

“十二五”国家重点图书出版规划项目

大数据技术与应用

丛书主编

朱扬勇 吴俊伟

Big Data

Technology and Application Series

黄冬梅 邹国良 等
编著

海洋 大数据



上海科学技术出版社



大数据技术与应用

海洋大数据

黄冬梅 邹国良 等
编著

上海科学技术出版社

图书在版编目(CIP)数据

海洋大数据 / 黄冬梅, 邹国良等编著. —上海: 上海科学技术出版社, 2016. 1

(大数据技术与应用)

ISBN 978 - 7 - 5478 - 2783 - 3

I. ①海… II. ①黄… ②邹… III. ①海洋—数据—研究 IV. ①P7

中国版本图书馆 CIP 数据核字(2015)第 196417 号

海洋大数据

黄冬梅 邹国良 等编著

上海世纪出版股份有限公司 出版
上海科学技术出版社

(上海钦州南路 71 号 邮政编码 200235)

上海世纪出版股份有限公司发行中心发行
200001 上海福建中路 193 号 www.ewen.co

苏州望电印刷有限公司印刷

开本 787×1092 1/16 印张 13.5

字数 300 千字

2016 年 1 月第 1 版 2016 年 1 月第 1 次印刷

ISBN 978 - 7 - 5478 - 2783 - 3/TP · 36

定价: 80.00 元

内容提要

本书从海洋大数据的历史及国内外研究与应用现状出发,分析了海洋大数据的获取与特征分类、海洋大数据的处理及应用的关键技术,展示了海洋大数据在风暴潮灾害评价与决策以及海洋溢油灾害的检测与防治两方面的典型应用,最后就关键技术与实际应用两个方面进行了发展趋势的展望。本书介绍了基于混合云的海洋大数据存储技术、海洋大数据的分析挖掘技术、海洋大数据的质量控制技术、海洋大数据的信息安全技术以及基于 Spark 云平台的海洋大数据动态分析与展示等技术。

本书立足于海洋信息化技术的基础研究、应用开发和海洋信息服务的实际项目与成果,实现理论与实践紧密结合,内容丰富,条理清晰,关键技术与典型应用相结合,具有一定的前瞻性。

本书适合于海洋信息化领域的相关技术人员阅读,可作为海洋信息化管理、决策和研究人员的参考书,也可作为相关专业研究生和高年级本科生的教材。

本书涉及内容成果获得如下项目的资助

1. 国家自然科学基金,“数字海洋”中海量复杂类型数据的质量检验及存储问题研究
2. 国家海域综合调查与评价重大专项,“数字海洋”上海示范区建设
3. 国家减灾防灾能力建设重大专项,城市风暴潮灾害辅助决策支持系统
4. 上海市科委项目,上海南汇城市风暴潮灾害辅助决策系统技术研究
5. 国家科技部海洋公益性行业专项,临港新城风暴潮灾害评估与对策辅助决策系统研究
6. 国家科技部海洋公益性行业专项,渤海海洋环境信息集成及动态管理技术示范应用
7. 国家科技部海洋公益性行业专项,苏北浅滩“怪潮”灾害监测预警关键技术研究及示范应用
8. 国家科技部海洋公益性行业专项,极地海洋环境监测网系统研发及示范
9. 国家海洋局南北极环境综合考察与评估专项,极地环境与资源信息集成与共享服务系统

大数据技术与应用

学术顾问

中国工程院院士 邬江兴

中国科学院院士 梅 宏

中国科学院院士 金 力

教授,博士生导师 温孚江

教授,博士生导师 王晓阳

教授,博士生导师 管海兵

教授,博士生导师 顾君忠

教授,博士生导师 乐嘉锦

研究员 史一兵

大数据技术与应用
编撰委员会

丛书指导

干 频 石 谦 肖 菁

主 任

朱扬勇 吴俊伟

委 员

(以姓氏笔画为序)

于广军	朱扬勇	刘振宇	孙景乐	杨 丽	杨佳泓	李光亚
李光耀	吴俊伟	何 承	邹国良	宋俊典	张 云	张 洁
张绍华	张鹏翥	陈 云	武 星	宗宇伟	赵国栋	黄冬梅
黄林鹏	韩彦岭	童维勤	楼振飞	蔡立志	熊 贇	糜万军

本书编委会

主 编

上海海洋大学 黄冬梅

上海海洋大学 邹国良

编 委

国家海洋局信息中心 石绥祥

国家海洋局东海分局 苏 诚

国家海洋局东海分局 龚茂珣

上海东海海洋工程勘察设计研究院 郭伟其

上海东海海洋工程勘察设计研究院 谢文辉

上海海洋大学 何世钧

上海海洋大学 王振华

上海海洋大学 郑宗生

上海海洋大学 贺 琪

上海海洋大学 梅海彬

上海海洋大学 袁小华

上海海洋大学 何盛琪


上海海洋大学 郑小罗

上海海洋大学 魏立斐



张明华 上海海洋大学
王 建 上海海洋大学
赵丹枫 上海海洋大学
张律文 上海海洋大学
杨蒙召 上海海洋大学
肖启华 上海海洋大学
张书台 上海海洋大学
包晓光 上海海洋大学
刘 爽 上海海洋大学
熊中敏 上海海洋大学
杜艳玲 上海海洋大学

丛书序



我国各级政府非常重视大数据的科研和产业发展,2014年国务院政府工作报告中明确指出要“以创新支撑和引领经济结构优化升级”,并提出“设立新兴产业创业创新平台,在新一代移动通信、集成电路、大数据、先进制造、新能源、新材料等方面赶超先进,引领未来产业发展”。2015年8月31日,国务院印发了《促进大数据发展行动纲要》,明确提出将全面推进我国大数据发展和应用,加快建设数据强国。前不久,党的十八届五中全会公报提出要实施“国家大数据战略”,这是大数据第一次写入党的全会决议,标志着大数据战略正式上升为国家战略。

上海的大数据研究与发展在国内起步较早。上海市科学技术委员会于2012年开始布局,并组织力量开展大数据三年行动计划的调研和编制工作,于2013年7月12日率先发布了《上海推进大数据研究与发展三年行动计划(2013—2015年)》,又称“汇计划”,寓意“汇数据、汇技术、汇人才”和“数据‘汇’聚、百川入‘海’”的文化内涵。

“汇计划”围绕“发展数据产业,服务智慧城市”的指导思想,对上海大数据研究与发展做了顶层设计,包括大数据理论研究、关键技术突破、重要产品开发、公共服务平台建设、行业应用、产业模式和模式创新等大数据研究发展的各个方面。近两年来,“汇计划”针对城市交通、医疗健康、食品安全、公共安全等大城市中的重大民生问题,逐步建立了大数据公共服务平台,惠及民生。一批新型大数据算法,特别是实时数据库、内存计算平台在国内独树一帜,有企业因此获得了数百万美元的投资。

为确保行动计划的实施,着力营造大数据创新生态,“上海大数据产业技术创新战略联盟”(以下简称“联盟”)于2013年7月成立。截至2015年8月底,联盟共有108家成员单位,既有从事各类数据应用与服务的企业,也有行业协会和专业学会、高校和科研院所、大数据技术和产品装备研发企业,更有大数据领域投资机构、产业园区、非IT

领域的的数据资源拥有单位,显现出强大的吸引力,勾勒出上海数据产业的良好生态。同时,依托复旦大学筹建成立了“上海市数据科学重点实验室”,开展数据科学和大数据理论基础研究、建设数据科学学科和开展人才培养、解决大数据发展中的基础科学问题和技术问题、开展大数据发展战略咨询等工作。

在“汇计划”引领下,由联盟、上海市数据科学重点实验室、上海产业技术研究院和上海科学技术出版社于2014年初共同策划了《大数据技术与应用》丛书。本丛书第一批已于2015年初上市,包括了《汇计划在行动》《大数据评测》《数据密集型计算和模型》《城市发展的数据逻辑》《智慧城市大数据》《金融大数据》《城市交通大数据》《医疗大数据》共八册,在业界取得了广泛的好评。今年进一步联合北京中关村大数据产业联盟共同策划本丛书第二批,包括《大数据挖掘》《制造业大数据》《航运大数据》《海洋大数据》《能源大数据》《大数据治理与服务》等。从大数据的共性技术概念、主要前沿技术研究和当前的成功应用领域等方面向读者做了阐述,作者希望把上海在大数据领域技术研究的成果和应用成功案例分享给大家,希望读者能从中获得有益启示并共同探讨。第三批的书目也已在策划、编写中,作者将与大家分享更多的技术与应用。


大数据对科学研究、经济建设、社会发展和文化生活等各个领域正在产生革命性的影响。上海希望通过“汇计划”的实施,同时也是本丛书希望带给大家一个理念:大数据所带来的变革,让公众能享受到更个性化的医疗服务、更便利的出行、更放心的食品,以及在互联网、金融等领域创造新型商业模式,让老百姓享受到科技带来的美好生活,促进经济结构调整和产业转型。



上海市科学技术委员会副主任

2015年11月

序



在生物学家眼中,海洋是生命的摇篮,生物多样性的展览厅;在地质学家心里,海洋是资源宝库,蕴藏着地球村人类持续生存的希望;在气象学家看来,海洋是风雨调节器,春、夏、秋、冬年复一年;在物理学家脑中,海洋是运动载体,风、浪、流汹涌澎湃;在旅行者看来,海洋是风景优美无边的旅游胜地;然而在信息学家脑海中,海洋是五花八门、瞬息万变、铺天盖地的大数据源。有人分析世界上现有的大数据中,环境监测数据占70%,而海洋环境监测数据量占到了其中的70%以上,似乎与海洋占地球的70%面积相吻合,其实随着卫星传感和网络等高新技术日益发展,天-空-海和海面-水中-海底立体观测所获取的数据将逐年翻倍增长,看来海洋大数据在21世纪将掀起一场惊涛骇浪的海洋信息革命。在这一场信息革命袭来之际,上海海洋大学数字海洋研究所/上海海洋灾害与安全环境数据工程研究中心的黄冬梅和邹国良等教授率领团队和国家海洋局东海分局合作者担当了上海大数据产业技术创新战略发展联盟牵头的《大数据技术与应用》丛书的《海洋大数据》一书编写,现已完稿,正待出版,如一朵海洋信息科学奇葩,甚是可喜可贺。

来自卫星、载人飞船、空间站、气艇、无人机、岸基雷达和观测站、船载探测平台、浮标、水下滑翔机、水下潜器和海底观测网等的资料时空无缝地丰富着海洋大数据。本书将告诉读者什么是海洋大数据,大数据不仅仅在于数据量大,更在于海量数据在网络和云计算技术支持下的快速处理、智能处理和智慧应用,把海洋大数据定义为以大数据驱动力智能的新兴海洋信息科学工程。来自海洋的自然属性的大数据必须交融来自人类海洋活动的社会、经济、历史、文化、法律数据,把看似有些不关联的数据,通过深层挖掘找出掩藏在人们想象背后的新关联现象,发现新规律。从哲学上来定义,海洋大数据是人类构建“海洋神经系统”的过程,而在这个系统中人类只是其中一个感测器而已,预示

着大数据的智能将超越人们的知识。

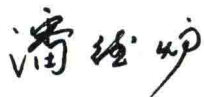
该书作者紧密围绕国家海洋发展战略,在海洋灾害防治、数字海洋建设、海洋信息安全建设等方面的关键技术研究与应用开发工作基础上,以海洋大数据的新理念加以提炼和分析,用他们的体会告诉读者海洋大数据区别于传统海量海洋数据的特征和分类;以数据驱动智慧的处理方法,挖掘的规则和模型,可为读者理解海洋大数据的内涵抛砖引玉。

更值得一提的是,本书还介绍了海洋大数据在灾害辅助决策系统和海洋溢油监测中的应用实践,让读者看到海洋大数据的潜在作用,为读者提供了海洋大数据应用的思路和榜样。

海洋是当今国际上政治、经济、外交和军事博弈的重要舞台,博弈中无非是对海洋环境认知能力的竞争、海洋资源开发能力的竞争和海洋权益维护能力的竞争。在这场错综复杂的三大能力的竞争中,哪个国家掌握了高科技制高点,哪个国家就掌握了主动权。如果把海洋信息比喻为登上这个制高点的翅膀,那么海洋大数据将是我国登上这个制高点的基底,这本书可谓海洋大数据惊涛骇浪下的一滴闪闪发亮的水滴,让读者可以从中受益。

著者集众贤之能,承实践之上,总结经验,理出体会,挥笔习书,言海洋大数据之理论,摆实践之范例,是一本值得一读的佳作。更欣慰的是本书的出版也看到了年轻的海洋信息学家的崛起和成长。

“百尺竿头,更进一步”,殷切期盼上海海洋大学海洋信息团队,在黄冬梅教授带领下,再接再厉,推海洋大数据之浪,为“透明海洋和智慧海洋”做出更大贡献。



2015年5月于杭州

前 言

海洋大数据作为全球大数据的重要组成部分,是实现海洋信息行业智能化管理和“互联网+”的基础与前提,也是实现我国“海洋强国”战略的支撑与保障。随着我国“空、天、地、底”海洋立体监测技术的发展和“数字海洋”建设的全面深入,海洋信息化已经逐步从“数字海洋”向“智慧海洋”发展,海洋数据从数量、增长速度、种类扩展三方面发生了飞跃式发展;同时海洋数据蕴含的价值也越来越高。如何充分挖掘海洋数据价值,为我国海洋经济与海洋科技发展提供更多的支撑与动力,利用大数据处理技术与方法来研究、开发和应用海洋大数据成为当务之急。

上海作为一个沿海国际大都市,海洋经济所占 GDP 的份额位居世界前列。通过海洋大数据的研究、开发和应用,合理利用和引导海洋地形地貌变化,有利于港口与航道发展,通过海洋经济数据的深度挖掘与加工,进一步促进上海国际航运中心的发展,从而推动海洋经济和海洋科技发展,增加海洋产业的贡献率。同时,上海地处东南季风的主风口,每年风暴潮灾害对上海的经济与社会发展产生不小的影响,海洋环境污染导致了东海海域赤潮、盐水入侵、海岸带侵蚀等海洋环境灾害的频发,严重影响上海的经济与社会环境的稳定。如何运用好海洋大数据构建和谐的海洋生态环境,将成为未来上海经济与社会可持续发展长远而持久的保证。

在这种形势下,响应上海大数据产业技术创新战略发展联盟牵头的《大数据技术与应用》丛书的编写要求,本书编写组在黄冬梅、邹国良组织下,策划编写了《海洋大数据》一书,由黄冬梅、邹国良等确定全书的章节目录结构和主要内容思路。具体分工为:第 1 章由黄冬梅、邹国良编著;第 2 章由王振华、苏诚、梅海彬、刘爽、张书台编著;第 3 章由苏诚、龚茂珣、王振华、郑宗生、刘爽编著;第 4 章由黄冬梅、魏立斐、王振华、何盛琪、贺琪、张明华、杜艳玲编著;第 5 章由赵丹枫、郭伟其、郑小罗、张律文、王建编著;第 6 章由

杨蒙召、何世钧、袁小华、谢文辉、邹国良编著；第7章由贺琪、肖启华、包晓光、熊中敏编著。最终全书由黄冬梅、邹国良负责统稿。

本书可作为相关专业研究生和高年级本科生的课程教材，对海洋行业管理部门的管理者和决策者亦有一定的参考价值，也可以作为海洋从业人员的参考丛书，还可以作为其他行业领域科技工作者的参考书，对普通大众了解海洋大数据有一定的帮助。海洋大数据涉及的数据种类和范围巨大，而且随着新技术的应用，其变化也超出了大多数人的预料，包括本书的作者。受编著者的能力和眼界的局限，书中难免有以点概面、挂一漏万的不足，欢迎广大读者批评指正，我们将一如既往地不断改进，今后写出更好的著作。

此外，本书的编写得到了潘德炉院士（中国工程院院士、国家海洋局第二海洋研究所研究员）的大力支持，为本书章节布局及写作思路提出了宝贵意见，并亲自为本书作序，在此表示衷心感谢；上海市数据科学重点实验室的朱扬勇和上海大数据产业技术创新战略发展联盟的吴俊伟、毛火华等专家对本书目录提纲的修改提出了许多宝贵意见，在此表示衷心感谢；编写组全体同仁为了完成本书的编写工作，付出了辛勤和努力，上海海洋大学数字海洋研究所的相关研究生也为本书查找了很多参考资料及对本书的格式等进行了修改，在此表示感谢。

本书作者

目录

第1章 海洋大数据的历史沿革	1
• 1.1 传统海洋数据	2
1.1.1 地理大发现	2
1.1.2 航海的需要	7
• 1.2 从传统海洋数据到海洋大数据	9
1.2.1 海洋经济的发展促进海洋大数据的产生	9
1.2.2 海洋灾害的防治促进海洋大数据的产生	17
• 1.3 海洋大数据发展现状	24
1.3.1 海洋大数据萌芽	24
1.3.2 海洋大数据国际现状	28
1.3.3 海洋大数据国内现状	32
参考文献	34
第2章 海洋大数据获取与特征	37
• 2.1 海洋大数据的获取	38
2.1.1 空基监测平台海洋数据的获取	38
2.1.2 陆基监测平台海洋数据的获取	42
2.1.3 海底监测平台海洋数据的获取	53

2.1.4	历史海洋数据	56
2.1.5	社会经济数据的收集与归档	57
• 2.2	海洋数据的特征	57
2.2.1	海洋数据的海量性	57
2.2.2	海洋数据的多类性	58
	参考文献	62
第3章	海洋大数据分类	65
• 3.1	被动产生的海洋大数据	66
3.1.1	海域使用管理数据	66
3.1.2	海洋环境管理数据	70
3.1.3	海岸带综合管理数据	71
• 3.2	主动产生的海洋大数据	72
	参考文献	76
第4章	面向海洋大数据应用的关键技术研究	79
• 4.1	海洋大数据的存储	80
4.1.1	云计算技术	81
4.1.2	海洋大数据专有云平台	82
4.1.3	海洋大数据存储模型	83
• 4.2	海洋大数据分析挖掘	85
4.2.1	时间序列相似性分析	85
4.2.2	时间序列预测	89
4.2.3	时空聚类	93
4.2.4	时空异常检测	94
4.2.5	监督分类与非监督分类分析	96
• 4.3	海洋大数据的质量控制	96
4.3.1	海洋大数据的生命周期	96
4.3.2	海洋大数据的质量要素	97
4.3.3	面向多模态海洋大数据的模糊质量评估模型研究	98