

论大数据视角下的地球空间信息学的机遇与挑战

李德仁^{1,2}, 张过², 蒋永华¹, 沈欣², 刘伟玲²

1. 武汉大学遥感信息工程学院, 湖北 武汉 430079;
2. 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉 430079

摘要

大数据时代已经到来, 并且已经深入人类生活的方方面面。作为地球科学与信息科学交叉融合催生出的地球空间信息学, 大数据时代的来临在为其提供更丰富的数据保障的同时, 也带来数据存储、管理、分析和挖掘方面的新挑战, 甚至造成了某种程度上的“数据爆炸”。从大数据视角, 梳理了当前地球空间信息学涉及的地理信息系统、智慧城市、遥感大数据和空间数据挖掘4个核心领域的瓶颈和挑战; 指出在大数据时代, 地球空间信息学可为地球科学研究提供更加精准、实时的空间信息框架和更加智能高效的信息处理手段, 从而服务于智慧城市、智慧地球建设和人类社会的可持续发展。而且, 大数据时代下, 地球空间信息学的发展面临着软件和硬件水平的双重考验。

关键词

大数据; 地理信息系统; 智慧城市; 遥感大数据; 空间数据挖掘

中图分类号: TP79

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022012

Opportunities and challenges of geo-spatial information science from the perspective of big data

LI Deren^{1,2}, ZHANG Guo², JIANG Yonghua¹, SHEN Xin², LIU Weiling²

1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
2. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract

The era of big data has arrived, and it has penetrated every aspect of human life. As the geo-spatial information science spawned by the intersection of earth sciences and information sciences, the advent of the era of big data provides it with richer prosperous data protection, but also brings new challenges in data storage, management, analysis, and mining, and even caused a certain degree of “data explosion”. From the perspective of big data, the bottlenecks and challenges in the four core areas of geographic information systems, smart cities, remote sensing big data, and spatial data mining

were sorted out. And it was pointed out that geo-spatial information science can provide more accurate and real-time spatial information frameworks and more intelligent and more efficient information processing methods for geoscience research, serving intelligent cities, smart earth construction, and sustainable development in the era of big data. Moreover, in the era of big data, the development of geo-spatial information science is facing the double test of software and hardware levels.

Key words

big data, geographic information system, smart city, remote sensing big data, spatial data mining

1 大数据时代的特征和挑战

1.1 大数据时代的特征

近年来,随着信息科技和网络技术的发展,人类在网络上交换的数据、传达的信息使得全球每日产生的数据超乎想象,全球的数据规模呈爆发式增长。2008年9月, *Nature*发表“big data”专刊文章,这标志着大数据概念的产生^[1]。2011年 *Science*发表对大数据进行探讨的专刊,这预示着大数据时代的到来^[2]。在2012年,美国正式开启了“大数据研究和发展倡议”计划^[3]。

2013年国家自然科学基金委员会专门开设了“大数据”项目群。大数据涵盖人类社会生活的方方面面,蕴含着巨大的社会、经济、科研价值,学者们常以未来世界的“石油”来表达它的重要性,它也早已成为企业界、科技界以及政界关注的热点领域。而在生活中,人类早已踏入了大数据时代。在互联网方面,Google在1 min内能收到200万次搜索查询请求,Facebook在1 min内能收到68万次搜索请求,百度每天需要处理的网页总数达到了1万亿,腾讯QQ和微信注册用户数分别高达8亿和5亿;在视频数据方面,我国摄像头总数超过2 000万,每个摄像头每小时可产生3.6 GB数据,数据量达到PB级甚至EB级;在交通出行方面,民航飞机单个引擎在1 h内产生

20 TB数据,北京市政交通一卡通每天的使用量约为4 000万次,产生30 GB的存储量;在用电量方面,国家电网年均产生数据约510 TB(不含视频),累计产生数据约5 PB;在医疗方面,单个病人的计算机断层扫描(computed tomography, CT)影像数据量达到了数十GB。各行各业的数据量都在不断地飞速增长。据国际数据公司(International Data Corporation, IDC)发布的《数据时代2025》预测,2025年人类的数据量将达到163 ZB,与2016年相比,数据量增加了近10倍^[4],并且涵盖结构化数据和非结构化数据,其中95%的数据是不精确的、非结构化的数据^[5]。

国际上针对大数据的工程技术研究走在科学研究的前面^[6]。如上所述,大数据最大的特征是数据量巨大,而它的价值重在其本身的隐含价值。现在对大数据特征的描述已经从3V、4V到5V,甚至新的3V^[7]。大数据的特征可总结为以下几点。

- 体量大(volume): 各行各业每天生产的数据的量达到EB甚至ZB级。
- 速度快(velocity): 对数据实时性要求高,要求数据响应速度快。
- 模态多样(variety): 数据种类众多。
- 真伪难辨(veracity): 数据来源较广,数据类型多样,数据存在缺失、歧义以及噪声等问题。
- 价值巨大(value): 大数据作为未来世界的“石油”,对于社会发展具有巨大的隐藏价值。

1.2 大数据时代带来的挑战

大数据体量大、速度快、模态多样等特征^[7]给当前各行各业应对大数据的存储、读取、处理、应用以及隐含规律的挖掘带来了严峻的挑战。当前的问题是,面对如今体量大、速度快、模态多样和真伪难辨的数据增长模式,如何有效地挖掘出大数据的巨大价值。地球科学方面的研究涉及政治、经济和文化发展,大数据时代也给地学研究带来了严峻的挑战,尤其是地球科学空间数据存储和潜在价值挖掘方面。下面从大数据时代的地理信息系统(geographic information system, GIS)、智慧城市、遥感大数据和空间数据挖掘等方面总结当今大数据时代下地球空间信息学的发展。

2 大数据时代的GIS

2.1 大数据GIS的机遇

GIS具有空间数据管理能力、空间分析能力以及地图数据可视化能力^[7]。虽然大数据时代给GIS的发展及应用带来了巨大的挑战,但大数据没有改变GIS的特征,而是对GIS进行了扩展,促使其进一步顺应时代的发展,这也是GIS进一步发展的机遇^[7]。在大数据时代, GIS应具有以下基本特征。

(1) 可扩展的动态数据管理模式^[7]

传统GIS的静态关系型数据存储模式已无法满足大数据时代GIS的需求,因此,在架构层面,大数据GIS需要满足扩展性,以实现动态无限增长的存储和查询的目标^[7]。目前,人们比较熟知的MapReduce

框架为当前动态数据管理模式的发展提供了很好的范例。

(2) 实现结构化数据和非结构化数据存储

关系型数据库难以满足大数据中非结构化数据的存储需求。近些年发展起来的键-值数据库(key-value store)是非关系型数据库的代表,该数据库可自由添加字段,满足非结构化数据的存储需求,但存在数据冗余的现象。虽然非关系型数据库在数据存储方面存在冗余问题,但非关系型数据库依然是空间数据库管理和存储的重要方式。为了满足数据存储的需求,未来将会出现兼顾关系型数据库和非关系型数据库的统一数据管理模型,实现结构化数据和非结构化数据的存储功能。

(3) 满足可视化分析

在“3S(包括遥感(remote sensing, RS)、GIS、全球定位系统(global positioning system, GPS))”中, GIS更注重数据的分析以及数据的可视化表达。在大数据时代,地球空间信息科学的内涵没有发生改变,但它的内容和形式变得更加丰富。目前, GPS数据可被理解为任何可以标识空间位置的数据, RS数据可被理解为多源传感器数据, GIS则将这些与空间相关的数据映射到空间基准下进行统一的管理、分析和显示^[7]。传统GIS可视化更加注重符号、尺度和三维等问题的可视化表达,但对于大数据时代下的GIS,若不对信息进行提炼和综合而直接以点、线、面等符号表达出来,不仅达不到信息传递的效果,反而会因为有效特征得不到表达而适得其反。在大数据时代,不同分辨率用于数字孪生的真三维室内外地上地下一体化实景模型将导致GIS数据的数据量和复杂度的提高,从而给数据探索、分析、理解和呈现带来巨大的挑战,这使得数据可视化基础逐渐成为大数据时代的显学。

2.2 大数据时代传统GIS的挑战

大数据时代传统GIS面临的挑战可总结为以下几点^[7]。

(1) 大数据体量大带来的挑战

海量的数据问题一直是GIS研究的一个重要课题,在从海量数据跨越到大数据的时代背景下, GIS发展也遇到了数据的问题,尤其是海量数据如何存储以及如何高效利用。GIS一般采用关系型数据库来存储数据,但是关系型数据库在海量数据存储方面未能解决可扩展性问题,因此, GIS的应用受到了限制^[3]。云计算和云存储以及云端服务模式的出现,为GIS解决大数据存储以及管理等问题带来了曙光。在云服务模式下,数据可以按照需求存储在不同的磁盘、不同的机器甚至不同的地点,但是这种模式在解决大数据体量大带来的存储问题的同时,也因为现有的存储模式带来了数据索引和数据查询方面的问题。因此,针对空间大数据的数据划分,基于内存的索引,历史、当前及未来时空索引的并发控制,以及基于多线程的并发连续查询等,也是亟须深入研究的方向^[3]。

(2) 大数据流质性带来的挑战

传统GIS的空间数据库存储的数据以静态数据为主,而大数据时代下GIS面临的是流数据(streaming data)的需求^[6]。流数据是由基于时间序列动态增加的数据观测值组成的序列,具有连续性和无限增长性的特征^[6-7]。如传统的ArcGIS文件型数据库,其存储量一般为2 GB,远远不能满足大数据的流质性特征。此外,现在的空间数据库的存储方式是静态的关系型数据,需要具有数据结构的详细定义以及限定的数据大小,并且数据的查询方式也是针对静态数据的查询,因此,现有GIS空间数据的存储以及管理方式都不再适用于大

数据时代的空间流数据。

(3) 大数据模态多样带来的挑战

在大数据时代,采集的数据通常包括结构化数据和非结构化数据,如图片、视频、文本、音频、3D模型、实景图片等,类型多样,且数据内容也千差万别,数据格式的多样性给数据存储带来了巨大的挑战^[6]。传统GIS数据结构一般包括栅格数据和矢量数据,很少涉及非结构化数据,而在大数据时代,非结构化数据的比重越来越大,如3D模型和实景图片等为人类的出行带来了极大的便利,却也给GIS分析和应用带来了一定的挑战。目前,常用的关系型数据库在面对非结构化数据时也难以进行有效管理,因此,如何实现对结构化数据和非结构化数据的统一管理、分析和利用也是GIS面临的一个重大挑战。

(4) 大数据难以挖掘的隐含价值带来的挑战

在大数据时代,流质性是导致大数据数据量庞大的原因,庞大的数据量给大数据带来了真实性以及多样性的挑战^[9]。大数据真正的价值在于其蕴含的潜在规律和知识。现实中的空间是一个复杂的系统,其具有4个方面的特征:非线性、多参数、不稳定性 and 不确定性。这对现有的统计方法提出了更高的要求,统计的本质是通过对样本进行描述来推出整体,进而陈述事实,最终进行确定性分析。而在大数据时代,数据本身的复杂性、不确定性和存在的噪声使得通过现有的统计学方法发掘数据隐含的规律和知识的难度加大。目前,数据挖掘技术的不断发展使得数据分析更加注重的是价值的发现而不是预期的描述,但对于大数据存在的不确定性问题,数据挖掘也无法解决。因此,对于大数据GIS,空间统计模型的选择、参数的训练至关重要,可进一步探索基于人工智能方法的空间数据挖掘算法。

2.3 大数据GIS的未来展望

在大数据时代,大数据的特征给GIS带来了挑战,未来GIS需要解决以下几个问题:一是面对超大规模的数据,如何进行高效的管理、查询和使用等;二是面对大数据的不确定性、非线性等,如何选择高效正确的统计模型等;三是如何发展出高效的空数据挖掘算法,以快速发现大数据中隐含的价值;四是如何高效地实现超大规模时空数据的可视化分析等。目前,大数据GIS亟须一整套科学的、系统的理论和方法来应对大数据背景带来的挑战。

3 大数据时代的智慧城市

3.1 智慧城市的概念和发展

1993年美国启动的“信息高速公路”计划和“九五”期间中国启动的“八金”工程标志着城市信息化建设的起步^[4]。1998年,时任美国副总统戈尔提出了“数字地球”的概念,这标志着城市信息化逐渐进入数字城市建设的新阶段^[9]。而我国的国家地理信息公共服务平台“天地图”代表了我国300多个城市初步建成数字城市的缩影,为数亿网民提供了高效的服务。2008年IBM公司提出了“智慧城市”新理念。2009年,IBM公司的彭明盛向美国政府提出了智慧城市建设的意见,它标志着智慧城市从数字阶段迈进智慧阶段^[5]。至此,智慧城市的概念逐步完善,智慧城市的基础设施建设也取得了理论和技术上的进步。与早期的信息基础设施和数字城市建设相比,智慧城市更加注重信息资源的集成、共享、整合和服

务,也更加强调城市管理方面的统筹与协调。智慧城市建立在数字城市的基础框架上,通过无所不在的传感器将它与现实城市关联起来,将海量数据存储、计算、分析和决策交由云计算平台处理,并按照分析决策结果对各种设施进行自动化的控制^[10-11]。智慧城市通过物联网将数字城市和现实城市进行有机的融合,形成虚实一体化的空间,达到数字型孪生城市的水平。为了便于理解,可用如下表达式理解智慧城市:
智慧城市=数字城市+物联网+云计算^[5] (1)

3.2 智慧城市中大数据带来的挑战

智慧城市是基于数字城市、物联网和云计算建立的现实世界与数字世界的融合,以实现对人和物的感知、控制和智能服务。它借助物联网将现实城市和数字城市连接在一起,并借助由海量传感器组成的物联网不断地采集城市中的人、车、物、水、电、气等海量数据,通过对数据进行存储、处理、查询和分析,为城市提供智慧服务的实时反馈。智慧城市对智慧的需求使得对大数据存储、处理、查询以及分析的实时性需求也越来越高,进而导致大数据时代下的智慧城市面临诸多挑战^[8,12]。

(1) 大数据存储的问题

目前来看,存储技术成本的下降速度与数据增长速度的差距仍较大,尤其在当下,海量数据的存储带来了巨大的成本障碍。例如天津市的安防系统,在理想情况下数据存储需要的空间是4.6 EB,需要投入500亿元的资金,由此可见大数据带来的存储成本之大。为了解决数据存储带来的巨大成本问题,我国采用两种方式降低成本,分别是缩短数据保存的时限和降低数据的存储质量。

(2) 大数据检索、信息提取的问题

在大数据时代,数据的检索和信息的

提取不再是对数据进行简单的数据存储和采集,而是对行为等关键信息一并进行提取和分析。随着卫星传感器的发送,各种类型的遥感卫星每天可产生高达PB级的数据,城市中摄像头每天可采集EB级的数据,并且城市采集的数据包括一些异常信息,如交通事故、盗窃等行为,可针对这些异常现象进行预警以及及时出警处理,从而全方位地保障人民的生命财产安全。但是如何实现信息的高效检索和语义信息提取,是当前信息系统在大数据时代下面临的重要难题,为此需寻求更好的解决方法。

(3) 大数据隐含知识挖掘问题

大数据中往往隐含着丰富的知识和规律,但这些知识和规律需要不断地挖掘才能发现,并且这些知识通常不是简单地通过统计分析、回归处理便可以得出。此外,大数据还具有难以实现自动化、难以实现继承管理等问题,尤其是面对空间数据时,这种问题更加突出,导致想要在空间大数据中有效地挖掘知识和规律变得异常困难。因此,对于智慧城市,如何挖掘这些有效信息一直是大数据面临的重要挑战。

3.3 大数据时代智慧城市的未来展望

云计算和数据挖掘等技术的发展为智慧城市有效应对大数据时代的挑战带来了技术支撑。智慧城市在面对各类庞大的数据量时,尤其是面对空间和视频等非结构化数据时,应充分发挥云计算的优势对数据进行有效的存储和管理以及快速检索。在数据存储层面,基于云计算基数,构建基于时空信息云平台的云存储与智能压缩算法,以解决数据存储问题。在数据检索和信息提取层面,基于云计算技术,建立从基础设施、数据、平台到服务的一体化

时空信息云平台,实现对空间大数据、视频大数据的有效管理^[10-11]。在数据挖掘层面,应用机器学习和人工智能方法构建神经网络以及决策树、遗传算法等方法,揭示数据间的内在关系,发现深层次的模式、规则和知识。智慧城市建设过程中产生的大数据问题既是下一代的科学前沿问题,也是推进智慧城市发展的源动力^[5,10]。

4 大数据时代的遥感大数据

4.1 大数据时代遥感大数据的发展

在遥感以及对地观测研究领域,对地观测技术的不断进步使得人类对地球的综合观测能力达到了空前的水平。并且,随着不同的波段、不同的成像方式以及不同分辨率数据的不断出现,遥感数据也逐渐实现了多元化^[13]。目前,遥感数据的数据量也在显著增加,甚至呈现出指数级的增长模式。获取数据的速度得到提高,时效性也得到了有效的提高。遥感数据也呈现出了明显的“大数据”特征。但现在存在的重要问题是,遥感信息处理能力仍旧十分低下,现有的遥感数据处理和分析仍未实现普适性,也仍未实现多源异构数据的处理。遥感信息处理技术与数据获取能力之间出现了严重失衡的现象,导致遥感信息的处理仍旧处于“数据到数据”的阶段,遥感信息知识获取能力不足,致使遥感信息得不到充分的应用,形成了“大数据,小知识”的悖论。更严重的现象是,由于海量遥感数据的堆积,遥感大数据占用了大量的存储空间,造成了某种程度上的“数据灾难”^[14-15]。

遥感大数据的价值与大数据的价值相符,重在数据隐藏的规律和知识,而不是其数据量大。同样,遥感大数据的目标

也和大数据一样,在于从数据中挖掘隐藏的知识以及对数据进行智能化和自动化分析,实现从遥感数据到知识的转变,突破“大数据,小知识”的瓶颈。遥感大数据的自动分析是进行遥感大数据信息挖掘、实现遥感观测数据向知识转化的前提,其主要目的是能够建立统一、紧凑、语义的遥感大数据标识,以便为后续的数据挖掘奠定基础^[16]。遥感大数据的自动分析主要包含数据的表达、检索和理解等方面。

4.2 大数据时代遥感大数据挖掘面临的挑战

数据挖掘的含义是从大量数据中搜索出隐藏的信息,数据挖掘是现阶段大数据处理的有效方法和重要手段。而针对遥感大数据挖掘,数据挖掘指从遥感大数据中发现地表的变化规律,探索自然和社会的变化趋势。遥感大数据挖掘可总结为以下几个步骤:数据获取、数据存储、数据分析和处理、数据可视化及数据挖掘。其中,耗费成本最大的是数据存储过程,这是目前数据存储技术的发展远远跟不上数据量增长的速度导致的。难度最大的是数据挖掘过程,这是因为它依赖于基于大数据和知识库的智能推理等理论和技术的支撑。数据挖掘过程中常用的方法有人工神经网络模型、决策树模型、云模型和深度学习等方法。

天空地传感技术的发展促进了遥感大数据的发展,新一代算力基础设施为遥感大数据的发展带来了基础支撑,如云计算平台,但同时也对遥感大数据的发展提出了新的挑战,一是面对每天生产出来的海量数据,如何对数据进行有效的存储;二是如何构建更高效的数据检索和存储模型,这对于数据应用来说至关重要;三是如何进行高效的数据挖掘,挖掘潜在的地学规律

和知识。因此,目前亟须寻求自动化的数据智能处理和挖掘方法,发展可以对空间地理分布的视频数据、影像数据进行数据融合和时空数据挖掘的新理论和新算法^[6]。

4.3 大数据时代遥感大数据挖掘的潜在应用

遥感大数据挖掘可实现地球各位置、各尺度变化规律的发掘,还可实现未知地球规律的发掘,如夜光遥感技术在社会科学方面、经济发展方面的应用等。未来10年,我国遥感数据的种类和数量将呈现飞速增长的趋势,而对地观测的深度和广度也会快速发展,因此,亟须开展遥感大数据的进一步深入研究,以使我国的遥感数据物尽其用^[6]。卫星成功发射和数据收集是遥感对地观测的第一步,而更大的挑战是如何实现高效地存储、分析和利用数据,为国家发展政策的制定、经济的发展提供支撑。目前,将遥感大数据转化成知识是主要的理论挑战和技术瓶颈,如果可以突破这一瓶颈,就可以提高遥感数据的利用效率,加强遥感在环境遥感、城市规划、地形图更新、精准农业和智慧城市等方面的应用。因此,重视和抓紧遥感大数据的研究不仅具有非常重要的学术价值,而且具有重要的现实意义^[16-17]。

5 大数据时代的空间数据挖掘

5.1 空间大数据挖掘的由来

1989年,从数据库中发现知识(knowledge discovery from database, KDD)的概念在美国底特律召开的国际人工智能联合会议(IJCAI)上被提出^[18]。

1994年, LI D R等人^[19]在加拿大渥太华举行的GIS国际学术会议上提出了从GIS数据库中发现知识(knowledge discovery from GIS, KDG)的概念。自此人们逐渐认识到,很多知识原来就隐藏在大量的数据之中。经过不断的努力, KDG已经发展成系统的空间数据挖掘理论。王树良^[20]率先从GIS空间数据中发现了用于指导GIS空间分析的知识,这进一步引导了地球空间信息从狭义向广义发展。1995年,在加拿大召开的第一届知识发现和数据挖掘国际学术会议上,数据挖掘的概念被提出。由于数据挖掘经常被认为是KDD的一部分,通常将其联合称为数据挖掘和知识发现(data mining and knowledge discovery)。李德仁等人^[16]把KDG概念进一步发展为空间数据挖掘和知识发现(spatial data mining and knowledge discovery),以克服空间灾难,用知识指导数据利用,奠定了空间数据挖掘和知识发现在地球空间信息学中的学科位置。此外,在不引起歧义的情况下,空间数据挖掘和知识发现可被简称为空间数据挖掘(spatial data mining, SDM)^[3,6,21]。

目前,空间数据挖掘已经渗入数据挖掘、知识发现以及地球空间信息学等相关学科的学术活动中^[4]。1997年,亚太地区、欧洲相继召开数据挖掘学术会议。2005年,现代数据挖掘及其应用国际学术会议召开,该会议现在被学者们列为数据挖掘领域的旗舰会议。空间数据挖掘的概念起源于国际GIS会议,空间数据挖掘后续逐渐成为各种GIS会议、国际摄影测量与遥感学会(International Society for Photogrammetry and Remote Sensing, ISPRS)等会议的主题,也逐渐成为SCI、EI等数据库的收录期刊的重要刊载内容^[21]。现在,空间数据挖掘已在人们认识和改造自然中发挥了重要的作用,也得到了国内外众多学者的青睐。

5.2 大数据时代空间数据挖掘价值

空间数据挖掘是从空间数据集中提取事先未知却潜在有用的一般规则的过程。空间数据挖掘可以为基于空间数据的应用提供有价值的知识,带来巨大的收益,是人类认识复杂系统的新思维、新手段,促进经济转型增长的新引擎,提升国家综合能力和保障国家安全的新利器,提升政府治理能力的新途径^[16-17,20]。

空间数据挖掘的价值可总结为以下几个方面^[21]。

(1) 助力提升数据竞争力

目前数据利用能力已成为对比各国综合国力和技术力量的重要指标。尤其在全球化时代,基于数据利用能力的数据流,能够在社会分工协作中引领技术流、物质流、资源流和人才流,增强空间数据的主权保护能力。空间数据挖掘可以实现数据资源的潜在价值,揭示新的知识发现,尽可能发挥数据资源的战略作用,全面提升数据的规模、质量和应用水平,进而提升国家竞争力,推动社会进步和可持续发展。

(2) 助推社会生产力发展

空间数据挖掘是一种空间决策支持技术,注重从数据中挖取未知却有用的知识,最大限度地提升数据资源的有效利用能力,实现更准确的监测、分析、预警、预测,为决策者提供极有价值的知识,提高决策的针对性、科学性和可靠性。数据挖掘已经渗透到社会经济活动中,以数据流推动社会生产要素的网络化共享、集约化整合、协作化开发和高效化利用,将影响社会分工协作的未来产业格局。高端智能的数据挖掘产业新生态,能够促进传统产业转型升级,培育新的经济增长点,驱动业务增值,提升经济运行水平和效率。

(3) 助力可持续发展

空间数据既有社会公益性,又有综合价值,是解决人口、环境、资源和灾害等重大社会可持续发展问题的基础资源。人类社会的可持续发展取决于合理利用资源和保护生态资源,有效地开发和利用数据资源能够帮助人类降低对物质和能量资源的巨量消耗。而空间数据挖掘即可实现以上需求。空间数据挖掘还能辅助发现新空间对象、浏览空间数据库、辅助空间决策、理解空间数据、发现空间联系以及空间数据与非空间数据之间的关系^[9],优化空间查询。

5.3 大数据时代空间数据挖掘的未来展望

面对每天增长的海量数据,空间数据挖掘由于空间数据库缺乏“谁(who)”“什么(what)”“时间(when)”“地点(where)”“原因(why)”“如何(how)”的信息和元数据,导致数据难理解、难组织、难整合、难迁移等问题,限制了空间数据挖掘的能力。目前,由于数据库技术的限制,多源高维数据给空间数据挖掘带来机遇的同时,也带来了挑战,尤其在大数据时代,空间数据挖掘已经渗透到国家安全、经济发展等各个层面,是保证在全球化时代海量空间数据得到有效利用的重要技术支持,未来也应投入更多的人力、财力以促进空间数据挖掘的发展。

6 总结

6.1 大数据时代地球空间信息学面临的挑战

大数据已经渗透到生活的方方面面,

人类也早已接触到大数据。人类社会每天以EB甚至PB的数据级别产生数据,但数据的实际利用率却很低,长时间的数据堆积甚至造成了某种程度上的“数据爆炸”。大数据时代下的GIS、智慧城市、遥感大数据、空间数据挖掘在面临EB甚至PB级的数据量时,都表现出了海量数据存储成本高、数据管理分析难、数据使用效率低、数据挖掘难等问题,其中最重要的是在海量数据下如何用好大数据这份科学界的“石油”。大数据时代下地球空间信息学面临的是软件水平、硬件水平的双重考验。

6.2 大数据时代地球空间信息学的发展展望

本文从大数据时代的GIS、智慧城市、遥感大数据和空间数据挖掘的发展、瓶颈等维度总结了大数据时代地球空间信息学的发展以及面临的问题。首先面临的是数据存储的问题,但这一问题的解决不仅依赖硬件水平的提高,也依赖数据存储算法以及云计算的发展;其次是数据检索提取,由于数据结构复杂,高效提取需要的数据是难题;最后是数据潜在信息挖掘,这是发挥大数据这份科学界的“石油”的重中之重。在多学科交叉的推动下,大数据时代地球空间信息学的发展模式要求软硬件必须同时发展。

参考文献:

- [1] HOWE D, COSTANZO M, FEY P, et al. Big data: the future of biocuration[J]. Nature, 2008, 455(7209): 47-50.
- [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] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 2016, 45(4): 379-384.
LI D R. Towards geo-spatial information science in big data era[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(4): 379-384.
- [4] 李德仁. 信息高速公路、空间数据基础设施与数字地球[J]. 测绘学报, 1999, 28(1): 1-5.
LI D R. NII, NSDI and digital earth[J]. Acta Geodaetica et Cartographic Sinica, 1999, 28(1): 1-5.
- [5] 李德仁, 姚远, 邵振峰. 智慧城市中的大数据[J]. 武汉大学学报·信息科学版, 2014, 39(6): 631-640.
LI D R, YAO Y, SHAO Z F. Big data in smart city[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6): 631-640.
- [6] 李德仁, 张良培, 夏桂松. 遥感大数据自动分析与数据挖掘[J]. 测绘学报, 2014, 43(12): 1211-1216.
LI D R, ZHANG L P, XIA G S. Automatic analysis and mining of remote sensing big data[J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(12): 1211-1216.
- [7] 李清泉, 李德仁. 大数据GIS[J]. 武汉大学学报·信息科学版, 2014, 39(6): 641-644.
LI Q Q, LI D R. Big data GIS[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6): 641-644.
- [8] 李德仁. 论时空大数据的智能处理与服务[J]. 地球信息科学学报, 2019, 21(12): 1825-1831.
LI D R. The intelligent processing and service of spatiotemporal big data[J]. Journal of Geo-Information Science, 2019, 21(12): 1825-1831.
- [9] GORE A. The digital earth: understanding our planet in the 21st century[J]. Photogrammetric Engineering & Remote Sensing, 1998, 65(5): 1-4.
- [10] 李德仁. 基于云计算的智慧城市运营脑[J]. 卫星应用, 2018(8): 8-12.
LI D R. Smart city operation brain based on cloud computing[J]. Satellite Application, 2018(8): 8-12.
- [11] 李德仁. 数字孪生城市 智慧城市建设的新高度[J]. 中国勘察设计, 2020(10): 13-14.
LI D R. Digital twin city smart city construction new height[J]. China Engineering & Consulting, 2020(10): 13-14.
- [12] 李德仁. 从测绘学到地球空间信息智能服务科学[J]. 测绘学报, 2017, 46(10): 1207-1212.
LI D R. From geomatics to geospatial intelligent service science[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1207-1212.
- [13] 李德仁, 王密. 高分辨率光学卫星测绘技术综述[J]. 航天返回与遥感, 2020, 41(2): 1-11.
LI D R, WANG M. A review of high resolution optical satellite surveying and mapping technology[J]. Spacecraft Recovery & Remote Sensing, 2020, 41(2): 1-11.
- [14] 李德仁, 丁霖, 邵振峰. 面向实时应用的遥感服务技术[J]. 遥感学报, 2021, 25(1): 15-24.
LI D R, DING L, SHAO Z F. Application-oriented real-time remote sensing service technology[J]. National Remote Sensing Bulletin, 2021, 25(1): 15-24.
- [15] 李德仁, 张华. 我国测绘遥感技术发展的回顾与展望[J]. 中国测绘, 2019(2): 24-27.
LI D R, ZHANG H. Review and prospect of mapping and remote sensing technology development in China[J]. China Surveying and Mapping, 2019(2): 24-27.
- [16] 李德仁, 王树良, 李德毅, 等. 论空间数据挖掘和知识发现的理论与方法[J]. 武汉大学学报·信息科学版, 2002, 27(3): 221-233.
LI D R, WANG S L, LI D Y, et al. Theories and technologies of spatial data mining and knowledge discovery[J]. Editorial Board of Geomatics and Information Science of Wuhan University, 2002, 27(3): 221-233.
- [17] 李德仁, 王树良, 史文中, 等. 论空间数据挖掘和知识发现[J]. 武汉大学学报·信息科学版, 2001, 26(6): 491-499.
LI D R, WANG S L, SHI W Z, et al. On spatial data mining and knowledge discovery (SDMKD)[J]. Editorial Board of Geomatics and Information Science of Wuhan University, 2001, 26(6): 491-499.
- [18] PEI J, TSENG V S, CAO L B, et al. Advances in knowledge discovery and data mining[M]. Heidelberg: Springer, 2013.

- [19] LI D R, DI K C, LI D Y. Knowledge representation and uncertainty reasoning in GIS based on cloud models[C]// Proceedings of the 9th International Symposium on Spatial Data Handling. [S.l.:s.n.], 2000.
- [20] 王树良. 基于数据场与云模型的空间数据挖掘和知识发现[D]. 武汉: 武汉大学, 2002.
WANG S L. Data field and cloud model based spatial data mining and knowledge discovery [D]. Wuhan: Wuhan University, 2002.
- [21] 李德仁, 王树良, 李德毅. 空间数据挖掘理论与应用(第二版)[M]. 北京: 科学出版社, 2013.
LI D R, WANG S L, LI D Y. Spatial data mining theories and applications[M]. Beijing: Science Press, 2013.

作者简介



李德仁(1939-), 男, 博士, 中国科学院院士, 中国工程院院士, 武汉大学教授、博士生导师, 国际欧亚科学院院士, 国际宇航科学院院士, 国际著名测绘遥感学家, 主要研究方向为地球空间信息科学。



张过(1976-), 男, 博士, 武汉大学测绘遥感信息工程国家重点实验室教授, 主要研究方向为航天摄影测量处理与应用。



蒋永华(1987-), 男, 博士, 武汉大学遥感信息工程学院副教授, 主要研究方向为星载光学卫星高精度处理、卫星在轨智能处理。



沈欣(1981-), 男, 博士, 武汉大学测绘遥感信息工程国家重点实验室副研究员, 主要研究方向为遥感卫星任务规划。



刘伟玲 (1990-), 女, 武汉大学测绘遥感信息工程国家重点实验室博士生, 主要研究方向为基于深度学习的遥感影像云检测。

收稿日期: 2021-10-26

通信作者: 张过, guozhang@whu.edu.cn

基金项目: 国家重点研发计划资助项目 (No.2018YFB0504905, No.2018YFC0825803, No.2016YFB0500801); 国家自然科学基金资助项目 (No.42171341, No.41971412, No.41601490, No.41901400, No.41801397); 科技部重点研发计划 (No.2018YFB0504905); 高分辨率对地观测系统重大专项 (No.GFZX0404130302); 民用航天技术预先研究项目 (No.D040107); 博士后创新人才支持计划 (No. BX2021222)

Foundation Items: The National Key Research and Development Program of China (No.2018YFB0504905, No.2018YFC0825803, No.2016YFB0500801), The National Natural Science Foundation of China (No.42171341, No.41971412, No.41601490, No.41901400, No.41801397), Key Research and Development Program of the Ministry of Science and Technology (No.2018YFB0504905), The Major Project of High Resolution Earth Observation System (No.GFZX0404130302), Civil Aerospace Technology Pre Research Project (No.D040107), Postdoctoral Innovation Talent Support Program (No.BX2021222)