CCF大专委2018年大数据发展趋势预测

Developing tendency prediction of big data in 2018 from CCF TFBD

周涛（1979-），男，博士，北京启明星辰信息安全技术有限公司教授级高级工程师、助理总裁，核心研究院院长，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。

卞超轶（1987-），男，北京启明星辰信息安全技术有限公司高级研究员，主要研究方向为大数据自身安全、大数据安全分析、AI+信息安全等。

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。

查礼（1974-），男，中国科学院计算技术研究所副研究员，中国计算机学会大数据专家委员会委员，《大数据》杂志编委。2003年于北京理工大学博士毕业后进入中国科学院计算技术研究所，一直从事分布式系统方向的研究工作。作为课题负责人承担过多项网格、云计算和大数据相关的国家级研究课题。发起并组织“Hadoop in China”大会（现已更名为中国大数据技术大会）。自2008年举办以来，参会人数逐年递增，目前已成为专注于大数据相关技术方向国内活跃的技术大会。曾两次获国家科技进步奖二等奖（2007年、2012年）。

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.2096-0271.2018008

1 引言

在2017年中国大数据技术大会（BDTC）开幕式上，中国计算机学会（CCF）大数据专家委员会（以下简称大专委）如期发布了2018年大数据十大发展趋势预测，引发了业界的广泛关注和持续传播。

本次大数据发展趋势预测经历了候选项征集和正式投票两个环节。在候选项征集环节，补充了若干体现大数据领域最新进展的候选项，并调整和删除了一些过时选项，最终形成的预测选项包括67项发展趋势选项和9项专项调研选项。在正式投票环节，投票范围面向大专委的正式委员和通讯委员，共收回选票82份。通过对这些选票的汇总和整理，形成了对2018年发展趋势的预测，与2017年预测结果的对比见**表1**。

通过对比不难发现，大专委对2018年大数据发展趋势预测的结果与2017年预测结果的重合度较高，10条预测项中有6条出现在2017年度的预测结果中。新出现的4条预测项反映了本次大专委预测结果的两大特点：一是人工智能在大数据应用中具有压倒性的优势，新增的4条预测项中，3条与人工智能相关（2018年预测排名中的第6条、第8条、第10条）；二是对大数据学科建设的依赖性增强，体现在新出现的另外一条预测项上（2018预测排名中的第4条），大专委的专家既对数据科学寄予厚望，又担心其学科突破进展缓慢。本文将对2018年大数据十大发展趋势预测进行简要的解读。

2 2018年大数据发展十大趋势

2.1 趋势一：机器学习继续成为大数据智能分析的核心技术

该项延续了2017年的预测结果，再次在投票中拔得头筹，可见其公认度之高和稳定。“大数据”一词原本是数据量大、数据样式复杂等特性的代名词，如今已经逐渐转变为预测分析、用户行为分析、态势感知等高级智能分析方法的运用。

大数据智能分析旨在从数据中挖掘提取潜藏的巨大价值，这正是大数据的核心意义所在。智能分析方法均以机器学习为核心，甚至可以说是机器学习技术的不同表现形式。机器学习（包括近年来兴起的深度学习、强化学习等）已是从事大数据行业的人员应具备的基础技能之一，它在

表1 大专委2017年、2018年大数据十大发展趋势预测对比

2018年预测排名	2017年预测排名	预测项
1	1	机器学习继续成为大数据智能分析的核心技术
2	2	人工智能和脑科学相结合，成为大数据分析领域的热点
3	4	数据科学带动多学科融合
4	–	数据学科虽然兴起，但是学科突破进展缓慢
5	9	推动数据立法，重视个人数据隐私
6	–	大数据预测和决策支持仍然是应用的主要形式
7	6	数据的语义化和知识化是数据价值的基础问题
8	–	基于海量知识的智能是主流智能模式
9	3	大数据的安全持续令人担忧
10	–	基于知识图谱的大数据应用成为热门应用场景

大量数据样本的支撑与分布式存储管理及计算处理等技术的支持配合下，成为将大数据转化为实际价值的核心手段的不二之选。

2.2 趋势二：人工智能和脑科学相结合，成为大数据分析领域的热点

与趋势一相同，该项也延续了2017年预测结果的排位，再次占据投票排行的榜眼位置。脑科学也称神经科学（Neuroscience），近年来在研究深度和宽度上有了重大突破，包含从对单个神经细胞的分子与细胞级的研究到对全脑神经网络的活动成像。人工智能与脑科学的结合可以追溯到20世纪四五十年代，人工神经网络的出现正是两个学科的最初也是最重要的成果之一。DeepMind公司在2017年12月发布的AlphaZero同时在围棋、国际象棋上展现出超越人类的强大智能，其中采用的卷积神经网络等深度学习技术的思想也是起源于人工神经网络及一些对人脑的初步研究结论，这说明了脑科学与人工智能结合的巨大潜力。

然而，脑科学研究与人工智能的真正融合还没有发生，因为研究者尚未完成对人脑神经结构的解析，不清楚百亿级的神经元如何交互，以完成高效的信息处理。脑科学的研究进展可能成为人工智能跨越发展的关键助推，如神经网络的自组织、自学习等，从而为大数据分析带来突破。因此，对人工智能与脑科学的结合研究将持续升温，成为相关领域的重要热点。

2.3 趋势三：数据科学带动多学科融合

该项是2017年预测结果趋势四——“多学科融合与数据科学兴起”的发展演进。数据科学从兴起逐渐成长为现实，专门的研究机构的建立以及相应的专业与学位的设立是这一过程的真实写照，数据科学家已然成为21世纪最受追捧的职业之一。但从本质上看，数据科学是一门综合统计、数据挖掘、机器学习、数据可视化、分布式系统、高性能计算等多项理论及技术，以从数据中提取潜在价值为目标的学科，它的存在本身就是多学科融合的典范。因此，数据科学的发展成熟必然会进一步推动相关学科的深入交叉融合。

此外，数据科学的发展对其他领域也产生了重要影响，包括经济学、医学、生物学、社会学等，它提供的数据处理及分析技术为研究者们提供了极大的帮助。人们发现越来越多的来自不同学科领域的问题可以采用类似的思想和方法进行研究，从而推动学科间的交流融合，促进共同发展。

2.4 趋势四：数据学科虽然兴起，但是学科突破进展缓慢

该项是十大预测中的新面孔。随着大数据技术的广泛应用，近年来数据学科已然兴起。国内外一些高校已经设立了相关专业，开设有关课程，逐步探索其发展方向。国外很多大学将数据科学与原有特色专业结合，在本校具有优势的领域中关注和实践数据科学。国内高校也纷纷设立了与大数据相关的专业或研究所，探索数据科学专业的未来发展。2016—2017年，经教育部批准，国内共有35所高校成功申请了“数据科学与大数据技术”本科专业。除了数学、通信和计算机等基础课程外，开设的专业课程主要分为3个方向，即大数据分析方向、大数据平台方向和深度计算分析方向。

学科是人类知识体系中的基本组成部

分，任何一个学科的发展都会经历萌生、形成、成长到成熟的过程。总体来说，大数据学科建设尚处于摸索阶段，还没有一个成熟的学科体系，相关课程体系及要求尚未完全达成共识，还需要进行进一步的技术研究、实践积累和理论提升，只有相应的知识被创造并逐步发展成系统化的理论与方法，才能形成一个有特色的学科。

趋势四与趋势三共同出现，反映了大专委的专家对大数据学科建设的矛盾心理。一方面，大专委的专家寄希望于在具体的应用技术之外，能够通过学科建设带动大数据的发展；另一方面，又对学科建设发展的进度持悲观看法。暂且不考虑这种矛盾性，这两项趋势预测同时出现，也体现了大专委越来越多的专家开始在技术之外，从科学的角度思考大数据的本质问题。

2.5 趋势五：推动数据立法，重视个人数据隐私

该项来自于2017年预测结果的趋势九。数据安全和个人隐私泄漏已然成为全球安全问题的焦点，近年来，有关数据和个人隐私数据泄露的安全事件频频爆出，如美国信用机构Equifax因遭到黑客袭击，大约1.43亿名用户的数据被泄露，相关内容包括社保号码、生日、地址等。所以，在2018年的十大趋势中，该项被更多的专家关注，一举进入前五。

要做到对数据加强保护，除了采用技术手段和行业自律外，还应加强法律建设和政府监管。2017年6月1日起,我国开始施行的《中华人民共和国网络安全法》用一个章节的篇幅专门规定网络信息安全保护相关条款，这对加强数据保护起到了非常积极的作用。但是，还要看到，由于技术的快速发展和现实情况的复杂多变，我国现行的法律法规中对网络信息保护的条款还不够，相互之间的协调也还存在一定的问题。因此，要从数据的全生命周期进行综合考虑，进一步推动数据立法，从法律层面对数据的采集、传输、流转、交易、使用和销毁等环节做出明确约束，使得个人数据隐私保护有法可依，以更好地对数据和个人隐私进行强有力的法律保护。

2.6 趋势六：大数据预测和决策支持仍然是应用的主要形式

这是有关大数据应用场景的预测。利用大数据做预测和决策支持是大数据的经典应用场景，也与机器学习和数据挖掘密切相关。典型做法是通过分析海量历史数据，找到现有现象之间的相关关系，建立相应的机器学习模型，并应用构建的模型预测未来，进而向决策者提供决策支持。

通过对海量的多维、异构数据进行融合分析，可以从时间、空间、网络等多个维度面向特定对象建立更全面和精准的画像，分析历史行为轨迹，预测未来发展态势。典型应用场景包括个性化推荐、资源配置优化、企业决策支持等。例如，电子商务企业通过分析用户的历史购买行为，进行精准的商品推荐；网约车企业通过历史数据对特定区域未来的客流量进行预测，进而实现车辆预先调度，达到整体资源利用最优化的目的。

2.7 趋势七：数据的语义化和知识化是数据价值的基础问题

该项在2017年的预测中排名第六，2018年的排名变化不大。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Linked of Data的发展，到Google知识图谱，再到Google Vault以及深度问答应用的出现，证明了数据的知识化组织和语义关联是发现、挖掘并有效管理大数据深层价值的前提。在可预见的未来，人们将面临更快的数据增长和更广的数据维度，面对这些海量复杂数据，数据的价值更容易被淹没。如何更好地发现和理解这些海量数据，依然会是未来持续关注的问题。

2.8 趋势八：基于海量知识的智能是主流智能模式

该项同样是趋势预测中的新面孔，可以作为趋势七的后续。2017年人工智能领域的一大热点是出现了像“AlphaGo”“AlphaZero”这种不基于人类已有知识的智能模式，但大专委的专家给出的预测中包含了基于海量知识的智能模式，这也体现了人工智能应用模式多样化的趋势。

计算机既能存储人们积累起来的知识和经验，又可以挖掘大数据中包含的信息，因此可以取代部分人脑的劳动。如果对人脑的研究有重大科学突破，机器很有可能成为人工大脑，像会思考的人一样处理信息。人工大脑的实现依赖于海量数据语义挖掘、信息抽取和知识库构建的创新及实用方法以及面向海量语义知识库（信息）的语义查询技术和方法。在趋势七的基础上，利用大数据实现基于海量知识的智能，也就顺理成章了。

2.9 趋势九：大数据的安全持续令人担忧

这是最近5年来连续出现在预测结果中的选项，只是每年的排名会有一些变化。大数据安全风险伴随大数据应用而生，人们在享受大数据福祉的同时，也遭受着前所未有的安全挑战。随着大数据应用的爆发，应用系统遭受攻击、数据丢失和个人信息泄漏的事件常有发生，而地下数据交易“黑灰产”也导致了大量的数据滥用和网络诈骗事件。这些安全事件，有的造成了个人的财产损失，有的引发了恶性社会事件，有的甚至危及了国家安全。可以说当前环境下，大数据平台与技术、大数据环境下的数据和个人信息、大数据应用等方面都面临着极大的安全挑战，这些挑战不仅对个人会产生重大的影响，更直接威胁到社会稳定和国家安全。

相对于业务功能，安全手段往往具有滞后性。现有大数据平台和技术主要围绕大容量、高速率的数据处理功能开发，在安全机制方面多通过调用外部安全组件、修补安全补丁的方式进行，存在整体安全规划不足、缺乏内建安全机制和安全措施协调不够等问题。因此，要想让大数据发挥作用，其安全保护仍然是一个要花大力气、持续解决的重要事项。

2.10 趋势十：基于知识图谱的大数据应用成为热门应用场景

该项首次出现在大专委的调查问卷中，就成功入选十大趋势，可见知识图谱在大数据领域的受关注程度。知识图谱是一种以符号形式描述物理世界中概念、实体及其关系的网状知识结构。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。然而，随着各领域数据的积累，海量复杂数据将不断加剧知识的碎片化和复杂

化，知识的碎片化会降低知识的价值，而知识的复杂化会降低知识的易用性。因此，需要一个能够有效管理领域知识的载体。知识图谱的出现，不仅可以将信息表达成更近似人类认知世界的形式，而且提供了一种更好的组织、管理和利用海量复杂数据的方式。现在基于知识图谱的大数据应用已经开始慢慢渗透到各行各业，例如，互联网金融中的反欺诈应用、企业的精准营销应用、生命科学中的药物发现应用、电信行业的客户关系发现应用等。预期未来基于知识图谱的大数据应用将会渗透到更多领域和场景。

3 大数据发展专项调研

3.1 最令人瞩目的应用领域

大数据的发展最直接的推动力来自于应用，最近5年大数据“最令人瞩目的应用领域”的专项调研结果见**表2**。前三甲一直都是互联网和电子商务、金融、健康医疗，但2018年金融超越互联网和电子商务，排名上升到首位，这在调研中还是首次出现。此外，城镇化和智慧城市的得票数也有所上升，其他选项的得票数与前四名相去甚远，不足以出现在排名中。这反映出随着国家智慧城市建设的推进，面向智慧城市的大数据应用受到了更多的关注。

3.2 取得应用和技术突破的数据类型

所谓“取得应用和技术突破的数据类型”是指当前的分析技术和应用形态还不成熟、在未来一年最有可能取得突破性进展的数据类型，最近4年的预测结果见**表3**。其中对2018年的预测集中在城市数据和视频数据，排名第三的语音数据及后续项目的得票数与前两名相去甚远。这可能与对这两类数据的处理还没有成熟的应用模式有关，而对语音、互联网、图形图像等数据的处理技术和应用模式已相对成熟，要想取得新的突破难度更大。

3.3 与大数据最匹配的概念

本项调查结果见**表4**。在对2018年的

表 2 大数据应用最令人瞩目领域调查结果对比

序号	2014年	2015年	2016年	2017年	2018年
1	互联网；电子商务	互联网；电子商务	互联网；电子商务	互联网；电子商务	金融
2	金融	金融	金融	健康医疗	互联网；电子商务
3	健康医疗	健康医疗	健康医疗	金融	健康医疗
4	舆情分析；情报分析	城镇化；智慧城市	城镇化；智慧城市		城镇化；智慧城市
5		社会安全；犯罪调查	舆情分析；情报分析		

表 3 取得应用和技术突破的数据类型调查结果对比

序号	2015年	2016年	2017年	2018年
1	社会化媒体数据	城市数据	城市数据	城市数据
2	视频数据	互联网交易相关数据	图形图像数据	视频数据
3	互联网日志与电商交易数据	企业数据	语音数据；视频数据	语音数据
4	语音数据；图形图像；设备测控数据	视频数据；图形图像数据		互联网公开数据
5	人体数据；宏观经济	人体数据		图形图像数据

预测中，“机器人和人工智能”得票数远高于其他候选项，以至于其他选项都不足以出现在该统计表中。人工智能呈现出的“一边倒”的优势，也反映了正处于风口上的人工智能的火热程度。

3.4 我国大数据发展的最主要推动者

本项关注到底是什么样的力量在推动大数据的技术、产业、应用的发展，调研结果见**表5**。可以看出，除了大型互联网公司和政府机构，其他的推动者都已经先后淡出了这个名单。这说明大专委的专家们已经形成了较为一致的看法：能够推动大数据发展的，要么是具备资金、技术和数据优势的互联网公司，要么是具备政策影响力的政府机构，其他机构对大数据发展的推动力都十分有限。

3.5 大数据发展阶段判断

本项借用Gartner技术成熟度曲线中对技术发展阶段的划分，评估人们对大数据当前发展阶段的看法，见**表6**。从这6个阶段的投票分布来看，第二阶段（即将快速增长）和第六阶段（稳步成长中）占投票数的63%，对比2017年的预测集中度有了进一步的提升，这表明整体上大专委的专家对大数据的发展前景持更加乐观的态度。事实上Gartner从2015年起，已经不在每年的新兴技术成熟度曲线中给出大数据的位置，Gartner对此的解释是大数据已经快速发展成为一项各个领域通用的基础技术，因此不再作为新兴技术进行定位。大专委的专家们给出的发展阶段判断与Gartner的判断有

表 4 与大数据最匹配的概念调查结果对比

序号	2016年	2017年	2018年
1	互联网+	智能计算或认知计算	机器人和人工智能
2	云计算	云计算	–
3	智慧城市	机器人和人工智能	–

表 5 我国大数据发展的最主要推动者调查结果对比

序号	2015年	2016年	2017年	2018年
1	大型互联网公司	大型互联网公司	大型互联网公司	大型互联网公司
2	政府机构	政府机构	政府机构	政府机构
3	国内大学和科研院所	创业企业	–	–
4	公共服务机构	–	–	–
5	创业企业	–	–	–

表 6 大数据发展阶段判断调查结果对比

发展阶段	2015年	2016年	2017年	2018年
极为初级	17%	33%	16%	14%
即将快速增长	31%	40%	24%	33%
爆发增长中	10%	9%	23%	16%
达到一个顶峰，上升乏力	18%	4%	7%	5%
达到一个顶峰，将下降和幻灭	5%	0	4%	1%
稳步成长中	20%	14%	26%	30%

一定的一致性。

4 结束语

本文介绍了CCF大专委对2018年大数据发展趋势预测的结果，并将最近几年的预测结果进行了对比分析，以便读者能够全面地了解大数据的发展趋势。

当前在各个领域通过采集、分析和运用数据提升能力的行为越来越普遍，大数据已经真正成为众多行业的底层关键技术。在国家战略层面，新一届政治局在2017年底就实施国家大数据战略进行了第二次集体学习，习近平总书记强调要通过大数据进行产业创新、打造数字经济、提升国家治理水平、改善民生以及保障国家数据安全。期待国内的大数据产业和技术能够实现快速、良性的发展，为社会创造更多的价值。 □