文章类型：论文|刊号（ISSN）：2630-4740 /（中图刊号）：715GL012

环境监测中异常数据识别与修复

李文婧　赵泽　田天
白洋淀流域生态环境监测中心
DOI:10.12238/eep.v7i5.2087

[摘　要] 随着民众环保意识的提高,使得环境监测数据的准确性和可靠性显得尤为重要,环境监测中的异常数据存在会误导数据分析甚至影响环境决策的有效性,因此需要不断探索和创新,推动环境监测领域的技术进步和应用发展；同时,需要加强跨学科合作与交流,共同应对环境监测领域的挑战和问题。基于此,本文首先阐述了环境监测中异常数据的定义、识别方法及其影响,然后深入探讨了异常数据的分析方法和整改策略,最后介绍了异常数据的识别和未来的发展趋势。
[关键词] 环境监测；异常数据；识别与修复
中图分类号：X84　文献标识码：A

Identification and repair of anomaly data in environmental monitoring

Wenjing Li　Ze Zhao　Tian Tian
Baiyangdian River Basin Ecological environment Monitoring Center

[Abstract] With the improvement of environmental awareness, the accuracy and reliability of environmental monitoring data are particularly important. The existence of common abnormal data in environmental monitoring can mislead data analysis and even affect the effectiveness of environmental decision, which needs to continuously explore and innovate to promote the technological progress and application development in the field of environmental monitoring. Meanwhile, it needs to strengthen interdisciplinary cooperation and communication to jointly tackle the challenges and problems in the field of environmental monitoring. This paper first expounds the definition, identification method and influence of abnormal data in environmental monitoring, then discusses the analysis method and rectification strategy of abnormal data, and finally introduces the identification of abnormal data and the future development trend.
[Key words] environmental monitoring; abnormal data; identification and repair

引言

环境监测是评价环境质量、防止环境污染的重要手段。在环境监测过程中,异常数据的出现是不可避免的,异常数据可能是由于测量误差、设备故障、环境变化等因素造成的,严重影响了环境监测数据的准确性和可靠性,因此,研究环境监测中异常数据的识别和修复具有重要意义。

1 环境监测异常数据的特征分析

1.1数据来源和类型

环境监测数据的来源非常丰富,包括城乡固定监测站点、移动监测车辆、机载无人机监测设备,以及通过遥感卫星等高科技手段获得的数据,这些站点和设备通常安装在工业区、主要自然保护区等关键环境监测区域,以实现对环境标准的连续或定期监测。在数据类型方面,环境监测数据涵盖了广泛的参数：物理参数数据,如温度、湿度、气压、风速、风向等,提供有关环境基本条件的直接信息；化学参数数据,如颗粒物浓度(PM2.5,PM10)和二氧化硫(SO_2),氮氧化物(NO_x),溶解氧(DO),pH等,揭示了环境中污染物的类型和浓度。此外,生物多样性指数、水生生物数量等生物参数数据,反映了生态系统的健康和生物的响应。不同类型的监测数据在环境监测中发挥着不同的作用,共同构成环境监测数据系统,为环境管理和决策提供科学依据。

1.2异常数据产生原因

产生异常数据的原因多种多样,包括设备故障等物理因素,外部干扰等环境因素,以及人为错误等人为因素。设备故障是产生异常数据的常见原因,因为传感器,分析仪器和其他设备在长时间运行后可能会磨损,老化或损坏,导致数据采集不准确或异常。外部干扰也会产生异常数据,天气变化、自然灾害、人为干扰等环境因素会干扰监测设备,导致异常数据,例如,雷电天气会导致监测设备暂时失灵,导致大量错误数据。此外,人为错误也是产生数据错误的重要原因,在数据收集,处理或分析过程中,人为错误会导致数据异常,例如采样人员不按照规范行事,造成

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

文章类型：论文|刊号（ISSN）：2630-4740 /（中图刊号）：715GL012

样品污染或混淆;数据录入错误数据录入等,这些原因可能导致观测数据的偏差或错误,从而影响环境观测的准确性和可靠性。

1.3异常数据特征

异常数据通常具有一些明显的特征,可以帮助识别和判断异常数据。数值异常是异常数据最直观的特征,异常数据的值通常与其他正常数据有很大差异,可以显示为很高或很低的值,这些异常值可能是由于设备故障,外部干扰或人为错误引起的,例如,在正常情况下,观察点的PM2.5浓度必须在一定范围内波动,如果数据突然出现远高于或远低于正常值,可以判断为异常数据。时间异常也是异常数据的一个重要特征,异常数据可能在时间分布上出现异常,如突然增加或减少,或周期性异常等,这些时间异常可能与环境因素的变化,监测设备故障,或人为操作错误有关,例如,夜间监测点长时间突然上升的噪声数据可能与附近工厂的夜间生产活动有关,因此可以被视为异常。空间异常和相关性是异常数据的重要特征,空间异常表现在某一地区的监测数据普遍偏高或偏低,这可能与该地区的环境条件、监测设备布局或数据传输等因素有关,相关异常表示为异常数据与其他相关参数数据之间存在异常关系,这可能与监测设备故障、数据处理错误或环境因素复杂性等因素有关[1]。

2 环境监测异常数据识别方法

2.1统计方法

统计方法是确定环境监测异常数据的最基本和最常见的一类方法,它主要是根据数据的统计特性,如平均值、标准偏差等来判断数据是否异常,常见的统计方法有阈值法、Z-score方法、箱线图(Box-plot)方法等。空气质量监测中,可以设置PM2.5浓度门槛,如果超过一定值(如75微克/立方米),则认为不正常,当实际监测的数据超过该门槛时,系统会发出警告。对于水质监测中的溶解氧(DO)数据,可以使用Z-score方法来确定异常值,首先计算所有DO数据的平均值和标准偏差,然后对每个数据点计算Z结果(即数据点与平均值之间的距离与标准偏差),当Z-score的绝对值超过一定值(如3)时,可以确定为异常数据。箱线图是一种分布式可视化数据的工具,可以帮助我们快速识别异常值,例如,在土壤重金属监测中,我们可以绘制各种重金属元素浓度的方图；箱图中的“箱”是指数据的平均值,上四分之一,下四分之一,而“需要”是指数据的最大值和最小值(或稍小的极值)。

2.2机器学习算法

随着机器学习技术的发展,越来越多的机器学习算法被应用于识别环境监测的异常数据,这些算法通常根据数据的历史模式进行学习,并试图预测数据的未来趋势,当实际数据与预测数据有很大差异时,可以判断为异常。在噪声监测中,随机森林算法可用于识别异常数据:首先,以历史噪声数据为训练集对随机森林模型进行训练；然后,对于新的噪声数据,使用模型进行预测,计算实际值与预期值之间的偏差:当偏差超过一定的门槛时,可识别为异常数据。在河流水质监测中,支持向量机算法可用于识别异常数据：首先,选取几个主要的水质参数(如DO、pH值、氨氮浓度等)作为显著变量,利用历史数据对SVM模型进行训练:然后,对于新的水质数据,使用模型进行分类预测,如果模型预测结果与实际结果有显著差异,则可视为异常数据。

2.3基于多源数据融合的异常数据识别方法

在实际环境监测中,通常从不同来源和不同类型的数据中获取数据,这些数据之间可能存在多余或互补的关系,基于多源数据整合的异常数据识别方法旨在综合利用这些数据,提高异常数据识别的准确性和可靠性。在空气质量监测中,可以同时从固定监测站,移动监测车辆,卫星和其他来源的遥感获取数据,这些数据可能会在时间和地点上变化和交叉,可以使用数据集成技术(如卡尔曼过滤器,贝叶斯网络等)来合并这些数据并生成综合空气质量指数(AQI),然后可以为AQI分配一个门槛,当AQI在某一地区超过时,可以知道该地区的空气质量数据是否异常。

3 环境监测异常数据修复策略

3.1常见的修复方法

在环境监测中,当发现异常数据时,应采取适当的维修策略来纠正或替换这些异常值。如果由于随机误差或孤立事件而产生异常数据,则可以简单地将异常值替换为数据点所在的平均值或平均序列；例如,在温度监测中,如果某个时间点的温度数据异常高,则可以将该异常值替换为该时间点之前和之后几个时间点的平均或平均温度。对于时间序列数据,可采用加法估计和修正异常值,常见加法包括线性加法、多边界加法、塑性加法等,例如在降水监测中,如果某段时间内的降水数据丢失或异常,则可采用相邻时间段的降水数据来估计该缺失或异常值。在某些情况下,可以依靠该领域专家的经验和知识来判断和修复异常数据,该领域专家可以根据对监测对象的了解和数据分析,确定哪些数据是合理的,哪些是不正常的,并提出适当的修复建议。

3.2基于模型预测的异常数据修复策略

基于模型的异常数据修复策略是利用历史数据和数学模型来预测和修复异常值,这种方法通常需要构建一个准确描述数据生成过程的模型,并使用该模型来预测异常数据的修复。对于具有明显或周期性趋势的环境监测数据,回归模型可用于预测和修正异常值；例如,在空气质量监测中,线性回归模型或非线性回归模型可用于预测PM2.5浓度,并将预测值与实际值进行比较。对于时间序列数据,可以使用时间序列模型(如ARIMA模型,LSTM神经网络等)来预测和修正异常值,这些模型能够捕获数据中的时间和动态相关特性,并据此进行预测和修正,例如在水位监测中,可以使用ARIMA模型来预测未来水位的变化,并将预测值与实际值进行比较[2]。

3.3基于多源数据协同的异常数据修复方法

基于多源数据格式的异常数据修复方法是指综合利用不同来源、不同类型的数据修复异常值,这种方法可以充分利用多源数据的集成和可重复性,提高异常数据修复的准确性和可靠性。通过数据集成技术对来自不同来源、不同类型的数据进行整合,

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

生成更全面、更准确的数据集，然后在此数据集的基础上对异常数据进行识别和修正，例如在空气质量监测中，可以同时从固定监测站、移动监测车辆、卫星遥感等多个来源获取数据，并通过数据融合技术将异常数据整合在一起，然后在此综合数据集的基础上进行识别和修正。利用来自多个来源的数据对同一被监测对象进行验证，确定并修正异常数据，例如，在土壤重金属检测中，可以通过实验室分析和现场快速检测获得土壤中重金属的浓度数据，然后对两种方法获得的数据进行比较和验证，确定并修正异常值，这种方法可以利用各种方法的优势和互补性，提高数据的准确性和可靠性。

4 环境监测中异常数据识别与修复未来趋势

4.1深度学习和异常数据识别

未来，深度学习将在异常数据识别领域发挥更重要的作用，深度学习模型能够处理复杂的非线性关系和高维数据，自动学习大量数据背后的规律和特征，通过训练深度学习模型，我们可以准确地识别和分类异常数据，例如，积累神经网络（CNN）和周期神经网络（RNN）等模型在处理图像和视频数据方面表现优异，可以应用于遥感图像中的异常识别。

4.2大数据和异常数据修复

随着大数据技术的不断发展，环境监测数据将变得更加丰富和复杂，大数据处理技术帮助更有效地处理和分析这些数据，从而更准确地修复异常数据。通过利用大数据平台，可以快速存储、查询和分析大量数据，结合机器学习算法和统计方法修复异常数据。此外，大数据技术可以帮助探索数据中潜在的模式和联系，为环境监测提供更全面、更深入的分析[3]。

4.3物联网和实时监控

随着物联网信息技术不断发展创新，物联网技术发挥的作用愈发强大，物联网技术的发展将为环境监测提供更准确的实时数据支持，通过部署各种传感器和监测设备，可以实时监测环境并收集数据，这些实时数据有助于我们更快地检测异常数据并采取相应的纠正措施，同时，物联网技术还可以实现设备之间的互联和协同工作，提高环境监测的效率和准确性。

5 结语

随着科学技术的快速发展，环境监测领域正在经历前所未有的变化和机遇，从异常数据识别中深度学习的精确应用，到大数据处理中的大数据技术，到实时监测物联网技术提供的无限可能性，以及人工智能技术在决策支持中的创新应用，这些技术正在以前所未有的速度推动环境监测领域向前发展。展望未来，有理由相信环境监测将变得更加智能、高效、准确，不仅可以实时捕捉环境变化，还可以利用先进的分析工具和技术准确地识别和修复异常数据，为环境保护和可持续发展提供有力的数据支持。但是，技术进步也带来了新的挑战，如何确保数据的准确性、安全性和隐私性，如何建立更有效的数据共享与合作机制，如何将环境监测与公共卫生和环境保护更紧密地结合起来，这些都是我们必须思考和解决的问题。

[参考文献]

[1]陆秋琴,魏巍,黄光球.环境监测系统中异常数据的识别和修复方法[J].安全与环境学报,2021,21(03):1300-1310.

[2]景永志,艾自东,田相臣.环境监测中异常数据识别与修复[J].环境工程技术学报,2024,14(03):1098-1104.

[3]李信茹,周民,米屹东,等.智慧环保体系在环境治理中的应用[J].环境工程技术学报,2021,11(05):992-1003.

作者简介：

李文婧(1988--),女,汉族,河北省衡水市人,本科,工程师,研究方向：环境监测。

通讯作者：

田天(1992--),男,汉族,河北省保定市人,本科,工程师,研究方向：环境监测。

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.