

Research on Precise Detection of Trace Organic Pollutants in Drinking Water Sources Based on Multi scale AI

Fenfang Tan

Hunan Changsha Ecological Environment Monitoring Center, Changsha, Hunan, 410000, China

Abstract

A multi-scale data fusion strategy is constructed using a dual layer architecture of “feature level fusion+decision level fusion”, which takes into account the complementarity and independence of features at different scales, and improves the accuracy and robustness of pollutant detection models. Feature level fusion is based on attention mechanism, which identifies the contribution of each scale feature through weight allocation, highlights core micro features, suppresses redundant information, and integrates them into a global feature vector through matrix operation, solving the problem of single scale feature information deficiency. Decision level fusion relies on the D-S evidence theory, using detection results from various scales as evidence sources, quantifying fusion credibility to eliminate uncertainty, and combining preliminary results from deep learning and machine learning models. Through iterative optimization of the synthesis process, a comprehensive decision result is output.

Keywords

multi-scale; AI technology; Drinking water source; Pollution monitoring

基于多尺度 AI 的饮用水源微量有机污染物精准检测研究

谭芬芳

湖南省长沙生态环境监测中心, 中国·湖南长沙 410000

摘要

采用“特征级融合+决策级融合”双层架构构建多尺度数据融合策略, 兼顾不同尺度特征的互补性与独立性, 提升污染物检测模型的精度及鲁棒性。特征级融合基于注意力机制, 通过权重分配识别各尺度特征贡献度, 突出核心微观特征、抑制冗余信息, 经矩阵运算整合为全局特征向量, 解决单一尺度特征信息缺陷。决策级融合依托D-S证据理论, 将各尺度检测结果作为证据源, 量化融合可信度以消除不确定性, 结合深度学习与机器学习模型初步结果, 通过迭代优化合成过程, 输出综合决策结果。

关键词

多尺度; AI技术; 饮用水源; 污染物监测

1 引言

随着对于污染物检测在精度以及稳定性方面要求的不断提高, 仅依靠单一尺度的数据很难全面地描绘出污染物的特性, 很容易因为所获取的信息不够全面而造成检测出现偏差, 微观特征对于污染物的识别以及定量来说是非常关键, 但缺少时空关联性方面的支持, 宏观特征可以反映出环境的规律, 但针对性却有所欠缺。传统的融合方法难以在特征互

补性以及独立性之间找到平衡, 无法有效地整合多维度的信息, 对检测模型的鲁棒性形成了制约。基于此, 迫切需要构建一个高效的多尺度数据融合架构, 实现不同层级特征的深度整合以及精准决策, 为污染物监测提供更为可靠的技术支撑。

2 饮用水源微量有机污染物特性

2.1 典型微量有机污染物种类及危害

饮用水源当中的微量有机污染物有着繁杂的种类, 按照来源可分成天然源性以及人为源性这两类。天然源性污染物主要包含藻类代谢所产生的藻毒素, 如微囊藻毒素、节球藻毒素等, 这些物质有较强的肝毒性, 长时间摄入会使肝细胞受到损伤甚至发生癌变, 人为源性污染物所占比例更高, 包含农药残留、抗生素、内分泌干扰物、持久性有机污染物这四大类别^[1]。农药残留里面的有机磷、拟除虫菊酯类物质, 会经由农田径流汇入水源, 对人体神经系统功能造成干扰,

【课题项目】长沙市2023年第一批自然科学基金项目《多尺度人工智能辅助下的饮用水源地微量有机污染物检测及生态健康风险评价》(课题编号: kq2402132)。

【作者简介】谭芬芳(1988-), 女, 中国湖南郴州人, 硕士, 高级工程师, 从事生态环境质量监测, 环境风险评价与预警研究。

抗生素类污染物大多源自医疗废水与养殖废水排放，长时间存在会促使环境中耐药菌滋生，对公共卫生安全构成威胁，内分泌干扰物比如双酚 A、邻苯二甲酸酯，可依靠食物链进行富集，扰乱人体内分泌系统，引发生殖发育异常。如表 1 所示。

表 1 污染类别

污染物类别	典型代表	主要来源	核心危害
藻毒素	微囊藻毒素-LR	蓝藻水华代谢	肝毒性、致癌性
农药残留	毒死蜱、氯氟氰菊酯	农田径流、农药施用	神经系统损伤
内分泌干扰物	双酚 A、邻苯二甲酸二丁酯	工业废水、塑料降解	生殖发育紊乱
持久性有机污染物	多氯联苯、多环芳烃	工业排放、燃烧产物	长期蓄积毒性、致癌致畸

2.2 微量有机污染物在饮用水源中的赋存形态与迁移规律

微量有机污染物在饮用水源中的赋存形态会受到自身理化性质以及水体环境参数的共同作用，主要呈现为溶解态、吸附态这两种形式，如多环芳烃、多氯联苯这类疏水性较强的污染物，容易吸附在悬浮颗粒物的表面形成吸附态，其赋存比例和水体中悬浮颗粒物的含量、有机质含量呈现正相关关系，而抗生素、部分农药等亲水性污染物，大多以溶解态的形式存在于水体之中，其稳定性会受到水体 pH 值、温度的影响。其迁移规律呈现出十分突出的时空差异性^[2]。在空间方面，靠近污染源的区域污染物浓度比较高，随着水流扩散呈现出梯度递减的态势，吸附态污染物会随着悬浮颗粒物沉降，容易在水体底泥中富集。在时间方面，丰水期因为受到径流稀释作用，污染物浓度普遍下降，枯水期水体流动性较差，污染物容易累积。微生物降解、光降解是污染物迁移转化的关键途径，在好氧环境下微生物对易降解污染物的分解效率比较高，光降解对芳香族污染物的转化作用更大。

3 多尺度检测数据获取与预处理

3.1 实验设计与样品采集

实验设计围绕覆盖饮用水源典型场景以及保障数据代表性展开，结合研究区域的水文特征与污染分布规律，运用分层抽样和随机抽样相结合的办法来布设采样点，研究区域挑选了河流型、水库型这两类典型的饮用水源地，按照上游、中游、下游以及近岸、中心区域来划分采样单元。同时，设置污染源附近的对照点与远离污染源的背景点，共布设 30 个采样点。样品采集严格依照水质监测技术规范执行，每个月采集一次水样，连续监测一年以便捕捉时空变化特征，采样时采用棕色玻璃瓶避光采集表层、中层、底层各 1L 水样，现场测定水温、pH 值、溶解氧等基础参数并做好记录，样品采集后马上加入固定剂抑制微生物活动，在 4℃ 冷藏运输，24 小时内完成实验室预处理，防止污染物形态转化与浓度

损耗，保证样品完整性与数据可靠性^[3]。

3.2 多尺度检测数据获取

多尺度检测数据获取是围绕微观、中观以及宏观这三个维度来进行的，形成了一个全方位的数据支撑体系（如表 2 所示）。在微观尺度方面，主要聚焦于污染物自身所有的固有特征，运用色谱—质谱联用技术来测定目标污染物的分子结构、官能团信息以及痕量浓度，借助光谱技术采集污染物的特征光谱数据，以此精准地捕捉单一污染物的微观属性。中观尺度更侧重于水体环境关联性数据，利用便携式监测设备在现场测定悬浮颗粒物含量、有机质含量、浊度等参数，分析水体环境对污染物赋存形态所产生的影响，在每一个采样点都设置 3 个平行样进行测定，以便降低误差，宏观尺度则是对多源数据进行整合，其中涉及了区域水文监测数据、污染源排放数据、气象数据以及历史监测数据等，构建起时空数据库。

表 2 尺度类型

尺度类型	数据类型	获取技术 / 方式	数据用途
微观尺度	污染物分子特征、痕量浓度	色谱 - 质谱联用、光谱技术	AI 模型微观特征提取
中观尺度	水体环境参数	便携式监测设备、平行样测定	污染物赋存与迁移关联分析
宏观尺度	时空多源数据	水文站监测、污染源排查、历史数据整合	区域污染规律与趋势分析

3.3 数据预处理流程

数据预处理依照去噪、标准化以及特征优化这样的核心逻辑展开，以此为 AI 模型训练供给高质量数据。首先要做的是进行异常值剔除以及缺失值填补，运用箱线图法来识别异常数据，结合监测过程记录去判断异常原因，对于由偶然误差致使的异常值给予剔除，而因系统误差导致的异常值借助相邻时间点数据插值进行修正，缺失值运用 K 近邻算法来填补，同时兼顾数据的时空关联性与随机性。其次实施数据标准化与归一化，针对不同尺度、不同单位的数据，运用 Z-score 标准化消除量纲影响，把数据映射到同一区间，防止高量级数据主导模型训练。最后开展多尺度特征筛选与降维，借助方差分析与互信息法筛选出与污染物检测高度相关的特征，剔除冗余信息，采用主成分分析法降低数据维度，在保留核心特征的情况下提升模型训练效率，保证预处理后的数据可契合多尺度 AI 模型构建需求。

4 基于多尺度 AI 的微量有机污染物检测模型构建

4.1 模型整体框架设计

模型的整体框架把多尺度数据当作输入，将 AI 算法作为核心，以精准检测作为目标，构建起“输入层—特征层—融合层—训练层—输出层”五级递进的结构，实现从多源数

据到污染物精准识别以及定量分析的全流程闭环。输入层承担着接收经过预处理的微观、中观、宏观多尺度数据的任务,借助数据接口来完成不同类型、不同维度数据的规范化导入,同时设置数据校验模块,对输入数据的完整性和一致性做二次核查,防止无效数据进入后续流程。特征层着重于多尺度特征的初步提取与分类,给各尺度数据分配独立的处理通道,保证不同维度特征可得到针对性的解析,融合层运用多尺度数据融合算法,实现跨尺度特征的深度整合,突破单一尺度数据的局限性。训练层配备深度学习与机器学习混合模型,完成模型训练、参数优化以及性能迭代。输出层负责输出污染物种类、浓度、分布等检测结果,同时生成检测精度、可靠性评估报告。该框架有模块化设计的优势,各层级可独立优化并且协同联动,保证了检测的精准性,又提高了模型对复杂水体环境的适应性,可有效应对饮用水源中微量有机污染物成分复杂、含量极低、干扰因素多的检测难题。

4.2 多尺度特征提取模块

多尺度特征提取模块运用分尺度解析以及针对性提取的设计理念,依靠 AI 算法探寻不同尺度数据里和污染物检测有关的核心特征,以此为后续的数据融合以及模型训练给予支持,微观尺度特征提取着重于污染物的固有属性。依据卷积神经网络搭建特征提取子模块,对色谱-质谱数据、光谱数据做深度剖析,精确捕获污染物分子结构、官能团振动频率、特征吸收峰等微观特征,借助多层卷积与池化操作,过滤光谱干扰、基线漂移等噪声的影响,提取具有唯一性的污染物特征向量,实现对不同种类污染物的初步区分。中观尺度特征提取围绕水体环境和污染物的关联性来开展,运用随机森林算法构建特征关键性评估模型,从悬浮颗粒物含量、有机质含量、pH 值、温度等环境参数之中,挑选出对污染物赋存形态、迁移转化有明显影响的关键特征,量化环境参数与污染物浓度的相关性,构建关联性特征集,为分析污染物在水体中的分布规律提供参考。宏观尺度特征提取侧重于区域时空特征的挖掘,基于长短期记忆网络提取水文数据、污染源排放数据、气象数据中的时空演变特征,捕捉污染物浓度的季节变化、空间扩散规律等宏观信息。

4.3 多尺度数据融合策略

多尺度数据融合策略运用一种双层融合架构,也就是“特征级融合+决策级融合”的架构,这种架构可兼顾不

同尺度特征所有的互补性以及独立性,提升模型的检测精度以及鲁棒性,特征级融合作为底层融合环节,采用注意力机制融合算法,对微观、中观、宏观特征集展开深度整合。借助注意力权重分配机制,可自动识别不同尺度特征对于污染物检测的贡献程度,对于污染物种类识别以及浓度定量有着关键作用的微观特征赋予较高权重,对于环境关联性以及时空规律性有着指示意义的中观、宏观特征分配适配的权重,如此可以有效突出核心特征、抑制冗余信息,依靠矩阵运算把多尺度特征融合成统一的全局特征向量,以此解决单一尺度特征信息不完整以及针对性不足的问题。决策级融合作为上层融合环节,基于 D-S 证据理论构建融合决策模型,将各尺度特征对应的检测结果当作独立证据源,借助证据合成规则对不同证据源的可信度进行量化与融合,消除单一尺度检测结果的不确定性,针对特征级融合之后的数据,分别输入深度学习与机器学习模型得出初步检测结果,结合各模型的检测精度以及稳定性指标确定证据权重,依靠多轮迭代优化证据合成过程,最终输出综合决策结果。

5 结语

双层融合架构借助特征级以及决策级的分层设计方式,构建起了一套完整的多尺度数据融合体系,切实突破了单一尺度特征和传统融合方法所存在的局限,特征级注意力机制可保障核心特征被精准提取出来,决策级 D-S 证据理论融合则可以降低检测结果的不确定性,此策略为提高污染物检测的精度以及鲁棒性提供了切实可行的途径,可广泛适配各类环境污染物检测场景。未来,可以对注意力权重分配算法和证据合成规则进行优化,结合实际检测需求来拓展应用范围,帮助环境监测技术实现升级迭代。

参考文献

- [1] 王永坤,姚宇,雷振,等. 再生水残余有机物富集浓缩方法的优化与构建——以微量有机污染物和生物毒性检测为导向[J]. 中国环境科学, 2025, 45 (08): 4311-4321.
- [2] 刘国梁. 试论水体中微量有机污染物的检测研究进展[J]. 决策探索(中), 2020, (07): 91.
- [3] 张博超,章丽萍,王大伟,等. 全二维气相色谱-飞行时间质谱(GC×GC-TOFMS)非靶向筛查水源中微量有机污染物[J]. 环境化学, 2023, 42 (04): 1109-1117.