对大数据的再认识

李国杰
中国科学院计算技术研究所 北京 100190

摘要
大数据已成为媒体与大众关注的新技术，大数据的应用也预示着信息时代将进入一个新阶段，但人们对大数据的认识有一个不断加深的过程。首先从“信息时代新阶段”、数据文化和认识论的高度阐述了对大数据的理解；接着通过对驱动效益和大成智慧的解释，探讨了如何正确认识大数据的价值和效益，并从复杂性的角度分析了大数据研究和应用面临的挑战；最后对发展大数据应避免的误区提出几点看法。

关键词
大数据；认识论；大成智慧；复杂性

Further Understanding of Big Data

Li Guojie
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract
Big data has become a new technology, which has drawn much attention of media and public. Widely applications of big data indicated that the information age will enter into a new stage. However, the understanding of big data is a process of deepening. The big data from the height of “new information age stage”, data culture and epistemology was expounded. Then how to correctly understand the value and benefit of big data through the explanation of driving effect and wisdom in cyberspace was discussed. The challenges for the research and application of big data technology from the angle of the complexity were analyzed. Finally, some views on avoiding the pitfalls when developing big data technologies were proposed.

Key words
big data, epistemology, wisdom in cyberspace, complexity

1 大数据兴起预示“信息时代”进入新阶段

1.1 看待大数据要有历史性的眼光

信息时代是相对于农业和工业时代而言的一段相当长的时间。不同时代的生产要素和社会发展驱动力有明显差别。信息时代的标志性技术发明是数字计算机、集成电路、光纤通信和互联网（万维网）。尽管媒体上大量出现“大数据时代”的说法，但大数据、云计算等新技术目前还没有出现与上述划时代的技术发明可媲美的技术突破，难以构成一个超越信息时代的新时代。信息时代可以分成若干阶段，大数据等新技术的应用标志着信息社会将进入一个新阶段。

考察分析100年以上的历史长河可以发现，信息时代与工业时代的发展规律有许多相似之处。电气化时代与信息时代生产率的提高过程惊人地相似。都是经过20~30年扩散储备之后才有明显提高，分界线分别是1915年和1995年[1]。笔者猜想，信息技术经过几十年的扩散储备后，21世纪的前30年可能是信息技术提高生产率的黄金时期。

1.2 从“信息时代新阶段”的高度认识“大数据”

中国已开始进入信息时代，但许多人的思想还停留在工业时代。经济和科技工作中出现的许多问题，其根源是对时代的认识不到位。18-19世纪中国落后挨打，根源是满清政府没有认识到时代变了，我们不能重犯历史性的错误。

中央提出中国进入经济“新常态”以后，媒体上有很多讨论，但多数是为经济增速降低做解释，很少有从时代改变的角度论述“新常态”的文章。笔者认为，经济新常态意味着中国进入了以信息化带动新型工业化、城镇化和农业现代化的新阶段，是经济和社会管理的跃迁，不是权宜之计，更不是倒退。

大数据、移动互联网、社交网络、云计算、物联网等新一代信息技术构成的IT架构“第三平台”[1]是信息社会进入新阶段的标志，对整个经济的转型有引领和带动作用。媒体上经常出现的互联网+、创客、“第二次机器革命”、“工业4.0”等都与大数据和云计算有关。大数据和云计算是新常态下提高生产率的新杠杆，所谓创新驱动发展就是主要依靠信息技术促进生产率的提高。

1.3 大数据可能是中国信息产业从跟踪走向引领的突破口

中国的大数据企业已经有相当好的基础。全球十大互联网服务企业中国占有4席（阿里巴巴、腾讯、百度和京东），其他6个Top10 互联网服务企业全部是美国企业，欧洲和日本没有互联网企业进入Top10。这说明中国企业在基于大数据的互联网服务业务上已处于世界前列。在发展大数据技术上，我国有可能改变过去30年技术受制于人的局面，在大数据应用上中国有可能在全世界起到引领作用。

但是，企业的规模走在世界前列并不表示我国在大数据技术上领先。实际上，国际上目前流行的大数据主流技术没有一项是我国开创的。开源社区和众包是发展大数据技术和产业的重要途径，但我们对开源社区的贡献很小，在全球近万名社区核心志愿者中，我国可能不到200名。我们要吸取过去基础研究为企业提供核心技术不够的教训，加强大数据基础研究和前瞻技术研究，努力攻克大数据核心和关键技术。

1 第一平台是指集中式的大型主机，第二平台是指服务器/客户端应用模式的互联网平台，第三平台是指移动互联网、云计算、社交网络、大数据、物联网等构成的新一代IT架构。

2 理解大数据需要上升到文化和认识论的高度

2.1 数据文化是一种先进文化

数据文化的本质是尊重客观世界的实事求是精神，数据就是事实。重视数据就是强调用事实说话、按理性思维的科学精神。中国人的传统习惯是定性思维而不是定量思维。目前许多城市在开展政府数据开放共享工作，但是发现多数老百姓对政府要开放的数据并不感兴趣。要让大数据走上健康的发展轨道，首先要大力弘扬数据文化。本文讲的数据文化不只是大数据用于文艺、出版等文化产业，而是指全民的数据意识。全社会应认识到：信息化的核心是数据，只有政府和大众都关注数据时，才能真正理解信息化的实质；数据是一种新的生产要素，大数据的利用可以改变资本和土地等传统要素在经济中的权重。

有人将“上帝与数据共舞”归纳为美国文化的特点之一，说的是美国人既有对神的诚意，又有通过数据求真的理性。美国从镀金时代到进步主义时期完成了数据文化的思维转变，南北战争之后人口普查的方法被应用到很多领域，形成了数据预测分析的思维方式。近百年来美国和西方各国的现代化与数据文化的传播渗透有密切关系，我国要实现现代化也必须强调数据文化。

提高数据意识的关键是要理解大数据的战略意义。数据是与物质、能源一样重要的战略资源，数据的采集和分析涉及每一个行业，是带有全局性和战略性的技术。从硬技术到软技术的转变是当今全球性的技术发展趋势，而从数据中发现价值的技术正是最有活力的软技术，数据技术与数据产业的落后将使我们像错过工业革命机会一样延误一个时代。

2.2 理解大数据需要有正确的认识论

历史上科学研究是从逻辑演绎开始的，欧几里得几何的所有定理可从几条公理推导出来。从伽利略和牛顿开始，科学研究更加重视自然观察和实验观察，在观察基础上通过归纳方法提炼出科学理论，“科学始于观察”成为科学研究和认识论的主流。经验论和唯理论这两大流派都对科学的发展做出过重大贡献，但也暴露出明显的问题，甚至走入极端。理性主义走向极端就成为康德所批判的独断主义，经验主义走入极端就变成怀疑论和不可知论[2]。

20世纪30年代，德国哲学家波普尔提出了被后人称为“证伪主义”的认识论观点，他认为科学理论不能用归纳法证实，只能被试验发现的反例“证伪”，因而他否定科学始于观察，提出“科学始于问题”的著名观点[3]。证伪主义有其局限性，如果严格遵守证伪法则，万有引力定律、原子论等重要理论都可能被早期的所谓反例扼杀。但“科学始于问题”的观点对当前大数据技术的发展有指导意义。

大数据的兴起引发了新的科学研究模式：“科学始于数据”。从认识论的角度看，大数据分析方法与“科学始于观察”的经验论较为接近，但我们要牢记历史的教训，避免滑入否定理论作用的经验主义泥坑。在强调“相关性”的时候不要怀疑“因果性”的存在；在宣称大数据的客观性、中立性的时候，不要忘了不管数据的规模如何，大数据总会受制于自身的局限性和人的偏见。不要相信这样的预言：“采用大数据挖掘，你不需要对数据提出任何问题，数据就会自动产生知识”。面对像大海一样的巨量数据，从事数据挖掘的科技人员最大的困惑是，我们想捞的“针”是什么？这海里

究竟有没有“针”？也就是说，我们需要知道要解决的问题是什么。从这个意义上讲，“科学始于数据”与“科学始于问题”应有机地结合起来。

对“原因”的追求是科学发展的永恒动力。但是，原因是追求不完的，人类在有限的时间内不可能找到“终极真理”。在科学的探索途中，人们往往用“这是客观规律”解释世界，并不立即追问为什么有这样的客观规律。也就是说，传统科学并非只追寻因果性，也可以用客观规律作为结论。大数据研究的结果多半是一些新的知识或新的模型，这些知识和模型也可以用来预测未来，可以认为是一类局部性的客观规律。科学史上通过小数据模型发现一般性规律的例子不少，比如开普勒归纳的天体运动规律等；而大数据模型多半是发现一些特殊性的规律。物理学中的定律一般具有必然性，但大数据模型不一定具有必然性，也不一定具有可演绎性。大数据研究的对象往往是人的心理和社会，在知识阶梯上位于较高层，其自然边界是模糊的，但有更多的实践特征。大数据研究者更重视知行合一，相信实践论。大数据认识论有许多与传统认识论不同的特点，我们不能因其特点不同就否定大数据方法的科学性。大数据研究挑战了传统认识论对因果性的偏爱，用数据规律补充了单一的因果规律，实现了唯理论和经验论的数据化统一，一种全新的大数据认识论正在形成。

3 正确认识大数据的价值和效益

3.1 大数据的价值主要体现为它的驱动效应

人们总是期望从大数据中挖掘出意想不到的“大价值”。实际上大数据的价值主要体现在它的驱动效应，即带动有关的科研和产业发展，提高各行各业通过数据分析解决困难问题和增值的能力。大数据对经济的贡献并不完全反映在大数据公司的直接收入上，应考虑对其他行业效率和质量提高的贡献。大数据是典型的通用技术，理解通用技术要采用“蜜蜂模型”：蜜蜂的效益主要不是自己酿的蜂蜜，而是蜜蜂传粉对农业的贡献。

电子计算机的创始人之一冯·诺依曼曾指出：“在每一门科学中，当通过研究那些与终极目标相比颇为朴实的问题，发展出一些可以不断加以推广的方法时，这门学科就得到了巨大的进展。”我们不必天天期盼奇迹出现，多做一些“颇为朴实”的事情，实际的进步就在扎扎实实的努力之中。媒体喜欢宣传一些令人惊奇的大数据成功案例，对这些案例我们应保持清醒的头脑。据Intel中国研究院首席工程师吴甘沙在一次报告中透露，所谓“啤酒加尿布”的数据挖掘经典案例，其实是Teradata公司一位经理编出来的“故事”，历史上并没有发生过[4]。即使有这个案例，也不说明大数据分析本身有什么神奇，大数据中看起来毫不相关的两件事同时或相继出现的现象比比皆是，关键是人的分析推理找出为什么两件事物同时或相继出现，找对了理由才是新知识或新发现的规律，相关性本身并没有多大价值。

有一个家喻户晓的寓言可以从一个角度说明大数据的价值：一位老农民临终前告诉他的3个儿子，他在他家的地中埋藏了一罐金子，但没有讲埋在哪里。他的儿子们把他家所有的地都深挖了一遍，没有挖到金子，但由于深挖了土地，从此庄稼收成特别好。数据收集、分析的能力提高了，即使没有发现什么普适的规律或令人完全想不到的新知识，大数据的价值也已逐步体现。

3.2 大数据的力量来自“大成智慧”

每一种数据来源都有一定的局限性和片面性，只有融合、集成各方面的原始数据，才能反映事物的全貌。事物的本质和规律隐藏在各种原始数据的相互关联之中。不同的数据可能描述同一实体，但角度不同。对同一个问题，不同的数据能提供互补信息，可对问题有更深入的理解。因此在大数据分析中，汇集尽量多种来源的数据是关键。

数据科学是数学（统计、代数、拓扑等）、计算机科学、基础科学和各种应用科学融合的科学，类似钱学森先生提出的“大成智慧学”[5]。钱老指出：“必集大成，才能得智慧”。大数据能不能出智慧，关键在于对多种数据源的集成和融合。IEEE计算机学会最近发布了2014年的计算机技术发展趋势预测报告，重点强调“无缝智慧（seamless intelligence）”。发展大数据的目标就是要获得协同融合的“无缝智慧”。单靠一种数据源，即使数据规模很大，也可能出现“瞎子摸象”一样的片面性。数据的开放共享不是锦上添花的工作，而是决定大数据成败的必要前提。

大数据研究和应用要改变过去各部门和各学科相互分割、独立发展的传统思路，重点不是支持单项技术和单个方法的发展，而是强调不同部门、不同学科的协作。数据科学不是垂直的“烟囱”，而是像环境、能源科学一样的横向集成科学。

3.3 大数据远景灿烂，但近期不能期望太高

交流电问世时主要用作照明，根本想象不到今天无处不在的应用。大数据技术也一样，将来一定会产生许多现在想不到的应用。我们不必担心大数据的未来，但近期要非常务实地工作。人们往往对近期的发展估计过高，而对长期的发展估计不足。Gartner公司预测，大数据技术要在5~10年后才会成为较普遍采用的主流技术，对发展大数据技术要有足够的耐心。

大数据与其他信息技术一样，在一段时间内遵循指数发展规律。指数发展的特点是，从一段历史时期衡量（至少30年），前期发展比较慢，经过相当长时间（可能需要20年以上）的积累，会出现一个拐点，过了拐点以后，就会出现爆炸式的增长。但任何技术都不会永远保持“指数性”增长，一般而言，高技术发展遵循Gartner公司描述的技术成熟度曲线（hype cycle）[2]，最后可能进入良性发展的稳定状态或者走向消亡。

需要采用大数据技术来解决的问题往往都是十分复杂的问题，比如社会计算、生命科学、脑科学等，这些问题绝不是几代人的努力就可以解决的。宇宙经过百亿年的演化，才出现生物和人类，其复杂和巧妙堪称绝伦，不要指望在我们这一代人手中就能彻底揭开其奥妙。展望数百万年甚至更长远的未来，大数据技术只是科学技术发展长河中的一朵浪花，对10~20年大数据研究可能取得的科学成就不能抱有不切实际的幻想。

4 从复杂性的角度看大数据研究和应用面临的挑战

大数据技术和人类探索复杂性的努力有密切关系。20世纪70年代，新三论（耗散结构论、协同论、突变论）的兴起对几百年来贯穿科学技术研究的还原论发起了挑战。1984年盖尔曼等3位诺贝尔奖得主成立以研究复杂性为主的圣菲研究所，提出超

2
技术成熟度曲线是指新技术、新概念在媒体上曝光度随时间变化的曲线，反映新技术从炒作到跌入低谷再到正常发展的规律，Gartner 公司每年发布一次。

越还原论的口号，在科技界掀起了一场复杂性科学运动。虽然雷声很大，但30年来并未取得预期的效果，其原因之一可能是当时还没有出现解决复杂性的技术。

集成电路、计算机与通信技术的发展大大增强了人类研究和处理复杂问题的能力。大数据技术将复杂性科学的新思想发扬光大，可能使复杂性科学得以落地。复杂性科学是大数据技术的科学基础，大数据方法可以看作复杂性科学的技术实现。大数据方法为还原论与整体论的辩证统一提供了技术实现途径。大数据研究要从复杂性研究中吸取营养，从事数据科学研究的学者不但要了解20世纪的“新三论”，可能还要学习与超循环、混沌、分形和元胞自动机等理论有关的知识，扩大自己的视野，加深对大数据机理的理解。

大数据技术还不成熟，面对海量、异构、动态变化的数据，传统的数据处理和分析技术难以应对，现有的数据处理系统实现大数据应用的效率较低，成本和能耗较大，而且难以扩展。这些挑战大多来自数据本身的复杂性、计算的复杂性和信息系统的复杂性。

4.1 数据复杂性引起的挑战

图文检索、主题发现、语义分析、情感分析等数据分析工作十分困难，其原因是大数据涉及复杂的类型、复杂的结构和复杂的模式，数据本身具有很高的复杂性。目前，人们对大数据背后的物理意义缺乏理解，对数据之间的关联规律认识不足，对大数据的复杂性和计算复杂性的内在联系也缺乏深刻理解，领域知识的缺乏制约了人们对大数据模型的发现和高效计算方法的设计。形式化或定量化地描述大数据复杂性的本质特征及度量指标，需要深入研究数据复杂性的内在机理。人脑的复杂性主要体现在千万亿级的树突和轴突的链接，大数据的复杂性主要也体现在数据之间的相互关联。理解数据之间关联的奥秘可能是揭示微观到宏观“涌现”规律的突破口。大数据复杂性规律的研究有助于理解大数据复杂模式的本质特征和生成机理，从而简化大数据的表征，获取更好的知识抽象。为此，需要建立多模态关联关系下的数据分布理论和模型，理清数据复杂度和计算复杂度之间的内在联系，奠定大数据计算的理论基础。

4.2 计算复杂性引起的挑战

大数据计算不能像处理小样本数据集那样做全局数据的统计分析和迭代计算，在分析大数据时，需要重新审视和研究它的可计算性、计算复杂性和求解算法。大数据样本量巨大，内在关联密切而复杂，价值密度分布极不均衡，这些特征对建立大数据计算范式提出了挑战。对于PB级的数据，即使只有线性复杂性的计算也难以实现，而且，由于数据分布的稀疏性，可能做了许多无效计算。

传统的计算复杂度是指某个问题求解时需要的时间空间与问题规模的函数关系，所谓具有多项式复杂性的算法是指当问题的规模增大时，计算时间和空间的增长速度在可容忍的范围内。传统科学计算关注的重点是，针对给定规模的问题，如何“算得快”。而在大数据应用中，尤其是流式计算中，往往对数据处理和分析的时间、空间有明确限制，比如网络服务如果回应时间超过几秒甚至几毫秒，就会丢失许多用户。大数据应用本质上是在给定的时间、空间限制下，如何“算得多”。从“算得快”到“算得多”，考虑计算复杂性的思维逻辑有很大的转变。所谓“算得多”并不是计算的数据量越大越好，需要

探索从足够多的数据，到刚刚好的数据，再到有价值的数据的按需约简方法。

基于大数据求解困难问题的一条思路是放弃通用解，针对特殊的限制条件求具体问题的解。人类的认知问题一般都是NP难问题，但只要数据充分多，在限制条件下可以找到十分满意的解，近几年自动驾驶汽车取得重大进展就是很好的案例。为了降低计算量，需要研究基于自举和采样的局部计算和近似方法，提出不依赖于全量数据的新型算法理论，研究适应大数据的非确定性算法等理论。

4.3 系统复杂性引起的挑战

大数据对计算机系统的运行效率和能耗提出了苛刻要求，大数据处理系统的效能评价与优化问题具有挑战性，不但要求理清大数据的计算复杂性与系统效率、能耗间的关系，还要综合度量系统的吞吐率、并行处理能力、作业计算精度、作业单位能耗等多种效能因素。针对大数据的价值稀疏性和访问弱局部性的特点，需要研究大数据的分布式存储和处理架构。

大数据应用涉及几乎所有的领域，大数据的优势是能在长尾应用中发现稀疏而珍贵的价值，但一种优化的计算机系统结构很难适应各种不同的需求，碎片化的应用大大增加了信息系统的复杂性，像昆虫种类一样多（500多万种）的大数据和物联网应用如何形成手机一样的巨大市场，这就是所谓“昆虫纲悖论”[6]。为了化解计算机系统的复杂性，需要研究异构计算系统和可塑计算技术。

大数据应用中，计算机系统的负载发生了本质性变化，计算机系统结构需要革命性的重构。信息系统需要从数据围着处理器转改变为处理能力围着数据转，关注的重点不是数据加工，而是数据的搬运；系统结构设计的出发点要从重视单任务的完成时间转变到提高系统吞吐率和并行处理能力，并发执行的规模要提高到10亿级以上。构建以数据为中心的计算系统的基本思路是从根本上消除不必要的数据流动，必要的数据搬运也应由“大象搬木头”转变为“蚂蚁搬大米”。

5 发展大数据应避免的误区

5.1 不要一味追求“数据规模大”

大数据主要难点不是数据量大，而是数据类型多样、要求及时回应和原始数据真假难辨。现有数据库软件解决不了非结构化数据，要重视数据融合、数据格式的标准化和数据的互操作。采集的数据往往质量不高是大数据的特点之一，但尽可能提高原始数据的质量仍然值得重视。脑科学研究的最大问题就是采集的数据可信度差，基于可信度很差的数据难以分析出有价值的结果。

一味追求数据规模大不仅会造成浪费，而且效果未必很好。多个来源的小数据的集成融合可能挖掘出单一来源大数据得不到的大价值。应多在数据的融合技术上下功夫，重视数据的开放与共享。所谓数据规模大与应用领域有密切关系，有些领域几个PB的数据未必算大，有些领域可能几十TB已经是很大的规模。

发展大数据不能无止境地追求“更大、更多、更快”，要走低成本、低能耗、惠及大众、公正法治的良性发展道路，要像现在治理环境污染一样，及早关注大数据可能带来的“污染”和侵犯隐私等各种弊端。

5.2 不要“技术驱动”，要“应用为先”

新的信息技术层出不穷，信息领域不

断冒出新概念、新名词，估计继“大数据”以后，“认知计算”、“可穿戴设备”、“机器人”等新技术又会进入炒作高峰。我们习惯于跟随国外的热潮，往往不自觉地跟着技术潮流走，最容易走上“技术驱动”的道路。实际上发展信息技术的目的是为人服务，检验一切技术的唯一标准是应用。我国发展大数据产业一定要坚持“应用为先”的发展战略，坚持应用牵引的技术路线。技术有限，应用无限。各地发展云计算和大数据，一定要通过政策和各种措施调动应用部门和创新企业的积极性，通过跨界的组合创新开拓新的应用，从应用中找出路。

5.3 不能抛弃“小数据”方法

流行的“大数据”定义是：无法通过目前主流软件工具在合理时间内采集、存储、处理的数据集。这是用不能胜任的技术定义问题，可能导致认识的误区。按照这种定义，人们可能只会重视目前解决不了的问题，如同走路的人想踩着自己身前的影子。其实，目前各行各业碰到的数据处理多数还是“小数据”问题。我们应重视实际碰到的问题，不管是大数据还是小数据。

统计学家们花了200多年，总结出认知数据过程中的种种陷阱，这些陷阱不会随着数据量的增大而自动填平。大数据中有大量的小数据问题，大数据采集同样会犯小数据采集一样的统计偏差。Google公司的流感预测这两年失灵，就是由于搜索推荐等人为的干预造成统计误差。

大数据界流行一种看法：大数据不需要分析因果关系、不需要采样、不需要精确数据。这种观念不能绝对化，实际工作中要逻辑演绎和归纳相结合、白盒与黑盒研究相结合、大数据方法与小数据方法相结合。

5.4 要高度关注构建大数据平台的成本

目前全国各地都在建设大数据中心，吕梁山下都建立了容量达2 PB以上的数据处理中心，许多城市公安部门要求存储3个月以上的高清监控录像。这些系统的成本都非常高。数据挖掘的价值是用成本换来的，不能不计成本，盲目建设大数据系统。什么数据需要保存，要保存多少时间，应当根据可能的价值和所需的成本来决定。大数据系统技术还在研究之中，美国的E级超级计算机系统要求能耗降低1 000倍，计划到2024年才能研制出来，用现在的技术构建的巨型系统能耗极高。

我们不要攀比大数据系统的规模，而是要比实际应用效果，比完成同样的事消耗更少的资源和能量。先抓老百姓最需要的大数据应用，因地制宜发展大数据。发展大数据与实现信息化的策略一样：目标要远大、起步要精准、发展要快速。

参考文献

[1] Erik B, Andrew M. 第二次机器革命. 蒋永军译. 北京: 中信出版社, 2014
Erik B, Andrew M. The Second Machine Age. Translated by Jiang Y H. Beijing: Citic Press, 2014

[2] 黄欣荣. 大数据对科学认识论的发展. 自然辩证法研究, 2014, 30(9): 83~88
Huang X R. The development of traditional epistemology base on big data. Studies in Dialectics of Nature, 2014, 30(9): 83~88

[3] Karl R P. 猜想与反驳: 科学知识的增长. 傅季重, 纪树立, 周昌忠等译. 上海: 上海译文出版社, 2015
Karl R P. Conjectures and Refutations: the Growth Scientific Knowledge. Translated by Fu J Z, Ji S L, Zhou C Z, *et al*. Shanghai: Shanghai Translation

Publishing House, 2015

[4] 卢明森, 鲍世行. 钱学森论大成智慧. 北京: 清华大学出版社, 2014
Lu M S, Bao S X. Qian Xuesin's View on Wisdom in Cyberspase. Beijing: Tsinghua University Press, 2014

[5] 吴甘沙. 漫谈大数据的思想形成与价值维度. http://www.chinainfo100.net/document/201404/article12793.htm, 2014
Wu G S. Discussion on thought formation and value dimension of big data. http://www.chinainfo100.net/document/ 201404/article12793.htm, 2014

[6] 徐志伟, 李国杰. 普惠计算之十二要点. 集成技术, 2012, 1(1)
Xu Z W, Li G J. A dozen essential issues of computing for the masses. Journal of Integration Technology, 2012, 1(1)

作者简介

李国杰, 男, 博士, 中国工程院院士。现任中国科学院计算技术所首席科学家, 曙光信息产业股份有限公司董事长, 中国计算机学会名誉理事长, 国家信息化专家咨询委员会信息技术与新兴产业专委会副主任, 中国科学院学位委员会副主席, 中国科学院大学计算机与控制学院院长, 中国科学技术大学计算机科学与技术学院院长等。主要从事计算机体系结构、并行算法、人工智能、计算机网络等方面的研究, 发表论文100多篇, 合著英文专著4本, 出版了报告论文集《创新求索录》。先后获得国家科学技术进步奖一等奖、二等奖, 首届何梁何利基金科学与技术进步奖等奖项。

收稿日期: 2015-04-14; **修回日期**: 2015-05-07
论文引用格式: 李国杰. 对大数据的再认识. 大数据, 2015001
Li G J. Futher understanding of big data. Big Data Research, 2015001