

大数据技术在自然资源测绘工程中的应用研究

王修龙, 赵尽晖

(山东万相资产评估事务所(普通合伙), 山东 东营 257000)

摘要 自然资源测绘工程面临海量多源数据处理的现实挑战, 卫星遥感、激光点云、矢量要素等异构数据规模已达 PB 量级, 传统技术手段已难以适应当前业务需求。本文以大数据技术与测绘工程的深度融合作为切入点, 对数据采集、处理分析等关键环节中的技术应用进行系统研究, 提出多源数据统一汇聚、实时流式接入、分布式并行计算、智能化识别分类等具体实现路径, 旨在提升数据处理效率与地物识别自动化水平, 为测绘行业技术革新提供理论参考。

关键词 大数据技术; 自然资源; 测绘工程; 分布式计算; 智能识别

中图分类号: P2; TP3

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2026.18.011

0 引言

自然资源测绘是获取地表空间信息的基础性工作, 其数据采集范围涵盖陆地、海洋、大气等多个领域, 数据量呈现爆发式增长态势。卫星遥感、激光雷达、无人机航测等新型技术手段的广泛应用, 使测绘数据规模从 TB 级别向 PB 级别发展, 数据类型涵盖影像、点云、矢量等多种形态, 传统处理模式已无法满足海量异构数据的存储、计算需求。大数据技术凭借其分布式架构、并行计算、智能分析等优势, 为测绘工程转型升级提供了技术支撑。基于此, 探索大数据技术与自然资源测绘工程的融合应用路径具有重要的现实意义。

1 大数据技术与自然资源测绘工程的内在契合性分析

自然资源测绘工程在数据获取阶段就呈现出鲜明的海量特征, 卫星遥感每日产生的影像数据以 TB 为单位计量, 地面激光扫描形成的点云信息密度可达每平方米数千至数万个坐标点, 各类传感器采集的多源异构信息持续汇聚成庞大的数据流。传统技术体系面对如此规模的信息在处理效率上存在明显局限, 数据存储空间不足只是表象, 更深层的困境在于无法从海量数据中快速提取有效信息, 大量采集成果因缺乏高效分析手段而沉淀为“数据孤岛”。大数据技术的核心优势恰恰体现在分布式存储架构能够承载 PB 级数据规模, 并行计算框架将原本需要数天完成的处理任务压缩至小时级别, 这种技术能力与测绘工程的实际需求形成了天然的对接关系。

自然资源管理本质上是一个动态监测过程, 森林覆盖率变化、土地利用类型转换、矿产资源储量波动都需要持续跟踪, 单次测绘只能反映某个时间切片的状态, 难以揭示资源演变的趋势规律^[1]。大数据技术具备的时序分析能力让多期测绘成果能够建立起时空关联, 历史数据不再是静态的存档记录, 而是转化为预测未来变化的知识基础, 测绘工程由此获得了从“测什么”向“为什么变化”深入的可能性, 资源管理决策也从经验判断转向数据驱动的科学模式。

2 大数据技术在自然资源测绘数据采集环节的应用实践

2.1 多源异构测绘数据的统一汇聚体系构建

测绘数据来源涵盖卫星遥感、航空摄影、地面激光扫描等多种渠道, 各类设备输出的数据格式存在明显差异, 影像数据以栅格形式存储, 矢量边界采用坐标序列表达, 点云信息则包含 xyz 三维坐标属性。要实现多源数据的统一管理, 首先需要建立元数据标准化描述规范, 对每批采集数据标注其空间参考系统、时间戳、传感器类型等核心属性, 将分散的数据特征信息集中记录在统一的索引库中, 后续检索时只需查询索引即可快速定位目标数据集。

数据接入层需要部署格式转换中间件, 针对不同来源设置专门的解析模块, 卫星影像进入系统时自动读取其元数据文件提取轨道参数, 无人机采集的照片则从文件头信息中获取拍摄位置, 转换模块将原始数据重组为系统内部约定的标准结构后再写入存储集群^[2]。存储架构采用分布式文件系统承载海量数据, 同时建

作者简介: 王修龙 (1988-), 男, 专科, 研究方向: 自然资源工程—测绘工程类。

立关系型数据库维护数据间的关联关系,影像文件本身存放在文件系统保证读写效率,其对应的地理范围、质量评价等属性则录入数据库便于条件筛选。

2.2 星载传感器与无人机平台的实时数据流接入

卫星过境时间窗口短暂,遥感数据下传后需要立即纳入处理流程才能满足应急监测需求,传统方式要等数据完整接收后再启动处理程序,时间延迟往往达到数小时。要实现数据的实时接入,地面接收站需要部署流式数据处理引擎,卫星下传的数据流边接收边切分成小的数据块,每个数据块到达后立即触发初步质检程序,检测其是否存在传输错误或云层遮挡,符合质量标准的数据块直接进入后续处理队列。

无人机作业时携带的存储卡容量有限,长时间飞行会面临数据存满的风险,机载设备配置无线传输模块后能够在飞行过程中将采集数据实时回传至地面工作站。地面接收端建立缓冲队列承接高速涌入的数据流,缓冲区设置容量阈值,当积累数据量达到设定值时批量写入持久化存储,避免频繁的磁盘写操作影响系统性能。数据流经过缓冲队列后进入分发环节,系统根据数据的地理位置属性将其路由到对应的区域数据库,同一测区的影像集中存储便于后续拼接处理^[3]。

2.3 地面调查数据的分布式协同采集机制

分布式外业采集产生的数据需实时汇聚以保证完整性与时效性。要建立协同采集机制,每个调查终端配置数据同步客户端,采集人员在现场完成一个测点的记录后,客户端自动将数据加密打包上传至中心服务器,上传过程采用断点续传技术,即使网络中断也能在信号恢复后继续传输未完成的数据包。

中心服务器接收各终端上传的数据后需要进行一致性校验,检查同一地块是否存在多个调查组重复采集的情况,发现重复数据时比对其采集时间戳,保留最新版本的记录丢弃过时数据。数据上传后触发自动质检程序,对照片清晰度、坐标精度等指标进行初步评估,不合格的数据推送提醒消息给采集人员要求补测。

2.4 海量遥感影像的自动化预处理流程优化

原始遥感影像存在辐射畸变、几何扭曲等问题,人工逐景处理耗时巨大且难以保证质量一致性。想要优化预处理流程,首先需要建立影像质量自动评估模

块,新接收的影像进入系统后先经过云量检测算法,计算影像中云层覆盖的像元比例,云量超过一定阈值(30%)的影像直接归入备选库不进入主处理流程,晴空影像则继续后续步骤。

辐射定标环节需要调用传感器特定的定标参数文件,系统根据影像元数据中记载的卫星型号自动匹配对应参数,将原始DN值转换为地表反射率,转换过程中需进行大气校正,常用方法包括利用影像中的暗像元区域估算大气路径辐射值后扣除。几何校正依赖高精度的控制点数据库,数据库中存储了全国范围的地面控制点坐标,系统根据影像覆盖范围自动提取相应区域的控制点,采用多项式拟合方法建立像元坐标与地理坐标的映射关系,重采样时选择双线性插值算法平衡处理速度与精度要求。预处理完成的影像按照标准分幅规则进行裁切,生成统一尺寸的图幅产品便于入库管理。

3 大数据技术在自然资源测绘数据处理与分析中的深化运用

3.1 分布式计算框架在测绘海量数据并行处理中的实现

单机处理能力面对TB级测绘数据已经捉襟见肘,一景高分辨率卫星影像的正射校正往往需要消耗数小时计算时间,批量处理时任务积压现象严重。想要突破计算瓶颈,需要部署分布式计算集群,主流分布式框架在计算模式与性能指标上各有侧重,表1对比了其技术特性与适用场景。将多台服务器组织成统一的计算资源池,任务调度器负责接收处理请求后将其拆解为若干子任务,每个子任务分配给集群中的空闲节点执行,节点完成计算后将结果返回调度器进行汇总。任务拆解时需要考虑数据的空间特性,影像处理任务按照图幅边界切分,每个节点处理完整的图幅避免出现接边问题,点云数据则按照空间网格划分,相邻网格分配给不同节点同时处理。

集群节点之间需要共享数据访问通道,避免重复传输相同的基础数据消耗网络带宽。系统搭建分布式文件系统作为数据交换中枢,原始测绘数据预先存储在文件系统中,各计算节点直接从文件系统读取所需数据块,处理完成后将结果写回指定目录,文件系统自

表1 分布式计算框架技术特性对比

框架名称	计算模式	处理延迟	内存依赖	典型应用场景
Hadoop MapReduce	批处理	Min ~ h 级	低	海量影像离线处理
Apache Spark	内存计算	S ~ min 级	高	迭代算法、中等规模数据
Apache Flink	流处理	Ms ~ s 级	中	实时数据接入与预处理
Dask	Python 并行	S ~ min 级	中	科学计算、栅格数据分析

动将数据分散存储在多个磁盘保证读写速度。计算过程中某个节点出现故障时,调度器需要感知节点状态变化,将未完成的子任务重新分配给其他健康节点^[4]。

3.2 机器学习算法在地物要素自动识别与分类中的应用

人工解译遥感影像需要操作员逐个像元判读地物类型,效率低下且主观性强,同一地物在不同解译员手中可能得到不同的分类结果。想要实现自动化识别,首先需要构建训练样本数据集,选取典型区域的遥感影像,邀请经验丰富的解译人员标注其中的地物边界,将林地、耕地、建设用地等类别的标准样本提取出来,每类地物至少准备数千个样本图斑确保样本的代表性覆盖不同季节、不同地形条件下的光谱特征差异。

样本数据准备完成后进入模型训练阶段,深度学习网络从样本影像中自动提取特征,卷积层识别像元的纹理、形状等视觉特征,全连接层学习不同特征组合与地物类别的对应关系。训练过程需要反复调整网络参数,设置合适的学习率控制参数更新幅度,学习率过大会导致训练不稳定,过小则收敛速度缓慢,实践中采用动态学习率策略,初期使用较大值快速逼近最优解,后期逐步降低学习率精细调整参数。模型训练完成后在验证数据集上测试其分类精度,计算混淆矩阵评估各类地物的识别准确率,发现某类地物误判率偏高时需要补充该类样本重新训练,直至整体精度满足业务要求后才能投入实际应用^[5]。

3.3 空间大数据挖掘在自然资源动态变化监测中的实践

资源变化往往呈现渐进式演变特征,森林退化、土地沙化等过程跨越数年甚至数十年,单次测绘只能捕捉某个时刻的状态快照,难以揭示变化的趋势规律。开展动态监测,需要建立多期测绘成果的时序数据库,将不同年份的土地利用现状图、植被覆盖度专题图等成果按照时间顺序组织存储,数据库中每个地块记录其历次调查的属性信息,形成完整的演变轨迹。数据入库前需要进行空间配准,确保不同时期的图层能够精确叠加,消除因坐标系差异导致的空间错位。

变化检测算法对比相邻时期的测绘成果,逐图斑分析其地类属性是否发生转变,林地转为建设用地、耕地退化为未利用地等变化类型自动标记出来。检测过程中需要设置变化阈值过滤微小差异,地块面积变动小于最小上图面积时判定为测量误差而非真实变化,植被指数波动幅度在正常季节变化范围内时不认为是退化现象。变化图斑提取后进入空间聚类分析环节,算法识别变化集中发生的区域,将空间上相邻的变化图斑归并为同一变化单元,计算每个单元的变化面积、

变化速率等统计指标,挖掘区域性变化的空间分布规律,为资源管理决策提供数据支撑。

3.4 大数据可视化技术在测绘成果智能表达中的创新

传统测绘成果以纸质地图或静态图件形式交付,使用者只能被动接受固定的表达方式,无法根据自身需求调整显示内容。提升成果的易用性,需要开发交互式可视化平台,将测绘数据加载到网络地图引擎,用户在浏览器中打开平台后能够自由缩放、平移视图。平台后台部署瓦片服务器,将大幅面地图切分为小尺寸瓦片,用户浏览时只加载视野范围内的瓦片,既保证了显示流畅度,又节省了网络传输流量。海量测绘数据同时展示会造成视觉混乱,平台需要提供多层次显示机制,将不同类型的测绘成果分层管理,地形晕渲、影像底图、专题要素等内容各自独立成层,用户勾选图层控制面板中的选项决定哪些图层显示,不需要的图层关闭后能够突出重点内容^[6]。大比例尺浏览时展示详细的地物边界,缩小到小比例尺后自动简化图形细节,采用符号化表达减少屏幕上的视觉元素数量。

4 结束语

大数据技术在自然资源测绘工程中的深度融入,重塑了传统测绘的数据采集、处理、分析全流程作业模式。分布式存储突破了海量数据管理的空间瓶颈,并行计算框架显著提升了数据处理效率,机器学习算法实现了地物要素的自动化识别,时序数据挖掘揭示了资源演变的内在规律。测绘成果从静态图件向动态可视化平台转变,极大地提升了数据的应用价值。随着物联网传感器网络的普及,测绘数据将呈现实时化、精细化特征,大数据技术需要在流式计算、边缘处理等方向持续深化,人工智能与测绘业务的结合也将更加紧密,智能化测绘体系的构建仍需持续探索实践。

参考文献:

- [1] 高翔.大数据测绘地理信息系统资源整合与应用[J].智慧中国,2026(01):66-67.
- [2] 张章强.自然资源管理中测绘技术的应用[J].大众标准化,2025(18):154-156.
- [3] 孔艳.大数据技术在测绘地理信息中的应用[J].华东科技,2025(08):85-87.
- [4] 刘国生,吕文斌,杜可炎.大数据技术在测绘地理信息中的应用[J].中国宽带,2024,20(12):141-143.
- [5] 郭党五.探析大数据技术在测绘地理信息中的应用[J].通讯世界,2024,31(10):163-165.
- [6] 吴鑫磊.探析大数据在测绘地理信息方面的应用[J].中国战略新兴产业,2024(05):56-58.