

大数据与镜像化生存：对大数据时代的哲学反思

陈 明

摘 要：大数据时代是和大信息、大计算机技术等问题联系在一起的，它们为大数据时代提供了技术支撑。大数据为我们打开了一个新的世界——大数据的镜像化世界，为我们提供了一种以统计学为基础的对“相关关系”进行研究的新的数据密集型科学思维范式。大数据还为我们提供了一种不同的信息方式，一种不同于“数字化生存”的“镜像化生存”。这是一种通过机器和技术“拟像”和“拟真”的生存，它并不能代替人类本真的生存；现实世界的“因果性”仍然是“镜像世界”数据“相关性”的基础和根基，人类世界仍然是大数据世界的意义来源，毕竟“人不是数据”。人不可能完全被数据化，数据化的镜像化生存不是人的基本生存状态；人类完全可以驾驭大数据。

关键词：信息方式；数字化生存；大数据时代；镜像化生存

作者简介：陈明，女，讲师，语言学硕士。（黑龙江大学 应用外语学院，黑龙江 哈尔滨，150080）

中图分类号：G206 **文献标志码：**A **文章编号：**1008-6552（2015）06-0143-06

大数据问题发端于计算机科学领域和信息科学领域，引起了科学研究的第四种范式——不同于传统计算机科学的数据密集型科学。大数据问题不仅仅局限于科学技术领域，它还在教育、传媒、商业、社会等领域产生了诸多影响，甚至改变了我们的生存方式。某种意义上，我们已经进入了“大数据”时代。哲学作为时代精神的精华，必须对大数据时代的本质做出自己的回答。

一、信息方式和技术的双重革命：从数字化生存到大数据时代

大数据应用涉及到“数字化”和“数据化”的问题。“数字化”和“数据化”是不同的信息方式。前者是信息社会的技术基础，后者是信息社会的信息基础，二者对信息社会的形成缺一不可。大数据问题的解决需要信息方式和信息技术的双重革命，我们进入了一种不同于数字化的生存的大数据时代。

按照舍恩伯格的^[1]理解，“数据”（data）是“已知”的“事实”。“数据代表着对某件事物的描述，数据可以记录、分析和重组它。”也就是所谓的对事物的“数据化”，它“是指一种把现象转变为可制表分析的量化形式的过程。”^[1]

大数据的发展与现代信息系统和信息方式息息相关。“大数据发展的核心动力来源于人类测量、记录和分析世界的渴望。”^{[1] (64)} 大数据不仅仅使人们关注技术，而是开始关注信息本身了。大数据为了得到量化的信息，必须对事物进行计量和记录。而计量和记录是数据化的前提，并为之奠定了良好的基础。但是大数据时代的到来需要新的计量、记录和分析技术。

大数据时代的到来与计算机技术的发展密不可分。计算机的出现带来了数字测量和存储设备，带来了计量和记录的革命，大大提高了数据化的效率。计算机也使得通过数学分析挖掘出数据的更大价值变成了可能。简而言之，“数字化带来了数据化，但是数字化无法取代数据化。数字化是把模拟数据变成计算机可读的数据，和数据化有本质上的不同。”^{[1] (66)}

数字化是信息社会的技术基础。数字化的理论基础是美国数学家克劳德·艾尔伍德·香农（Claude Elwood Shannon）在20世纪40年代所证明的“采样定理”。该定理说：在一定条件下，用离散的序列可以完全代表一个连续函数。不过“采样分析是信息缺乏时代和信息流通受限制的模拟数据时代的产

物”，^{[1](17)}是对应小数据范围才有用的方法。

数据化是对数字化的拓展与推进，是信息社会的信息基础。“虽然数字化是促成向大数据转变的重要原因，但仅有计算机的存在却不足以实现大数据。”^{[1](15)}数据化关注的焦点集中在数字信息更复杂、更高级的存在形态上。与数字化相比，“数据化意味着我们要从一切太阳底下的事物中汲取信息，甚至包括很多我们以前认为和‘信息’根本搭不上边的事情。这样，我们就激发出了这些数据此前未被挖掘的潜在价值。”^{[1](15)}只有新兴技术工具的使用才使这一切成为可能。

大数据本质上是一种信息系统和信息方式，它需要一种对其进行解码的计算机技术。通过计算机的解码我们会解释数据包含的语义内容，获得有关自然界和人类社会的认知“镜像”。但大数据的信息形式又是很难被结构化和形式化的，因此需要不同于传统形式逻辑的计算方法，需要对数理逻辑的变革和计算机技术的革命。对于数据化来说，“数字化文件不再是一幅图像，而是以字母形式出现、按一定结构组成的数据”。^[2]对于大数据技术来说，仅仅数字化是不够的。还必须处理许多非结构化、半结构化和许多其他类型的海量数据，而这需要不同于以往计算机的新技术和新算法。

大数据和云计算技术的发展密不可分。“在云计算出现之前，传统的计算机是无法处理如此量大、并且不规则的‘非结构数据’的。以云计算为基础的信息存储、分享和挖掘手段，可以便宜、有效地将这些大量、高速、多变化的终端数据存储下来，并随时进行分析与计算。大数据与云计算是一个问题的两面：一个是问题，一个是解决问题的方法。通过云计算对大数据进行分析、预测，会使得决策更为精准，释放出更多数据的隐藏价值。数据，这个21世纪人类探索的新边疆，正在被云计算发现、征服。”^{[1](序言)}正是计算机技术的发展才使大数据时代真正的到来，展现出大数据的分析、运用和发展前景，使人类过一种数据化的生活得以可能。

大数据时代还和电子媒介技术密不可分。大数据要把一切都“数据化”并储存起来以便随时进行量化分析，离不开电子媒介技术。电子媒介能把一切事物转换成数据和信息储存起来。通过电子媒介，过去作为我们“人的延伸的”的技术都会“转换成信息系统”。^[3]电子媒介具有无穷的转换能力，它“将自然完全变形或转换成人技术的世界”。^{[3](95)}过电子媒介的转换，我们可以转换意识，“将我们的意识迁移到电脑中去”，使电脑作为人的意识的延伸。通过“给人的各种感觉编制程序，使之更接近于人的意识。”^{[3](97)}不仅如此，电子媒介甚至“可以将整个世界的场景迁移到电脑的存储器之中。”^{[3](95)}总之，电子媒介技术使大数据时代得以可能，同时使“整个生活转换成信息的精神形态”，使“全球和人类大家庭被转换成统一的意识”得以可能。^{[3](97)}这种意识表现为舍恩伯格所说的“大数据思维”和量化一切的目标。

总之，大数据问题是数字化技术（信息技术）领域的新问题，它不同于传统计算机科学和信息科学，它需要一种新的科学技术——数据密集型的科学技术。其实是一种新型的数据处理技术——大数据信息技术。伴随着“分析与数据环境关联性、海量并行处理机构（MPP）、云计算、网格计算以及 MapReduce”等信息技术的发展，收集和处理海量数据已经成为可能。^[4]这些技术需要面对超大量的数据、多样化的数据和快速化的数据，对它们进行整理、挖掘和分析获得新的信息和知识，最终获得关于世界、自然和人类社会的“镜像”。和传统的数字化技术相比，大数据信息技术分析面临大量的、实时的（快速的）和多样化的（许多半结构化和非结构化的）数据，这需要建立新的数据库和新的数据分析平台。

大数据时代对信息技术和计算机科学提出了新的挑战，可能会引发计算机领域的革命。对我们当今以数字化为基础的信息社会更可能是一次“新的浪潮”。大数据造成了数字化信息时代的“信息爆炸”。不仅如此，信息的量变还导致了质变，大数据时代的信息方式应运而生。“随着计算机技术全面融入社会生活，信息爆炸已经积累到了一个开始引发变革的程度。它不仅使世界充斥着比以往更多的

信息，而且其增长速度也在加快。信息总量的变化还导致了信息形态的变化——量变引发了质变。最先经历信息爆炸的学科，如天文学和基因学，创造出了‘大数据’这个概念”。^{[1](37)}正是计算机技术和电子媒体技术的发展使“大数据时代”的到来成为可能，使人类对大数据的分析和挖掘成为可能。因此，不能离开计算机的发展和算法的发展来理解来孤立地理解大数据问题，大数据的分析以技术的发展为前提，大数据时代首先是由大数据技术造成的。

大数据技术造成了不同于尼葛洛庞帝所说的“数字化生存”的新的生活方式。人类“数字化”生存是指“我们生活在一个计算型的社会，因为我们相信世界可以通过数字和数学而获得解释。”大数据技术造成了“数据化生存”，舍恩伯格对这种生存进行了描述：“明天，我们的下一代，一群被‘大数据观念’陶冶长大的家伙，会发自肺腑地认为‘量化一切’并从中学习对于社会是至关重要的。”^{[1](73)}

二、大数据时代与人的镜像化生存

哲学作为世界观、认识论、方法论和价值论的统一，必须思考大数据给我们提供了一种什么样的世界观、什么样的认识论和方法论。大数据本身作为计算机科学领域的技术问题对哲学家而言只是提供了一种反思的前提，我们必须从世界观、认识论、方法论和价值论角度对大数据时代进行反思，在此基础上探讨大数据时代人的生存方式的变化。

从世界观上，大数据时代认为能量化一切，从最不可能的地方提出数据。“然而通过数据化，在很多情况下我们就能全面采集和计算有形物质和无形物质的存在，并对其进行处理。将世界看作信息，看作可以理解的数据的海洋，为我们提供了一个从未有过的审视现实的视角。它是一种可以渗透到所有生活领域的世界观。”^{[1](73)}不仅如此，通过大数据的量化，世界变成还可以被统计预测所控制的信息。“世界就是一堆桀骜不驯的信息的总和，一旦划定、命名、区分和量化（数据化？）之后，就能被人类驯服并加以利用。”^{[1](110)}

从认识论和方法论上，大数据时代引起了相关的思维变革。第一，要尽可能分析与事物相关的“全部”数据，而不是之前的随机抽样，即“样本=总体”。第二，要乐于接受数据的繁杂，而不应过分追求其精确性。第三，重视大数据呈现的“相关关系”，而不要执于探索事物间的因果关系。“数据化代表着人类认识的一个根本性转变。有了大数据的帮助，我们不会再将世界看作是一连串我们认为或是自然或是社会现象的事件，我们会意识到本质上世界是由信息构成的。”^{[1](73)}大数据时代我们应该建构一种基于数据信息“相关关系”预测的统计学意义上的概率认识论，它本质上与人工智能技术、机器算法和数据挖掘技术等领域密切相关，是一种大数据时代由数据决定的“信息认识论”。这种认识需要建立一种“大数据思维”，这种思维本质上是一种“工程师思维”和机器思维，“所以只会用工程师的观点解决问题——把所有决策简化成一个逻辑问题。数据成为了一切决策的主宰”。^{[1](112)}

在价值论上，大数据通过新型计算机技术和算法系统挖掘数据之间的关联和数据的潜在价值，为我们提供一种信息价值论的新方向。数据作为信息价值具有自己的特点，它可以重复使用而没有损耗，“不同于物质性的东西，数据的价值不会随着它的使用而减少，而是可以不断地被处理。而且信息不会像其他物质产品一样随着使用而有所耗损。”^{[1](77)}不仅如此，它还可以因不同的使用方式产生新的价值，可以释放出新的“隐藏价值”。“判断数据的价值需要考虑到未来它可能被使用的各种方式，而非仅仅考虑其目前的用途。”在大数据时代我们要通过技术手段“来释放数据的隐藏价值。”^{[1](78)}简言之，数据的价值是开放的、遵循“相关关系”并随着数据挖掘方式而不断生成，它有时甚至超过了数据提供者自身的想象。信息价值仿佛有着自身生成的规律，它已经不受信息提供者的控制。它是一个自律性的独立封闭系统，信息、数据的价值获得了一种类似本体的地位。

除了世界观、认识论、方法论及价值论上的革新以外，大数据还为我们人类提供了一个不同的时代

和崭新的生存方式,^[5]也即戴维-杰勒恩特所预言的“镜像世界”和镜像化生存。镜像世界和镜像化生存是建立在大数据的大量化(Volume)、多样化(Variety)和快速化(Velocity)三大特征之上的,反映了大数据概念的本质。

大数据反映了人类的镜像化生存。这里所说的镜像化生存是指以计算机、网络等硬件为基础的,以数字化数据及其运算来表征或显示物质世界中各种真实关系的生存方式。而人类在计算机网络里完成的那种本应在现实世界里完成的社会实践,我们也可以称之为“镜像实践”。

正是大数据时代营造出一个与真实存在的世界确实相关联的镜像世界,所以它才赋予了自己无与伦比的价值属性:因为存在着这样的真实关联,人们可以借用计算机设备和相关计算技术,通过研究这样一个镜像而间接获得对自然界和人类社会的了解及真知。显然,大数据为我们提供了一个独立的“数据化”的“镜像世界”。因为大数据的“镜像化”,往日的计算机虚拟世界不再是一个虚幻世界,而是通过数据化“拟像”建立了这个世界与真实世界之间的关联性和相关性,这种关联和相关为人们研究真实世界提供了一个前所未有的视角。

总之,大数据给我们带来一种真实世界和镜像化世界之间相互作用的“综合生存”,以至于我们有时甚至将“数据之镜”中的世界当成了真实的世界。当然,镜像化的真实并不是现实世界的真实,现实世界是不可能被完全镜像化的。现实世界存在本身揭示了大数据的界限:大数据技术挖掘、整理和分析获得的信息世界总是不完全的,它只是对真实世界的“形式化的”模拟、逻辑化的建构,它不可能代替真实世界。人们可能对大数据的“镜像”会产生“眩晕”,但是真实世界自身的存在力量总是会使大数据的世界发生“内爆”,从而使人们“溢出”镜像世界回到真实世界。简言之,大数据的镜像化世界仍然要根植于人的现实生活世界,并从人的生活世界获得其存在的意义。

三、对大数据生存的哲学反思

大数据主张对一切进行量化,进而运用统计学寻求数据之间的“相关关系”和“数理关系”。数据作为事物的“量”的规定,反映了事物一定的“质”。数据关联作为事物之间“量”的关系,当然也反映了事物之间的某种“质”的“联系”,揭示了事物和事物之间互动的某种规律。在此意义上,用大数据对事物进行统计学的把握,并对事物的发展趋势做出预测是有道理的。但是数据作为事物的量的规定,只是事物的一个方面。数据“相关关系”和“数理关系”只是事物之间的“量的关系”,它只是事物之间真实联系的“模拟”和“近似”,也即仅是一种“镜像性”的关联。它总是不完全的,它是对事物之间的联系的数量“抽象”,因此,大数据的量化的“相关关系”分析不能取代真实世界的“质”的“因果关系”分析,至多只是体现为一种对因果关系分析的“正面启发法”,有助于揭示事物之间真实的“质”的关联。“在大多数情况下,一旦我们完成了对大数据的相关关系分析,而又不满足于仅仅知道‘是什么’时,我们就会继续向更深层次研究因果关系,找出背后的‘为什么’。”^{[1](58)}

“在大数据时代,即使很多情况下,我们依然指望用因果关系来说明我们所发现的相互联系,但是,我们知道因果关系只是一种特殊的相关关系。相反,大数据推动了相关关系分析。相关关系分析通常情况下能取代因果关系起作用,即使不可取代的情况下,它也能指导因果关系起作用。”^{[1](58)}我们不认为相关关系能够取代因果关系,相反,它必须以事物之间的“因果关系”为基础。如果把它作为理解事物的全部,就会造成对真实世界的扭曲。大数据的“量化世界”和“量化关系”世界仅仅是真实世界和真实关系的一种“量”的抽象,它既是对真实世界的“镜像”反映,也是对它的镜像扭曲。真实世界的“质”和因果联系是不能被完全“量化”和“数理关系化”的。

大数据只是一个相对性的概念,它相对于传统的数据处理技术是“大”的,它表现为数据的“大量化”,数据类型的“多元化”,以及具有实时特征的“快速化”。相对于人类的生活实践和真实世界

的存在，大数据实践和大数据世界作为镜像实践和镜像世界，它总是不足的、是不够“大”的。

大数据技术的核心就是预测。这种预测基于一种机器内部的算法——云计算技术。大数据不是要教机器像人一样思考，相反，它是要人像机器一样思考。“它是把数学算法运用到海量的数据上来预测事情发生的可能性。”大数据技术是一种基于独立的机器系统和算法系统之上的预测系统，“这些预测系统之所以能够成功，关键在于它们是建立在海量数据的基础之上的。此外，随着系统接收到的数据越来越多，它们可以聪明到自动搜索最好的信号和模式，并自己改善自己。在不久的将来，世界许多现在单纯依靠人类判断力的领域都会被计算机系统所改变甚至取代。”^{[1](52)}人最后将被还原为一个可计算的算法系统和自动机器，人被数据分析和算法完全量化了，变成了“数据人”。人完全无法驾驭数据，反倒是数据控制了人。更确切地说，是数据之间的“相关关系”系统控制了人。

相关关系分析是大数据预测的核心。“相关关系的核心是量化两个数据值之间的数理关系。”^{[1](50)}相关关系意味着一种数据自身独立生产的“意义结构”，它意味着“黑格尔式的线性、深度因果观逐渐被一种非线性、平面式的结构观所取代。这种结构不再强调意义的真实来源，而是强调意义如何被生产。”^[4]数据之间的相关关系结构变成自律性的能指符号网络，数据的“所指”（真实世界）意义已经变成了数据能指关系网络中的一个变量。真实世界被“数据”之间的关系所“拟像”了，真实世界成了一个由数据“拟真”的世界，也即通过数据之间的关系被“模拟”，从而看上去“像”是“真”的；真实世界的意义依靠数据的“相关关系”决定的预测。大数据为我们描绘的世界是一个“拟真”和“拟像”的世界，它在某种程度上建构并生产着（掩盖着）真实世界的真实关联。简言之，大数据“数据化”的信息世界是对真实世界的“拟像”和“拟真”，它需要技术的解释，而技术又需要人的解释。人的生活世界才是大数据世界的意义之根。这里，鲍德里亚对资本主义拟真世界的批判，同样适用于大数据化的信息世界。“拟真世界是最大的伪象征构境，是以比真实更真实的拟真逻辑埋葬了象征交换。它以人与人之间的虚假的象征符号关联替代了物性功用存在链。这样，人类存在将永远跌落于符码控制和诱惑的黑暗之中。”^[6]当然，人类同样也可能跌落于大数据造成的镜像世界中而不能自拔，这主要体现在以下这些问题中：

无所不在的数据采集会不会侵犯个人的隐私？基于相关关系的行为预测，会不会侵犯人们的自由意志？“一个用数据进行预测的时代”，“会对人类自由和尊严产生什么影响呢？”^{[1](32)}大数据公司会不会形成新的数据垄断，会不会对数据进行滥用？“大数据时代也需要新的规章制度来保卫权势面前的个人权利。”^{[1](78)}人们会不会过于信赖数据，形成数据独裁？“如何来监管一个算法系统呢？”^{[1](78)}大数据造成的“数据库的话语”会不会变成“在后现代、后工业化的信息方式下对大众进行控制的手段”？^[7]如何合理地利用大数据使之变成强大的武器，挣脱大数据所造成的镜像世界的困境等等。这些问题都是不容回避的哲学问题。

四、结 语

大数据给我们带来了一种新的世界观——数据化的“镜像世界”，这个世界本质上是由云计算等新的计算机技术提供的大数据的海量信息世界。“大数据标志着人类在寻求量化和认识世界的道路上前进了一大步。过去不可计量、存储、分析和共享的很多东西都被数据化了。拥有大量的数据和更多不那么精确的数据为我们理解世界打开了一扇新的大门。社会因此放弃了寻找因果关系的传统偏好，开始挖掘相关关系的好处。”^{[1](90)}通过大数据提供的这个世界，我们获得了一种镜像化的生存。但本质上这是一种机器和技术“拟像”和“拟真”的生存，它并不能代替人类本真的生存。现实世界的“因果性”仍然是“镜像世界”数据“相关性”的基础和根基，人类世界仍然是计算机和机器世界的意义来源，毕竟“人不是机器”。

大数据带来的思维变革本质上是机器思维和机器学习，而不是人的思维变革；是计算机工程思维而不是人类的科学认知思维。单纯对数据进行“相关关系”分析可能会导致数据的独裁和控制，“因为数据量极为庞大，最后做出决策的将是机器而不是人类自己。”^{[1](112)} 要避免这种情况，“其解决之道在于挖掘因果逻辑而非相关性”。“放弃对因果性的追求，就是放弃了人类凌驾于计算机之上的智力优势，是人类自身的放纵和堕落。如果未来某一天机器人和计算完全接管了这个世界，那么这种放弃就是末日之始。”^{[1](译者序)} 无疑，大数据是一个很重要的概念，代表了很重要的趋势，但大数据不是万能的。人不可能“完全”变成数据，大数据对于人类的生存来说不是“全数据”。“我们能收集和处理的只是世界上极其微小的一部分。这些信息不过是现实的投影——柏拉图洞穴上的阴影罢了。因为我们无法获得完美的信息，所以做出的预测本身就不可靠”。^{[1](131)}

现实世界和人类世界永远抵抗着、驾驭着、引领着大数据的“量化”，建立在机器和算法系统之上的大数据预测系统永远要依靠人类社会这一开放的、复杂的历史系统才能获得其可理解性和存在的意义。“大数据提供的不是最终答案，只是参考答案，为我们提供暂时的帮助，以便等待更好的方法和答案出现。这也提醒我们在使用这个工具的时候，应当怀有谦恭之心，铭记人性之本。”^{[1](131)}

参考文献：

- [1] [英] 维克托·迈尔-舍恩伯格. 大数据时代——生活、工作与思维的大变革 [M]. 盛杨燕, 周涛译. 杭州: 浙江人民出版社, 2012: 64.
- [2] [美] 尼葛洛庞帝. 数字化生存 [M]. 胡泳, 范海燕译. 海口: 海南出版社, 1996: 216.
- [3] [加] 马歇尔·麦克卢汉. 理解媒介——论人的延伸 [M]. 何道宽译. 北京: 商务印书馆, 2004: 93.
- [4] 卞友江. “大数据”概念考辨 [J]. 新闻研究导刊, 2013 (5).
- [5] 贾利军, 许鑫. 谈“大数据”的本质及其营销意蕴 [J]. 南京社会科学, 2013 (7).
- [6] 张一兵. 反鲍德里亚: 一个后现代学术神话的祛序 [M]. 北京: 商务印书馆, 2009: 31.
- [7] [美] 马克·波斯特. 信息方式 [M]. 范静哗译. 北京: 商务印书馆, 2000: 132.

(上接第142页)

- [8] 马克思恩格斯全集 (第25卷) [M]. 北京: 人民出版社, 1974: 712-713.
- [9] 马克思恩格斯全集 (第42卷) [M]. 北京: 人民出版社, 1979: 167.
- [10] [美] 唐纳德·沃斯特. 自然的经济体系——生态思想史 [M]. 北京: 商务印书馆, 1999: 426.
- [11] 马克思恩格斯全集 (第23卷) [M]. 北京: 人民出版社, 1974: 66.
- [12] 世界自然保护联盟等. 保护地球——可持续生存战略 [M]. 北京: 中国环境科学出版社, 1992: 31.
- [13] 习近平. 在中央政治局第六次集体学习时的讲话 [EB/OL]. http://news.xinhuanet.com/video/2013-05/24/c_124761554.htm. 2013-05-24.