

CCF大专委2020年 大数据发展趋势预测

Developing tendency prediction of big data in 2020 from CCF TFBD



周涛 (1979-), 男, 博士, 阿里巴巴集团安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。



程学旗 (1971-), 男, 中国科学院计算技术研究所研究员、副所长, 中国科学院大学讲座教授, 中国科学院网络数据科学与技术重点实验室主任, 大数据分析系统国家工程实验室副主任。国家杰出青年科学基金获得者, 中共中央组织部“万人计划”科技领军人才, 曾获得中国青年科技奖、中国科学院青年科学家奖和中国计算机学会青年科学家奖。主要研究方向为网络数据科学、大数据分析系统、互联网搜索与挖掘、网络空间信息内容安全等。其科研成果在全国大范围应用, 并发挥实际效果, 带领团队3次获得国家科技进步奖二等奖。发表学术论文200余篇, Google Scholar引用14 000余次, 3次获得国际顶级学术会议最佳学术论文奖。兼任中国计算机学会大数据专家委员会秘书长、中国中文信息学会信息检索专业委员会主任等。



陈宝权 (1969-), 男, 北京大学博雅特聘教授, 前沿计算研究中心执行主任, IEEE Fellow。中国计算机学会常务理事, 第七届教育部科学技术委员会信息学部委员。主要研究方向为计算机图形学与可视化, 国家“973”项目“城市大数据计算理论与方法”首席科学家。

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.20200009

1 引言

自2012年起,中国计算机学会(CCF)大数据专家委员会(以下简称大专委)面向全体委员,发起了一年一度的大数据趋势预测活动。站在年底展望来年大数据领域的发展趋势,已经成为大专委的一项品牌活动。2019年12月5日,在中国大数据技术大会(BDTC)开幕式上,CCF大专委正式发布了2020年大数据十大发展趋势预测。本次预测汇集了160余名大专委委员的投票结果,参与投票的委员来自高校、科研院所、相关部委、大型央企、民营企业等不同团体,具有广泛的代表性。本文将介绍本次趋势预测结果,分析历年预测结果的变化趋势,并对本次趋势预测体现出的新变化进行解读。

2 2020年大数据十大发展趋势预测解读

按照得票数排序,大专委对2020年大数据十大发展趋势的预测结果见表1。

表1 2020年大数据十大发展趋势预测

排名	预测项目
1	数据科学与人工智能的结合越来越紧密
2	数据科学带动多学科融合,基础理论研究的重要性受到重视,但理论突破进展缓慢
3	大数据的安全和隐私保护成为研究和应用热点
4	机器学习继续成为大数据智能分析的核心技术
5	基于知识图谱的大数据应用成为热门应用场景
6	数据融合治理和数据质量管理工具成为应用瓶颈
7	基于区块链技术的大数据应用场景渐渐丰富
8	对基于大数据进行因果分析的研究得到越来越多的重视
9	数据的语义化和知识化是数据价值的基础问题
10	边缘计算和云计算将在大数据处理中成为互补模型

同往年相比,本次的趋势预测结果体现出以下共性。

- 大数据与人工智能的共生关系受到持续认可。反映在预测结果上,2020年与人工智能相关的预测项(表1中的第1条、第4条)已经连续4年出现在十大趋势中。

- 对学科突破的期待心态依然存在。这体现在第2条预测项上,专家们认可数据科学对多学科融合的推动作用,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。

- 既要挖掘数据价值,又要在此过程中兼顾数据安全和隐私保护。与数据安全相关的预测项(表1中的第3条)是十大趋势中的常青树,历年都会有与之相关的预测项。伴随着2018年欧盟《通用数据保护条例(GDPR)》落地引发的关注、2019年我国网络安全和信息化委员会办公室《数据安全管理办法(征求意见稿)》的出台,2020年与数据安全相关的选项再度进入前三名。

- 从数据到知识的途径依然是关注热点。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,因此基于知识图谱的大数据应用以及与知识自动发现和挖掘相关的候选项(表1中的第5条、第9条)得到了较多关注。

- 大数据与区块链的结合稳中有升。区块链是一项出现已有10年的技术,在2019年下半年再度成为关注热点。体现在本次预测结果上,与区块链相关的预测项(表1中的第7条)排名较2019年上升一位。

对于上述在往届趋势预测结果中曾经出现过的预测项,本文不再对其进行详细解读,如需了解可参阅2019年的趋势预测分析^[1],本文重点关注2020年趋势预测结果中的新面孔(表1中的第6条、第8条、第10条)。

2.1 趋势六：数据融合治理和数据质量管理工具成为应用瓶颈

数据融合技术是多源信息协调处理技术的总称，数据治理是运用不同的技术工具对大数据进行管理、整合、分析并挖掘其价值的行为^[2]。数据融合治理是大数据应用的基石，如果数据在融合中存在属性偏差或信息损失，或者融合后的数据质量低下，上层应用的价值将无从保障。在行业大数据应用实践中解决了数据有无问题后，对数据质量的管理将会成为最迫切的挑战。目前业界还缺乏通用、有效的数据融合治理与数据质量管理工具，这将成爲大数据应用向深层次发展的瓶颈。

2.2 趋势八：对基于大数据进行因果分析的研究得到越来越多的重视

大数据时代“一切皆数据”，被数字化的事物和流程越来越多。利用统计方法对数据进行相关性分析，成为科学决策和预测的重要手段。然而相关性不等于因果性，许多在统计上具有强相关性的事物，在逻辑上并不存在直接或间接的因果性。如果无法分析出相关性背后的因果关系，不考虑结论的可解释性，必然会影响决策的质量和范围。例如，利用医疗大数据和人工智能算法，深度神经网络对病理图像处理的准确性已经达到甚至超过普通医师^[3]，但受限于深度学习的黑箱特性，目前仍然无法用深度神经网络取代医师的诊断结论。专家预测：对数据中的因果性、对结果可解释性的研究，将会受到更多的重视。

2.3 趋势十：边缘计算和云计算将在大数据处理中成为互补模型

边缘计算是指靠近数据源的处理模

式，是一种分散式处理框架。过去大数据的概念往往和云计算绑定在一起，但在实际应用中，将数据放在终端上进行部分处理的方法具有实时性高、对网络带宽占用少、更有利于隐私保护等优点。随着终端处理能力的增强，将部分计算任务部署在终端上，与云端任务进行合理的分层解耦，成为一种可靠性更高、计算成本更低、实时性更强的计算框架。预期在未来的大数据处理模式中，边缘计算和云计算将成为互补模型，共同发展。

3 历次趋势预测演变分析

自大专委2012年开展活动以来，已经连续8年对大数据领域的发展趋势进行了预测。如果以3年为一个周期，对每年的趋势预测结果进行归类 and 对比分析，可以清晰地感受到8年来大数据趋势预测结果的变化情况，这也体现了大数据发展阶段的变化。

- 2013—2015年：在该阶段的趋势预测项中，大数据概念刚刚产生，专家们关注大数据如何落地、如何从“概念”走向价值以及大数据与传统行业的跨界融合。这期间的典型预测项包括“数据的资源化”（2013年，第1项）、“大数据从概念走向价值”（2014年，第1项）、“大数据分析成为数据价值化的热点”（2015年，第1项）、“与各行业的结合，跨领域应用”（2015年，第3项）等。

- 2016—2018年：在该阶段的趋势预测中，大数据概念已经被各行业所接受，专家们关注大数据产业发展的推动力，从希冀政策法规过渡到依赖学科进展，最终聚焦在人工智能上。这期间的典型预测项包括“《促进大数据发展行动纲要》驱动产业生态”（2016年，第6项）、“人工智能

与脑科学相结合,成为大数据领域热点”(2017年,第2项)、“多学科融合与数据学科兴起”(2017年,第4项)、“机器学习继续成为大数据智能分析的核心技术”(2016年、2017年均第1项)等。

● 2019—2020年:在该阶段的趋势预测中,基于大数据的行业应用已经广泛兴起,专家们开始关注数据科学的基础作用、大数据在具体领域中发挥实效以及大数据应用落地中的具体技术障碍。这期间的典型预测项包括“数据科学与人工智能的结合越来越紧密”“数据的语义化和知识化是数据价值的基础问题”以及本次趋势中新出现的3条候选项。

4 大数据发展专项调研分析

在十大趋势预测之外,2020年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,2020年预测的投票结果延续了2019年预测的态势,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,这表明互联网领域的大数据应用相对成

熟,大家已经习以为常,其他领域的新应用更容易受到关注。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,视频数据首次排在了首位,这与健康医疗、智慧城市领域的主流数据类型是一致的。

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、图计算位列前三名,在“将取得突破性进展的技术环节”投票上,数据分析、数据语义理解、脱敏和隐私保护排在前三名,这与前面的“十大趋势调研”的结果是吻合的。

在“大数据市场处于哪个阶段”判断上,2020年的调研结果与2019年相比有一个有趣的变化。认为大数据的发展处于前两个阶段(初级、即将快速扩张)的得票率总计下降7%,认为处于中间3个阶段(爆发增长中、上升乏力、下降和幻灭)中每一项的得票率与2019年恰好相同,3项的得票率在最近两年均为21%、9%和1%,而认为处于最后一个阶段(稳步成长)的得票率上升7%。可以看到,专家们对于大数据的发展趋势更加乐观了,3个增长阶段的得票率综合达到了83%。从投票结果看,给读者的感觉是大数据已经度过了下降和幻灭期,进入了相对稳定的增长阶段。

在“我国大数据发展的最主要推动

表2 2020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	视频数据、城市数据及地理数据、互联网公开数据、语音数据
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	数据分析、数据语义理解、脱敏和隐私保护
5	大数据市场处于哪个阶段	稳定增长、即将快速扩张、爆发增长
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、会买数据
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、5G

者”投票中,投票结果已经维持多年不变,仍然是只有大型互联网公司和政府机构。互联网公司在自身业务发展中拥有了大数据,而政府机构在社会治理中也积累了大数据,其他领域的推动力量显得单薄。反映在“数据资源流转上的举措”上,投票结果跟往年相比有了细微的变化,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,同时购买数据的意愿得到了提升,首次进入前三名;但销售数据的意愿并没有提升,由此看来,距离交易双方的意愿达成还有一定差距。

在“大数据的最佳拍档概念”投票中,排名靠前的选项是数据科学、机器人和人工智能、智能计算和认知计算、5G。其中,前3个选项之间本身就具有较强的关联性,且在最近的趋势调研中已经连续出现。值得强调的是2020年的新面孔“5G”。2019年被称为我国的5G元年,国内电信运营商开启了对5G通信协议的支持,大家预期未来会产生一些基于5G的新应用形态,从而产生新的大数据应用领域。笔者也期待着未来大数据与5G融合产生的化学反应。

5 结束语

虽然这是一项已经连续开展了8年的趋势预测活动,人们仍然能够从每年的预测结果中感受到大数据领域的一些新变化。本次趋势预测中出现的3个新预测项中,“数据融合治理和数据质量管理工具”是在解决了数据的存储规模、访问速度、计算能力后,大数据应用深入发展面临的难题;“大数据因果分析”相对于直接的、表层的相关性分析,需要挖掘更深层次的逻辑关系;“边缘计算和云计算成为互补模型”则是大数据应用向低成本、低时延、保护隐私方向发展的更契合实际的落地方

案。总体来看,这3个新预测项针对的问题都是大数据应用发展到较深入的阶段才会遇到的新问题。笔者认为,一项事物发展到一定阶段,总会产生与这个阶段相适应的特定问题;本次趋势预测中新问题的出现,也标志着大数据应用发展到了一个新阶段。

大数据从概念兴起到应用落地,已有约10年的历史。在当前所处的数字经济时代,数据已经成为各行各业发展的基石。笔者期待着数据科学能取得理论突破,也希望基于大数据的应用能够更深层次、更加充分地体现大数据的价值。在本次趋势预测中,无论是在数据科学层面,还是在应用工具层面,都有一些需要攻克的难关。笔者期待着这些困难能够被逐渐攻破,从而让大数据的发展上升到一个新高度。

参考文献:

- [1] 周涛,潘柱廷,程学旗. CCF大专委2019年大数据发展趋势预测[J]. 大数据, 2019, 5(1): 112-118.
ZHOU T, PAN Z T, CHENG X Q. Developing tendency prediction of big data in 2019 from CCF TFBD[J]. Big Data Research, 2019, 5(1): 112-118.
- [2] 马朝辉, 聂瑞华, 谭昊翔, 等. 大数据治理的数据模式与安全[J]. 大数据, 2016, 2(3): 83-95.
MA Z H, NIE R H, TAN H X, et al. Research on data schema and security in data governance[J]. Big Data Research, 2016, 2(3): 83-95.
- [3] 萧毅, 夏晨, 张荣国, 等. 人工智能技术在医学影像中的应用讨论[J]. 第二军医大学学报, 2018, 39(8): 813-818.
XIAO Y, XIA C, ZHANG R G, et al. Discussion of artificial intelligence application in medical imaging[J]. Academic Journal of Second Military Medical University, 2018, 39(8): 813-818. □

2018年《大数据》高被引论文Top10

排名	论文信息
1	马小峰, 杜明晓, 余文兵, 等. 基于区块链的供应链金融服务平台[J]. 大数据, 2018, 4(1): 13-21. MA X F, DU M X, YU W B, et al. Blockchain based supply-chain finance service platform[J]. Big Data Research, 2018, 4(1): 13-21.
2	祝烈煌, 董慧, 沈蒙. 区块链交易数据隐私保护机制[J]. 大数据, 2018, 4(1): 46-56. ZHU L H, DONG H, SHEN M. Privacy protection mechanism for blockchain transaction data[J]. Big Data Research, 2018, 4(1): 46-56.
3	钱卫宁, 金澈清, 邵奇峰, 等. 区块链与分享型数据库[J]. 大数据, 2018, 4(1): 36-45. QIAN W N, JIN C Q, SHAO Q F, et al. Blockchain and sharing database[J]. Big Data Research, 2018, 4(1): 36-45.
4	蔡维德, 郁莲, 袁波, 等. 面向大数据的区块链在清算系统中的应用[J]. 大数据, 2018, 4(1): 22-35. TSAI W T, YU L, YUAN B, et al. Big data-oriented blockchain for clearing system[J]. Big Data Research, 2018, 4(1): 22-35.
5	闫树, 卿苏德, 魏凯. 区块链在数据流通中的应用[J]. 大数据, 2018, 4(1): 3-12. YAN S, QING S D, WEI K. Application of blockchain in data circulation[J]. Big Data Research, 2018, 4(1): 3-12.
6	李崇纲, 许会泉. 冒烟指数: 大数据监测互联网金融风险[J]. 大数据, 2018, 4(4): 76-85. LI C G, XU H Q. Smoke index: big data technologies monitor Internet financial risks[J]. Big Data Research, 2018, 4(4): 76-85.
7	胡鹏飞. 金融科技在互联网金融行业性风险防范领域的应用[J]. 大数据, 2018, 4(1): 117-123. HU P F. Application of FinTech in internet financial industry risk prevention[J]. Big Data Research, 2018, 4(1): 117-123.
8	郑广勇, 杨桢, 曹瑞芳, 等. 宏基因组大数据分析的质量控制流程规范[J]. 大数据, 2018, 4(3): 3-12. ZHENG G Y, YANG Z, CAO R F, et al. Quality control of big data analysis for metagenomics[J]. Big Data Research, 2018, 4(3): 3-12.
9	王宏志, 梁志宇, 李建中, 等. 工业大数据分析综述: 模型与算法[J]. 大数据, 2018, 4(5): 62-79. WANG H Z, LIANG Z Y, LI J Z, et al. Survey on industrial big data analysis: models and algorithms[J]. Big Data Research, 2018, 4(5): 62-79.
10	朱扬勇, 叶雅珍. 从数据的属性看数据资产[J]. 大数据, 2018, 4(6): 65-76. ZHU Y Y, YE Y Z. Defining data assets based on the attributes of data[J]. Big Data Research, 2018, 4(6): 65-76.

数据来源: 中国知网; 统计日期: 2020-01-01