

“大数据—小数据”问题：以小见大的洞察

研究成果：“大数据—小数据”问题：以小见大的洞察

作者：陈国青，张瑾，王聪，卫强，郭迅华

发表期刊：《管理世界》，2021年第2期，2203-213+14

一、本文研究背景与意义

随着大数据、人工智能、移动互联等新兴科技与社会经济生活的深度融合，数字经济正逐渐成为一种重要的经济形态。以科技创新为核心动力的数字经济的发展，在数字空间重构了经济社会活动和管理决策场景，同时也催生了大量的数据，例如，线上购物中的消费者偏好、社交网络上的富媒体内容动态、共创环境下的价值创造活动、平台生态圈内的参与者行为、虚拟化生产中的数字组装日志、智能交通中的时空轨迹、用户直连制造（C2M）中的需求订单、交易市场支付结算的数字货币等等。与土地、劳动力、资本、技术等传统要素一样，数据也成为了一种重要的生产要素，进而使得大数据赋能成为了产业创新和管理决策的基础性驱动机制。

在此背景下，大数据决策范式作为新型决策范式应运而生。大数据扩大了人们以全景和细粒度的方式观察现实世界全貌的可能性，从而使人们能够在决策过程中全面了解当前事物。但是，在许多情况下，由于数据可获性、成本、时间、能力和心理因素等各种原因，人们经常仅能接触到有限且部分的数据（即小数据）。换句话说，尽管人们希望对大数据有一个全局的了解，但是可能常常不得不依靠他们可以掌握或处理的小数据来进行决策。显然，此时的决策质量在很大程度上取决于小数据的质量，这里，小数据的获取过程如同“成像”过程（例如使用成像算法和设备），成像效果直接影响着人们对大数据的认识。

在决策过程中，人们希望通过成像获得的小数据可以还原大数据的“真容”。在此，大数据与小数据之间的差异就引出了决策信息不对称性问题。为解决这种不对称性，本文从语义反映的角度提出“大数据—小数据”问题，寻求“以小见大”的洞察，帮助人们获得良好的小数据映像。毋庸置疑，在个人、组织及政府的各级决策中，解决这一问题对学界、业界和政界都具有重要意义。首先，如果无法很好地从小数据中捕获对于大数据的洞察，则不对称的大小数据可能会误导决策者，造成决策者对大数据“雾里看花”的错觉。也可能无法通过小数据对大数据的全貌形成有效的反映呈现，容易造成“盲人摸象”的残缺和局限。其次，数据作为重要的生产要素，随着其规模的指数性增长，要素效率问题受到关注。而“大数据—小数据”问题的提出和求解则可以优化数据要素的使用效率，提升数据在数字空间中的核心价值。再者，“大数据—小数据”问题在更为复杂多样的环境中可能具有其它形式和类型，探究“大数据—小数据”问题将进一步拓宽大数据驱动的决策科学的创新空间。

二、主要内容

本文提出了大数据决策中的一个重要问题，即“大数据-小数据”问题。这一问题是指在给定大数据集合的情况下，找到具有特定规模的小数据子集，以使小数据的语义尽可能接近地反映大数据的语义。也就是说，“大数据—小数据”问题是在“语义反映”的意义上，寻求获得小数据使其语义与大数据语义尽可能相近。进而，本文围绕小数据如何代表性反映、一致性反映和多样化反映大数据的语义来讨论“大数据—小数据”问题的3种类型。

具体而言，代表性语义反映是一种体现显式语义特性的“大数据—小数据”问题，这种反映通常是在原生事实的层面上，通过数据实例之间的相似性以内容覆盖的形式来表示。例如，当需要从所有搜索结果中浏览一小部分条目时，当需要从所有企业博文中读取一小部分文章时，当需要从所有客户反馈中阅看一小部分评论时，当需要从所有舆情专报中审视一小部分报告时，……，林林总总，人们遇到了依据小数据认识全局进行决策的情形。此时，小数据通过部分具体的数据实例内容来反映大数据整体的数据实例内容。

一致性语义反映是一种体现隐式语义的“大数据—小数据”问题。这种反映通常是在间接模式的层面上，通过相关属性上的数值分布来表示。例如，在线上购物环境中，消费者可以在阅看一小部分产品评论后就可以掌握全体评论中对产品不同属性的情感极性分布情况，从而避免产生有偏的购物决策。其它场景还如企业口碑的详略画像、受众声音的宏微聆听、媒体报道的点面呈现、政策分析的繁简要义等等。此时，语义反映强调小数据集合在特定属性特征上的取值模式与大数据集合的一致性。

多样性语义反映同样也是一种体现隐式语义特征的“大数据—小数据”问题，这种反映通常也是在间接模式的层面上，通过分组的信息熵以结构覆盖的形式来表示。例如，人们可以在浏览新闻过程中知晓多角度的报道，可以在竞品搜索时获取到更加丰富的选择，也可以在政策制定时考虑各类人群及其诉求，等等。此时，语义反映侧重小数据集合反映大数据集合的类别多样性，即体现大小数据集合语义在类别结构上的相近性。

在作者团队一系列研究的基础上，本文在最后针对上述3种不同类型的“大数据—小数据”问题，给出了相应的优化解决方法，并概述了这些方法的基本思路、优势以及管理决策效果。

三、主要结论与政策建议

本文针对决策信息在大小数据间的不对称性，提出了“大数据—小数据”问题，并从“语义反映”的角度定义了相关概念的科学内涵。同时，围绕代表性、一致性和多样性，阐释了“大数据—小数据”问题的3种类型，并给出了相应的领域情境、问题复杂性和可能的求解路径。

在数字经济环境中，对于“大数据—小数据”问题的重视和解决程度既关乎着数据要素的有效利用，也影响着大数据决策和赋能的效果。所以，对于学界、业界和政界来讲，相应的学术研究探讨、企业数字化转型、政策决策和治理等领域需要进一步加强大数据决策思维，鼓励“大数据-小数据”问题求解的方法创新和使能创新。

四、边际贡献与未来拓展

本文的主要贡献旨在通过定义和求解“大数据—小数据”问题揭示大数据决策中存在的信息不对称性。通过小数据反映大数据，进而实现“以小见大”的洞察，对基于数据进行科学决策具有重要意义，也是对数字经济中数据要素的更深度发掘。

在数字经济中，随着数据要素和数智化作用的日益显现，也将出现更多的“大数据—小数据”问题的应用场景和有效实践。进一步的研究拓展可在本文的基础上，继续探索在新场景下“大数据—小数据”问题的建模、求解及其赋能的不同形式，洞察和解构大数据中的深层次语义，提升大数据驱动的管理决策和价值创造水平。

供稿：科研事务办公室 编辑：高晨卉 责编：吴淑媛 赵霞