

基于数据增强的图卷积神经网络的浓度预测方法

罗丽平^{1, 2, 3, 4}, 曾腾¹

(1. 广西民族大学物理与电子信息学院, 广西 南宁 530006; 2. 多模态信息智能感知处理与应用广西高校工程研究中心, 广西 南宁 530006; 3. 广西智语人形机器人重点实验室, 广西 南宁 530006; 4. 广西智能视觉协作机器人工程研究中心, 广西 南宁 530006)

摘要: 为解决电容式浓度预测在样本稀疏与噪声干扰下精度下降的问题, 本研究提出一种基于数据增强的图卷积神经网络 (GCN, Graph Convolutional Network) 的浓度预测方法。对射频识别 (RFID, Radio Frequency Identification) 无线传感系统采集的实验数据, 先采用插值提高数据密度, 再通过卡尔曼滤波从观测序列中提取隐变量并与原特征拼接以扩充维度; 然后构建含深度残差连接的 GCN, 学习电容与浓度之间的非线性映射, 并采用不同图卷积方法进行性能评估。实验结果表明, 该方法浓度预测的平均相对误差 (MRE, Mean Relative Error) 为 2.35%, 明显低于线性递减权重的粒子群优化等现有方法。此外, 在不同的图卷积方法下, 其预测性能仍保持稳定。由此可见, 本研究提出的基于数据增强的图卷积神经网络架构, 能有效解决数据稀疏、含噪情况下盐溶液浓度检测的准确性与鲁棒性问题, 具有向多盐体系与复杂工况推广应用的潜力。

关键词: 浓度预测; 数据增强; 插值; 卡尔曼滤波; 图卷积网络 (GCN); 残差连接

中图分类号: TP212.9; TP183; X832

文献标志码: A

doi: 10.11959/j.issn.2096-3750.XXXX.

A Data-Augmentation-Based Graph Convolutional Neural Network for Concentration Prediction

LUO Liping^{1,2,3,4}, ZENG Teng¹

1. School of Physics and Electronic Information, Guangxi MINZU University, Nanning 530006, China

2. Engineering Research Center of Multi-modal Information Intelligent Sensing, Processing and Application, University of Guangxi, Nanning 530006, China

3. Guangxi Key Laboratory of ZHIYU Humanoid Robots, Nanning 530006, China

4. Guangxi Engineering Research Center of Intelligent Vision and Collaborative Robotics, Nanning 530006, China

Abstract: To address the accuracy degradation of capacitive concentration prediction under sparse samples and noise interference, this study proposes a concentration prediction method based on a data-augmented Graph Convolutional Network (GCN). Experimental data collected from a Radio Frequency Identification (RFID) wireless sensing system are first processed through interpolation to increase data density. A Kalman filter is then applied to extract hidden variables from the observation sequence, which are concatenated with the original features to expand the feature dimension. A GCN with deep residual connections is constructed to learn the nonlinear mapping between capacitance and concentration. Different graph convolution strategies are further evaluated for performance comparison. Experimental results show that the proposed method achieves a mean relative error (MRE) of 2.35%, which is significantly lower than existing approaches such

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 罗丽平, luoliping@gxmzu.edu.cn

基金项目: 国家自然科学基金 (No.62561006); 广西科技重大专项 (桂科 AA23073006); 广西重点研发计划 (桂科 AB25069215)

Foundation Items: Nature Science Foundation of China (No.62561006), Guangxi Key Science and Technology Major Project (No. AA23073006), Guangxi Key Research and Development Program (No.AB25069215)

as the linearly decreasing weight particle swarm optimization. Moreover, the prediction accuracy remains stable across various convolution strategies. These results demonstrate that the proposed data-augmented GCN effectively enhances the accuracy and robustness of salt solution concentration detection under sparse and noisy data conditions, with potential for extension to multi-salt systems and complex environments.

Key words: concentration monitoring, data augmentation, interpolation, Kalman filtering, graph convolutional network (GCN), residual connection

0 引言

近年来,随着工业自动化和智慧水务的发展,污水处理与浓度检测受到广泛关注^[1]。其中,盐水浓度作为污水处理的重要参数,其实时、精准的预测对于污染预警、水质调控和节能运行具有重要作用^[2]。相比于传统的实验室取样的原子吸收光谱法^[3-4]与电化学分析法^[5],基于无线射频识别(RFID, Radio Frequency Identification)的传感器系统因其非接触、低成本和易部署等优点,在环境监测中具有广泛的应用前景^[6]。

2019年Makarovaite等人提出了一种基于超高频(UHF, Ultra-High Frequency)RFID的无源标签天线检测方法,通过分析标签谐振频率随介电参数变化的响应,实现对封闭容器内水性及有机液体介电特性的识别,为低成本、非接触式液体监测提供了新的技术途径^[7]。在此基础上,后续研究者进一步尝试构建RFID标签的响应特征与目标参数之间的映射模型,建立“频率-浓度-电容”三元关系,以实现间接的浓度预测^[8-9]。然而,在实际监测环境中,标签响应信号常受温度、氢氧离子浓度指数(pH, Power of Hydrogen)及介质波动等多种干扰因数的影响^[10-11],导致原始数据存在数据缺失、噪声和非线性等问题,严重的约束了模型的预测性能和系统的稳定性。

为解决数据稀疏和噪声干扰对盐溶液浓度预测的挑战,深度学习依托端到端表征学习与噪声鲁棒特征提取,可从多源时序信号中自动建模浓度映射,成为提升精度与稳定性的有效途径。已有研究通过构建融合物联网与深度学习的空气质量监测框架,实现了污染物浓度的精确预测^[12]。注意力机制与LSTM模型相结合,对矿区粉尘浓度进行短时滚动预测,进一步提升了浓度预测的时效性与精度^[13]。尽管深度学习在浓度预测中展现了潜力,但在处理RFID传感数据时仍面临挑战。近年来,图

卷积网络(GCN, Graph Convolutional Network)因其在处理非欧几里得空间数据和捕捉节点间复杂依赖关系方面的优势^[14-15]。首先,RFID传感数据通常表现为离散的、低维度的采样点,缺乏天然的显式图结构拓扑^[16]。在样本稀疏的情况下,直接构建的图邻接矩阵往往连接性不足,导致图卷积过程中的信息聚合效率低下,难以充分挖掘电容参数与浓度之间的潜在关联。其次,实际工况下的RFID信号往往伴随显著的环境噪声和随机波动,标准GCN模型缺乏针对时序噪声的鲁棒特征提取机制,容易将噪声作为特征进行传播,从而降低预测精度^[17]。此外,传统的GCN架构在网络层数加深时容易出现过平滑现象或梯度消失问题^[18],限制了模型对高度非线性映射关系的拟合能力。

鉴于上述局限性,仅依靠标准GCN无法满足高精度、高稳定性的浓度监测需求。因此,本研究提出了一种结合数据增强与改进GCN的预测框架。针对数据稀疏与噪声问题,采用插值算法与卡尔曼滤波构建增强型特征空间,通过提取隐变量丰富节点信息并抑制噪声;针对深层网络训练困难问题,引入深度残差连接机制,在保留图结构信息的同时提升非线性表达能力。实验结果表明,所提出的方法在浓度预测中取得了显著的效果,平均相对误差(MRE, Mean Relative Error)达到2.35%,明显优于现有方法。

1 RFID无线传感系统

本文采用文献[9]提出的基于UHF RFID标签的非侵入式传感系统进行数据采集,实现液体介电特性的非接触性探测。该传感模型通过将两个RFID的标签贴附于管道外部,利用期间的电磁耦合作用形成等效的电容结构。当被测溶液的介电常数发生变化时耦合电容特性随之发生变化,从而影响标签的匹配电容。具体结构如图1所示:

图1展示的无线感知系统由工作环境—传输

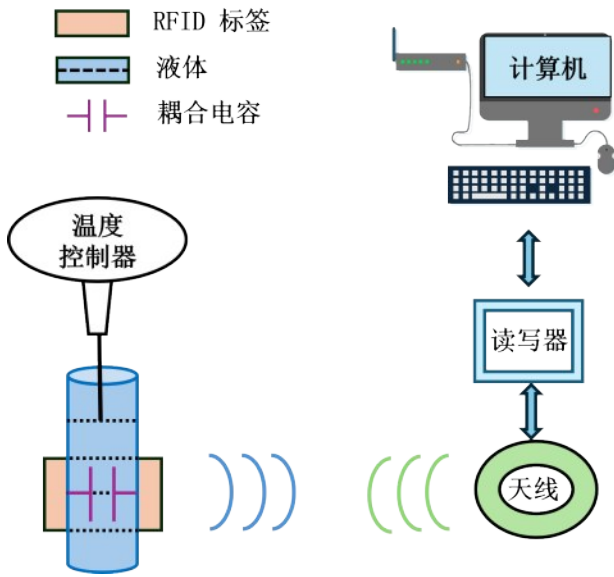


图1 基于UHF RFID的非侵入式无线传感系统

天线—读写器—PC端应用软件—温控单元构成。工作环境包含被测溶液与容器及一对对称贴附的RFID标签，两标签与介质共同形成等效缝隙圆柱电容，其介电性质变化经标签匹配电容的改变间接表征浓度。读写器由接口、射频、控制、电源四模块组成：接口模块与上位机进行指令与数据交互；控制模块管理整机工作状态；射频模块经外接传输天线激励无源标签并回收数据；电源模块为系统供能。上位机采用UHFReader18完成读写器配置、工作频段设定与频扫数据记录。为抑制温漂并提升读数稳定性，温控单元将溶液温度恒定在20°C。

该系统将两个标签作为电容，通过耦合电容的变化来反映介质的介电性质。由于它们附着在圆柱形管道上，因此两个标签被视为圆柱形电容器。在管道环境中，文献[9]对狭缝圆柱电容（SCC, Slit Cylindrical Capacitance）模型进行了描述与分析。该模型耦合电容的表达式如下式：

$$C = \frac{2\epsilon_0\epsilon_1\epsilon_r LR_1}{\sqrt{(\epsilon_1 R_2)^2 - [\epsilon_r(R_1 - R_2)]^2}} \cdot \arctan\left(\sqrt{\frac{\epsilon_1 R_2 - \epsilon_r(R_1 - R_2)}{\epsilon_1 R_2 + \epsilon_r(R_1 - R_2)}} \cdot \tan\left(\frac{\varphi}{2}\right)\right) \quad (1)$$

其中， ϵ_0 为真空介电常数， ϵ_1 和 ϵ_r 表示已知和未知的相对介电常数， R_1 和 R_2 为结构的内外径。 L 为内置电感， φ 为标签所对的圆心角。系统最终读出的匹配电容 C_M 受上述耦合电容的影响，其关系由式

(2) 给出：

$$C_M = \frac{1}{Lw^2} - C + \frac{1}{Rw} \sqrt{(CRw)^2 - 1} \quad (2)$$

Cole-Cole模型被广泛应用于描述溶液的介电特性，其表达如式(3)：

$$\epsilon_r^* = \epsilon' - j\epsilon'' \\ = \epsilon_\infty(x) + \sum_{n=1}^N \frac{\Delta\epsilon_n(x)}{1 + (j\omega\tau_n(x))^{1-\alpha}} + \frac{\delta_0}{j\omega\epsilon_0} \quad (3)$$

其中 ϵ_r^* 是复介电常数， ϵ' 和 ϵ'' 分别表示其实部和虚部。 n 表示Cole-Cole模型的阶， x 则表示浓度。 $\epsilon_\infty(x)$ 是高频极限介电常数，表示材料在高频下的介电响应。 $\Delta\epsilon_n(x)$ 是色散强度的大小。 $\tau_n(x)$ 是弛豫时间常数，它描述了材料的弛豫行为。模型通过考虑溶液的电学特性，利用频率、松弛时间等参数来描述溶液的介电响应。由公式(1)-(3)可知，溶液浓度的变化会引起介电参数发生改变，进而影响介电常数和耦合电容，最终通过天线的共轭匹配自动调整匹配电容 C_M 。

为便于分析，可以对模型做适当简化：忽略导电损耗项，并假设只存在一个占主导地位的弛豫过程($n=1$)。在这样的前提下，模型就会自然地化简为经典的Cole-Cole模型：

$$\epsilon^* = \epsilon_\infty(x) + \frac{\Delta\epsilon_1(x)}{1 + (j\omega\tau_1(x))^{1-\alpha}} \quad (4)$$

ϵ^* 的实部记作 ϵ' ，其计算方法如下：

$$\epsilon' = \epsilon_\infty(x) + \frac{\Delta\epsilon_1(x) \left[1 + (\omega\tau(x))^{1-\alpha} \sin\left(\frac{\pi\alpha}{2}\right) \right]}{1 + 2(\omega\tau(x))^{1-\alpha} \sin\left(\frac{\pi\alpha}{2}\right) + (\omega\tau(x))^{2(1-\alpha)}} \quad (5)$$

根据文献[19]，盐溶液的参数 a 取0.05。当 $a \approx 0$ 时， $\sin((\pi a)/2)$ 项几乎可以忽略，因此 ϵ' 可以简化为：

$$\epsilon' \approx \epsilon_\infty(x) + \frac{\Delta\epsilon_1(x)}{1 + (\omega\tau(x))^{2(1-\alpha)}} \quad (6)$$

在以下分析中，仅考虑复介电常数的实部 ϵ' 。之所以可以做此简化，是因为在903 - 927.4 MHz的频率范围内，虚部 ϵ'' 相对较小，且随溶质浓度变化的幅度有限[20]。相比之下， ϵ' 对浓度变化更为敏感，在决定体系电容特性方面起主导作用。因此，采用 ϵ' 就能够提供足够的信息，实现对浓度的准确估计。

2 数据获取与数据增强

本文搭建了一套无创的 RFID 实验平台来获取数据，用于支撑后续的特征扩展与模型训练。实验选用氯化钙 (CaCl₂) 标准溶液，浓度范围为 2000 - 10,000 mg/L，共配制 9 组样品，浓度间隔为 1000 mg/L。该范围涵盖了微生物从开始承受渗透压胁迫到活性显著受抑的完整过渡区间，是工业废水处理中需重点监测以实现早期预警的关键工艺段。溶质采用高纯度 CaCl₂ (≥99.5%)，溶剂全程使用去离子水。称量过程使用精度为 ±1 mg 的电子天平。将称取的 CaCl₂ 先溶解于 500 mL 去离子水中，再转入容量瓶定容至 1 L，充分混匀后保存备用。

实验平台如图 2 所示，该 RFID 系统采用博洋科技的 RFID101 超高频 (UHF) 读写器，并配套使用 UHFReader18 软件。RFMicron 3200AFR 标签配备尺寸为 90 × 24.1 mm 的金属天线，工作频率覆盖 903 - 927.4 MHz，与 IEEE 标准规定的 UHF 频段一致。实验设置的频率步长为 0.4 MHz，因此每种浓度溶液可获得 62 个频点的数据，其余实验条件与文献[21]保持一致。溶液容器为玻璃圆筒，相对介电常数为 8，内壁半径 10.9 mm，外壁半径 12.45 mm。温度控制器将溶液恒定保持在 20°C，以避免温度变化对相对介电常数的影响。对每一种浓度，系统都会进行一次完整的频率扫描；在每个频点记录对应的匹配电容，记为 $C_M(x, f)$ 。由此得到的原始数据集由频率、浓度与电容三元组构成。

实验测量仅在少量离散的浓度点上进行，导致浓度维度的数据分布较为稀疏，从而削弱了时序建

模的效果。为解决这一问题，研究在浓度维度采用插值方法进行数据增强，生成随浓度连续变化的“浓度-频率”特征集合，其具体流程如图 3。该方法不仅提升了数据集的连续性与完整性，也为后续神经网络建模提供了更充足的输入空间支撑。

为缓解原始实验样本稀疏带来的识别偏差与模型不稳定性，本研究采用数据插值进行有监督的数据增强:在不引入额外实验噪声与外推风险的前提下，基于现有 558 组样本，仅在观测区间内生成致密的拟合点。具体而言，我们选择了四类常用且代表性强的插值方法用于对比与互证：线性插值(强调局部保形与实现简洁)、样条插值(以分段多项式平滑提高一阶与二阶连续性)、多项式插值(在适度阶数控制下逼近全局趋势)以及基于回归分析的插值(以参数化关系捕捉潜在结构并具备一定抗噪性)。四种方法分别在相同的原始样本上构建插值函数，对于每一个固定的频率通道，构建浓度的映射函数。基于原始的离散点，利用插值函数拟合处浓度与信号轻度之间的连续曲线，并以 0.1 g/L 的步长进行从采样。并据此独立生成增强数据集；增强后的样本规模扩展至 5,022 组，显著提升了输入空间的覆盖度与梯度信息的可用性。

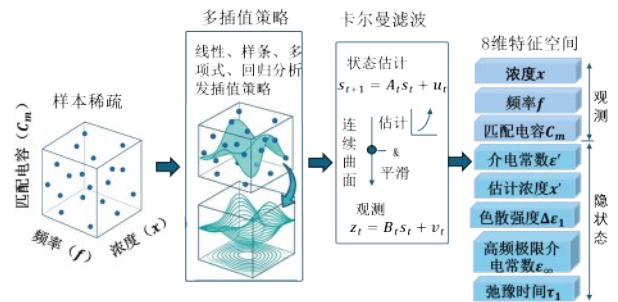


图 3 数据增强策略流程示意图

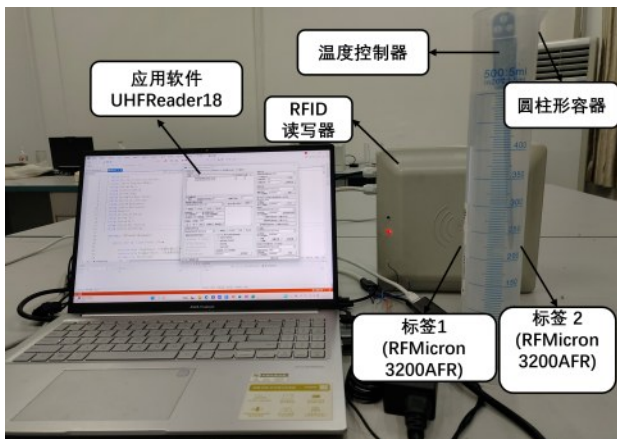


图 2 实验平台

在完成插值增强后、进入模型训练之前，我们基于文献[21]对插值后的序列进行卡尔曼滤波，并提取式 (6) 的 5 个隐变量 $s = [\varepsilon', x, \Delta\varepsilon, \varepsilon_\infty, \tau]^T$ 。在假设 $n = 1$ 的条件下， $\Delta\varepsilon, \varepsilon_\infty, \tau$ 均为与浓度相关的介电参数。为刻画这些参数与浓度 x 之间的对应关系，本文采用二次多项式对 $\Delta\varepsilon, \varepsilon_\infty$ 和 τ 进行拟合。设待估计的状态向量为 $s = [\varepsilon', x, \Delta\varepsilon, \varepsilon_\infty, \tau]^T$ 。则状态转移函数可表示为：

$$f(s) = \begin{cases} \varepsilon' &= \varepsilon_\infty + \frac{\Delta\varepsilon_1}{1 + (\omega\tau_1(x))^{2(1-\alpha)}} \\ x^* &= x \\ \Delta\varepsilon_1 &= a_1x^2 + a_2x + a_3 \\ \varepsilon_\infty &= b_1x^2 + b_2x + b_3 \\ \tau_1 &= c_1x^2 + c_2x + c_3 \end{cases} \quad (7)$$

其中, $a_i, b_i, c_i (i \in \{1, 2, 3\})$ 表示需要拟合的系数。为描述介电参数与浓度之间的关系, 引入二次多项式模型。之所以采用这种形式, 是因为其结构简洁, 却能够较好地刻画非线性变化特征。离子溶液的介电特性会随浓度变化而呈现明显的非线性趋势, 而二次多项式在模型复杂度与拟合精度之间取得了较为理想的平衡, 不仅计算量较低, 也能有效反映这种非线性依赖关系。

观测模型用于刻画系统状态与观测结果之间的内在联系, 其表达式为

$$h(s_i) = C_M(x, \omega) \quad (8)$$

其中, $h(s_i)$ 表示非线性观测函数在本研究中, RFID 传感系统的观测量为 C_M 。结合公式 (1)、(2) 和 (6), 可以进一步得到对应的测量模型表达式。

$$\begin{aligned} & C_M(x, \omega) \\ &= \frac{1}{L\omega^2} - \frac{2\varepsilon_0\varepsilon'}{l} R_1 \sqrt{(\varepsilon_1 R_2)^2 - (\varepsilon'(R_1 - R_2))^2} \\ & \cdot \arctan \left(\tan \frac{\varphi}{2} \sqrt{\frac{\varepsilon_1 R_2 - \varepsilon'(R_1 - R_2)}{\varepsilon_1 R_2 + \varepsilon'(R_1 - R_2)}} \right) \\ & + \frac{1}{R\omega} \sqrt{\frac{2\varepsilon_0\varepsilon' l R_1}{(\varepsilon_1 R_2)^2 - (\varepsilon'(R_1 - R_2))^2}} \\ & \cdot \sqrt{\left[\arctan \left(\sqrt{\frac{\varepsilon_1 R_2 - \varepsilon'(R_1 - R_2)}{\varepsilon_1 R_2 + \varepsilon'(R_1 - R_2)}} \tan \frac{\varphi}{2} \right) R\omega \right]^2 - 1} \end{aligned} \quad (9)$$

该方法的状态转移矩阵与观测矩阵均通过雅可比矩阵法获得。具体而言, 状态转移矩阵由状态转移函数 $f(s)$ 在当前状态估计处求偏导得到, 用于在扩展卡尔曼滤波中近似状态演化的线性关系; 观测矩阵则由非线性观测函数 $C_M(x, \omega)$ 对状态变量求偏导得出, 以实现观测方程的局部线性化。其他具体实现细节参考文献[21]。

运行卡尔曼滤波后, 提取状态5维隐状态与原始3维观测按时间对齐后拼接, 形成8维输入特征用于后续深度学习模型训练。据此通过4中插值方法以及分别进行卡尔曼滤波生成了4个独立的数据

集。该步骤用卡尔曼增益降噪抗漂, 并注入介电实验。在后续实验中, 模型分别在每个数据集上进行训练与评估, 旨在验证模型在不同数据分布假设下的鲁棒性。

3 深度残差连接的GCN设计

经过前述的数据增强策略, 原本稀疏且含噪声的观测数据已经被映射为包含物理观 $C_M(x, f)$, 以及介电隐变量 $(\Delta\varepsilon, \varepsilon_\infty, \tau)$ 的高维特征空间。然而卡尔曼滤波本质是一种基于时序的局部状态估计方法, 它有效的揭示了单一样本点背后的物理机制, 却难以显式捕捉不同频率点与不同浓度样本在特征空间中的全局拓扑依赖关系。在实际的射频传感过程中, 介电响应随频率和浓度的变化构成了一个连续且复杂的非线性流形, 样本点之间并非孤立存在, 而是通过某种内在的物理规律相互耦合。传统的回归模型 (如 SVR 或 MLP) 往往基于样本独立同分布的假设, 忽略了这种节点间的结构化信息, 导致在处理复杂非线性映射时泛化能力受限。

其具体网络机构如图4所示, 该架构通过在图卷积层旁路引入线性残差连接, 兼顾了局部特征的保持与邻域信息的聚合, 从而实现端到端的高精度浓度反演。具体而言, 本文将数据增强后的数据定义为图的节点特征矩阵 X , 其中 N 为输入的样本数量, 输入特征维度为7维。隐藏层中的图卷积层均为64层, 输出层为1用于输出预测的浓度。

为了捕捉所有样本间的全局依赖, 我们构建了全连接的边索引集合, 即图中任意两个节点间均存在边连接, 使得信息流能在整个特征空间内高效传递。网络主体由三个连续的双路残差图卷积块堆叠而成。与仅包含恒等映射的传统残差网络不同, 本模型在每一层并未简单相加输入与输出, 而是设计了“图聚合”与“线性变换”两条并行路径。每层节点特征更新规则形式化描述如下:

$$H^{(1)} = \text{ReLU}(\text{GCNConv}_1(X, A) + \text{Linear}_1(X)) \quad (10)$$

$$H^{(2)} = \text{ReLU}(\text{GCNConv}_2(H^{(1)}, A) + \text{Linear}_1(H^{(1)})) \quad (11)$$

$$H^{(3)} = \text{ReLU}(\text{GCNConv}_2(H^{(2)}, A) + \text{Linear}_1(H^{(2)})) \quad (12)$$

其中 H 节点特征矩阵, A 是图的邻接矩阵。这种设计既利用 GCN 捕获了节点间的邻域拓扑信息, 又通过 Linear 层保留了节点自身的特征变换能力, 有

效避免了深层 GCN 的信息同质化问题。为了衡量模型的预测效果，我们设计了一种自定义的损失函数——归一化平均绝对残差损失 (NMAR)。其表示为式(13)：

$$NMAR(y_{\text{pred}}, y_{\text{true}}) = \frac{\sum |y_{\text{pred}} - y_{\text{true}}|}{\sum |y_{\text{true}}| + \epsilon} \quad (13)$$

其中， y_{pred} 为模型预测值， y_{true} 为真实标签， ϵ 为防止除零错误的小常数。该损失函数的设计确保了在进行浓度预测时，模型不仅优化了预测精度，同时提高了在数据稀疏和噪声干扰下的稳定性从而影响标签的匹配电容

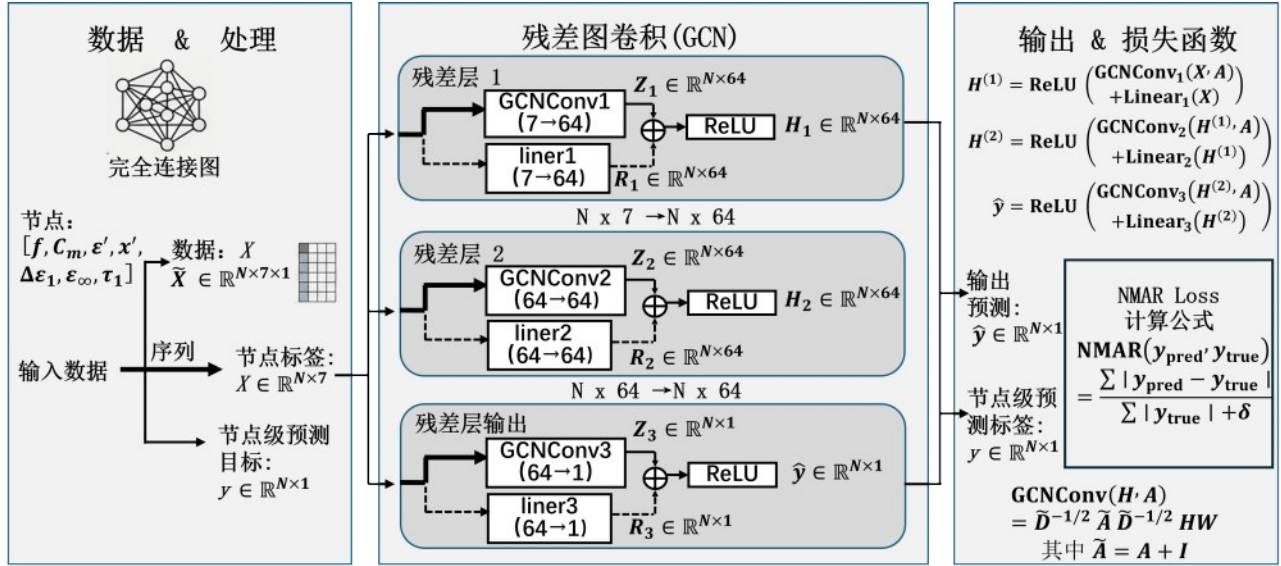


图4 深度残差连接的图卷积神经网络结构框图

4 实验结果与分析

4.1 模型训练

为了客观评估模型在实际应用中的泛化性能，避免因数据同源性导致的过拟合问题，本研究采用了基于独立实验批次的数据集划分策略。具体而言，分别进行了多组独立的射频传感实验。值得注意的是，训练集与测试集均独立进行了相同标准的数据增强预处理 (5022 组样本)，确保了输入特征分布的一致性，从而有效验证模型在面对全新、未知实验样本时的鲁棒性。

通过设计好的深度残差连接图卷积神经网络进行模型训练。目标通过预测数据集中的浓度值，输入特征维度为7、隐藏层的维度为64、输出维度为1。优化器选用 Adam，其学习率为0.01。训练过程采用自定义的 NMRE 的损失函数，**最大训练周期为 2000 轮**。其训练不同数据集的损失函数如图 5 所示：

从图 5 中可以看出，4 种插值方法的数据集在

深度残差连接的图卷积神经网络中，其训练损失持续下降，并表现出良好的趋势。约在第 1250 轮时，所有插值方法的数据集训练损失均低于 0.05，训练至第 2000 轮时，损失稳定在 0.03 左右波动，表明该网络能够有效学习数据间的特征信息。

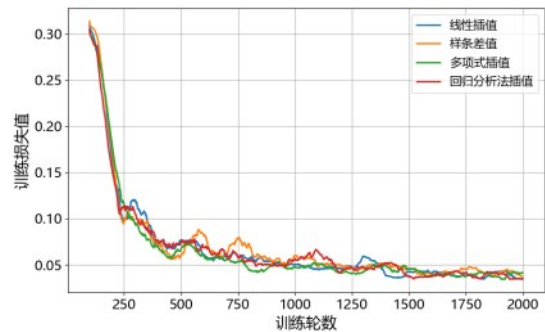


图5 不同数据集下训练损失函数对比图

为探究不同拖布聚合机制对浓度预测精度的影响，并验证所提框架在特征提取层面的通用性，本研究对比分析了 3 种主流的图卷积算子，分别是

GCNConv、ChebConv 和 SAGEConv，并在同一数据集上验证其收敛情况。其中，GCNConv 作为经典的频域方法，基于拉普拉斯矩阵的一阶近似实现高效的局部特征平滑，是目前应用最广泛的基准模型^[22]；ChebConv 同样基于谱图理论，但通过引入切比雪夫多项式滤波器，能够灵活的调整多项式阶数以显示的捕捉领域内的结构信息^[23]；而 SAGEConv 则代表了典型的空域归纳式学习范式，不同与前两者的谱视角，它侧重于通过采样与聚合函数直接恶化空间令居介电的特征^[24]。图6给出了上述3种不同的图卷积算子针对同一数据集，相同超参数的配置下的训练损失收敛轨迹，直观反应他们在处理射频传感信号的适应性差异。

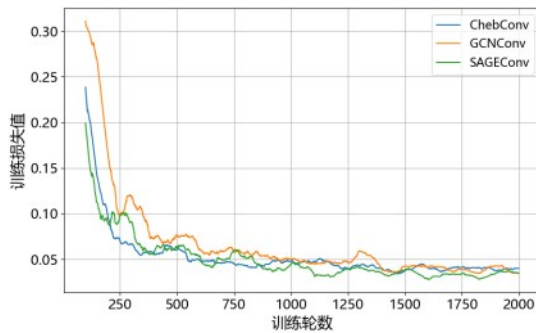


图6 不同图卷积方式的训练损失函数对比图

从图6可以看出，即使改变图卷积方式，训练损失依然较低。在训练过程中，GCNConv 的损失明显高于ChebConv 和 SAGEConv，而ChebConv 在1000轮后训练损失稳定低于0.05，三种图卷积方式最终都收敛至0.03左右波动。这表明，残差连接的图卷积网络能够有效学习浓度与电容数据的特征，并且适应不同的图卷积方式。

4.2 浓度预测结果分析

在训练结束后，使用训练好的模型对不同时序的数据进行浓度估计。其预测浓度与真实浓度的比较如图7所示，上半部分为整体趋势，下半部分则为局部放大，用数据点的形式画出以便清晰观察期预测情况。其中序列1浓度逐渐上升和序列2逐渐下降分别对应模拟污水处理过程中盐分逐渐积累和稀释或者处理过程。

从整体效果上看，基于深度残差连接的图卷积架构，即使更换不同的图卷积方式，均可以较好的跟踪浓度浓度变化的趋势。图7(a)展示是随时间

逐步上升的序列1，图7(b)则层现逐渐下降的过程。在2000-8000 mg/L的浓度区间内，模型的预测精度较高。值得注意的是，尽管在约6000 mg/L处观察到了轻微的局部波动，但其幅度远小于高浓度区的偏差，且处于随机测量噪声的统计容限之内，未对整体趋势跟踪造成影响。相比之下，当浓度超过8000 mg/L后预测曲线的波动较大，这可能与模型在高浓度条件下无法完全捕捉到更高浓度下的关系特征。另一方面，介电参数在该区间往往趋于饱和状态，当浓度继续变化时，介电性质的幅度越来越小，敏感性随之下降。在实际的工程实践中，当盐浓度超过8500 mg/L,会对污水处理的过程造成明显干扰，例如抑制微生物活性，降低除磷效率；相比之下，较低的盐浓度(≤ 8000 mg/L)对系统性能影响较小^[25]。这也从应用角度说明，将盐浓度监测重点放在2000 - 8000 mg/L范围内具有现实意义。

为验证算法在实际浓度预测中的性能，生成随机序列3和4，模拟在实际工业环境中的非单调波动，以反映更真实的条件。结果如图8所示：

图8(a)和8(b)显示了数据的时间顺序被随机化时的结果，为所提方法的时间建模能力提供了进一步的验证。所有数据集的估计曲线保持平滑，没有明显的波动。每条曲线都能准确捕获浓度的变化。误差分析采用MRE评估,其计算公式如式(14)：

$$MRE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (14)$$

其中， \hat{y}_i 和 y_i 分别表示预测浓度和真实浓度， n 表示预测的样本量。表1给出了所提方法在不同图卷积算子下不同序列的具体预测MRE，可见其预测MRE均小于2.5%。从不同的图卷积算子方面可以看出，在各个序列上的平均MRE，SAGEConv 效果最佳，GCNConv 次之。由于GCNConv 应用较为广泛，其随机序列3和4的MRE分别为2.28%、2.41%，取其平均2.35%作为该方法的最终MRE。

为了评估不同插值方法的数据对所提残差图卷积网络架构浓度预测的影响，每个插值方法得到的数据不同时序的预测MRE取其平均作为该插值方法对应的预测MRE有表2给出：

由表2可以看出，深度残差连接的图卷积网

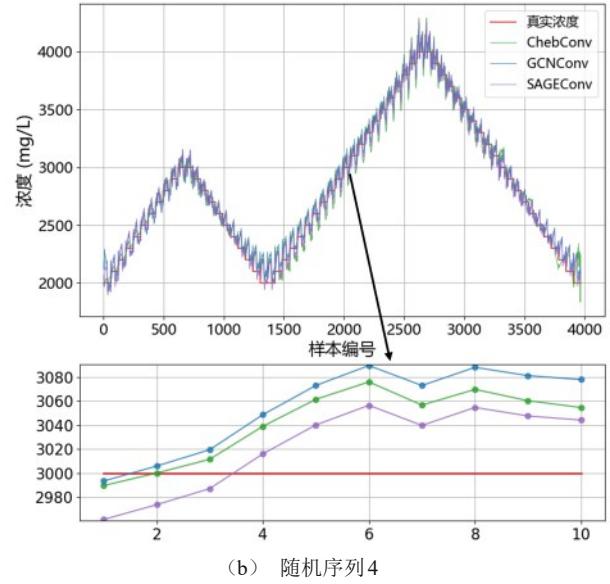
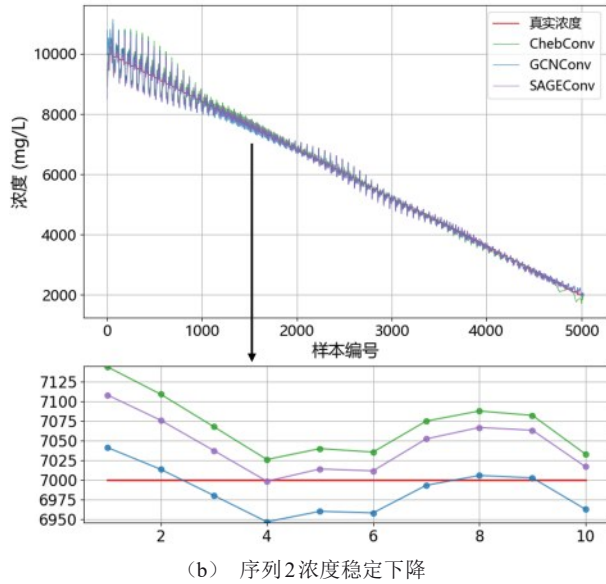
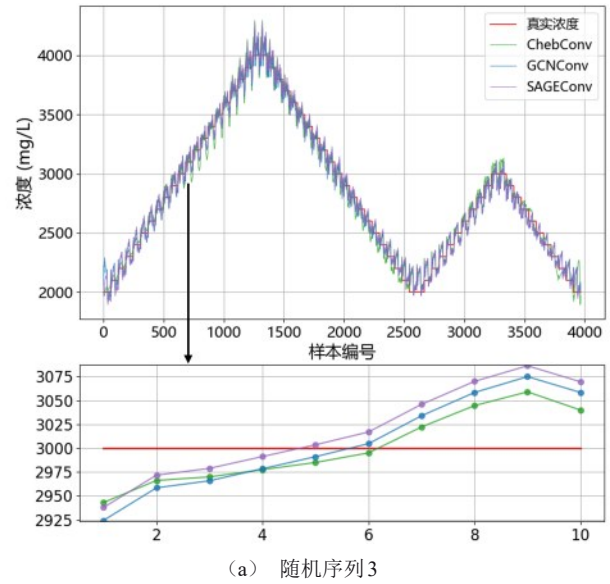
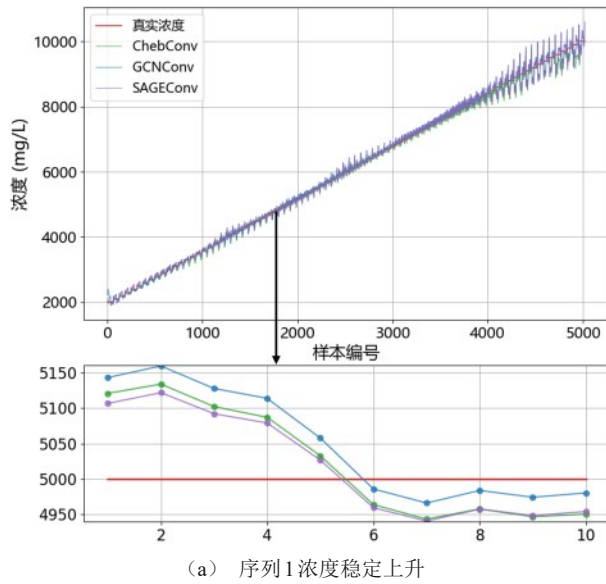


图7 预测浓度和真实浓度的比较

图8 预测浓度和真实浓度的比较

络，对不同的插值方法均可以达到较好的预测效果。其中样条差值方法在该架构下的不同图卷积算子下的平均MRE最高，为2.11%。而回归分析法的中和表现最佳仅为1.76%。从数据增强策略的角度而言，对于不同的插值方法，该架构均能较好的捕捉到浓度的变化，这在一定程度上验证了其性能的鲁棒性。

为更加群面的检验所提结构在不同条件下的稳健性，补充了分析性的可视化结果，用于从不同角度观察模型的整体表现。图9给出了浓度-频率二维空间中的相对误差分布情况，其中Z轴周表示相对误差的大小，从空间的角度层显模型误差分析的整

表1 不同卷积方法和序列的MRE

MRE(%)	序列1	序列2	序列3	序列4
ChebConv	1.67	2.31	2.25	2.40
GCNConv	1.74	1.86	2.28	2.41
SAGEConv	1.73	1.89	2.21	2.16

表2 不同卷积方法和序列的MRE

MRE(%)	线性	样条	多项式	回归分析
ChebConv	2.16	2.12	1.88	1.91
GCNConv	2.06	2.14	1.74	1.72
SAGEConv	2.00	2.07	1.71	1.64

体分布特征。

图9的结果进一步印证了上述分析。观察到在

较宽的浓度与频率范围内，相对误差始终维持在较低水平，空间分布均匀，没有明显的异常区域。这种一致性表明，该方法对浓度和频率变化并不敏感，具备良好的适应性和稳定性。整体来看，该框架在绝大多数工作区间内都表现出可靠而一致的性能，少量精度波动也处于可接受范围之内，充分体现了方法在实际应用中的稳健性与实用价值。

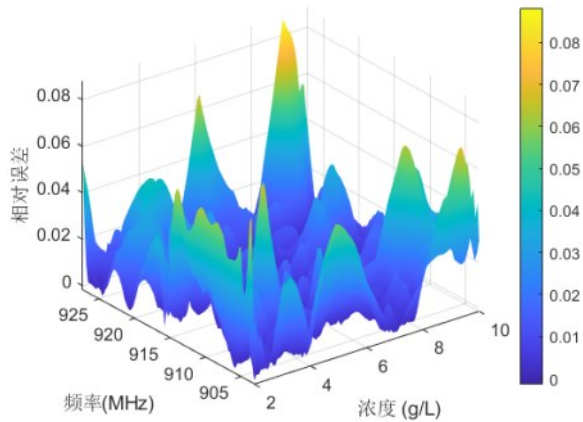


图9 预测浓度与真实浓度的相对误差

表3总结了这项工作的性能，并与近年来的其他研究进行了比较。

表3 所提方法和其他方法的比较

文献	溶液	方法	频率(GHz)	浓度(mg/L)	MRE (%)
[8]	Water/ Glucose	LDW-PSO	0.902-0.928	25-375	4.15
[26]	Water/ NaCl	校正曲线拟合	1.88-1.94	0-10000	3.73
本文	Water/ CaCl ₂	残差图卷积网络	0.903-0.927	2000-10000	2.35

表3将本研究提出的深度残差链接的图卷积框架与现有的代表性传感方法进行了基准对比。为了确保对比的学术严谨性，必须指出各方法在核心机制上的本质差异：文献[8]采用基于物理模型的优化方法，利用线性递减权重的粒子群优化算法(LDW-PSO, Linear Descending Weights Particle Swarm Optimization)求解逆问题；文献[26]则依赖于高灵敏度的矩形波导谐振腔硬件，通过校准曲线拟合实现检测。对比结果表明，尽管实验配置不同，本研究提出的方法在无需复杂迭代计算(相比[8])且仅使用低成本标签(相比[26])的前提下，

在工程关键的2000-8000 mg/L区间内实现了更优的精度(MRE 2.35%)，充分验证了深度图表示学习在解决非线性传感难题上的算法优势。

5 结论

本研究提出了一种融合卡尔曼滤波数据增强与深度残差图卷积神经网络(GCN)的浓度预测框架，并通过严谨的实验证据证实了其在样本稀疏与噪声干扰下的优越性。针对稀疏采样挑战，模型利用插值扩充与全连接图拓扑的全局聚合能力，在有限离散样本下成功重构了频率响应的连续规律，实现了仅2.35%的MRE，显著优于线性递减权重粒子群优化(LDW-PSO)等传统方法，证明了其在数据匮乏条件下的特征拟合能力；针对噪声干扰，通过卡尔曼滤波提取高维隐变量与残差学习机制构建的双重抗噪体系，有效抑制了信号波动，不同图卷积算子下MRE均稳定在2.5%以内的实验结果，充分佐证了该框架具有卓越的鲁棒性与泛化稳定性。综上所述，该方法有效突破了传统非线性映射模型在复杂工况下的精度瓶颈，为低成本、高精度的实时工业水质监测提供了可靠的理论依据与技术支撑。

未来研究将探索多参数联合建模策略，例如耦合流速、温度和pH值。这些参数可能存在复杂交互作用，联合建模能更有效捕捉其关联性，从而提升模型精度与泛化能力。

参考文献：

- [1] CRINI G, LICHTFOUSE E. Advantages and disadvantages of techniques used for wastewater treatment [J]. Environmental Chemistry Letters, 2019, 17(1): 145-155.
- [2] WEN J, DONG H R, ZENG G M. Application of zeolite in removing salinity/sodicity from wastewater: a review of mechanisms, challenges and opportunities [J]. Journal of Cleaner Production, 2018, 197: 1435-1446.
- [3] PERELONIA K B S, BENITEZ K C D, BANICOD R J S, et al. Validation of an analytical method for the determination of cadmium, lead and mercury in fish and fishery resources by graphite furnace and cold vapor atomic absorption spectrometry [J]. Food Control, 2021, 130: 108363.
- [4] ZHOU C, YU W T, YANG Y, et al. A new ion concentration detection method based on spectral matching and fusion for zinc leaching liquid [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 72: 1-12.

- [5] MONTERROSO S C C, CARAPUÇA H M, SIMÃO J E J, et al. Optimisation of mercury film deposition on glassy carbon electrodes: evaluation of the combined effects of pH, thiocyanate ion and deposition potential[J]. *Analytica Chimica Acta*, 2004, 503(2): 203-212.
- [6] QIAN X Q, LI Z, MENG Z Z, et al. Flexible RFID tag for sensing the total minerals in drinking water via smartphone tapping [J]. *IEEE Sensors Journal*, 2021, 21(21): 24749-24758.
- [7] MAKAROVAITE V, HILLIER A J R, HOLDER S J, et al. Passive wireless UHF RFID antenna label for sensing dielectric properties of aqueous and organic liquids [J]. *IEEE Sensors Journal*, 2019, 19(11): 4299-4307.
- [8] XIE S, MA C L, FENG R H, et al. Wireless glucose sensing system based on dual-tag RFID technology [J]. *IEEE Sensors Journal*, 2022, 22(13): 13632-13639.
- [9] FENG R H, XIANG X W, XIE S, et al. Sensing system for mixed inorganic salt solution based on improved double label coupling RFID [J]. *IEEE Sensors Journal*, 2023, 23(12): 13565-13574.
- [10] MAZZARACCHIO V, FIORE L, NAPPI S, et al. Medium-distance affordable, flexible and wireless epidermal sensor for pH monitoring in sweat [J]. *Talanta*, 2021, 222: 121502.
- [11] KHAN M R R, KHALILIAN A, SEO J, et al. Highly reliable passive RFID-based inductor-capacitor sensory system strengthened by solvatochromism for fast and wide-range lactate detection [J]. *IEEE Sensors Journal*, 2022, 22(12): 12228-12236.
- [12] AKINOSHIO T D, OYEDELE L O, BILAL M, et al. A scalable deep learning system for monitoring and forecasting pollutant concentration levels on UK highways [J]. *Ecological Informatics*, 2022, 69: 101609.
- [13] LI L, ZHANG R X, SUN J D, et al. Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm [J]. *Journal of Environmental Health Science and Engineering*, 2021, 19(1): 401-414.
- [14] JIA M S, GABRYS B, MUSIAL K. A network science perspective of graph convolutional networks: a survey [J]. *IEEE Access*, 2023, 11: 39083-39122.
- [15] FU S C, LIU W F, ZHOU Y C, et al. HpLapGCN: hypergraph p-Laplacian graph convolutional networks [J]. *Neurocomputing*, 2019, 362: 166-174.
- [16] JIN G Y, LIANG Y X, FANG Y C, et al. Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(10): 5388-5408.
- [17] WU Z H, PAN S R, CHEN F W, et al. A Comprehensive Survey on Graph Neural Networks [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(1): 4-24.
- [18] LI G H, MÜLLER M, QIAN G C, et al. DeepGCNs: Making GCNs Go as Deep as CNNs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 6923-6939.
- [19] GULICH R, KÖHLER M, LUNKENHEIMER P, et al. Dielectric Spectroscopy on Aqueous Electrolytic Solutions [J]. *Radiation and Environmental Biophysics*, 2009, 48(1): 107-114.
- [20] CHEN T, HEFTER G, BUCHNER R. Dielectric Spectroscopy of Aqueous Solutions of KCl and CsCl [J]. *The Journal of Physical Chemistry A*, 2003, 107(20): 4025-4031.
- [21] FENG R H, LIN X Y. RFID-based real-time salt concentration monitoring with adaptive EKF [J]. *Sensors*, 2025, 25(12): 3826.
- [22] FU S C, LIU W F, ZHANG K, ZHOU Y C. Example-Feature Graph Convolutional Networks for Semi-Supervised Classification [J]. *Neurocomputing*, 2021, 461: 63-76.
- [23] SEMLANI Y, RELAN M, RAMESH K. PCN: A Deep Learning Approach to Jet Tagging Utilizing Novel Graph Construction Methods and Chebyshev Graph Convolutions[J]. *Journal of High Energy Physics*, 2024, 2024(7): 1-15.
- [24] VAN BELLE R, VAN DAMME C, TYTGAT H, DE WEERDT J. Inductive Graph Representation Learning for Fraud Detection[J]. *Expert Systems with Applications*, 2022, 193: 116463.
- [25] 吴海宁. 高盐化工废水处理工艺与参数优化的研究[D]. 广州: 华南理工大学, 2010.
- WU H N. Research on treatment technology and parameter optimization of high-salinity chemical wastewater [D]. Guangzhou: South China University of Technology, 2010.
- [26] GENNARELLI G, ROMEO S, SCARFI M R, et al. A microwave resonant sensor for concentration measurements of liquid solutions [J]. *IEEE Sensors Journal*, 2013, 13(5): 1857-1864.