

基于人工智能的环境监测数据自动分析与预测技术探讨

谢颂辉*, 赵冉, 牛忠林, 曹丽萍, 齐西仓

(山东天衡检测有限公司, 菏泽 274000)

摘要: 随着环境污染问题的加剧, 环境监测在生态保护与可持续发展中扮演着至关重要的角色。但传统环境监测手段在数据采集、处理与预测方面存在局限性, 如依赖人工操作, 难以实时处理大规模数据以及无法有效应对数据的高复杂性问题。因此, 引入基于人工智能的环境监测数据自动分析与预测技术显得尤为重要。本文围绕人工智能技术在环境监测中的应用展开研究, 重点探讨了数据清洗、异常值处理和多源异构数据的集成与标准化处理等关键技术, 以提升环境数据的质量与分析基础。通过深度学习模型(如 LSTM), 研究了人工智能在空气质量预测、水质变化分析等环境监测典型任务中的优越表现, 并结合案例进行了定量评估, 验证其实用性。为环境监测领域的智能化发展提供了新的思路与技术支持。

关键词: 环境监测; 人工智能; 数据清洗

0 引言

如今公众对环境质量的关注度日益提升, 对环境监测数据的准确性和时效性要求也越来越高。传统的环境监测方法, 虽然在一定程度上能够反映环境质量状况, 但受限于人力、物力和技术手段, 往往难以满足大规模和高精度等环境监测需求。鉴于此背景, 本研究致力于探讨如何运用人工智能技术提高环境监测数据处理的自动化水平和精确性, 以弥补传统方法的不足。期望本文能够为环境监测领域的智能化进步提供新的视角。

1 环境数据分析与预处理

1.1 环境数据的质量控制与清洗

在环境监测数据的分析过程中, 数据质量的控制是确保人工智能模型能够高效运作的关键环节^[1]。环境数据来源于多种传感器、监测站及第三方平台, 这些数据通常具有较高的复杂性和多样性, 包括不同的时间分辨率、空间分辨率以及测量误差等问题。异常值和缺失值是环境数据中最常见的质量问题, 如未得到有效处理, 不仅会降低模型的预测精度, 还可能导致分析结果的显著偏差^[2-3]。研究如何高效地进行数据清洗, 确保数据的高质量 and 一致性, 是提高人工智能模型预测准确性和可靠性的核心任务^[4]。

传统的统计学方法如四分位距法(interquartile range, IQR), 通过计算数据分布的四分位差距(即中间 50% 数据的范围), 来识别远离正常范围的异常点^[5-6]。它操作简单且对偏态分布的数据较为敏感, 适合于分布相对均匀的小规模数据集。而 Z-score(标准分数)法则是一种基于正态分布假设的异常值检测方法, 通过评估数据点与总体均值的偏离程度, 将偏离超过一定阈值(如 3 倍标准差)的数据点标记为异常值。其公式如下:

$$Z = \frac{x - \mu}{\sigma}$$

其中 x 为数据点, μ 为数据的均值, σ 为数据的标准差。当 $|Z| > 3$ (常用阈值为 3) 时, 通常该数据点被认为是异常值。由于其广泛适用性和对偏差的高度敏感性, 这种方法特别适合处理大规模数据集, 且在环境监测的多种场合中得到了广泛应用。但在数据维度较高或分布极为复杂的情况下, 传统统计方法可能无法全面捕捉异常点的所有特征^[7-8]。在这种情况下, 基于机器学习的异常值检测技术显示出更强大的适应性, 如孤立森林(isolation forest)等基于机器学习的异常值检测技术, 能有效识别并剔除高维数据中的异常点。孤立森林是一种近年来被广泛采用的机器学习技术, 通过构建多棵随机树, 能够更早期地将异常点“孤立”出来, 从而高效地识别出离群点^[9]。它能适应高维复杂数据, 并在多源异构的环境监测数据中展现出优异

* 通信作者: 谢颂辉, 高级工程师, 研究方向为检验检测、环境监测、环境工程。E-mail: thjc01@163.com

的性能, 特别适合非线性、动态变化的环境数据清洗, 已逐渐成为环境监测数据处理中不可或缺的工具。

IQR、Z-score 和孤立森林三种常用方法的检测结果具有显著差异。Z-score 方法检测出的异常值数量最多(20个), 表明其对离群点具有极高的敏感性。IQR 方法检测出的异常值数量适中(15个), 适合于数据分布规则的场景^[10-11]。孤立森林方法则检测出的异常值数量最少(10个), 但其准确性较高, 尤其在处理高维数据时表现优异。这证明了孤立森林在处理复杂环境监测数据时的有效性。图1直观地展示了三种方法在异常值检测上的差异, 为选择合适的清洗工具提供了重要参考。这表明, 不同检测方法的选择应根据数据特性和清洗需求灵活调整。当需要快速筛选大规模数据中的潜在异常值时, 可优先考虑 Z-score 法, 而在分析复杂多维环境数据时, 孤立森林因其鲁棒性强, 更适合作为主要工具^[12]。缺失值的修复同样是数据清洗中的重要环节。在环境监测中, 由于设备故障、数据传输中断或外部环境干扰, 缺失值问题十分常见。缺失值不仅会破坏数据的完整性, 还可能显著降低模型的训练效果和预测能力。针对不同场景, 研究者通常采用多种方法来填补缺失值, 例如简单插值法(均值或中位数插值)、回归预测法以及基于机器学习的插补方法^[13]。实际应用中, 通过对异常值检测和缺失值修复的系统性清洗, 可以显著提升环境监测数据的质量, 为后续的人工智能模型构建提供坚实的基础。

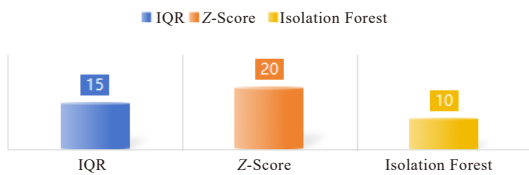


图1 异常值检测方法对比

1.2 环境数据集成与标准化处理

在环境监测领域, 数据的来源多样且格式各异, 这给数据分析和模型训练带来了不小的挑战。环境数据集成在工作中将来自不同监测站点、传感器和第三方平台的数据进行整合, 形成一个统一的数据集。这些传感器可能使用不同的测量原理和标准, 导致数据格式、单位、采集频率等方面的差异。这种异构性在数据集成过程中提出了挑战, 因此如何有效地对这些数据进行集成和标准化处理, 成为环境数据分析中的一个关键问题^[14]。由于环境数据通常来自不同类型的传感器、监测平台和时间序列, 其数据格式和结构可能高度异构。因此数据集成的目标是通过规范化处理, 使得这些异构数据能够在同一平台上进行统一的分析与建模。在数据集成的过程中, 首先需要

对不同数据源进行统一的时间和空间对齐, 通过插值算法(如线性插值、样条插值等)对不同时间点的观测值进行平滑处理, 使得不同来源的数据能够对应到同一时间序列上^[15]。

在完成时间和空间的对齐后, 数据的标准化处理成为集成过程中的重要步骤。数据标准化(data standardization)是指将数据转换为统一的尺度, 以消除不同测量单位和量纲对模型性能的影响。环境数据中, 不同传感器可能使用不同的测量单位, 如温度的单位可能为摄氏度或华氏度, 湿度的单位可能为百分比或帕斯卡等^[16-17]。这些不同的单位会导致数据在分析和模型训练过程中出现偏差。因此标准化处理的核心任务是这些不同单位的数据转换为一个统一的尺度, 以便进行后续的分析 and 建模。常见的数据标准化方法包括 Z-score 标准化、Min-Max 标准化和小数定标标准化等。Z-score 标准化通过将数据转换为均值为 0, 标准差为 1 的正态分布, 消除了不同量纲的影响; Min-Max 标准化将数据缩放到一个指定的范围(如 0 到 1), 使得所有数据都处于同一尺度下; 小数定标标准化则通过移动数据的小数点位置, 将数据缩放到一个统一的尺度。在环境数据集成与标准化处理中, 需要根据数据的实际情况选择合适的标准化方法, 以确保数据的一致性和可比性。这些标准化方法可以有效减少由于特征尺度差异而带来的偏差, 使得各个特征在模型训练过程中能够平等地发挥作用^[18]。

2 基于人工智能的预测模型构建应用

2.1 深度学习模型在环境监测中的构建应用

深度学习作为人工智能的一个重要分支, 在许多领域取得了显著成果, 尤其是在处理复杂的时序数据方面。环境监测数据通常具有较强的时序性和非线性特征, 长短期记忆网络(long short-term memory, LSTM)在这方面展现了极大的潜力。与传统的机器学习方法相比, LSTM 能够有效捕捉数据中的长期依赖关系, 因此在环境数据的建模与预测中具有重要的应用价值^[19]。LSTM 是一种特殊的循环神经网络(RNN), 其独特的门控机制使得它能够在长时间跨度内保持信息的流动, 并避免传统 RNN 模型在处理长序列时出现的梯度消失或爆炸问题。

在实践中随着训练轮次的增加, 损失值逐渐下降, 从初始值 0.5 下降到 0.2, 呈现出明显的收敛趋势。这表明 LSTM 模型的优化过程有效, 模型的学习能力得到验证。通过分析损失曲线, 可以判断模型在当前轮次下的训练状态是否良好, 以及是否存在过拟合或欠拟合问题。由于损失值的下降速度在后期趋于平缓, 表明模型已接近收敛, 后续训练可以适当减少轮次以节省计算资源。在环境

监测中, 气温、湿度、气压等气象变量常常随着时间发生变化, 且相互之间存在复杂的动态关联。LSTM 能够通过“记忆细胞”保存关键的历史信息, 并通过遗忘门屏蔽不重要的输入, 从而在预测任务中保留对未来有用的时序依赖。LSTM 的这种能力使得它在处理具有长期依赖关系的环境数据时, 比传统的时序预测模型(如 ARIMA、GARCH 等)具有更高的准确性和鲁棒性。

在 LSTM 模型的构建中, LSTM 模型不仅能够处理单一变量的时间序列预测, 还能够同时处理多个相关变量的联合预测^[20]。环境监测数据通常会受到噪声、缺失值等因素的影响, 因此数据清洗和标准化是构建高效模型的基础。通过对数据进行去噪、填补缺失值和标准化处理, 可以确保输入数据的质量, 从而为模型的训练打下坚实的基础。且 LSTM 的训练过程中, 选择适当的优化算法(如 Adam 优化算法)是至关重要的, 能够有效提高模型的收敛速度, 减少训练时间。为了防止过拟合, 训练中常使用正则化方法如 L2 正则化和 Dropout, 确保模型的泛化能力。

2.2 LSTM 模型在环境监测中的典型应用场景

在环境监测领域, LSTM 模型凭借其处理时序数据的优越性能, 在多种具体应用场景中展现了强大的能力。第一, 在气象监测和预测方面, LSTM 通过对气温、湿度、气压和降水量等多维变量的动态建模, 能够生成精准的短期和中长期气象预测结果。这种预测不仅能够为农业、水资源管理和能源调度提供参考, 还可用于极端气象事件(如暴雨和干旱)的早期预警, 提高灾害应对的效率和精准度。第二, 在空气质量监测和预警方面, LSTM 通过结合历史空气质量指数(AQI)数据与相关气象变量(如风速、风向和温度), 可以精准预测未来空气污染程度及其空间分布。这种预测能力对于提前识别重污染天气的潜在风险具有重要意义, 能够为环境管理部门提供决策依据, 从而采取有效措施降低污染物浓度, 并减少对公众健康的影响。第三, 在水质监测与污染控制方面, LSTM 模型可以分析并预测水质参数(如溶解氧、pH 值和化学需氧量)的变化趋势, 帮助识别潜在的污染源并跟踪污染物扩散路径。

3 结束语

本文通过对人工智能技术在环境监测数据处理与预测中的研究与应用分析, 揭示了其在提升数据质量、预测准确性和实时响应能力方面的巨大潜力。以 LSTM 模型为代表的深度学习算法, 凭借其对时序数据的强大建模能力, 在气象预测、空气质量监测及灾害预警等任务展现了广泛的应用前

景。相信未来通过研究人员持续的技术创新与实践探索, 为环境保护和可持续发展提供强有力的技术支撑。

参考文献

- [1] 苑振杰, 王芳, 闫世博. 人工智能技术在大气环境监测中的应用[J]. 中国战略新兴产业, 2025, (3): 77-79.
- [2] 刘佩佩, 朱邱晗, 辛瀚. 人工智能赋能体育馆实时监测与智控优化设计[J]. 物联网技术, 2025, 15(2): 60-64.
- [3] 高松, 焦正, 吴克食, 等. 恶臭气体监测技术研究进展[J]. 自然杂志, 2025, 47: 1-10.
- [4] 宫轲楠, 胡泊, 徐玄, 等. 人工智能在计量学中的应用进展与展望[J]. 宇航计测技术, 2024, 44(6): 28-41.
- [5] 李峻涛, 张小英. 单片机在智能家居环境监测系统中的应用与优化[J]. 数字通信世界, 2024, (12): 166-168.
- [6] 岳明洋, 韩东爽. 人工智能在海南地区生态监测中的应用研究[J]. 造纸装备及材料, 2024, 53(12): 123-125.
- [7] 罗伟锦, 邓昌忠, 许斌锋, 等. 顶管基坑施工监测时间序列联动规律研究[J]. 土工基础, 2024, 38(6): 1044-1049.
- [8] 周卓涵, 范从勇, 李静. 基于无人机遥感技术的环境监测研究[J]. 黑龙江环境通报, 2025, 38(2): 79-81.
- [9] 方永睿. 人工智能技术在环境监测中的应用[J]. 清洗世界, 2024, 40(11): 181-183.
- [10] 陈镇茂. 新污染物共排放对生态环境监测和管理的影响分析[J]. 皮革制作与环保科技, 2024, 5(21): 39-41.
- [11] 王爱娟. 新形势下生态环境监测质量问题与对策研究[J]. 皮革制作与环保科技, 2024, 5(21): 65-67.
- [12] 郑林, 蔡其晓. 新形势下生态环境监测质量问题与对策研究[J]. 皮革制作与环保科技, 2024, 5(21): 87-89.
- [13] 罗焯. 环境监测数据的数值修约规则探讨[J]. 中国环境监测, 2024, 40(1): 63-67.
- [14] 覃琼霞, 王维. 基于生态环境治理现代化探讨生态环境监测的数字化转型实践[N]. 山西科技报, 2024-11-07(B6).
- [15] 张中桥. 固相微萃取法在环境监测中的应用[J]. 黑龙江环境通报, 2025, 38(2): 142-144.
- [16] 胡敬平, 陈然, 侯慧杰, 等. “双碳”战略下融合人工智能的环境学科人才培养模式探索[J]. 高等工程教育研究, 2024, (6): 85-88.
- [17] 尤卫军, 曹彦嵘, 蒋晓彤, 等. 应用数据治理的生态环境监测系统研究[J]. 浙江工业大学学报, 2024, 52(6): 633-640.
- [18] 张仲华, 袁金良. 完善流域生态环境监测工作的思考[J]. 黑龙江环境通报, 2024, 37(12): 74-76.
- [19] 徐丹, 宫振. 从地理环境和设备仪器分析建设长白山 GNSS 监测站的方案[J]. 中国设备工程, 2024, (23): 186-188.
- [20] 严鹏. 生态环境监测体系建设研究[J]. 中国环保产业, 2025, (1): 65-67.