基于大数据与系统思维来探讨现代计算理论与技术发展

摘要： 大数据泛指巨量的数据集, 因可从中挖掘出有价值的信息而受到重视。随着计算机软硬件的更新换代和网络技术的迅猛发展, 现代计算理论与技术真正步入了“大数据”时代。大数据时代发展下，人类的思维方式也将产生巨大的改变，因此我们必须从以往的小数据思维迅速转换成大数据思维，以适应这场急速的变革。大数据思维具有整体性、多样性、平等性、开放性、相关性和生长性等特征，从本质上来说它是一种复杂性思维，并且维正一步步在从逻辑判断发展为系统思考，大数据思维和系统思维获得了技术上的实现，因而影响更加巨大和深远。为了进一步提高现代计算理论与技术水平，还需要从当前存在的问题出发，采取有效的措施方法，推动现代计算理论与技术全面发展。下面本文主要结合基于大数据与系统思维，通过大数据与系统体系的分析，从数据科学、工业4.0以及信息物理系统的角度对于现代计算进行探讨。

关键词：大数据；大数据思维；系统思维；现代计算；信息处理；工业4.0

Summary : Big data refers to a large number of data sets, which can be used to mine valuable information. With the upgrading of computer hardware and software and the rapid development of network technology, modern computing theory and technology have really entered the era of ' big data '. With the development of the era of big data, the way of thinking of human beings will also have great changes, so we must quickly change from small data thinking to big data thinking to adapt to this rapid change. Big data thinking has the characteristics of integrity, diversity, equality, openness, relevance and growth. In essence, it is a kind of complex thinking, and it is developing from logical judgment to systematic thinking step by step. Big data thinking and systematic thinking have been realized technically, which has a greater and far-reaching impact. In order to further improve the level of modern computing theory and technology, it is necessary to take effective measures to promote the comprehensive development of modern computing theory and technology from the current problems. Based on big data and system thinking, this paper discusses modern computing from the perspective of data science, industry 4.0 and information physics system through the analysis of big data and system system.

Keywords : Big data ; big data thinking ; systematic thinking ; modern calculation ; information processing ; industry 4.0

1 引言

近些年,由于计算机、物联网等信息化技术以及传感技术的发展，使得现代生活中出现了“一切皆可数据化”的思维,数据的产生方式由“人机”、“机物”的二元世界向着融合社会资源、信息系统以及物理资源的三元世界转变，数据规模呈膨胀式发展。

例如，互联网领域中，谷歌搜索引擎的每秒使用用户量达到200万，Twitter每天的推特量已经超过了3. 4亿；科研领域中，仅某大型强子对撞机在一年内积累的新数据量就达到15 PB左右；电子商务领域中,作为世界连锁性企业沃尔玛，其每小时可处理的客户交易可超过100万笔，相应为数据库注入超过2. 5 PB的数据；航空航天领域中，仅一架双引擎波音737在横贯大陆飞行的过程中，传感器网络便会产生近240 TB的数据。

综合各个领域，目前积累的数据量已经从TB级上升至PB、EB甚至已经达到ZB级别，其数据规模已经远远超出了现有计算机所能够处理的量级，而且全球的数据量正以每18个月翻一倍的速度呈膨胀式增长。对此全球著名的管理咨询公司McKinsey首先提出了“大数据时代”的到来，其认为数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。

大数据时代的到来颠覆了工业界、学术界对传统数据的认知，同时也引起了数据获取、存储、分析、挖掘以及可视化等技术的变革。本文在具体介绍大数据和系统思维内涵的基础上归纳总结了大数据处理 (包含大数据采

集、存储以及挖掘等) 的技术体系, 并从数据科学、工业4. 0以及CPS的角度，对现代技术计算理论和计算发展进行探讨和展望。

2 大数据思维和系统思维概述

2.1 大数据和大数据思维概述

维基百科中指出，大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间限制的数据集。全球著名的管理咨询公司McKinsey则将数据规模超出传统数据库管理软件的获取、存储、管理以及分析能力的数据集称为大数据；研究机构Gartner将大数据归纳为需要新处理模式才能增强决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产；徐宗本院士则在第462次香山科学会议上的报告中,将大数据定义为“不能够集中存储、并且难以在可接受时间内分析处理,其中个体或部分数据呈现低价值性而数据整体呈现高价值的海量复杂数据集；

“大数据”并不等同于“大规模数据”， Viktor Mayer - Schönberger和Kenneth Cukier在《“大数据”时代》中提出：大数据应具有4V特性，即Volume（数据量大）、Velocity（数据处理速度快）、Variety（数据具有多样性）和Value（数据价值密度低）。

“大数据”并不是一个空的概念,其出现对应了数据产生方式的变革。如果从事件发生的三要素来看,需要具备时间、地点以及人物要求,事件才能完整。但是对于“大数据”而言,其产生方式已经分别在这三要素上突破了限制,即传统数据产生方式的变革导致了具有4V特性的“大数据”的出现。为此，从事件发生三要素的独特视角，清晰、全面地分析大数据产生的特点以及变化。大数据的产生如图1所示。

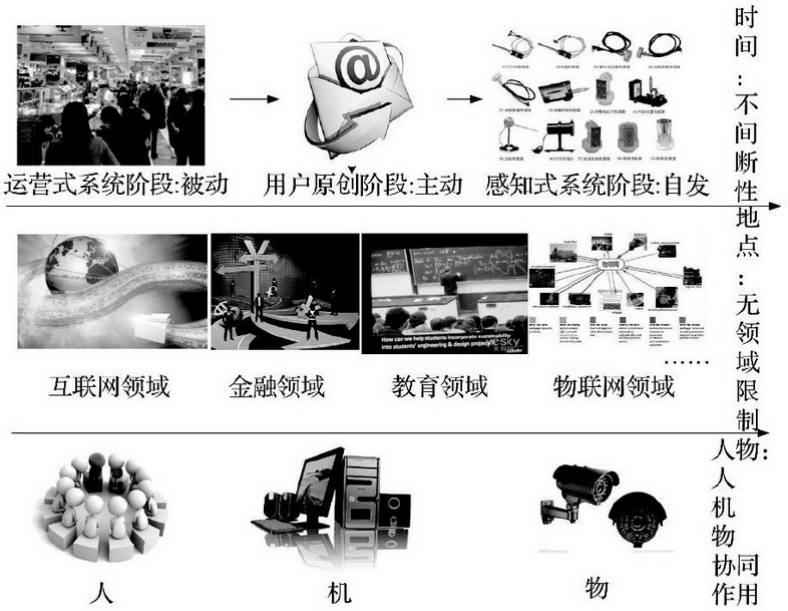

图1 大数据的产生

（1）时间: 不间断性。数据产生方式经历了被动、主动以及自发式的历程, 其已经脱离了对活动的依赖性,突破了传统时间的限制,具备了持续不间断产生的特性。

（2）地点: 无领域限制。大数据已经出现在各种领域,包括互联网、 金融、医疗、教育、科研、航空航天以及物联网等，甚至已经分布在了我们能够想象到的生产生活的各种领域。领域的扩展已经为“大数据”的形成提供了重要基础。

（3）人物: 人、机、物协同作用。众所周知,人物是传统事件发生的重要因素, 而对于数据的产生,其主体已经从传统的“人”的概念扩展到“人”、“机”、“物”以及三者的融合。随着云计算、物联网等信息技术的发展,“人”、

“机”及“物”的规模逐渐扩大,相互之间的作用越来越明显,数据的产生方式也已经由“人机”或“机物”的二元世界向着融合社会资源、信息系统以及物理资源的三元世界转变。

通过以上分析可知,数据产生的三要素已经发生了历史性的变革,人、机、物协同作用下,不间断、 无领域限制的数据产生方式已经突破了传统数据的概念,其必然导致数据性质的变革。

在大数据的时代背景下, 大数据思维孕育而生, 大数据正在改变我们的一切，其中最重要的是从改变我们的思维方式开始。简单来讲, 大数据思维就是一种收集处理大数据, 发现大数据的价值, 并应用大数据来看待问题解决问题的一种新的思维模式, 其有如下几个原理：1) 数据核心原理。从“流程”核心转变为“数据”核心。2) 数据价值原理。由功能是价值转变为数据是价值。3) 全样本原理。从抽样转变为需要全部数据样本。4) 关注效率原理。由关注精确度转变为关注效率。5) 关注相关性原理。由因果关系转变为关注相关性。6) 预测原理。从不能预测转变为可以预测。7) 信息找人原理。从人找信息, 转变为信息找人。

2.2 系统和系统思维概述

系统 (System) 一词来源于古希腊语, 其含义是“由部分组成的整体”。现代的定义是“由若干元素按一定关系组合、具有特定功能的有机整体, 其中元素又称为子系统”。科学的系统研究必须确定系统的元素, 划定系统的边界。一般的系统具有如下几个特征：1.整体性；2.层次性；3.关联性；4.功能性；5.有序性；6.平衡性；7.创新性；

系统思维就是在思考时, 把认识对象看成一个系统, 从系统内部的要素、要素和要素之间、系统和环境之间的相互联系和相互作用中综合地考察认识的一种思维方式。

系统思维一般有如下方法:

(1) 整体法。所谓整体法就是在分析分析和解决问题的时候, 眼光始终着眼于整体, 把整体放在第一位, 而不是让系统内部的元素结构活动凌驾于整体之上。

(2) 折叠结构法。折叠结构法就是系统思维时要求注意系统内部结构的合理性。

(3) 折叠要素法。要素法指的是对要素进行周全和充分的考察, 充分认识这些要素发挥的作用以确保系统内部的良好状态。

(4) 折叠功能法。功能法指的是为了使一个系统呈现出最佳的态势, 从大局出发来调整或是改变系统内部各部分的功能与作用。

2.3 大数据思维和系统思维的联系

(1) 整体性联系。前文提到, 大数据思维具有全数据取代随机抽样数据的原理。大数据时代追求的不是针对性的样本而是全部数据，而这所有数据正好刻画了研究对象的整体，整体性原理是系统论的一个根本原理，因此所谓“全体”其实就是系统科学中整体性的科学描述。

(2) 关联性联系。大数据分析强调相关关联，知道“是什么”;小数据挖掘获得因果关联, 了解“为什么”；经验来自数据因果的探究，是线性关联，而创新源于数据相关的发现，是非线性关联。用大数据探究系统的关联性，我们得到的是相关和因果并存的结论，因此二者在关联性方面可以联系起来。

(3) 动态性联系。大数据时代，信息更新速度快，数据分析和应用的时效性高。大数据失去时效性，就意味着失去价值。系统同样具有动态性特点：任何系统随时间变化而变化，系统与外部环境不断进行物质、能量、信息的交换。大数据分析的目的之一是帮助我们弄清系统动态演化的规律、方向和动力。

3 大数据与系统技术体系分析

大数据出现颠覆了传统数据计算处理的一系列技术，如大数据获取方式的改变导致数据规模迅速膨胀，相对于传统的数据库系统,其索引、查询以及存储都面临着严峻的考验,而且怎样快速地完成大数据的分析也是传统数据分析方法无法解决的。为此针对规模大、速度快、数据多样、价值密度低的大数据，本文将大数据系统处理技术体系总结如图2所示。

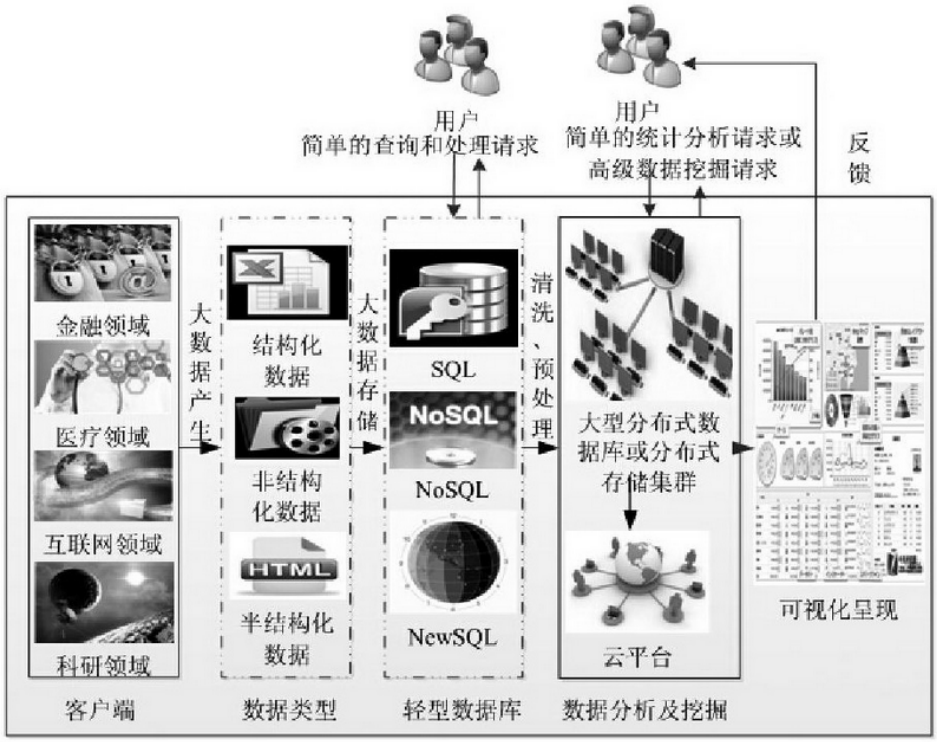

图2 大数据系统处理技术体系

如图2所示,大数据处理技术体系主要涉及大数据的采集技术、存储技术、分析及挖掘技术、可视化呈现技术4个部分。

1) 大数据的采集: 来自于不同领域的大数据, 其特点、数据量以及用户数目不同,按照结构特点, 可划分为3种类型: 结构化数据、半结构化数据以及非结构化数据。大数据采集的挑战是并发数高、 流式数据速度快。其特点如表1所示。

表1 大数据系统不同数据类型及特点分析

数据类型	举例	特点
结构化数据 (structured)	二维表	先有结构后有数据、行数据
半结构化数据[33] (semi-structured)	HTML 文档、XML 文档[34]、SGML 文档	先有数据后有模式、无规则性结构
非结构化数据 (unstructured)	图形、文本、声音、视频	模式具有多样性

2) 大数据的存储: 改进的轻型数据库可用于完成大数据的存储并响应用户的简单查询与处理请求，与此相关的大数据轻型数据库总结如表2所示。; 而当数据量超过轻型数据库的存储能力时,则需要借助于大型分布式数据库

或存储集群平台, 且随着互联网技术和云计算技术的发展,建立在分布式存储基础上的云存储已经成为大数据存储的主要趋势。大数据存储的主要挑战是数据异构、结构多样、规模大。与此相关的大数据轻型数据库总结如表2所示。

表2 大数据存储的轻型数据库

分类	举例		
	现属公司	数据库名称	主要特点
SQL	EMC	Greenplum[37]	关系型数据库集群
	HP	Vertica[38]	分布式 MPP 列式数据库 具有数据库内分析功能
	Teradata	Aster Data[39]	结合 SQL 与 MapReduce
NoSQL	Google	HBase[40]	分布式、面向列、开源
	10gen	MongoDB[41]	操作简单、完全免费、 源码公开、随时下载
	Facebook	Cassandra[42]	分布式网络服务、高扩展
	VMware	Redis	超高性能的键值数据库
NewSQL	Google	Spanner[43]	可扩展、全球分布式
	Google	Megastore[44]	融合 NoSQL 的可扩展性 和传统的关系型数据库
	Google	F1[45]	动态扩展、并行 SQL 执行引擎

3) 大数据的分析及挖掘 : 大数据的分析涉及简单的统计分析以及分类汇总,其挑战在于导入数据量大,查询请求多; 而大数据挖掘涉及数据的分类、聚类、频繁项挖掘等,其算法复杂, 计算量大。

4) 大数据可视化: 大数据的挖掘及分析结果将在显示终端以友好、形象、易于理解的形式呈现以供专业人士分析结果的准确性或为用户提供决策信息支持。大数据呈现的挑战在于数据维度高、呈现需求多样化。

大数据系统处理环节中各技术功能的相互配合使用可为大数据价值的有效实现提供技术基础。在此过程中，现代计算的平台也在逐步演化，最早的计算资源是只能由专业人员使用的大型机,之后发展成个人电脑走进千家万户,现为了满足海量数据运算的需要,这些小型的服务器又通过网络搭建集群提供更强大的计算资源,且为了方便管理、部署及提高资源使用率,虚拟化技术应运而生。最终所有的IT资源都会迁移到"云"中。其现代计算平台演变如图3所示。

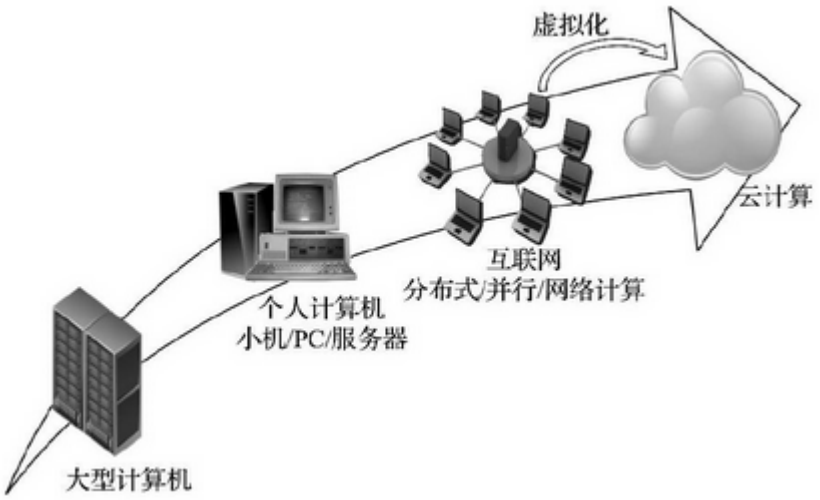

图3 现代计算平台的演化

云计算在未来将成为重要的计算模式，其将根据需求实现资源的有效分配以及应用。

4 现代计算发展趋势及挑战

大数据的出现以及其相关技术在近几年的迅速突破使得大数据在改变人类生产生活方式中逐渐承担重要角色，美国政府甚至将其称为“未来的石油”，可见大数据的重要性。目前大数据的膨胀式发展已经改变了人类的思维方式，“一切皆可数据化”的思维已经出现，并且必然会在以后的科学研究中占据主导地位。同时大数据在人类生产方式上的应用将会加速工业4. 0的到来，而大数据在人类生活方式上的应用也会助阵CPS系统价值的展现。为此,本部分将从改变思维方式、改变生产方式以及生活方式的三个角度阐述大数据的发展趋势，并分析其发展所面临的挑战。

4.1 现代计算发展趋势

4.1.1 改变思维方式

在2007年，图灵奖的获得者Jim Gary提出了科学的第四范式———“数据密集型科学”，之前的三种科学范式分别为实验科学、理论科学以及计算科学，第四范式的提出标志着数据对于科学研究的重要性的提升，其实质是科学研究将从以计算为中心向以数据为中心转变，即数据思维的到来。

“数据密集型科学”一经提出就得到了领域内研究学者的广泛关注,如微软在2009年10月发布了《e-science科学研究的第四种范式》，其对Jim Gary的观点进行了应用扩展,首次全面描述了快速兴起的数据密集型科学的研究，并将一个完整的科学研究分为四个部分，分别是数据收集、数据整理、数据分析以及数据可视化，其强调大量收集的数据需要有效分析才能实现其价值，e-science提供了一种新的科学思维，即各种工具的使用都应用于解决科学研究中海量数据问题，由此可见大数据的发展已改变了科学研究的思维方式，为了促进大数据的认知以及发展，微软研究院已经于2012年10月23日发布《第四范式: 数据密集型的科学发现》 中文版。

大数据的发展不仅改变了科学思维，也必然会引起企业以及政府、个人的思维方式的变革，维克托·尔耶·舍恩伯格在《大数据时代: 生活、工作与思维的大变革》中指出对于大数据时代，应放弃对因果关系的渴求，而更关注相关关系，正如其在福布斯·静安南京路论坛上的演讲所述: “在大数据时代,人们每天醒来，要想的事情就是这么多大数据可以用来做什么，其价值可以体现在哪些方面，而且是否可以找到一个别人从未涉及的事情使得思路以及想法成为重要的资产”。由此可见，大数据时代必然会引起思维的转变，而且思维的转变越快，越能在如今竞争激烈的社会中抢占先机。

4.1.2 改变生存方式

在21世纪,信息技术突飞猛进的今天，物联网、嵌入式技术、传感技术等的发展，为人类更全面地感知客观存在的物理世界提供了基础；而互联网、云计算等信息技术的发展更是改变了人类通信与管理信息的方式。随着技术的发展以及工具的更新换代，人类也提出了更高的生存需求，美国国家科学基金委员会在2006年提出了CPS的概念；2007年，不同机构及研究学者对其进行了定义，包括LEE,Baheli,Sastry以及Krogh等，强调计算元素以及物理元素，实体与虚拟网络的关系，并注重通信、计算以及控制能力，尽管不同定义的描述不同但是都明确了CPS的内涵： Cyber与Physical的深度融合后形成的智能系统。CPS的含义如图4所示：

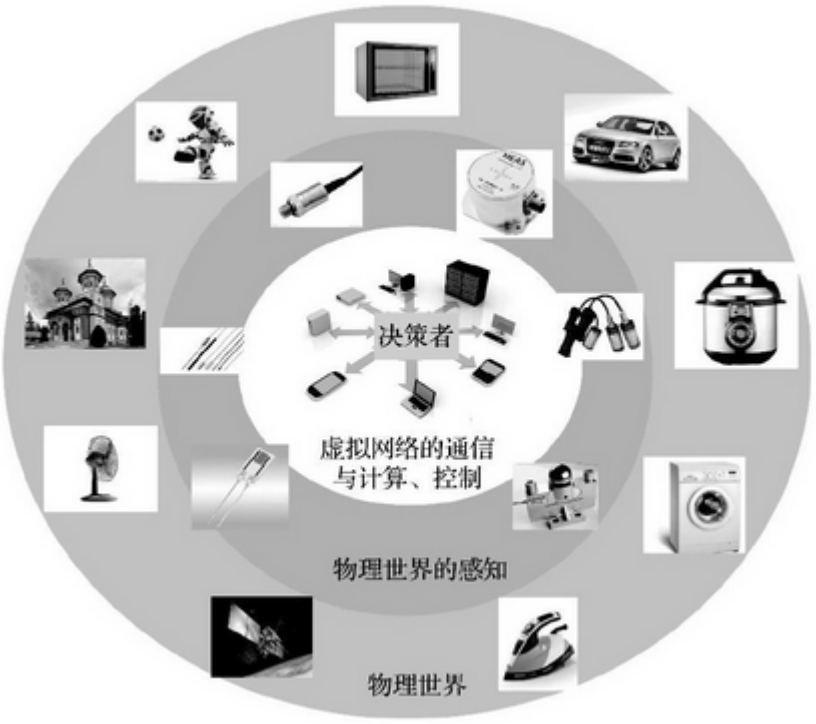

图4 信息物理系统运行

最外面一层是物理实体,其代表我们生活的物理世界; 中间一层为感知层,包括了传感器等具有采集功能的设备; 第三层为计算机等具有计算功能的设备,其负责实现对采集数据的分析以及可视化呈现; 最里面一层为决策层(具有决策能力的人或者其他事物) ,其通过感知以及分析结果做出决策,并作用于物理实体。CPS的运行图体现了在感知基础上,“人”、“机”、“物”的深层融合。可应用于无人机、自主导航的汽车等以实现物理实体的自主工作,医疗领域中可应用于自动手术,物联网领域中可实现生活中的智能家居以及智慧城市等。上述CPS的成功实现,最重要的基础就是系统中收集的大量数据的有效分析以及处理,其是决策支持的重要来源。由此可见,大数据的产生以及有效分析是CPS的重要资源和基础,结合其他技术的发展,将为改变人类生存方式提供重要动力。

4.1.3 改变生产方式

1950年至今,电子技术和计算机等信息技术的发展开创了“信息时代”,使得产品更为丰富,功能性更强; 而随着科技的进一步发展,科技的进步也必定引起生产方式的变革。为此,德国提出了 “工业4. 0”,即第四次工业革命,以智能制造为主导实现生产制造人机一体化,“工业4. 0”的提出预示着革命性的生产方式的诞生,而实现“工业4. 0” 的基础就是大数据的分析以及CPS的推广,其标志着生产制造业必须转向以数据分析为中心。由此可见,大数据的发展将在生产方式改变中起到关键作用。四次工业革命演化如图5所示。

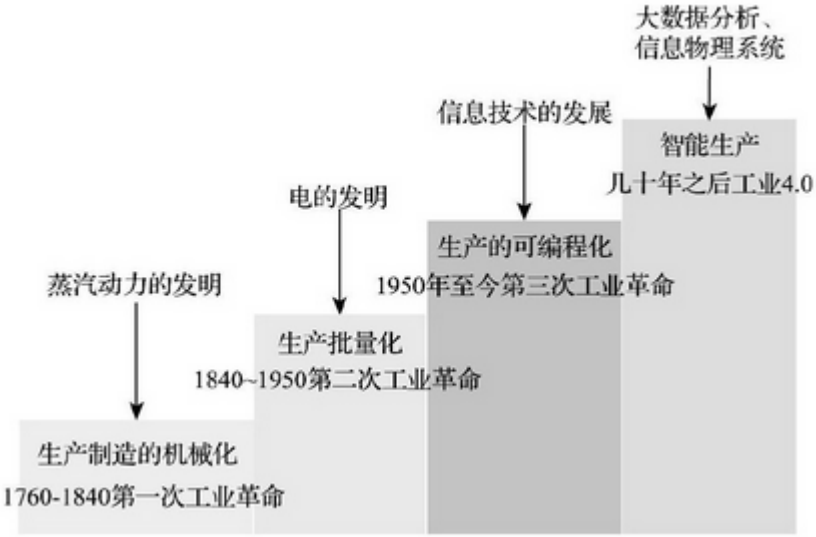

图5 四次工业革命演化

“工业4. 0”将要达到的目标是通过物联网系统实现智能工厂,即每件产品、零部件都会包含大量的信息，包括何时生产、可以用多久、是否需要替换等，通过非人为干预的智能方式实现自主处理。 由此可见，大数据将在改变生产方式中扮演重要角色，由大数据到决策的实现将加速工业4. 0时代的到来。

4.2 现代计算发展的挑战

大数据其规模大、速度快以及结构多样的特点，为传统数据的分析、存储以及管理技术带来的挑战不言而喻。对大数据处理流程的挑战总结如表3所示。

表3 大数据发展的挑战

研究主题	挑战
大数据预处理及集成	广泛的异构性、时空特性、数据质量
大数据分析	先有数据后有模式、动态增长、先验知识的缺乏、实时性
大数据硬件处理平台	硬件异构性、新硬件
性能测试基准	系统复杂性高、案例多样性、数据规模庞大、系统的快速演变
隐私保护	隐性数据的暴露、数据公开与保护、数据动态性
大数据管理的易用性	可视化、人机交互、数据起源技术、海量元数据的高效管理
大数据的能耗	低功耗、新能源

大数据的规模大，其质量影响算法的效率以及精度，大数据预处理作为数据分析的第一步，至关重要。而大数据来源的多样性，使得数据具有广泛的异构性、时空特性等，其为大数据预处理及集成带来严峻的考验；大数

据规模动态增长使得大数据的模式获取困难，加之先验知识的缺乏，如何在规定的时间内返回有价值的分析结果也是研究学者设计算法时不得不考虑的问题；这种需求也给大数据的计算系统提出了挑战。以数据为中心的系统结构要消除不必要的数据存放、通信和计算。但是与之相对应的系统并未完全实现这一思路，所以这也是大数据计算系统未来所必须解决的问题。同时，作为大数据处理的支撑技术，包括隐私保护、硬件平台以及大数据管理、能耗等也有很多难题需要突破。大数据发展的挑战也为大数据的发展指明了方向，需要大数据相关工作者突破领域限制，共同努力。

5 结论

本文在大数据和系统思维的基础上，对于大数据处理中各个环节涉及的关键技术进行了归纳与分析，并从数据科学、工业4. 0以及信息物理系统的角度, 展望了大数据的发展对于改变人类思维方式、生产方式以及生活方式的重要作用。研究学者应在 “大数据”大热的趋势下，冷静分析，按照自身定位以及需求，定义科学问题，以切入切实可行的研究方向，并通过把握大数据处理流程中的关键技术, 建立持续的研究体系，以把握大数据这一发展机遇,充分利用大数据创造大价值。

参考文献

[1]程学旗,靳小龙,王元卓,郭嘉丰,张铁赢,李国杰.大数据系统和分析技术综述[J].软件学报,2014,25(09):1889-1908.DOI:10.13328/j.cnki.jos.004674.

[2]彭宇,庞景月,刘大同,彭喜元.大数据:内涵、技术体系与展望[J].电子测量与仪器学报,2015,29(04):469-482.DOI:10.13382/j.jemi.2015.04.001.

[3] 邬贺铨.大数据时代的机遇与挑战[J].求是,2013(04):47-49.

[4] 黄欣荣.大数据时代的思维变革[J].重庆理工大学学报(社会科学),2014,28(05):13-18.

[5] 韩海庭,吴晖,孙圣力,等. 现代计算理论在征信领域的应用研究[J]. 征信,2020,38(6):14-21. DOI:10.3969/j.issn.1674-747X.2020.06.004.

[6] 王建民.工业大数据技术综述[J].大数据,2017,3(06):3-14.

[7] 王彦夫.大数据思维和系统思维[J].信息系统工程,2017(09):137+139.

[8] 何文韬,邵诚.工业大数据分析技术的发展及其面临的挑战[J].信息与控制,2018,47(04):398-410.DOI:10.13976/j.cnki.xk.2018.8085.

[9] 张维明,唐九阳.大数据思维[J].指挥信息系统与技术,2015,6(02):1-4.DOI:10.15908/j.cnki.cist.2015.02.001.

[10] 孙海燕,冶栋玉,曹娟.浅谈“大数据”时代背景下计算机信息处理技术[J].信息系统工程,2016(04):112.

[11] 武延军.大数据时代已经来临——人机物融合的大数据时代[J].高科技与产业化,2013(05):46-49.

[] Hansson, S.O. Technology and Mathematics.Philos. Technol. **33,** 117–139 (2020). https://doi.org/10.1007/s13347-019-00348-9

[] Tucker, Allen and Belford, Geneva G.. "computer science". Encyclopedia Britannica, Invalid Date, https://www.britannica.com/science/computer-science. Accessed 21 December 2021.