

引文格式:刘纪平,张福浩,徐胜华.政务地理空间大数据研究进展综述[J].测绘学报,2017,46(10):1678-1687. DOI:10.11947/j.AGCS.2017.20170320.
LIU Jiping, ZHANG Fuhao, XU Shenghua. Progresses and Prospects in Geospatial Big Data for E-government [J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1678-1687. DOI: 10.11947/j.AGCS.2017.20170320.

政务地理空间大数据研究进展综述

刘纪平,张福浩,徐胜华
中国测绘科学研究院,北京 100830

Progresses and Prospects in Geospatial Big Data for E-government

LIU Jiping, ZHANG Fuhao, XU Shenghua
Chinese Academy of Surveying and Mapping, Beijing 100830, China

Abstract: In recent years, geospatial big data have attracted great attention from industry, academia, research and government sectors, and even have triggered a lot of industry changes. Geospatial big data for E-government provide new means for government information management and decision making. This paper analyzes the concepts and characteristics of geospatial big data for E-government, mainly reviews the key technologies in geospatial big data for E-government, including data integration, storage and management, data mining and analysis, visualization, and decision support service technologies. Finally, the bottleneck problems and technical challenges of geospatial big data for E-government are discussed.

Key words: geospatial big data for E-government; data integration; storage and management; data analysis and data mining; visualization; decision support service technologies

Foundation support: The National Key Research and Development Program of China(No. 2016YFC0803101); The Special Scientific Research Fund of Public Welfare Profession of China (No. 2015123032); Basic Research Fund of CASM (No. 7771701)

摘 要:近年来,地理空间大数据引起了产学研及政府部门的高度关注,甚至引发了许多变革。面向电子政务的地理空间大数据技术为政府信息管理和决策提供了新的手段。本文在阐述政务地理空间大数据的内涵及特点的基础上,重点梳理了政务地理空间大数据处理的关键技术,包括政务地理空间大数据整合、存储与管理、挖掘与分析、可视化以及决策支持服务技术,最后分析了当前和未来政务地理空间大数据所面临的机遇和挑战。

关键词:政务地理空间大数据;数据整合;存储与管理;分析与挖掘;可视化;决策支持服务

中图分类号:P208 **文献标识码:**A **文章编号:**1001-1595(2017)10-1678-10

基金项目:国家重点研发计划(2016YFC0803101);测绘地理信息公益行业专项(201512032);中国测绘科学研究院基本科研业务费(7771701)

随着物联网、移动互联网、云计算的蓬勃发展,信息技术与人类社会活动的方方面面不断交叉融合,遍布各地的智能传感设备、移动设备、电子商务网站、社交网络与社会媒体每时每刻都在产生价值丰富、类型各异的数据,大数据时代已然来临,并已经迅速发展成为科技界、工业界、企业界甚至政府部门关注的热点,日益对全球生产、流通、分配、消费活动以及经济运行机制、社会生活方式和国家治理能力产生重要影响。Nature 和

Science 等国际顶级学术刊物也相继出版专刊探讨大数据带来的机遇和挑战。当前,许多国家的政府和国际组织都认识到了大数据的重要作用,纷纷将开发利用大数据作为夺取新一轮竞争制高点的重要抓手,实施大数据战略。2012 年 3 月,美国政府发布了《大数据研究与发展倡议》,来推进从大量的、复杂的数据集合中获取知识和洞见的能力;2013 年 6 月,八国峰会签署了《八国集团开放数据宪章》,标志着开放政府数据成为全球趋

势;2015年9月,国务院发布了《关于促进大数据发展的行动纲要》,标志着大数据在我国上升到国家战略层面^[1-2]。在电子政务由电子政府向智慧政府转变的背景下,政府数据的共享融合是电子政务大数据应用的前提,也是打造智慧政府生态圈的基础。

“大数据”概念中所蕴含的“基于全面数据进行决策”“从全覆盖的实时数据流中发现规律、预测趋势”等观念、方法与技术对政府管理精细化与综合决策构成了深远的影响。地理空间大数据不仅是电子政务不可或缺的重要组成部分,同时也是政务信息的重要载体,是电子政务资源整合的基础框架,政务地理空间大数据已成为各级政府部门进行业务管理和宏观分析决策的有力工具。政府部门的交通数据、气象数据、社会经济统计数据等各类地理空间大数据已被广泛研究,大量研究成果表明利用这些大数据不仅可以反映国家、区域、城市中存在的问题,还可以用来解决空气污染、交通拥堵、能耗增加和规划落后等带来的挑战^[3-5]。对于海量且瞬息万变的政务地理空间大数据来说,其蕴含的价值已然超乎人类的想象。本文将在阐述政务地理空间大数据的内涵及特点的基础上,从数据整合、存储与管理、挖掘与分析、可视化、决策支持服务等方面分析政务地理空间大数据的理论与技术发展现状,并结合其面临的机遇和挑战,对政务地理空间大数据做出展望。

1 政务地理空间大数据的内涵及特点

政务地理空间大数据是面向政务信息管理和政府综合决策的地理空间大数据。利用多源空间数据汇集整合、分布式空间数据混合存储管理等技术,以地理空间数据为基础,整合政府机关掌握的与经济、社会管理和公共服务相关的信息,将孤立、无序、分散的数据资源转化为逻辑一致、有机关联的政务信息资源,最后通过深度融合、挖掘与分析,为政府部门提供智能化决策支持服务^[6-9]。

政务地理空间大数据除了具有一般大数据体量大(volume)、类型多样(variety)、处理速度快(velocity)、价值密度低(value)、真实性高(veracity)等特征外,还具有其自身的特点,主要体现在:①全样本持续更新,传统的政务信息资源主要通过调查手段获取,表现为抽样的、样本反馈的、主观的、结果的、结构化的以及断点的数据类型;而当前的政务地理空间大数据是以监测数据

为主,数据呈现出全样本的、监测记录的、客观的、过程的、非结构化的以及连续的数据类型特点,数据资源更加全面、丰富,时效性更强。②多源异构,政务地理空间大数据来源多样,既包含办公文档、文本、报表、图像、音视频等非结构化数据或半结构化数据,也包含专业部门的遥感影像、导航定位、地图等结构化地理空间数据。伴随移动互联、感知技术的成熟,产生了大量与位置相关的时序数据、轨迹数据、社交网络中的图片、语音、视频等数据,这些多源异构的数据在政府部门决策中也发挥了重要作用。这些通过不同渠道获取的数据在精度、尺度、数据存储、空间关系、语义表达等方面均存在较大差别。③多粒度多层次,政务地理空间大数据来源于国家级、省级和市县等不同层次的政府机构和部门,涵盖了宏观、微观等不同粒度的数据,且不同部门间碎片化、零散化、低效率的数据交互和分析普遍存在。④权威可靠,政务地理空间大数据涉及工商、税务、交通、医疗、教育、水利、农业、地震、环境保护等领域,数据来源于政府各个权威部门,事关百姓生活的方方面面,数据质量较高同时数据的潜在价值十分可观。⑤安全性高,政务地理空间大数据具有较高的安全性,与普通大数据不同,政务地理空间大数据往往与国家安全、国家利益密切相关。

2 政务地理空间大数据关键技术

政务地理空间大数据的产生方式不断扩展,数据量呈现出爆炸式增长,由于数据具有多源异构、跨领域多部门、服务方式多样性等特征,导致数据之间的关系变得错综复杂且存在关联、交叉和融合,严重影响了政务信息管理和政府空间决策。因此,如何最大程度挖掘政务地理空间大数据的大价值,其关键在于对数据进行有效的整合、存储、挖掘与分析以及可视化分析,并提供高效的决策支持服务。

2.1 政务地理空间大数据整合技术

政务地理空间大数据按照表现特征可以分为图形、图像、数值、文字、音频视频等数据类型,按照数据结构可以分为矢量、栅格、文本、多媒体等数据类型,按照信息应用可分为基础地理空间数据、专题空间数据、政务专题数据等。这些数据在时间和空间上具有不同分辨率,关系极为复杂,来源于不同部门及层次的政务地理空间大数据不仅在数据格式、分类系统、空间尺度、数据模型、数据

表达、时空参考、编码规范等方面存在不一致问题^[10-11],还存在数据不连续、缺乏关联等不足。政务地理空间大数据整合是通过多源、多尺度、多种类、多分辨率、多时相、动态化的地理空间数据

的整合,实现对政务文本、表格、视频、空间信息等多类型数据的分类、提取、关联和汇聚,最终为政务信息化建设和电子政务提供一致的、连续的、相关联的地理信息支持。

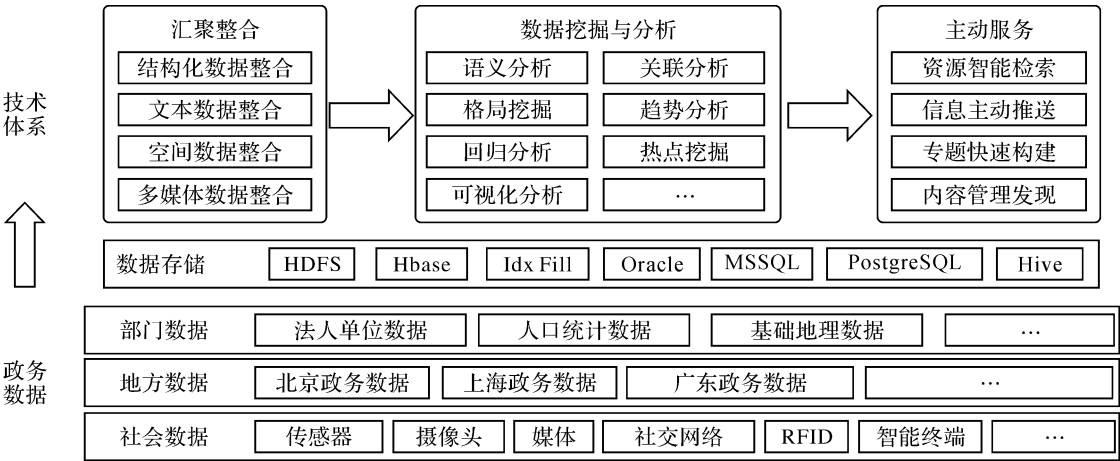


图 1 政务地理空间大数据的关键技术

Fig.1 Key technologies of geospatial big data for E-government

空间数据整合一词“Conflation”最早由Saalfeld提出,用于改善美国人口调查局的人口普查地图数据库的几何精度^[12]。随后,文献[13—14]提出了空间数据融合是将具有重叠区域的不同来源的地理信息合并的过程,目标是处理不同空间数据间的位置、语义、拓扑冲突,使得数据更加精准、冗余最少。文献[15]提出空间数据整合是将不同数据间的同名对象进行识别,再进行数据的几何、属性变换。而政务地理空间大数据整合结合其自身的特点,需要重点解决矢量栅格数据一体化整合、多尺度地理空间数据整合、统计数据和空间数据整合、文本信息和空间信息整合等问题^[9]。相关学者提出了基于语义层面的数据整合、基于本体的信息资源整合、多源异构和多尺度数据的几何纠正与语义转换、分类数据整合、碎片化数据整合等^[16-23]方法。当前,随着基于位置的社交网络的发展,带来了丰富的众源网络地理空间数据^[24-25],不仅丰富了移动位置服务、智能导航、应急救援响应等领域的地理空间数据资源,还可以辅助政府部门分析城市规划发展、社会生活所面临的问题和挑战,但同时也面临着大量的数据整合问题需要解决,例如,文献[26]提出基于本体的方法融合道路网数据和专业地理数据,实现多源道路网数据和专业地理数据的匹配整合;文献[27]利用自然语言处理工具和数据挖掘

技术实现交通轨迹数据与位置签到数据的融合。

2.2 政务地理空间大数据存储与管理技术

政务地理空间大数据类型复杂,传统的存储和管理模式已经无法满足复杂数据的处理需求。从最简单的数据存储到分布式数据库存储慢慢发展起来,分布式和并行计算技术可以为地理空间大数据处理提供一些较好的解决方案。当前,国内外的研究和主要着眼于海量地理空间数据的快速存储、查询,并取得了较大进展^[28],基于Hadoop的地理空间矢量数据的分布式存储和查询,实现了大数据的实时访问、存储和查询^[29-31],对于政务地理空间大数据的存储和查询拥有较大优势。

大多数政务地理空间大数据是非结构化和半结构化,数据管理要求已经超越传统的关系型数据库的有效范围,需要能够提供可扩展的灵活的数据库技术来管理和访问数据。NoSQL用于超大规模政务地理空间大数据的存储,可以很容易适应新的数据类型,这些类型的数据存储不需要固定的模式,无需多余操作就可以横向扩展,并且不会受第三方数据供应商的内容结构变化的影响。HBase是三大NoSQL数据库之一,HBase是一个既支持结构化也支持非结构化数据存储的分布式数据库,分布式计算操作利用了MapReduce计算框架,因此拥有实时读写数据、

高效准确存储和管理数据的优势,且 HBase 采用列稀疏存储的列/行型矩阵,列值中不会存储空 NULL 的元素,可以大大节省存储空间^[32]。除此以外,政务地理空间大数据的存储一般还结合分布式数据仓库(Hive)、分布式关系型数据库(PostgreSQL)、全文搜索引擎以及分布式文件系统(HDFS)等,实现政务结构化数据和非结构化数据的混合式存储^[33-35]。针对政务地理空间大数据的查询问题,传统的并行计算和空间数据库技术难以应对政务数据量大、类型多样的空间查询处理和分析,且可扩展性差、支持类型单一。当前,基于 MapReduce 框架设计并行化、高效的空间查询算法已经成为国内外学者关注的热点^[36],并为政务地理空间大数据的查询分析提供了思路。MapReduce 及其开源系统 Hadoop 目前已经广泛应用于地理空间科学领域,大量的空间查询和分析算法移植到 Hadoop 平台,并产生了一系列成果^[37-39]。此外,运用近似计算的思想针对网络相关的查询进行近似处理也出现了很多成果^[40-41],相关研究提出了基于特定损失函数对道路网络数据进行化简并以简化的网络为基础进行相关查询的近似处理方法,以提高查询处理的实时性,同时有效约束处理过程中产生的误差。

2.3 政务地理空间大数据挖掘与分析技术

政务地理空间数据挖掘与分析是指从数据中提取出未知的、隐含的、有用的信息及知识,以最大程度的辅助政务部门的空间决策^[42]。充分利用大数据挖掘与分析技术,从海量的政务地理空间大数据中获取对政府部门最有价值的信息,可以为政府部门提供更加准确有效的决策服务。

政务地理空间大数据挖掘与分析涉及概率论、空间统计学、规则归纳、模糊集、云理论、粗糙集、大数据计算、人工智能、机器学习、深度学习、回归分析、聚类分析、趋势分析、关联分析、空间分析、格局分析等理论和方法^[43-46]。与传统的数据挖掘分析相比,政务地理空间大数据更加强调在隐含未知情形下对空间数据本身分析上的规律挖掘,例如进行政务地理要素统计分析、多源信息融合的组织机构空间分布格局分析、顾及时空非稳定性的地理加权回归分析、众源网络地理信息挖掘分析等。政务地理要素统计分析通过基础地理空间数据、地理国情普查数据、社会经济数据,采用空间叠加分析、聚类分析、网络分析、关联分析

等方法,结合基于 Hadoop 的海量数据处理技术可以展现交通网络、居民地与设施、地理单元等地理要素的基本状况^[47-48];多源信息融合的组织机构空间分布格局分析首先利用中文分词、语义分析和地理编码技术实现海量组织机构信息与空间信息的融合,再采用基于面的空间聚类分析方法、多中心模型拟合方法分析组织机构空间分布格局^[49]。顾及时空非稳定性的地理加权回归分析可以有效探测局部时空非平稳性,进而分析事物随时空位置变化而变化的特征,政府相关部门可以用来分析城市扩展的影响因素、各种疾病与社会自然因素的相关性、空气质量的影响因素等^[50-52];众源网络地理信息挖掘可以充分挖掘社交网络和社会化媒体中蕴含的有价值的信息,这些社会资源为政府分析决策提供了重要支撑^[53]。

2.4 政务地理空间大数据可视化技术

大数据可视化技术正逐步成为大数据时代的热点,从人作为分析主体和需求主体的角度出发,强调人机交互、符合人的认知规律的分析,将人所擅长的但机器不具备的认知能力融入到分析中^[54]。政务地理空间大数据的可视化可以为政府部门更加形象、直观地展示值得关注的重要政务事项,为领导决策提供可视化信息服务。

政务地理空间大数据的可视化分析,包含了政务地理区域统计分析可视化、政务空间位置聚类分析可视化、空间位置关联分析可视化等。例如,基于多层次网络统计地图可视化,将单一类型数据的网格化分析上升到多类型、多尺度的政务信息资源在网格内整合分析,有利于在指定区域内开展综合性政务数据空间分布、对比和关联分析,在政府区域规划、突发事件应急中可快速实现信息提取和资源整合;基于影响力叠加算法的热度图可视化和基于 k-means 算法的聚类图可视化可以有效反映政务区域范围内数量级较大的数据点的某种特征的空间态势分布情况。此外,针对政务地理空间大数据中时空信息的可视化,常与地图制图学相结合,通过对时空维度和相关信息建立可视化表征,对时空相关的规律和模式进行展示^[55]。基于时空立方体的方法用三维的方式将时空和事件展现出来,相关成果已应用于城市交通数据、飓风数据、轨迹数据等可视化分析中^[56-57]。另一方面,在政府外网也催生了大量基于位置的文本可视化分析、图可视化分析等方法。其中,文本可视化通过挖掘文本中蕴含的如主题

聚类、逻辑结构、动态演化规律等语义特征,可以进行直观的展示^[58-59];针对图可视化,相关学者围绕放射图^[60]、空间填充法^[61]、图聚簇简化可视化^[62]、多尺度交互^[63]等进行了大量研究。

2.5 政务地理空间大数据决策支持服务技术

政务地理空间大数据的最终目的是为政府管理提供决策支持服务,而面对海量的地理空间大数据,关键在于如何为政府部门寻找自己感兴趣的信息,实现智能化服务。智能检索和主动推送技术在政务综合决策中最为重要,也是政务地理空间大数据决策支持服务研究领域的热点,可以帮助工作人员从海量的数据信息中寻找有用的信息并推荐给政府部门作出合理高效的决策。

政务地理空间大数据的智能检索是在整合各种政务信息资源的基础上实现业务统一组织,依据办公人员的浏览习惯、访问记录、查询历史,对政务信息资源、网站、地图、视频等进行个性化的搜索,并主动提供相关的信息内容。其中关键在于对用户建模以及用户查询意图识别。UCAIR (user-centered adaptive information retrieval) 作为智能客户端系统,采用决策理论框架系统实现智能搜索,可以通过智能客户端进行隐式用户兴趣挖掘,且用户模型可以随时进行学习、更新^[64]。文献[65]讨论了通过分析日志文件,计算用户和对应行为的相似度,通过认知情感理论构建用户模型,并开发长期和短期模型,用于存储用户搜索相关信息。文献[66]提出了映射用户查询到类别体系,将个性化用户模型和普通用户模型相结合来识别用户兴趣。文献[67]基于用户查询意图识别给出了个性化搜索框架,提出了利用搜索语料库的公共 LDA 模型构建用户模型的方法以及新的用户查询意图识别方法。文献[68]提出了以地理空间事件为纽带,用本体驱动的地理事件相关的信息自动检索方法。文献[69]针对海量的文本、图片、视频等多模态数据检索系统深入的研究了基于深度学习的跨模态检索方法。文献[70]利用 Wikipedia 的文章和类别信息确定用户的查询意图,从而辅助信息检索。

政务地理空间大数据的主动推送服务是根据用户的兴趣偏好进行个性化的政务信息推荐。大数据时代政务信息资源虽然巨大,但蕴含用户偏好的历史记录往往较为稀疏,并不能十分精确计算用户的个性化决策需求,且政府部门一般需要更加多样性的信息推荐,这对于主动推送系统的

可解释性和可扩展性均提出了更高的要求。政务地理空间大数据的主动推送服务可以借鉴时下流行的推荐系统及相关算法,如传统的基于内容的推荐算法、协同过滤的推荐算法、混合推荐算法,以及近年来伴随社交网络和位置服务而发展的基于社交网络的推荐算法、兴趣点的推荐算法等。基于内容的推荐算法主要包含内容分析、用户文件学习、信息过滤 3 方面内容,比较典型的推荐系统有 Personal Web Watcher、CiteSeer 等^[71];协同过滤推荐算法主要基于用户的历史反馈信息进行推荐,综合利用机器学习和统计的方法进行学习预测,包含聚类模型、潜在语义模型、关联规则、深度学习等方法^[72-73];基于社交网络的推荐算法主要包括基于内存的社交网络推荐算法和基于矩阵分解的社交网络推荐算法,这类算法可以有效改进传统协同过滤算法的数据稀疏和冷启动问题^[74-75];基于兴趣点的推荐算法是利用用户的签到历史记录以及其他偏好,为用户提供个性化的兴趣点推荐服务,近年来成为数据挖掘和机器学习领域的热点^[76-77],单纯利用签到数据的兴趣点推荐方法较少,大多融合额外信息来改进兴趣点推荐算法的性能,如兴趣点地理位置、社交网络信息、时间信息、兴趣点的内容和用户评论信息等。这些推荐算法有机融合政务地理空间大数据中,可以帮助政府部门实现重要信息资源的主动推送服务,提升服务的精准性。

3 政务地理空间大数据面临的机遇及挑战

大数据技术作为电子政务发展的推进器,不仅给技术进步带来深刻变革,而且给政府协同和公共服务模式带来了创新性发展。“用数据说话、用数据决策、用数据管理、用数据创新”已经成为提升政府治理能力的新途径。在国家政策的推动下,政府电子政务中运用大数据技术已经初步建立,但相比较于专业的基础地理信息,现有的技术工具与处理方法不能高效地对政务地理空间大数据进行分析处理;同时,政府部门也对空间信息决策服务的形式、内容与能力提出了更高的要求,给面向决策服务的政务地理空间大数据带来新的机遇与挑战。

(1) 政务业务自治与碎片化条件下的多源信息资源智能聚合。单一数据源已越来越难以满足应对很多政策领域日益复杂问题的需求,在数字时代,政府治理的关键在于整合多数据源的公共

管理模式。现有的政务数据整合方法虽然在一定程度改善了政务信息孤立存储、相互隔离的状况,但依然面临着政务业务自治与碎片化条件下的多源信息资源智能聚合的挑战,政务信息资源依然存在较少顾及语义特征、缺乏有效的关联机制等现实问题,需要研究语义约束的政务数据多层次关联规则,建立政务需求与政务地理空间大数据的统一描述模型和关联约束,以及基于关联分析的政务地理空间大数据渐进求精的智能聚合方法,实现满足政府需求的政务信息主动检索与精准发现。此外,在大数据时代,数据的收集建立呈现出与传统信息数据建立原则不同的更加开放聚合的特点,不同渠道、不同维度的信息可以通过有效的整合形成丰富完整的信息档案,不但政府自己可以维护信息,居民、企业、社会组织等也可以参与到信息的维护中去。大数据平台的建设超越了原先分立的部门架构,数据的归属权不再被认为是某些部门所专有,而是从一开始就具有开放、共享的属性,不同部门可以使用大数据平台中的不同政务信息,而居民、企业和社会组织也可以依据相应的法律法规获取有效的政务信息。

(2) 多数据中心数据一体化存储管理。分布式数据库和并行计算技术虽然为解决政务地理空间大数据高效、快速、实时的存储和管理提供了较好的思路,但是政务地理空间大数据呈现多样、动态、异构的特点,并分布式存储于多个数据中心,亟须构建多数据中心数据一体化存储模型,解决海量空间数据、非空间数据的一体化混合存储、实时管理以及快速增长的增量数据的扩展问题。政务地理空间大数据所固有的数据密集性、计算密集性、并发访问密集性等特征,需要设计更加高效的面向政务地理空间大数据的高性能空间查询和管理算法以帮助人们从大规模的地理空间数据中获取有价值的知识,跨越多核和节点并行处理、高度的扩展性和容错性、支持异构环境以及较低的分析延迟和成本等,实现集数据整合、知识化管理、数据分析挖掘和可视化展示等多位一体的存储与管理。

(3) 面向政府决策的政务地理空间大数据深度分析。政务地理空间大数据呈现出粒度减小、精度提高、范围扩大的趋势,需要重点研究面向实体描述的时间序列分析与空间趋势发现、主题聚焦的互联网地理信息获取与挖掘、基于用户行为分析的地理信息主动服务等深度分析方法。未来

面向政务地理空间大数据的深度分析面越来越广,层次越来越深,要求实现决策分析的自动定制和主动提供。知识化、智能型深度分析研究将推动政务信息系统由以数据管理为主要功能向以知识管理、空间决策为主要功能转变、由数据驱动向知识和模型协同驱动转变、由定量模型分析向基于知识的智能决策方法转变、由空间信息分析应用向更广阔的多领域空间决策分析应用转变。

(4) 政务地理空间大数据智能化、主动化服务和跨部门协同应用。在社会化感知条件下,政务地理空间大数据以流式产生、管理和应用,政务地理空间大数据的业务已向资源检索、专题快速构建、数据个性化服务、信息主动推送等智能化和主动化的模式发展,相关应用呈现出跨部门协同治理与合作的趋势,设计开发面向政府协同决策和跨部门协同应用的地理空间大数据分析服务软件平台,以辅助领导科学决策为目的,并针对当前热点和应用需要,建设民生服务、大气质量监测、公共安全与应急响应、城镇化发展产业结构调整等专题应用智能化和主动化信息服务,释放政府大数据价值,可以为建设阳光政府、服务型政府提供重要信息保障。

4 结 语

我国电子政务地理信息服务的整体水平还处于发展的初级阶段,需要进一步扩展和深化,有待探索、研究的问题还很多,从技术研究到实际应用与发达国家相比还有较大差距。当前,大数据、云计算等信息技术与地理信息服务的融合给电子政务的发展带来了新的机遇和挑战。本文基于近年来在政务地理空间大数据方面研究的总结,阐述了政务地理空间大数据的内涵及特点,对政务地理空间大数据的整合、数据存储与管理、挖掘与分析以及决策支持服务涉及的理论问题、关键技术、发展趋势等方面作了分析和展望,并指出了政务地理空间大数据混合存储与动态管理、数据整合、数据挖掘与分析、决策支持服务等存在的问题及发展趋势,以期推动地理空间大数据技术在电子政务信息资源中的应用。

致谢:中国测绘科学研究院政府 GIS 研究中心董春研究员、王勇副研究员以及中国矿业大学环境与测绘学院刘春阳博士生等为本文提出了宝贵的修改意见,在此一并致谢。

参考文献：

- [1] DOCTOROW C. Big Data: Welcome to the Petacentre[J]. Nature, 2008, 455(7209): 16-21.
- [2] REICHMAN O J, JONES M B, SCHILDHAUER M P. Challenges and Opportunities of Open Data in Ecology[J]. Science, 2011, 331(6018): 703-705.
- [3] ZHENG Yu, CAPRA L, WOLFSON O, et al. Urban Computing: Concepts, Methodologies, and Applications[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2014, 5(3): 38.
- [4] ZHENG Yu. Trajectory Data Mining: An Overview[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2015, 6(3): 29.
- [5] ZHENG Yu. Methodologies for Cross-domain Data Fusion: An Overview [J]. IEEE Transactions on Big Data, 2015, 1(1): 16-34.
- [6] 刘纪平, 刘钊, 王亮. 基于功能协同的电子政务空间信息服务[J]. 测绘学报, 2006, 35(4): 299-302.
LIU Jiping, LIU Zhao, WANG Liang. Research on the Spatial Information Service for E-government Based on Function Collaboration [J]. Acta Geodaetica et Cartographica Sinica, 2006, 35(4): 299-302.
- [7] 刘纪平, 张福浩, 王亮, 等. 面向大数据的空间信息决策支持服务研究与展望[J]. 测绘科学, 2014, 39(5): 8-12, 17.
LIU Jiping, ZHANG Fuhao, WANG Liang, et al. Research and Prospect on Spatial Decision Support Service Based on Big Data[J], Science of Surveying and Mapping, 2014, 39(5): 8-12, 17.
- [8] LIU J P, XU S H, ZHANG F H, et al. Research and Implementation of Government Geographic Information Service Platform [J]. Journal of Earth Science and Engineering, 2011, 1(2): 48-54.
- [9] 刘纪平, 张福浩, 王亮, 等. 电子政务地理信息服务[M]. 北京: 测绘出版社, 2014.
LIU Jiping, ZHANG Fuhao, WANG Liang, et al. Geographic Information Service for E-government [M]. Beijing: Surveying and Mapping Press, 2014.
- [10] 赵彬彬. 多尺度矢量地图空间目标匹配方法及其应用研究 [D]. 长沙: 中南大学, 2011.
ZHAO Binbin. A Study on Multi-scale Vector Map Objects Matching Method and Its Application [D]. Changsha: Central South University, 2011.
- [11] 张云菲. 多源道路网与兴趣点的一致性整合方法[D]. 武汉: 武汉大学, 2015.
ZHANG Yunfei. Methods for Congruent Conflation of Multi-source Road Networks and POIs [D]. Wuhan: Wuhan University, 2015.
- [12] SAALFELD A. Conflation Automated Map Compilation[J]. International Journal of Geographical Information Systems, 1988, 2(3): 217-228.
- [13] MICHALOWSKI M, AMBITE J L, THAKKAR S, et al. Retrieving and Semantically Integrating Heterogeneous Data from the Web[J]. IEEE Intelligent Systems, 2004, 19(3): 72-79.
- [14] LONGLEY P A, GOODCHILD M F, MAGUIRE D J, et al. Geographic Information Systems and Science[M]. 2nd ed. [S.l.]: John Wiley & Sons, 2005.
- [15] CASADO M L. Some Basic Mathematical Constraints for the Geometric Conflation Problem[C]//Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Lisboa: Instituto Geográfico Português, 2006: 264-274.
- [16] HEARST M A, LEVY A Y, KNOBLOCK C, et al. Information Integration[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(5): 12-24.
- [17] MAY W, LAUSEN G. A Uniform Framework for Integration of Information from the Web[J]. Information Systems, 2004, 29(1): 59-91.
- [18] SESTER M, VON GÖSELN G, KIELER B. Identification and Adjustment of Corresponding Objects in Data sets of Different Origin[C]// Proceedings of the 10th AGILE International Conference on Geographic Information Science. Denmark: Aalborg University, 2007.
- [19] DALYOT S, DAHINDEN T, SCHULZE M J, et al. Integrating Network Structures of Different Geometric Representations[J]. Survey Review, 2013, 45 (333): 428-440.
- [20] 陈军, 王东华, 商瑶玲, 等. 国家 1: 50 000 数据库更新工程总体设计研究与技术创新[J]. 测绘学报, 2010, 39(1): 7-10.
CHEN Jun, WANG Donghua, SHANG Yaoling, et al. Master Design and Technical Development for National 1: 50 000 Topographic Data-Base Updating Engineering in China[J]. Acta Geodaetica et Cartographica Sinica, 2010, 39(1): 7-10.
- [21] DOERR M, HUNTER J, LAGOZE C. Towards A Core Ontology for Information Integration [J]. Journal of Digital Information, 2003, 4(1).
- [22] BUTENUTH M, GÖSELN G V, TIEDGE M, et al. Integration of Heterogeneous Geospatial Data in a Federated Database[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2007, 62(5): 328-346.
- [23] 谢潇. 语义感知的地理视频大数据自适应关联组织方法[J]. 测绘学报, 2016, 45(10): 1260. DOI: 10.11947/j. AGCS.2016.20160327.
XIE Xiao. A Semantics-aware Self-adaptive Associated Organization Method of GeoVideo Big Data [J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(10): 1260. DOI: 10.11947/j. AGCS.2016.20160327.
- [24] 唐炉亮, 杨雪, 牛乐, 等. 一种众源车载 GPS 轨迹大数据自适应滤波方法[J]. 测绘学报, 2016, 45(12): 1455-1463. DOI: 10.11947/j. AGCS.2016.20160117.

- TANG Luliang, YANG Xue, NIU Le, et al. An Adaptive Filtering Method Based on Crowdsourced Big Trace Data[J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(12): 1455-1463. DOI: 10.11947/j.AGCS.2016.20160117.
- [25] ZHENG Yu, XIE Xing, MA Weiying. GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory[J]. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2010, 33(2): 32-39.
- [26] DU Heshan, ANAND S, ALECHINA N, et al. Geospatial Information Integration for Authoritative and Crowd Sourced Road Vector Data [J]. *Transactions in GIS*, 2012, 16(4): 455-476.
- [27] LI Jun, QIN Qiming, HAN Jiawei, et al. Mining Trajectory Data and Geotagged Data in Social Media for Road Map Inference[J]. *Transactions in GIS*, 2015, 19(1): 1-18.
- [28] 范建永, 龙明, 熊伟. 基于 HBase 的矢量空间数据分布式存储研究[J]. *地理与地理信息科学*, 2012, 28(5): 39-42. FAN Jianyong, LONG Ming, XIONG Wei. Research of Vector Spatial Data Distributed Storage Based on HBase [J]. *Geography and Geo-Information Science*, 2012, 28(5): 39-42.
- [29] WANG Yonggang, WANG Sheng. Research and Implementation on Spatial Data Storage and Operation Based on Hadoop Platform [C] // *Proceedings of the 2nd IITA International Conference on Geoscience and Remote Sensing*. Qingdao: IEEE, 2010: 275-278.
- [30] GUNARATHNE T, WU T L, QIU J, et al. MapReduce in the Clouds for Science [C] // *Proceedings of the 2nd International Conference on Cloud Computing Technology and Science*. Indianapolis: IEEE, 2010: 565-572.
- [31] DUTTA H, KAMIL A, POOLEERY M, et al. Distributed Storage of Large-scale Multidimensional Electroencephalogram Data Using Hadoop and HBase [M] // FIORE S, ALOISIO G. *Grid and Cloud Database Management*. Berlin Heidelberg: Springer, 2011: 331-347.
- [32] 丁琛. 基于 HBase 的空间数据分布式存储和并行查询算法研究[D]. 南京: 南京师范大学, 2014. DING Chen. *Research on Distributed Storage and Parallel Query Algorithm of Spatial Data in HBase*[J]. Nanjing: Nanjing Normal University, 2014.
- [33] THUSOO A, SARMA J S, JAIN N, et al. Hive: A Warehousing Solution over a Map-reduce Framework[J]. *Proceedings of the VLDB Endowment*, 2009, 2(2): 1626-1629.
- [34] HAN Bo, ZHANG Yanxia, ZHONG Shoubo, et al. Astronomical Data Fusion Tool Based on PostgreSQL[J]. *Research in Astronomy and Astrophysics*, 2016, 16(11): 178.
- [35] JAYALAKSHMI T S, CHETHANA C. A Semantic Search Engine for Indexing and Retrieval of Relevant Text Documents[J]. *International Journal of Advance Research in Computer Science and Management Studies*, 2016, 4(5): 7-11.
- [36] 王珊, 王会举, 覃雄派, 等. 架构大数据: 挑战、现状与展望[J]. *计算机学报*, 2011, 34(10): 1741-1752. WANG Shan, WANG Huiju, QIN Xiongpai, et al. *Architecting Big Data: Challenges, Studies and Forecasts*[J]. *Chinese Journal of Computers*, 2011, 34(10): 1741-1752.
- [37] TerraFly[EB/OL]. <http://www.TerraFly.FIU.edu>.
- [38] AJI A, WANG Fusheng. High Performance Spatial Query Processing for Large Scale Scientific Data [C] // *Proceedings of the on SIGMOD/PODS 2012 PhD Symposium*. Scottsdale, Arizona, USA: ACM, 2012: 9-14.
- [39] 张书彬, 韩冀中, 刘志勇, 等. 基于 MapReduce 实现空间查询的研究[J]. *高技术通讯*, 2010, 20(7): 719-726. ZHANG Shubin, HAN Jizhong, LIU Zhiyong, et al. *Research on Implementing Spatial Queries Based on Map Reduce*[J]. *Chinese High Technology Letters*, 2010, 20(7): 719-726.
- [40] TAO Yufei, SHENG Cheng, PEI Jian. On K-skip Shortest Paths [C] // *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*. Athens, Greece: ACM, 2011: 421-432.
- [41] TAO Yufei, HU Xiaocheng, CHOI D W, et al. Approximate MaxRS in Spatial Databases[J]. *Proceedings of the VLDB Endowment*, 2013, 6(13): 1546-1557.
- [42] 李德仁, 王树良, 李德毅. *空间数据挖掘理论与应用* [M]. 2 版. 北京: 科学出版社, 2013. LI Deren, WANG Shuliang, LI Deyi. *Spatial Data Mining Theory and Application* [J]. 2nd ed. Beijing: Science Press, 2013.
- [43] 王树良, 丁刚毅, 钟鸣. 大数据下的空间数据挖掘思考[J]. *中国电子科学研究院学报*, 2013, 8(1): 8-17. WANG Shuliang, DING Gangyi, ZHONG Ming. *On Spatial Data Mining under Big Data*[J]. *Journal of China Academy of Electronics and Information Technology*, 2013, 8(1): 8-17.
- [44] ZHANG Junbo, ZHENG Yu, QI Dekang. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction [C] // *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco, California, USA: AAAI, 2017.
- [45] LUO Ping, LIN Fen, XIONG Yuhong, et al. Towards Combining Web Classification and Web Information Extraction: A Case Study [C] // *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris, France: ACM, 2009: 1235-1244.
- [46] 奚雪峰, 周国栋. 面向自然语言处理的深度学习研究[J]. *自动化学报*, 2016, 42(10): 1445-1465. XI Xuefeng, ZHOU Guodong. *A Survey on Deep Learning for Natural Language Processing* [J]. *Acta Automatica Sinica*, 2016, 42(10): 1445-1465.

- [47] 陈岚. 基于因子分析和聚类分析的省级政府门户网站评估[J]. 电子政务, 2010(2): 95-100.
CHEN Lan. Evaluation of Provincial Government Portal Based on Factor Analysis and Cluster Analysis[J]. E-Government, 2010(2): 95-100.
- [48] 何占军, 刘启亮, 邓敏, 等. 显著空间同位模式的多尺度挖掘方法[J]. 测绘学报, 2016, 45(11): 1335-1341. DOI: 10.11947/j.AGCS.2016.20150371.
HE Zhanjun, LIU Qiliang, DENG Min, et al. A Multi-scale Method for Mining Significant Spatial Co-location Patterns[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(11): 1335-1341. DOI: 10.11947/j.AGCS.2016.20150371.
- [49] 赵仲秋, 季海峰, 高隽, 等. 基于稀疏编码多尺度空间潜在语义分析的图像分类[J]. 计算机学报, 2014, 37(6): 1251-1260.
ZHAO Zhongqiu, JI Haifeng, GAO Jun, et al. Sparse Coding Based Multi-scale Spatial Latent Semantic Analysis for Image Classification[J]. Chinese Journal of Computers, 2014, 37(6): 1251-1260.
- [50] LIU Jiping, XU Shenghua, ZHANG Fuhao, et al. A Hybrid Genetic-ant Colony Optimization Algorithm for the Optimal Path Selection[J]. Intelligent Automation & Soft Computing, 2017, 23(2): 235-242.
- [51] LIU Jiping, YANG Yi, XU Shenghua, et al. A Geographically Temporal Weighted Regression Approach with Travel Distance for House Price Estimation[J]. Entropy, 2016, 18(8): 303.
- [52] 赵阳阳, 刘纪平, 徐胜华, 等. 一种基于半监督学习的地理加权回归方法[J]. 测绘学报, 2017, 46(1): 123-129. DOI: 10.11947/j.AGCS.2017.20150470.
ZHAO Yangyang, LIU Jiping, XU Shenghua, et al. A Geographic Weighted Regression Method Based on Semi-Supervised Learning[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(1): 123-129. DOI: 10.11947/j.AGCS.2017.20150470.
- [53] 胡庆武, 王明, 李清泉. 利用位置签到数据探索城市热点与商圈[J]. 测绘学报, 2014, 43(3): 314-321. DOI: 10.13485/j.cnki.11-2089.2014.0045.
HU Qingwu, WANG Ming, LI Qingquan. Urban Hotspot and Commercial Area Exploration with Check-in Data[J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(3): 314-321. DOI: 10.13485/j.cnki.11-2089.2014.0045.
- [54] KEIM D, QU Huamin, MA K L. Big-data Visualization[J]. IEEE Computer Graphics and Applications, 2013, 33(4): 20-21.
- [55] HEY T, GANNON D, PINKELMAN J. The Future of Data-intensive Science[J]. Computer, 2012, 45(5): 81-82.
- [56] DEMŠAR U, VIRRANTAUŠ K. Space-time Density of Trajectories: Exploring Spatio-Temporal Patterns in Movement Data[J]. International Journal of Geographical Information Science, 2010, 24(10): 1527-1542.
- [57] TOMINSKI C, SCHUMANN H, ANDRIENKO G, et al. Stacking-based Visualization of Trajectory Attribute Data[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(12): 2565-2574.
- [58] VIEGAS F B, WATTENBERG M, FEINBERG J. Participatory Visualization with Wordle[J]. IEEE Transactions on Visualization and Computer Graphics, 2009, 15(6). DOI: 10.1109/TVCG.2009.171.
- [59] CUI Weiwei, WU Yingcai, LIU Shixia, et al. Context Preserving Dynamic Word Cloud Visualization[C] // Proceedings of 2010 IEEE Pacific Visualization Symposium. Taipei: IEEE, 2010: 121-128.
- [60] GOU Liang, ZHANG Xiaolong. TreeNetViz: Revealing Patterns of Networks over Tree Structures[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(12): 2449-2458.
- [61] 任磊, 王威信, 滕东兴, 等. 面向海量层次信息可视化的嵌套圆鱼眼视图[J]. 计算机辅助设计与图形学学报, 2008, 20(3): 298-303, 309.
REN Lei, WANG Weixin, TENG Dongxing, et al. Fisheye View for Visualization of Large Tree by Packing Nested Circles[J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(3): 298-303, 309.
- [62] SELASSIE D, HELLER B, HEER J. Divided Edge Bundling for Directional Network Data[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(12): 2354-2363.
- [63] ABELLO J, VAN HAM F, KRISHNAN N. ASK-GraphView: A Large Scale Graph Visualization System[J]. IEEE Transactions on Visualization and Computer Graphics, 2006, 12(5): 669-676.
- [64] SHEN Xuehua, TAN Bin, ZHAI Chengxiang. Implicit User Modeling for Personalized Search[C] // Proceedings of the 14th ACM International Conference on Information and Knowledge Management. Bremen, Germany: ACM, 2005: 824-831.
- [65] ZHOU Xiaoming, CONATI C. Inferring User Goals from Personality and Behavior in A Causal Model of User Affect[C] // Proceedings of the 8th International Conference on Intelligent User Interfaces. Miami, Florida, USA: ACM, 2003: 211-218.
- [66] LIU Fang, YU C, MENG Weiyi. Personalized Web Search by Mapping User Queries to Categories[C] // Proceedings of the 11th International Conference on Information and Knowledge Management. McLean, Virginia, USA: ACM, 2002: 558-565.
- [67] 林博. 基于用户查询意图识别的个性化搜索方法研究[D]. 湘潭: 湘潭大学, 2015.
LIN Bo. Research of Personalized Search Based on User Query Intent Identification[D]. Xiangtan: Xiangtan University, 2015.
- [68] 刘纪平, 栗斌, 石丽红, 等. 一种本体驱动的地理空间事

件相关信息自动检索方法[J]. 测绘学报, 2011, 40(4): 502-508.

LIU Jiping, LI Bin, SHI Lihong, et al. An Automated Retrieval Method of Geo-spatial Event Information Based on Ontology[J]. Acta Geodaetica et Cartographica Sinica, 2011, 40(4): 502-508.

[69] 冯方向. 基于深度学习的跨模态检索研究[D]. 北京: 北京邮电大学, 2015.

FENG Fangxiang. Deep Learning for Cross-Modal Retrieval[J]. Beijing: Beijing University of Posts and Telecommunications, 2015.

[70] HU Jian, WANG Gang, LOCHOVSKY F, et al. Understanding User's Query Intent with Wikipedia[C]// Proceedings of the 18th International Conference on World Wide Web. Madrid: ACM, 2009: 471-480.

[71] BOLLACKER K D, LAWRENCE S, GILES C L. Discovering Relevant Scientific Literature on the Web [J]. IEEE Intelligent Systems and their Applications, 2000, 15(2): 42-47.

[72] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based Collaborative Filtering Recommendation Algorithms [C]// Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China: ACM, 2001: 285-295.

[73] WANG Hao, WANG Naiyan, YEUNG D Y. Collaborative Deep Learning For Recommender Systems[C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2015: 1235-1244.

[74] JAMALI M, ESTER M. TrustWalker: A Random Walk Model for Combining Trust-based and Item-based Recommendation[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France: ACM, 2009: 397-406.

[75] MA Hao, KING I, LYU M R. Learning to Recommend with Social Trust Ensemble[C]// Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Boston, MA, USA: ACM, 2009: 203-210.

[76] CHENG Chen, YANG Haiqin, KING I, et al. Fused Matrix Factorization with Geographical and Social Influence in Location-based Social Networks[C]// Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Ontario, Canada: AAAI, 2012(12): 1.

[77] YE Mao, YIN Peifeng, LEE W C, et al. Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation [C] // Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing: ACM, 2011: 325-334.

(责任编辑:宋启凡)

收稿日期: 2017-06-19

修回日期: 2017-08-28

第一作者简介: 刘纪平(1967—),男,博士,研究员,博士生导师,研究方向为政务地理空间大数据、政务地理信息服务、应急地理信息服务。

First author: LIU Jiping (1967—), male, PhD, research fellow, PhD supervisor, majors in geospatial big data for E-government, government geographic information services, emergency geographic information services.

E-mail: liujp@casm.ac.cn

通信作者: 徐胜华

Corresponding author: XU Shenghua

E-mail: xushh@casm.ac.cn