

大数据与大数据经济学

俞立平

(宁波大学商学院 现代管理研究中心, 浙江 宁波 315211)

摘要: 本文从大数据的发展现状分析入手, 讨论了大数据对传统经济学的挑战, 首次提出大数据经济学的概念。认为大数据经济学包括大数据计量经济学、大数据统计学和大数据领域经济学, 并分析了大数据经济学与信息经济学、信息技术等相关学科的关系, 最后对大数据经济学发展前景进行了展望, 认为大数据经济学不仅将理论科学、实验科学、复杂现象模拟统一在一起, 而且将自然科学和社会科学统一在一起, 将理论与实践应用实时地统一在一起, 大数据经济学具有“智能经济学”的特点。

关键词: 大数据; 大数据经济学; 大数据计量经济学; 大数据统计学

中图分类号: F01 **文献标识码:** A **文章编号:** 1002-9753 (2013) 07-0177-07

Big Data and Big Data Economics

YU Li-ping

(Business school of Ningbo University, Modern Management Center, Ningbo 315211, China)

Abstract: This paper analyzes the challenge of big data to traditional economics from big data development. It first gives the definition of big data economics. Big data economics contains big data econometrics, big data statistics and big data field economics. The author still analyzes the relationship of big data economics and information economics, information and technology subject etc. Big data economics combines with not only theory science, experimental science, and complex phenomena simulation, but also natural science and social science. Big data economics is intelligent economics.

Key words: big data; big data economics; big data econometrics; big data statistics

一、引言

2012年, Twitter上每天发布超过4亿条微博, Facebook上每天更新的照片超过1000万张, Farecast公司用将近10万亿条价格记录来预测机票价格, 准确率高达75%, 采用该系统购票, 平均每张机票可节省50美元。Gartner预测未来5年全球大数据将会增加8倍, 其中80%是非结构化数据。2013年世界上存储的数据将达到1.2ZB

(1ZB = 1024EB, 1EB = 1024PB, 1PB = 1024TB, 1TB = 1024GB), 如果将这些数据刻录到CD只读光盘上, 并堆起来, 其高度将是地球到月球距离的5倍。2011年, 麦肯锡公司对全世界大数据的分布作了一个研究和统计, 中国2010年新增的数据量约为250PB, 而欧洲约为2000PB, 美国约为3500PB, 大数据已经深深地充斥了人类经济社会的许多角落。

收稿日期: 2013-01-27 修回日期: 2013-03-28

基金项目: 技术创新金融研究(12FGL007)。

作者简介: 俞立平(1967-)男, 江苏姜堰人, 博士, 宁波大学商学院教授, 研究方向: 计量经济、科学计量。

著名未来学家阿尔文·托夫勒 (1980)^[1] 很早就在其经典著作《第三次浪潮》中, 将大数据热情地赞誉为“第三次浪潮的华彩乐章”, 但是大数据成为高频词是最近一两年的事情。随着社交网络、物联网、云计算的兴起, 数据规模越来越大, 2011 年 5 月, 全球知名咨询公司麦肯锡 (McKinsey and Company) 发布了《大数据: 创新、竞争和生产力的下一个前沿领域》^[2] 报告, 标志着“大数据”时代的到来, 指出“数据已经渗透到每一个行业和业务职能领域, 逐渐成为重要的生产因素; 而人们对于海量数据的运用将预示着新一波生产率增长和消费者盈余浪潮的到来”。2012 年世界经济论坛发布了《大数据、大影响》^[3] 的报告, 从金融服务、健康、教育、农业、医疗等多个领域阐述了大数据给世界经济社会发展带来的机会。2012 年 3 月, 奥巴马政府发布《大数据研究和发展倡议》^[4], 投资 2.5 亿美元, 正式启动大数据发展计划, 计划在科学研究、环境、生物医学等领域寻求突破。据 Gartner 公司 2012 年 8 月发布的技术发展生命周期^[5] 趋势图 (图 1), 大数据不到两年时间内成为新技术发展的热点。一时间大数据蜂拥袭来, 那么什么是

大数据? 大数据对传统经济学会带来哪些冲击? 传统经济学应该如何面对大数据带来的挑战?

对于什么是大数据, 目前业界并没有公认的说法。Dumbill (2012)^[6] 采用 IBM 公司的观点, 认为大数据具有“3V”特点, 即规模性 (Volume)、多样性 (Variety)、实时性 (Velocity)。以 IDC 为代表的业界认为大数据具备“4V”特点, 即在 3V 的基础上增加价值性 (Value)。NetApp 公司^[7] 认为大数据应包括 A、B、C 三大要素, 即分析 (Analytic)、带宽 (Bandwidth) 和内容 (Content)。所谓大分析 (Big Analytics), 指通过对大数据进行实时分析后带来新的业务模式, 帮助用户获得洞见, 从而更好地进行客户服务; 高带宽 (Big Bandwidth) 指快速有效地消化和处理大数据; 大内容 (Big Content) 一方面指大数据包括结构化、半结构化数据与非结构化数据, 另一方面则是指对数据的存储扩展要求极高, 能轻松实现数据的恢复、备份、复制与安全管理。Gartner 认为, 大数据需要新的处理模式才能具有更强的决策力、洞察发现力和流程优化能力, 满足海量、高增长和多样化信息资产的需要。

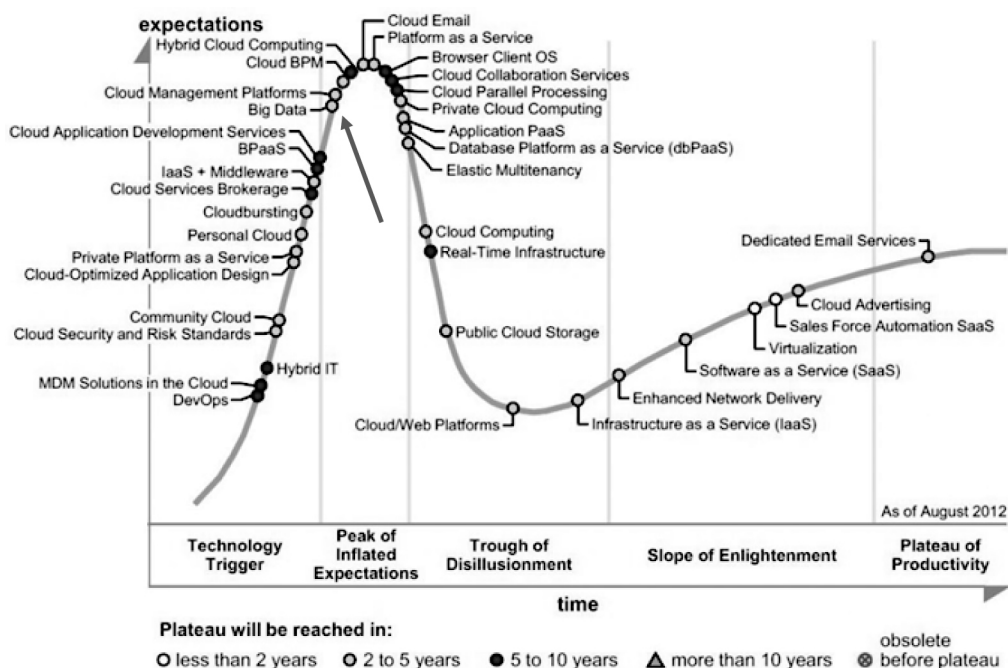


图 1 Gartner 新技术生命周期

大数据是工业传感器、互联网、移动数码等固定和移动设备产生的结构化数据、半结构化数据与非结构化数据的总和,大数据重在实时的处理与应用,以获得所需要的信息和知识,从而实现商业价值以及为公共管理服务。数据挖掘和人工智能等应用工具在大数据处理中发挥着重要作用,现代信息技术是大数据赖以存在和发展的主要支撑力量。

二、大数据给经济学带来的影响

Victor (2012)^[8]在其最新著作《大数据时代——生活、工作与思维的大变革》中指出,大数据时代,思维方式要发生3个变革:第一,要分析与事物相关的所有数据,而不是依靠分析少量数据样本;要总体,不要样本。第二,要乐于接受数据的纷繁复杂,而不再追求精确性。第三,不再探求难以捉摸的因果关系,应该更加注重相关关系。杨华磊(2013)^[9]分析了高频数据对传统经济学研究范式的冲击,出现了“非主流经济学就是致力研究异常现象的经济学”,当然高频数据与大数据不是一回事,两者之间存在交集。那么,大数据给经济学带来了哪些影响呢?

(一) 大数据研究对象变成了总体

传统经济学研究中,由于搜集数据的条件所限,人们往往对数据进行抽样,用少量样本来进行研究,这一传统一直延续至今,并且成为经济学研究的主流做法,但是抽样的质量对研究结果影响很大,比如公众对政府统计部门公布的物价指数和基尼系数引发的怀疑。在大数据时代,很多场合下已经无需进行针对样本的研究,直接将总体作为研究对象,从而很大程度上改变了数据来源方式,对数据的处理也产生了深远的影响。

(二) 大数据不需要基于假设检验的研究

传统的经济学研究,往往根据研究内容提出数个假设,然后再采用数学模型基于统计检验来验证假设。但在大数据时代,由于有足够的变量、足够的数据,可以采用人工智能来进行数据挖掘和知识发现,得到的结论是成百上千的,和传统经济学研究需要验证假设的数量永远不是一个数量级。在大数据时代,如果继续采用传统的

假设检验方法进行研究,永远是不充分的、不完备的、无法满足需要的。大数据时代重在对数据处理结果的多样化结果进行分析,可以是基于经济学的,也可以是基于应用的,从而辅助人们决策。

此外,由于变量的完备性要求使得传统的基于假设验证的研究有时变得十分尴尬。比如,研究研发投入对企业绩效的影响,需要考虑的不仅仅是研发投入,还要考虑企业资本结构、竞争水平、人员素质、行业特点、管理能力等诸多因素,研究者重点关注的是研发投入的弹性系数,但却得到了其它所有数十个变量的弹性系数,从而使研究重心不容易掌握。

(三) 大数据使得因果关系变得不太重要

传统经济学是一门解释科学,重在对经济现象进行解释,了解他们的因果关系,但在大数据时代,这样做是远远不够的,大数据甚至可以发现事物发展潜在的规律,以供经济学家解释,具有一定的“智能性”,某种程度上超越了经济学研究的因果关系。

大数据并没有改变因果关系,但是使传统经济学的因果关系变得不太重要。比如经济学家在预测房价时,无非是根据住房价格变化的影响因素来进行分析,比如经济发展水平、人均收入、土地价格、宏观房产政策、地点等因素。但谷歌预测房价时,根据住房搜索查询量变化进行预测,结果比不动产经济学家的预测更为准确及时。IBM日本公司,通过检索关键词“新订单”、“雇员”、“生产”等来预测采购经理人指数,仅用6小时就得出结果,并且和专业的采购指数分析师们计算的结果基本一致。大数据并没有改变因果关系,但使因果关系变得意义不大,很多时候因果关系成为“正确的废话”。

(四) 传统的因果关系有时无法验证

弄清事物之间的内在联系和作用机制,一直是传统经济学的核心。但有时因果关系是没有办法验证的。比如新产品上市,人们往往倾向于购买新产品,这样对旧产品的需求会下降,那么旧产品价格应该立即回落,这是其一。从另外一个角度,如果大家都认识到这一点,就会贪便

宜购买旧产品，短期内会造成旧产品供不应求，反而导致旧产品涨价。究竟是涨是跌，要看这两种因素谁弱谁强，采用传统经济学研究方法是难以验证这两种效应的，只能验证两种效应作用的综合结果。

实际情况是，在大数据时代，西雅图 Decide.com 公司分析了近 400 万商品的超过 250 亿条价格信息，发现新产品上市时，短期内旧产品价格是上涨的，过一段时间才逐步回落。采用大数据，既可以知道多少人购买旧产品，也能知道多少人购买新产品，以及旧产品价格变化的规律。在这种情况下，我们知道所有的因果关系，却难以检验，并且没有意义，知道结果更重要。

(五) 传统经济学研究具有滞后性

传统经济学对于新生事物是不敏感的，必须等事情发生并且成长到一定规模以后才能搜集到足够数据进行相关研究。在大数据时代，可以通过海量数据对经济行为进行分析，一旦有新情况、新动态立即予以关注，从而实现对新事物的早期干预和分析，因此具有前瞻性。大数据本身就具有智能，可以辅助经济学发现知识。

(六) 大数据对基于统计检验的计量经济学冲击很大

建立在回归和统计检验基础上的计量经济学以其严谨的逻辑成为经济学研究的重要方法论，迄今为止，诺贝尔经济学奖获得者有近半数计量经济学家，但大数据动摇了这一根基，比如采用普通回归研究自变量 X 于因变量 Y 的关系，对于 X 回归系数采用 t 检验，一般认为相伴概率小于 0.05（特殊情况可以放大到 0.1）就说明两变量相关。其实在这种情况下，犯两变量不相关错误的可能是 5%，以 CNNIC 发布的《第 31 次中国互联网络发展状况统计报告》^[10] 为例，2012 年底我国网民数量达 5.64 亿人，假设我们研究网民平均受教育年限（ X ）与上网时长（ Y ）的关系，5% 就是 2820 万人，此时我们还能漠视这 5% 的错误吗？同样，如果 t 检验的相伴概率为 0.95，那么很明显说明平均受教育年限与上网时长不相关，但同样会犯错误，即有 5% 的可能性

平均受教育年限（ X ）与上网时长（ Y ）是相关的，会涉及 2820 万网民，这同样是不能忽视的。

(七) 大数据对经济学建模提出挑战

传统的经济学研究，往往采用 1 个或少数几个数学模型来进行研究，但任何模型都各有长处，也各有其局限，没有包治百病万能的数学模型。比如动态面板容易使投入变量的弹性系数估计变小，空间面板容易出现空间矩阵设置方法不当导致结果偏误，面板变系数模型难以和空间面板结合使用，面板联立方程模型对方程形式的要求极高，面板向量自回归模型难以和空间面板融合等等。在研究同一问题时，可用模型其实较多，有没有最佳模型呢？这恐怕是个无解的问题。实际情况是，迄今为止传统经济学研究得出的结论，至多只能说明采用甲模型的结论，并不具有普适性，换个乙模型结论可能立即就变了，其实研究结论是脆弱的。

此外，在研究同一个问题时，即使采用同一模型，由于模型的变量选择、估计的方法、参数设置、滞后期选择等不同，也会导致估计结果相差很大。

在大数据时代，借助云计算和分布式处理等现代信息技术，往往可以采用成百上千的模型来进行研究。Google 公司在预测 2009 年美国甲型 H1N1 流感爆发时间时，把 5000 万条美国人常用的检索词条和美国疾控中心 2003 - 2008 年期间季节性流感传播数据进行比较，希望通过搜索记录判断这些人是否得了流感，先后共采用了 4.5 亿个不同的数学模型，预测结果和官方数据的一致率高达 97%，但比官方节省了两周时间，从而为政府采取相关措施赢得了宝贵的时间。

在传统经济学研究中，由于研究对象错综复杂，直接影响与间接影响因素众多，变量的完备性被认为是不可能的事情，往往只能选取少数变量来进行研究，达到一个相对满意的结果。在大数据时代，我们可以获取越来越多的变量，从而使遗失变量的可能性降到最低，这样在研究中由原来的数个变量可能会变成数十个甚至成百上千的变量，在这样的情况下，对原有的建模技术就

带来了巨大挑战,对计量经济学的发展将会产生深远影响。

(八) 大数据给经济学研究工具和手段发生变化

传统经济学研究,一个团队、数台电脑、几个软件就能进行像样的研究,很少有运算需要动用大中型服务器的,但在大数据时代,经济学研究发生了巨大的变化,在人员组成上,不光要有经济学家和领域专家,还要有大数据维护专家、大数据建模专家;在计算工具上,需要广泛借助云计算,几台电脑根本解决不了问题;从合作关系上,需要广泛与政府、大数据拥有者、云计算服务商等合作,不然难以进行研究。大数据时代,经济学研究必须依靠跨学科团队,传统的少数几个学者就能进行研究的模式已经难以为继。

(九) 大数据彻底改变了传统的统计调查方式

大数据彻底改变了传统的统计调查方式,比如对于经济指数、物价指数的计算,完全可以采用全新的模式,彻底摒弃传统方式。对于统计学中的异常点,以往的处理方式往往是丢弃,或者是平滑,但在大数据时代,由于样本众多,异常点成为宝贵的资源和研究对象广受重视。传统的统计数据是经过加工后的结构化的数据,在大数据时代,人们更加重视原始数据和非结构化数据,因为如果统计数据已经经过加工,那就变成了二手数据,如果一手数据加工过程出现问题必然导致后续处理出现误差。此外,大数据还使间隔时间较短的高频数据研究成为可能。

三、大数据经济学

(一) 大数据经济学的定义与研究内容

考虑到大数据给传统经济学带来的巨大冲击和影响,迫切需要对此进行研究,斯坦福大学的教授、沃尔玛全球电子商务的高级副总裁、WalmartLabs 的共同创立者 Anand Rajaraman (2012)^[11]发明了一个新词 Econinformatics,指将计算机科学和信息技术应用于经济学领域,特别指应用于大数据的经济分析。由于该词和 Information Economics 的意义相近,翻译成中文后更容

易混淆,加上其和 Ecoinformatics (生态) 相近,因此并不是一个好的名词。本文提出大数据经济学 (Big Data Economics 或 Economics of Big Data),给出如下定义:

大数据经济学是在经济学研究和应用中采用大数据并且采用大数据思想对传统经济学进行深化的新兴交叉学科。大数据经济学不仅要研究如何建模、管理和应用大数据,而且要深入研究传统经济学如何应对大数据带来的挑战并进行改良,大数据经济学需要经济学家、领域专家和信息技术专家等密切合作,对人文社科与自然科学的跨学科研究提出了更高的要求,并且对整个经济学、社会学、公共管理等将带来革命性变革。大数据经济学的研究内容包括:

第一,大数据计量经济学 (Big Data Econometrics)。这是和传统计量经济学对应的一个学科,也是大数据经济学下面的子学科。在大数据背景下,经济学建模与分析方法与传统计量经济学完全不同,迫切需要采用全新的思路和方法进行研究。对信息技术专家们而言,大数据经济学仅仅是算法和建模问题,但是如果没有经济学理论指导,没有经济学家的思维,必然会导致研究方向的迷失。一些大数据领域的学者认为“要相关,不要因果”,这是非常要不得的,传统经济学理论至今仍然到处闪烁着智慧的光芒,对经济现象的深入见解时刻发挥着重要的作用,所以大数据背景下的经济学分析不能主要靠信息技术的建模专家来进行,必须继续依靠大数据计量经济学家。

第二,大数据统计学 (Big Data Statistics)。如前所述,大数据给统计学带来的挑战是革命性的,在某些领域,传统统计学所采用的抽样调查方式必将彻底淘汰。此外,传统统计学所要求的精确数据与数据加工方式可能是画蛇添足甚至败笔之举,人们更加重视对一手数据而不是经过加工过的二手统计数据进行分析。大数据时代,人们更加关注原始数据、关注半结构化甚至非结构化数据,浏览记录、查询关键词、微薄文字、照片等等都是宝贵的数据资源。在大数据时代,传

统统计学也必须进行变革,对数据储存手段、处理设备、处理方法都提出了新的要求。

第三,大数据领域经济学。包括大数据生态经济学、大数据环境经济学、大数据金融学、大数据城市经济学、大数据工业经济学、大数据农业经济学、大数据交通经济学、大数据建筑经济学、大数据商业经济学、大数据信息经济学、大数据人口经济学等学科,借用大数据的思想和技术来进行各应用经济领域的研究。

在以上大数据经济学的各学科中,大数据统计学是基础,大数据计量经济学是研究方法,而大数据领域经济学是具体的运用,他们之间存在着密切的共生关系。

大数据由于是基于总体的,很大程度上解决了传统宏观经济学与微观经济学缺乏较强逻辑联系的问题,此外大数据对传统计量经济学带来的一个有益之处就是,结构化的大数据更加接近正态分布,这样就降低了小样本假设检验失效问题。

(二) 大数据经济学与传统经济学的关系

大数据经济学刚刚提出,现在讨论其与其它学科的关系也许为时尚早。大数据经济学与传统经济学是一种互补共存关系。在大数据经济学诞生之初,由于大数据经济学理论和技术尚不成熟,虽然大数据经济学发展很快,但仍然以传统经济学为主。随着大数据经济学的发展,两者会达到某种均衡。毕竟,大数据不能解决所有的经济学问题,一些研究仍然无法获得大数据,需要采用传统经济学解决的问题留待传统经济学解决,需要大数据经济学解决的问题由大数据经济学解决。

(三) 大数据经济学与信息经济学的关系

大数据是现代信息技术高速发展的产物,因此必须研究大数据经济学与信息经济学的关系。传统信息经济学 (Information Economics) 包括两个部分:宏观信息经济学与微观信息经济学,严格意义上讲,这两者并没有必然的关系。Machlup (1962)^[12]和 Porat (1977)^[13]是宏观信息经济学的创始人,又称为情报经济学、信息产业经济

学,主要从产业经济学角度研究信息这一特殊商品的生产、流通、利用以及经济效益的一门新兴学科,研究视角集中在信息化与产业经济学,是经济学的重要分支。

Stigler (1961)^[14]和 Arrow (1972)^[15]是微观信息经济学的创始人,又称为理论信息经济学,研究不对称信息理论、信息商品的分析、信息成本 and 价格、信息市场分析、信息搜寻理论等,提出用不完全信息理论来修正传统的市场模型中信息完全对称的假设,又称契约理论或机制设计理论。

大数据产业自身发展的经济学问题仍然属于信息经济学的范畴,不属于大数据经济学。随着大数据的迅猛发展,据世界经济论坛预测,大数据会为全球带来 440 万个 IT 岗位,其中 190 万个在美国,另外每一个大数据的岗位会催生 3 个非 IT 就业岗位,也就是说未来会推动美国产生 600 万个就业岗位,这类问题就是信息经济学的研究范畴。

(四) 大数据经济学与信息技术及其它相关学科的关系

毋庸置疑,大数据经济学离不开现代信息技术,是现代信息技术发展到大规模计算与存储阶段的必然结果,甚至在信息技术专家眼里,大数据仅仅是一种技术。但是大数据经济学更是一种思想,只不过现代信息技术使这种思想成为可能。大数据经济学必须以现代信息技术为基石,重在研究其在经济学领域中的应用,因此大数据经济学是一个学科跨度很大的学科,包括经济学、管理工程、统计学、信息技术、情报学、心理学等相关学科。

四、大数据经济学发展展望

本文首次提出大数据经济学的概念,大数据经济学将是 21 世纪经济学的重大进展之一。它是随着大数据在人类经济社会中的应用而产生的,目前尚处于萌芽阶段,其实践远远超越理论,可以预见的是,不久的将来是大数据经济学的理论建构和高速发展期,借助高度发达的现代信息技术,大数据经济学理论可以随时得到检验和修

正, 这样一开始大数据经济学就处在一个很高的研究和应用水平上, 其发展速度远远高于其它任何新兴学科, 这也是现代信息技术对新兴学科的重要贡献之一。

从学科分类上, 目前的经济学包括理论经济学与应用经济学两个一级学科, 可以预计的是, 随着大数据经济学的日益发展与成熟, 大数据经济学将成为和理论经济学与应用经济学并列的一级学科, 是经济学一级学科中的“小弟弟”。

图灵奖得主 Jim Gray 2007 年在美国国家科学研究委员会发表演讲, 指出科学研究共经历了 4 个阶段: 数千年前, 人类注重采用实验科学来描述自然现象; 几百年前, 人类注重理论科学; 几十年前, 人类转向计算科学, 模拟复杂现象; 而今天, 人类进入数据探索阶段, 将理论科学、实验科学、复杂现象模拟趋于统一。Jim Gray 的结论主要针对自然科学, 对大数据经济学而言, 不仅是将理论科学、实验科学、复杂现象模拟统一在一起, 而且将自然科学和社会科学统一在一起, 将理论研究与实践应用实时地统一在一起, 大数据经济学将是智能经济学。

参考文献:

- [1] 阿尔文·托夫勒 (Alvin Toffler) 著. 第三次浪潮 [M]. 黄明坚译. 北京: 中信出版社, 2006.
- [2] Big data: The Next Frontier for Innovation, Competition, and Productivity [EB]. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation, 2013. 1. 24.
- [3] Big Data, Big Impact: New Possibilities for International Development [EB]. <http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>.
- [4] Big Data Across the Federal Government [EB].

http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet.pdf, 2013. 1. 24.

[5] Gartner's 2012 Hype Cycle for Emerging Technologies Identifies "Tipping Point" Technologies That Will Unlock Long-Awaited Technology Scenarios [EB]. <http://www.gartner.com/newsroom/id/2124315>.

[6] Edd Dumbill. What Is Big Data [EB]. <http://strata.oreilly.com/2012/01/what-is-big-data.html>, 2013. 1. 24.

[7] NetApp. Big Data Solutions for Government [EB]. <http://www.netapp.com/us/solutions/industry/government/bigdata.html>, 2013. 1. 24.

[8] 维克托·迈尔·舍恩伯格, 肯尼斯·库克耶著. 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.

[9] 杨华磊. 高频数据对传统经济学研究范式的冲击 [EB]. <http://blog.sciencenet.cn/blog-456786-656901.html>, 2013. 1. 24.

[10] 第 31 次中国互联网络发展状况统计报告 [EB]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwjbg/201301/t20130115_38508.htm, 2013. 1. 24.

[11] Welcome to Econinformatics: Home of Economics & Big Data [EB]. <http://econinformatics.com/blog/2012/01/23/welcome-to-econinformatics-home-of-economics-big-data>.

[12] Machlup F. The Production and Distribution of Knowledge in the United States [M]. Princeton University Press, New Jersey, 1962.

[13] Porat M U. The Information Economy [M]. Washington, DC, Government Printing Office, 1977.

[14] Stigler G J. The Economics of Information [J]. The Journal of Political Economy, 1961, 69 (3): 213-225.

[15] Arrow K J. The Economics of Information [M]. Harvard University Press, 1984.

(本文责编: 瑞 源)