经济全貌(必须指出,这些数据反映的更多是重工业运行情况)。同样,智能电表采集的社会用电情况不经意间反映了房屋空置比例。数据的这一奇特特性亟需跨界思维去挖掘。

在大数据的开放式创新中,不只是需要技术的开源,更需要思想的开源。如果能够把世界各地、各行各业

的跨界数据思维及其实践内容档案化,加入检索功能,数据智慧就能得到积累和传播,真正让大数据之光普照大众、惠及我们的地球和城市。

6 结束语

文章从5个方面阐述了大数据的开放式创新。我们期待通过开放式创新,中国能够出现一万个、十万个甚至百万个数据思维公司,他们如群星般璀璨,秉持知行合一,或净化环境,或改善民生,或推动产业转型升级,或提升社会治理,形成一股巨大的力量,实现大数据在中国的繁荣!

参考文献

- [1] Linked Data [EB/OL]. (2016-07-27)[2009-06-18]. <http://www.w3.org/DesignIssues/LinkedData.html>
- [2] 吴甘沙. 大数据技术发展的十个前沿方向[J/OL]. 大数据, 2015(2) [2015.08.28]. <http://www.j-bigdataresearch.com.cn/CN/10.11959/j.issn.2096-0271.2015023>
- [3] MOODY D, WALSH P. Measuring the Value Of Information: An Asset Valuation Approach [C]// Proceedings of Seventh European Conference on Information System (ECIS'99), Copenhagen Business School, Frederiksberg, Denmark, 1999
- [4] Reply. The Valuation of Data as an Asset: A Consumption-Based Approach [EB/OL]. [2014-04-22]. <https://www.reply.eu/Documents/13903>
- [5] LANEY D. Why and How to Measure the Value of Your Information Assets [EB/OL]. [2015-08-04]. <https://www.gartner.com/doc/3106719/measure-value-information-assets>

作者简介



吴甘沙,曾任英特尔中国研究院院长,现任驭势科技有限公司CEO;研究方向为大数据系统架构、隐私保护/多方安全计算、数据货币化。