

大数据挖掘技术在环境污染溯源中的应用

张园园

(盐城市响水生态环境局, 江苏 盐城 224600)

摘要 大数据挖掘技术通过整合多源环境数据, 构建了从污染源识别到迁移路径分析的完整技术体系, 显著提升了环境污染溯源的精准性与效率。大数据背景下污染源的数据体系与技术架构包括通过多源数据采集与融合、大数据处理平台架构。大数据挖掘技术在污染源中关键技术的核心方法包括污染源识别与特征分析、污染物迁移路径与空间分布分析、智能溯源与预警系统。大数据污染源技术的应用场景包括水体污染源、大气污染源、土壤污染源。基于此, 本文对大数据挖掘技术在环境污染溯源中的应用进行了研究, 以期相关人员提供参考。

关键词 大数据挖掘技术; 环境污染溯源; 空气污染溯源; 水污染溯源; 土壤污染溯源

中图分类号: TP31; X5

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2025.28.009

0 引言

随着工业化进程加速, 环境污染的复杂性与跨区域特性日益凸显, 传统监测手段难以应对多源异构数据的爆炸式增长。大数据挖掘技术通过物联网传感器、卫星遥感等手段构建“空—天—地”一体化监测网络, 结合机器学习、时空聚类算法, 实现了污染源的动态解析与贡献率量化。这一技术转型为环境治理提供了数据驱动的科学路径。

1 环境污染溯源的定义

环境污染溯源是指通过环境监测数据、污染物特征分析及模型模拟等科学技术手段, 追踪和定位造成环境污染的源头, 并解析其迁移路径和贡献程度的过程。该技术体系旨在明确污染物的来源、类型、排放途径及空间分布, 其核心在于综合运用化学分析、同位素示踪、生物标记、遥感监测及大数据模型等多种方法, 构建从污染排放到环境影响的完整证据链, 从而为明确污染责任、制定精准治理措施及提升环境监管效能提供科学依据, 最终实现从被动响应到主动防控的治理模式转变^[1]。

2 环境污染溯源面临的挑战

环境污染溯源当前面临多重挑战, 其技术层面存在数据壁垒突出、多源异构数据融合困难以及模型泛化能力不足等问题, 难以在复杂环境下实现精准解析。管理层则体现为跨部门跨区域协同机制不畅、标准体系不完善以及基层执法能力薄弱, 导致溯源与治理衔接不足。与此同时, 成本高昂制约了先进技术在中西部和经济欠发达地区的普及, 而部分污染物的瞬时

性、隐蔽性特征以及监测数据与公众实际感受的偏差, 进一步增加了精准溯源的难度^[2]。

3 大数据背景下污染源的数据体系与技术架构

3.1 多源数据采集与融合

3.1.1 数据来源

环境污染溯源工作的数据来源广泛, 旨在构建一个多维度、立体化的信息采集体系。其核心数据基础由空天地一体化的监测网络所提供, 这包括部署于地面的物联网传感器、移动监测设备, 可进行大范围扫描的卫星遥感平台, 以及能够实现精细化局部巡查的无人机。此外, 环境管理业务中产生的各类资料也至关重要, 例如政府发布的生态环境状况公报、环境统计报告、企业的排污许可信息及相关台账。同时, 农业部门的农用地土壤污染状况详查成果、耕地环境质量类别划分资料, 以及水利部门掌握的水体底泥监测数据、灌溉水源信息等, 共同构成了溯源分析的重要依据。在特定情况下, 甚至需要对历史遗留工业场地的土地利用变迁、周边居民的问卷调查信息进行收集与分析。这些多源异构的数据经过专业的融合与预处理后, 为精准识别污染源、解析迁移路径提供了坚实的数据基础^[3]。

3.1.2 数据融合与预处理技术

数据融合与预处理技术是通过系统性的方法将来自不同来源、格式和结构的异构数据整合为统一、高质量、可用于分析的数据集的关键技术流程。该技术体系首先对多源原始数据进行清洗以处理缺失值、异常值和噪声数据, 并通过标准化与归一化等方法消除

量纲差异。在此基础上,运用数据集成与转换技术将异构数据在语义和格式上进行对齐与整合,并借助特征提取与选择方法构建有效的特征表示。根据融合层次的不同,可依次采用数据级融合、特征级融合或决策级融合策略,以逐步提炼信息并支持最终的分析与决策。该技术为环境污染精准溯源提供了坚实的数据基础^[4]。

3.2 大数据处理平台架构

大数据处理平台架构是一个为高效处理海量、多源、异构数据而设计的分布式系统框架,其核心目标是实现从数据采集、存储、计算到分析与应用的全链路管理。该架构通常采用分层设计,依次包括数据采集层、数据存储层、数据处理层、数据服务层以及数据应用层。数据采集层负责从各类数据源实时或批量抽取数据;数据存储层利用分布式文件系统和多种数据库技术实现数据的可靠持久化;数据处理层则根据业务场景选择批处理、流处理或混合模式对数据进行计算加工;数据服务层将数据能力封装成接口供上层调用;最终在数据应用层通过可视化工具或业务系统释放数据价值。整个架构确保了数据在各环节间高效、稳定、安全地流动,为基于数据的决策与创新提供支撑。

4 大数据挖掘技术在污染溯源中关键技术的方法

4.1 污染源识别与特征分析

4.1.1 源解析模型

源解析模型是基于质量守恒定律与统计学原理,通过建立污染物从排放源到环境受体的定量关系来识别和量化各类污染源贡献率的技术体系。该类模型主要包括以污染源为起点的扩散模型和以受污染区域为对象的受体模型两大类。扩散模型基于污染物排放强度、气象条件及地形因素模拟其迁移扩散过程,而受体模型则通过分析环境介质中污染物的化学特征来反推其来源。正定矩阵因子分解等受体模型无需预先获取源成分谱即可解析出污染源的贡献比例与成分特征,其优势在于能够处理复杂的多源混合污染问题。随着技术进步,源解析模型已从二维结构发展到能够整合时空尺度信息的三维乃至四维模型,并通过耦合地理信息系统与人工智能算法不断提升对污染源解析的精确度与动态模拟能力,为精准追溯污染责任主体和制定有效治理方案提供了核心科学依据^[5]。

4.1.2 同位素溯源技术

同位素溯源技术是一种基于自然界中稳定同位素比值差异来追溯物质来源与迁移路径的高精度分析方

法。该技术利用不同地区在气候、地质、水文等环境因素影响下形成的独特同位素“指纹”,通过高精度质谱仪测定样品中间位素组成并与已知源进行比对,从而精准识别污染源或农产品原产地。在环境污染领域,该技术可解析大气颗粒物水体污染物及土壤重金属的具体来源及其贡献比例;在食品安全领域则广泛应用于鉴别葡萄酒乳制品等高价产品的产地真实性。随着区块链与大数据技术的融合,同位素溯源正朝着构建全球化加密数据库与智能分析平台的方向发展,为跨区域环境治理与商品真实性认证提供更为可靠的技术支撑^[6]。

4.2 污染物迁移路径与空间分布分析

4.2.1 时空关联分析与扩散建模

时空关联分析与扩散建模通过整合地理信息系统的空间分析能力与污染物扩散数值模型,构建污染物在环境中迁移转化的精确模拟框架。该技术利用时空统计方法分析污染源与监测点之间的关联规律,结合高斯扩散模型或拉格朗日粒子模型等数值工具,模拟污染物在大气或水体中的传输过程。其中高斯模型适用于稳态均匀条件,其浓度计算依赖源强、风速及扩散参数;而拉格朗日模型通过追踪粒子运动轨迹可有效处理复杂地形下的非均匀扩散。通过耦合气象水文数据与地理高程信息,模型能够量化污染物在不同时空尺度的浓度分布,并借助克里金插值等空间插值技术生成连续分布图,实现从源解析到迁移路径的全过程动态可视化。

4.2.2 机器学习与数据挖掘算法

机器学习与数据挖掘算法是数据分析领域的核心技术,二者既有区别又紧密关联。数据挖掘是从大量数据中自动发现潜在有价值模式的过程,其典型任务包括关联规则分析、聚类分析以及异常检测等。机器学习则侧重于构建能够从数据中自主学习的算法模型,以完成预测或分类等任务,主要分为监督学习、无监督学习和强化学习等类别。在实际应用中,数据挖掘流程常依赖机器学习算法来构建预测模型,而数据挖掘所发现的模式与知识又有助于优化机器学习模型。当前,这些算法已广泛应用于客户细分、风险评估、销售预测及个性化推荐等多个场景,成为支撑数据驱动决策的关键工具^[7]。

4.3 智能溯源与预警系统

4.3.1 实时监测与动态预警

实时监测与动态预警是通过集成传感器网络、数据传输通信、智能分析算法和预警发布机制,构建对

目标对象状态进行持续性感知、即时性分析和主动性响应的技术体系。该系统依托部署于现场的各类传感设备实时采集物理、化学或行为参数,利用有线或无线通信网络将数据稳定传输至处理中心,并运用边缘计算或云端分析平台对数据进行清洗、融合与模型计算,从而精准识别异常状态或预测风险趋势。一旦监测值超越预设阈值或算法判定出现异常模式,系统便通过可视化界面、声光设备或移动终端等多渠道自动触发分级预警信号,形成从感知、分析、决策到反馈的闭环管理,实现对安全隐患的早发现、早预警和早处置。

4.3.2 人工智能辅助决策

人工智能辅助决策是指利用人工智能技术增强人类决策者处理复杂问题能力的新型决策范式,其核心在于通过数据驱动和算法分析,为决策者提供全面、精准和前瞻性的决策支持。该系统通常构建于多层架构之上,通过数据采集层汇聚多源信息,借助知识表示层构建领域模型,并依托推理决策层运行机器学习等算法以生成决策建议,最终通过人机协同的交互反馈层优化决策流程。该技术通过将人类的经验智慧与机器的高速计算、模式识别等能力深度融合,显著提升了在复杂多变环境下决策的科学性、时效性和准确性。其应用已广泛渗透至医疗诊断、智能制造、城市管理、金融风控等诸多领域,推动各行业的决策模式从传统的经验驱动向数据驱动和智能驱动转型。

5 大数据污染溯源技术的应用场景

5.1 水体污染溯源

水体污染溯源是通过综合运用水质指纹识别、多元统计模型及空天地一体化监测网络,精准识别污染物来源并解析其迁移路径的技术体系。其核心技术包括基于三维荧光光谱或紫外-可见光谱的水质指纹识别技术,通过比对污染水体与已知污染源的特征光谱数据库实现精准溯源,一次分析仅需约 21 分钟;同时结合多元统计模型定量计算生活源、工业源及农业面源等不同污染源的贡献比例,并依托物联网传感器、无人机巡航及无人监测船等构建的实时监测网络,实现从污染预警、溯源分析到执法取证的全链条闭环管理。该技术已成功应用于太湖水源地保护、深圳电镀工业园区偷排查处等实际场景,有效支撑了水环境的精准治理与高效监管。

5.2 大气污染溯源

大气污染溯源是通过集成卫星遥感、地面监测、移动走航和无人机巡查等天空地一体化监测手段,结

合数值模型与人工智能算法,精准识别并追踪大气污染物来源、迁移路径及贡献率的全过程。该技术体系利用高分辨率卫星影像锁定区域污染高值区,通过走航车和无人机进行精细化巡查以定位至企业乃至排放工段,并借助受体模型、扩散模型及伴随算法等定量解析不同污染源的贡献比例,实现从宏观区域到微观企业的精准管控。这一技术支撑了重污染天气的应急响应与常态化治理,推动大气污染防治从经验判断向数据驱动决策的科学转变。

5.3 土壤污染溯源

土壤污染溯源是通过系统性的技术方法确定土壤中污染物的来源、迁移路径及贡献程度的过程。该技术体系综合运用环境地球化学、同位素示踪、空间统计分析 & 数值模型等手段,对土壤中的重金属、有机污染物等进行定性识别和定量解析。其核心在于通过污染源指纹特征比对、扩散模拟及多源受体模型计算,建立从污染排放到土壤环境的因果关联,进而精准界定主要污染源及其空间影响范围,为制定精准治理策略和落实环境管理责任提供科学依据。

6 结束语

大数据挖掘技术已深度嵌入环境污染溯源的全链条,其未来发展需聚焦技术融合与机制创新。长远来看,通过强化复合型人才培养与国际合作,推动从“污染清除”到“生态增值”的转型,大数据挖掘技术将为全球环境治理提供更智能、可持续的解决方案。

参考文献:

- [1] 沈航,闫赫.基于大数据分析的大气污染防治策略研究[J].生态与资源,2024(09):105-107.
- [2] 姜锦玉,陈磊杰,巨小芹.大数据在环境污染防治管理中的应用研究[J].皮革制作与环保科技,2024,05(16):82-83,86.
- [3] 马弋惠,姚安晓.基于大数据技术的生态环境监测与评价[J].皮革制作与环保科技,2024,05(15):46-47,50.
- [4] 丁方纪,楚焕庆.环境保护领域中信息数字化分析技术的应用研究[J].信息与电脑(理论版),2024,36(14):12-15.
- [5] 全吉昌,任樊杰.基于大数据的污染源在线自动监测数据分析与决策支持系统研究[J].生态与资源,2024(06):98-100.
- [6] 何希,霍志彬.大数据时代水污染监测及防治探讨[J].中国矿业,2024,33(S1):186-190.
- [7] 徐中华.大数据在生态环境治理领域的应用与前景分析[J].产业与科技论坛,2024,23(12):35-37.