

基于机器学习耦合作用的水环境质量分析

郑伟¹ 郭文卿² 隋静² 孙波³ 吴秋兰^{3*}

1. 莱芜职业技术学院, 中国·山东 济南 271100

2. 泰山职业技术学院, 中国·山东 泰安 271000

3. 山东农业大学, 中国·山东 泰安 271000

摘要: 在水环境质量分析碰到传统方法难以处理非线性耦合关系、高维异构数据整合不够及动态响应滞后等问题的情形下, 现有分析模式无法契合需求, 为应对水质状态动态变化复杂化、风险预判时效性要求提高及监测数据爆发式增多的问题, 本文首先完整讲解机器学习的技术特征与基本原理, 明确其在非线性建模、海量数据处理及自适应学习方面的显著长处; 其次着重钻研机器学习在水环境质量分析中的实际作用, 并给出相应的水环境质量分析办法, 达成了特征选择提升分析模型精度、多源数据融合带动水质预测、时空耦合剖析水质演变规律以及智能模型动态监控水质风险等策略的整合运用, 为相关人员提供了实际操作的借鉴。

关键词: 机器学习; 耦合作用; 水环境质量

Analysis of Water Environment Quality Based on the Coupling Effect of Machine Learning

Zheng Wei¹, Guo Wenqing², Sui Jing², Sun Bo³, Wu Qiulan^{3*}

1. Laiwu Vocational and Technical College, China Shandong Jinan 271100

2. Taishan Polytechnic, China Shandong Tai'an 271000

3. Shandong Agricultural University, China Shandong Tai'an 271000

Abstract: In the context of water environment quality analysis encountering problems such as the difficulty of traditional methods in handling nonlinear coupling relationships, insufficient integration of high-dimensional heterogeneous data, and lagging dynamic responses, the existing analysis models fail to meet the demands. To address the issues of complex dynamic changes in water quality status, increased requirements for the timeliness of risk prediction, and the explosive growth of monitoring data, this paper first comprehensively explains the technical features and basic principles of machine learning, highlighting its significant advantages in nonlinear modeling, processing of massive data, and adaptive learning. Secondly, it focuses on exploring the practical application of machine learning in water environment quality analysis and presents corresponding analysis methods. It achieves the integrated application of strategies such as improving the accuracy of analysis models through feature selection, promoting water quality prediction through the fusion of multi-source data, analyzing the evolution laws of water quality through spatio-temporal coupling, and dynamically monitoring water quality risks with intelligent models, providing practical references for relevant personnel.

Keywords: Machine learning; Coupling effect; Water environment quality

0 引言

随着全球工业化与城市化进程的飞速发展, 水环境系统正遭遇自然因素波动、多介质交互影响及人类活动干扰的复杂难题, 传统水环境质量分析方法依靠的主要是统计学模型与专家经验, 在处理非线性关系、高维异构数据及动态演变过程时局限性逐渐凸显。机器学习技术依托其海量数据处理优势、自适应学习特性及强大的非线性拟合能力, 为水环境质量分析供给了新的技术办法。如今多算

法配合、多源数据融汇的深度探究相较于早期的单一模型使用, 机器学习正不断拓展传统方法范畴, 其通过探究数据内在法则、组建动态预测框架及甄别关键影响因子, 能完成对水质态势更精准的勾勒与更前瞻的风险评估, 尤其是面临“数据驱动决策”的背景, 机器学习与水环境科学的高度结合作用, 不仅实现了分析范式从“经验主导”到“数据智能”的转型, 更为科学支撑了水环境质量的系统性管理。

1 机器学习概述

机器学习是借助算法让计算机从数据中自动摸索规律,并凭借规律对新数据作出预测或决策的人工智能技术分支,核心为构建数字模型,依靠训练数据调整模型参数,使模型得以获取输入和输出间的复杂映射关联,而无需人工专门写出编程规则。常见机器学习方法有监督学习(如决策树、神经网络、支持向量机)、强化学习及无监督学习(如降维、聚类)等,其中监督学习凭借可直接利用标注数据开展精准预测^[1]。

2 水环境质量分析中机器学习的作用

2.1 提升水质预测精准度与效率

水质预测是水环境管理的核心要素,传统方法(如线性回归、时间序列模型)依靠简化假设,不易精准刻画多因素(如流量、温度、土地利用)对污染物浓度非线性作用的变化进程;机器学习凭借设立输入跟输出的高维映射模式,能够自我学习指标间的复杂联动关系,对照传统模型要人工筛选变量然后设定固定参数。机器学习借助数据驱动自动做出特征权重与模型结构优化,能保持预测精度的同时降低分析周期,达成对水质短期起伏与长期走向的迅速反应,给应急决策及资源调配提供实时依据^[2]。

2.2 挖掘复杂数据中的潜在关联

水环境系统包含水体质量指标(如COD、氨氮、溶解氧)、环境因素(如降水、水温、pH值)、人类行为(如排污量、土地利用类型)等多源异构数据,传统分析举措不易识别深层次的关联格局,机器学习借助无监督学习与深度神经网络,能够从巨量数据中自主挖掘隐藏的关联特性,给予数据支撑以理解水环境系统的整体性,引领分析从“单指标孤立评价”迈向“多要素协同解析”^[3]。

2.3 实现水质动态监测与风险预警

水质风险体现出不确定性与突发性,传统基于固定阈值的监测方法难以及时捕捉异常状态的演变过程。机器学习通过接入实时多源监测数据,依托在线学习机制动态更新模型参数,可以完成对水质状态的连贯追踪与异常模式的迅速鉴别,孤立森林算法可快速甄别水质指标中的离群数值,能利用序列建模,循环神经网络预估未来短时间内水质恶化的走向。此动态监测及预警能力不仅增强了风险响应的时效,还能借助持续学习达成适应环境条件长期变化的效果,助力水环境安全主动防控达成智能化工具支撑^[4]。

3 基于机器学习耦合作用的水环境质量分析策略

3.1 多源数据融合驱动水质预测

水质监测数据、气象参数、水文条件、土地利用类型及人类活动排放等多源因素均会影响水环境质量。单一类型数据不易完整呈现水质状态动态变化,导致传统预测模型面对复杂场景的适应性较差。而通过汇聚异构多源数据,可为水质预测提供更充分的输入内容,提高模型对非线性关系的捕捉本领与预测结果的准确性。

实际操作时,首先收集社会经济数据、水质监测数据、气象数据、水文数据及遥感影像数据,将这些数据统一变更为相同量纲与时间分辨率,确认数据实现时空维度一致,而后采用特征交叉与嵌入技术,将不同来源的特征汇聚到统一特征空间,例如采用主成分分析方法提取气象与水文数据的共性成分,或借助图神经网络处理数据来构建水质指标与周边土地利用类型的空间关联图谱。

其次,持续采用多模态融合模型。如依靠注意力机制的深度神经网络,对多源特征实施加权融合,强化关键信息对预测目标的关联程度,还可采用集成学习方法,把单一数据源训练的子模型所产生的预测结果进行二次优化。

依靠多种类型数据融合,水质预测模型可综合权衡环境因子的时间滞后效应和交互作用,明显提高对短时波动与长期趋势预测的精确水平,尤其是在恶劣气候或突发污染情形当中,可更准确地展现水质动态的响应机制。

3.2 特征选择优化分析模型精度

水环境数据一般包含大量冗余特征与无关变量,直接输入机器学习模型会造成过拟合风险增加、掩盖关键信息及计算效率降低,降低模型预测结果的稳定性与对水质核心驱动因素的辨识能力。采用有组织的特征选择,可去除干扰特征并留存对水质指标最具解释意义的核心变量,从而改良模型架构并增进预测精度。实际操作中首先借助过滤式方法求取各特征与目标水质指标的统计关联强度,初步选出与目标关联度高的候选特征群,而后引进包裹式方法(从初始特征集合中不断地选择特征子集),遵照特定机器学习模型训练后的结果,持续除去对模型性能效果最差的特征,逐步改良特征子集。更进一步结合嵌入式方法,在模型训练进程中自主探寻特征的权重系数,辨别对预测目标影响程度最深的因子。对于高维稀疏的数据而言,可利用主成分分析或非负矩阵分解做降维处理,选用低维潜在特征替换原始变量。

其次,特征选择操作完成后,将经优化的特征集输入

到深度神经网络或梯度提升模型（如图1）进行训练，通过审视特征选择前后的模型性能指标，检验特征筛选对提高精度的成效。此方法可切实消除无关特征的干扰，集中剖析核心驱动因子的作用机制，使模型凭借更少计算资源取得更高的预测精准度。

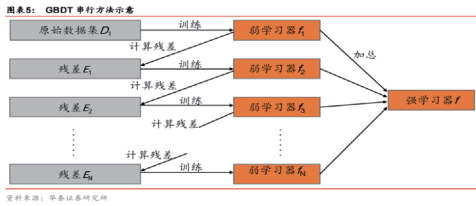


图1 梯度提升模型

3.3 时空耦合分析水质演变规律

流域地形、土地利用类型及污染源分布的差异对水质指标在空间维度上产生影响，在时间维度上体现长期趋势变化与季节性波动的结合特征，传统分析方法不易同时捕捉时间与空间维度的交互影响。采用时空结合分析，可以展现水质变化的时间动态性规律与空间异质性，为区域开展差异化治理提供科学参考。

首先打造时空数据立方体（方法如图2），将水质监测数据依照时间序列与空间网格进行归整，形成囊括空间坐标、时间戳与指标值的数据三维结构，继而选取时空卷积神经网络或时空图神经网络。时空卷积神经网络借助三维卷积核同时获取时间步长内的序列特征与空间邻域内的局部关联模式；时空图神经网络则将监测站点、流域单元设为节点，以水力连通性或空间距离构建图结构，利用图卷积层将节点间的时空依赖信息传递。

其次将卷积神经网络与长短期记忆网络的混合架构结合，借助长短期记忆网络抓取水质指标的长时间序列关联，卷积神经网络采集空间网格内的局部特性，借助门控机制达成时空特征的动态融合，开展模型训练的期间。采用时空注意力机制，自行赋予空间区域与不同时间步长对预期目标的权重，发挥关键时段与关键区域的影响作用。实施时空耦合分析，可以度量水质指标的空间扩散路径及时间累积效应，比如确定污染物从上游到下游的移动速度以及季节性农业面源污染的频发时期，为制定恰当的流域治理策略提供数据保障。

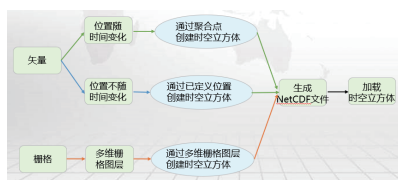


图2 时空数据立方体制作方法

3.4 智能模型动态监测水质风险

水质风险呈现出突发性、快速扩散与不确定性的特性，传统静态监测方法借助人工巡检与固定阈值，难以察觉异常状态并及时预警潜在风险，依靠智能模型做动态监测，可达成对水质情况的实时追踪、异常类型的快速判定与风险档次的智能区分，有力提升风险响应的时效水平与精准水平。

首先设置多源实时监测设备，不断采集外部干扰因子（如降雨强度、排污量）、环境参数（如水温、溶解氧、pH值）和水质指标（如电导率、浊度、重金属浓度）的高频数据资料，之后运用流式计算框架（如 Apache Flink）对实时数据做预处理，含有缺失值补齐。继续创建在线学习模型，如在线支持向量机或增量式随机森林，模型在接纳新数据后自动修正参数，不用再次训练整体数据，达成对环境变化的快速适应效果。

其次，运用孤立森林算法对实时数据流中的离群点（如重金属浓度骤增）进行检测，或借助循环神经网络（RNN）对未来短时间内水质的趋势（如未来3小时溶解氧下降幅度）进行预测，若预测得出的值超出动态阈值则启动风险预警。借助历史风险事件数据构建类别划分模型，把水质态势分为低、中、高风险层级，并匹配具体的风险源类型（如表1），为应急决策供给针对性意见。采用智能模型开展动态监测，可达成从“事后响应”到“事前预警”的转变，极大减少水质风险对生态系统与人类健康的潜在威胁。

表1 风险源类型及应对

风险等级	核心监测指标异常阈值范围	典型关联风险源类型	响应时效要求
低风险	指标波动幅度 ≤ 10% 且无离群点	自然因素	24小时内响应
中风险	单一指标超基准值 10% ~ 30%	农业面源	12小时内响应
高风险	多指标协同异常或单一指标超30%	工业泄漏	1小时内紧急处置

4 结语

本文提出多源数据融合促进预测、特征甄选增强模型精度、时空联合解析演变规律及智能模型动态管控风险等做法，可有效提升水质分析的时效性、前瞻性与精准性，发挥融合机器学习在特征提取、非线性建模及动态响应方面的长处。此策略模式为复杂水环境情况里的质量检验提供了可运用的实践借鉴，有关工作者未来应积极寻觅机器学习算法跟领域知识的深度融合渠道，以助力水环境质量分析向“机理 - 数据协同驱动”的更高级阶段发展，最终促成水质保障安全能力的全面增强。

参考文献:

[1] 李晶, 顾立业. 数据挖掘技术在水环境质量监测中的应用研究[J]. 清洗世界, 2025, 41 (09): 119-121.

[2] 刘凯, 刘锋. 基于机器学习模型的流域水质评价与预测[J]. 环境污染与防治, 2025, 47 (07): 1-10.

[3] 梁远仁. 城镇化进程中的城郊流域水环境质量演变机制研究[D]. 福建农林大学, 2025.

[4] 刘瑶, 胡静. 基于时间序列、主成分聚类 and 机器学习耦合作用的水环境质量研究: 全过程视野下的解析[J]. 环境影响评价, 2023, 45 (04): 116-124.

基金项目: 本文系山东省教育厅职业院校教师访学研修资助项目阶段性成果。

作者简介: 郑伟 (1981.12-), 男, 汉族, 山东省济南市, 研究生, 副教授, 研究方向: 流域生态综合治理。