

大数据已成为基础通用技术

Big Data: the Basic General Technology

王德政/WANG Dezheng
汪绍飞/WANG Shaofei
王梅/WANG Mei

(中兴通讯股份有限公司,广东深圳
518057)
(ZTE Corporation, Shenzhen 518057, China)

中图分类号:TN929.5 文献标志码:A 文章编号:1009-6868(2018)05-0054-003

摘要: 大数据已在各行业中落地应用,从某种意义上说,已成为“隐身”基础通用技术。一方面数据需要被更严密地组织与管理,才能切合各行业上层应用的需要;另一方面大数据需要与人工智能(AI)、物联网(IoT)等技术紧密结合,才能更好地为行业应用挖掘其自身价值。大数据与数据治理、人工智能、物联网等技术的融合,将促进大数据技术的普及,为行业应用的开发降低成本。

关键词: 大数据;数据治理;AI;IoT

Abstract: Big data has been applied in various industries, and it has become the "invisible" basic general technology. On the one hand, the data needs to be more closely organized and managed to meet the needs of upper-layer applications in various industries. On the other hand, big data needs to be closely integrated with artificial intelligence (AI) and Internet of things (IoT) to better excavate its value for industry applications. The integration of big data and data governance, AI, and IoT will promote the popularity of big data technology, and reduce the costs of industry applications.

Key words: big data; data governance; AI; IoT

1 大数据技术的发展趋势

大数据已经成为一种通用的基础技术。作为基础数据平台,一方面数据需要被更严密地组织与管理,才能切合上层应用的需要;另一方面大数据必须与各行业紧密结合,才能被深度挖掘出价值。

各行各业在数据组织与管理上普遍存在数据标准不统一、数据质量低、数据管理困难等挑战。这些挑战一方面需要通过管理手段来解决,另一方面也需要通过合理的技术手段进行辅助。

大数据作为基础技术与基础平台,必须与各个行业应用相结合,在解决实际问题中创造价值。从行业角度看,几乎所有的行业都可以结合大数据构建更好的系统解决方案;从技术角度看,大数据与人工智能、物联网之间有较为明显的技术融合趋势,未来这些领域将以数据为纽带,一体化协同发展。

2 大数据治理技术的发展趋势

大数据应用在其组织与管理上

收稿日期:2018-07-20
网络出版日期:2018-10-13

遇到了很多困难与挑战:

(1)数据缺少业务定义,难以进行业务分析;

(2)数据往往来自不同的系统,缺乏统一的标准,互通受阻;

(3)数据质量参差不齐,其完整性、准确性、一致性、时效性难以保证,分析结果的可信度大打折扣;

(4)数据间的内在关联未能建立,跨业务、跨领域的分析难以开展;

(5)数据使用中安全与隐私保护不足,存在违反法律法规的风险。

大数据要真正成为企业的核心资产,其治理是必由之路。从2017年伊始,大数据治理已成为产业生态圈中的研发重点。不少企业正在积极开展实践,一般步骤为“建立组织架构和规范→梳理应用需求→梳理企业数据信息→引进大数据治理

技术平台→治理数据”,治理后的数据成为企业资产为数据应用与数据运营提供基础。

大数据治理的最终目标不仅仅是为了管理数据,更是为了应用数据。数据只有被使用不断流通,才能最大限度地发挥其价值。因此大数据治理不能只关注数据本身,还要面向业务需求,根据用户所需开展治理活动。大数据治理是一个系统工程,总体来看需要具备6方面的关键能力与技术:

(1)统一元数据管理。元数据指“所有系统、文档和流程中包含的所有数据的语境,是生数据的知识”^[1],统一元数据管理已成为大数据治理的重要一环,为大数据的质量提升提供基础,使大数据的维护管理工作更加有效。

(2)数据标准管理。大数据治理要能够有序开展,必须建立统一的数据标准,为元数据的统一、数据的集成融合、数据的质量提升等提供依据。

(3)大数据质量管理。数据质量是数据应用的基础,通过大数据质量管理,可以获得干净、可靠的数据。这是大数据治理的重要目标,也是发挥大数据价值的必要前提。

(4)主数据管理。主数据是跨系统、跨模块、跨部门、跨地区、有高质量要求、高时效要求、被各项业务反复使用的基础性和敏感性数据。主数据是业务信息系统的神经中枢,是业务运行和决策分析的基础,也被认为是黄金数据^[3]。通过对主数据的管理,保证了其完整性、一致性、精确性、及时性,这样才能更好地支撑跨部门、跨应用的数据融合的一些应用需要。

(5)大数据集成^[4]。大数据集成不仅仅是将数据在物理上集中存储起来,还要依据数据标准,统一元数据的定义,将外部数据加工转换为业务所需要的目标数据,建立数据之间的内在关联。

(6)大数据安全与隐私保护。数据信息化为人类生产和生活带来便利的同时,也带来前所未有的数据安全与隐私威胁,大到国家安全、企业经营,小到个人隐私,都需要从不同的角度加强数据安全与隐私威胁。

通过大数据治理,获得及时、准确、可靠、安全脱敏后的高质量数据,可为大数据深入广泛的应用、企业的数字化转型提供强有力的抓手。

3 大数据与 AI 协同的发展趋势

人工智能(AI)在广义上是指任何能够让计算机通过图灵测试的方法和系统;而狭义上则是指通过研究人类智能产生的方式来让电脑模拟人的智能。

进入 21 世纪,随着神经网络算法

的不断优化,面向图形处理器(GPU)的编程接口带来了计算力的提升,这使得可以针对结构更复杂(多层神经元)的网络高效完成训练。传统的神经网络也因其复杂度和层数大幅度增加而改名为深度学习。我们可以把深度学习理解为以数据为基础的复杂神经网络学习系统,是传统神经网络在数据模式驱动下的演进和发展^[4-5]。

Google 在 2016 年将其战略从“移动优先”转变为现在的“AI 优先”。在移动时代,通过其知识图谱、自然语言处理、翻译、语音识别、图像识别、地图等相关产品积累大量的数据和技术,为今天 Google 的 AI 优先战略构建坚实的基础。所以,Google 本质上是一个以数据为基础的公司,是一个大数据公司。

我们可以看到:随着互联网/移动互联网的发展,数据量迅速增加。云计算和大数据的兴起,使得计算机存储和处理数据的能力快速提升。从某种意义上说,大数据为 AI 提供数据处理能力,而 AI 为大数据提供应用场景。例如:当数据治理涉及跨多个系统与业务时,往往需要借助 AI 技术以提升其处理能力,才能满足快速激增的海量数据以及快速发展的大数据应用需要。

大数据与 AI 相互融合,已经成为事实。AI 技术通过大数据,获取突破性成果,而以 AI 技术为特征的大数据应用则遍地开花,逐步渗透到各个行业和各个领域。我们面对的是一个以大数据应用为标志,以人工智能技术为特征的新时代。为客户提供融合大数据平台的 AI 解决方案是适应当前技术发展趋势并且具备良好的市场需求的产品。

4 大数据与 IoT 协同的发展趋势

物联网本身不是新概念,在 20 世纪 90 年代就已被提出。最近,以物联网为基础的智慧生活、智慧城市、

智慧地球等设想正在不断成为现实,物联网成为各技术巨头竞相布局的技术高地。

物联网广泛融合了大量的现有技术,涉及到通信、大数据、人工智能、数据挖掘、云计算、自动化、电子、材料等众多领域。其最核心的特性是连接与数据:在连接方面,物联网拓展了传统通信网络的功能和范围,将其延伸到更为广泛的物理世界;在数据方面,物联网接入了种类繁多的海量设备,极大地拓展了网络信息数据的来源渠道。据统计,新近全球创建、获取和复制的数据总量中,20% 来自物联网,而且增速最快。凭借越来越无处不在的连接、越来越丰富的数据,物联网正成为各类技术进步的新动力和助推器^[6]。

简单的、局部的物联网孤岛应用,其数据类型简单、数据量小,很难形成规模和产业效应,影响力极其有限。而大数据存储、大数据分析、云计算、人工智能等新兴数据存储和处理技术的出现,满足了物联网的大数据存储、智能化处理的要求,大大加速了物联网的前进步伐,加快了物联网孤岛应用的融合。

在存储方面,大数据拥有丰富的分布式云存储系统,可以满足物联网在大规模数据存储方面的要求^[7]。例如:Hadoop 分布式文件系统(HDFS)、分布式列存储系统(HBASE)、亚马逊 S3 云存储、微软 azure 云存储等。

在处理方面,面对不同的应用场景,大数据提供了不同的大规模数据处理框架,可以实现对物联网大规模数据的离线和实时分析,发掘物联网大规模数据中更多潜在的价值,催生物联网更多的应用,如:MapReduce、Spark、Storm 等。

在交换及共享方面,为解决物联网应用碎片化、孤岛化的痛点,业界相关标准化组织正致力于对物联网的数据模型和业务流程进行规范化和标准化,如:开放移动联盟(OMA)提出的轻量级机器到机器(LWM2M)

架构;欧洲电信标准化协会(ETSI)发起的、由多个标准化组织共同成立的物联网国际化标组织 oneM2M 提供的 oneM2M 架构;由微软、英特尔、三星、高通、思科等多家企业组成的开放连接基金会(OCF)提出的物联网设备标准等。这些架构和标准的提出,有力促进了物联网的发展^[8]。

物联网提供了广泛海量的连接,可以获得大规模的数据;而大数据以及 AI 的相关技术提供了智能分析能力,包括机器学习以及深度学习,可以用于分类、预测、自动决策以及视频、图像、语音的识别等。随着物联网采集数据的增多,智能化程度的提高,从这些数据中挖掘更多价值的需求就更加迫切,物联网与大数据之间的技术融合将更加紧密。

5 结束语

一个技术的成熟,只有当大众意识不到其存在时,才算是真正成熟,例如:历史上文字的发明、金属冶炼的发明等。这些深刻改变人类社会的技术,在当今社会或被“视而不

见”,或被视为如砂石一般理所当然存在的外部环境。

随着市场应用的深入,大数据与其他技术结合得越强,其作为基础技术的特征也就越显著。或许在不久的将来,大数据将“无迹可寻”,但同时又无处不在。

致谢

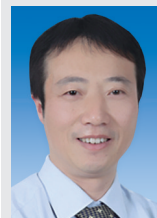
本篇文章得到中兴通讯网管及服务规划部高级工程师郭海生、工程师周永康的帮助,谨致谢意!

参考文献

- [1] 麦考.元数据仓储的构建与管理[M].北京:机械工业出版社,2004
- [2] 中国信息通信研究院.数据资产管理实践白皮书[C]//大数据产业峰会.北京:中国信息通信研究院、数据中心联盟,2018
- [3] 董欣.大数据集成[M].北京:机械工业出版社,2017
- [4] 孙运雷.物联网服务质量动态保障方法研究[D].北京:北京邮电大学,2014
- [5] 邓雪峰.设施农业物联网系统建模与模型验证[D].北京:中国农业大学,2016
- [6] 刘畅.物联网中多层/跨层接入管理关键技术[D].大连:大连理工大学,2014

- [7] 李开复.人工智能:李开复谈AI如何重塑个人、商业与社会的未来图谱[M].北京:文化发展出版社,2017
- [8] 吴军.智能时代:大数据与智能革命重新定义未来[J].榆林科技,2017(1):66

作者简介



王德政,中兴通讯股份有限公司中心研究院总工;主要负责大数据平台的规划,曾先后参与中兴通讯 WCDMA 核心网、3G 平台、IMS、统一网管,以及大数据平台等系统的规划与研发工作。



汪绍飞,中兴通讯股份有限公司物联网平台项目经理;目前主要从事物联网平台方面的研发工作,曾参与软交换、IMS、大数据平台 DAP 和物联网平台的研制工作;发表论文 10 余篇,参与《大数据架构师指南》的编写。



王梅,中兴通讯股份有限公司网管及服务系统部平台项目经理;主要负责数据管理与数据治理领域,并曾参与公司统一网管、大数据平台 DAP 和政企 UOC 平台的研制工作。