

大数据时代的认识论问题探究

刘 畅

北京交通大学马克思主义学院, 北京

收稿日期: 2022年3月18日; 录用日期: 2022年4月17日; 发布日期: 2022年4月24日

摘 要

大数据本质上是利用数据进行信息处理的新兴技术, 当人们利用这种新兴技术在认识世界时, 暴露出认识中的精确性与混乱性、因果性与相关性等问题。由问题所带来的风险对个人信息、和谐社会环境造成威胁, 从大数据运行规则的调整以及法律救济途径的完善对风险予以规制是善用大数据的应有之义。

关键词

大数据, 认识论, 风险分析, 风险规制

Exploration of Epistemology in the Era of Big Data

Chang Liu

Marxism School of Beijing Jiaotong University, Beijing

Received: Mar. 18th, 2022; accepted: Apr. 17th, 2022; published: Apr. 24th, 2022

Abstract

Big data is essentially an emerging technology that uses data to process information. When people use this emerging technology to understand the world, they expose problems such as accuracy and confusion, causality and correlation in understanding. The risks brought by the problems pose a threat to personal information and the environment of a harmonious society. It is necessary to regulate the risks from the adjustment of the operation rules of big data and the improvement of legal remedies.

Keywords

Big Data, Epistemology, Risk Analysis, Risk Regulation

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,有关大数据的话题正在世界范围内广泛传播。随着互联网技术的深入发展以及智能手机的推广应用,人们在赛博空间里的活动、行为都有迹可循,被记录成数据。在虚拟的网络世界中数据日益海量增长,传统的计算处理模式无法高效捕捉、管理和处理海量数据,而以云计算为代表的大数据处理模式在处理信息时决策能力、流程优化能力、洞察能力都大幅提升。人类俨然已经进入大数据时代。在这样的时代背景下,“反馈经济”、“分享经济”等新经济、新商业模式逐渐形成,大数据应用正在创造巨大的商业价值和经济效益,成为新世纪的矿产和石油。与此同时,大数据时代的到来,也给人们的生产方式和生活方式带来变革,冲击着人们传统的思维模式和认识世界的方式。哲学作为时代精神的精华,能够切准大数据时代脉搏、凝聚实践智慧。从哲学视角去探讨大数据时代的认识论问题是我们获得有益启示的途径。

2. 大数据时代的认识论基础

哲学上的认识论问题是指探讨人类认识的本质,认识发展的规律和认识的结果等问题。认识论问题长久以来一直是哲学领域研究的核心问题。以西方近代哲学上的经验论和唯理论之争为例,二者争论的焦点在于他们对认识的起源和认识的本质问题上存在根本分歧。唯理论的代表人物笛卡尔从“普遍怀疑”出发,运用理性去怀疑除“我在怀疑”以外的一切命题,提出了“天赋观念”的主张。以“天赋观念”为出发点,经过逻辑推理便能够获得具有普遍必然性的知识。唯理论强调认识的普遍必然性,并把理性当作认识的起源。经验论的拥护者洛克则反驳了笛卡尔的“天赋观念”,人类不可能察觉不到任何天赋于心灵中的观念。心灵的最初状态只是一块白板,人类所有的知识都建立在经验基础上。经验论则把经验当作认识的起源。后世众多哲学家对唯理论和经验论之争都发表了他们独到的见解,试图终结二者之间的争论。康德通过“先验自我”来克服二者之间的对立,他探讨的是认识何以可能的问题,在主客体关系上,把认识论归结为独立于客观的人的主观行为。黑格尔则将本体论、认识论与逻辑学统一起来,认识到人的认识是一个发展的过程,同时在各个方面呈现着客观世界的辩证法。而在马克思主义哲学史上,马克思将实践引入认识论,把辩证法应用于反映论,创立了能动的反映论,科学地解释了认识的本质是主体在实践基础上或通过实践对客体的能动的、创造性的反映。列宁在集中批判经验主义的基础上,阐述了马克思主义的唯物主义反映论的认识论原理,为马克思主义“哲学基本问题”奠定了认识论基础。实践证明,马克思主义哲学克服了唯理论和经验论在认识论上的局限性,指出在人类的认识中,一定是物质第一性,同时人类的思维是联系发展,用内在矛盾推动人类认识发展的。是我们认识世界改造世界的科学的指导理论。

进入大数据时代,弗洛里迪将大数据表述为“大、种类多样、复杂、纵向及(或)分散式的数据集合,由工具、传感器、网络互动、电子邮件、视频、点击数据流以及其它现在乃至将来均可使用的数字来源构成。”^[1]美国 Gartner 公司将大数据表述为:“大数据是指数量巨大、速度快捷、种类繁多的信息财富,这些数据需要新的技术手段来处理,以便提高决策制定、领悟发现以及过程优化等能力。”^[2]IBM 则对大数据的特点进行了概括,得出了 5V 的结论:Volume (大量)、Velocity (高速)、Variety (多样)、Value (低价值密度)、Veracity (真实性)。大数据作为一种新兴技术,借助于云计算对大量多样化的数据进行分

布式高效处理与存储，人们通过大数据技术获得了前所未有的新的认知。当前的大数据技术不仅是新一轮科技革命进程中的一个重要环节，预示着未来时代转型的方向。而且也深刻地影响着人们传统的认识论视角。基于以上分析，大数据时代的认识论基础为数据。数据作为一种能够反映客观世界的物质存在，在认识论上表现为能动的反映论。而大数据并不采用随机分析即抽样调查的方式，而是将全体数据进行分析处理，获得相关结论。这种对结论的认识也是源自经验。数据因此兼具经验论和唯理论认识论的特点。

3. 大数据时代的认识论问题及风险分析

在大数据时代，技术条件已经出现了质的突破，认识能力也极大提升。人们在获取尽可能多的数据的基础上，能够深入到认识客体的微观层面，绘制出一幅更加立体、更加具有时效性的认识模型。然而，与以往不同的是，大数据在对信息的认识过程中存在偏爱混乱性而排斥精确性，推崇相关性而逐渐摒弃因果性的情形，这将引发一定的社会风险。

3.1. 认识中的精确性与混乱性问题

对精确性的追求是以往信息匮乏时代、生产力落后条件下的产物。在大数据时代，“精益求精”往往不如“拥抱混乱”更能高效认识未知世界。认识运动的规律表明，主体在认识客体时首先要将感性认识上升到理性认识，要观察现象，尽可能多地占有丰富的感性材料。在观察现象的基础上，对大量的感性材料进行“去粗取精、去伪存真、由此及彼、由表及里”的分析。即运用抽象思维的方法，从混沌的表象中提炼出本质的规定[3]。长久以来，我们在认识事物时倾向于将其从混乱的现象中提炼出来，追求精确性，追求唯一的真理。在这个过程中，混乱是贬义的，是应该竭力避免、逐层剥离出去的。而在大数据时代，正在形成一种新的认识。在效率面前，追求唯一的真理是对注意力的分散。要想获得大规模数据带来的利益，混乱应该是一种标准化的途径，不应该排斥混乱。“真理”成为一种拓扑学，它不属于一个点，而是属于某种区域或类型。人们对真理的追求不再是“有的放矢”而是“彗星式”放矢[4]。这也许与我们的直觉相矛盾，经过实践证实，接受数据的不够精确与不够完美，从越混乱庞杂的数据中，我们反而能够更好地预测未来，认识世界。以测量葡萄园的温度为例，现有两种方案。第一种方案是只使用一个刻度精确的温度测量仪进行测量工作。第二种方案是使用更多数量的温度测量仪，至少每 100 棵葡萄树拥有一支。但对测量仪的精确性不做过高要求，允许出现混乱。相比之下，方案二最后提供的测量数据反而比方案一更加准确。多样的混乱优于单一的精确定，由于数据样本多，即使出现混乱、错误的的数据也不会对整体造成太大偏差，其创造的价值大于其造成的不利影响。

3.2. 认识中的因果性与相关性问题

在自然科学中有一条重要的认识论原则就是因果关系，人们通过对因果关系的探究将事物理解为合规律性的。在社会科学中，原因与结果是揭示事物之间引起与被引起关系的一对范畴。因果关系就是包括时间先后次序在内的一种现象必然引起另一种现象的普遍联系。因果关系属于唯物辩证的思维方式，用唯物辩证的思维方式去研究人类认识发展的规律是辩证唯物主义认识论的要求。不论是在自然科学领域、还是在社会科学领域，对因果关系的研究素来是人类认识事物之间相互关系的重要环节。在大数据时代，人们不再重点寻找事物之间的因果关系，而是更加强调对事物相关关系的分析。相关关系的核心是量化两个数据值之间的数理关系。相关关系强是指当一个数据值增加时，另一个数据值很可能也会递增。例如，谷歌通过大数据成功预测了 H1N1 流感在美国范围的传播，这种预测手段高效且及时。相比之下，疾控中心则在流感爆发数日后才后知后觉，具有滞后性。谷歌采用的就是把握数据之间的相关性

方法，在一个特定的地理位置，通过谷歌检索特定词条的人越多，则能够推测出该地在相应时期内有更多流感患者。很显然，相关关系绕开了事物背后复杂的因果链条，直截了当地找出事物之间的相关性。这种“只需知其然、不需知其所以然”的数据分析模式，能够减少过去因人为直觉给事物之间强加因果所带来的主观偏见，将事物之间的关系更加客观、清晰地呈现出来，可以帮助我们更加高效地捕捉现在、预测未来。

3.3. 大数据时代的认识论风险分析

认识活动的发展过程是从感性认识到理性认识，再从理性认识到实践，实践、认识、再实践、再认识，循环往复以至无穷。主观和客观的矛盾贯穿认识运动的全过程。因此，人们若想获得关于事物正确的认识往往要经历多次反复。而大数据时代，人们利用大数据去获得对事物的认识时，需要去适应大数据的运行机制，即混乱性优于精确性、相关性优于因果性。事实上，利用大数据去认识世界时背后是效率至上原则占主导地位的。在不考虑因果性只追求相关性的前提下去“高效”认识世界时，从方法论角度看是有益的。而从认识论角度看是值得反思的，不仅大数据的理性问题尚待讨论，而且人们在“效率优先”原则主导下，往往不会采取多次反复的实践检验，认识活动发展过程的完整性也因此被割裂。关于技术理性的研究，马克斯·韦伯分成工具理性和价值理性对待。马尔库塞确立了技术理性的概念，他将技术理性视为工具理性。工具理性着眼于效率、计算、功能效用及其手段，而忽视目的的合理性。价值理性则重视行为本身的价值，不注重行为的手段与后果。大数据在进行数据分析时采取相关性而非因果性的标准，过于追求效率，从工具理性视角看它很难符合合理性。而当大数据广泛运用于人们的实际生活中时，人们不由自主地被卷入数据共享的时代。人们在互联网上的一举一动都变得有迹可循，被数据记录，也被数据监视。大数据威胁着个人隐私、机密，乃至身份。从价值理性视角看，大数据存在着伦理问题。

4. 大数据时代的认识论风险规制

从认识论的视角来看，大数据在追求相关性、拥抱混乱性的道路上不可避免地产生技术异化现象，对人们的现实生活造成数据监视、信息泄露等社会风险。对大数据引发的社会风险急需进行风险规制。

4.1. 坚持相关性和因果性相结合

在商业化迅速发展的时代，经济至上、效益至上的理念催生了大数据善用相关性而非因果性的运行规则去发掘隐藏在市场中的商业价值。然而，从大数据技术工具理性的视角看，过于高效未必符合合理性。实际上，相关性与因果性并非是相互对立的两种运行规则。虽然事物之间存在着相关性并不必然等于二者之间就存在着逻辑上的因果性，但某些相关性的背后的确潜藏着某种因果关系。对因果关系的思考尽管不是以营利为目的的商业主体所追求的，然而对于科研工作者而言研究事物之间的因果关系才是其核心目标所在。相关性研究呈现出“知其然”的预测，因果性研究揭示“知其所以然”的道理。运用大数据去分析事物之间的关系，只有将因果性与相关性相结合，通过所呈现出的相关性去把握事物背后的因果性，才能做到“知行合一”，从而更加有效地进行认识活动，从事实践活动，实现认识与实践的统一。

4.2. 依法规制精确性中的混乱性

人们利用大数据在对个人信息的处理上，在认识世界改造世界的过程中，威胁着个人的隐私、机密、身份，不仅涉及到对个人财产权的侵犯可能，还涉及到对个人人格权侵犯的可能。从价值理性的视角看，存在着伦理问题。从现实的视角看，存在着法律风险。通过对大数据技术运用的认识论层面的反思，建

构相关法律对风险予以规制，已成为适应大数据时代风险社会的要求。一方面，从民法救济角度，通过立法明确数据采集者在进行数据采集时及时向被采集者履行告知义务，对即将采集的信息予以披露。在实际生活中，拥有信息采集手段的信息采集者善于隐瞒其信息采集行为，通常不会主动披露信息采集的细节。即便是在网站中针对用户在使用产品前提供的附有“已同意”选项的知情条款，往往成了流于形式的对对方的免责条款。在这种情况下，从法律的公平原则出发，应制定举证责任倒置的法律规定。当用户个人信息被泄露、隐私权遭到侵犯时，用户无需主动收集受到侵犯的证据。侵权者负有提供证据证明自己无过错的责任，如果没有充分证据证明自己无过错，则需承担法律责任。另一方面，从刑法救济角度，明确侵权主体的具体概念和类别，使得用户在受到权利侵犯时不因找不到侵权主体而维权失败。侵权主体明确后涉及到侵权行为的责任认定问题。刑法应加大对国家机关、事业单位等公共部门的问责权限，坚决打击公权力肆意侵犯私权利的行为。在定罪量刑标准上，对危害个人信息类的犯罪，可将个人所遭受的损失、人身受到损害的程度以及侵权主体的违反所得纳入定罪量刑的司法考量范围中。

5. 结语

对大数据技术的全方位运用浪潮正在席卷全球，围绕大数据时代的认识论问题进行理解、探究，是我们善用大数据的基础。大数据时代，新的认识方式正在颠覆人们传统的认识论理念。以开放包容的心态接纳新事物、新技术、新理念的同时，在实践中也要用批判的视角去一分为二地看问题。对于其背后的风险应及时采取技术完善、法律完善的途径予以规制，趋利避害，使大数据更好地服务于人类社会。

基金项目

北京交通大学基本科研业务费研究生创新项目“基于利益相关者视角的数字信任问题研究”(2021YJS184)。

参考文献

- [1] 方环非. 大数据: 历史、范式与认识论伦理[J]. 浙江社会科学, 2015(9): 114.
- [2] 黄欣荣. 大数据对科学认识论的发展[J]. 自然辩证法研究, 2014, 30(9): 84.
- [3] 陈先达, 杨耕. 马克思主义哲学原理[M]. 第5版. 北京: 中国人民大学出版社, 2019: 151.
- [4] 余乃忠. 大数据时代的认识论重塑[J]. 江海学刊, 2019(5): 133.