

大数据可视化在环境质量综合评估中的应用与优化

石彬兵

深圳市朗玥环保科技有限公司 广东深圳 518049

摘要: 为解决环境质量综合评估中多源数据处理复杂、评估结果传达低效等问题,本文聚焦大数据可视化技术在该领域的应用与优化。首先构建“数据处理-模型构建-可视化呈现-评估应用”的全流程框架,分析当前可视化应用中存在的多源数据融合不足、动态交互性欠缺等痛点;其次提出融合多维度指标的分层可视化模型,设计基于GPU加速的实时渲染算法与交互式分析模块;最后以某省环境监测数据为样本开展实验验证。结果表明,优化后的可视化系统在数据处理效率、异常识别准确率及决策支持能力上均显著提升,为环境质量综合评估提供高效技术支持。

关键词: 大数据可视化; 环境质量综合评估; 多源数据融合

一、引言

(一) 研究背景

随着工业化与城市化进程加速,大气污染、水体富营养化等环境问题日益凸显,环境质量综合评估成为生态环境保护决策的核心环节。当前环境监测体系已形成“地面站点+卫星遥感+移动监测”的多源数据采集网络,单区域日监测数据量可达TB级,涵盖PM2.5、COD、重金属等百余种指标^[1]。传统评估方法依赖统计报表与静态图表,难以直观呈现数据关联规律与动态变化趋势,导致决策层难以快速精准把握环境质量状况。

大数据可视化技术通过图形化编码将海量复杂数据转化为直观信息,具备多维度展示、动态交互等优势,为环境质量评估提供全新解决方案。例如美国EPA开发的VERDI系统可实现空气质量模型数据的时空可视化与交互分析,国内部分地区也构建了环境监测可视化平台,但在多源数据融合呈现、个性化分析等方面仍存在不足。因此,开展大数据可视化在环境质量综合评估中的应用与优化研究具有重要现实意义^[2]。

(二) 国内外研究现状

国外在环境数据可视化领域起步较早,技术应用较为成熟。除VERDI系统外,欧盟的Copernicus大气监测

服务平台采用热力图、时序曲线等可视化形式,实现全球空气质量实时监控与趋势预测,其基于WebGL的渲染技术支持百万级数据实时展示^[3]。在学术研究方面,Smith等(2023)提出基于知识图谱的环境污染物关联可视化模型,有效提升了污染源追溯效率。

国内研究聚焦于应用场景拓展与技术适配。王雪等(2020)构建了区域环境质量可视化评估体系,整合大气、水、土壤监测数据实现综合展示,但未解决多源数据时空校准问题。李刚等(2019)提出环境数据可视化清洗算法,提升了数据质量,但在动态交互分析模块设计上较为简单。现有研究仍存在三大痛点:一是多源数据融合可视化时易出现维度冗余;二是动态监测数据渲染延迟超过100ms,难以满足实时评估需求;三是可视化结果与评估模型结合不紧密,决策支撑能力不足。

(三) 研究内容与技术路线

本文研究内容包括:(1)构建环境质量综合评估的多源数据处理体系,实现监测数据的清洗、融合与标准化;(2)设计分层可视化模型,实现宏观区域概况、中观指标分布、微观异常溯源的多维度展示;(3)优化可视化渲染与交互算法,提升系统性能;(4)以实际监测数据为样本开展实验验证。

技术路线采用“问题导向-模型构建-算法优化-实验验证”的思路:首先通过文献分析与实地调研明确应用痛点;其次基于Hadoop平台构建数据处理模块,结合ECharts与deck.gl设计可视化模型;然后采用GPU加速与数据降维技术优化系统性能;最后通过对比实验与专家评估验证优化效果。

作者简介: 石彬兵(1982年12月),女,壮族,广西桂林市,现任职于深圳市朗玥环保科技有限公司数据分析工程师。深耕大数据与数据分析领域,具备扎实的专业功底,熟练掌握Hadoop、Spark分布式计算框架,精通Python、SQL数据处理技术及ECharts等可视化工具。

二、相关理论与技术基础

(一) 环境质量综合评估指标体系

参考《环境空气质量标准》(GB3095-2018)与《地表水环境质量标准》(GB3838-2002),构建“大气-水-土壤-生态”四维评估指标体系。大气环境选取PM2.5、PM10、SO₂等6项核心指标;水环境选取COD、氨氮、溶解氧等5项指标;土壤环境选取重金属含量、pH值等4项指标;生态环境选取植被覆盖率、生物多样性等3项指标。采用层次分析法确定指标权重,其中大气与水环境权重分别为0.35、0.30,土壤与生态环境权重分别为0.20、0.15。

(二) 大数据可视化核心技术

数据预处理技术:采用ETL工具实现多源数据整合,通过Z-Score标准化消除量纲差异,运用孤立森林算法识别异常数据,数据清洗准确率可达98.2%。**可视化渲染技术:**结合矢量图形与栅格图像优势,采用deck.gl的分层渲染架构,通过HeatmapLayer、ContourLayer等实现多维度数据展示,WebGL2硬件加速支持百万级数据实时渲染,帧率提升至30fps以上。**交互分析技术:**基于事件驱动模型设计钻取、筛选、联动等交互功能,支持用户从区域概况下钻至监测站点明细数据,实现“宏观-微观”联动分析。

(三) 可视化评估指标体系

从数据呈现、交互性能、决策支撑三个维度构建评估指标:数据呈现维度包括信息覆盖率($\geq 95\%$)、指标关联度(≥ 0.8);交互性能维度包括渲染延迟($\leq 50\text{ms}$)、响应准确率($\geq 99\%$);决策支撑维度包括异常识别效率($\leq 30\text{s/个}$)、方案匹配度(≥ 0.85)。通过熵权-TOPSIS法实现可视化效果综合评价。

三、大数据可视化在环境质量综合评估中的应用现状与问题

(一) 应用场景分析

实时监测可视化:某省环境监测中心构建的可视化平台,通过地图叠加热力图展示PM2.5浓度分布,红色区域代表重度污染,绿色代表优良,管理人员可实时掌握区域污染态势。但该平台仅支持单指标展示,无法同时呈现PM2.5与O₃的协同变化关系。

趋势分析可视化:采用折线图展示某流域COD浓度月变化趋势,结合柱状图呈现不同污染源贡献占比,为污染治理方案制定提供数据支撑。但传统折线图难以展示数据突变点与异常原因,需人工辅助分析。

应急预警可视化:在突发性污染事件中,通过动态

箭头展示污染物扩散路径,结合等高线图呈现浓度梯度变化。某化工园区泄漏事件中,可视化系统实现扩散范围预测,但响应延迟达150ms,影响应急决策时效性。

(二) 核心问题剖析

多源数据融合可视化不足:现有系统多采用“数据并列展示”模式,如分别呈现卫星遥感数据与地面监测数据,未实现时空校准与关联分析。某案例中,卫星监测的PM2.5高值区域与地面站点数据存在偏差,因未考虑地形遮挡因素,导致评估结果误差达12%。

动态交互与深度分析欠缺:多数平台以静态展示为主,交互功能局限于缩放、平移,缺乏个性化分析工具。用户需手动导出数据至Excel进行二次分析,完成一次污染源追溯平均耗时40分钟,效率低下。

性能瓶颈突出:海量数据渲染时易出现卡顿现象,某城市1000个监测站点的实时数据可视化时,帧率降至12fps,低于人眼舒适阈值(24fps)。同时,多用户并发访问时系统响应延迟显著增加,峰值达200ms。

评估模型与可视化脱节:可视化结果仅作为评估数据的“展示窗口”,未融入评估算法。例如,环境质量等级划分仍需人工代入公式计算,无法通过可视化界面直接调整参数并实时更新评估结果。

四、大数据可视化优化方案设计

(一) 多源数据融合可视化模型构建

设计“数据层-融合层-可视化层”三级架构:数据层采用HDFS分布式存储多源数据,通过Kafka实现实时数据流传输;融合层构建时空校准模型,采用动态时间规整(DTW)算法实现不同监测频率数据的时间同步,结合克里金插值法完成空间数据融合,数据关联误差降至5%以下;可视化层设计多维度融合展示模块,如“卫星影像+地面站点+地形数据”叠加展示,通过颜色编码污染物浓度,纹理编码地形类型,直观呈现地形对污染扩散的影响。

(二) 分层可视化与交互优化

构建“宏观-中观-微观”分层可视化体系:宏观层采用仪表盘展示区域环境质量综合指数,以红黄绿三色标识等级;中观层通过专题图呈现单指标空间分布,如大气污染热力图、水质等级分区图;微观层实现监测站点明细数据展示,支持点击站点查看实时监测曲线与历史数据对比。

优化交互功能模块:新增“智能筛选-关联分析-模型计算”联动功能,用户可通过下拉菜单选择评估指标与时间范围,系统自动生成指标关联热力图;设计污

染源追溯工具，点击污染高值区域即可显示潜在污染源、扩散路径及贡献占比，分析耗时缩短至5分钟内。新增自定义报表功能，支持用户导出个性化分析报告。

（三）性能优化策略

渲染性能优化：采用GPU加速的并行渲染算法，将数据分片后分配至多个计算核心同时处理，结合LOD（细节层次）技术，根据缩放比例动态调整数据精度，百万级数据渲染帧率提升至35fps。引入数据缓存机制，对高频访问的历史数据进行本地缓存，缓存命中率达82%，响应延迟降至40ms。

系统架构优化：采用微服务架构将数据处理、可视化渲染、交互分析模块解耦，实现负载均衡。多用户并发访问时，通过容器化技术动态扩展资源，支持500用户同时在线操作，峰值响应延迟控制在80ms内。

（四）评估模型与可视化融合设计

将层次分析法（AHP）与可视化系统深度融合，在界面设置指标权重调整控件，用户可通过滑块修改权重值，系统实时更新综合评估指数与等级划分结果。设计评估结果敏感性分析模块，通过tornado图展示各指标对评估结果的影响程度，为决策提供量化依据。

五、实验验证与效果分析

（一）实验设计

实验数据：选取某省2024年1-6月环境监测数据，包括1200个地面站点实时数据、30景卫星遥感影像、50条河流监测数据，数据总量达1.2TB，涵盖大气、水、土壤等18项评估指标。

实验环境：服务器配置为Intel Xeon E5-2690 v4处理器、64GB内存、NVIDIA Tesla V100显卡；客户端为主流浏览器Chrome 120.0；对比对象为优化前可视化系统与传统评估方法。

评估指标：选取数据处理效率、可视化效果、决策支持能力三类指标，其中数据处理效率包括数据清洗耗时、渲染延迟；可视化效果包括信息覆盖率、用户满意度；决策支持能力包括异常识别准确率、问题解决耗时。

（二）实验结果与分析

数据处理效率提升：优化后系统数据清洗耗时从120s降至35s，效率提升70.8%；百万级数据渲染延迟从150ms降至42ms，满足实时评估需求。多用户并发测试中，500用户同时访问时响应延迟为78ms，远低于优化

前的200ms，系统稳定性显著提升。

可视化效果优化：信息覆盖率从88%提升至96%，实现多源数据融合展示；邀请20名环境评估专家开展满意度调查，优化后系统满意度达92.5分（满分100分），较优化前提升28分。专家反馈显示，分层可视化与关联分析功能有效提升了数据解读效率。

决策支持能力增强：异常识别准确率从85%提升至96%，成功识别出32处潜在污染源，较传统方法多识别8处；问题解决耗时从40分钟缩短至5分钟，决策效率提升87.5%。在某流域污染治理案例中，基于优化后系统制定的治理方案使COD浓度下降35%，优于传统方案的22%。

六、结论与展望

（一）研究结论

本文构建了大数据可视化在环境质量综合评估中的应用框架，明确了多源数据融合、动态交互等核心技术要点。针对现有应用痛点，提出融合时空校准的多源数据可视化模型、分层交互设计及GPU加速优化策略，实现评估模型与可视化系统的深度融合。实验验证表明，优化后的系统在数据处理效率、可视化效果及决策支持能力上均实现显著提升，为环境质量综合评估提供了高效、精准的技术支撑。

（二）不足与展望

研究不足：一是高维小众指标（如新型污染物）的可视化表达仍需优化；二是系统在极端天气下的污染物扩散模拟精度有待提升。未来可从三方面展开研究：一是引入深度学习算法，实现污染物浓度预测与可视化的实时联动；二是结合VR/AR技术构建沉浸式环境评估场景；三是拓展跨区域数据共享可视化功能，支撑区域协同治理决策。

参考文献

- [1] 王雪, 李娟, 张伟. 大数据可视化技术在区域环境质量评估中的应用[J]. 环境科学研究, 2020, 33(7): 1652-1660.
- [2] 李刚, 王萌, 刘敏. 环境质量综合评估中的数据挖掘与可视化方法研究[J]. 计算机应用与软件, 2019, 36(10): 231-236+251.
- [3] 张敏, 陈浩, 赵静. 基于大数据的空气质量数据可视化研究[J]. 中外企业家, 2015(3): 249-250+253.