

数据的价值层次和设计模式

李满海, 辛向阳

(澳门科技大学, 澳门 999078)

摘要: **目的** 对数据产生价值的过程进行分析, 探讨将数据当作设计对象的研究范式。**方法** 以数据驱动型企业的案例为基础, 通过归纳总结进行分类研究, 对采集、处理和应用 3 个数据加工环节如何产生价值进行案例说明。**结果** 得出数据用于发现和解决问题的 CIC 通用模式, 以及数据用于重新定义问题的 CIC 双环模式, 提出不同数据加工方式和用途会产生不一样的数据价值; 并且提出数据价值根据 CIC 复杂度可以分为 4 个层次, 数据价值层次越高, 数据产生的价值越大。**结论** 随着互联网和大数据技术的发展, 用户生成内容让数据有新陈代谢的能力, 用户参与让数据的生产和消费同时进行, 数据越用越有价值; 在大数据时代, 数据不仅仅用于发现和解决问题, 其更大价值还体现在重新定义问题上。

关键词: 数据; 大数据; 设计; 价值; 用户生成内容

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2019)02-0134-04

DOI: 10.19554/j.cnki.1001-3563.2019.02.023

The Value Hierarchy and Design Mode of Data

LI Man-hai, XIN Xiang-yang

(Macau University of Science and Technology, Macau 999078, China)

ABSTRACT: The work aims to analyze the process of generating the value of data, and explore the research paradigm with data as the design object. Based on the cases of data-driven enterprises, a classification study was conducted through induction and summary, and how the three data processing links like collecting, interpreting and contextualizing generated value was illustrated. The CIC common model used by data for discovering and solving problems and the CIC double-loop model used by data for redefining problems were obtained. It was proposed that different data values could be generated with different data processing methods and uses. Moreover, the data value could be divided into four levels according to CIC complexity. The higher the level of data value, the greater the value generated by data. With the development of Internet and big data technology, the UGC(User-Generated Content) enables data to have the ability of metabolism. Users' participation facilitates the production and consumption of data at the same time. The more frequently used the data, the more valuable they are. In the era of big data, data are not only used to discover and solve problems, but also have greater value in redefining problems.

KEY WORDS: data; big data; design; value; UGC(User-Generated Content)

在计算机科学领域, 数据结构设计和数据库设计都有几十年的设计历史了, 各项设计方法和理论比较成熟, 产品也很丰富; 而在设计学领域, 将数据看作可以被设计的对象是近十年来发生的事情, 还处于探索阶段。在大数据的时代背景下, 随着信息可视化的研究和实践, 学术界对数据的理解和方法的运用已经初有成效, 比如清华大学美术学院向帆对数据的艺术

应用^[1]。在行业领域, 当阿里的芝麻信用、联通的智慧足迹以及新零售等数据驱动的新兴商业生态逐步成熟的时候, 设计师所服务的产业环境、工作中所接触的产品属性、设计流程都和传统产品设计大相径庭。在反思设计对象本体属性、改变设计流程、重新确定判断标准之前, 有个事情需要探讨, 那就是数据的价值层次以及数据的通用设计模式。

收稿日期: 2018-09-15

作者简介: 李满海 (1979—), 男, 重庆人, 澳门科技大学博士生, 主要研究方向为设计学。

1 设计的对象：数据

数据是事物客观性和人类主观性的纽带，也是人类认识世界的桥梁。和语言文字一样，数据是人工产物，始于原始人在采集、狩猎等生产活动，是用于记录事物的性质和交流的工具。设计是人类构思、规划和生产产品，进而达成个体或集体某种目的的过程^[2]。传统上对数据的设计，主要强调数据的采集、存储、处理和应用等各环节的管理，比如关注调研样本数量、覆盖面、精准度等，目的是更好地去发现和解决问题。随着物联网、人工智能等数据技术的发展，以及直播视频、微信语音等内容的日趋多样化，数据量爆炸式增长，数据价值呈现指数化增长。下面从数据的用途开始切入，围绕价值纬度去探讨数据的设计模式。

1.1 数据用于发现和解决问题

人们在探索问题真相的过程中，面对未知问题会提出假设，数据可以被用于验证假设，直到问题被发现；而且通过逻辑推理，数据也可以解决已知的具体

问题。数据用于发现和解决问题有 3 个基本的加工环节：采集（Collecting）、处理（Interpreting）和应用（Contextualizing），简称 CIC 通用模式，见图 1。

以莫里的航海图^[3]为例。当时船长们认为海洋是一个不可预知的世界，船只根据传统航海经验宁愿在水上绕弯而不走直线，虽然船长们知道这并不完全正确，但是也只好如此。为了解决这个问题，莫里率先梳理了记录特定日期和地点的风、水与天气的所有航海日志，按经纬度划分出了海洋位置，按月份标出温度、风速和风向，同时和海上船只建立了互动协作机制，每艘舰船收集和反馈线路上的航海数据，并定期向海里扔掷和捞取标有日期、位置、风向以及当时洋流情况的瓶子以获取更多的数据，莫里将修正后的最新航海图作为回报提供给船只。通过绘制和分析 120 万个数据点，出版海洋自然地理与气象专著和航海导航图，揭晓了风向和洋流都非常利于航行的一些天然航线。这些数据经过采集、处理和应用产生了价值，解决了航海的航线问题，不仅航海路程减少了 1/3 左右，而且增加了安全感。

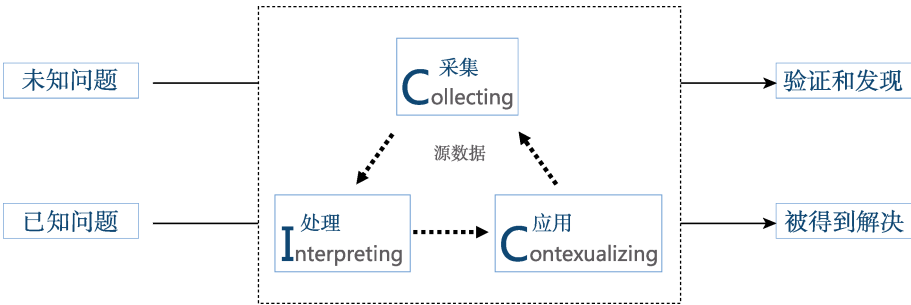


图 1 数据用于发现和解决问题的 CIC 通用模式
Fig.1 The CIC common mode used by data for discovering and solving problems

数据用于发现和解决问题，有的只需要 1 个环节，有的只需要其中 2 个环节，有的则需要 3 个环节。不同数据加工方式会产生不一样的数据价值，这里分为 3 个价值层次。

1) 点层次价值，是指数据经过 CIC 通用模式中的 1 个环节所产生的价值。例如，只有中国联通可以采集到所有联通手机用户的通话记录和位置数据；例如，通过二值掩码等多种图像合成技术，可以将一张模糊不清的心脏扫描原图处理成肌肉纤维清晰可见的图片，帮忙医生做准确的诊断^[4]；例如，Amazon 网上书店根据消费者在网站上的浏览和购买产生的行为数据，发现消费者的喜好，进而推荐与此喜好相关的新图书以促进图书销售。

2) 线层次价值，是指数据经过 CIC 通用模式中的任意 2 个环节所产生的价值。例如，阿里巴巴将每日运营数据的采集和处理做成阿里指数，即根据阿里巴巴网站每日运营的基本数据，包括每天网站浏览量、每天浏览的人次、每天新增供求产品数、新增公

司数和产品数这 5 项指标统计计算，得出行业大盘和产业基地的变化趋势。随着不断修正数据处理指标和优化采集数据，阿里指数的合理性和准确性也不断增强。阿里指数为电子商务的发展提供了很好的参考价值。例如，Alpha Go 采用搜索算法、机器学习和神经网络等计算机技术，对三千万盘人类顶级棋手对弈数据进行分析，通过价值网络来进行整体局面判断，不断调优决策网络，得到有胜算的最佳落子的位置。

3) 面层次价值，是指数据经过 CIC 通用模式的完整 3 个环节所产生的价值。与点层次和线层次的最大区别，是数据在整个过程中是动态连贯的，即采集环节的数据是实时变化的，处理环节要及时处理，应用环节要给用户呈现最新数据。例如，百度迁徙地图通过基于地理位置信息的开放接口，采集手机用户的每天 35 亿次的实时定位数据，经过数据隐私的脱敏处理，在迁徙地图平台上采用图形化映射出手机群体用户位置变化的轨迹，以区域和时间两个维度，将中国春节前后人口大迁徙的轨迹与特征，以动态、即

时、直观的可视化方式呈现出来。不仅解决交通领域实时监控人流的问题,而且基于往年的历史数据,与交通运输部联合发布春运预测报告,给铁路、公路等运输部门提供车辆合理调度的参考依据。

1.2 数据用于重新定义问题

设计师面临着两种问题:确定性和不确定性。在自然科学领域,科学家和工程师们基本是在确定的环境中,使用确定的要素,针对确定的需求去找到合理的解决方案^[5]。在人文社科领域,由于世事反复无常而非线性,各种因素层层叠叠相互影响,遇到的问题无法被确切描述,这类诡异(wicked)的问题没有明确的限制条件,初始的时候无法像自然科学领域那样找到最优的解决方案,只有通过解决或部分解决才能被明确定义^[6]。

数据用于重新定义问题,最大的特点是有用户的参与。用户在参与的过程中继续生成新的数据,这些UGC(User-Generated Content,用户生成内容)数据^[7]经过采集、处理和应用,会发现未曾想到的新问题并

找到解决办法。用户参与的每个探索过程都至关重要,用户和数据共同去重新定义不确定的问题。例如,阿里巴巴的发展过程,是重新定义电子商务的过程,在阿里巴巴成功之前,没有可参照的样本。从淘宝到天猫,从支付宝到蚂蚁金服,从双十一到新零售,整个过程都是商家、消费者、供应商等参与者帮助马云一起去定义什么是电子商务。中间走了一些弯路,比如出现大量假货的信任危机,但是通过有效的数据监控找到源头,问题得到快速解决。原阿里巴巴负责数据分析的副总裁车品觉提到,大数据的力量来自触类旁通的关联,以前数据用于做分析判断,现在更多地用在了商业创新上^[8]。

将“源数据的CIC加工环”和“UGC数据的CIC加工环”打通融合一起,双环之间通过采集环节的“有意识地再获取数据”来衔接,即在初始源数据的基础上,主动创造条件和用户产生互动,使得用户在使用数据的过程中再产生更多的数据,形成一个数据源源不断的动态闭环,原先的问题被重新定义。该过程,简称CIC双环模式,见图2。

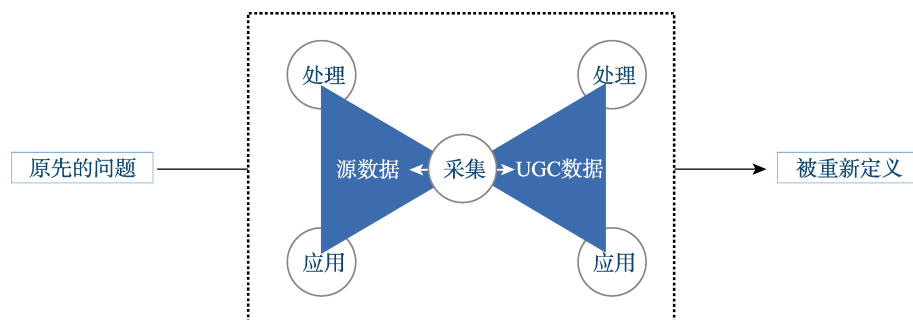


图2 数据用于重新定义问题的CIC双环模式

Fig.2 The CIC double-loop mode used by data for re-defining problems

数据通过CIC双环模式产生的价值,称为体层价值。体层次与面层次最大的区别点是:面层次只对源数据进行CIC加工,而体层次在面层次的基础上提供与用户互动的机会,给用户创造再次产生数据的可能性,比如数据应用平台化。用户不同程度地参与数据的生产过程,数据的生产和消费是同时进行的,用户既是数据的使用者也是数据的生产者。体层次的价值由所包含的诸多点线面层次的价值共同构成。

数据是非物质形态,虽然数据价值的存在和流通需要附加在物质载体上,但是独立于物质载体,不具备比如外观、尺寸、材质等物质特征^[9]。数据在被使用的过程中,内容不会损耗减少,不像实物产品在消费中以自身的消耗和磨损作为代价。数据这两个特征有利于人们使用数据去定义人文社科领域中的不确定问题,这是因为数据为设计师提供了试错机会,允许边实践边验证,不断调优,每个尝试都可以采用不同方法,每个探索都在部分解决问题,从简单问题到

复杂问题,最后从因果分析转变到关联分析^[10],最终寻找到更好的解决办法。

例如,今日头条是一个没有小编的新闻引擎系统,用户输入自己的兴趣爱好,系统自动输出匹配的新闻信息。刚开始系统不知道用户的新闻喜好,若采用传统方法,需要了解该用户的工作重点及教育经历等信息,设计师可以通过现场观察、访谈调研等方式,再给一些新闻内容进行测试验证。然而,今日头条直接通过数据分析用户的新闻偏好。用户在注册的时候输入自己的兴趣和爱好,假如选择了喜欢看娱乐新闻,那么今日头条会先推送娱乐新闻给该用户,同时后台采集用户的使用习惯数据,经过一段时间的行为数据观察,比如发现用户实际上在体育频道新闻上浏览更多的时间,这时,今日头条会调整优化,更多推送体育新闻给该用户。从用户的视角来说,用户会觉得今日头条很人性,知道自己喜欢什么;从今日头条的视角来说,推送新闻的精准度更高了,每篇的阅读量也随着提高。

2 数据的价值层次与 CIC 复杂度

上述提到, 数据产生的价值从低到高可以分为4个层次: 点层次、线层次、面层次和体层次。数据的价值层次越高, 数据产生的价值越大。

数据的价值层次跟 CIC 复杂度有关系, 数据被人加工处理的复杂度越高, 数据产生的价值越大。点层次、线层次和面层次的数据价值侧重发现和解决问题, 体层次的数据价值侧重重新定义问题, 见图3。

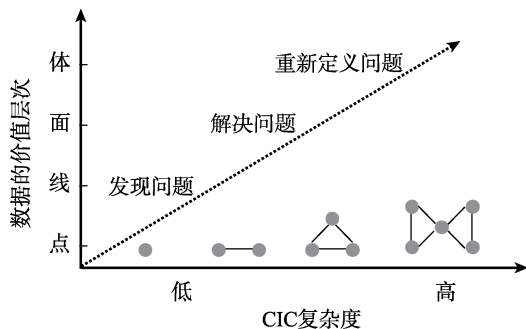


图3 数据的价值层次与 CIC 复杂度
Fig.3 The value hierarchy of data and CIC complexity

CIC 的每个环节都能产生价值, 通过单个环节或两个环节的数据加工, 可以发现规律趋势, 但达不到将分析结果应用在运营决策或机会预测的数据洞察要求, 只有打通采集、处理和应用3个环节, 数据才能面向某个特定行业提出一套能解决问题的方案。

以中国银联为例, 当全国所有银联卡的详细消费数据被实时采集, 那么就实现了数据的点层次价值, 因为采集环节所存储的大量源数据本身就已产生价值; 当成立银联智惠公司对实时采集到的银联卡消费数据进行深度清洗和挖掘处理, 萃取出全国各大主要城市的核心商圈和品牌, 那么就实现了数据的线层次价值, 因为经过采集和处理两个环节后的数据, 可以以更高的价格卖出; 当银联智惠公司不仅实时采集和清洗处理, 而且还为零售、金融、房地产、旅游、互联网等行业提供大数据应用产品及解决方案, 那么就实现了数据的面层次价值, 因为经过采集、处理和应用3个环节后, 数据产生了更大的价值。

数据越用越有价值, 特别是用户在使用数据的过程中新产生的数据, 这些行为数据会产生新的价值。例如, Google 会记录下用户使用搜索的过程, 包括搜索了哪个关键词, 点击了哪条搜索结果。Google 对流感趋势的分析^[11]正是基于这些用户的搜索行为数据。正如阿里巴巴首席技术官王坚提到: Google 数据分析的对象不是搜索结果数据本身, 而是人们怎么使用搜索, 其本质上是想要理解人的在线行为^[12]。

3 结语

近年来随着大数据概念的兴起, 许多企业在数据

相关方面投入很大, 但是麦肯锡等咨询机构调查发现, 大部分投资回报率很低。采用本文论述的数据价值层次去解释该现象, 就会发现这些企业对数据价值的挖掘, 大部分停留在了点线面层次。有的企业搭建了 IT 系统, 有能力采集到企业经营过程中的各项数据, 但是没有想清楚如何处理和应用; 有的企业简单处理了这些数据, 比如可视化呈现成数据报表, 停留在了验证决策假设和监控运营效果层面。

本文基于数据价值层次提出的 CIC 通用模式和 CIC 双环模式, 从商业角度来说, 可以指引这些企业更充分地挖掘出数据价值; 从学术角度来说, 这两个设计模式将数据作为设计对象, 探讨了其研究范式, 特别是数据用于解决人文社科领域中的诡异问题提出了思路。

参考文献:

- [1] 向帆. 视觉文献的视觉化设计——全国美展获奖油画作品视觉化工具 Award Puzzle 设计探索[J]. 装饰, 2016(7): 92—94.
XIANG Fan. Award Puzzle, a Visualization Platform for Chinese National Art Exhibition Awarded Oil Painting Works[J]. Zhuangshi, 2016(7): 92—94.
- [2] SABINE J, JURGEN F. Designing Business and Management[M]. An Imprint of Bloomsbury Publishing Plc, 2016.
- [3] CHARLES L L. Matthew Fontaine Maury: the Pathfinder of the Seas[M]. Literary Licensing LLC, 2013.
- [4] CHEN Wei, YAN Zhi-cheng, ZHANG Song, et al. Volume Illustration of Muscle from Diffusion Tensor Images[J]. IEEE Transactions on Visualization and Computer Graphics, 2009, 15(6): 1425—1432.
- [5] BUCHANAN R. Wicked Problems in Design Thinking[J]. Design Issues, 1992, 8(2): 3—20.
- [6] MARC S. Co-Design as a Process of Joint Inquiry and Imagination[J]. Design Issues, 2013, 29(2): 16—28.
- [7] 赵宇翔, 范哲, 朱庆华. 用户生成内容(UGC)概念解析及研究进展[J]. 中国图书馆学报, 2012(5): 68—81.
ZHAO Yu-xiang, FAN Zhe, ZHU Qing-hua. Conceptualization and Research Progress on User-Generated Content[J]. Journal of Library Science in China, 2012(5): 68—81.
- [8] 车品觉. 决战大数据: 驾驭未来商业的利器[M]. 杭州: 浙江人民出版社, 2018.
CHE Pin-jue. Big Data: the Business Revolution[M]. Hangzhou: Zhejiang People's Publishing House, 2018.
- [9] DATTA P P, CHRISTOPHER M. Information Sharing and Coordination Mechanisms for Managing Uncertainty in Supply Chains: a Simulation Study[J]. International Journal of Production Research, 2011, 49(3): 765—803.
- [10] MAYER S V, CUKIER K. Big Data: a Revolution That Will Transform How We Live, Work and Think[M]. John Murray Publishers Ltd, 2013.
- [11] GOOGLE. Google Flu Trends[EB/OL]. [2018-08-08]. <http://www.google.org/flutrends>.
- [12] 王坚. 在线[M]. 北京: 中信出版集团, 2016.
WANG Jian. Being Online[M]. Beijing: China CITIC Press, 2016.