生态与农村环境学报 2022，38（9）：1101-1111
Journal of Ecology and Rural Environment

DOI：10.19741/j.issn.1673-4831.2021.0668

陆晓松，王国庆，李勖之，等. 场地环境大数据采集和机器学习方法在污染智能识别中的应用研究进展[J].生态与农村环境学报，2022，38(9)：1101-1111.

LU Xiao-song，WANG Guo-qing，LI Xu-zhi，*et al*. Research Progress of Big Data of Site Environment Acquisition and Machine Learning Method in Pollution Intelligent Identification[J].Journal of Ecology and Rural Environment，2022，38(9)：1101-1111.

场地环境大数据采集和机器学习方法在污染智能识别中的应用研究进展

陆晓松[1,2]，王国庆[1,2①]，李勖之[1,2]，杜俊洋[1,2]，孙 丽[1,2]（1. 生态环境部南京环境科学研究所，江苏 南京 210042；2. 国家环境保护土壤环境管理与污染控制重点实验室，江苏 南京 210042）

摘要：由于大数据技术的快速发展，用于分析挖掘场地污染特征和成因机制的数据量和类型也大幅增加，传统的场地环境数据获取、清洗和挖掘方法难以满足大数据的存储和处理要求。近年来，采用机器学习算法对场地多源异构数据进行挖掘，实现地块尺度、区域尺度的污染识别已成为研究热点。系统综述了场地污染智能识别大数据的获取、处理和挖掘方面的现状和不足，提出了利用5G和互联网、终端信息采集、网络爬虫、自然语言处理方法获取场地环境数据的应用对策。针对场地多源数据集成和融合的关键技术措施以及未来我国场地污染智能识别模式进行展望。

关键词：场地污染识别；生态环境大数据；机器学习；信息技术

中图分类号：X53 **文献标志码**：A **文章编号**：1673-4831(2022)09-1101-11

Research Progress of Big Data of Site Environment Acquisition and Machine Learning Method in Pollution Intelligent Identification.

LU Xiao-song[1,2], *WANG Guo-qing*[1,2①], *LI Xu-zhi*[1,2], *DU Jun-yang*[1,2], *SUN Li*[1,2] (1. Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment, Nanjing 210042, China; 2. State Environmental Protection Key Laboratory of Soil Environmental Management and Pollution Control, Nanjing 210042, China)

Abstract: Due to the rapid development of big data technology, the amount and type of data used to analyze and mine site pollution characteristics and formation mechanisms have also increased significantly. The traditional methods of acquiring, cleaning and mining site environmental data are difficult to meet the requirements of big data storage and processing. In recent years, it has become a research hotspot to use machine learning algorithms to mine multi-source heterogeneous site data to realize pollution identification at site and regional scales. The status and deficiencies of big data acquisition, processing and mining on intelligent identification of site pollution are systematically reviewed. Then, the application countermeasures to obtain site environmental data using 5G and the Internet, terminal information collection, web crawler, and natural language processing methods are proposed. Finally the key technologies of site multi-source site data integration and fusion and the future intelligent identification mode of site pollution in China are prospected.

Key words: site pollution identification; big data on the ecological environment; machine learning; information technology

随着城市化和产业转移进程的加快，大量工业企业被关闭或迁移，遗留下数量众多的污染场地[1-2]。掌握场地土壤和地下水环境特征并识别污染状况，是污染场地风险管控的基础，也是控制场地污染和保障环境安全的重要前提，同时有助于对工业企业用地开展针对性的环境管理。由于污染场地具有污染物来源复杂，污染深度、空间分布变异性大等特点，采用布点采样调查、检测分析和模型预测的方法识别场地污染和风险，往往存在边界判定模糊、风险预测偏离较大、成本较高的问题；并且对于场地污染的成因机制，例如与源汇的关系、各特征指标的影响程度等，不能清楚描述。因此，国内外研究者希望通过构建区域经济、土壤和地质背景，以及行业类别、生产工艺、产排污特征、敏感

收稿日期：2021-11-02

基金项目：国家重点研发计划(2018YFC1800202)

① 通信作者 E-mail：nies.sepa@163.com

受体等场地污染识别指标体系[3-4]，采用机器学习方法，结合地统计方法和污染迁移模型，构建场地污染风险的识别和预测模型，以便能够更快速、高效地识别场地污染和风险，减少场地调查评估和环境管理成本。

2016 年，环境保护部印发了《生态环境大数据建设总体方案》，明确大数据、云计算基础能力等在大气、水、土壤、噪声等环境管理信息化方面的建设要求[5]。2016 年 5 月，国务院印发《土壤污染防治行动计划》，明确提出要“提升土壤环境信息化管理水平，利用环境保护、国土资源、农业等部门相关数据，建立土壤环境基础数据库，构建全国土壤环境信息化管理平台”[6]。因此，基于已有的数据成果开展大数据挖掘，是进行场地土壤污染智能识别和成因分析等研究的迫切需求。

随着国家重点研发计划“土壤污染成因与治理技术专项”的实施，利用大数据和机器学习方法驱动场地污染识别与风险管控已逐步成为该领域的研究热点。通过大数据技术对场地基本信息、污染特征信息等进行更全面的获取，结合机器学习方法，可以发现传统理化模型难以得出的规律，形成全新的场地污染识别模式。

1 我国生态环境大数据应用

随着信息技术的发展，获得生态环境数据的信息来源、数据量都得到大幅扩展，我国生态环境信息化发展已开始向大数据方向转变。大数据无论从数据量、数据类型、获取和存储方式，还是结果展示及决策分析上，都有别于传统数据应用（表 1）[7]。在数据规模与日俱增的情况下，采用传统数据清洗和挖掘算法处理数据的效率受到极大限制。采用传统的环境信息编码、线面分类法与多维树状分类的方式已无法满足多源异构的环境大数据获取和存储的需求[8]。与传统的生态环境信息化管理应用相比，生态环境大数据具有多源异构、海量源数据的特点，采用数理统计分析、情报检索、机器学习、专家经验判别和模式识别等处理体系和方法，对看似与生态环境领域无关的数据进行深度挖掘，实现更多的数据价值，形成支持决策[9]，已成为大数据技术的核心。

表 1 生态环境大数据应用与传统数据应用的区别

Table 1 The differences between ecological environment big data applications and traditional data applications

项目	传统数据应用	大数据应用
数据来源和类型	来源单一、分布集中的采样监测数据；经过整理汇总的结构化数据	来源多样的数据，包含直接获取的与生态环境无直接关联的海量多源异构数据
数据量	数据量通常为 GB 级；数据集成简单	数据量从 TB 到 PB 级，甚至更大；持续更新；数据集成困难
数据存储方式	关系型数据库（RDBMS）	分布式文件存储系统（HDFS），NoSQL 数据库
应用方法	数据统计分析，构建数值模型	深度挖掘数据价值，数值模型与大数据模型相互验证
结果展示形式	数据分析结果，以及图表展示	可视化展示、虚拟现实等各种直观的表达形式
决策支持	基于数据分析结果，根据人为经验形成决策	通过机器学习和深度学习方法直接获得有效决策，减少人工业务操作

“十二五”以来，随着新《环境保护法》《大气污染防治行动计划》《水污染防治行动计划》等陆续颁布，我国大数据在生态环境保护领域的研究与应用得到了快速发展。上述生态环境保护法规和制度促进了各省（区、市）围绕环境质量监测、污染源监管、移动执法、排污收费等核心业务，依托云计算、大数据、遥感及地理信息、物联网、视频监控等新技术，尝试建立生态环境信息中心，以统筹环境数据资源，不断发展环境保护信息化水平[10-12]。

相比于大气和水环境大数据，我国场地污染识别领域的土壤环境大数据研究起步较晚，目前的各类研究对土壤环境大数据的概念界定依旧模糊，对于信息化发展和大数据应用之间的关系认知不明确，特别是基于场地土壤环境数据采集、处理、分析、访问及应用等层级设计的大数据架构研究和应用十分缺乏。

2 场地环境大数据采集技术的应用

2.1 移动互联和手持终端数据采集技术

已有的场地污染调查数据以调查报告的文本数据为主，主要通过资料收集、人员访谈和现场踏勘方式获取。根据场地污染调查和风险评估过程，需对场地及周边情况、涉及危废的设施/构筑物情况、场地利用类型及周边情况以及场地污染特征进行调查分析（图 1）。还有研究通过检索获取场地污染识别相关文献，统计和归纳各地区、行业的企业产品及原辅材料、生产工艺、产排污情况等，构建场地污染识别的指标体系[3-4]。由于场地污染识别所需数据来源和结构不同，采用传统的人工检索和摘录方法获取数据分析和挖掘所需的结构化数据，存

在效率低、规范性差等问题。

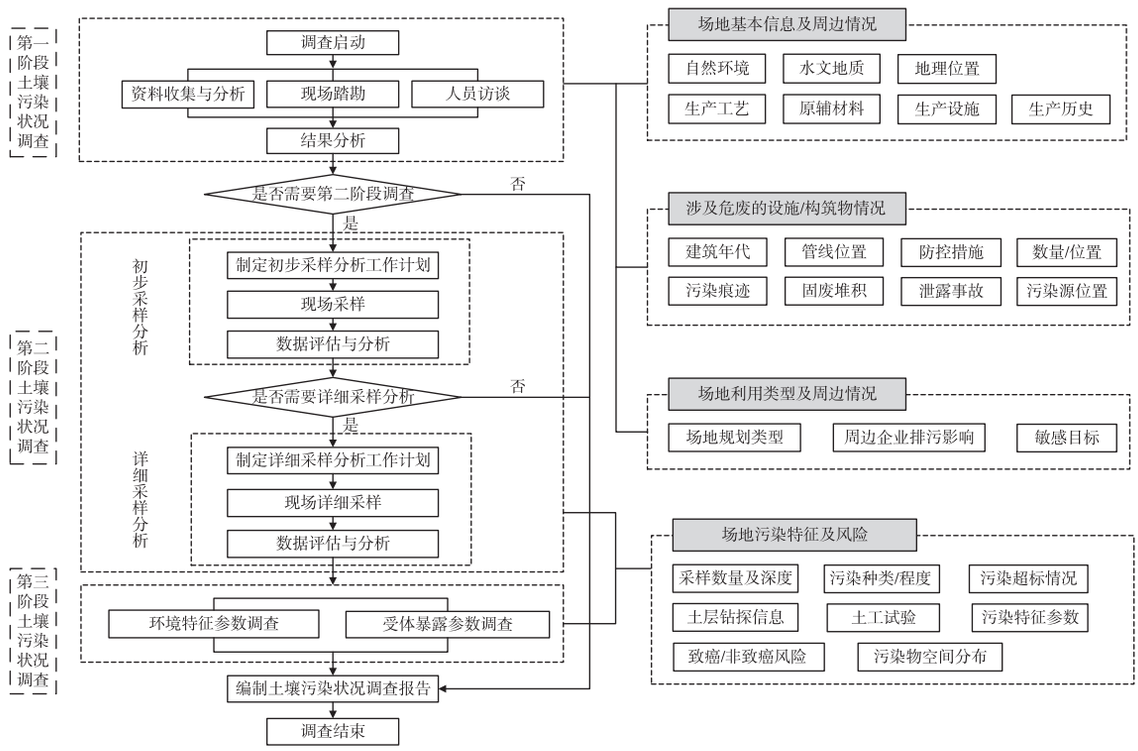

图 1　场地环境调查和风险评估过程需要获取的信息

Fig. 1　Information required for the site environmental investigation and risk assessment process

企业生产年限、产品和原辅材料，污染源信息、迁移途径和敏感受体等数据，通常需要通过现场调查、资料收集和人员访谈等方式获取。基于移动互联网、全球定位系统的手持终端信息采集技术能够用于准确、高效地采集场地信息，与传统的纸质和人工记录采集信息方式相比，其自动化程度高，能够形成完备的电子档案。在场地特征指标获取的方法上，使用手持终端将预先设定的信息逐一录入，替代传统的采样记录单、人员访谈记录等方式，能够有效提高数据采集效率、记录规范性和准确性，便于形成结构化的场地调查数据集。近年来，场地调查信息化管理技术呈现快速发展趋势，我国于“农用地土壤污染状况详查”和“重点行业企业用地土壤污染状况调查”期间开始使用手持终端和信息管理平台采集、存储和管理调查数据。我国已有部分省（区、市）的生态环境部门、大型企业等，相继设计和开发了场地调查和信息管理平台，用于支持关闭搬迁企业地块、化工园区和典型行业企业及周边土壤污染状况调查等工作，促进了我国土壤环境管理的信息化水平。

2.2　基于网络爬虫和文本分类的数据采集技术

通过传统的手工方式收集整理场地相关公共数据，由于效率和及时性较低，已不能满足大数据分析和挖掘的需求。随着网络爬虫技术的发展，通过互联网直接爬取数据的技术已日趋成熟，极大地提高了数据获取的效率。网络爬虫是通过模拟人类与浏览器交互访问互联网的过程，并仿照复制、粘贴的方法采集网页中呈现出的各种内容，通过相应的程序解析出需要的文本、图片和视频等形式的数据（图 2）。目前，流行的爬虫工具包括基于 Java 语言的 Nutch、Heritrix，基于 Python 语言的 Scrapy、Crawley 和 PySpider，以及基于 Php 脚本语言的 Phpspider 和 Beanbun 等[13-14]。其中，Scrapy 框架是一个较为成熟的开源网络爬虫框架，继承了 Python 语言高效、简单的特点，已被广泛应用于大数据挖掘研究。

自然语言处理（natural language processing，NLP）是数据挖掘领域最重要、最具代表性的组成部分之一，在文本处理、机器翻译和问答舆情分析等任务中的运用日益广泛和成熟[15]。NLP 的处理流程通常包括文本获取、语料预处理、特征化处理、模型训练和评估等过程（图 3）。文本语料在输送给语言模型前一般需进行分词、词性标注和命名实体识别等预处理，目前可使用的中文分词和文本预处理

的开源工具包括 Jieba、ANSJ、THULAC、LTP 等[16]。传统的 NLP 模型主要基于规则和统计的框架,其中,基于规则的方法通过采用正则表达式表示需要匹配的字符串,操作简单,灵活性好,但只适用于表达规范的文本,文本特征抽取效果高度依赖于制订的规则[17]。基于统计的方法包括隐马尔科夫(HMM)和条件随机场(CRF)等模型,通过建立语言模型对输入的语句样本进行单词的划分,并对划分结果进行概率计算,获得概率最大的预测结果[18]。随着深度学习算法的不断发展,深度神经网络模型由于具有强大的文本表征能力、学习能力等优点,近年来已成为 NLP 在各领域的研究热点[19]。

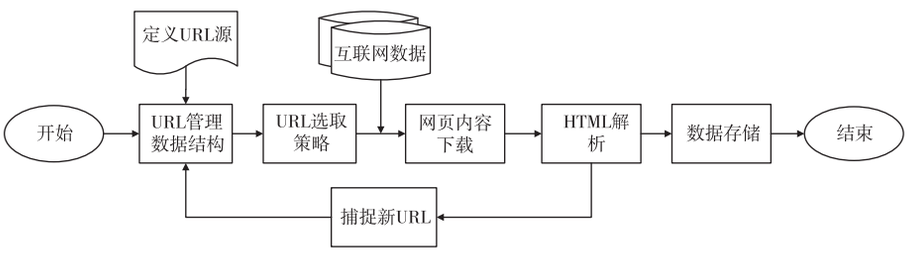

图 2 互联网爬虫框架工作流程

Fig. 2 Flowchart of internet crawler framework

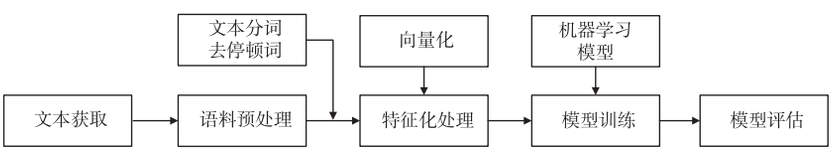

图 3 基于自然语言处理的数据挖掘流程

Fig. 3 Flowchart of data mining based on natural language processing

利用已有的多源异构数据,包括图像、文本等非结构化数据,采用 NLP 和机器学习的方法,通过特征提取、数据类型转换等,获得对于场地污染识别有价值的数据,已成为场地环境大数据获取的重要途径。WANG 等[20]分别采用基于规则和基于统计的方法,从土壤调查报告、实地调查报告和相关手稿等文本数据源中,有效提取土壤环境相关字段和信息,并将其转化为结构化数据。通过网络爬虫高效获取互联网公共数据,结合 NLP 非结构化文本信息抽取,形成简单化、易操作的数据获取方法,能够为场地环境数据采集提供有效途径。

2.3 基于大数据框架构建场地污染智能识别信息系统

随着 Hadoop 和 Spark 等不断发展,基于分布式框架设计已成为大数据应用的重要标签。大数据技术以空间换时间的方式,将单机算法在空间上并行化,弥补其单机计算资源不足的缺点,使得通过存储、分析和计算手段处理海量复杂、冗余数据的效率得到大幅提升[21]。其中,Hadoop 是目前应用最为广泛的分布式大数据处理框架,它以 HDFS、Yarn 和 MapReduce 为核心组件,分别负责大数据的分布式存储、资源调度和计算[22],具备可靠、高效、可伸缩等特点。在大数据 Hadoop 的广义生态圈中,还包括 Flink、Zookeeper、Sqoop、Hive、HBase、Flume、Pig 等组件和工具,用于完善集群管理和分布式协作。由于 Spark 本身并没有提供分布式文件系统,大多依赖于 Hadoop 的分布式文件系统 HDFS,并且继承了 MapReduce 的线性扩展性和容错性,将计算的中间数据与结构均优先保存在内存资源中,具有计算效率更高、兼容性更广、容错能力更强的优势[23]。MapReduce 和 Spark MLlib 都是基于分布式架构中用于数据计算和支持机器学习的数据处理模块,不仅包括分类、回归、聚类、协同过滤和降维等常用算法[24],还包括一些高层次的 API,使机器学习算法在实际大数据处理工作中得到简化[25]。有研究者针对生态环境大数据,特别是多源遥感数据、环境监测连续数据,利用 Hadoop 高效的数据存储和计算能力,开展机器学习算法的应用研究[26-29]。目前,国内有研究者针对场地污染识别需求,开始探索开发场地环境大数据应用平台[30]。在未来的研究中,基于大数据框架设计和开发场地污染智能识别系统,能够提高大量非结构化数据存储和解译算力,有助于实现多源异构数据清洗、融合、数据挖掘和决策分析(图 4)。

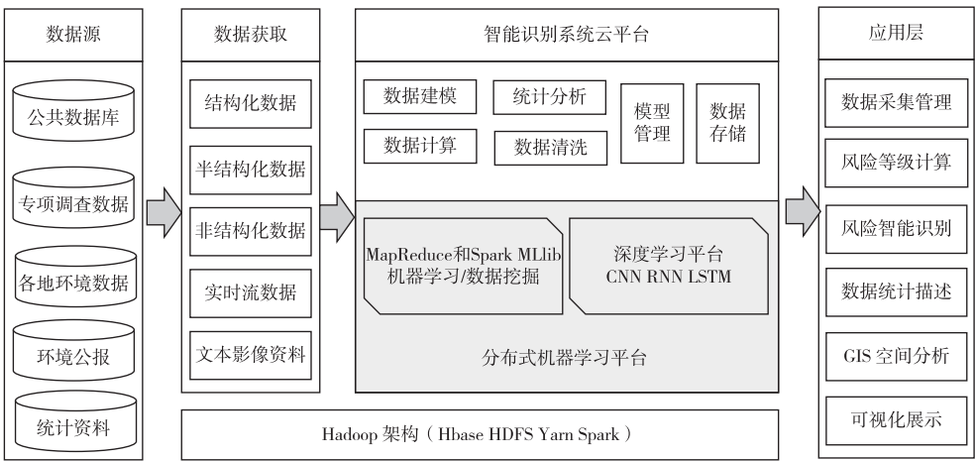

图 4　大数据智能识别云平台架构

Fig. 4　Architecture diagram of big data intelligent identification cloud platform

3　机器学习方法在我国场地污染识别中的研究进展

3.1　场地污染识别常用的机器学习算法

机器学习是计算机基于数据构建概率统计模型并运用模型对数据进行预测和分析的学科。该技术方法使用数据挖掘、人工智能、模拟仿真、关联分析等现代技术手段，在解决复杂环境问题方面展示出明显优势[31-32]。根据学习任务，机器学习算法可分为回归、分类和聚类；根据学习方式或有无标签，可以分为监督式学习、非监督式学习、半监督式学习和增强学习，用于分类和回归任务的算法主要为监督式学习[33]。机器学习方法作为重要的大数据挖掘方法，已被广泛运用于大气、海洋、矿山等环境污染预测和特征识别等方面[34-35]。人工神经网络（ANN）、SVM 和随机森林等机器学习算法在大数据分析和挖掘方面具有很高的热度[36]。上述机器学习算法的函数逼近、模式识别、回归计算等已被广泛应用于生态环境领域，包括环境质量变化预警预报和综合评价等，并取得了较好效果。朴素贝叶斯算法（naive Bayes，NB）是基于概率论的分类算法，由于具有简单易用和高效的特性，在基于文本分类中得到广泛应用[37]。

3.2　地块尺度场地污染识别研究

在地块尺度上的污染识别，已有研究多通过构建场地特征指标与污染物含量之间的线性和非线性关系，并将其与三维建模等方法结合，对污染物含量及其空间分布进行预测。任加国等[38]基于某重金属和 PAHs 复合污染场地的少量分析测试数据，运用多元统计方法分析两类土壤污染物之间的关联性，并利用已知数据建立 BP 神经网络模型，预测缺失土壤样本中重金属和 PAHs 含量。LIU 等[39]采用随机森林与普通克里格相结合的 RFOK 模型，通过建立污染物含量与地形要素、样点环境要素和遥感数据等多源环境数据之间的非线性关系对某大型砷渣站点土壤砷空间分布进行预测。此外，还有研究通过数学模拟、机器学习方法，对场地土壤和地下水的电阻成像（ERT）、污染羽分布等理化性质进行反演，获取污染源在地块尺度上的空间分布以及范围。能昌信等[40]通过 Sobel 算子提取场地电阻率数据的边缘特征，并将其与深度卷积神经网络（CNN）算法结合，用于污染场地电阻率层析成像的反演，能够显著提高场地污染面积、位置的识别精度。王玉玲等[41]将聚类算法 K-means、模糊 C 均值（FCM）和混合高斯模型（GMM）3 种聚类算法引入 ERT 监测系统，用于识别垃圾填埋场渗滤液污染范围。

上述研究主要利用机器学习的回归和聚类算法，对场地污染的源-汇关系进行挖掘，重视特征指标与污染类型、程度和范围之间的“因果关系”。然而，受限于场地样本数量、数据获取的方法和途径，不同地区和行业类别的场地污染驱动因素和敏感指标存在很大差别，针对单一污染场地开展的污染识别的研究结果，很难被运用到区域或者不同行业类别的场地污染特征识别中。

3.3　大数据在区域尺度场地污染识别中的研究

我国于 20 世纪 80 年代开始，相继开展了全国土壤环境背景值调查、土壤污染状况调查、多目标区域地球化学调查、农产品产地土壤重金属污染防治普查、农用地土壤污染状况详查、重点行业企业

污染状况调查等多次全国尺度的土壤环境调查。上述数据既包含了结构化数据,又有文本报告、矢量和栅格图件等非结构化数据,具有大数据海量、多源异构的特点,但分布在生态环境、国土资源、农业等多个政府部门[42]。由于上述专项调查数据大多属于内部数据,很难全面整合应用于场地污染识别和成因挖掘分析。有研究者通过工商登记网站、行业企业信息网站、文献检索数据库等途径,获取场地地理位置、企业规模、行业类别、生产年限、地块使用历史等基本信息,用于构建我国场地污染识别的指标体系和方法。例如:王鑫等[43]从中国农药信息网获取农药企业生产信息,用于构建场地土壤污染快速识别的指标体系;李强等[44]基于近 50 篇国内外文献资料,收集了我国 13 个省份的冶炼企业场地污染数据,分析了生产过程、生产环节和污染装置 3 类场地生产信息对应的可能产生的污染物及其潜在风险;还有研究者从全国排污许可证管理信息平台和绿网公共数据库等获取企业生产工艺、原辅材料和污染排放信息,作为潜在的敏感指标,用于区域尺度和不同行业企业场地污染特征和成因驱动因素的分析挖掘[45]。

土壤类型、土壤理化性质和土地利用等基础数据也被用于场地污染状况和变化趋势等研究。张健琳等[46]基于检索的文献、土地利用类型等信息,采用 Meta 分析方法量化了不同情景模式下金属矿开采场地对周边土壤污染的影响。郭长庆等[45]基于高分辨率遥感影像、土地利用/覆盖数据、土壤类型和环境专题数据等信息,分析我国能源开采和加工行业场地时空变化趋势等。YANG 等[47]以土地覆盖、矿山或冶炼厂距离、道路距离、地形高程等作为影响因子,利用地理探测器模型识别驱动因子对土壤重金属累积的影响强度,识别湖北某地级市土壤重金属污染主要来源和贡献率。

此外,谷歌和百度搜索引擎 API、OSM 电子地图等提供了数据共享方式,支持研究人员和开发者通过网络爬虫方式获取兴趣点(POI)数据。JIA 等[48]基于谷歌搜索引擎 API 获取我国长三角地区 7 000 多个企业地理位置、名称等基本信息,基于 HMM 模型对文本进行分词,采用 SVM、NB 和 ANN 这 3 种机器学习算法预测所有企业的行业类别,采用双变量局部莫兰指数分析不同行业类别与土壤 Cd 和 Hg 浓度测量值之间的关系,并揭示其形成原因。黄国鑫等[30]以南方某地级市为研究区,基于 NLP 和机器学习方法,分别采用改进型朴素贝叶斯、随机森林和 XGBoost 等分类模型,利用 POI 数据对企业的行业类别和地块污染进行预测;结果表明,改进型朴素贝叶斯模型能够更有效地预测疑似土壤污染企业,具有较好的准确率和召回率。

4 场地污染识别数据挖掘分析方法的完善

4.1 构建和优化场地污染识别指标体系

由于不同的研究在场地调查数据获取的难易程度、数据挖掘分析角度和数据处理方式等方面存在较大差异,对于文本、图像等非结构化数据解译方式各异,已有研究所构建的场地污染识别的特征指标体系各不相同;并且往往重视特征指标数量和模型训练数据规模,而忽视污染过程的“因果关系”,用于区域尺度和不同行业类别的场地污染识别存在精准性不高、科学性不足的问题。因此,综合考虑场地污染过程、产排污特性,建立全面和统一的场地污染识别指标体系,是对区域尺度和不同行业企业场地污染进行识别的基础。

基于场地污染识别的清单方法,是区域尺度和不同行业企业地块污染识别的重要方法。该方法通过建立场地特征指标与土壤污染特征之间的关联,在减少采样分析成本和缩短周期的前提下,实现场地污染识别[49]。我国生态环境部于 2019 年制定发布了基于在产、关闭搬迁企业地块污染源-污染途径-受体模式的风险筛查与风险分级相关技术规定,涉及土壤和地下水的企业环境风险管理水平(在产企业)、地块污染现状、污染物迁移途径和污染受体 4 类[50]。该指标体系的建立为我国场地污染识别提供了重要依据。由于该技术规定采用专家打分法设定指标赋分,具有一定主观性。李天魁等[51]基于污染物迁移转化多介质模型,采用不确定性分析、灵敏度分析、案例实证分析等数值模拟方法,对关闭搬迁企业地块风险筛查方法开展综合评估,并对指标内部分级与赋分调整提出建议。传统的指标权重确定和赋分方法还包括层次分析、模糊隶属度分析、灰色关联分析、相关分析和主成分分析等[52-53]。基于机器学习的特征选择方法(过滤法、封装法和嵌入法),筛选敏感指标和获取评价指标重要性,已成为确定指标权重的重要方法[54]。例如,嵌入法通过随机森林等模型训练结果,获取特征指标的贡献率或重要性,适用于指标权值确定和优化[55]。针对我国重点行业企业土壤污染调查中获得的大量地块的基础信息和采样检测结果,采用机器学习方法优化污染识别指标体系和权重赋分,期望更有效地提高场地污染识别的准确性和适用性。

4.2　小样本和不平衡数据预处理和挖掘方法

由于场地调查数据获取困难且缺少历史资料，或是在线数据采集系统不稳定等因素，可能导致数据缺失或异常，可用于模型训练的数据量较小。针对小样本数据机器学习问题，一是可以采用最近邻、聚类、回归分析等方法，补齐数据集中的缺失值和剔除异常数据，以扩充有效的模型训练数据集[56]；二是在数据挖掘阶段，选择弱监督学习、迁移学习和元学习等，利用先验知识来弥补监督学习信息的不足[57]。

此外，研究者收集到的少量场地调查数据往往存在样本数据不平衡的问题，例如，通过互联网或文献数据库检索得到的某些重点行业企业场地调查信息，土壤污染超标的地块样本往往较多，未污染的地块往往较少。基于不平衡数据的学习一直是各领域数据挖掘的难题，在利用高度不平衡数据训练分类模型时，分类器很容易倾向于多数类而忽略少数类[58-59]，少数类样本易被作为噪声而抛弃[60]。目前，对于不平衡数据的挖掘任务，主要针对预处理和挖掘算法两个方面进行处理。其中，数据预处理方法包括利用采样法再平衡样本空间以缓解其不平衡程度；采用特征选择和特征提取（PCA、SVD 等）方法筛选出使分类模型性能更好的特征子集。数据挖掘方法包括半监督聚类算法、权重强化监督学习 Boosting 算法、代价敏感模型等，用于提高模型性能和泛化能力[61]。因此，重视评估数据质量是否能满足污染识别需求，以及数据处理和模型应用的数值机理研究，更有助于提高场地污染识别数据挖掘的准确性和适用性。

5　场地环境大数据及污染智能识别展望

5.1　土壤污染信息智能采集技术

目前，大数据在气候变化预测、生态监测网络与模拟、区域大气污染治理等生态环境领域得到初步应用[9]。我国《生态环境大数据建设总体方案》《生态环境监测网络建设方案》《土壤污染防治行动计划》都对生态环境监测和土壤环境监测网络构建提出了明确要求。我国大气环境监测及预警方面取得的成果最为显著，已建立了覆盖全国的大气污染监测网络[7]。与大气、水等环境监测网络相比，我国土壤环境监测仍处于试点和筹建阶段，亟需推进基础能力、技术支撑、信息化管理和制度创新等方面建设[62]。

在未来的研究中，借助于大数据技术在信息采集、预处理、存储与管理方面的优势，重视 5G、物联网、全球定位等信息化技术在土壤环境监测网络中的作用，能够有助于实现土壤污染智能识别和综合决策。在“十三五”期间，我国已实现了以移动终端、信息化平台代替传统方法的数据采集和处理方式，具备了高效、规范和准确地获取场地基本信息和土壤污染调查数据的技术能力。基于光离子、X 射线荧光光谱、污染传感等的场地污染快速检测技术也日趋成熟，为构建同时检测土壤中重金属、有机污染物的仪器系统提供了基础[63]；结合土壤污染遥测技术的研究和应用，能够提升土壤环境立体监测以及快速掌握污染物在土壤中时空分布的能力[64]。

5.2　基于深度学习的污染智能识别方法

随着人工智能的三要素算力、算法和数据交替突破的迭代发展，计算机视觉（CV）技术在光学字符识别（OCR）、边缘提取、手写数字识别，以及人脸识别、动态背景检测、图像生成等更复杂的领域都有不断拓展[65]；同时，图像非结构化数据的特征提取和挖掘在生态环境领域受到重视。已有研究利用 SIFT、Gabor、HOG、Haar 等图像颜色、纹理等特征提取算法，使用 PCA、聚类的光谱波段选择，与空间纹理、边缘特征提取算法相结合的方法识别污染源和污染范围[66-69]。近年来，基于 CNN 的深度学习算法模型 AlexNet、VGG、GoogLeNet、ResNet 等在图像识别的预测性能方面得到不断提升，使得深度学习在 CV 和图像处理领域占据着不可替代的地位[70-71]。在未来的研究中，基于无人机遥感、高光谱图像等，利用深度学习算法自动地进行抽象、隐式学习，实现非结构化数据高层语义特征提取和目标识别[15-16]，可以为场地污染智能识别提供新的方法和模式。

6　结论

随着可获取的相关数据源和数据量的大幅增加，分布式数据存储和数据分析计算等信息技术水平的提高，我国生态环境大数据应用正处于快速发展阶段。将海量的多源异构数据和信息进行链接，通过数据分析挖掘的方式驱动管理决策，将成为促使生态环境管理向智能化、数字化和精准化转变的重要驱动力。随着《土壤污染防治行动计划》的发布、我国多部门和各地区土壤环境调查的开展，将大数据应用于场地污染识别的方法受到研究者更广泛的关注。基于大数据技术，探索区域尺度和行业企业的场地污染识别方法将成为研究热点之一。利用物联网、5G 等信息技术手段，以及大数据集成

和融合的思维范式,有助于解决“数据孤岛”问题,实现场地环境数据的高质量和深度挖掘。机器学习、深度学习和自然语言处理等方法的应用,作为多源异构数据集成、清洗和挖掘的核心,提供了高质量的数据和探究场地污染成因机制的重要方法。在未来,利用大数据和深度学习方法,对源-汇关系、污染传感和遥感等信息的特征提取和数据融合,实时快速地进行综合分析和决策推断,能够为场地污染识别提供更加准确和高效的新方法。

参考文献:

[1] 骆永明.中国污染场地修复的研究进展、问题与展望[J].环境监测管理与技术,2011,23(3):1-6.[LUO Yong-ming.Contaminated Site Remediation in China: Progresses, Problems and Prospects[J].The Administration and Technique of Environmental Monitoring,2011,23(3):1-6.]

[2] 姜林,钟茂生,张丽娜,等.基于风险的中国污染场地管理体系研究[J].环境污染与防治,2014,36(8):1-10.[JIANG Lin, ZHONG Mao-sheng, ZHANG Li-na, *et al*. Establishing a Risk Based Framework for Contaminated Site Management in China[J]. Environmental Pollution & Control,2014,36(8):1-10.]

[3] RAMPANELLI G B,BRAUN A B,VISENTIN C,*et al*.The Process of Selecting a Method for Identifying Potentially Contaminated Sites:A Case Study in a Municipality in Southern Brazil[J]. Water,Air,& Soil Pollution,2021,232(1):1-20.

[4] 张秋垒,黄国鑫,王夏晖,等.基于案例推理和机器学习的场地污染风险管控与修复方案推荐系统构建技术[J].环境工程技术学报,2020,10(6):1012-1021.[ZHANG Qiu-lei, HUANG Guo-xin, WANG Xia-hui, *et al*. Construction Technology for Site Pollution Risk Control and Remediation Scheme Recommendation System Supported by Case-based Reasoning and Machine Learning[J]. Journal of Environmental Engineering Technology, 2020, 10(6):1012-1021.]

[5] 环境保护部办公厅.关于印发《生态环境大数据建设总体方案》的通知[EB/OL](2016-03-08)[2021-11-02].http://www.zhb.gov.cn/gkml/hbb/bgt/201603/t20160311_332712.htm.[General Office of the Ministry of Environmental Protection of China.Notice on the Issuance of the Overall Plan for the Construction of Ecological and Environmental Big Data[EB/OL].(2016-03-08)[2021-11-02].http://www.zhb.gov.cn/gkml/hbb/bgt/201603/t20160311_332712.htm.]

[6] 国务院.国务院关于印发土壤污染防治行动计划的通知(国发[2016]31号)[EB/OL].(2016-05-31)[2021-11-02].http://www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm.[The State Council of China.Circular of the State Council on Printing and Distributing the Action Plan for the Prevention and Control of Soil Pollution[EB/OL].(2016-05-31)[2021-11-02].http://www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm.]

[7] 张哲.基于大数据应用的环境数据资源中心设计研究[D].北京:清华大学,2017.[ZHANG Zhe.Research on the Design of Data Center of Environmental Information Based on Applications of Big Data[D].Beijing:Tsinghua University,2017.]

[8] 刘锐,谢涛,卫晋晋.浅谈环境大数据的组织与管理[C]//2016全国环境信息技术与应用交流大会暨中国环境科学学会环境信息化分会年会论文集.北京:[出版者不详],2016:116-120.[LIU Rui,XIE Tao,WEI Jin-jin.Study on Environmental Information Classification.Proceedings of 2016 National Environmental Information Technology and Application Exchange Conference.Beijing:[s. n.],2016:116-120.]

[9] 赵苗苗,赵师成,张丽云,等.大数据在生态环境领域的应用进展与展望[J].应用生态学报,2017,28(5):1727-1734.[ZHAO Miao-miao,ZHAO Shi-cheng,ZHANG Li-yun,*et al*.Applications of Eco-environmental Big Data: Progress and Prospect[J]. Chinese Journal of Applied Ecology,2017,28(5):1727-1734.]

[10] 康佳文,杨培林.内蒙古自治区环境信息化总体规划的战略构想研究[J].环境科学与管理,2015,40(6):178-181.[KANG Jia-wen,YANG Pei-lin.Strategic Conception of Inner Mongolia Autonomous Region Environmental Information Planning[J].Environmental Science and Management,2015,40(6):178-181.]

[11] ZHANG F J.A Review of Development of Environmental Protection Information during the 12th Five-year Plan Period and Its Outlook[J]. Meteorological and Environmental Research, 2017, 8(3):67-70.

[12] 李云婷,严京海,孙峰,等.基于大数据分析与认知技术的空气质量预报预警平台[J].中国环境管理,2017,9(2):31-36.[LI Yun-ting,YAN Jing-hai,SUN Feng,*et al*.Air Quality Forecasting Platform Based on Big Data Analytics & Cognitive Technology[J]. Chinese Journal of Environmental Management, 2017, 9(2):31-36.]

[13] 李乔宇,尚明华,王富军,等.基于Scrapy的农业网络数据爬取[J].山东农业科学,2018,50(1):142-147.[LI Qiao-yu, SHANG Ming-hua,WANG Fu-jun,*et al*.Data Crawling from Agricultural Internet Based on Scrapy[J].Shandong Agricultural Sciences,2018,50(1):142-147.]

[14] 马联帅.基于Scrapy的分布式网络新闻抓取系统设计与实现[D].西安:西安电子科技大学,2015.[MA Lian-shuai.Design and Implementation of Distributed Netnews Crawling System Based on Scrapy[D].Xi'an:Xidian University,2015.]

[15] 黄春梅,王松磊.基于词袋模型和TF-IDF的短文本分类研究[J].软件工程,2020,23(3):1-3.[HUANG Chun-mei,WANG Song-lei.Research on Short Text Classification Based on Bag of Words and TF-IDF[J].Software Engineering,2020,23(3):1-3.]

[16] 王颖洁,朱久祺,汪祖民,等.自然语言处理在情感分析领域应用综述[J].计算机应用,2021.DOI:10.11772/j.issn.1001-9081.2021071262.[WANG Ying-jie,ZHU Jiu-qi,WANG Zu-min,*et al*.Review of Applications of Natural Language Processing in Sentiment Analysis[J].Journal of Computer Applications,2021.DOI:10.11772/j.issn.1001-9081.2021071262.]

[17] 杨晶.基于领域词库的新闻提取技术的研究及应用[D].武汉:湖北大学,2018.[YANG Jing.The Research and Application of News Extraction Technology Based on Domain Lexicon.Wuhan:Hubei University,2018.]

[18] 李枫林,柯佳.基于深度学习的文本表示方法[J].情报科学,

2019,37(1):156-164.[LI Feng-lin,KE Jia.Text Representation Method Based on Deep Learning[J].Information Science,2019,37(1):156-164.]

[19] 谷文静.基于混合神经网络的语言文本分类方法[J].电子设计工程,2021,29(19):44-48.[GU Wen-jing.Language and Text Classification Method Based on Hybrid Neural Network[J].Electronic Design Engineering,2021,29(19):44-48.]

[20] WANG D S,LIU J Z,ZHU A X,*et al*.Automatic Extraction and Structuration of Soil-environment Relationship Information from Soil Survey Reports[J].Journal of Integrative Agriculture,2019,18(2):328-339.

[21] MALLIOS X,VASSALOS V,VENETIS T,*et al*.A Framework for Clustering and Classification of Big Data Using Spark[C]//On the Move to Meaningful Internet Systems:OTM 2016 Conferences.[s.l.]:[s.n.],2016.

[22] 张国华,叶苗,王自然,等.大数据 Hadoop 框架核心技术对比与实现[J].实验室研究与探索,2021,40(2):145-148,176.[ZHANG Guo-hua,YE Miao,WANG Zi-ran,*et al*.Comparison and Implementation of Core Technologies of Big Data Hadoop Framework[J].Research and Exploration in Laboratory,2021,40(2):145-148,176.]

[23] 杨青,张亚文,张琴,等.基于 Hadoop 的多维关联规则挖掘算法研究及应用[J].计算机工程与科学,2019,41(12):2127-2133.[YANG Qing,ZHANG Ya-wen,ZHANG Qin,*et al*.Research and Application of a Multidimensional Association Rules Mining Algorithm Based on Hadoop[J].Computer Engineering & Science,2019,41(12):2127-2133.]

[24] 胡福平,徐美华,沈华明.SVM 的并行计算结构研究及 FPGA 实现[J].微电子学与计算机,2018,35(6):79-83.[HU Fu-ping, XU Mei-hua, SHEN Hua-ming. Research on Parallel Computing Architecture of SVM and Implementation on FPGA[J].Microelectronics & Computer,2018,35(6):79-83.]

[25] JOY R,SHERLY K K.Parallel Frequent Itemset Mining with Spark RDD Framework for Disease Prediction[C]//2016 International Conference on Circuit,Power and Computing Technologies (ICCPCT).March 18-19,2016.Nagercoil,India:IEEE,2016:1-5.

[26] 王义武,杨余旺,于天鹏,等.基于 Spark 平台的 K-means 算法的设计与优化[J].计算机技术与发展,2019,29(3):72-76.[WANG Yi-wu,YANG Yu-wang,YU Tian-peng,*et al*.Design and Optimization of K-means Algorithm Based on Spark Platform[J].Computer Technology and Development,2019,29(3):72-76.]

[27] 张波.基于 Spark 的 K-means 算法的并行化实现与优化[D].武汉:华中科技大学,2015.[ZHANG Bo.The Parallelization and Optimization of K-means Algorithm Based on Spark[D].Wuhan:Huazhong University of Science and Technology,2015.]

[28] 刘建涛.黄河三角洲典型地表类型遥感协同提取方法及生态环境遥感评价研究[D].北京:中国科学院大学(中国科学院遥感与数字地球研究所),2018.[LIU Jian-tao.A Study on Collaborative Extraction Method of Typical Land Surface Types and the Evaluation of Ecological Environment in the Yellow River Delta Using Remote Sensing[D].Beijing:University of Chinese Academy of Sciences (Institute of Remote Sensing and Digital Earth,Chinese Academy of Sciences),2018.]

[29] 王超,安贝贝,刘兰玉.基于 Hadoop-MPP 架构的智慧监测大数据平台[J].信息与电脑(理论版),2021,33(13):124-126.[WANG Chao,AN Bei-bei,LIU Lan-yu.Intelligent Monitoring Big Data Platform Based on Hadoop-MPP Architecture[J].China Computer & Communication,2021,33(13):124-126.]

[30] 黄国鑫,朱守信,王夏晖,等.基于自然语言处理和机器学习的疑似土壤污染企业识别[J].环境工程学报,2020,14(11):3234-3242.[HUANG Guo-xin,ZHU Shou-xin,WANG Xia-hui,*et al*.Natural Language Processing and Machine Learning-based Suspected Soil Contamination Enterprise Identification[J].Chinese Journal of Environmental Engineering, 2020, 14 (11):3234-3242.]

[31] 何清,李宁,罗文娟,等.大数据下的机器学习算法综述[J].模式识别与人工智能,2014,27(4):327-336.[HE Qing,LI Ning,LUO Wen-juan,*et al*.A Survey of Machine Learning Algorithms for Big Data[J].Pattern Recognition and Artificial Intelligence,2014,27(4):327-336.]

[32] 张润,王永滨.机器学习及其算法和发展研究[J].中国传媒大学学报(自然科学版),2016,23(2):10-18,24.[ZHANG Run,WANG Yong-bin.Research on Machine Learning with Algorithm and Development[J].Journal of Communication University of China (Science and Technology),2016,23(2):10-18,24.]

[33] 崔琴芳.基于机器学习的矿区土壤重金属含量遥感估算及监测方法研究[D].西安:长安大学,2020.[CUI Qin-fang.Study on Methods of Estimation and Monitoring Soil Heavy Metal Content in Mining Areas Based on Machine Learning Using Remote Sensing Technology[D].Xi'an:Changan University,2020.]

[34] LARY D J,ALAVI A H,GANDOMI A H,*et al*.Machine Learning in Geosciences and Remote Sensing[J].Geoscience Frontiers,2016,7(1):3-10.

[35] BAI Y,LI Y,ZENG B,*et al*.Hourly $PM_{2.5}$ Concentration Forecast Using Stacked Autoencoder Model with Emphasis on Seasonality[J].Journal of Cleaner Production,2019,224:739-750.

[36] 周永章,王俊,左仁广,等.地质领域机器学习、深度学习及实现语言[J].岩石学报,2018,34(11):3173-3178.[ZHOU Yong-zhang,WANG Jun,ZUO Ren-guang,*et al*.Machine Learning,Deep Learning and Python Language in Field of Geology[J].Acta Petrologica Sinica,2018,34(11):3173-3178.]

[37] 赵博文,王灵矫,郭华.基于泊松分布的加权朴素贝叶斯文本分类算法[J].计算机工程,2020,46(4):91-96.[ZHAO Bo-wen,WANG Ling-jiao,GUO Hua.Weighted Naive Bayes Text Classification Algorithm Based on Poisson Distribution[J].Computer Engineering,2020,46(4):91-96.]

[38] 任加国,龚克,马福俊,等.基于 BP 神经网络的污染场地土壤重金属和 PAHs 含量预测[J].环境科学研究,2021,34(9):2237-2247.[REN Jia-guo,GONG Ke,MA Fu-jun,*et al*.Prediction of Heavy Metal and PAHs Content in Polluted Soil Based on BP Neural Network[J].Research of Environmental Sciences,2021,34(9):2237-2247.]

[39] LIU G,ZHOU X,LI Q,*et al*.Spatial Distribution Prediction of Soil as in a Large-scale Arsenic Slag Contaminated Site Based on an Integrated Model and Multi-source Environmental Data[J].Environmental Pollution,2020,267:115631.

[40] 能昌信,孙晓晨,徐亚,等.基于深度卷积神经网络的场地污染非线性反演方法[J].中国环境科学,2019,39(12):5162-5172.[NEN Chang-xin,SUN Xiao-chen,XU Ya,*et al*.A Site Pollution Nonlinear Inversion Method Based on Deep Convolutional Neural Network[J].China Environmental Science,2019,39(12):5162-5172.]

[41] 王玉玲,王蒙,闫岩,等.基于聚类算法的ERT污染区域识别方法[J].中国环境科学,2019,39(3):1315-1322.[WANG Yu-ling,WANG Meng,YAN Yan,*et al*.An ERT Pollution Area Identification Method Based on Clustering Algorithm[J].China Environmental Science,2019,39(3):1315-1322.]

[42] 郭书海,吴波,张玲妍,等.土壤环境大数据:构建与应用[J].中国科学院院刊,2017,32(2):202-208.[GUO Shu-hai,WU Bo,ZHANG Ling-yan,*et al*.Soil Environmental Big Data:Construction and Application[J].Bulletin of Chinese Academy of Sciences,2017,32(2):202-208.]

[43] 王鑫,于东升,马利霞,等.基于万维网大数据的农药场地土壤污染快速预测方法研究[J/OL].土壤学报,2021.https://kns-cnki-net.webvpn.las.ac.cn/kcms/detail/32.1119.P.20210910.1450.010.html.[WANG Xin,YU Dong-sheng,MA Li-xia,*et al*.Research on the Method of Rapid Prediction of Soil Pollution in Pesticide Polluted-sites Based on Network Big Data[J/OL].Acta Pedologica Sinica,2021.https://kns-cnki-net.webvpn.las.ac.cn/kcms/detail/32.1119.P.20210910.1450.010.html.]

[44] 李强,何连生,王耀锋,等.中国冶炼行业场地土壤污染特征及分布情况[J].生态环境学报,2021,30(3):586-595.[LI Qiang,HE Lian-sheng,WANG Yao-feng,*et al*.The Characteristics and Distribution of Soil Pollution in Smelting Industry Sites in China[J].Ecology and Environmental Sciences,2021,30(3):586-595.]

[45] 郭长庆,迟文峰,匡文慧,等.1990—2020年中国能源开采和加工场地多源数据综合制图与时空变化分析[J].地球信息科学学报,2022,24(1):127-140.[GUO Chang-qing,CHI Wen-feng,KUANG Wen-hui,*et al*.Mapping and Spatio-temporal Changes Analysis of Energy Mining and Producing Sites in China Using Multi-source Data from 1990 to 2020[J].Journal of Geo-information Science,2022,24(1):127-140.]

[46] 张健琳,瞿明凯,陈剑,等.中国西南地区金属矿开采对矿区土壤重金属影响的Meta分析[J].环境科学,2021,42(9):4414-4421.[ZHANG Jian-lin,QU Ming-kai,CHEN Jian,*et al*.Meta-analysis of the Effects of Metal Mining on Soil Heavy Metal Concentrations in Southwest China[J].Environmental Science,2021,42(9):4414-4421.]

[47] YANG Y,YANG X,HE M J,*et al*.Beyond Mere Pollution Source Identification:Determination of Land Covers Emitting Soil Heavy Metals by Combining PCA/APCS,GeoDetector and GIS Analysis[J].CATENA,2020,185:104297.

[48] JIA X L,HU B F,MARCHANT B P,*et al*.A Methodological Framework for Identifying Potential Sources of Soil Heavy Metal Pollution Based on Machine Learning:A Case Study in the Yangtze Delta,China[J].Environmental Pollution,2019,250:601-609.

[49] 潘洪来.制订我国污染场地土壤风险筛选值的几点建议[J].中国资源综合利用,2018,36(9):80-81,85.[PAN Hong-lai.Suggestions on Formulating Soil Risk Screening Values for Contaminated Sites in China[J].China Resources Comprehensive Utilization,2018,36(9):80-81,85.]

[50] 李勖之,姜琼,王国庆,等.不同国家农用地土壤环境标准比较与启示[J/OL].环境科学,2021.https://doi-org-443.webvpn.las.ac.cn/10.13227/j.hjkx.202106203.[LI Xu-zhi,JIANG Rong,WANG Guo-qing,*et al*.A Comparative Study of Soil Environmental Standards for Agricultural Land among Different Countries[J/OL].Environmental Science,2021.https://doi-org-443.webvpn.las.ac.cn/10.13227/j.hjkx.202106203.]

[51] 李天魁,刘毅,谢云峰.关闭搬迁企业地块风险筛查方法评估:基于EPACMTP模型的研究[J].中国环境科学,2018,38(10):3985-3992.[LI Tian-kui,LIU Yi,XIE Yun-feng.Assessment of the Risk Classification Method for Closed Industrial Contaminated Sites:A Study Based on EPACMTP Model[J].China Environmental Science,2018,38(10):3985-3992.]

[52] 陈衍泰,陈国宏,李美娟.综合评价方法分类及研究进展[J].管理科学学报,2004,7(2):69-79.[CHEN Yan-tai,CHEN Guo-hong,LI Mei-juan.Classification & Research Advancement of Comprehensive Evaluation Methods[J].Journal of Management Sciences in China,2004,7(2):69-79.]

[53] 张馨用.基于模糊综合评价的长三角地区大气污染现状研究[J].国土与自然资源研究,2021(6):26-29.[ZHANG Xin-yong.Study on Air Pollution in Yangtze River Delta Based on Fuzzy Comprehensive Evaluation[J].Territory & Natural Resources Study,2021(6):26-29.]

[54] 崔鸿雁,徐帅,张利锋,等.机器学习中的特征选择方法研究及展望[J].北京邮电大学学报,2018,41(1):1-12.[CUI Hong-yan,XU Shuai,ZHANG Li-feng,*et al*.The Key Techniques and Future Vision of Feature Selection in Machine Learning[J].Journal of Beijing University of Posts and Telecommunications,2018,41(1):1-12.]

[55] 彭志江.面向小样本数据的特征分析技术研究[D].成都:电子科技大学,2021.[PENG Zhi-jiang.Research on Feature Analysis Technology for Small Sample Data[D].Chengdu:University of Electronic Science and Technology of China,2021.]

[56] 彭云聪,秦小林,张力戈,等.面向图像分类的小样本学习算法综述[J/OL].计算机科学,2022.https://kns.cnki.net/kcms/detail/50.1075.TP.20220111.1414.014.html.[PENG Yun-cong,QIN Xiao-lin,ZHANG Li-ge,*et al*.Survey on Few-shot Learning Algorithms for Image Classification[J/OL].Computer Science,2022.https://kns.cnki.net/kcms/detail/50.1075.TP.20220111.1414.014.html.]

[57] 胡西范,陈世平.基于机器学习的小样本学习综述[J].智能计算机与应用,2021,11(7):191-195,201.[HU Xi-fan,CHEN Shi-ping.A Survey of Few-shot Learning Based on Machine Learning[J].Intelligent Computer and Applications,2021,11(7):191-195,201.]

[58] BLAGUS R,LUSA L.SMOTE for High-dimensional Class-imbalanced Data[J].BMC Bioinformatics,2013,14:106.

[59] PROVOST F,FAWCETT T.Robust Classification for Imprecise Environments[J].Machine Learning,2001,42(3):203-231.

[60] 金秋.神经网络在小样本数据集的研究及应用[D].成都:电子

科技大学，2020.[JIN Qiu. Research and Application of Neural Network in Few Shot Learning[D]. Chengdu: University of Electronic Science and Technology of China, 2020.]

[61] 向鸿鑫，杨云.不平衡数据挖掘方法综述[J].计算机工程与应用，2019，55(4)：1-16.[XIANG Hong-xin, YANG Yun. Survey on Imbalanced Data Mining Methods[J]. Computer Engineering and Applications, 2019, 55(4): 1-16.]

[62] 王夏晖.我国土壤环境质量监测网络建设的重大战略任务[J].环境保护，2016，44(20)：20-24.[WANG Xia-hui. The Major Strategic Task of the Construction of Soil Environmental Quality Monitoring Network in China[J]. Environmental Protection, 2016, 44(20): 20-24.]

[63] 骆永明，滕应.中国土壤污染与修复科技研究进展和展望[J].土壤学报，2020，57(5)：1137-1142.[LUO Yong-ming, TENG Ying. Research Progresses and Prospects on Soil Pollution and Remediation in China[J]. Acta Pedologica Sinica, 2020, 57(5): 1137-1142.]

[64] 刘文清，杨靖文，桂华侨，等."互联网+"智慧环保生态环境多元感知体系发展研究[J].中国工程科学，2018，20(2)：111-119.[LIU Wen-qing, YANG Jing-wen, GUI Hua-qiao, *et al*. Study on the Development of Multi Perception System for "Internet Plus" Smart Environmental Protection[J]. Engineering Science, 2018, 20(2): 111-119.]

[65] TANAKA H, HIRAKAWA Y, KANEKU S. Recognition of Distorted Patterns Using the Viterbi Algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1982, PAMI-4(1): 18-25.

[66] SHU Y F, CHEN Y G, XIONG C W. Application of Image Recognition Technology Based on Embedded Technology in Environmental Pollution Detection[J]. Microprocessors and Microsystems, 2020, 75: 103061.

[67] CHEN N, SONG Z L. Application of UAV Precise Navigation in Environmental Safety Monitoring and Detection[C]// 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). October 21-24, 2018. Bangkok, Thailand: IEEE, 2018: 31-36.

[68] ZHANG D J, HUANG H, SHENTU Y C, *et al*. An Image Processing Method for Floating Benzene Slick Detection on Water Surface [C]// OCEANS 2017-Aberdeen. June 19-22, 2017. Aberdeen, United Kingdom: IEEE, 2017

[69] ZHANG X, WANG Y C, ZHANG N, *et al*. Spectral-spatial Fractal Residual Convolutional Neural Network with Data Balance Augmentation for Hyperspectral Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10473-10487.

[70] ARROYO E, SELKER T. Attention and Intention Goals Can Mediate Disruption in Human-computer Interaction[J]. Lecture Notes in Computer Science, 2017, 6947: 454-470.

[71] 贺园园，胡小敏，梁腾飞.基于深度机器学习的霾污染监测技术[J].计算机测量与控制，2020，28(8)：18-22.[HE Yuan-yuan, HU Xiao-min, LIANG Teng-fei. Haze Pollution Monitoring Technology Based on Deep Machine Learning[J]. Computer Measurement & Control, 2020, 28(8): 18-22.]

作者简介：陆晓松(1988—)，男，江苏南通人，助理研究员，博士，主要从事土壤质量与生态环境方面的研究。E-mail: luxiaosong2014@163.com

(**责任编辑：**李祥敏)