

·科技纵横捭阖·

## 从卷宗社会到数据社会

2012年3月22日,美国总统奥巴马政府宣布投资2亿美元拉动大数据相关产业发展,将“大数据战略”上升为国家战略,甚至还将大数据定义为“未来的新石油”。自此,大数据风起云涌,并得到广泛使用。那么,究竟什么是大数据?根据笔者理解,如果说互联网是关于“物”的,那么大数据就是关于“人”的<sup>[1]</sup>。可是,关于人的数据早就有了,怎么现在才提出大数据的概念?

大数据并非数据大。在统计学中,样本选取得越多,得到的统计结果就越接近真实结果。在工业社会中,统计学在社会学中的重要应用就是社会的治理。关于某个人的统计数据就是其“卷宗”,顾名思义,卷宗就是档案。因此,工业社会实际上是“卷宗社会”。在工业社会中一个人若想成为“正常”的“社会人”,就必须落入统计学得出的高斯曲线之内,而在高斯曲线二端的就是“异常”的人群<sup>[2]</sup>,也就是那些困厄潦倒或者大富豪之类的人群。如果说工业社会对社会的治理是相对于人群的,那么信息社会就是在工业社会治理的基础上,体现出更为个性化的治理,即某个人可供采集的数据样本就更大了。所以也可以将信息社会称之为“数据社会”。

互联网的出现,尤其是各种社交媒体,使得统计样本的采集更具个性,更为精致,也更加连续;充斥在网络中的海量数据便构成了所谓的“大数据”。如果能将它们“提纯”并迅速处理成有用信息,无异于掌握了一把能打开另一个世界的钥匙。譬如,相关大数据已经应用到交通、医疗、气象、电商等方面。这些精致化的定制服务是卷宗社会所不具备的,因此大数据延展了对社会的治理。由于直接与人相关,大数据成了“金矿”,有待人们进一步挖掘,并从中寻求各种机会。

美国物理学家惠勒曾提出“万物源于比特”这样的论题。假如这个论题成立,那么整个社会也变成了计算的。早在2009年,便有15位学者在《Science》杂志撰文,首次提出计算社会科学的概念<sup>[3]</sup>。2014年又出版了《计算社会科学导论:原理与应用》一书,旨在从信息处理角度,



利用先进的计算机和信息技术对社会学进行跨学科研究并首次对计算社会科学进行了完整而统一的介绍,其中包括计算社会科学的概念,提出了相关的定义和专业术语,描述了计算社会科学各个领域的研究范围和研究内容,包括信息挖掘、社会化网络、社会复杂性理论和社会仿真等。此外,还详细介绍了社会信息挖掘、社会网络分析、社会复杂性理论和社会仿真模型4种研究方法,解释了如何利用计算方法对不同层面的社会复杂性进行区分和分析。从社会计算科学角度看,大数据将人的行为的结构化数据与其背景和人口统计学的信息相结合,产生出很多成果并应用,如有的放矢的广告和营销等。人们可以把社会媒体大数据中提取出的自然语言文本的情感挖掘视为一种数据挖掘的自然延伸。由于大数据的无限开放性,其未来的应用潜力更大。因此,从信息哲学的角度看,大数据在认识论上也适应新的社会发展。

人的行为维度具有无限可能性,但人的资源却又是有限的。那么如何协调无限和有限的关系?传统的数据由于属性有限,个体参与度较低,其价值预期比例大,即数据处于压缩状态,而无法协调无限和有限的关系。大数据体现的结果就是将传统的数据“解压缩”,使其数据密度大大减小,从而放大个体数据的效应。由于有了海量的数据和强大的计算处理能力,有了人与人、人与物的互联互通,就可以成就人的行为的无限可能性。电子商务在这一点上体现得淋漓尽致。大数据是关于人的,它不单是一个技术问题,也是一个管理问题。认识到这一点就要破除传统的管理办法,将数

据打通,使其不断更新,避免产生“数据孤岛”现象。那么,首先就要给出“全量数据”,也就是说,关键的数据不能缺失;其次,那些关键信息不能靠专家规定。

中国是人口大国,社交媒体产生的数据以指数形式递增。我们需要未雨绸缪,需要思考如何处理这些庞大的数据。大数据都要被计算机处理,因此关于人的数据一定要有原始大数据的“元数据”——它们是为机器服务的,必须通过元数据的语义标示并赋予其意义,才能被机器处理。因此,若想从数据中发现知识,就必需有大量的元数据。元数据就好比影视剧中的“桥段”,将机器中的原始数据与人的行为连接起来,大数据挖掘技术恰恰就是针对元数据的。尽管如此,大数据还是有其无法企及的地方。大致可以归纳为以下几点:不能没有有效的商业模式,不能替代管理的决策力,不能保证消除“噪音”,不能进行无目的的知识发现,不能一次建模终身受益,不能替代领域专家,不能忽略数据标注,等等。

大数据并非一个终极阶段,它的出现不过是人类历史进程中的一个环节,其重要意义在于它在计算机技术的基础上为整个人类带来变革中的一步。回顾历史,计算机从20世纪50年代起就在人类历史上开始了潜移默化的革命,其根本标志就是“数字化”,以及物理世界和虚拟世界的无缝连接。未来,我们依然任重道远。

### 参考文献

- [1] 刘钢. 大数据: 创造崭新的生存环境[N]. 学习时报, 2014-12-08(7).
- [2] Ian Hacking. The Taming of Chance [M]. Cambridge: Cambridge University Press, 1990.
- [3] David Lazer, Alex Pentland, Lada Adamic, et al. Computational social science [J]. Science, 2009, 323(5915): 721-723.

文/刘钢

作者简介 中国社会科学院哲学研究所, 研究员。图片为本文作者。

(编辑 王丽娜)