

Application of Intelligent Algorithms in Data Analysis of Environmental Engineering Laboratory

Shuang Chen Tangpu Ma Jie Wang Mengting Wang

Hubei Liangqing Agricultural Technology Co., Ltd., Wuhan, Hubei, 430000, China

Abstract

In the stage of conducting environmental engineering experiments, the data scale is large, the structure is complex, and the fluctuation degree is severe. In terms of accuracy and efficiency, traditional analysis methods are no longer suitable for practical needs. With the advancement of artificial intelligence technology, intelligent algorithms have gradually played a key role in the analysis of environmental experimental data by leveraging their outstanding non-linear modeling, adaptive learning, and efficient computing capabilities. Through preprocessing, modeling, prediction, and optimization of experimental data, intelligent algorithms not only improve the quality of data analysis, but also promote the visualization and automation of experimental results. In terms of multi-source data integration, trend recognition, and pollutant diffusion prediction, their performance is very significant. With the continuous improvement of algorithm theory and hardware technology innovation, the application of intelligent algorithms in environmental engineering experiments will become increasingly common, providing higher levels of support for improving experimental efficiency and result reliability.

Keywords

intelligent algorithm; Data processing; Environmental engineering; experimental analysis

智能算法在环境工程实验室数据分析中的应用

陈爽 马堂普 王洁 王梦婷

湖北良顷农业科技有限公司, 中国·湖北 武汉 430000

摘要

在环境工程实验开展阶段, 数据规模庞大、结构错综复杂、波动程度剧烈, 就准确性和效率而言, 传统分析方法已难适应实际需求, 伴随人工智能技术的进步, 智能算法借助其突出的非线性建模、自适应学习与高效计算能力, 渐渐在环境实验数据的分析中发挥关键作用。经过对实验数据进行预处理、建模、预测与优化操作, 智能算法既提升了数据分析的质量, 也推动实验结果朝着可视化与自动化迈进, 在多源数据整合、趋势识别以及污染物扩散预测等方面, 其表现十分显著, 伴随算法理论持续完善以及硬件技术的革新, 智能算法在环境工程实验中的应用将愈加普遍, 为实验效率提升和结果可靠性提供更高水准的支持。

关键词

智能算法; 数据处理; 环境工程; 实验分析

1 引言

环境工程实验涉及数量众多、复杂又多源的数据, 面对高维、非线性以及动态的数据特征, 传统的数据处理方法往往存在一定局限, 智能算法凭借其出色的泛化能力与自适应特质, 渐渐被引入环境实验数据的分析进程里, 在实验数据预处理、模型构建、结果预测等诸多阶段, 智能算法皆表现出强大的辅助价值, 为环境工程领域的研究提供了新颖的技术路径及方法支撑。

2 智能算法与环境实验数据的契合性

2.1 高维数据处理优势显著

智能算法在高维数据处理中优势十分明显, 尤其以主成分分析、线性判别分析、自编码器而降维工具为代表, 能于不损失关键特征信息的情形下压缩数据的维度, 剔除那些无关变量与噪声, 提高后续建模的效率水平, 递归特征消除、信息增益等特征选择算法, 可依照数据重要性对变量进行排序, 帮助实验人员更高效地提取有效信息, 提升实验参数的设置水平, 进而提升整体分析的质量与科学决策的水平。

2.2 非线性关系建模能力强

环境实验的各类变量间, 往往不符合线性相关的假设要求, 以大气污染预测为例, 除了风速、温湿度、排放量等直接因素外, PM2.5 浓度也受其作用, 也受地形、气压、

【作者简介】陈爽(1996-), 女, 中国吉林四平人, 从事环境工程研究。

光照强度等间接变量的调控,构建起高度非线性、多重耦合的复杂架构。传统回归方法往往难以对这些关系进行有效建模,导致预测的偏差幅度较大,神经网络、支持向量机、随机森林等智能算法方法具有良好的非线性建模能力,能通过多层结构或者核函数技术描绘变量间的非线性映射关系,BP神经网络通过多层感知机制,认知输入变量与输出指标间的复杂关联,支持向量回归在小样本情形下仍能维持出色的泛化能力,依靠智能算法进行建模,可切实增强实验数据拟合效果与预测的精确性,给环境监测和污染控制提供更可靠的科学佐证。

2.3 动态变化捕捉更加灵敏

环境实验数据往往展现出显著的时序及动态特征,数据于时间维度展现出连续性与周期性,诸如水质参数昼夜的起伏波动、大气污染浓度季节的动态变化等,传统静态分析方法很难精准捕捉此动态演化过程,智能算法所采用的融合时间序列建模能力结构,诸如循环神经网络、长短期记忆网络、门控循环单元等类型,专门为处理时序数据量身打造,有能力学习历史数据的依赖关联,预测未来趋势的起伏变化。在连贯的水质监测实验期间,利用既往的监测数据,LSTM模型能预测未来某时刻的溶解氧或氨氮浓度,实现预警效用,基于Kalman滤波和贝叶斯网络的动态模型正逐步在实验控制和调度系统里得到应用,提高响应时效同时提升处理精密度,依靠智能算法所拥有的动态建模能力,实验室得以实时掌控环境变化态势,实现对异常波动快速地响应及分析,极大增强实验的实用价值与前瞻意义^[1]。

3 常用智能算法在实验数据分析中的应用

3.1 神经网络在浓度预测中的作用

在环境工程实验的操作里,水体污染物浓度的预测属于常见且关键的任务范畴,多种因素影响到化学需氧量、生化需氧量、总氮、总磷等污染指标的变动,关联复杂,波动频繁出现,对于这类非线性、多变量的数据处理,传统线性或回归模型的精度欠佳,无法满足实际预测要求。人工神经网络,尤其像前馈型和深度这两类神经网络,由于具备自学习及自适应能力,在该领域得到了普遍的应用,通过对历史监测数据反复训练,神经网络可以识别输入参数跟污染物浓度之间潜在的非线性规律,实现对未来时段的精确预判,在河流水质监测相关实验中,采用水温、pH值、流速和前期污染浓度作为输入变量,神经网络可输出精度高的COD预测数值,逐步把卷积神经网络引入到水体图像识别与污染监测方面,进一步扩大了其于环境实验里的应用范畴。

3.2 支持向量回归在趋势判断中的应用

环境实验数据一般都伴随大量噪声和不规则起伏,准确判断参数变化趋势,成为实验分析关键一环,支持向量回归是基于统计学习理论的机器学习方法,尤其在处理小样本、高维度、非线性特征明显的的数据上适用,其核心的突出

优势是引入核函数构建高维映射,促使数据在高维空间达到线性可分的情形,从而构造出稳定的预测模型。在环境数据趋势分析中,SVR可识别污染物浓度随时间、季节或人为干预出现的演化趋势,在开展温室气体排放监测的实验里,SVR模型有能力精准捕捉CO₂或CH₄浓度的上下走势,为政策制定及控制措施的评估提供数据依托,跟传统回归方法对比,SVR对异常值跟干扰项的敏感度偏低,呈现出更突出的稳健性与泛化能力,在实际应用里呈现出良好的可解释及一致特性^[2]。

3.3 聚类算法在样本分类中的价值

环境工程实验中经常会碰到大量来源不同、类型多样的实验样本,像来自不同水体区域的采样数据、不同排放源的废气样本等,面对尚无明确分类标签的实验数据时,聚类算法属于无监督学习方法的范畴,可自动探寻数据的内在架构,识别出相似的样本,完成有效的分组跟归类。常见聚类方法有K-means算法、层次聚类与密度聚类等,进行水质分析工作之际,借助K-means算法,可根据各类污染因子浓度对不同水样自动归类,从而识别出污染的类型与源头,在空气质量实验的操作阶段,将不同采样点的PM_{2.5}、SO₂、NO₂等数据进行聚类,可以找出污染物浓度分布的空间差异及规律,为后续开展分析和区域治理提供借鉴依据。聚类结果同样可辅助开展模型构建,比如结合监督学习对不同类别的样本分别进行预测模型训练,由此改善整体分析的精度与工作效率,引入聚类算法,不仅增强了实验数据的有序性与逻辑性,也给实验结果的深度挖掘提供了多维度的支持。

4 智能算法在实验分析流程中的角色

4.1 实验前数据清洗与预处理

环境工程实验产出的大量原始数据,一般存在着数据缺失、数值异常以及格式不一致等问题,直接削弱后续数据分析的精准度与可靠性,数据清洗及预处理过程里,智能算法发挥着重大作用,在缺失值填补这一方面,基于决策树的算法可分析数据相互间的内在关系,合理估算并填补缺失之处,避免简单插值引起的偏差情形;以无监督深度学习模型形式存在的自编码器,可自动挖掘数据的潜在特征,把异常数据从正常数据中分离出来,以此实现对异常值的平滑处置。自编码器也适用于数据降噪方面,提高信号与噪声的比例,面对格式及单位相异的数据,智能算法可自动进行识别与转换工作,让数据实现统一,后续模型训练因数据预处理阶段质量的提升有了坚实基础,切实增强整体实验分析的精准度与可靠性^[3]。

4.2 实验中实时监测与反馈机制构建

伴随环境监测技术的发展,实验数据的采集设备持续走向智能化,实时监测成为常态,智能算法进行嵌入式应用,尤其是涉及到卷积神经网络与深度学习模型,对实时数据处

理及反馈起到了关键性作用,在水质监测相关实验里,部署基于 CNN 的图像识别模型,可以实时识别水样中藻类的种类及数量,为水体生态状况评估实时提供数据方面的支撑。环境传感器采集到的多维度数据,经过智能算法进行处理后,可迅速识别异常变动与潜在隐患,实现自动示警与反馈纠偏,该机制不但提高了实验响应速度水平,还大幅减少了人工分析所面临的负担,提升实验自动化水平及精细化管理能力,构建实时反馈机制,为环境实验的动态调控及决策筑牢了技术根基。

4.3 实验后分析结果的可视化与优化

实验告一段落,智能算法在结果分析与优化阶段也起着关键作用,环境工程实验的相关数据往往维度繁杂,若直接展示,很难清晰呈现其内在规律,借助像主成分分析、t-分布随机邻域嵌入这样的降维与可视化技术,可把高维数据投射至二维或三维空间,实现数据结构的直观呈现效果,为研究人员快速识别样本聚类、异常点及趋势特点提供助力。广泛应用遗传算法、粒子群优化等智能优化算法来调整与优化实验参数,通过模拟进化流程或群体协同的搜索方法,在多目标、多约束的条件下,这些算法可找到最优实验方案,提高实验开展效率与结果的精准度,智能算法跟可视化分析工具结合,为实验报告及科研交流工作提供丰富的数据依托,提升了环境工程实验的科学价值及实用方面意义。

5 智能算法应用中的挑战与发展趋势

5.1 数据质量对模型性能的影响不容忽视

输入数据质量在很大程度上对智能算法效果起着关键作用,环境工程实验数据来源呈现出多样性,涉及传感器监测获取的数据、实验室检测得到的结果以及遥感影像等,模型预测和分析能力的发挥,直接与数据的准确性和完整性相关,传感器精度方面的限制、采样频率存在的不一致以及数据传输过程里的丢包或误差,皆会让数据质量有所降低。不同实验设备与方法的标准存在不统一现象,数据格式及单位的差别较为明显,进一步增加了数据预处理的难度,没有统一的数据标准及规范,时常导致模型训练阶段的异常波动,影响模型的稳定性和泛化性能发挥,数据标准处理、实时校准与异常数据自动检测及修正,成为环境工程领域智能算法应用必须重点关注的难题,采用物联网技术和大数据平台结合模式,实现连续、高精度、标准化的数据采集及管理,会极大推进智能算法的稳定执行和精准实施。

5.2 模型泛化能力有待进一步提升

在环境工程相关实验里,数据往往展现出区域的差异

和时间的动态性,不同实验条件加上不同环境背景,数据特征差异明显,致使部分智能算法在特定数据集上表现极为出色,不过迁移至其他实验场景时性能急剧下降,即出现过拟合现象,模型的泛化能力欠佳,对智能算法实现广泛应用造成阻碍。应对这一难题挑战,跨域学习、迁移学习成为当下研究的热门方向,跨域学习通过借助源域现有的数据知识,有效增强目标域模型的适应水平,降低对海量标注数据的倚赖,迁移学习通过知识迁移与参数微调,帮助模型迅速适配新的数据分布与环境的特性,联邦学习的兴起为既保护数据隐私又提升模型泛化提供了新方案,加强模型结构设计创新力度、引入正则化策略、增加训练数据的多样性,均能提升智能算法在环境工程实验中的适应性与鲁棒程度。

5.3 多算法集成推动分析精度提升

由于环境工程实验的数据有着复杂性高、变量繁杂且影响因素相互交错的特点,单一智能算法往往无法满足多维度、多目标分析的需要,作为融合多个基础模型优势的技术,通过加权投票、堆叠与融合等相关策略,获得更稳固精准的预测及分类结果。集成方法诸如随机森林、梯度提升树,在实验数据分类和污染物浓度预测里成效突出,深度强化学习跟环境反馈机制相联合,可实现实验参数的动态优化与智能抉择,助力实验进程实现自动化与智能化升级,多算法的集成显著提升了模型的综合性能,也提升了系统的容错及稳定特性,避免了单一模型受噪声及异常数据影响的缺陷,当计算能力得以提升且算法理论渐趋成熟,多算法融合会变成智能算法在环境工程实验领域实现高精度分析与智能控制的重要路径。

6 结语

智能算法为环境工程实验室的数据分析提供了全新的解决之道,不光提升了数据处理的效率及准确性,同样拓宽了实验设计及分析的技术操作手段,随着算法理论与计算平台进一步地发展,对于环境科学研究而言,智能算法将发挥更核心的功效,为环境保护及可持续发展提供强有力的数据后盾。

参考文献

- [1] 田江,高炽扬, and 李亚伟. “基于智能算法的测试数据自动生成模型研究.” 信息安全与技术 (2010).
- [2] 杨恺, and 高效伟. “高超声速气动热环境工程算法.” 导弹与航天运载技术 4(2010):5.
- [3] 任伯帆,龙腾锐, and 王昆平. “复杂非线性函数最优化问题的一种实用智能算法.” 工程数学学报 22.3(2005):7.