

自然资源大数据挖掘分析关键技术研究

李姗姗

武汉吉威空间信息技术研究院有限公司

DOI:10.12238/gmsm.v3i5.870

[摘要] 按照实施国家大数据战略的要求,本文研究运用大数据存储、计算、服务引擎、模型构建、数据挖掘和可视化等技术,构建大数据挖掘分析能力,有效挖掘自然资源数据价值,从而精准支撑各级自然资源主管部门在自然资源信息服务、业务审核、形势分析和决策支持等领域的应用,并将应用成果向其他行业共享,创新自然资源管理方式,强化政府管理和服务的科学性和前瞻性,提升自然资源管理智能决策支持能力。

[关键词] 自然资源; 大数据; 挖掘分析

中图分类号: P96 文献标识码: A

引言

“十三五”规划明确提出:“实施国家大数据战略,推进数据资源开放共享,运用大数据提升国家治理现代化水平,建立健全大数据辅助科学决策和社会治理的机制,推进政府管理和社会治理模式创新”。当前,自然资源管理领域国土、规划、海洋、林业等各类海量、异构、多源基础空间数据存在存储和计算效率不佳、应用深度和多样性不足、需求响应灵活程度不高、数据应用缺乏体系化等问题,传统的技术架构以及缺乏强有力后台支撑的系统已经很难满足自然资源时空大数据的存储、计算和应用需求,亟需构建一套适用于自然资源行业管理的数据计算和应用技术体系,从而提升自然资源数据应用水平和服务共享水平。

面向自然资源管理改革、“数字政府”改革建设和社会经济持续发展的新形势、新需求,自然资源调查评价、政务管理、社会服务、党风廉政等工作对于信息化的应用需求进一步深化,自然资源机构改革调整及自然资源业务的进一步整合,自然资源数据挖掘分析工作迎来了良好的机遇。因此,开展自然资源大数据挖掘分析关键技术应用研究,立足山水林田湖草海资源的规划、开发利用和管理,实现“用数据说话、用数据决策”的服务,用灵活快速的手段,为自然资源

决策提供支撑,为政府及社会提供自然资源信息服务,对提升自然资源管理与服务水平具有非常重要的意义。

1 总体设计

结合系统研发标准化方法,工程总体框架主要包括基础框架层、模型库层、系统层和应用层等内容。其中,基础框架为自然资源大数据挖掘分析工作提供基础算力支撑,模型库是挖掘分析算法聚合,系统是挖掘分析过程实现的主要手段,应用层是挖掘分析工作的落脚点,管理机制是工作有效推进的行政保障。

1.1 自然资源大数据基础框架

自然资源大数据基础框架主要包括自然资源大数据混合存储框架及自然资源大数据混合计算框架,综合利用RDS分布式空间数据库、TDS非关系型数据库和DFS分布式共享文件系统进行多源异构数据综合存储,引入Geo_HPC、Hadoop及Spark等技术支撑多种自然资源数据计算场景。

1.2 自然资源一体化大模型库

按照“按需定制、灵活搭建、标准统一、开放共享”的设计原则,采用面向对象的方法,设计由基础模型库和业务模型库构成的自然资源一体化大模型库。

1.3 自然资源大数据挖掘分析系统面向业务应用和宏观决策建立大数

据分析处理建模,实现公共基础数据、专业基础数据、业务管理数据、公共政务数据的快速接入、融合、分析及挖掘,并对挖掘分析成果进行可视化设计与分发,将可视化成果在门户网站中进行展现。自然资源大数据挖掘分析系统包括模型构建子系统、数据集成子系统、分析计算子系统、服务发布子系统及成果可视化子系统五部分内容。

1.4 自然资源大数据挖掘分析应用

围绕自然资源“两统一”职责,结合机构改革职能调整、核心业务升级等实际情况,按照“模型构建、数据源接入、服务发布、实时计算、成果可视化”的技术流程,从“资源资产、规划配置、管制利用、保护修复”等场景开展自然资源大数据挖掘分析应用,并支持输出体系化、多元化的分析结果,支撑业务决策和政策制定,为全省自然资源科学管理提供系统支撑。

2 关键技术

2.1 创建自然资源大数据混合存储计算生态链

针对自然资源相关管理部门数据海量多源异构、计算场景复杂多样的状况,在充分利用现有通用IT基础设施的情况下,本文提出的自然资源政务大数据混合存储和混合计算生态链构建,搭建了大数据基础框架,打破了自然资源研究

领域空间大数据计算和数据挖掘分析割裂的局面,让算力、算法和业务无缝协同,并在全省范围内支撑了土地管理、土地整治、规划管理、矿产管理等业务应用,既节省了建设成本,又能高效满足用户基于海量数据进行业务管理、辅助决策的多样化需求。

数据存储机制上,依据数据结构、数据量、数据特点和应用需求,采用差异化混合存储策略,综合利用RDS分布式空间数据库、TDS非关系型数据库和DFS分布式共享文件系统进行多源异构数据综合存储。在此基础上,通过数据库访问接口、文件访问接口以及数据服务方式实现各类数据资源共享访问,构建面向用户一致透明操作的综合数据库。数据计算机制上,引入基于HPC的并行计算框架进行模型化计算,提供了MapReduce、Spark、MPI、OpenMP、CUDA等多种计算架构,满足不同粒度的空间分析算法需求,可支撑“小数据+大数据”的业务审核实时计算场景及“大数据+大数据”的复杂分析批处理场景,在一体化在线调度机制下,实时调配计算资源并支持弹性化扩展,按需为用户提供高效可靠的计算服务。

2.2 打造自然资源大数据挖掘分析模型服务引擎

为此,工程按照“模型构建、数据源接入、服务发布、实时计算、成果可视化”的技术流程,基于SaaS模式,利用基于插件的应用服务工具快速搭建技术打造服务引擎,构建了自然资源一体化模型库,针对汇总统计、关联分析和深度挖掘等不同深度的挖掘分析层次,全流程实时在线发布各类自然资源模型服务,从而实现自然资源海量数据应用模型和算法的业务化运行,大大提升了自然资源数据对业务应用和管理决策的支撑力度。

将使用频率高、输入类型少、算法稳定的计算模型固化为插件,构建自动化、服务端运行等业务构件的注册机制,实现基础模型的集成管理;利用知识发现(KDD)与数据挖掘(DM)相结合的时空大数据挖掘分析技术,通过对基础模型

的业务组装,形成各类业务模型。面向自然资源管理领域各类标准化业务场景和探索性的深度挖掘场景,如土地利用历史回溯、国土空间规划双评价、土地资源精准化配置等,接入各类资源库、服务库中的数据源,在线发布模型服务,开展各类密集型数据实时分析计算。通过交互式的可视化展现方式,洞察数据内部时空变化规律,可支持大屏、移动终端、PC等跨终端的可视化应用。

2.3 探索自然资源数据成果融合与快速集成应用的方法

因此,面向自然资源业务管理需求和“数字政府”改革建设需求,工程通过采用数据融合和可视化等技术,探索如何将各种类型的自然资源数据成果进行快速集成应用的方法,构建全省可互动、可寻址、可定向的自然资源智慧屏,打破了“好看不好用”的数据大屏痛点,通过交互式触屏操作,实现大数据、大分析,更好发挥数据规模优势,从而推进具有全自然资源行业特色、与“数字政府”建设内容相得益彰的自然资源政务信息化建设。

工程依托于自然资源一体化数据库成果,面向自然资源领域资源资产、规划配置、管制利用和保护修复的业务体系管理需求,建设相应的专题库、主题库,在特征级和决策级开展数据融合;通过多种可视化的自然资源要素表达载体,将枯燥、乏味的数据转换为生动、直观、可视化的图表、动画等成果形式,无缝接入大屏、手机端、PC端、平板端等各类终端设备,展示大数据本身及其挖掘分析的结果,直观地传达对象的关键方面与特征;技术架构采用改进的MVC架构,将View层单独抽取出来,由前端构建单页面应用,后端仅提供Model层和Controller层,以资源访问API形式提供服务,在提升健壮性与灵活性的同时,也提高了共享效率;同时,基于自然资源大数据基础框架、微服务架构和矢量瓦片等技术促进挖掘分析成果在各厅属单位的共享应用,从基础层、数据层和服务层与省“数字政府”基础设施、政务大数

据中心、应用支撑平台和粤政图等内容准确对接,在不同行业部门之间共享多种指标服务、地图服务和可视化方案。

3 结束语

《自然资源科技创新发展规划纲要》指出,要运用现代化、信息化、智能化技术手段,发展自然资源大数据技术,挖掘和预测自然资源时空演变规律与发展趋势。本研究面向自然资源数据管理和应用需求,运用了多项当代大数据信息技术,搭建了大数据挖掘分析计算框架和挖掘分析模型库,初步形成自然资源管理决策支持的数据应用模式,为掌握自然资源家底、摸清自然资源工作情况及辅助自然资源管理决策提供依据,取得了良好的应用效果。进一步地,在现有成果基础上,验证挖掘分析数据源、挖掘过程、挖掘结果的正确有效,深化自然资源大数据挖掘分析应用,是后续的重要展望。

[参考文献]

- [1] FANG C L, GUAN X L, LU S S, et al. Input-output efficiency of urban agglomerations in China: an application of data envelopment analysis (DEA)[J]. *Urban Studies*, 2013, 50(13):2766-2790.
- [2] Fang S F, Zhu Y Q, Lida Xu, et al. An integrated system for land resources supervision based on the IOT and cloud computing [J]. *Enterprise Information Systems*, 2015:1-17.
- [3] 龙瀛, 吴康, 王江浩, 等. 大模型: 城市和区域研究的新范式[J]. *城市规划学刊*, 2014, (006):52-60.
- [4] 潘军. 多元地学空间数据融合及可视化研究[D]. 吉林大学, 2005.
- [5] 王国胤, 刘群, 于洪, 等. 大数据挖掘及应用[J]. *科技与出版*, 2017, 11(275):2.
- [6] Xing J, Sieber R E. A land use/land cover change geospatial cyberinfrastructure to integrate big data and temporal topology [J]. *International Journal of Geographical Information Science*, 2016, 30(3-4):573-593.