

大数据视域下的海洋信息化建设构想

宋德瑞^{1,2,3}, 曹可^{2,3}, 张建丽^{2,3}, 景昕蒂^{2,3}, 王相海¹, 孙书翰¹, 解鹏飞^{2,3}

(1. 辽宁师范大学城市与环境学院 大连 116029; 2. 国家海洋环境监测中心 大连 116023;

3. 国家海洋局海域管理技术重点实验室 大连 116023)

摘要:美国的“大数据研究与发展计划”提出后,引发了世界各国大数据研究的高潮。我国也在近期将大数据上升到国家政策的高度,大数据产业得到全面支持。海洋大数据作为海洋综合管理的基础,对于维护国家海洋权益、准确掌握海洋状况具有重要意义。文章在我国大数据技术快速发展的背景下,分析了国内外大数据技术发展的基本情况,提出了基于大数据资产化管理模式、大数据思维模式、大数据可视化模式来构建海洋大数据平台思路,进而挖掘海洋数据中隐藏的“智慧”,全面提升海洋综合管控能力。

关键词:大数据;数据资产化;海洋信息化;数据可视化;数据挖掘

中图分类号:G2;P7

文献标志码:A

文章编号:1005-9857(2017)09-0050-05

The Conception of Marine Information Construction from the Perspective of Big Data

SONG Derui^{1,2,3}, CAO Ke^{2,3}, ZHANG Jianli^{2,3}, JING Xindi^{2,3},
WANG Xianghai¹, SUN Shuhan¹, XIE Pengfei^{2,3}

(1. School of Urban and Environmental, Liaoning Normal University, Dalian 116029, China; 2. National Marine Environmental Monitoring Center, Dalian 116023, China; 3. State Oceanic Administration Key Laboratory of Marine Management Technology, Dalian 116023, China)

Abstract: After the presentation of America's "Big Data Research and Development Program", it aroused a surge in the world's big data research. China has also recently raised big data to the level of national policy. Big Data industry has been fully supported. Being a basis of comprehensive marine management, oceanic big data has important sense to protect national maritime rights and interests and accurately grasps the developing conditions of ocean. Based on the background of the rapid development of large data technology in China, this paper analyzed the basic situation of the development of large data technology both at home and abroad, and put forward the idea of building large ocean data platform based on large data asset management mode, large data thinking mode and large data visualization model. The hidden "wisdom" of ocean big data was also mined, to enhance the comprehensive marine management and control capacity.

Key words: Big Data, Data assets, Marine informatization, Data visualization, Data mining

收稿日期:2017-03-13;修订日期:2017-07-19

基金项目:海洋公益性行业科研专项(201005011、201405028);国家海洋局海域动态监视监测与评价业务化项目。

作者简介:宋德瑞,高级工程师,硕士,研究方向为海域监测业务信息化研究

随着互联网、云计算^[1]等技术的高度发展,移动智能设备的快速普及,爆炸式增长的大数据时代已经来临,据 IDC 的研究估计^[2]:到 2020 年,全球数据使用量将达到 35.2 ZB,大概需要 376 亿个 1 TB 硬盘来存储数据。2012 年 3 月美国奥巴马政府公布了“大数据研究和计划”^[3],提高政府从海量复杂数据中获取知识和远见的能力。继美国之后欧盟、日本、韩国等也纷纷提出大数据相关举措,法国政府在其发布的《数字化路线图》,日本也公布了发展开放公共数据和大数据为日本新 IT 国家战略的核心。我国于 2012 年批复了“十二五”国家政务信息化建设工程计划,开始关注大数据的研究,并构建我国大数据产业链和大数据研究平台,2014 年上海市率先实行政府部门数据对外开放。

随着各国大数据国家战略的实施,大数据技术成为当前最热的关注研究领域。2003—2006 年谷歌工程师发表了有关 Map reduce、GFS 和 Big Table 等核心技术的学术论文,引发了雅虎、Facebook 等公司开始关注该领域。目前在数据处理和存储方面应用最为广泛的为 ApacheHadoop 系列开源平台,Google 公司针对大数据的需求又提出了 Dremel 和 Pregel 技术^[4],而后 IBM 结合云计算和自主计算提出“认知计算”^[5];在大数据分析方面,除了传统的 BI 技术,人工智能技术领域的统计分析(R HIPE)、机器学习(K 均值聚类算法、Apriori 算法、FP-Growth 算法等^[6])、数据挖掘(IBM SPSS、SGIlineSet、Oracle Darwin,开源的有 Weka 等商业工具)、自然语言处理(OpenNLP、FudanNLP 和哈工大的 LTP^[7])、知识与推理等方法也发挥着巨大的作用;TableauSoftware 公司在数据可视化分析领域走在前列,主要有 Tableau Desktop、Tableau Server、Tableau Reader、Tableau Public 等,其创新应用 VizQL 数据库实现增加了用户的体验^[8];针对 Hadoop 和 MapReduce 计算框架架构下数据加密、访问控制、区别隐私保护和审计^[9]等方面,IBM 已成功实现了同态加密技术,解决了云环境下大数据的加密保护问题^[10]。

海洋信息化作为国家信息化的重要基础^[11]。在大数据研究的热潮下,如何运用大数据思维和技

术,管理海量的海洋大数据,实现海洋大数据信息共享,发掘海洋大数据对国民生产生活的最大价值,获取海洋大数据资产的最大权益,是海洋信息化战略面临的重要问题。本文从数据资产、数据思维和数据可视化 3 个维度加以分析论述。

1 数据资产化

大数据及各类大数据技术的不断发展,大数据被认为具有 4V 特点,即:规模性(volume)、多样性(variety)、实时性(velocity)^[12]和价值性(value)^[13]的特点。大数据的价值性特点在现代社会和经济活动中愈来愈凸显,故大数据被称为新的经济资产类别^[14]。

李志刚认为大数据的价值是指数据正在成为一种新型资产,一种形成竞争力的重要基础^[15]。Mayer-Sch 认为,只有以需求为导向,通过不同数据的重组、挖掘再生的能支持预测的大数据才会变成资产^[16]。海洋数据资产化是指整合海洋数据,针对不同的需求,运用数据重组或者数据挖掘技术,实现海洋数据的决策支持,实现海洋数据的最大价值,逐步建立数据财产权理论,从功能价值转向数据价值,如图 1。借鉴维克托·迈尔的数据、思维和技术的重要性,考虑到我国海洋洋信息化建设中存在的信息共享困难、重复建设、盲目建设问题及现有的技术孤岛、资源孤岛和信息孤岛^[11]等问题,建立以数据资产化管理模式为主导的海洋数据服务需从以下方面入手。



图 1 数据资产化价值

(1)构建海洋数据资源库。海洋数据设计的专题领域较多^[17],目前海洋数据多按获取部门分散存储于不同的系统和数据库,各个系统和数据库的存储格式大不相同,增加了海洋数据的共享难度。海洋数据资产化管理的基础是建立海洋数据资源库,制定海洋大数据信息化的标准和规范,通过数据松散耦合方式,逻辑整合分散存储于各个数据库和系统的海洋数据,以便对数据进行组织和分析,通过

数据资产交换模式奠定海洋大数据资产化基础。

(2)增强海洋大数据技术研究。因海洋数据存在的复杂性、多源化、敏感性、不同比例尺等不同问题,对数据进行资产化分类,总体分为自有产权、他人产权(共享)、公共产权(互联网)三大类,在此基础上,加强海洋数据预处理技术,积极探索海洋数据挖掘方法和分析方法,从海洋大数据中获取更多的信息,提升数据资产价值,实现数据资产价值的二次升值。

(3)增强海洋数据服务功能。海洋数据价值的实质是产品和服务^[18],海洋大数据资产化的关键是可以生产多少有效的信息和服务。海洋信息和服务应从公众需求、海洋管理、海洋研究、海洋环保、海岛保护、海洋科技等方面入手,开发出既可服务于单一部门的基础信息,又可服务于综合部门的分析信息,并且通过整合海洋大数据,分析海洋大数据之间的关联关系,以提供更多专题的海洋信息产品和服务。

2 大数据思维

所谓思维就是人类头脑活动的内在程序,是一种习惯性的思考问题和解决问题的模式,并由此引导自我行为。按照舍恩伯格的说法:“所谓大数据思维,是指一种意识,认为公开的数据一旦处理得当就能为千百万人急需解决的问题提供答案”。大数据思维导致信息化出现了三大改变^[19]:全部样本代替了随机抽样的样本;在大趋势正确的前提下效率较精度重要;不再单纯追求因果相关而追求事物演变的主流趋势。

海洋数据是典型的大数据,涉及海洋水文、生态环境、气象气候、地质地貌、海洋化学、经济、海洋物理化学等方面,其数据类型有矢量、图像、视频、文本等,拥有的资产难以估算,但多源化、复杂化、动态性、敏感性等特点加大了海洋信息化的难度。2015年3月,马云提出人类正由IT(Information Technology)时代进入了DT(Data Technology)时代,他认为二者最大的区别在于:IT时代以“我”为中心,DT时代则以“别人”为中心,让别人更强大,开放和承担更多的责任^[20]。DT更加注重的是用户和客户的需求,从大数据的思维的来看,就是要多

挖掘出数据的价值,满足各类用户的需求。

近年来,我国海洋管理部门积极提倡海洋电子政务工程,研发了海洋环保、海域使用、海岛保护、海洋预报等多个业务系统^[21],如全国海洋生态环境监督管理系统、国家海域动态监视监测管理系统、国家海岛监视监测系统等,不断发展和完善了国家海洋局政府网站、各海洋专题服务网站,海洋信息化发展迅速。从海洋大数据发展思维来看,我国在数据获取、存储管理和分享技术取得了一定的成就,但是随着海洋经济的快速发展,海洋信息需求不断增大,需求产品和服务不断增多,海洋信息的采集、存贮、检索、分析、交换和集成等简单的信息化过程已经不能满足时代的发展要求,而是需要通过海洋信息的整合过程,挖掘更大数据价值,特别是在大数据思维的背景下,需要建立全球视野的大数据平台。

我国海洋信息化需要借鉴Google、百度、腾讯等网络公司运营数据资产的成功经验,坚定数据就是资源这一实质,将体现数据价值凝聚于信息化之中,从用户需求出发,在数据架构、数据获取、数据传输、数据存储、数据管理、数据分析以及数据安全等方面重新构建以数据为核心的信息化体系,挖掘出海洋数据的综合利用价值。当前基于大数据的挖掘技术是研究热点,传统数据分析方法的升级和拓展应用于大数据分析是研究难点。

3 数据可视化

数据可视化主要是借助于图形化手段,清晰有效地传达与沟通信息,数据可视化主要是将复杂的数据借助空间图形,便于人们理解和观察,同时采用不同的空间分析方法还可以发现大数据内部隐含的规律和发展趋势,便于研究者判断和模拟未来的发展趋势。在大数据时代,大数据可视化分析的研究与发展将为科学新发现创造新的手段和条件^[22],数据可视化与可视分析技术水平以及信息可视化表达能力,体现着信息化建设成果的先进性和有效性^[23]。在大数据环境下,利用各种技术分析数据,用形象直观的方式展示结果,这样能够快速发现数据中蕴含的规律特征^[24]。数据信息可视化参考模型见图2^[25]。

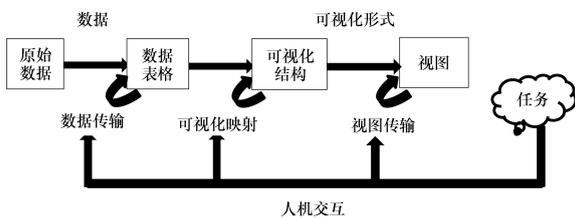


图2 信息可视化参考模型

海洋大数据的可视化,是海洋大数据思维的直观展现,也是进一步提升海洋大数据资产价值的重要手段之一。因海洋数据具有时态性、动态多变性、模糊性等,海洋数据的可视化不但要将数据以图形形式表现出来,实现实时空间查询、选择、分析等功能,而且要借助二维和三维的地理信息软件或可视化数值分析工具,如: ArcGIS、Skyline、Supermap 等,实现海洋大数据多维图表统计功能,同时结合数据挖掘方法实现空间分析,增强多维动态模拟与预测可视化技术的研究,具体可从下面3个方面来研究。

(1)海洋数据实时动态的可视化即数据本身的可视化。采用位置信息、区域统计、多维图表等方式,关注多源、动态、有效等海量数据的可视化,选取合适的符号,实现海洋数据的实时动态美观的可视化。

(2)海洋数据挖掘方法与可视化的结合。结合(非)监督分类、直方图分类、聚类、粗集分类、智能分类等分析算法,空间自相关、叠加、提取、领域分析等空间分析工具,还包括栅格的重分类、栅格表面运算(坡向、坡度、曲率、等值线等)等传统数据分析方法,同时采用选择分类、回归分析、聚类、关联规则、特征、变化和偏差分析、Web 页挖掘等数据挖掘方法挖掘出海洋信息最大的潜在价值。

(3)海洋数据多维动态模拟与预测的可视化。以时相为索引,采用海洋数据动态模拟技术,直观表达基于时间段(点)的连续抽样动态变化,提高认知效果,使得海洋要素的变化规律更易显现。

4 结论

自从美国的“大数据研究与发展计划”提出后,大数据在各国和各个领域受到了极大关注。随着“信息化”上升到国家战略的高度,基于海洋大数据的信息化必将成为国家信息发展的重要基础。海

洋信息化需进入“智慧”阶段,立足本土,走向深蓝,以数据资产化管理模式为主导运用大数据思维建立海洋大数据服务平台,鼓励海洋行业内外建立本机构、本单位的数据资产化管理模式,在此基础上以松耦合模式迅速扩充海洋基础信息库,大力发展海洋数据挖掘和可视化技术,创新开发数据产品和服务,逐步利用行业外、社会公众力量,形成“百花争鸣,百花齐放”态势,挖掘出海洋数据的最大价值,提升海洋信息化整体能力水平和实力,进而全面提升海洋综合管控能力。

参考文献

- [1] SONSINSKY B. Cloud computing bible [M]. Indianapolis: John Wiley & Sons Inc, 2011.
- [2] HAO Yulu, SONG Meina, HAN Jing, et al. A Cloud computing model based on hadoop with an optimization of its task scheduling algorithms[C]//The 13th International Conference on Enterprise Information Systems, 2011: 524-528.
- [3] Office of Science and Technology Policy, Executive Office of the president. Obama administration unveils “big data” initiative: announces MYM200 million in New R & D investments [EB/OL]. (2012-05-29) [2017-01-06]. <http://www.whitehouse.gov>.
- [4] 刘军. Hadoop 大数据处理[M]. 北京: 人民邮电出版社, 2013: 45-60.
- [5] 王博. IBM 亮剑新互联时代[EB/OL]. (2014-04-14) [2017-01-06]. <http://www.dooland.com/magazine/online.php?pid=MTAzMDEx>.
- [6] HARRINGTON P. 机器学习实战[M]. 李锐, 李鹏, 曲亚东, 译. 北京: 人民邮电出版社, 2013: 184-239.
- [7] 百度. 自然语言处理[EB/OL]. (2014-05-23) [2017-01-06]. <http://baike.baidu.com/view/18784.htm?fr=aladdin-BaiDu>.
- [8] WAN R. Data Visualization star: tableau software[EB/OL]. (2013-07-01) [2017-01-06]. <http://www.kuailiyu.com/article/3901.html>.
- [9] ZHANG Xuyun, LIU Chang, SURYA N, et al. Privacy reservation over big data in cloud systems[J]. Security, Privacy and Trust in Cloud Systems, 2014(3): 239-257.
- [10] GENTRY C. Fully homomorphic encryption using ideal Lattices [C]//Proceedings of the 41st annual ACM symposium on theory of computing (STOC'09). USA: Stanford University and IBM Watson, 2009: 169-178.

(以下内容转至第 77 页)

据,及时发现国际航运市场变动,并预测可能的趋势,做好风险防范工作,并为有关企业提供针对性的资金支持,做到锦上添花,雪中送炭。

4 结语

中国的商业银行发展航运金融,有着自己独特的优势,但也存在着明显的缺陷和不足。航运金融业务的发展,需要航运业与金融业之间相辅相成,互相协调。因此,在发展航运金融业务的过程中,商业银行要结合现实情况,抓住发展机遇,发挥自身优势,弥补自身不足,积极转型,从而更好地为我国发展海洋经济的战略服务。

参考文献

- [1] 甘爱平,曲林迟.航运金融学[M].上海:格致出版社,2010:1-474.
- [2] 王文佳,钟芳燕,戴丽琼,等.舟山群岛新区航运金融服务业务发展思路及对策研究[J].海洋开发与管理,2013,30(11):

107-110.

- [3] 王泽宇,韩增林,李洋.现代物流业对大连建设东北亚航运中心的拉动效应分析[J].海洋开发与管理,2005,22(2):31-35.
- [4] 陈继红,韩玲冰,张明香.上海航运金融服务发展现状与对策[J].水运管理,2012,34(9):16-20.
- [5] 张海波,刘霖涌.我国商业银行发展海洋金融业务的对策研究:以中国建设银行宁波分行为例[J].特区经济,2014(6):100-102.
- [6] 王磊.上海航运、金融发展与经济增长关系的实证研究[J].武汉交通职业学院学报,2010,12(2):8-12.
- [7] 王仁祥,邓平.创新型城市建设中金融支持度的评价与指标体系设计[J].海南金融,2008(2):21-24.
- [8] 上海海洋大学航运金融课题组.航运金融[M].北京:中国金融出版社,2012:1-232.
- [9] 殷明,郑士源,章强.航运金融市场价格传导及买卖价差的形成机制:基于干散货运输现货与衍生市场的实证分析[J].上海金融,2015(2):77-81.
- [10] 陈扬,王学锋,郑士源.“航运金融”的研究边界及其重点命题的探讨[J].上海金融,2012(8):107-109.

(上接第53页内容)

- [11] 许莉莉,汤海荣,张燕歌.海洋信息化标准体系研究[J].中国标准导报,2015(1):49-51.
- [12] DUMBILL E. What Is Big Data [EB/OL]. (2013-01-24) [2017-01-06]. <http://strata.oreilly.com/2012/01/what-is-big-data.html>.
- [13] 俞立平.大数据与大数据经济学[J].中国软科学,2013(7):177-183.
- [14] DICKERSON R, LU J, LU J, et al. Stream feeds-an abstraction for the worldwide sensor web[C]//The Internet of Things. Springer Berlin Heidelberg, 2008:360-375.
- [15] 李志刚.大数据:大价值、大机遇、大变革[M].北京:电子工业出版社,2012.
- [16] MAYER-SCHNBERGER V, CUKIER K. Big Data: A Revolution That Will Transform How We Live, Work and Think [M]. Eamon Dolan/Houghton Mifflin Harcourt. 中文版《大数据时代》.盛扬燕,周涛,译.杭州:浙江人民出版社,2013.
- [17] 刘高岭.海洋信息元数据目录服务系统的设计与实现[D].青岛:中国海洋大学,2008.
- [18] 祁冬梅,于婷,邓增安.IODE海洋数据共享平台建设及对我

国海洋信息化进程的启示[J].海洋开发与管理,2014,31(3):57-61.

- [19] 维克托·迈尔·舍恩伯格,肯尼思·库克耶.大数据时代:生活、工作与思维的大变革[M].盛扬燕,周涛,译.杭州:浙江人民出版社,2012.
- [20] 马云.IT时代和DT时代的区别是什么? [EB/OL]. (2015-03-22). [2017-01-06] <http://tech.qq.com/a/20150322/019206.htm>
- [21] 国家海洋局规划司.海洋信息工程总结[EB/OL]. [2017-01-06] http://www.cgs.gov.cn/dzss/z_t_more/xxh/news/036.htm.
- [22] 李国杰.大数据研究的科学价值[J].中国计算机学会通讯,2012,8(9):8-15.
- [23] 戴勤奋,魏合龙,王圣洁,等.大数据时代的海洋地质信息化建设[J].海洋地质前沿,2014,30(1):67-70.
- [24] 陈明.大数据可视化分析[J].计算机教育,2015(5):028.
- [25] CARD SK, MACKINLAY JD, SHNEIDERMAN B. Readings in Information Visualization: Using Vision To Think [M]. San Francisco: Morgan-Kaufmann Publishers, 1999:1-712.