 DOI: https://doi.org/10.12349/iser.v6i3.5209

The construction of big data of environmental monitoring platform and data analysis method

Jianfeng Huang Bimao Wei Yanlin Tang

Laibin Ecological Environment Monitoring Center, Guangxi Zhuang Autonomous Region, Laibin, Guangxi, 546100, China

Abstract

With the increasing serious problem of global climate change and environmental pollution, environmental monitoring is becoming more and more important. In order to effectively deal with environmental problems, the construction of environmental monitoring big data platform has become a necessary measure. This paper first discusses the basic construction method of environmental monitoring big data platform, and analyzes the key technologies involved in the platform, such as data collection, storage, processing and visual display. Subsequently, this paper focuses on environmental monitoring data analysis methods, including data preprocessing, pattern recognition, anomaly detection and predictive analysis techniques. Through the analysis of practical cases, this paper discusses the application effect of these methods in environmental monitoring, and proposes improvement measures. The research results show that the environmental monitoring big data platform can provide timely and accurate decision support for environmental management departments, so as to effectively promote environmental protection work.

Keywords

environmental monitoring; big data platform; data analysis; predictive analysis; data processing

环境监测大数据平台的构建与数据分析方法研究

黄坚锋 韦必帽 唐燕林

广西壮族自治区来宾生态环境监测中心，中国 · 广西 来宾 546100

摘 要

随着全球气候变化和环境污染问题的日益严重，环境监测变得越来越重要。为有效应对环境问题，环境监测大数据平台的构建成为一个必要的举措。本文首先探讨了环境监测大数据平台的基本构建方法，分析了平台中涉及的关键技术，如数据采集、存储、处理及可视化展示。随后，本文重点讨论了环境监测数据分析方法，包括数据预处理、模式识别、异常检测和预测分析等技术。通过实际案例的分析，本文探讨了这些方法在环境监测中的应用效果，并提出了改进措施。研究结果表明，环境监测大数据平台能够为环境管理部门提供及时、准确的决策支持，从而有效促进环境保护工作。

关键词

环境监测；大数据平台；数据分析；预测分析；数据处理

1 引言

随着社会经济的快速发展和工业化进程的加速，环境污染问题日益严重，尤其是大气、水质和土壤污染等问题，严重影响了人类的生存和发展。为了有效监测和治理这些问题，环境监测的任务变得尤为重要。传统的环境监测方法依赖人工巡查和单一设备采集数据，效率低下，难以应对日益复杂的环境问题。因此，基于大数据技术的环境监测平台成为一种新的解决方案。

大数据技术具有强大的数据采集、存储、处理和分析能力，能够帮助环境监测部门实时获取来自多个渠道的数据，并对这些数据进行深入分析。通过大数据平台，环境监测不仅能够实现数据的高效处理，还能够通过智能分析提供决策支持，进而推动环境保护政策的制定与实施。然而，尽管大数据平台在环境监测中具有巨大潜力，但在实际应用过程中仍面临诸如数据共享难度大、数据质量不高、分析方法不成熟等问题。本文旨在探讨如何构建环境监测大数据平台，并研究适合该平台的数据分析方法。

2 环境监测大数据平台的构建

2.1 平台的基本构成

环境监测大数据平台的构建涉及数据采集、数据存储、数据处理和数据展示四个基本环节。首先，平台需要依托先进的传感器技术和网络通信技术，完成对环境数据的实时采集。环境监测的对象涵盖空气质量、水质、土壤污染等多个

【作者简介】黄坚锋（1978–），男，中国广西岑溪人，本科，工程师，从事土壤与农业化学（环保方面）研究。

方面，数据来源丰富，因此，数据采集系统需要具备多维度、多元化的数据采集能力。其次，数据存储是平台的重要组成部分，环境监测数据量庞大且多样化，因此需要使用分布式存储技术，如 Hadoop、NoSQL 数据库等，确保数据存储的高效性和可扩展性。数据存储的设计不仅需要考虑数据量的激增，还需要支持数据的快速查询和检索。

在数据处理方面，平台需要采用强大的数据处理和分析系统。数据处理过程包括数据清洗、格式转换、去噪声等步骤，为后续的分析提供高质量的基础数据。此外，平台还需要为用户提供数据可视化功能，使得环境监测数据可以直观地呈现给管理人员，为决策提供依据。最后，平台的维护和升级也是构建过程中不可忽视的环节，随着数据量和数据种类的增加，平台需要具备良好的可扩展性和灵活性，支持日后的数据扩展和技术升级。

2.2 数据采集与传输技术

环境监测的首要环节是数据采集，而高效的数据采集离不开先进的传感器技术和通信技术。随着传感器技术的不断发展，各类传感器设备能够精准地监测空气质量、土壤污染、水质变化等环境要素。传感器设备通过物联网技术连接到数据采集系统，实时将数据传输至平台。

为保证采集数据的准确性和实时性，平台通常会采用无线传感器网络（Wireless Sensor Networks, WSN）与远程监控技术结合的方式。WSN 具有分布广泛、响应迅速、实时性高的特点，能够覆盖大面积区域，实时传输各类环境数据。此外，随着 5G 技术的逐渐普及，传感器设备与云端平台的连接速度将大大提升，进一步改善数据的实时传输效果。为了实现高效的数据传输，环境监测平台需要选择适合的数据传输协议，确保数据传输过程中的低延迟和高可靠性。

2.3 数据存储与处理

数据存储是环境监测大数据平台的核心技术之一。在海量环境数据的采集过程中，传统的关系型数据库已无法满足其存储需求，因而需要采用分布式数据库和大数据处理框架。常用的存储系统包括 Hadoop、HBase、Cassandra 等，这些分布式存储技术可以有效处理大规模数据，支持快速地存储和查询操作。

与此同时，数据处理技术也是平台建设的重点。环境监测数据通常包括时间序列数据、多维度数据以及异构数据，如何对这些数据进行处理和分析，成为平台建设中的重要问题。大数据技术，如 MapReduce、Spark 等数据处理框架，能够对海量数据进行并行计算和分析，极大提高了数据处理的效率。

3 环境监测数据分析方法

3.1 数据预处理

数据预处理是环境监测大数据分析的基础环节。由于传感器和监测设备在采集数据时可能受到环境干扰，导致数据噪声、缺失或不一致，预处理过程尤为重要。数据预处理的主要任务包括数据清洗、去噪、补全缺失数据等。常见的数据清洗方法包括重复数据删除、异常值检测、数据规范化等，确保输入数据的质量和准确性。

为了提高数据的可用性和准确性，环境监测平台通常采用先进的数据处理算法，如机器学习中的数据挖掘技术，对数据进行进一步处理。例如，使用聚类分析和分类算法对采集到的环境数据进行分类，发现潜在的环境问题，为后续的分析提供基础。

3.2 模式识别与异常检测

环境监测数据的分析通常需要识别出潜在的模式或趋势，并对异常情况进行监测和预警。通过模式识别技术，可以从大量的环境监测数据中发现数据之间的规律，进而预测未来的环境变化趋势。常见的模式识别方法包括神经网络、支持向量机（SVM）、决策树等，这些方法能够帮助我们更好地理解复杂的环境数据结构和特征。神经网络通过多层结构来提取数据的特征，适用于大规模复杂数据的分析；支持向量机（SVM）利用分类和回归的理论对数据进行模式识别，能够处理高维数据并且具有较好的泛化能力；决策树则通过树状结构对数据进行划分，能够直观地展示数据中的规律，便于环境监测决策的快速做出。

异常检测是环境监测数据分析中的另一项关键技术，它能够及时发现数据中的异常值，为管理部门提供环境风险预警。例如，在空气质量监测中，通过分析 PM2.5、CO2 等污染物的浓度数据，可以检测到污染水平的异常升高。当数据出现突变或偏离常态时，系统能够自动报警，提醒相关部门采取应急措施。为了提高异常检测的精度和效率，可以结合多种方法，如基于统计学的控制图方法、基于机器学习的孤立森林算法等，通过对环境数据的实时监测和分析，进一步提升预警系统的可靠性和响应速度。

3.3 预测分析

预测分析是环境监测平台中的重要数据分析方法。通过历史数据的分析，结合回归分析、时间序列分析等技术，平台能够预测未来的环境变化趋势，为环境管理和政策制定提供科学依据。回归分析是通过建立变量之间的数学模型，研究环境因子（如气温、湿度等）与空气质量、土壤质量、水体污染等环境指标之间的关系，从而进行趋势预测。时间序列分析则主要通过分析数据的时间序列特征，识别其中的周期性变化趋势，预测未来的环境数据波动。

例如，在空气质量预测中，环境监测平台可以基于历史的气象数据和污染物浓度数据，结合回归分析、ARIMA 模型等方法，预测未来几天或几周内的空气质量状况。对于水质监测，平台可以通过历史水质数据分析，结合水流量、降雨量等因素，预测未来水质的变化趋势，进而为水污染防治提供指导。

随着人工智能和机器学习技术的不断发展，预测分析能力得到了大幅提升。深度学习算法、强化学习模型等先进技术在环境监测中的应用，进一步提高了预测的精确度和适应性。深度学习算法通过多层神经网络对数据进行深度特征提取，能够从复杂的大数据中发现隐藏的规律，预测环境变化趋势时具有较高的准确性。而强化学习模型则通过不断地与环境交互，优化决策过程，使得环境监测平台在不断变化的环境条件下仍能有效预测并做出及时响应。

4 环境监测平台应用实例与实践效果

4.1 平台在环境污染监控中的应用

环境监测大数据平台在环境污染监控中的应用效果已经取得了显著的进展。随着传感器技术、物联网和大数据分析技术的不断发展，环境监测平台能够实时收集和分析来自不同监测点的环境数据。以某城市的空气质量监控项目为例，该平台通过传感器网络采集空气中的污染物浓度数据，并将数据上传至云平台进行实时分析处理。平台能够通过大数据分析技术对这些数据进行深入挖掘，识别出污染源，定位污染区域，并根据污染物的浓度变化情况，及时发布预警信息，从而提醒公众和相关管理部门做好应对措施。这种实时、精确的监测不仅有效提高了空气质量监控的精度，还增强了政府对空气污染状况的掌控力，使得污染源的追踪和治理工作得以迅速开展。此外，平台还能够根据分析结果提供污染的趋势预测，协助制定更为科学的污染治理方案，进一步提升环境管理的效率和效果。

4.2 平台在灾害监测与预警中的应用

环境监测大数据平台不仅在常规的环境监控方面发挥重要作用，还在自然灾害监测和预警中有着广泛的应用。平台通过集成多种数据源，如气象数据、地震数据、水文数据等，能够对洪水、地震、台风等自然灾害进行有效监测和预警。通过对这些数据的实时采集、处理与分析，平台能够及时识别潜在的灾害风险，并通过预测模型提供灾害的可能发生时间、地点和影响范围等关键信息，从而帮助政府和相关部门提前部署灾害应急响应措施。

例如，在洪水监测和预警中，平台通过实时监测降雨量、河流水位、土壤湿度等数据，结合历史数据分析，能够预测未来一段时间内洪水发生的风险。通过对比不同气象模型和地理信息，平台能够准确判断洪水发生的地点及其可能影响的范围，及时向相关地区发布洪水预警。进一步，平台还可以根据实时数据的变化，动态调整预警模型，确保预警信息的准确性和时效性。

5 环境监测大数据平台的挑战与发展趋势

5.1 面临的数据质量与隐私保护问题

尽管环境监测大数据平台在数据采集和分析方面取得了显著进展，但仍然面临数据质量、数据隐私保护等问题。传感器设备的采集准确性、数据的标准化和统一性等问题仍需进一步改进。此外，随着大数据技术的发展，如何保护个人隐私，避免信息泄露，成为亟待解决的问题。

5.2 数据共享与合作机制的建设

为了更好地利用环境监测数据，需要建立更加开放和透明的数据共享机制。目前，不同地区和部门之间的环境监测数据仍然存在一定的信息孤岛现象。因此，建立统一的数据标准和共享平台，推动跨部门、跨地区的合作，将是未来环境监测大数据平台发展的重要方向。

6 结语

随着技术的不断发展，环境监测大数据平台将更加智能化和精准化。通过加强数据采集、存储和分析能力，结合先进的预测分析和智能预警技术，环境监测大数据平台将成为环境保护和可持续发展的重要支撑。未来，随着数据共享机制的完善和隐私保护技术的提升，环境监测大数据平台将在推动环境治理、促进生态文明建设方面发挥越来越重要的作用。

参考文献

[1] 姚绪荣,吴岚.辽宁省国土资源信息系统（LLRIS）分析模型研究[J].测绘科技动态,1994,(04):11-18.

[2] 胡侃,夏绍玮.基于大型数据仓库的数据采掘:研究综述[J].软件学报,1998,(01):54-64.

[3] 许祖润,吴世英,刘海涛,等.用单片微机和 PC 微机组成高速大数据量核数据获取和控制系统的方法[J].核电子学与探测技术,1998,(01):30-33+25.

[4] 周傲英,邱越峰,田增平,等.数据库技术:回顾与展望[J].计算机科学,1999,(04):17-21.