

大数据环境下的管理信息系统发展研究

杨丽彬¹, 李海林^{1,2}, 张飞波¹

1. 华侨大学信息管理系, 福建 泉州 362021; 2. 华侨大学现代应用统计与大数据研究中心, 福建 厦门 362021

摘要

大数据时代的到来给管理信息系统带来了一定的冲击和挑战,因此有必要研究大数据对现代管理信息系统的作用及影响,使得管理信息系统的性能和社会服务作用得到进一步提升。首先从数据的内涵与特征来理解大数据的概念,接着从数据处理过程的角度来讲述大数据对管理信息系统业务流程的变化和影响,同时,从管理信息系统的构成来解析大数据对于管理信息系统的推动与发展作用。最后,针对基于大数据环境下的信息安全和信息人才短缺等问题进行了探讨,并提出了相关的解决方案。

关键词

大数据;管理信息系统;数据挖掘;信息处理

中图分类号:C931.6

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2016010

Research on development of management information system in big data environment

YANG Libin¹, LI Hailin^{1,2}, ZHANG Feibo¹

1. Department of Information Management, Huaqiao University, Quanzhou 362021, China

2. Research Center for Applied Statistics and Big Data, Huaqiao University, Xiamen 361021, China

Abstract

The age of big data brought the management information system some shocks and challenges. So it was necessary to investigate the effects and impacts of the big data for management information system, which made the performance and social service of management information system more powerful. The connotation and features were given in advance to understand the conception of big data and the impacts of big data for business process of management information system were illustrated. Meanwhile, the promoting functions of big data for management information system were analyzed from its constitution. Finally, some problems including information security and talent shortage were discussed in the big data environment, and some solutions for them were also proposed.

Key words

big data, management information system, data mining, information process

1 引言

近年来,信息技术迅猛发展,尤其是以互联网、物联网、信息获取技术、社交网络等为代表的技术发展日新月异,促使手机、平板电脑、PC等各式各样的信息传感器随处可见,虚拟网络快速发展,现实世界快速虚拟化,数据的来源及其数量正以前所未有的速度增长。根据市场研究公司IDC 2012年12月发布的《数字宇宙研究报告》(Digital Universe)显示,2012年全球数据总量达到2.8 ZB(相当于2 800亿GB),预计到2020年全球数据总量将会超出预期,达到40 ZB(相当于40 000亿GB),若以现有的蓝光光盘为计量标准,那么40 ZB的数据全部存入蓝光光盘,所需要的光盘总重量将达到424艘尼米兹号航母的总重量。从现在到2020年,基本上每两年数据量就会翻一番^[1]。而这些数据约80%是非结构化或半结构化类型的,甚至更有一部分是不断变化的流数据。因此,数据的爆炸性增长态势以及其数据构成特点使得人们进入了“大数据(big data)”时代。

如今,IBM、EMC、Oracle、Intel等跨国IT界巨头纷纷进军大数据行列,发布自己的大数据战略以及产品。大数据不单单是在IT行业引起大量热烈关注,各个国家对其也有很高的重视,有些甚至提高到战略发展的高度。比如,2012年3月29日,美国奥巴马政府就公布了“大数据研发计划(big data research and development initiative)”;2013年8月2号,澳大利亚政府信息管理办公室(AGIMO)发布《公共服务大数据战略》,提出了指导大数据发展的六大原则和实现目标拟采取的6项举措;2013年8月30日,日本总务省公布2014年的任务和拟采取的重要信息化措

施,其中就包括推进ICT增长战略,通过促进大数据和开放数据的利用来实现经济复苏;2013年10月31日,英国商务、创新和技能部发布《英国数据能力战略》,旨在使英国成为从数据中提取见解和价值的世界领导者。学术界更是呈现百家争鸣的热况,就中国而言,已经分别于2011年、2012年和2013年的7月举办了3次大数据世界论坛(big data world forum),同时2013年12月在北京举行了第一届中国计算机学会(CCF)大数据学术会议和国内最大规模、最具影响力的中国大数据技术大会(big data technology conference, BDTC)。另外,《Nature》、《Science》分别于2008年和2011年针对大数据发表专刊;《纽约时报》、《华尔街日报》也都推出大幅报道。在其他的电商行业、社交平台上也都有大数据的踪影存在。

从这些大数据的研究与应用可以看到,现今对于大数据的研究主要还处于宏观层面,部分的微观性研究也多是针对技术方面,极少有针对管理方面的研究。另外,大数据与其他相对已经成熟的技术结合的研究也是一块缺失。对于大数据的应用和研究仍处于初级阶段。

近些年由于各种新技术、新理论的不断涌现,管理信息系统也在随之改变。近5年来,射频识别(RFID)技术、地理信息系统(GIS)、面向服务的架构(SOA),尤其是物联网和云计算的运用发展,给管理信息系统带来了新的发展点。通过整理我国近5年在管理信息系统领域的部分学术研究成果,发现管理信息系统与新理论、新技术的结合给管理信息系统的发展带来了新的推动力。例如,物联网与管理信息系统结合应用于逆向物流管理方面,解决了逆向管理信息系统中数据采集更新以及共享的难题;将GIS与管理信息系统结合应用于城建部分的配电网建设,能满足配电

网“安全、经济、优质”的要求^[2]；SOA与管理信息系统结合应用于社区管理信息，使得系统具备了适合基层工作业务流程要求的能力，具有更强大的生命力^[3]；云计算与管理信息系统结合应用于高校资源管理，有利于降低教育信息化成本，加强信息的整合与共享^[4]；RFID技术与管理信息系统结合用于冷库的仓储管理，能够做到对冷库内货物进行实时跟踪和精确定位，使得冷库的管理走向透明化和自动化^[5]。另外，国家自然科学基金委员会管理科学学部管理科学与工程很早就开始资助管理信息系统的研究^[6]。例如，“十五”至“十二五”期间，“信息化条件下的管理模式与方法研究”、“信息系统技术对管理模式的影响规律”、“电子商务/政务管理的理论与方法”和“智能健康信息服务管理”等分别被纳入优先或重点资助领域和发展方向。

管理信息系统经过40多年的发展，已经成为了一套相对成熟的集技术、管理思想、系统等于一体的信息管理体系，并且成功地应用于各个领域。无论是在企业、政府、公共服务事业，还是在航空、交通等领域都有不俗的表现。大数据时代的到来给人们的生活带来方方面面影响的同时，也影响着管理信息系统的发展。对此，国家自然科学基金委员会信息科学学部在2014年项目申报指南中明确指出^[7]，将大数据管理系统评测基准的理论与方法作为大数据技术和应用中十大挑战性科学问题之一，并拟以重点资助。大数据对管理信息系统数据处理过程中的数据采集、清洗、集成、存储、分析以及数据可视化都产生了影响，并对管理信息系统的硬件、软件、人才培养以及管理思想方面都有一定的推动作用，也存在着不少的挑战，例如，人才短缺、数据共享、信息安全等相关问题；反之，管理信息系统的发

展也将会加快大数据的发展进程。

2 大数据影响信息的处理过程

管理信息系统是一个以人为主导的，综合利用计算机硬件、软件和网络设备进行信息的采集、传递、存储、加工、整理和挖掘工作，以提高组织的经营效率为目标的信息系统。信息的整个加工处理过程是管理信息系统中最重要的一环。大数据的出现使得信息的采集、存储、清洗、集成、数据挖掘等工作内容都发生了不同程度的变化。

2.1 数据的采集源与采集技术

大数据拓宽了数据的采集源。传统的数据大多是来自于人为主动产生的数据，比如QQ、微博、照片等。但是随着电子商务、物联网、互联网、大数据的发展，机器数据日益显现它的价值，这就导致了各种公司及机构开始重视收集机器数据，无论是散落在不同地理位置的传感器收集的数据、移动互联设备中的数据信息（如通过手机或平板电脑获得的个人位置信息）、网络点击数量（如淘宝的客户浏览数据）的数据，还是射频ID数据，都成为新的热门数据来源。其中射频识别数据（RFID data）可以算是大数据发展以来的新宠，将RFID嵌入产品，进行实时数据分析，是未来包括物流、交通、零售、动物保护研究等在内的物联时代的智能首选。

数据源的变化也带来了数据采集技术的变化。大数据推动下产生的新的数据采集技术主要体现在以下几个方面。

- 对于互联网企业，采用分布式架构的系统日志采集方法，满足每秒百兆字节的日志数据采集的要求。如Facebook的

Scribe, Scribe将信息从各个日志源分别收集,然后存储于中央存储系统中,等待集中式的处理分析,以“分布收集、统一处理”的方式来提高数据采集的速度。

- 对于网络上的数据,采用网络数据采集方法,主要是指垂直搜索引擎和网络爬虫等。由于数据的海量性,通常对于采集后的数据需要进行分拣和二次加工,所以要做到既快速又尽可能准确地把握数据的采集。例如,国内的“火车采集器”就是将垂直搜索引擎、网络雷达、信息追踪与自动分拣和自动索引技术结合起来的应用,达到海量数据采集与后期处理相结合的目的,以此应对数据的海量性。

- 对于网络流量,可以采用深度分组检测(DPI)、深度流检测(DFI)或是盲检测(BLINC)的宽带管理技术。现在也有学者提出将DPI和DFI技术相结合的联合检测方法,使得其具备检错和纠错的能力,提高网络流量检测的准确性^[9]。

- 对于保密性要求较高的企业经营数据和科学研究数据,采用系统接口方式以确保数据的安全。

2.2 数据集成与清洗技术

数据集成是用各种手段和工具,在已有的数据基础上按照一定逻辑关系对数据进行统一的规划和组织,以实现数据资源共享的目标^[9]。数据集成主要是将异构性的数据进行处理,以便在数据挖掘阶段能够进行统一的操作。对于大数据时代数据以非结构化、半结构化数据占主要部分的情况,数据快速、有效、实时、动态地集成成为一个重要研究方向。因此,分布式的协作策略、动态式的实时集成成为最新的信息技术方面的研究热点。

大数据时代用于企业数据挖掘的数据增多,而数据质量参差不齐,往往存在数据

冗余、不完整、稀疏甚至错误的现象。在数据挖掘分析前,先进行预处理,可以减少数据的错误率,这是保证数据挖掘成果高准确性和高效率的有效且必要的方法。而数据清洗就是这样一个减少错误和不一致性的过程。面对海量的数据,盲目收集所有数据显然是一个很大的负担,所以数据清理的研究趋势是能够从最小的数据资源里获取最准确、最正确的知识。

2.3 数据存储的发展

大数据时代,传统的数据存储系统遇到了瓶颈,已经不能满足数据指数型暴增所带来的海量存储的需求。数据存储正朝着横向和纵向无限扩容的方向前进,即对存储系统能力的提升和存储量的扩容,使得存储系统能够满足大数据“类型”的需求和“量”的需求。另外,为了提高系统的扩展性、降低系统维护复杂度,提出了以对象作为基础的存储形式。在这样的发展和需求背景下,云存储成为了首选。云存储就是通过应用软件将网络中大量存在的各种不同类型的存储设备集合起来,协调各设备的工作来满足大数据对于数据存储的要求。

云存储主要可以分为两类,一类就是生活中常会用到的一些个人的云存储应用,比如最初的酷盘,后来的360云盘、百度云等,主打个人资料的上传与共享。另一类是企业级的云存储应用,比如EMC推出EMC ATMOS云存储,主打大规模非结构化数据的存储、归档和访问,以面向云的设计,确保企业和服务提供商在保持庞大规模数据的同时高效运行;IBM推出的IBM SmartCloud Virtual Storage Center,是一种存储虚拟化平台,有助于快速迁移到基于云且面向服务的敏捷存储环境,支持数据的大规模增长,降低复杂性。

2.4 数据挖掘与分析

数据分析是管理信息最重要的一环,也是创造价值的一环。新类型的数据不断涌现,如实时数据、动态数据、时空数据、Web数据,这些都是值得关注的数据库。对应这些不同类型的数据,出现了不同的数据分析方法。具体来说,分为以下几种类型。

- 实时数据,作为一种动态的数据流,如微博、微信等,是大数据发展以来数据分析非常重要的一个方面。主要采用的是动态的数据流处理方式。

- 动态数据主要是通过时间序列从时间维度来分析发现相关的变化规律。零售、电信、金融行业对此都已经有了广泛应用,如零售行业通过分析历史数据在时间维度上的规律来发现企业潜在的商机,并预测未来客户群体以及客户的潜在需要,有针对性地制定营销策略和采购计划。

- 时空数据由于各种移动传感器、定位系统的应用而变成了热点。经过几年的发展,在时空挖掘领域已经出现了许多价值的工作,如时空模式发现、时空异常检测、时空预测和分类、时空聚类以及时空推理与数据挖掘相结合^[10]。

- Web数据,大数据时代对于互联网的重视尤为突出,因此大数据的数据量以及其重要程度使得Web数据研究技术进一步发展。这个发展就是基于云计算的Web数据挖掘,同时也为Web数据挖掘成功找到了适应当前网络需求的可行方案。这个需求指的是对于当前互联网海量数据的地理分布、异构、动态以及高复杂性这些特性的应用需求^[11]。

大数据对于数据分析方面的影响不单是在分析数据类型变化及相对应的数据挖掘方法的变化上,还给分析处理带来新

的要求——实时分析^[12]。从现有的技术和模式来看,主要是从流处理、批处理以及两者相结合的方面入手。数据的价值会随着时间的流逝而减少,基于此,流处理将数据当作流,将连续不断的数据当成数据流,获得实时近似的结果。MapReduce模型是批处理最具代表性的模型,通过廉价的服务器集群实现大数据的并行处理^[13],将问题分开处理,以此避免因大规模数据传输造成的通信负担。两者的结合是近些年来的一又热点,可以相互取长补短,使得数据分析能够更加实时化。

大数据时代的数据分析应用于方方面面,每个人的生活中或多或少都在应用着分析的成果,从试验性的接触到熟练掌握,人们渐渐发现数据分析带来的好处,对其需求也渐渐扩大。作为卖家,需要从各种营业数据里分析自己销售商品的未来走向,找出什么是卖点;作为买家,需要对各种产品的参数进行深度比较分析,获得最优于自己的购买方案。因此,数据分析已经不是大企业的专利,渐渐地应开始走向平民化。对于这点,IT巨头微软公司已经有了准确的认识。2013年10月,微软全球高级副总裁、大中华区董事长兼首席执行官贺乐斌在“大数据媒体日”发布大数据愿景:推动数据分析平民化,希望基于标准化的产品,使所有人都能够在任何时间、任何地点利用数据,并更好地做出决策。

2.5 辅助决策

在大数据背景下,人类的决策方式将不可避免地发生改变。大数据将把数学算法运用到海量数据上,通过让数据做主的方式进行科学决策,以此来修正人类的偏见和直觉。将信息系统和辅助决策进行有效的结合是大数据时代的必然结果,而这将为辅助决策带来新的变革。特别地,由

于信息之间的联系越来越紧密,孤立地看待问题必将被时代淘汰,单项决策支持的辅助决策在大数据时代能发挥的作用已经越来越小,综合、全面地看待问题,是大数据对决策者提出的新要求,也是决策支持在大数据时代的发展。信息系统的接入在消除信息孤岛的同时,也使得辅助决策实现了从单项决策支持向企业级决策支持的转变^[14,15]。

传统的辅助决策系统往往不能实现信息检索,而是主要进行智能决策。在大数据时代下,决策者将面对海量决策信息,信息检索的实现是帮助决策者进行信息分析与筛选的重要环节。信息系统现有的检索功能对这一环节的实现起着至关重要的作用,如何将信息系统的检索功能高效地运用到辅助决策中,是未来的系统开发者面临的挑战。

大数据时代的到来使得各行各业都不可避免地受到了冲击与改变,决策者在面临大量信息的同时也对辅助决策系统提出了更高的要求。信息系统对辅助决策的改变提高了辅助决策系统对海量数据的分析与处理能力,实现了从单一决策的制定到综合决策支持的改变,同时如何高效地兼并信息系统与辅助决策系统是每个从业人员面临的共同挑战,需要进一步研究。

2.6 数据可视化

数据展示是数据分析后将结果传递给人的一个过程,大数据分析得到的结果之间的关联关系更加复杂,传统的展示方式已经不能满足要求。大数据带来了更为直观、互动的方式,以便人们更好地理解。可视化技术的引入就是其中重要的一项。

可视化技术是将数据转化为图形,以此来给予人们深刻而意想不到的洞察力。可视化技术是能够放大人类感知的数据的图形

表达方式,聚焦于关键的信息特征,以压缩信息的方式使得复杂信息能够快速被人理解^[16]。同时,从数据到图像的转变也能从中发现单纯数据处理时不易发现的信息。信息设计师David McCandless 在TED上说道:

“通过可视化,可以把信息变成一道可用眼睛来探索的风景线,一种信息地图。当你迷失在信息中时,信息地图非常实用。”

3 大数据对管理信息系统基本构成的推动性

管理信息系统的主要功能包括:给企业提供能反映实际情况的信息来支持企业决策;辅助管理者监督控制来有效利用资源;用过去的信息预测未来和为不同管理层提供不同的报告来控制企业行为。管理信息系统是由硬件、软件和人构成,它的应用也代表了一种企业管理思想。所以,管理信息系统是硬件、软件、人以及管理思想的有机结合体,而大数据对管理信息系统的这几个方面产生了影响。

3.1 大数据推动管理信息系统的硬件发展

管理信息系统的硬件主要包括计算机硬件和网络管道两方面。大数据的海量数据特性首先冲击的是计算机硬盘的存储容量,而其庞大的数据量所需的计算次数又给计算性能带来了新要求,另外,半结构化与非结构化的数据占多数的特性也要求数据处理方式的转变。

CPU的架构已经从单核转向多核,现今的发展主要是通过软件架构来完美组合多核CPU,使得能够更好地利用多核CPU的并发机制来展示价值。对于硬盘存储,磁性介质硬盘(HDD)因其读写速率的限

制,不能很好地适应大数据的要求,进而出现了基于闪存的固态硬盘(SSD)。固态硬盘有许多独特的优点,比如它的抗震性、低耗性、小体积,尤其是良好的读写性能,能为管理信息系统提供优良的存储性能。对于内存部分,典型代表PCM-SCM的出现,使得内存同时具备了处理速度快和不容易丢失的内存与磁盘的两方面优点。

IDC曾预测“大数据会显著推动基础架构横向扩展”。显然,硬件企业已经认识到这一点,并开始积极地靠近。以Intel公司为例,Intel公司为实现容忍网络隔离而推出具有良好横向可扩展性的x86架构,另外在存储领域里也推出了基于开放架构的云计算体系。Intel公司的至强处理器强劲的计算能力,加上数据直接I/O技术(DDIO)性能和SATA接口的固态硬盘(SSD)可以提高原始存储的读写率,更好地适应并满足大数据的存储、处理速度要求。另外,大数据的发展和市场需求,势必引起大数据一体机研发热潮,使得大数据存储、处理、软硬一体化得到充分利用,进而更好地为大数据研究和发展提供硬件平台支持,这在一定程度上解决了原有架构的扩展瓶颈,进一步推动了大数据向产业化发展,并使得大数据技术在相关行业中得到充分应用。

管理信息系统的硬件还有重要的一个方面就是网络管道。大数据对于数据传输速度更快、数据量更大的要求,促使管道技术向宽带化演进。另外,在管道架构方面,为了实现系统时延性的降低,网络管道趋于扁平化。软件定义网络和多管道组合技术也将使管道技术走向虚拟化与智能化。

3.2 大数据推动管理信息系统的软件发展

管理信息系统的软件方面主要包括操

作系统、数据处理系统、数据库系统、应用软件等,可以从文件系统、数据处理系统、数据库系统这几方面看到大数据影响下的管理信息系统的软件发展。

文件系统是一种存储和组织计算机数据的方法,是操作系统用于明确磁盘或分区上文件的方法和数据结构,使对其访问与查找更加明确与便利。现如今的文件系统基本已经都采用分布式文件系统,经过这几十年的发展,该系统技术已经趋于成熟。在大数据的冲击下,分布式文件系统应大容量、高性能、高可用等要求出现了Hadoop分布式文件系统(HDFS),它的优点是高度容错性和高吞吐率,并且可运行在通用硬件上。另外一个比较热门的分布式文件系统是QFS(quantcast file system),这是一个开源的文件系统。与HDFS相比,它的性能更加优异,从容量来看,QFS可以节省50%的磁盘空间;从吞吐率来看,QFS的读写速度是HDFS的两倍;另外,QFS更容易与系统软件兼容,实现命令式的终端。

数据处理系统模式主要就是批处理与流处理两种。而现代数据处理最重要的就是数据处理的实时快速,也就是数据实时处理。对此,出现了内存计算,希望以此达到实时处理或是近似于实时处理的效果。内存计算的意思就是在服务器的内存里处理超大量的实时数据,从而达到在分析和交易中提供即时结果的效果。另外,大数据时代下的数据挖掘算法研究成为管理信息系统决策支持的研究重点。深度学习是大数据时代又一重要的人工智能方法,是一种深度神经网络的机器学习模型,其成为了现代人工智能研究的发展方向。然而,大数据背景下的深度学习具有模型庞大、计算量大、耗时长等特点,使得分布式管理系统通常需要并行计算来加速数据的处理与分析。

大数据的数据规模性、低密度价值性、类型多样性等特性及数据在不同领域应用的时间、处理方式的差异,推动了数据库系统的前进。现今, NoSQL和NewSQL数据库阵营迅速崛起。NoSQL就是“not only SQL”, 是对关系型数据库的一种补充, 即非关系型数据库。NoSQL指的是非关系型数据库, 首要解决的问题就是如何对大数据进行有效处理。NoSQL普遍采用简单的数据模型, 通过元数据和应用数据分离技术以及弱一致性技术来处理大量数据, 实现较高的吞吐率, 并可以应用于便宜的PC服务集群上, 实现可扩展性^[17]。NewSQL是支持关系型数据库的, 同时以SQL作为其主要的接口, 但是内部结构产生了较大变化: 它丢弃了原有的单线程服务的锁制和昂贵的恢复机制, 采用冗余机器来实现复制和恢复故障。如今的数据库系统已经在这两个概念领域里呈现出百花齐放的景象。

3.3 大数据推动新体系架构的研发

信息系统体系结构是对信息系统各构成要素及其关系的描述, 信息系统体系结构包括信息系统的概念、信息系统的基础设施架构、信息系统的信息资源结构和信息系统的软件架构等。基于云计算的军事信息系统架构^[18]以云计算资源管理“虚拟化”分布式数据管理和分布式数据处理等作为基础支撑技术, 对已有军事综合电子信息系统体系架构进行改造, 能够满足信息系统的高可靠性和高可扩展性要求, 该体系结构为了对数据进行高效的利用, 发生了以下改变: 采用分布式数据处理技术, 大幅提高了大规模数据访问和情报处理容量, 提高了系统处理的实时性、可靠性; 采用虚拟数据存储的方式, 提高系统访问数据实时性和可靠性, 降低系统反应

时间, 提高作战效能。

除了以上几个适应于大数据的信息系统体系结构框架之外, 国内外还有很多学者和组织都在这方面做了许多有益的尝试, 使得信息体系结构具有更好的数据处理与分析性能。如今数据就是新能源, 高效地利用数据已经成了企业乃至国家的重点研究课题, 研发具有更高大数据适应性的信息系统架构具有很高的社会及现实意义。

3.4 大数据推动数据收集分析的人才发展

由于现今用于各种大数据分析的策略性产品还没有完全成熟, 不能单靠一家大数据服务供应商解决所有问题。如果一家企业外购这类产品, 要做的并不只是简单的购买、安装、运行, 更重要的是要有优秀的系统管理员来持续这样的应用过程, 这样才能使系统发挥应有的效能。只有充分地了解建立这些环境的实现方法, 才能应用科学系统为企业服务。管理信息系统重要的一个部分是数据挖掘, 而在大数据时代这点尤为突出, 即怎样利用现有的系统从众多类型各异的数据中找出真正的价值。而采集数据和清洗数据作为数据分析的前提也显得尤为重要。

大数据时代的管理信息人才要掌握整个系统运行的知识、大数据应用的基本原理以及管理信息系统的整个管理思想和运作模式。除此之外, 还要掌握的就是数据挖掘的能力, 要有掌握机器学习、知识图谱的应用能力, 能够熟练应用各种数据挖掘模型以及统计学的专业技能等。总而言之, 管理信息人才要利用现有的大数据产品, 做好信息管理工作, 从海量的大数据中发掘有用信息。

企业里面大多数人都是管理信息系统的用户, 但基本都缺乏上述的能力, 所

以大数据时代的企业具备专业的技术和管理人员更加重要。大数据推动了专业数据方面的技术和管理人才的发展。

3.5 大数据推动企业管理思想转变

就管理信息系统本身而言,它通过信息收集、存储、传输、加工和输出等实现辅助企业事务处理和管理职能以达到管理的目的。所以,以往一般企业的决策只是将管理信息系统做一个辅助参考,并不以数据挖掘的结果为真正的决策,而主要采用以目标驱动决策的方法进行决策。但是大数据给决策带来了新的思潮——数据驱动决策。根据美国麻省理工学院的一个团队对北美330家企业调查走访的结果^[19]可知,把自身企业定义为数据驱动的企业,能更好地完成既定的财务和运营目标,尤其是处在行业排名前1/3的企业。数据驱动企业平均比非数据驱动企业的生产力高出5%,利润高出6%。可见,数据驱动能给企业发展带来推动作用。

数据驱动决策并不是说管理者可以完全依赖数据,管理者要做一个提问的专家,要清楚地理清企业脉络并具有长远发展眼光,能够总体控制方向,然后由数据来告知答案。比如,零售商的管理者会提出:谁是优质客户;什么样的广告可以刺激购买等。当然也有些问题不是数据可以回答的,比如,企业在竞争对手中的地位。总而言之,一些由数据可以回答的问题在渐渐地表达数据本身的意见,最终实现数据驱动决策。

有学者提出,将传统的“目标驱动策略”与“数据驱动策略”结合起来,形成双向决策模型,从而构建了“评估与预测”和“监测与预警”的技术创新管理模型^[20]。这是管理决策思想的另一个创新,新与旧的结合使企业能更好地从传统“目标驱动

决策”逐渐转向“数据驱动决策”,让企业管理者逐渐相信数据说的“话”。

4 问题与措施

大数据在给管理信息系统带来推动作用的同时也带来了许多问题,最突出的问题为以下几个方面。

4.1 数据安全问题与相应的措施

数据安全问题作为大数据冲击下的产物,对于管理信息系统也尤为重要。由于信息管理系统里的数据都是企业的重要资产,尤其是一些核心信息,一旦泄露就会给企业带来巨大的经济损失。然而,很多管理信息系统都涉及跨部门设计,甚至有可能链接外部数据或使用云端,导致数据存在着安全隐患。

面对这样一个挑战,可以从两方面着手,一方面是从技术角度,完善管理信息系统;另一方面是从管理角度,建立完备的管理制度。

从技术上保障数据安全可以从以下3个方面入手。

(1) 建立完善的外部环境,加固受攻击面

大数据时代的管理信息系统的外部环境往往依赖于网络,数据存储一般都放在云端。操作系统漏洞或网络环境问题有可能威胁整个系统。如亚马逊EC2业务在2008年2月15日的云计算宿主离线导致的事件:大规模的服务中止,并抹去了一些客户的应用数据。所以要加强操作系统和网络环境的安全保障,对此,可以从网络资源分配、防火墙建设、入侵检测技术以及及时打补丁与杀毒来实现。如惠普用Fortify进行代码扫描来保证代码的无安全漏洞性;整合应用生命周期,将传统惠

普服务器、惠普云、数据中心解决方案及防病毒、防火墙产品整合,达到加固受攻击面的效果。

(2) 对于系统内部要严格做好安全措施,加强风险管理

云安全技术是这一方面比较先进的技术。云安全也是基于“云”的创新,主要是通过网状的客户端对网络中的各种软件进行检测来发现其中的异常,最终达到将各种病毒和木马的解决方案与各个客户端共享的目的。它融合了网络计算、并行处理以及未知病毒的判断等诸多新兴的技术和概念。为了避免一旦出现安全问题,系统无法恢复数据的问题,要及时、有效地做好数据的备份。另外,要建立完善的数据库访问机制,避免关键数据的不必要外泄。

(3) 要主动防护信息,变被动为主动,从企业安全环境的评估、转型、优化以及管理入手,全面做到智能防护

如惠普推出的惠普数据中心安全防护服务——HP ArcSight ESM 6.0c和惠普TippingPoint新一代入侵防御系统等产品 and 解决方案,都是以智能主动防护为目标的。

从管理上保障数据安全,可以从以下4个方面入手。

(1) 完善数据资源保障制度

根据现有的国家信息安全标准来建立企业的安全保障制度,这个安全保障制度包括数据资源安全的评估流程和评估标准、针对日常安全维护工作的维护机制以及应对突发情况的快速响应机制。

(2) 完善数据资源共享安全制度

根据企业数据资源的机密程度划分不同的等级,依据划分的等级来设置企业数据资源的访问权限,以此建立相应的访问制度。存储在云端的数据,按照数据级别分区域存储。

(3) 完善机密数据资源保护制度

设置企业信息机密区,减少人员的接

触,并制定机密数据接触使用的相关规定;对于可接触人员签订相应的保密协议,防止泄密事件发生;对于技术上的存储安全措施(防火墙的设置)设定相应的制度,从制度上要求执行。

(4) 完善数据资源审计制度

建立定期对系统进行检测与安全评估的制度,从制度上落实数据资源的审计工作;对于存储于云端的数据,引入第三方信息安全审计机制,完成云端存储的风险评估工作^[21]。

4.2 专业的技术和管理型人才缺失与相应的措施

现今,大数据的人才处于紧缺状况。据麦肯锡全球研究院(McKinsey Global Institute)研究预测:在未来几年,仅在美国本土就可能有将近14万~19万的具备深入分析数据能力人才的缺口出现,至于能够通过分析大数据来为企业做出有效决策的数据管理人员和分析师的缺口将达到近150万人。

面对这样的现状,企业可以选择不同的应对措施。

- 企业外包。普通的金融、能源公司可以选择外包的形式,选择相对成熟的大数据服务厂商。

- 内部培训。对于企业内部的数据管理人员与分析师,企业可以采取技术培训的方法。

- 与高校人才合作。网络公司需要从根本上采取措施,不单单是员工的技术培训,可以从人才引进方面着手:与高校合作,培育企业所需要的人才。比如,IBM公司的“全球千所高校合作学术计划”,学生可以在学校通过专业课程的培训获得最基础的专业知识,包括市场营销、数学、统计学等;而同时通过IBM公司,学生可以获得

大数据分析软件、课程教材、研究案例等,通过IBM公司的数据专家针对性的专业讲座来提高专业水平。这项计划旨在培养学生的大数据和分析学方面的能力,使之能胜任大数据的相关工作。

4.3 其他问题

除了上述数据安全问题 and 人才紧缺问题之外,大数据的发展还引出了许多其他方面的问题。在数据处理方面,虽然已经有了新的技术突破,但是大数据应用的问题并没有完全解决。例如,怎样有效而快速地做好大量数据的清洗工作;怎样掌握“量”与“质”的平衡(若是清洗过程中粒度太粗,就可能无法达到清洗的预期效果;若是清洗过程中粒度太细,就可能将有用的信息过滤掉);实时处理方法虽然效果不错,但是该方法对应用的数据类型有一定的限制,如何将这些数据处理方法有效结合,创造出一个通用型的实时处理方式来打破这个瓶颈;数据中心的大部分电能用来维持服务器闲置时的供电,如何解决这方面的能耗问题;硬件方面,新旧硬件的交替期可能出现的木桶效应和不匹配问题。这些问题都有待解决。2013年6月,美国爆发的“棱镜事件”带来的是隐私方面的不安全感。信息公开共享与隐私的矛盾问题也摆在了眼前。

5 结束语

大数据确实给人们的生活带来了方方面面的影响,也影响着现代管理信息系统的发展,同时,大数据新兴概念、技术和方法给管理信息系统的创新和应用增添了新的视角,对管理信息系统的数据处理流程和相关基本构成都产生了积极的推动作用。

现如今,对于大数据的研究和应用尚处于初级阶段,管理信息系统研究者应该紧跟历史潮流,抓住机遇,做好一切准备,实现现代管理信息系统处理大数据的能力,使其能更好地为高校、企业和政府提供决策性支持,进而增加它们的创造力和竞争力。然而,对于大数据带来的人才转型、数据安全、新旧硬件技术的磨合以及能源等方面的问题,管理信息系统开发者都应该重点关注,这也是管理信息系统发展中的重要研究方向。

参考文献:

- [1] GANTZ J, REINSEL D. The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the FarEast[EB/OL]. [2012-12]. <http://www.emc.com/leadership/digital-universe/2012view/index.htm>.
- [2] 贺超, 庄玉良. 基于物联网的逆向物流管理信息系统构建[J]. 中国流通经济, 2012(6): 30-34.
HE C, ZHUANG Y L. Study on reverse logistics management information system based on internet of things[J]. China Business and Market, 2012(6):30-34.
- [3] 温平川, 邹晓晖. 基于SOA的社区管理信息系统设计与实现[J]. 重庆邮电大学学报(自然科学版), 2011, 23(6):765-769.
WEN P C, ZOU X H. Design and realization of community management information system based on SOA[J]. Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition), 2011, 23(6): 765-769.
- [4] 孙凌宇, 欧阳春娟, 冷明, 等. 云计算与高等教育管理信息服务系统构建[J]. 山西财经大学学报, 2012, 34(1):251-257.
SUN L Y, OUYANG C J, LENG M, et al. Cloud computing and higher education management information service system[J]. Journal of ShanXi Finance and Economics

- University, 2012, 34(1): 251-257.
- [5] 张庆英, 何宁英. 基于RFID的冷库管理信息系统[J]. 武汉理工大学学报(信息与管理工程版), 2011, 33(4):553-555.
ZHANG Q Y, HE N Y. RFID-based cold storage management information system[J]. Journal of Wuhan University of Technology (Information & Management Engineering), 2011, 33(4):553-555.
- [6] 吴金南, 刘作仪. 基于国家自然科学基金项目的管理信息系统研究进展与分析[J]. 管理学报, 2013,10(8):1201-1207.
WU J N, LIU Z Y. Research progress on management information systems: an analysis of research projects funded by NSFC[J]. Chinese Journal of Management, 2013, 10(8):1201-1207.
- [7] 国家自然科学基金委. 2014年度国家自然科学基金基金项目指南[M]. 北京: 科学出版社, 2014.
The National Natural Science Foundation of China. 2014 Guidelines for National Natural Science Fund Project[M]. Beijing: The Science Publishing Company, 2014.
- [8] 叶文晨, 汪敏, 陈云寰, 等. 一种联合DPI和DFI的网络流量检测方法[J]. 计算机工程, 2011, 37(10): 102-104.
YE W C, WANG M, CHEN Y H, et al. Network flow inspection method of joint DPI and DFI[J]. Computer Engineering, 2011, 37(10):102-104.
- [9] 黄晓斌, 钟辉新. 大数据时代企业竞争情报研究的创新与发展[J]. 图书与情报, 2012(6):9-14.
HUANG X B, ZHONG H X. On the innovation and development enterprises competitive intelligence analysis in the big-data era[J]. Library and Information, 2012(6):9-14.
- [10] 刘大有, 陈慧灵, 齐红, 等. 时空数据挖掘研究进展[J]. 计算机研究与发展, 2013, 50(2):225~239
LIU D Y, CHEN H L, QI H, et al. Advances in spatiotemporal data mining[J]. Journal of Computer Research and Development, 2013, 50(2): 225-239.
- [11] 程苗. 基于云计算的Web数据挖掘[J]. 计算机科学, 2011, 38(10A):146-149.
CHENG M. Web data mining based on cloud-computing[J]. Computer Science, 2011, 38(10A): 146-149.
- [12] 孟小峰, 慈祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146-169.
MENG X F, CI X. Big data management: concepts, techniques and challenges[J]. Journal of Computer Research and Development, 2013, 50(1):146-169.
- [13] 覃雄派, 王会举, 杜小勇, 等. 大数据分析-RDBMS与MapReduce的竞争与共生[J]. 软件学报, 2012, 23(1):32-45.
QIN X P, WANG H J, DU X Y, et al. Big data analysis-competition and symbiosis of rdbms and mapreduce[J]. Journal of Software, 2012, 23(1):32-45.
- [14] 孙强, 张雪峰. 大数据决策学论纲: 大数据时代的决策变革[J]. 华北电力大学学报(社会科学版), 2014(4): 33-37.
SUN Q, ZHANG X F. An outline of the data decision making: the age of big data decision change[J]. Journal of North China Electric Power University (Social Science), 2014(4): 33-37.
- [15] 张祥. 大数据时代企业管理决策的发展探究[J]. 东方企业文化, 2015(24):171.
ZHANG X. Explore the development of big data era of enterprise management decisions[J]. Oriental Enterprise Culture, 2015(24):171.
- [16] 洪文学, 王金甲. 可视化和可视化分析学[J]. 燕山大学学报, 2010, 34(2):95-99.
HONG W X, WANG J J. Survey on visualization and visual analytics[J]. Journal of Yanshan University, 2010, 34(2):95-99.
- [17] 覃雄派, 王会举, 李芙蓉, 等. 数据管理技术的新格局[J]. 软件学报, 2013, 24(2):175-197.
QIN X P, WANG H J, LI F R, et al. New landscape of data management technologies[J]. Journal of Software, 2013, 24(2): 175-197.
- [18] 方建勇. 一种基于云计算技术的军事信息系统体系架构[J]. 计算机技术与发展, 2013(12): 235-239.

FANG J Y. A military information system architecture based on cloud computing technology[J]. Computer Technology and Development, 2013(12): 235-239.

[19] ANDREW M, ERIK B. Big data: the management revolution[J]. Harvard Business Review, 2012, 90(10): 60-68.

[20] 朱东华, 张巍, 汪雪锋, 等. 大数据环境下技术创新管理方法研究[J]. 科学学与科学技术管理, 2013(4): 172-180.

ZHUDH, ZHANG Y, WANG XF, et al. Research

on the methodology of technology innovation management with big data[J]. Science of Science and Management of S&T, 2013(4):172-180.

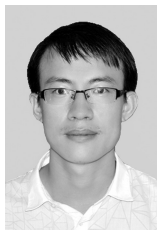
[21] 吴金红, 张飞, 鞠秀芳. 大数据:企业竞争情报的机遇、挑战及对策研究[J]. 情报杂志, 2013(1): 5-9.

WU J H, ZHANG F, JU X F. Big data: opportunities, challenges and strategies of enterprise competitive intelligence[J]. Journal of Intelligence, 2013(1): 5-9.

作者简介



杨丽彬(1982-), 女, 华侨大学信息管理系讲师, 主要研究方向为数据分析与知识管理。



李海林(1982-), 男, 博士, 华侨大学信息管理系副教授、硕士生导师, 主要研究方向为数据挖掘与决策支持。



张飞波(1993-), 女, 华侨大学信息管理系本科在读, 主要研究方向为管理信息系统和数据分析。

收稿日期: 2015-09-03

基金项目: 国家自然科学基金资助项目(No.61300139); 福建省社会科学规划项目(No.2013C018); 华侨大学中青年教师科研提升资助计划基金资助项目(No.ZQN-PY220)

Foundation Items: The National Natural Science Foundation of China (No.61300139), Social Science Planning Project of Fujian Province (No.2013C018), Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University (No.ZQN-PY220)