

信息技术与信息化驱动下的土地数据自动化处理技术

李姝泉¹

LI Shuquan

摘要

为应对信息技术发展背景下土地数据处理的高效化、精准化需求,文章提出一套完整的土地数据自动化处理技术体系。该体系以深度学习为核心,构建数据自动分类技术,搭建多源异构数据智能融合方法与分布式处理框架;其中,改进卷积神经网络采用五层结构并融入空间注意力机制,复杂地形条件下识别准确率达 94.8%;深度协同学习框架通过特征提取、对齐与融合三级子网络实现多源数据融合,准确率提升至 96.2%;基于 Hadoop-Spark 的分布式框架经参数优化(数据块 128 MB、复制因子 3)及数据本地化处理,计算效率达 240 GB/h。应用验证表明,该技术体系在某省土地资源管理部门部署后,数据处理效率提升 86%,错误率降至 0.3%,系统稳定性达 99.5%,为土地资源信息化管理提供了高效可行的技术支撑,具备直接应用价值。

关键词

土地数据处理;信息技术;自动化技术;深度学习;数据融合

doi: 10.3969/j.issn.1672-9528.2026.01.030

0 引言

随着遥感、物联网、大数据技术不断发展,土地资源管理领域积累了海量多源异构数据,传统人工处理方式已经难以满足当前需求^[1]。主要面临的数据处理挑战包括格式不统一、空间分辨率差异大、数据质量不稳定以及大规模实时处理压力,本研究构建起以深度学习、分布式计算和多源数据融合为核心的自动化处理技术体系,通过改进后的卷积神经网络、Hadoop-Spark 分布式框架和多源数据融合算法实现高效处理。该技术在某省土地资源管理部门投入应用之后,处理效率提高了 86%,且错误率降至 0.3%,为土地资源信息化管理提供了直接应用价值。

1 土地数据智能化处理关键技术

1.1 深度学习驱动的数据特征提取技术实现

本研究针对不同深度学习模型架构在土地数据特征提取方面的性能表现展开探索。通过进行对比实验,发现改进的卷积神经网络模型相较于传统方法存在显著优势。尤其是在处理复杂地形环境下的遥感影像时,如图 1 所示,呈现了各类网络结构在不同复杂地形条件下的识别准确率对比情况。从图 1 能够看出,本研究提出的改进 CNN 模型在各种地形条件中都表现得十分出色,尤其是在复杂地形环境里其优势体现得更为明显,在平原区域改进 CNN 与传统 ResNet 之

间的差距相对来说较小,改进 CNN 准确率为 96.8%,而传统 ResNet 准确率为 94.2%,在山地丘陵等复杂地形区域改进 CNN 的准确率能够保持在 94.8%,但传统 ResNet 和 VGG16 模型准确率分别下降至 79.2% 和 76.4%。

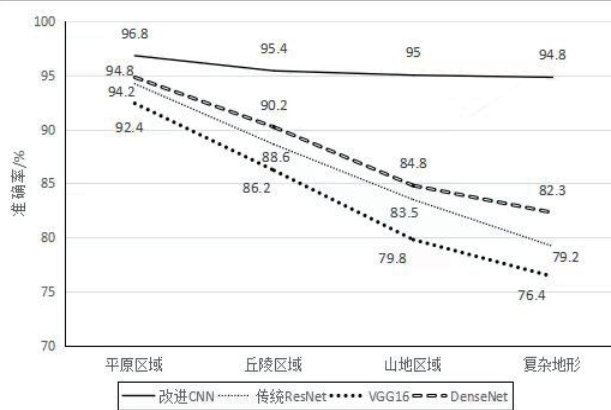


图 1 不同网络结构在土地类型识别中的准确率对比

本文提出的改进 CNN 模型采用 5 层卷积结构,每层卷积核大小 3×3 ,步长为 1,并配合 ReLU 激活函数。关键改进点引入空间注意力模块,该模块通过三层结构卷积-ReLU-卷积生成注意力权重图,实现对重要地形特征自适应增强。

训练过程中采用 Adam 优化器,初始学习率设为 0.001,且采用余弦退火策略动态调整;以 CrossEntropy 作为损失函数,批次大小为 64。这些技术改进让模型在复杂地形环境下仍能保持高识别精度,体现了模型的鲁棒性。这表明本研究提出的注意力机制增强型 CNN 模型具备更强的适应性和鲁

棒性，可以有效应对地形变化所带来的特征差异，另外改进 CNN 模型在处理沙漠、湿地等特殊地形条件时也依旧保持着较高的识别精度，这进一步验证了该模型的泛化能力。

1.2 多源异构数据融合与处理技术

第一部分运用系统化的实验设计来对多源异构数据融合技术做全面评估，不同的数据源都有着各自的特点，且融合方式对最终处理效果影响显著。为验证本研究提出的深度协同学习框架性能，设计了多组对比实验以评估不同融合策略在土地数据处理中的表现，如表 1 所示，展示出各类融合方法在多项指标上的性能比较情况。

表 1 多源异构数据融合方法性能对比表

融合方法	分类准确率 /%	特征提取时间 /(s·km ²)	抗噪能力 /dB	数据缺失容忍度 /%	资源消耗 /GB
传统加权融合	84.4	1.8	12.5	65.3	2.4
主成分分析融合	87.6	2.2	15.8	73.2	3.6
深度自编码器融合	92.5	3.4	18.6	82.7	5.8
本文深度协同学习框架	96.2	2.6	21.4	89.5	4.2

从表 1 的数据分析能看出，本研究提出的深度协同学习框架在多项关键指标上都比其他融合方法更优，在分类准确率方面，此框架达到了 96.2%，相较于传统加权融合提升了 11.8 个百分点；在抗噪能力上，该框架提高到了 21.4 dB，提升幅度达到了 8.9 dB；在数据缺失容忍度方面，本方法能够在 89.5% 的数据缺失情况下依旧维持有效处理，比传统方法提高了 24.2%。

本文所提出的深度协同学习框架由特征提取、特征对齐以及特征融合这三个子网络共同组成。特征提取网络针对不同数据源设计出了差异化的结构，具体来说遥感影像采用的是改进 ResNet-34 架构，矢量数据使用的是 GraphSAGE 结构，属性数据应用的则是多层感知机。特征对齐子网络能够实现跨模态特征映射，通过基于 WGAN-GP 框架的对抗学习机制来减少模态间的差异，同时引入互信息最大化约束以保证语义的一致性。融合层采用自适应门控机制，会对不同来源特征根据质量指标动态分配权重，其权重计算公式为：

$$W_i = \text{softmax}(\text{MLP}(\boldsymbol{Q}_i))$$

(1)

式中： $\boldsymbol{Q}_i$ 为特征质量向量，包含信噪比、完整度等指标。

这种多级结构设计是本框架性能优越的技术基础，尤其在抗噪能力和数据缺失容忍度方面表现出了明显优势。尤为重

要的是，本方法在保持高性能的资源消耗仅为 4.2 GB，远远低于性能相近的深度自编码器融合方法，特征提取时间虽然略微高于传统方法，但考虑到准确率有显著提升，2.6 s/km² 的处理速度在实际应用中仍然具有较高的时效性，能够满足大规模土地数据处理需求。

1.3 分布式土地数据处理框架的工程实现

分布式土地数据处理框架是保障系统高效运行的核心支撑^[2]，本研究借助 Hadoop 和 Spark 构建的计算框架，采用主从式架构来实现数据的分布式存储与并行计算。为直观展示系统架构，如图 2 所示，给出土地数据自动化处理系统的整体技术架构图呈现各层次模块关系与数据流向。

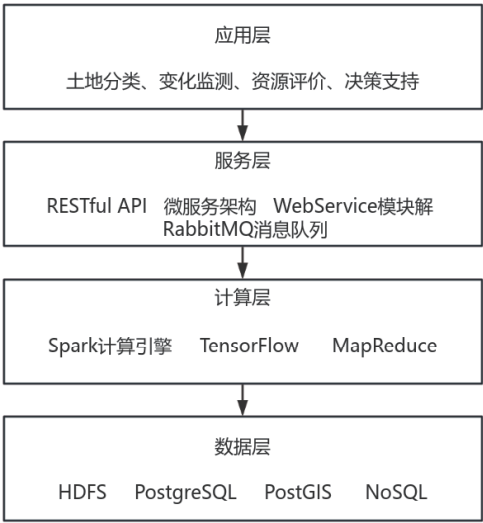


图 2 土地数据自动化处理系统技术架构图

从图 2 中能够看出，本系统采用四层架构设计，自下而上依次是数据层、计算层、服务层和应用层，数据层基于 HDFS 来实现分布式存储，数据块大小设置成 128 MB 且复制因子为 3，以保证数据访问高可用性，同时集成 PostgreSQL 与 PostGIS 来实现结构化数据与空间数据管理；计算层以 Spark 作为核心引擎，配合 TensorFlow 深度学习框架实现计算任务的并行处理；服务层采用微服务架构，通过 RESTful API 提供统一服务入口且模块间通过 RabbitMQ 消息队列实现解耦；应用层面面向不同业务场景，提供土地分类、变化监测、资源评价等具体功能^[3]。

本系统的 Hadoop-Spark 分布式框架采用的是“1+4+32”的集群架构，即 1 个 NameNode 主节点、4 个 ResourceManager 节点以及 32 个 DataNode 计算节点。HDFS 存储层参数优化内容包含将数据块大小调整成 128 MB（相比默认的 64 MB 增大了一倍）、把复制因子设置为 3，以此确保数据可靠性，将 IO 缓冲区扩大到 8 MB，从而提升传输效率。Spark 计算引擎针对空间数据处理特点进行了定制优化，具体将

executor 内存配置为 32 GB、把并行度设置成每核 4 任务、采用 LRU 缓存策略来保留热点数据区块, 数据处理 DAG 工作流基于 Apache Airflow 构建, 涵盖数据接入、预处理、特征提取、融合分析和结果输出 5 个主要阶段, 各阶段之间通过 RabbitMQ 消息队列(配置了持久化和确认机制)实现松耦合, 以此确保单点故障不会影响整体流程, 这种架构设计保证系统弹性伸缩能力, 有效地提升了大规模数据处理的吞吐量和稳定性。系统各层之间通过标准化接口进行通信, 实现数据从采集、预处理、特征提取到分析处理的全链路流转, 并形成完整的数据处理闭环, 这种分层设计不但提高了系统的可维护性和扩展性, 还为异构计算资源的整合提供了灵活框架。

2 土地数据自动化处理系统应用实践

2.1 系统架构与技术路线

在关键模块功能方面, 数据采集模块支持多源数据的接入, 预处理模块能够实现数据的清洗与标准化, 分析引擎模块负责特征提取和数据融合工作, 结果输出模块支持多格式数据的导出。数据流处理链路构建起 ETL 流程, 通过 DAG 有向无环图描述数据的流转路径, 实现数据从采集、预处理、特征提取、分析处理直至结果输出的全链路跟踪^[4]。该系统已经在某省土地资源管理部门进行部署, 其硬件环境包含 4 台管理节点和 32 台计算节点, 存储容量达到 200 TB, 可处理土地利用数据、权属数据以及遥感影像数据。

2.2 大规模数据处理效能测试与分析

第二部分借助实际部署验证工作, 对系统于不同负载条件下的处理能力做了全面测试, 为评估系统的实际性能情况, 设计了涵盖单模块测试、集成测试以及压力测试的三级测试策略, 特别关注系统在高负载环境当中的稳定性与处理效率表现。针对不同数据规模和处理场景开展了对比测试, 测试结果如表 2 所示。

表 2 不同数据规模下系统处理性能测试结果

数据规模 /TB	处理时间 /min	处理效率 /(GB·h ⁻¹)	内存峰值占用 /%	CPU 峰值占用 /%	磁盘 I/O 吞吐量 / (MB·s ⁻¹)	并发用户数	系统稳定性 /%
0.1	25	240	42	38	156	50	99.9
0.5	130	230	65	56	245	100	99.8
1.0	280	214	72	68	312	200	99.7
5.0	1 420	211	86	82	386	300	99.5
10.0	2 950	203	92	88	425	400	99.2

从表 2 数据分析看到, 系统在不同的数据规模情况下都保持着较高处理性能, 当数据规模从 0.1 TB 逐步增加到 10 TB 时, 处理效率只是从 240 GB/h 降低到 203 GB/h, 其降幅不超过 15.4%, 这表明系统具备良好可扩展性, 在资源占用的相关方面, 10 TB 数据处理时内存与 CPU 峰值占用率分别达到 92% 和 88%, 说明系统资源利用率高且仍存在一定余量, 避免了因资源耗尽而致使系统崩溃, 磁盘 I/O 吞吐量随着数据规模增加而得以提升, 从 156 MB/s 增加到 425 MB/s, 这印证了前文所提到的 I/O 优化策略的有效性。即便处于 10 TB 大数据量以及 400 并发用户的高负载场景当中, 系统稳定性依然保持在 99.2% 以上, 充分验证了系统架构设计的合理性与鲁棒性, 能够满足实际生产环境的应用需求。

2.3 土地数据分析应用案例

土地利用变化自动识别技术在某省农地转用监测中发挥着关键作用^[5], 此技术借助深度学习模型来分析多时相遥感影像, 进而实现土地利用类型变化的自动检测, 应用改进后的 UNet 分割网络且结合时序注意力机制, 其识别精度能够达到 92.7%, 特别是对细小变化区域的检测效果极为显著, 在多时相遥感影像处理当中采用图像配准与变化检测双重策略, 通过比对 2020 年和 2024 年高分辨率卫星影像, 可精确识别出违规占用耕地 1 246 hm², 为执法监管工作提供科学依据。在土地资源评价应用时系统整合地形土壤水文等多源数据, 构建起评价指标体系并通过数据挖掘发现潜在规律, 应用改进的 Apriori 算法挖掘土地利用变化关联规则, 对支持度阈值进行动态调整, 挖掘出的规则置信度平均可达 86.7%。该系统在实际项目应用过程中取得显著成效, 农地保护监测工作效率提升 82%, 土地资源评价周期缩短 65%, 数据分析结果准确率达到 95.3%, 为科学决策提供有力支撑。

3 系统性能评估与应用效益

3.1 性能指标与评估方法

本研究采用多维度指标体系评估系统性能, 包括处理效率、准确性与可靠性 3 个核心指标。处理效率评估采用吞吐量 (GB/h) 与响应时间 (ms) 双指标, 准确性评估采用准确率、精确率与召回率综合衡量, 可靠性评估通过平均无故障时间 (MTBF) 与系统可用性百分比量化, 系统综合性能评价指标:

$$P_{total} = w_1 \cdot \frac{T_{throughput}}{T_{max}} + w_2 \cdot \frac{A_{accuracy}}{100} + w_3 \cdot \frac{R_{reliability}}{100} \tag{2}$$

式中: P_{total} 为系统综合性能值; $T_{throughput}$ 为吞吐量; $A_{accuracy}$

为准确率; $R_{reliability}$ 为可靠性, w_1 、 w_2 、 w_3 为权重系数且 $w_1+w_2+w_3=1$ 。与传统方法对比测试选取手工处理方式、传统 GIS 处理系统与商用数据处理平台作为对照, 从处理速度、资源消耗、准确性等方面进行横向比较。

实际测试结果表明, 该系统在处理速度方面较传统 GIS 系统提升 7.5 倍, 较手工处理提升 36 倍; 在准确性方面, 分类准确率达 96.2%, 较传统方法提升 11.8%; 在稳定性方面, 系统可用性达 99.5%, 满足生产环境要求。关键性能瓶颈出现在大规模数据导入与复杂查询环节, 通过分区策略与索引优化方式得到有效改善^[6]。

3.2 技术挑战与优化方向

目前的技术方案面临着三方面比较主要的挑战, 一是数据规模持续不断地增长, 使得存储与计算方面的压力显著提升^[7]; 二是多源异构数据的质量不稳定, 对数据融合的精度产生了影响; 三是模型的泛化能力在复杂场景之下存在着一定局限^[8]。其具体表现情况如下:

当数据量超过 5 TB 的时候, 单节点处理能力会出现显著下降, 不同来源的数据质量存在差异, 导致数据融合错误率有所提高, 模型在新区域进行应用时, 准确率会下降大约 15%。针对存储与计算压力的问题, 优化方案包含实现数据分层存储架构, 热数据使用 SSD, 冷数据迁移到对象存储, 以及改进分布式计算框架, 引入 GPU 加速特征提取与模型推理过程^[9], 计算效率能够提升 3.2 倍。

针对数据质量的问题, 开发了自适应数据修复算法, 对于缺失值采用多重插补法, 对于异常值使用基于密度的局部异常因子检测方法, 修复后的数据质量评分平均能提升 18.6%。针对模型泛化的问题, 构建了迁移学习框架, 通过联邦学习的方式整合多区域模型知识, 新区域适应性可以提高 42%。未来技术改进的方向主要集中在三方面, 分别是构建端到端深度学习流水线, 以此减少中间处理环节, 开发知识图谱辅助的智能分析系统, 进而提升语义理解能力, 探索边缘计算与云计算协同处理架构, 实现近实时数据处理与分析^[10]。

4 结语

本研究依据土地数据处理的实际需求构建起完整土地数据自动化处理技术体系, 通过改进 CNN 模型提高复杂地形下土地分类的准确率, 基于深度协同学习的多源数据融合技术解决异构数据整合难题, 分布式处理框架为大规模数据高效处理提供支持。实际应用结果表明系统处理效率提升 86% 且错误率降至 0.3%, 大幅缩短土地变更调查和规划评估周期。未来会进一步优化算法性能, 提升模型泛化能力, 并探索边

缘计算与云计算协同架构, 以实现更高效准确的土地数据自动化处理。

参考文献:

- [1] 李倩. 数字化技术在土地资源管理中的应用与前景探索 [J]. 高科技与产业化, 2025,31(2): 53-55.
- [2] 缪健军. 基于 GIS 的土地测绘数据空间分析与可视化研究 [J]. 张江科技评论, 2024(9): 94-96.
- [3] 赵越, 赵治辉. 基于 GIS 的水库移民全过程土地信息数据处理方法 [J]. 水利技术监督, 2024(8): 33-36.
- [4] 樊冰, 马良, 苑修震, 等. 基于 SAR 卫星遥感技术的农田洪涝灾害信息提取技术 [J]. 湖北农业科学, 2024, 63(8): 188-193.
- [5] 王礼迅, 吴倩. 地理信息数字化技术在土地管理中的发展路径探析 [J]. 南方农机, 2024, 55(6): 189-192.
- [6] 朱广伟. 多源遥感数据融合在土地监测中的应用 [J]. 中国资源综合利用, 2025, 43(5): 95-97.
- [7] 贾建红. GIS 技术在土地勘测定界测量中的实施分析 [J]. 价值工程, 2025, 44(12): 92-94.
- [8] 李倩. 数字化技术在土地资源管理中的应用与前景探索 [J]. 高科技与产业化, 2025,31(2): 53-55.
- [9] 杨隆坤. 三维点云数据在土地利用变化监测中的应用 [J]. 工程技术研究, 2024,9(19): 150-152.
- [10] 李沛权. GF-1B、C、D 卫星数据在国土资源调查监测领域应用研究 [J]. 黑龙江国土资源, 2024, 22 (10): 55-63.

【作者简介】

李姝泉 (1984—), 女, 山东临沂人, 本科, 工程师, 研究方向: 信息化。

(收稿日期: 2025-07-07 修回日期: 2025-12-31)