

当代社会生活的大数据化：困境与反思

郝庭帅

提 要：大数据技术以前所未有的方式将人、事、物三者之间的关系数量化、数字化，它深深渗透并影响着当代社会生活。本文试图探讨“大数据”作为一种新的认知方式和生活方式的贡献与挑战，反思其局限与困境。第一，大数据推进了复杂事物的分析理解，但这一研究范式主要强调相关关系而忽视因果关系，形成了以结果和效用为导向的研究思路忽略了社会现象背后的机制，影响了其使用效果。第二，大数据技术将信息数据化和商业化，个体丧失了对自身信息的选择权和控制权，诱发了触犯个体隐私的伦理问题。第三，大数据与功利主义结合更大限度上推广大数据决策，但这一决策模式弱化人的自由选择和传统道德规范的价值。

关键词：大数据 社会生活 预测 隐私 功利主义

一、“大数据”的兴起

在二十世纪后半叶，自计算机出现以来，数据信息处理分析不完整的情况普遍存在。例如，在气象学领域能够被及时处理与分析的数据只是所拥有数据的很小一部分。又如，基因组学和大型强子对撞机产生海量数据，因人工或单台计算机无法在合理时间内完成截取、管理、处理，或整理成为人类所能解读的信息。但随着信息技术的进步，工程师们采用了分布式架构，进行分布式数据挖掘，依托云计算的分布式处理、分布式数据库和云储存、虚拟化技术等，实现了对海量数据的处理，大数据由此兴起。

在数据分析专业领域，人们普遍认为大数据具有 4 个“V”的特点，即 Volume（大量）、Velocity（高速）、Variety（多样）、Veracity（真实性）。与小型数据集的数据分析相比，处理海量数据即将各个小型数据集合并后进行数据分析，得到了许多预设之外的信息。由于人们可以从各种各样类型的数据中快速找到要素在数据中

的联系，在数据来源的表象上反映出事物间的相关性，大数据逐渐备受关注。大数据技术不但可用于天文学、基因组学和大气科学等专业学科的数据计算与处理，还应用于工业的技术革新、物流与交通、公共医疗与个人健康、预测商业趋势和商品价格走势、预测人的社会行为及其偏好等。本文试图对迅猛发展的大数据及其运用进行讨论，说明大数据技术在推进认识复杂事物、便捷社会生活的同时，暗含着不易解决的研究难题和伦理困境。本文试图探讨大数据作为新的认知方式和生活方式的贡献与挑战，反思其局限与困境。第一，大数据推进了复杂事物的分析理解，但这一研究范式主要强调相关关系而忽视因果关系，形成了以结果和效用为导向的研究思路忽略了社会现象背后的机制，影响了其使用效果。第二，大数据技术将信息数据化和商业化，个体丧失了对自身信息的选择权和控制权，诱发了触犯个体隐私的伦理问题。第三，大数据与功利主义结合更大限度上推广大数据决策，这一决策模式弱化人的自由选择和传统道德规范的价值。

二、作为新的科学研究范式

下面首先介绍一组大数据应用的实际案例以呈现大数据运用的基本模式和特点。本文认为是由于对大数据的数据挖掘形成的新的数据分析思路导致的科学范式的改变，从而起到了不同于传统的预测效果，特别是，大数据下的研究者更重视变量的相关关系而非因果关系。维克托（Viktor Mayer Schnberger）和肯尼思（Kenneth Cukier）（2013）在《大数据时代》一书中大量描绘了大数据分析的实际案例。我们以谷歌关于流感预测的研究分析为案例进行初步讨论。

谷歌将 5000 万条美国人最频繁检索的词条和美国疾控中心在 2003 年至 2008 年间季节性流感传播时期的数据进行了对照以关注的特定检索词条的使用频率与流感在时间和空间上的传播之间的相关关系。为了测试这些检索词条，他们总共处理了 4.5 亿个不同的数学模型。将得出的预测与 2007 与 2008 年美国疾控中心记录的实际情况进行对比后发现，他们的软件发现了 45 条检索词条的组合，将它们用于一个特定的数学模型后，他们的预测与官方数据的相关性高达 97%。2009 年甲型 H1N1 流感爆发的时候，与习惯性滞后的官方数据相

比，谷歌则要更有效，更及时（维克托、肯尼思，2013：3）。无独有偶，MIT媒体实验室人类动力学实验室的研究者也用手机预测疾病传播，并成功地区分出了感染了流感的人群（维克托、肯尼思，2013：118）。

（一）科学预测及其边界

从上述案例可以发现大数据分析一个重要的应用效果是对未来变化趋势做出预测，这甚至是大数据的研究人员视为核心的东西（涂子沛，2013，2014；赵国栋等，2013）。“预测”一词是典型的人的认知系统的词语，但并非所有大数据都是为了预测。如天文学和基因组学对大数据技术的利用，就是纯粹基于数据量级过大，只有大数据技术可以进行处理。

大数据分析能否得到预设之外的新发现，要看这门学科的理论的完整度。例如天体物理学的理论对天体之间的关系的解释已经相对完整，大数据所提供的所有相关关系都在理论控制的范围内，这时并不存在预测行为，大数据技术本身只是处理庞大数据的工具。而如果利用大数据确实找到了理论外的相关关系，预测到了某天体理论值外的运动，那这件事情也并不能说是预测，反而是更高限度的对理论在未知层面的补充，那这种相关关系的存在本身就会得到研究者的重视，作为探究因果关系的基石。同样，对于基因组学来说，大数据技术的运用是为了效率不得不为之的，否则无法处理庞大的数据，但对于这门人类还所知甚少的科学，任何发现都有可能是理论之外的，那这时任何相关关系都有可能被视为因果关系用来建立解释研究对象的理论，这种相关关系的发现也并不算是预测了什么。以上两个例子说明并非所有大数据技术的运用都能起到预测的效果，尤其是在自然关系的探求上。

我认为，现在人们所重视的大数据预测有两个核心特征或条件。第一，大数据应用所在领域的理论有待完善的空间，大数据技术找到的相关关系是已知理论不能解释的，具有附加应用价值的。第二，大数据应用是与人的行为和生活有关的。例如大气科学，和前两个例子一样，也是因为数据量太庞大，不得不采用大数据技术，而且研究的也是自然规律。但是它具备以上两个条件，大数据分析有可能找出现在已知理论之外的相关关系，而这种关系可以提高天气预测的准确度，从而影响了人们的生活，所以大数据预测天气可视为预测。同样地在地质学上可以用大数据预测地震，也符合这个标准。相反，基因组学在得到实验数据阶段只满足第一个条件，不满足第二个条件的，所以不是预测。但当它运用于基因排序和疾病预测的时候，

就满足了第二个条件。

即使这样还是有不少学者认为，大数据除了解决人类社会生活的具体问题，还是变革性的科研手段，可以为原来手段受限制的学科增加新的维度和研究方式。如周涛（2013a）认为，“在科学范式发生转变之后，基于实验室小规模控制实验的半定量科学，将走向基于大规模而非控制数据分析的定量科学，如社会学、心理学、管理学等。”

（二）由分析性到综合性之争

传统的科学研究范式被大数据技术改变后，研究者和使用者对于研究范式的选择存在差异，也包含着某些角色的错位。电子科技大学周涛教授认为，“这种科学范式的变化并不复杂，传统的模式是利用数据建立理论，再用数据和理论给出解释，继而利用这个解释或是理论做出预测，接着利用这些预测来检验理论是否正确，最后实行控制和干预。但是在大数据时代传统的科学范式发生了转变，有数据和理论可以解释，没有理论，单是有数据本身，虽然得不到解释，但是可以做到预测和控制”（周涛，2013b）。如谷歌利用搜索记录对流感的预测，其所掌握的只有数据，但是却得到了高效的预测。因此倾向于大数据应用的实际效果的人认为理论被终结了。美国斯坦福大学逻辑组的研究员扎芬恩（Zavain Dar，2014）认为，“从分析方法论转向综合方法论的转变其潜在意义将是巨大的。以往，这些方法局限在人工智能及信息获取等计算机科学子域，一切数字化以后，我们将不再需要预先设定的结构和公理这些知识性甚至哲学性的负担了，基础架构的普及就可以加速综合型问题解决的普及。”

对于扎芬恩（Zavain）的观点，我认为即使同意其所说的综合型问题确实可以被如此解决，大数据的效用可以被肯定，但是结构和公理真的是冗余吗？凯文（Kevin Kelly）在《失控》一书中写到接球的案例，即接球这个行为也可以被看成预测。一个球抛向接球者，接球者成功接住了它，这是不是接球者对球的运动轨迹的预测。在接球时，接球者脑中并不存在牛顿经典力学的公式（ $F = ma$ ）及其解读。即使是三五岁的孩子也有能力接住朝他飞来的球，无论这是否证明接球者脑中存在由其经验和感觉建立起来的球的运动轨迹的模型，在接球的过程中牛顿力学公式（ $F = ma$ ）对接球者而言是没有意义的。可能懂得力学原理还不如多练习两次于接球这件事影响大，但难道能因此认为牛顿力学公式（ $F = ma$ ）不值得探究，是没有价

值的公理吗？

回想谷歌的流感预测，诚然传统的疾病传播规律并没有被利用，但是从数据到预测这个过程中就不存在理论吗？搜索的动机、词语的运用背后是心理学、社会心理学、语言学、传播学和人类动力学等学科有关联的研究对象，对这些背后的因果关系的探究可能对于预测流感这件事没有直接的帮助，但是并不能否定这些公理的价值，也否定这些相关关系作为重要研究素材的价值。

（三）由因果关系到相关关系

大数据分析的效果就是寻找到数据间的相关关系。虽然因果关系本身就是人类认知所追求的，但推崇大数据这一技术特点的学者认为我们应该降低对因果关系的追求。这是基于大数据技术自身，尤其是机械式学习导致的，技术的使用者往往容易采用以结果和效用为导向的研究思路（周涛，2013c）。如政府通过大数据分析可以知道一座城市的经济状况和商业活跃度和这座城市的人的通话记录的广泛度有关，越是频繁地与城市以外的人联系，并且联系越广泛和分散，这座城市的经济就越好。如果市政府把这个相关关系视作因果关系的话，出于发展经济的目的，市政府可以制定一个经济政策，例如对长途电信的费用进行补贴，来鼓励市民多和城市以外的人通信。显然，这样的政策的有效程度是值得怀疑的，但是，我们无法确定政府或大的企业在制定政治和经济决策时，有没有将相关关系理解为因果关系，从而误解或夸大了某些数字的意义。

然而，在运用大数据分析的结论时，是否也该警醒或反思大数据相关性分析的结论的陷阱和局限？以谷歌流感预测为例，有研究者指出这三年来谷歌流感预测并没有交出令人满意的答卷。英国《新科学家》周刊网站报道称，在过去3年，该系统一直高估与流感相关的就医量，在这类数据最有用的流感季节高峰期尤其预测不准确。在2012年和2013年的流感季节，它预测的就医量是美国疾控中心（CDC）最终记录结果的两倍。

可以说，这在很大程度上源于大数据分析忽视了对因果关系的探究。当认为相关关系重要于因果关系的观点，最终影响到了大数据应用本身的效果时，相关关系还有什么意义呢？我认为大数据下相关关系和因果关系之间张力的来源，一方面是已知理论外的模糊领域，另一方面是会对人的生活产生重大影响的应用效果，使用者的价值诉求和研究者的价值诉求因这两方面产生歧义，所以形成了相关关系和因

果关系谁更重要的争论。

同时，角色的错位使得为因果关系辩护的人很难发声，因为这完全是来自专业学科之外的新手段。研究流行病传播的专家无法批评谷歌预测流感的手段，只能以其结果效用进行评判，因为这是两个领域完全不同的研究方法。他也很难从人们特定的搜索词语中，学习到可用于医学的信息，所以他在坚持自己工作意义的同时，承认探究因果关系在自身领域的重要性，但也不反对大数据所呈现的相关关系，只是不会也没有能力对这些相关关系里可能存在的因果关系进行研究。

三、个体信息的“大数据化”

“量化一切”是信息时代的主题。大数据的出现为数据化带来了利益驱动，促使数字化的个人信息转向可分析的数据，以获得更多的产品和服务。大数据分析存在对个人隐私的窥视，这种对隐私的侵犯因为大数据分析的效果而不同于以往的隐私安全问题。

(一) 个人信息的数据化、商业化

从前人们并不关心的个人信息，随着技术的发展这些信息被有意或无意地记录了下来，通过大数据分析生产出了新的产品或服务，反馈了我们的社会生活，这是事情正向发展的自然过程。但是原本对信息的数字化并不带有数据挖掘的动机，只是技术实现的必然趋势，例如短信，数字化是为了完成信息的传递，并不是为了大数据对其文本内容的挖掘。从这个角度看几乎所有个人数字化信息都不具有数据化的初衷，因为数字化只是借道于计算机的运算、储存和传输的二进制语言，而数据化只是把现象转变为可分析和重组的量化形式。

大数据的出现为数据化带来了利益驱动，促使数字化的个人信息转向可分析的数据，以获得更多的产品和服务。姜奇平（2012）在《爆发》一书的推荐序中指出，这些被数据化的个人信息包括：个人的地球上的全部运动轨迹（通过LBS采集）（如微信）；个人的全部消费记录（通过在线支付采集）（如支付宝、网银、银联）；个人的全部交往记录（通过SNS采集）（如人人网、Facebook）；个人的全部言行记录（通过邮件、手机通话、时间表、视频监控等采集）。这里还不包括未

来十年会有长足发展的可穿戴设备、体征监测设备、智能家居和智慧汽车等，可以想象不到二十年的时间，人的全部行为都会被不同程度的记录，而且是数据化的。

个人是否还拥有这些信息呢？要讨论这个问题，需要先回答这些问题：谁的信息？个人的信息，甚至可以说是现代社会每个个体的信息。谁在数字化这些信息？电脑、智能手机、银行卡、摄像头等在将信息数字化。谁保管这些信息？谷歌、百度保管用户的搜索信息，微信、微博、人人保管用户的交往信息，移动联通保管用户的通信信息，银行保管储户的消费和存款信息，医院保管患者的健康信息，公共场所的管理者和政府部门保管人们的监控录像信息等。谁加工这些信息？本身具有大量数据和大数据处理技术的互联网公司在处理这些信息，专业的数据分析公司、政府部门和科研单位也在处理这些信息。谁使用这些信息？互联网公司本身、与人的生活相关的各种提供商品或服务的公司、政府部门和科研单位在使用这些信息。

上述每个环节的主体都是有差别的，正是这些主体的差别和每个环节的合理性，导致人们对自身的信息的保护意识大大下降。例如，人们不会给一个在大街上卖药的推销员留自己的电话，但是人们会给医院留自己的电话，当然人们是知道医院贩卖患者的信息是不道德的，甚至是违法的。实际上，在商业伦理不健全的情况下，这种事情是经常发生的，像孕妇在医院刚生产完，就会接到电话问需不需要制作手模和胎毛笔等，这就是典型的信息从生产到使用之间的某环节为了利益而出卖了信息。

但是大数据时代的情况又与传统的情况不尽相同，这主要是因为信息加工者采用了大数据分析技术。首先是信息的产生是伴随数字化过程自然发生的，这使得这些信息难引起个人的注意，而数据的加工者向各种信息保管者购买了信息，混杂的信息在大数据分析后得出了与初始信息完全无关的新信息，因此这种关系是无法直观理解的。所以使用者使用时，个人无法知道其是何时何地以何种途径泄露了自己的信息，甚至人们可能不知道有人正在使用自己的信息。例如，传统情况下，如果推销者知道了某人有某种病，向其推荐药物，那这个人就能猜测到是医院出卖了他的信息；但是在大数据技术下，人们不知道推销者是如何获知自己患病的，有可能是微博，有可能是网购记录，甚至有可能是导航里的位置信息，不知道源头人们就很难采取自我保护，以前人们可以换一家医院或找监管部门投诉，但是现在人们连要投诉谁都不知道，甚至连数据加工者本身可能都搞不清楚是哪些信息出卖了用户。

在这种情况下，人们虽然在某种程度上还拥有这些信息，可以查看自己的通信记录、消费记录等，但是无法阻止别人也在使用自己的数据。作为信息的生产者，人们不能阻止信息的产生，因为这些信息是伴随个人的行为而来的。在很大程度上人们也没办法阻止信息数字化，电脑和智能手机几乎绑定了人的生活，一旦使用就很难不再使用。即使有人不使用智能设备、银行卡和汽车，但是家里的电表水表、出行的飞机票火车票和公共场所的监控录像等这些都是不由个人选择的关于个人信息的数据信息。政府从前只是根据人们的水表电表来收取水费电费，现在政府可以通过人们的水电数据知道其有没有非法改建为集体宿舍，几点起床和睡觉，起不起夜等。

在大数据时代的背景下，数据化的大趋势看来是不可避免的，因为大数据的应用效果有足够的利益驱使企业和政府进行数据化和数据整合，而个人的数据又不可避免的会被使用。首先，人们无法避免自己在互联网上的公开信息被使用，如微博、社交网络，这些用户自己公开的信息是任谁都可以使用的^①；其次，人们无法阻止政府和寡头企业利用自己的数据，而恰恰是他们掌握了人们大量的数据，像谷歌的大数据项目几乎都在使用用户的搜索记录，而政府更是会以各种名目滥用甚至窃取个人的信息，如棱镜计划^②。人们似乎还拥有着自己的数据，但是其他人也同时拥有着你的数据。个人对自己的信息丧失了控制权，正是一切大数据与隐私之间伦理问题的原罪。

（二）隐私与窥视——数据分析的原罪

复杂网络（Complex Network）和人类动力学^③的研究者认为人类行为存在既有

① 实际上，网络公开信息的使用规则正在因个人数据保护法律的逐步完善而改变。早在1995年，欧洲议会和欧盟就提出了“被遗忘权”的概念，意指任何公民可以在其个人数据不再需要时提出删除要求。2014年5月13日，欧洲法院裁定，普通公民的个人隐私拥有“被遗忘权”，并据此要求国际网络搜索引擎巨头谷歌必须按照当事人要求删除涉及个人隐私的数据。

② 棱镜计划（PRISM）是一项由美国国家安全局（NSA）自2007年小布什时期起开始实施的绝密电子监听计划。美国情报机构一直在九家美国互联网公司中进行数据挖掘工作，从音频、视频、图片、邮件、文档以及连接信息中分析个人的联系方式与行动。

③ 人类动力学（Human Kinetics）是一门新兴的交叉学科，主要由统计物理学家发起并推动。人类动力学关注人类日常的行为模式，力图挖掘新的统计规律并建立相应的动力学模型。复杂网络（Complex Network），具有自组织、自相似、吸引子、小世界、无标度中部分或全部性质的网络，研究的内容主要包括：网络的几何性质，网络的形成机制，网络演化的统计规律，网络上的模型性质，以及网络的结构稳定性，网络的演化动力学机制等问题。

的规律，而利用现在广泛的数字化信息采集和大数据分析人们可以掌握其中的规律，继而预测大部分的人类行为。除了学者在这方面努力探究外，企业和政府也在积极利用新科技带来的好处，但他们并不像研究人员一样致力于探究人类行为的本质规律，而是希望通过预测人的某些行为，帮助自己更好地完成工作。零售行业希望卖出更多的商品，所以他们就像沃尔玛一样开始利用大数据关联商品的销售，他们也会通过你的搜索记录、浏览记录、消费记录和微博等数据，利用大数据得出你的关联商品，进行营销。美国洛杉矶的警察局和加州大学合作，利用大数据技术分析了过去几十年这一地区犯罪的时间空间属性，得到了最佳的巡逻模型。这是充满新希望的人类科技，但是工具依然呈现它的两面性，如何在这个时代保证隐私和个人自由，成为新的难题。

人们该如何对待这样的关联商品营销？2012年《纽约时报》曾发表一篇文章，报道塔吉特（Target）公司有一个分析项目，可确定一位顾客何时怀孕，并将购买与妊娠有关的物品的优惠券送给其中一位少女，该少女父亲得知后非常恼怒，痛骂该公司经理。人们该如何看待这件事？第一，大数据分析得出的相关关系是高概率的，但不是一定的，那未怀孕的女性收到这类优惠券就是不尊重。第二，对怀孕的女性来说，连自己的家人可能还不知道的隐私，门口的推销员却得知了，也是不尊重的。即使假设这位少女怀孕的信息并没有被使用，她没有收到优惠券，营销系统出于她过小的年龄弃用了她孕妇的标签。这种情况是否可以接受呢？不管人们接不接受，后果已经出现。首先，是人们已经无从知道自己的某些隐私已经被知道了，像在被人窥视，而自己并没发觉；再则，即使大数据分析出的数据没有被人使用，也不能证明这些被大数据分析出的关于个人的数据就是合理无害的。

第一，人们只知晓自己产生的信息的意义，不知晓大数据处理后的信息的意义，所以人们可以控制哪些信息被自己产生，但不能控制哪些信息被大数据解读出来。所以说，人们所同意的是某些服务商拥有特定的信息，即我自己产生的信息，但是人们并未同意这些服务商或其他机构拥自己所产生的信息以外的信息，即大数据分析的数据。例如，从用户在脸书网（Facebook）上的一百多个点赞行为判断出某位用户是同性恋。在这个案例中，假设我是当事人，我在社交网络中的点赞行为，自然是我认为公开的信息，我也知道互联网上的所有人都可以看到这个点赞，但我同意这个信息被他人知道，我未同意的是有人知道我是同性恋这个信息，就像我不会在社交网络里直接说我是同性恋一样，我留下的公开信息是我同意的，而大数据通

过我留下的公开信息解读出来的是我未同意的。

第二，大数据分析是寻找数据中的相关关系，所以它一定有其预设目的。这些目的可能是关联到某种商品和服务，也可能是预测某种行为和爱好。数据库收集了关于个人的信息，即使没有使用也并不代表这些个体的身份特征不存在，恰恰可能是因为某人具有某种特征，这个信息才没有被使用。如某人同时是未成年和孕妇，营销公司可能为了避免纠纷将系统设置成非目标客户，所以关于她的数据并非真的没有被使用，只是她恰好在影响之外而已，但她身份特征的数据信息确是存留在数据库里。

从以上两点看，大数据所涉及的隐私问题区别于以往的传统隐私和网络隐私的问题。以前的隐私问题主要是信息的出卖和隐私信息的使用不当，主要集中在信息保管者和使用者身上，通过建立监管和个人信息使用标准可以解决大部分实际问题。但是正如上文所讨论的，大数据技术所带来的隐私问题的核心是数据加工者利用人们同意的公开的个人信息生产出了人们未同意的个人信息，这些信息里有多少是人们会同意商家或机构可以拥有的，有多少是不会同意的，甚至不想任何人知道的，这些数据的加工者和使用者并不知道，而个人也丧失了对这些信息选择和控制的权利。这样的时代，哪些信息还能被称为隐私？在共享经济巨大的利益驱使下，我们会变成透明人吗？

（三）个体行为的选择与“被选择”

大数据时代，可穿戴的体征监测设备通过你的体征数据和过去八小时的运动情况可以预测出“你饿了”这个事实，甚至早于你自己发现“你饿了”，随后通过你的饮食喜好、经济状况、时间安排和地理位置帮你推荐附近最适合你的餐厅，并把这个消息推送给你，这一切都是系统自动化的采集分析，并没有与你有任何的意识交流，当你得到这个信息推送的时候，是突然被告知“你饿了”，甚至早于你自己察觉。这其中并没有第二个人，只是机器与你互动。这样的生活确实很方便，但是方便的同时人们是否也在失去一些东西？大数据时代，人们是在选择还是在被选择？亚马逊网站推荐的图书总是用户的最爱，美食软件总能找到适合每个人口味的餐厅和菜品，淘宝网总是知道消费者下一件想要的商品是什么。人们在选择商品，还是大数据在帮商品寻找客户，或许这并不重要，有人认为只要系统能帮人做出最好的选择，到底是谁选择的又有什么关系！事情真的应该这样对待吗？大数据给出的是

不是最佳选择，人们又应不应该遵从最佳选择呢？

如上面这个案例中，假设我是当事人，我起码有三个选择被影响了。第一，我是否询问智能系统关于就餐的事；第二，我是否吃饭，或马上吃饭；第三，我是否去系统为我选择的餐厅就餐。第一个选择的干扰是最好解决的，即系统设定为被动模式，不主动建议，而是等候人的问询。这对基于大数据预测的智能服务系统来说，将会是很重要的守则，因为如果任由系统随意发表建议，那人做选择的主体性会受到很大的影响。第二个问题也伴随第一个问题解决，因为询问是在我做出是否吃饭的选择之后。第三个问题是一个大数据预测的代表性问题，即大数据推荐的选择是否就是最佳选择，人们是否应该遵从这种最佳选择。我可以相信大数据给我的是不错的选择，起码是不会令我后悔的选择，甚至可以假定它就是最适合我的选择。但我不认为它是最佳选择，因为大数据的分析方式是根据一个既设目的找相关关系，这里目的一定是寻找适合我的餐厅，而利用的数据包括我的行为习惯和与我相似的人的行为习惯和过往选择和评价。

但是目的之外的收获是大数据没法计算的。比如去另一家餐厅的路上有一家我会很感兴趣的商店，又或者会遇到一见倾心的人，这种事情自然也是大数据能预测的。但是当大数据预测的目的只是餐厅这个主题时，它没有任务去预测那些意外，那所谓的最佳选择也就变成了目的范围内的最佳选择。再比如导航系统，最开始时是简单的路径计算，融入大数据后，可以根据路况和未来路况帮用户找到耗时最短的路程。假如某人的目的不是用最短的时间到达目的地，而是想兜风或欣赏风景，那导航给他的路径就不是最佳选择，又或者他的目的是大数据的既设目的，路径是最佳路径，但是并不代表如果他走了别的路就没有收获，人类发展很多时候不正是靠着走错路才走出了新的道路？

显然如果人们习惯了大数据帮自己做决策和选择，那他的生活应该是便捷和有质量的，但是少了意外，少了犯错的机会，少了体验痛苦、饥饿、烦躁、无聊的机会。大数据可以帮人们找到适合自己的结婚对象，但是人们也相应地丧失了失恋的机会，失恋无疑对人的成长是有很大影响的，但是这些都在这个大数据预测服务的目的之外。你还觉得你应该交出选择权吗？上文提到的案例都是理想型的大数据服务，但是存不存在可能会影响到个人自由的大数据预测呢？上文提到的案例已经多多少少影响了人们的自由了，至少是选择的自由。

四、大数据决策的功利主义

大数据在商业领域广为使用，下面尝试探索大数据未来可能更极端化的运用和影响。尽管当下的科技还不成熟，但是可以猜想科技发展的某些趋势。无论是可穿戴设备和体征监测设备，还是基因工程和人工智能，都将很大程度上和大数据技术相融合，互助发展。在这种趋势下技术与功利主义相融合，诱发更大程度的数据决策，人的量化的实现则将这种功利主义数据决策推向了人本身，最终形成的选择导向无疑是对自由主义的颠覆和造成传统道德伦理的意义丧失。

这种猜想是基于大数据技术本身与功利主义的相融，大数据技术具有后果主义和工具主义的导向。大数据的算法就能被看成一个具有工具理性的黑匣子。周涛（2013）指出“大数据分析的主要手段，如机器学习，解决问题的方式是训练所有可能的模型和拟合所有可能的参数，问题从一个端口进去，答案从另一个端口出来，中间则是一个黑匣子”，只在乎完成目标的必要手段。下面结合功利主义的一般结构，来讨论这个问题（程炼，2008）。

1. 福利主义：“功利主义者都持有一种福利主义的价值观，这种价值观说，福利是伦理考虑的唯一基本因素，福利的提升或降低是仅有的善和恶，凡是不影响福利的升降的事情都与道德无关。”
2. 个体福祉的基本性：“功利总是相对于可受益或受损的对象而言的，更确切地讲，功利必须落实到个体对象上。”
3. 功利的定量化：“功利是一个定量的或者至少原则上可定量的概念。正功利与负功利是可以换算的。正价值和负价值相互抵消之后总有一个净价值存在。”
4. 普遍性：“当我们考察一个行动带来的功利时，我们需要考虑行为对所有受到影响的个体的功利。”
5. 不偏不倚：“功利主义要求平等地看待每个受到影响的人的利益。”
6. 功利的最大化：“正确的行为就是将功利最大化的行为。”

在这个结构上，大数据技术与福利主义的概念似乎不能直接关联。大数据可以

落实到个体对象上，甚至比人更小的个体上，满足个体福祉的基本性；大数据本身就是定量的技术，满足功利的定量化；大数据是全体数据，多样而混杂，最具普遍性；大数据不具备人类情感，满足客观公正；大数据擅长于寻找人类行为和事物发展的相关关系，满足功利最大化的计算要求。那福利如何在大数据技术上体现呢？假设使用边沁对福利的理解——趋乐避苦，将大数据分析的目的视为大数据与增加快乐减少痛苦有关系，但是这个命题中如何将快乐和痛苦解释为计算机语言是最大的障碍。

我现在假设一个数值维度相对单一的环境，尝试论证大数据能否完成功利主义的计算。如上文中的汽车导航系统，但是它的目的变了，不是为自己寻找到达目的地用时最短的路径，而是为这座城市的所有汽车寻找到达所有目的地的总和时间和最短的路径。在这个假想实验中，先假定有且仅有到达目的地的所用时间长短影响每个人的福利，且每单位分钟对应的功利是相等的，无视人在其中的差别。默认每辆车一个人，或者系统有手段得知每辆车里有几个人。这时，这个城市的全交通导航系统应该是这样表现的。大数据分析可以进行全交通的分配，如果 A 车走 B 路径需多耗时 2 分钟与 M 车走 N 路径少耗时 5 分钟两条数据有相关关系，那系统就会如上导航，总体节约 3 分钟，即获得 3 个单位的功利。以此扩展到所有的车辆，最后得出用时总和最小的所有路径。但是导航和司机的实际驾驶还是会有出入，比如有的司机不小心开错了，有的司机认为导航指示的路径太远自找捷径，这都要求大数据系统马上计算出新的最优路径。是不是只要按照导航显示的路径行驶就是功利主义的正确的行为，我想应该不是的。因为人无法真正百分之百地按照系统规划的时间空间属性驾驶，总会有细微的误差，又或者说是大数据无法将预测修正到百分之百符合实际情况，所以人完成的是不断在变化的功利主义的行为，没有完成一个真正的正确的行为。但是可以认为人们完成了规则功利主义下的正确的行为，因为导航规划的路径作为一个能够实现最大功利的原则，被人们遵守了。如果是全自动驾驶的汽车遵循导航的路径，我认为完成的应该就算是功利主义的正确行为了，虽然全自动驾驶也会有细微的误差，但是这里主要的区别是人的随意性和难以量化。

即使这是个不成熟的实验，但是也足以证明大数据技术是功利主义在可实践性上的一块有利基石。大数据反驳了三个对不可实践性的反驳。第一，即使面临的行动选择再多，决定行动的后果的因素再多，对于大数据来说都只是计算量而已，简易的技术问题，不会产生无从选择、无法计算的问题；第二，大数据是全体样本，

对所有人的行为和后果都充分了解，不会产生行为人不知道后果对其他人的影响的问题；第三，有人认为人们已经认识到的某些行动规则是可以基于之前的功利判断沿用的，但是古典功利主义认为只有一条关于正确行为的规则，因此，人们每次行动都要重新做计算。大数据对之前计算的结果没有依赖，无论如何都要根据当下的实际情况和目的再算一次，所以正符合功利主义的要求，重新计算并不成为负担。以上的论证只能证明大数据在功利主义下有可实践性，与人的行为的功利主义的可实践性没有关系。

从这个实验假想看，功利主义的实践貌似是完成了。但是这个实验设置的前提是很苛刻的，几乎完全放弃了功利的定义这个核心。如果真的以趋乐避苦作为功利的标准，来进行这个实验，情况就要复杂很多，且无法完成。系统要计算每个人对于路程用时的心理感受，行驶期间所看所听所感的苦乐度，关键还是苦乐度无法真正量化为一种具有标准的客观值，人类也不具备一种量化的表现形式。这也是功利主义理论渴望统一所有价值所要面临的必然挑战。

体征监测设备或者为量化人的生理状态提供了可能，如果可以对人体的电波和各种激素、酶实现监测的话，人们或许可以用综合性的生理指标来定义快乐和痛苦。通过大数据分析，人们也能预测什么事在什么情况下会对特定的人产生多少的快乐或痛苦。如果这样的话，是否就真的完成边沁心目中的功利主义了呢？人的情感和一切有别于动物的特点是不是都可以用生理指标来量化呢？穆勒并不这样认为，虽然同是功利主义的拥护者，穆勒否认“一切快乐经验都是在内在性质上相同的，而差别只在于量上的不同”，而认为“人类真正的幸福，要求比单纯的感觉享受更多的东西。人们一方面需要低级的肉体快感，更重要的是，他们也需要文化、智力、友谊、知识和创造力，后一类事物是人类幸福不可缺少的要素”（转引自程炼，2008：154－155）。穆勒将功利主义从边沁的定量中解放了出来，这种将幸福的要素从质上的区分，也正对应了上文对大数据预测预设目的之外的行为或特点存在的合理性的辩护。大数据的预设前提没办法涵盖全部的幸福要素，或是单一目标的全部效用可能，走错路对于导航来说效用是负的，但是实际情况中不一定没有收获，有可能哪怕错误都是有价值的。

举例来说，像见义勇为这种行为，如何计算功利？假设在大数据技术和体征监测技术都完善的情况下，一个人看见另一个人落水，应不应该救？如果是按边沁的可量化的快乐来计算，完全可以依靠大数据预测救与不救的结果，大数据知道当事

两个人的身体状况和幸存或亡故后所有相关人的苦乐值，结合概率给出一个判断。当事者遵循这个判断就是规则功利主义的正确行为，反之，就是错误的行为。在这里救人与不救人本身没有道德意义了，有道德意义的是是否遵循了大数据给出的功利计算的判断。如果用穆勒的观点，就要考虑到见义勇为这件事作为文化在人类社会的作用，大数据无法对未来如此大的随机性进行判断，所以单可以掌握人的苦乐度还不可以完成这个功利主义计算。

如果大数据技术可以充当规则功利主义中的那一组原则的话，那遵循大数据的判断，就是规则功利主义的正确行为。当水边有很多人可以救落水者的时候，并不是谁去救了就有道德，不去救就没道德，而是由大数据告诉所有人，应该由谁去救，大家都遵循这个判断。或许因为某人是游泳教练，身体素质好，心理素质强，成功救人的概率最高；或许因为某人是孤家寡人，即使因救人而死，不会有太多人难过。总之，救人在这里已经不是道德上的事情，只不过是大数据眼里数据库里各种数值的比拼。而这种对功利主义的遵循，使得人们也没有了除了遵循大数据判断外的道德选择。那在大数据面前，现在人们是数据库里一个个属性标签，未来可能是一组数值，最后甚至可能是全部的生命体征，社会的一切选择与道德评判都归决于这些数值。当代社会中，数据决策的影响越来越大，一组组量化值和属性标签越发成为人们做选择和决策的依据，而正如上文所讨论的，大数据技术可能会将这种趋势发展向更极端的境地。

五、结语

无疑，大数据逐渐成为当代社会生活的重要内容，深刻影响着人们的行为方式。即使现在人们也会认为两米以上身高的人就应该去打篮球一样，人们会觉得大数据推荐的工作是最适合自己的，吃的、玩的、恋爱对象、结婚对象都是理所应当的数值匹配。人们是否会在这种值化下放弃了多元道德的选择，承认一个没有伦理的运算世界。如果大数据技术可以在青少年时期甚至更早就预测到某些孩子是同性恋的话，这些孩子的父母要如何看待和解决这个问题，外界的干涉会不会对这些孩子本身的成长形成导向，原本高概率成为同性恋的孩子因为这些外界的引导，最终会形成如何的性取向？这种导向的影响更直接反映在孩子的兴趣爱好、能力专长和职业

选择上，如果大数据预测某个孩子以后会更擅长某项技能，从事某种职业获得成功的概率最大，那父母是否会因此而对孩子做出引导，这种导向到底是有益于其个人的发展的，还是影响其自由选择的。这种导向会不会促使政府或公司等权力机构对组织成员采取更为极端的家长式的管理，而这种管理正是基于大数据对人的自由选择形成了导向，并最终成为标准化的选择模式。人们会不会在这种导向下忘记了自由的真实含义，家长式管理会不会将大数据的高概率推向一定，没有小概率的世界将会是什么样子？自由主义和传统伦理在大数据面前还能走多久？又会走向哪里？这些问题值得我们对大数据的普遍应用及其社会后果予以深刻反思。

参考文献：

- 艾伯特，拉斯洛，巴拉巴西，2012，《爆发》，马慧译，北京：中国人民大学出版社。
- 埃里克，西格尔，2014，《大数据预测》，周昕译，北京：中信出版社。
- 程炼，2008，《伦理学导论》，北京：北京大学出版社。
- 基思，威利茨，2013，《数字经济大趋势》，徐俊杰、裴文斌译，北京：人民邮电出版社。
- 姜奇平，2012，《推荐序》，载《爆发》，艾伯特·拉斯洛·巴拉巴西著，马慧译，北京：中国人民大学出版社。
- 凯文，凯利，2010，《失控》，陈新武、陈之宇等译，北京：新星出版社。
- 洛克，1964，《政府论（下篇）》，叶启芳、瞿菊农译，北京：商务印书馆。
- 涂子沛，2013，《大数据》，南宁：广西师范大学出版社。
- ，2014，《数据之巅：大数据革命，历史、现实与未来》，北京：中信出版社。
- 维克托，迈尔舍，恩伯格、肯尼思·库克耶，2013，《大数据时代》，盛杨燕、周涛译，杭州：浙江人民出版社。
- 威廉，詹姆斯，2013，《实用主义》，陈羽纶、孙瑞禾译，北京：中国青年出版社。
- 约翰，穆勒，2008，《功利主义》，徐大建译，上海：上海人民出版社。
- ，1959，《论自由》，许宝骙译，北京：商务印书馆。
- 扎芬恩·达尔，2014，《大数据将改变人类解决问题的方式》，载36氪网站，2月10日（<http://www.36kr.com/p/209580.html>）
- 赵国栋、易欢欢、糜万军、鄂维南，2013，《大数据时代的历史机遇》，北京：清华大学出版社。
- 周涛，2013a，《大数据的1.0版本，2.0版本和3.0版本颠覆性变化下的商业革命》，人民论坛第15期。
- ，2013b，《大数据商业革命和科学革命》，《视听界》第03期。
- ，2013c，《译者序》，载《大数据时代》，维克托·迈尔舍·恩伯格·肯尼思·库克耶著，盛杨燕、周涛译，浙江人民出版社。

作者单位：北京大学哲学系
责任编辑：艾云