

# CCF大专委2019年 大数据发展趋势预测

*Developing tendency prediction of  
big data in 2019 from CCF TFBD*



**周涛** (1979- ), 男, 博士, 阿里巴巴集团安全全部教授级高级工程师、资深算法专家, 主要研究方向为大数据安全分析、事件关联分析、入侵检测等。



**潘柱廷** (1969- ), 男, 北京永信至诚科技股份有限公司教授级高级工程师、高级副总裁, 中国计算机学会 (CCF) 常务理事、中国网络安全协会人才培养教育工作委员会副主任、CCF大数据专家委员会委员兼副秘书长、CCF计算机安全专家委员会常务委员、中国互联网协会常务理事、云安全联盟 (CSA) 中国区理事。长期从事信息安全技术、战略研究和教育工作。



**程学旗** (1972- ), 男, 大数据分析系统国家工程实验室副主任, 中国科学院计算技术研究所研究员、副总工程师、副所长, 中国科学院网络数据科学与技术重点实验室主任, 中国计算机学会大数据专家委员会秘书长, 国家杰出青年科学基金获得者。先后主持并完成了十余项国家自然科学基金、国家重点基础研究发展计划 (“973” 计划)、国家高技术研究发展计划 (“863” 计划)、国家信息安全重大专项以及中国科学院知识创新工程等科研任务。两次获得国家科技进步奖二等奖 (2012年个人排名第一、2004年个人排名第二), 获得第十二届中国青年科技奖、中国计算机学会青年科学家奖、中国科学院青年科学家奖等荣誉。主要研究方向为 Web 信息检索与数据挖掘。

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2019009

## 1 引言

2018年12月6日,中国计算机学会(CCF)大数据专家委员会(以下简称大专委)在2018年中国大数据技术大会(BDTC)的开幕式上,正式发布了2019年大数据十大发展趋势预测。作为自2012年起就持续开展的一项活动,大专委“大数据发展趋势预测”已经形成了良好的品牌效应。本次趋势预测结果一经发布,就引发了国内各大媒体的广泛传播。

本次大数据发展趋势预测面向大专委的正式委员和通讯委员,经历了候选项征集和正式投票两个环节。2018年委员们对趋势预测的参与热情有了显著提升,投票人数创历史新高。在候选项征集环节,有47位委员对候选项的设立积极建言献策,笔者团队根据大家的意见对2019年趋势预测的候选项进行了大幅度的修订,补充了若干体现大数据领域最新进展的候选项,调整和删除了一些过时选项,最终形成的预测选项包括60项发展趋势选项和9项专项调研选项。在正式投票环节,通过微信、邮件等方式

共收回选票130份。通过对这些选票的汇总和整理,形成了对2019年发展趋势的预测,见表1。

通过与2018年大数据发展趋势预测结果的对比可以发现,2019年大数据发展趋势预测结果具有以下特点。

- 人工智能呈现出压倒性优势。在十大发展趋势预测中,有6条与人工智能直接或间接相关,且涉及学科发展(第1条、第6条)、实现技术(第2条、第9条)、应用场景(第5条、第7条)等多个方面。

- 数据安全问题关注度持续提升。本次趋势预测中出现了两条与数据安全相关的预测项,其中既有与个人数据安全与隐私保护相关的预测项(第3条),又有与国家层面数据权属相关的预测项(第10条),说明大专委的专家们开始站在更高的层面上思考数据安全问题。

- 大数据学科建设受到重视。大专委的专家们认可数据科学对多学科融合的推动作用(第4条),也重视基础理论研究,但仍然不认为短期内能取得突破性进展。

- 区块链技术首次出现在预测结果中。区块链并不是一项新技术,在2018年年底甚至已经出现了降温的态势。在区块链大热的时候,该选项从未进入前十名,反倒是在当前这个时间点上被更多的专家

表1 2019年、2018年大数据十大发展趋势预测对比

2019年预测排名	2018年预测排名	预测项
1	-	数据科学与人工智能的结合越来越紧密
2	1	机器学习继续成为大数据智能分析的核心技术
3	-	大数据的安全和隐私保护成为研究和应用热点
4	3、4	数据科学带动多学科融合;基础理论研究受到重视,但未见突破
5	10	基于知识图谱的大数据应用成为热门应用场景
6	7	数据的语义化和知识化是数据价值的基础问题
7	-	人工智能、大数据、云计算将深度融合为一体化的系统
8	-	基于区块链技术的大数据应用场景渐渐丰富
9	8	大数据处理多样化模式并存融合,基于海量知识仍是主流智能模式
10	-	关键数据资源涉及国家主权

认可(第8条),背后的原因值得思考。

## 2 2019年大数据发展十大趋势预测

### 2.1 趋势一: 数据科学与人工智能的结合越来越紧密

该项是在本次候选项征集阶段,根据委员们反馈的意见新增的项目。一个候选项首次出现就成为趋势预测的冠军,这在历次调研中都没出现过,可见本预测项的受欢迎程度。

数据科学与人工智能虽然目前是两个独立的学科,但二者均与计算机、数学(特别是统计学)有密切的联系,问题空间也有一定的重合度。近年来,人工智能已经成为推动数据科学发展的核心驱动力,许多委员从事与两个学科相关的工作,例如为了应用人工智能技术而借助数据科学的理论和方法进行数据管理,或者为了挖掘数据的价值而借助人工智能技术进行数据分析。相信随着应用场景的拓展,二者之间的界限也会越来越模糊。

### 2.2 趋势二: 机器学习继续成为大数据智能分析的核心技术

该项在连续两年拔得头筹后,终于走下冠军宝座,以微弱劣势屈居亚军。这种连续排名靠前的阵势,本身就说明了大家对机器学习的认可。

大数据的价值是潜在的,不具备表象性。管理大数据的价值在于利用大数据,而如果没有机器学习技术对大数据进行分析,大数据的利用将无从谈起。随着机器学习与数据科学家的关系越来越紧密,对于数据科学领域的职业发展而言,掌握机器学习的基础技能将成为一种必需技能。

在大数据时代,依靠大数据管理和高性能计算的支持,机器学习将成为大数据智能分析的核心技术。

### 2.3 趋势三: 大数据的安全和隐私保护成为研究和应用热点

该项目是本次新增的候选项。在往年的调研中,趋势项“大数据的安全持续令人担忧”曾经连续5年入选十大趋势预测。2018年笔者根据候选项征集结果,对该项目进行了扩展,补充了与隐私保护相关的内容,调整后的项目依然延续了往年的热度,成为排名第3位的趋势项,这也说明大家对数据安全是一贯重视的。

2018年,一个标志性的事件使得数据安全与隐私保护成为政府、学术界和产业界共同关注的焦点,这就是欧盟《通用数据保护条例(GDPR)》的推出。GDPR引发了全球各行各业,特别是互联网巨头的高度关注。GDPR中的相关条款(如适用范围的扩大、对数据主体权利的提升、对数据控制者和处理者严格的问责制度、对数据画像的特别限制等)对现有的数据安全机制提出了更高的要求,这也使得人们对数据安全和隐私保护问题的关注度得到了提升。

### 2.4 趋势四: 数据科学带动多学科融合; 基础理论研究受到重视,但未见突破

该项是2018年预测结果趋势项第3条“数据科学带动多学科融合”与趋势项第4条“数据学科虽然兴起,但是学科突破进展缓慢”的合集。由于本次投票中两个趋势项得票相同、内容相关,故在这里合并为一个趋势项。这两个趋势项的排名与去年类似,依然成为较为靠前的趋势项。

在大数据时代,许多学科表面上看来

研究的方向大不相同,但是从数据的视角来看,其实是相通的。随着社会的数字化程度逐步加深,越来越多的学科将在数据层面趋于一致,可以采用相似的思想进行统一的研究。“数据科学发现范式”成为多学科通用的研究范式,因此数据科学对多学科融合的推动作用受到了委员们的认可。

作为一门与数学、计算机等学科相关的交叉学科,虽然数据科学已经初具规模,国内也出现了相关的专业设置、课程设置、标准教材,但数据科学自身仍然缺少突破性的理论成果。对科学问题的认识和求解需要一个过程且有不确定性,近期仍然很难取得重大突破。该趋势项的产生说明大专委的专家对大数据学科建设的矛盾心理依然存在。

## 2.5 趋势五: 基于知识图谱的大数据应用成为热门应用场景

该项首次出现在2018年大数据发展趋势预测的结果中,2019年预测结果的排名由2018年的第10位大幅攀升至第5位,说明人们对知识图谱的关注度有了进一步的提升。

知识图谱是一种以符号形式描述物理世界中的概念、实体及其关系的网状知识结构。基于知识图谱建立大数据表述的实体间的关联关系,并以此为基础开展各类个性化的应用成为发展趋势。当前知识图谱技术主要应用于智能语义搜索(如Knowledge Vault)、移动个人助理(如Google Now、Apple Siri)以及深度问答系统(如IBM Watson、Wolfram Alpha)等。随着智能音箱、语音助手、智能客服、知识问答等应用的成熟,普通人在日常生活中已经不知不觉地享受到知识图谱带来的种种便利,预期未来基于知

识图谱的大数据应用将会渗透到更多的领域和场景。

## 2.6 趋势六: 数据的语义化和知识化是数据价值的基础问题

该项在近3年的趋势调研中连续出现,每年的排名变化不大,由2018年的第7位上升至2019年的第6位。该趋势项可以看成趋势五“基于知识图谱的大数据应用成为热门应用场景”背后的理论基础。

数据语义化是通过符号变换将文档转换成机器可“理解”的符号的过程;数据知识化是在语义化的基础上,进一步挖掘并展示数据深层含义的过程,这两个过程是知识自动发现和挖掘的基础。从大数据中获得知识和价值是人们利用大数据的一个基本需求。在当前热门的大数据应用中,从知识图谱到多种自然语言问答应用的出现,可以推断广大用户在大数据时代获取信息时,越来越需要数据和信息的知识化组织和语义关联。

## 2.7 趋势七: 人工智能、大数据、云计算将高度融合为一体化的系统

该项也是在候选项征集阶段,根据大专委委员们反馈的意见新增的项目。本趋势项集齐了“ABC(artificial intelligence、big data、cloud computing)”三大热门技术,这使得它首次出现便跻身最终的预测项中。

该项主要体现了工业界的发展趋势。当前无论是公有云还是私有云,云服务提供商都倾向于提供一体化的平台,为用户提供统一的人工智能分析建模、大数据计算以及资源分配与共享管理功能,从而增加便利性、降低使用成本、丰富业务场景。反映在云服务内容上,无论是国

外亚马逊的AWS、微软的Azure、谷歌的GoogleCloud, 还是国内的阿里云、腾讯云, 都已经不满足于仅提供基础设施即服务 (infrastructure as a service, IaaS) 层虚拟化的能力, 而是更多地提供大数据存储及智能分析的软件即服务 (software as a service, SaaS) 能力, 这将大大加快云用户在此基础上拓展业务能力的步伐。

## 2.8 趋势八: 基于区块链技术的大数据应用场景渐渐丰富

与区块链相关的趋势项首次出现在预测结果中。区块链不是一项“新兴”的技术, 它已经存在了很多年; 区块链也不是一项“热门”的技术, 事实上在近期咨询机构的报告中, 区块链已经渐渐走下巅峰, 有了降温的趋势。但在区块链火热的时候, 该选项从来没有出现在预测结果中, 反倒是在这个时间点上被更多的委员们认可, 这反映了专家们对新事物的冷静判断力。

2018年CCF区块链专业委员会正式成立, 致力搭建产业界和学术界互动的专业平台, 这也推动了本项目成为2019年的趋势项之一。区块链具有去中心化、难以篡改、记录可溯源等优点, 这使得它在交易、认证、流程管理等领域具有广泛的应用场景。相信随着更多的应用驱动, 更多的基于区块链的大数据应用将会涌现。

## 2.9 趋势九: 大数据处理多样化模式并存融合, 基于海量知识仍是主流智能模式

该项是两个趋势项“大数据处理多样化模式并存融合”与“基于海量知识仍是主流智能模式”的合集, 同样也是由于得票相同的原因进行了合并。其中, 趋势

项“基于海量知识仍是主流智能模式”在2018年趋势预测中排名第8位, 在2019年的趋势预测中微调至第9位, 变化趋势不明显。

在大数据处理模式方面, 专家们认为批量计算、流式计算和内存计算等多种大数据计算模式将同时存在, 一些技术将趋于融合。现实中的需求是多样化的, 不同业务场景中数据的量级、产生的速度、对时延的容忍度、计算的模式 (历史、近线、实时) 等差异巨大, 这就需要有多样化的模式满足差异化的需求。

在数据工程领域, 知识是更高层次的数据, 海量知识来源于对海量数据的语义挖掘、信息抽取和知识库构建。通过从数据中提炼信息和知识, 可以消除原始数据中的不确定性、补充信息的上下文、降低特定问题搜索空间。在海量知识的基础上进行检索和推理, 是当前火热的各类“智能助手”背后的核心技术, 这也是该趋势项能够持续入选的主要原因。

## 2.10 趋势十: 关键数据资源涉及国家主权

该项首次入选十大趋势, 说明专家们已经不仅仅局限在从个人、机构的视角考虑数据安全问题, 而是开始站在国家层面思考数据安全问题。

在信息时代, 数据已经像石油一样, 成为重要的战略资源。但是在数据主体方面, 有相当一部分数据资源掌握在各类企业中, 这是与传统自然资源相比最大的差异。一些互联网巨头掌握的数据资源非常多, 如果其丧失数据权属, 可能会危及社会秩序和国家安全。为此, 国家互联网信息办公室于2017年制定了《关键信息基础设施安全保护条例 (征求意见稿) 》, 将一旦遭到破坏、丧失功能或者数据泄露, 可能严

重危害国家安全、国计民生、公共利益的信息系统，纳入了关键信息基础设施保护范围，这也推动了委员们对数据权属问题的高度重视。

### 3 大数据发展专项调研

与往年的调研保持一致，本年度的趋势调研仍然包括9个专项调研项目。对于每个调研项目，保留得票数远高于其余候选项的条目作为调研结果，具体见表2。

在“最令人瞩目的应用领域”投票中，健康医疗排在了首位，而人们习以为常的互联网、电子商务退居到第4位，这表明大数据的应用领域会更加丰富。这对整个行业而言是个利好消息。与之相对应的是，在“将取得重大应用和技术突破的数据类型”投票中，城市数据排在了首位，其次是视频数据，而往年经常出现在前面的互联网公开数据的票数很小，在表2中没有列出来。

在“最令人瞩目的学科和技术”投票中，人工智能、自然语言处理/知识工程、图计算位列三甲，在“将取得突破性进展的技术环节”投票中，数据语义理解、数据分析、大数据安全排在前三位，

这与前文的“十大发展趋势预测”的结果是吻合的。

在“大数据市场处于哪个阶段”的判断上，参照常规的技术成熟度曲线，将大数据市场的发展分为极为初级、即将快速扩张、爆发增长、上升乏力、下降和幻灭、稳步成长6个阶段。可以看到，大专委的专家认为大数据的发展还是趋于乐观的，大部分把选票投给了即将快速扩张、爆发增长和稳步成长3个阶段。从投票结果来看，大数据已经度过了下降期和幻灭期，进入了相对丰富的增长阶段。

在“我国大数据发展的最主要推动者”投票中，投票结果已经维持多年不变，仍然只有大型互联网公司和政府机构，推动者的丰富度和力度还有待加强。反映在“数据资源流转上的举措”上，投票结果跟往年比依旧没有显著的变化，大家还是更倾向于自己收集数据，或者为已经收集的数据提供服务，产业界期待的数据交易选项并没有得到太多的选票。在“专家个人的工作重点”投票中，排名前三位的分别是算法等基础技术研究、应用实践、人才教育，其他选项（如基础理论研究）的票数较少，这也反映了国内大数据人才在各领域的分布情况。

在“大数据的最佳拍档概念”投票中，

表2 2019年大数据趋势专项调研结果

序号	调研项目	调研结果
1	最令人瞩目的应用领域	健康医疗、城镇化/智慧城市、金融、互联网/电子商务
2	将取得重大应用和技术突破的数据类型	城市数据、视频数据
3	最令人瞩目的学科和技术	人工智能、自然语言处理/知识工程、图计算
4	将取得突破性进展的技术环节	数据语义理解、数据分析、大数据安全
5	大数据市场处于哪个阶段	即将快速扩张、爆发增长、稳步成长
6	我国大数据发展的最主要推动者	大型互联网企业、政府机构
7	在数据资源流转上的举措	会自己收集数据、会利用数据提供服务
8	专家个人的工作重点	算法等基础技术研究、应用实践、人才教育
9	大数据的最佳拍档概念	数据科学、机器人和人工智能、智能计算和认知计算

排名前三位的选项（数据科学、机器人和人工智能、智能计算和认知计算）的票数非常接近，且远远领先于其余选项。这3个选项之间本身就有很强的关联关系。考虑到人工智能在近两年投票中的压倒性优势，出现这样的结果也在意料之中。

## 4 结束语

作为一项已经连续开展了7年的趋势预测活动，笔者期待在每年的预测结果中能够产生一些新热点，给读者一些新感受，但结果却往往事与愿违，特别是最近3年的趋势预测结果。每年的预测结果中大约有60%的趋势选项会跟上一年完全重合，其余的变化往往来源于已有候选选项的相近概念，很难产生新热点。

笔者也在思考出现这种局面的原因。可能正像对“大数据发展阶段”的判断那样，大数据已经成为各行各业的底层共性

技术，进入了一个稳定发展的时期。在这种情况下，所谓的“热点”，其实已经成为支撑领域发展不可或缺的“支点”，而相对稳定的支点是一个行业发展成熟的重要标志。在这种情况下，除非是出现变革型的理论或技术突破，否则整个领域将会在几个基础的支点支撑下平稳发展。

## 参考文献:

- [1] 周涛, 潘柱廷, 杨婧, 等. CCF大专委2017年大数据发展趋势预测[J]. 大数据, 2017(1): 102-108.  
ZHOU T, PAN Z T, YANG J, et al. Developing tendency prediction of big data in 2017 from CCF TFBD[J]. Big Data Research, 2017(1): 102-108.
- [2] 周涛, 卞超轶, 潘柱廷, 等. CCF大专委2018年大数据发展趋势预测[J]. 大数据, 2018(1): 77-84.  
ZHOU T, BIAN C Y, PAN Z T, et al. Developing tendency prediction of big data in 2018 from CCF TFBD[J]. Big Data Research, 2018(1): 77-84. □