

2014年大数据发展趋势预测

关键词：调研方法 发展趋势 对比解读 应用领域

CCF大数据专家委员会

在2013年12月5日举行的2013中国大数据技术大会上，CCF大数据专家委员会发布了《2014年大数据发展趋势预测》报告。报告一经发布就受到各界的广泛关注，在微博、微信和各网站上被迅速传播和解读。为了让广大读者更清晰、更全面地了解此报告的内容与思想，CCF大数据专家委员会在本刊发布独家的官方解读。

报告的调研方法

2013年度的调研方法和2012年类似，首先向大数据专家委员会的全体委员征集趋势选项。报告工作组在征集到的选项基础上，对2012年相关候选项以及CCF大数据专家委员会发布的《中国大数据技术与产业发展白皮书（2013年）》¹中有关发展趋势的部分内容进行整理，形成了57个候选项。然后，大数据专家委员会全体委员对候选项进行投票。参与候选项提议和投票活动的大数据专家委员会正式委员共111位（包括来自学术界66位、产业界28位、政府部门4位、海外

人士13位），还有11位来自数学、物理、金融、管理等领域的特邀委员。本次调研报告重点关注对2014年中国大数据发展趋势的预测，具体包括四个分项：2014年度十大发展趋势、最受瞩目的应用领域、最受关注的学科技术与最有可能取得突破的大数据技术。

2014年十大发展趋势

1. 大数据从“概念”走向“价值”

作为第一大趋势，共得到93票，涉及3个候选项，即大数据向更多新领域扩张（67票）、更多数据驱动的商业模式（13票）、资本高度关注（13票）。前一选项是指会有更多领域开始有意识地收集和积累大数据，并从中挖掘以前不曾触及的新价值，从而使该领域产生重大变化；中间选项是指除了广告和电商这两种已有的数据驱动应用的商业模式之外，互联网、金融等将会成为大数据应用的新型商业模式，特别是基于海量数据

的信用体系和风险控制；后一选项是指资本将高度关注大数据领域，相关的融资、并购、首次公开募股（IPO）将会纷纷出现。

从2012年开始，大数据就成为一个社会各界都高度关注的热词，很多应用领域都在讲大数据的概念，比如金融大数据、安全大数据、制造业大数据、物流大数据等等。与2012年不同的是，有更多的领域关注大数据，而且一些领域已经开始实实在在地落实大数据的价值，开始探索大数据驱动的业务模式。

2. 大数据处理架构的多样化模式并存

这一趋势共得到71票，由3个候选项融合而来。其中，大数据处理多样化模式并存融合（34票），意味着Hadoop/MapReduce框架一统天下的模式将被打破，实时流计算、分布式内存计算、图计算框架等将与之并存；大数据存储与管理（23票）意味着尽管分布式文件系统、数据索引与查询技术、查询语言、实时/流式数据存储管理

¹ <http://www.ccf.org.cn/sites/ccf/xhdtory.jsp?contentId=2774966444490>。

等都不是新问题,但大数据体量大、速度快、模态多等特性带来的挑战终将引起大数据存储与管理的质变;内存计算成为提高处理性能的主要手段之一(14票),这意味着大数据的实时处理需求迫切。同时,内存价格将不断降低,内存计算将成为解决实时性大数据处理问题的主要手段。

大数据的应用需求多种多样,而相关技术还远远没有达到成熟和稳定的程度,在2014年乃至之后的几年内,不太可能存在可以“一统天下”的大数据系统架构。我们会持续看到更多丰富多彩的、创新引领突破的技术架构与应用模式。

3.大数据安全与隐私

大数据安全和隐私越来越重要,得到53票。其中主要候选项包括大数据安全(33票)和大数据个人数据保护和隐私立法(20票)。相对于传统的数据模式,大数据更容易成为网络攻击的目标,大数据分析技术也更容易被黑客利用。如今的网络和数字化生活使得犯罪分子、网络黑客以及并无恶意的大数据服务提供商可轻易获得用户的有关信息,而且犯罪分子的犯罪手段也变得更多、更高明且更加不易被追踪和防范。同时,用户的隐私会越来越多地融入到各种大数据中,而各种数据来源之间的无缝对接以及越来越精准的数据挖掘技术,使得大数据拥有者

能够掌控越来越多的用户和越来越丰富的信息。在挖掘这些数据价值的同时,隐私泄漏存在巨大风险。由于系统故障、黑客入侵、内部泄密等原因,数据泄漏随时可能发生,从而造成难以预估的损失。因此,大数据时代,因数据而产生的安全保障问题、隐私问题非常严峻。目前尚没有较好的解决办法。今后,在安全和隐私方面的投资会增大,而有偿的隐私保护服务有可能会被大众所接受。

4.大数据分析可视化成为热点

在大数据务实发展的同时,人们对大数据发展趋势的思考也越来越具体。大数据分析及其可视化技术将成为热点,共得到61票,其中大数据可视化33票,实时分析处理集成14票,大数据分析的革命性方法14票。大数据带来新一轮的技术革命,对于大数据查询和分析的实用性和实效性,人们能否及时获得决策信息变得非常重要,决定着大数据应用的成败。而基于大内存的计算模式或将成为大数据实时处理的重要手段。同时,针对大数据计算的挑战,面向大数据的数据挖掘、机器学习和人工智能非常需要并很可能会产生新的理论突破。

尤其值得关注的是,大数据的复杂性决定了数据可视化技术将成为大数据时代的显学。对大数据

进行分析后,为了方便用户理解,需要有效的可视化技术,其中交互式的展示和超大图的动态化展示值得重点关注。如果一个机构要尝试迈向大数据模式,那么一定要有相当分量的数据可视化投入。

5.大数据产业成为战略性新兴产业

大数据产业成为战略性新兴产业,这一趋势得到了大数据专家委员会委员的广泛认可,共得71票,由3个候选项构成。其中,大数据产业成为国家战略新兴产业(31票)是指,各个国家都将推动大数据的战略性新兴产业,将利用大数据来提高国家的经济决策和社会服务能力,保障国家安全;大企业大数据(27票)选项预示着大企业将成为大数据技术研发与应用最为活跃的群体;政府大数据(20票)选项则预示着一些拥有大数据的政府部门将纷纷利用积累的数据,采用大数据技术进行分析,从而产生突出的社会经济效益。

从产业拓展的角度看,大数据是继云计算、物联网之后的一个巨大的新产业领域。而且其所蕴含的机会和挑战将大大多于云计算和物联网。大数据产业(数据产业)具有很强的蜂箱效应²,除了产业自身的经济蕴藏量之外,还将大大撬动其他产业的跨越升级。

6.数据的商品化和数据

² 蜂箱效应:在农田果园边上的蜂箱,不能仅仅用蜂箱产的蜜来衡量其价值,还要看到蜜蜂在田园中授粉为农作物增产带来的决定性作用。

共享联盟化

这一趋势涵盖了数据资源化、私有化、进而商品化(28票)和数据共享联盟(31票)两个候选项,总计59票。一方面,数据既然是重要的资源,一定具有其特定的价值属性,因而数据私有化和数据的独占问题将是客观存在的现实。因此,一部分专家认为,如何界定数据的权属,以及如何让数据在有数据权属保护的情况下进行商品化交易是未来一个重要的发展趋势。

另一方面,数据共享联盟是大数据专家委员会2012年讨论的议题。然而,由于数据共享联盟的生态环境尚未建立,在2013年进展缓慢。但是,大数据专家委员会依然希望数据共享联盟能够逐步壮大,成为产业和学术环环相扣的支撑环节以及产业发展的核心环节。

7. 基于大数据的推荐和预测将逐步流行

很多大数据科普书中,都对大数据进行了推荐和预测。大数据专家委员会认为,中国大数据科研和产业在2014年将开始真正落到实处,具体体现是,基于大数据的推荐和预测将逐步在很多领域发挥实际作用。这一趋势得到76票,包括3个候选项:大数据预测和决策支持(31票),是指通过对海量历史数据的分析,帮助捕捉现在和预测未来,进而通过大数据进行决策支持;精确个性化推荐系统(27票),是指依靠技术革新,推荐系统的推荐结果将更加精确,更加

了解用户,更加个性化,从而成为能够带来效益的大数据商业应用;基于社交网络的广告营销模式会初现端倪(18票)。在积累了多年的用户和数据基础后,基于社交网络的广告营销模式具备了爆发商业应用的机会。

8. 深度学习与大数据智能成为支撑

要挖掘大数据背后的大价值,最核心的支撑技术是什么?大数据专家委员会预测深度学习和大数据智能(总计56票)将是此问题的答案。深度学习成为大数据智能分析的核心技术(28票),预示着在大数据时代,依靠高效能计算的支持,深度学习有望成为大数据智能处理的核心技术之一;基于海量知识的智能(18票),意味着将要兴起的下一个浪潮——智能机器会依赖于对捕捉到的数据进行分析来做判断和决策;利用群体智慧与众包计算方式将使大数据智能成为可行的技术(10票)。总体而言,面对大数据,基于设备、人力和知识的分布式计算将会推动深度学习、数据智能等技术成为数据价值利用的支撑手段。

9. 数据科学的兴起

该项趋势获得47票,包括3个选项:与大数据相关的新学科,即数据科学兴起(27票),出现新型的大数据系统评测基准(10票),各类学科的交叉以及传统学科对数据的广泛依赖将迅速产生数据科学学科(10票)。同时,类似波色

子的发现,数学、生物、物理、化学、材料等领域在一定程度上依赖数据科学才取得了突破性进展。

应当看到,数据科学作为一门科学,还有很多根本问题没有解决,甚至还有很多问题没有被提出。数据科学真正兴起并成为一支柱学科,需要学术界更多的努力。

10. 大数据生态环境逐步完善

虽然大数据的良性生态环境目前还没有形成,但它正在逐步的完善过程中,这一趋势获得61票。具体包括开源成为主流(23票),大数据将与云计算、物联网、移动互联网等热点新兴计算相互交融(20票),大数据教育(18票)。

发展趋势对比解读

对比2013年度和2014年度预测,可以看到大数据发展越来越务实。如表1所示。

在核心技术方面,从2013年度比较笼统的基于大数据的智能和革命性方法,变为2014年度4个指向性非常明确的技术趋势预测,即有别于Hadoop的多模式架构并存、大数据可视化、推荐和预测、深度学习。

在技术生态方面,大数据安全和隐私问题依然得到高度关注,而数据科学的兴起依然是大数据专家委员会委员们的期望和关注点。

在产业生态方面,从2013年度笼统的“更大的数据”变为2014年度着重关注大数据的价值和应

表1 2013年度与2014年度十大趋势预测对比

	2013年度预测	2014年度预测
核心技术方面	4.基于大数据的智能化的出现 5.大数据分析的革命性方法	2.大数据架构的多样化模式并存 4.大数据分析可视化 7.基于大数据的推荐与预测流行 8.深度学习与大数据智能成为支撑
技术生态方面	2.大数据的隐私问题突出 6.大数据安全 7.数据科学兴起	3.大数据安全与隐私 9.数据科学的兴起
产业生态方面	1.数据的资源化 3.大数据与云计算等深度融合 8.数据共享联盟 9.大数据新职业 10.更大的数据	1.大数据从“概念”走向“价值” 5.大数据产业成为战略性新兴产业 + 6.数据商品化与数据共享联盟化 10.大数据生态环境逐步完善

用。价值和应用自然也会带来战略性新兴产业地位。数据资源化和权属问题依然受到关注，并在数据共享联盟问题上成为矛盾焦点。大数据与其他新兴IT技术和现象的密切相关得到广泛认可。而大数据相关的新职业则从预测项中消失，这反映出这个职业的独特性还有待确认。

最受瞩目的应用领域

在该分项调研中，90张选票中有80张关注了应用领域，可见专家们对应用领域的重视程度。候选的应用领域包括互联网与电子商务、金融、健康医疗、生物信息与制药、城镇化与智慧城市、社会安全和犯罪侦查、网络空间安全和对抗、商情与舆情分析、网络通信、制造业、工业大数据、企业管理大数据、能源领域大数据、气候与环境模拟、经济统计和模拟、教育领域大数据、前沿基础科学

研究（如高能物理、生物、天文、海洋、地震等）等。

其中互联网与电子商务得60票，金融得34票，健康医疗得31票，商情与舆情分析得23票。结合57个候选应用领域的得票数，我们汇总出2014年最令人瞩目的六大应用领域：网络大数据、金融大数据、健康医疗大数据、企业大数据、政府管理大数据和安全大数据。

从投票情况来看，大数据在基础自然科学研究领域、超复杂系统（气候、经济）分析领域、工业制造和管理等方面的应用关注度相对不高。原因有两个方面：一方面是大数据专家委员会的委员大部分来自IT界，大家相对更加关心像互联网大数据、电子商务大数据等IT领域的应用；另一方面，在制造业、工业能源等领域，大数据产学研用的活跃度要比网络大数据、金融大数据等滞后一些。

最受瞩目的学科和技术

设立该分项的主要目的是调研学术界和产业界对大数据学科及其关键技术的关注范畴及其热度。结果汇总如下：

- 大数据学习与挖掘（75票）。其中，深度学习32票，机器学习22票，数据挖掘21票；
- 大数据计算架构和系统（44票）。其中，分布式计算23票，内存计算21票；
- 大数据分析可视化（47票）。其中，自然语言、知识工程13票，可视化13票，复杂网络12票，图计算9票。

最先取得突破的技术环节

这一问题共收到61张选票。排名前三位的2014年最有可能取得突破的技术是：大数据挖掘与关联分析（28票）、存储结构和系统（20票）、数据采集和数据化（19票）。

CCF大数据专家委员会发布的《2014年大数据发展趋势预测》从筹备调研到形成报告，花了将近3个月时间。与2012年相比，本次调研参与人员更多、选项更多，对报告结果的梳理也更有体系，解读更加具体。但即便如此，该报告对发展趋势的把握与解读难免有不到位甚至是有偏差的地方。欢迎读者向CCF大数据专家委员会提出宝贵的建议和意见。■

（程学旗 潘柱廷 靳小龙 执笔）

附表：57个候选项的原始得票（总票数90票）

大数据整体态势	
T1 大数据向更多新领域扩张(67)	T2 数据资源化、私有化，进而商品化(28)
T3 2014年与大数据相关的最令人瞩目的应用领域： 互联网、电子商务(60)，金融（股市预测、金融分析等）(34)，健康医疗（流行病监控和预测等）(31)，舆情分析(23)，城镇化、智慧城市(17)，社会安全和犯罪侦查(13)，网络空间安全和对抗(11)，企业管理大数据(9)	大数据用于教育(6)，前沿基础科学研究（如高能物理、生物、天文、海洋、地震等）(5)，生物信息、制药(5)，制造业、工业大数据(4)，网络通信(3)，能源（石油、电力等）(1)，气候/环境模拟(1)，经济统计和模拟(0)
大数据应用	
T4 大数据预测和决策支持(31) T5 大企业大数据(27) T6 政府大数据(20) T7 大数据加强军队(6) T8 犯罪侦破与预防(13)	T9 MOOC和教育大数据(18) T10 精确个性化推荐系统(27) T11 基于社交网络的广告营销模式会初现端倪(18) T12 大数据生态系统(9) T13 分析平民化(10)
T14 最令人瞩目的大数据相关学科和技术： 深度学习(32)，数据挖掘(21)，分布式计算、MapReduce、Hadoop等技术(23)，内存计算、实时分析系统和技术(21)，机器学习(22)，自然语言处理、知识工程、语义搜索和智能问答(13)，可视化(13)，复杂网络(12)，图计算(9)，人机协同计算(8)，流计算(7)，索引和查询技术(7)，高性能计算(2)，统计学(2)，高性能网络在大数据领域的应用(1)，迁移学习(1)	T15 最可能先取得重大突破的技术环节： 分析：挖掘、关联等（类化学式，产生新的数据形态）(28)，存储结构和系统(20)，采集和数据化(19)，处理：筛选、采样、压缩、清洗等（类物理式，保持数据本身形态）(10)，数据语义理解(10)，可视化(9)，人机交互和穿戴类传感器(8)，计算系统(6)，大数据安全、攻击和防御(4)，数据科学基础理论(4)，传输系统及其效能(3)，脱敏和隐私保护(0)
大数据学术与技术	
T16 数据科学兴起(27) T17 所有的科学正迅速变成数据科学(10) T18 数据质量和数据清洁(14) T19 更复杂异构多样的数据(15) T20 基于海量知识的智能(18) T21 大数据分析的革命性方法(14) T22 大数据处理多样化模式并存融合(34) T23 内存计算成为提高处理性能的主要手段(14) T24 网络传输压力突出(9) T25 众包成为支撑技术(13) T26 数据的语义化和知识化(13)	T27 大数据专用软硬件系统(11) T28 大数据可视化(33) T29 实时分析处理集成(14) T30 自组织计算(3) T31 机器学习的范式发生变化(2) T32 语音会获得爆发式应用(4) T33 深度学习成为大数据智能分析的核心技术(28) T34 大数据系统评测基准(10) T35 大数据生命周期管理(7) T36 大数据存储与管理(23) T37 大数据安全(33)
大数据产业和业态	
T38 大数据产业成为国家战略新兴产业(31) T39 大数据个人数据保护和隐私问题立法(20) T40 数据主权是国家主权(6) T41 出现真正的大数据资源中心(4) T42 国家关键数据资源(8) T43 数据共享联盟(31) T44 开源成为主流(23) T45 资本高度关注(13) T46 产业政策(7) T47 数据权属立法启动(8)	T48 数据产品丰富(4) T49 数据外包(4) T50 产业垂直整合(2) T51 大数据新职业(12) T52 大数据与云计算、物联网(20) T53 数据价值评估模型(2) T54 更多数据驱动的商业模式(13) T55 大数据供应链(1) T56 大数据标准(7) T57 大数据处理生态系统(7)