

基于PCA-Informer⁺模型的周期性甲烷菌体浓度预测研究

杨帆^a, 毛腾跃^{a*}, 占伟^b

(中南民族大学 a. 计算机科学学院; b. 资源与环境学院, 武汉 430074)

摘要 厌氧发酵是一种重要的生物技术,然而现有的检测技术无法实时测量发酵周期中甲烷菌体的浓度值,导致难以准确监测发酵状态.针对此问题,提出了基于改进Informer深度学习模型的周期性甲烷菌体浓度预测方法.首先,基于Informer构建基础预测模型;其次,利用PCA主成分分析,将特征变量从8维降低至3维,优化模型的输入,提高预测效率;然后,根据周期中每个时间点的重要性建模设计WeightedMSELoss损失函数,以更好地适应周期性甲烷菌体浓度预测任务;最后,融合特征变量、位置编码和周期编码,提高模型捕获长期依赖的能力.实验结果表明:Informer相较于长短期记忆网络(LSTM)、循环神经网络(RNN)、门控循环单元(GRU)在周期性甲烷菌体浓度预测任务上效果最好,且基于Informer改进的PCA-Informer⁺模型,平均绝对误差(MAE)、均方根误差(RMSE)相较于原Informer模型分别下降了26%、11%,模型效率提高了18%,实现了较为快速准确的甲烷菌体浓度预测.

关键词 Informer模型;菌体浓度预测;主成分分析

中图分类号 TP391.4 文献标志码 A 文章编号 1672-4321(2025)03-0393-07

doi:10.20056/j.cnki.ZNMDZK.20250313

Prediction of periodic methane concentration based on PCA-Informer⁺

YANG Fan^a, MAO Tengyue^{a*}, ZHAN Wei^b

(South-Central Minzu University, a. College of Computer Science; b. College of Resources and Environment, Wuhan 430074, China)

Abstract Anaerobic fermentation is an important biological technology, but the existing detection technology can not measure the concentration of methane bacteria in the fermentation cycle in real time, which makes it difficult to accurately monitor the fermentation state. To solve this problem, a periodic methane concentration prediction method based on improved Informer deep learning model was proposed. Firstly, the basic prediction model is constructed based on Informer. Secondly, based on PCA, the feature variables are reduced from 8 to 3 dimensions to optimize the input of the model and improve the prediction efficiency. Thirdly, according to the importance of different time steps in the cycle, the WeightedMSELoss loss function is designed to better adapt to the periodic methane culture task. Finally, feature variables, location coding and period coding are integrated to improve the ability of the model to capture long-term dependencies. The experimental results show that: Compared with long short-term memory network (LSTM), recurrent neural network (RNN) and gate recurrent unit (GRU), Informer has the best performance in predicting periodic methane concentration. Moreover, The PCA-Informer⁺ model is based on Informer improvements, with MAE and RMSE reduced by 26% and 11% respectively, and model efficiency increased by 18%, achieving relatively fast and accurate prediction of methane concentration.

Keywords Informer model; prediction of methane concentration; Principal Component Analysis

收稿日期 2024-02-28

* 通信作者 毛腾跃(1974-),男,副教授,博士,研究方向:空间通信、智能控制,E-mail:734813320@qq.com

基金项目 国家民委中青年英才培养计划(MZR20007);湖北省技术创新计划重点研发专项(2023BAB087);新疆维吾尔自治区区域协同创新专项(科技援疆计划)(2022E02035)

厌氧发酵在当今工业中具有重要意义,厌氧发酵为可再生能源生产提供了途径,同时有助于减少废物对环境的影响.江苏连云港某生物化工有限公司的主产品是工业酒精,利用厌氧发酵技术,该公司将废料转化为沼气作为副产品并用于发电.为了进一步缩短厌氧发酵周期,提高经济效益,我校团队与该公司合作,搭建了IBB(IBioBooster)生物智能强化控制系统(以下统称IBB系统).IBB系统的工作原理为:抽取少量大反应池中的废料液到IBB系统反应罐中,保证IBB系统反应罐与大反应池中废料成分相同.将甲烷菌种投入到IBB系统反应罐中,并加入适量的营养物质,通过控制IBB系统反应罐的温度、pH、搅拌速率等来营造一个甲烷菌体繁殖速率最快的环境.当甲烷菌体浓度达到最大时,将IBB系统反应罐中菌体培养液投放回大反应池,此时甲烷菌体浓度较大,迅速进行厌氧发酵产生沼气.然而,由于甲烷菌体浓度只能离线测量,且测量结果具有滞后性,目前在IBB系统培养甲烷菌体培养液的过程中,只能凭人工经验来判断菌体浓度达到最大的时间.

随着人工智能的发展,近些年基于机器学习和深度学习模型的预测方法广泛应用于各个领域.在生物发酵领域,如黄慧玲等^[1]提出利用机器学习构建海水浴场粪便污染指示细菌的浓度预测模型的想法,但其仅提出想法而未进行实际预测模型搭建工作;赵亭亭等^[2]以牛粪发酵产甲烷过程为研究对象,提出一种厌氧发酵产气预测混合模型,该混合模型仿真下的甲烷产率与实验数据吻合良好,短期预测准确率达90%以上,但在预测未来较长时间的数据时,准确率较低;高学金等^[3]提出一种基于注意力LSTM的多阶段发酵过程质量预测方法,并将该方法应用到青霉素发酵仿真数据.在其他领域,董浩等^[4]基于Informer构建了PM2.5浓度预测模型,并将该模型与LSTM和RNN模型进行对比,实验证明其提出的基于Informer的PM2.5预测模型效果最好,预测准确率约93%;WEI等^[5]基于Informer构建了风电功率预测模型,该模型选取风速、环境温度等作为特征变量,预测准确率达90%以上;WANG等^[6]通过将卷积网络与Informer模型相结合,提出了船舶运动姿态数据预测方法,预测准确率在95%左右.虽然上述几位学者所做工作不属于生物厌氧发酵领域,但是其所做工作与生物发酵过程中的菌体浓度预测同属于多维长时间序列预测任务,证明了Informer模型在长时间序列预测任务中的优异表现,同样也为生物菌体浓度预测提供了方法指导.

本研究目的在于高效准确地预测厌氧发酵过程中甲烷菌体浓度,在Informer模型的基础上,首先通过添加PCA降维来提高模型预测效率,然后更换根据甲烷培养周期中每个时间点的重要性来建模的WeightedMSELoss损失函数,最后利用特征融合,融合特征变量、位置编码和周期编码,共同作为模型的输入.以上操作使得改进后的模型更适用于本文周期性甲烷菌体浓度预测任务.

1 Informer模型

北京航空航天大学团队针对Transformer在长时间序列预测任务上存在的三个问题进行改进得到了一种多维长时间序列预测模型^[7],名为Informer,其模型结构如图1所示.

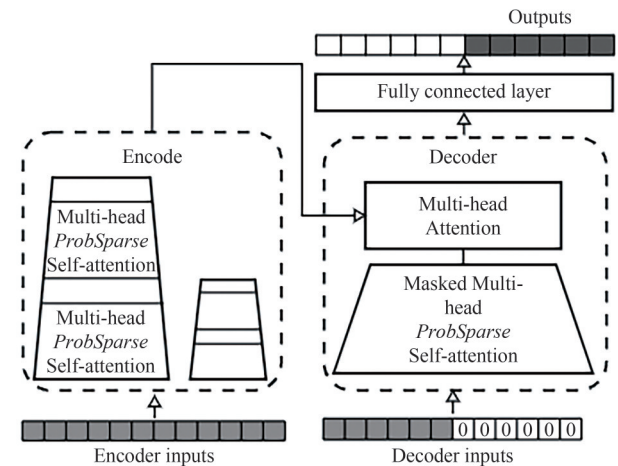


图1 Informer模型结构

Fig. 1 The structure of Informer

针对Transformer在长时间序列预测任务中自注意力机制需要计算每一个时刻与其他时刻的相似程度^[8],导致时间复杂度较高的问题,提出了概率稀疏自注意力机制(ProbSparse Self-attention)^[9],其公式见式(1).式中, $A(Q, K, V)$ 表示计算得到的注意力, Q, K, V 为通过输入特征变量线性变换得到的大小相同的稀疏矩阵, \bar{Q} 与 Q 大小相同,但只包含了利用概率选取的重要性较高的 $\log L$ 个时刻的数据.该方法将时间复杂度由 $O(N^2)$ 降低为 $O(L \log L)$.

$$A(Q, K, V) = \text{Soft max} \left(\frac{\bar{Q} K^T}{\sqrt{d}} \right) V \quad (1)$$

针对Transformer在长时间序列预测任务中内存占用过大的问题,提出了自注意力蒸馏机制(Self-attention Distilling)^[10-11],通过添加一维卷积层、ELU激活函数、最大池化层逐渐缩小每一层的数据规模.

Self-attention Distilling 的公式见式(2),其中 X_j^t 表示第 j 层,第 t 个时刻的 attention 值,MaxPool 为最大化操作.

$$X_{j+1}^t = \text{MaxPool}(\text{ELU}(\text{Conv1d}([X_j^t]_{AB}))) \quad (2)$$

针对 Transformer 解码器的输入依赖于编码器的输出^[8],每次预测一个值,逐步预测速度慢的问题,提出了一次性预测未来所有值(One Forward Operation).原理为将编码器输入数据中的部分值与全 0 占位符拼接,全 0 占位符的长度与需要预测的数据长度相等,将拼接序列输入到解码器中一次输出所有预测值.

综上,Informer 模型针对 Transformer 的改进使得它不仅预测效率更高,而且对长时间序列的预测也更准确.IBB 系统中甲烷菌体浓度预测任务,属于长时间序列预测任务,故本文选取 Informer 作为基础模型来进行甲烷菌体浓度预测研究.

2 基于 Informer 改进的 PCA-Informer⁺模型

2.1 PCA-Informer⁺模型结构

针对甲烷菌体浓度预测任务,对原 Informer 模型进行三处改进得到 PCA-Informer⁺模型,其结构如图 2 所示.第一,由于甲烷菌体培养过程中特征参数较多,且某些特征可能包含冗余信息等问题提出了利用 PCA 主成分分析方法对数据进行降维操作,将特征变量从 8 维降至 3 维;第二,根据甲烷菌体培养周期中每个时间点的重要性建模,设计 WeightedMSELoss 损失函数;第三,融合特征变量、位置编码、周期编码,共同作为模型的输入,提高模型对本文预测任务的拟合度.

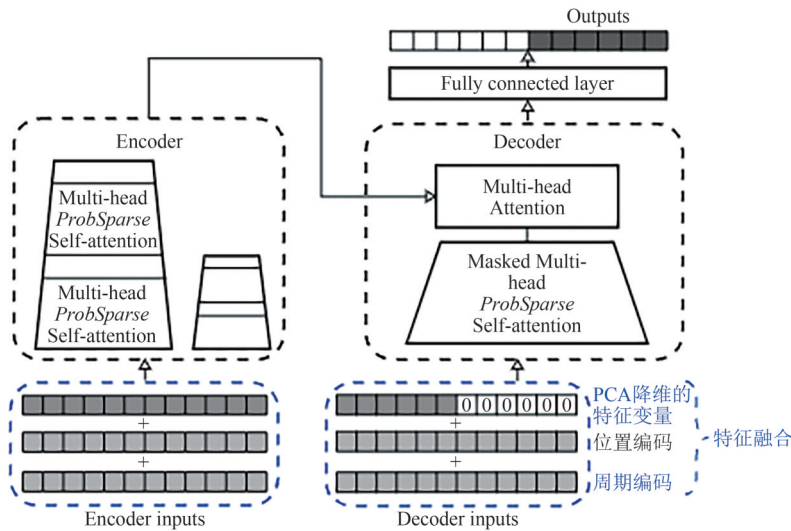


图 2 PCA-Informer⁺模型结构

Fig. 2 The structure of PCA-Informer⁺

2.2 PCA 降维

在甲烷菌体浓度预测任务中,由于特征变量数量较多,某些特征可能包含冗余信息等问题,利用 PCA 主成分分析方法对数据进行降维处理,从而在保留大部分环境参数信息的同时减少特征的数量^[12-13],优化模型的输入.利用 PCA 降维的步骤如图 3 所示:

(1)数据标准化:对甲烷菌体培养过程中的碳氮比、温度、PH 等 8 个特征变量进行标准化处理,使得每个特征的均值为 0,方差为 1,以确保所有特征变量具有相同的尺度,消除不同特征之间的量纲差异.

(2)协方差矩阵计算:计算标准化后的多维时间序列数据的协方差矩阵.

(3)特征值分解:PCA 对协方差矩阵进行特征

值分解,求出协方差矩阵的特征值($\lambda_1, \lambda_2, \dots, \lambda_d$)和对应的特征向量(v_1, v_2, \dots, v_d).特征向量 v 表示数据中的主成分,而特征值 λ 表示每个主成分的重要性.

(4)主成分选择:根据特征值的大小,可以选择保留最重要的主成分,从而降低数据的维度,并以特征值的比例计算每个主成分的贡献率.设定阈值 90%,选择累计贡献率大于阈值的前 k 个主成分.计算得出:主成分 1-3 的贡献率分别为 64.5%、22.4%、7.9%,前 k 个主成分累计贡献率如图 4 所示,前三个主成分累计贡献率为 94.8% 大于阈值 90%,且其余主成分贡献率均小于 5%,因此,保留前 3 个主成分作为模型的输入.

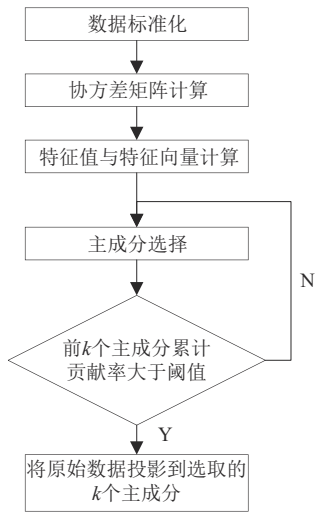


图3 基于PCA主成分分析降维流程图

Fig. 3 Flow chart based on PCA dimension reduction

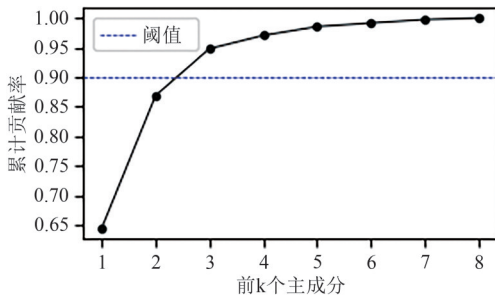


图4 前k个主成分累计贡献率

Fig. 4 The cumulative contribution rate of the first k principal components

(5)数据投影:将原始数据投影到所选的主成分上来进行降维.这将生成一个新的低维表示,其中每个样本都由主成分的线性组合构成.

2.3 WeightedMSELoss 损失函数

Informer 模型的原损失函数为 MSE, 见式(3). MSE 损失函数常用于回归分析任务,用来衡量模型预测值与实际值之间的平方误差,对于连续数值的预测具有较好的表现.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2, \quad (3)$$

将 Informer 模型原损失函数替换为新设计的 WeightedMSELoss, 见式(4). 该损失函数根据甲烷菌体培养周期中每个时间点的重要性来建模, 根据时间点在当前周期中的位置来设置不同的权重, 对不同时间步的预测误差进行加权, 以更好地反映长期依赖关系.

$$WeightedMSELoss = \frac{1}{N} \sum_{i=1}^N (weight_i \times (Y_i - y_i)^2), \quad (4)$$

式中: $weight_i$ 为权重值, Y_i 为甲烷菌体浓度的预测值, y_i 为甲烷菌体浓度的真实值, N 为测试集样本总数量.

甲烷菌体浓度随时间(逐小时)的变化曲线见图5. 分析可知: 每个培养周期甲烷菌体浓度变化趋势基本相同. 甲烷菌体培养过程可以划分为四个时期: 延滞期、对数期、稳定期和取菌期. 延滞期为刚开始培养, 菌种适应新环境, 此阶段菌体浓度基本不变. 对数期, 菌体活跃度较高, 菌体数量呈现对数增长. 稳定期, 菌体浓度达到最大, 由于细胞的溶解作用及营养物质等条件的制约, 菌体浓度呈缓慢降低趋势. 取菌期, 当甲烷菌体培养达到稳定期后, 便将 IBB 系统培养罐中的菌体培养液输送回大反应池, 此时 IBB 系统培养罐中菌体浓度急剧下降.

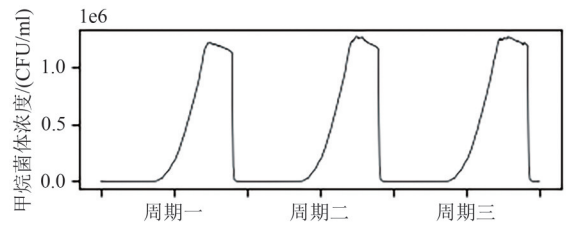


图5 甲烷菌体浓度变化曲线

Fig. 5 Methane concentration curve

在 WeightedMSELoss 损失函数设计中, 延滞期时间步权重最低, 对数期时间步权重值最高. WeightedMSELoss 权重设计方法为: 用 1 初始化权重向量, 将对数期权重调整为 2, 将稳定期权重调整为 1.5, 然后归一化权重向量, 使得权重之和为 1.

2.4 特征融合

PCA-Informer⁺模型基于 Informer 模型原有的时间特征提取部分进行改进, 针对甲烷培养周期进行深入分析, 并添加了专门针对周期性特征的提取方法. 利用 DataEmbedding 函数, 对三个角度进行特征融合, 如图6所示:

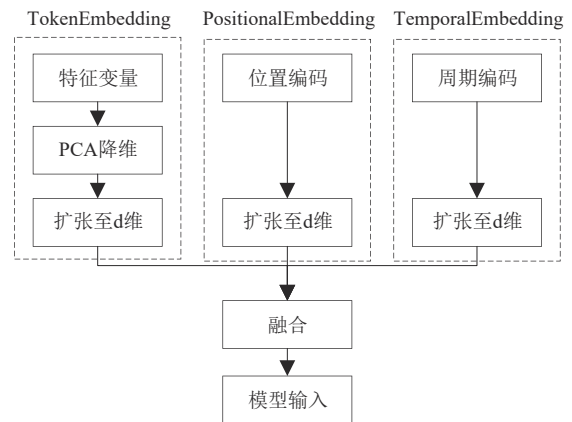


图6 特征融合流程图

Fig. 6 Feature fusion flow chart

(1)TokenEmbedding: 用来将 PCA 降维后的特征变量序列转化为向量表示, 并使用 1 个一维卷积

层将其从3维扩张到 d_{model} 维.

(2)PositionalEmbedding:甲烷菌体浓度预测属于多维时间序列预测问题,为了保证序列数据输入到模型以后顺序关系不丢失,对其位置信息进行编码,然后将位置编码也扩张至 d_{model} 维.PositionalEmbedding函数中使用 \sin 和 \cos 函数来生成位置编码,公式见式(5-6).式中:position为位置, j 表示维度, d_{model} 表示扩张的维度.

$$\text{PE}_{(\text{position}, 2j)} = \sin \frac{\text{position}}{(2L_x)^{2j/d_{\text{model}}}}, \quad (5)$$

$$\text{PE}_{(\text{position}, 2j+1)} = \cos \frac{\text{position}}{(2L_x)^{2j/d_{\text{model}}}}. \quad (6)$$

(3)TemporalEmbedding:在甲烷菌体培养过程中,每个周期的菌体浓度变化趋势接近,对周期信息进行编码,并将周期编码也扩张至 d_{model} 维.

将上述三步得到的同维度的特征变量数据、位置编码和周期编码相加.最后对所有嵌入信息的总和和应用1个dropout层,防止过拟合,将其正则化输出作为预测模型的输入.

3 实验与结果分析

3.1 数据来源

学校团队在江苏连云港某生物化工有限公司搭建了IBB系统,实验选取该IBB系统的历史数据作为数据集,这些数据包括甲烷菌体发酵过程中的菌体浓度和发酵罐环境参数,如碳氮比、温度、pH、溶解氧浓度等8个特征变量.IBB系统每一小时记录一条数据,共计4800组数据,每240组数据为一个甲烷菌体培养周期,共计20个周期.

3.2 数据预处理

在数据预处理阶段,分为三步:首先,在数据采集过程中,可能会出现数据丢失等意外情况导致数据集中某些值为空,分析得知甲烷菌体浓度预测任务中,菌体浓度和环境培养参数均为线性变化的,故针对空值采取线性插值法处理;其次,若系统传感器发生故障等意外可能会导致记录的数据与真实值不符,对于这些异常值,使用前后两个数据的平均值来代替异常;最后由于数据中不同特征变量之间数值相差较大,这些尺度差异可能会导致模型在训练过程中受到数据范围较大的特征的影响,从而影响模型的性能.因此,首先需要对数据进行Z-Score标准化操作^[14],其公式如式(7),这种标准化方法使得数据的分布具有均值为0和标准差为1的特

性,有助于模型更好地处理数据并加速收敛.

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (7)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}, \quad (8)$$

式(1)中 x_i 为原始数据, μ 为原始数据的均值, σ 为原始数据的标准差, σ 的计算公式见式(8).

将标准化后的数据按顺序以7:1:2的比例划分为训练集(3360组)、验证集(480组)和测试集(960组).

3.3 实验环境

本实验基于windows系统实现,硬件配置:CPU为13th Gen Intel(R) Core(TM) i5-13490F 2.50 GHz,显卡为NVIDIA RTX 4070,内存32 G;软件配置:系统版本为Windows10专业工作站版,编程语言为Python 3.11,深度学习框架为Pytorch 2.0.1.

3.4 模型性能评价指标

为了更加合理的对比不同的模型性能,本文采用平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Squared Error, RMSE)作为模型评价的指标.其中,MAE衡量了预测值与真实值之间的平均绝对差异,能够客观地反映出模型在每个时间点上的平均预测误差.RMSE考虑了预测值与真实值之间的平均平方差异,并且通过取平方根将其转换为与目标变量相同的单位^[15].本文MAE与RMSE的值均为在进行逆标准化操作之前计算得到,MAE和RMSE的计算公式见式(9-10):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |Y_i - y_i|, \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2}, \quad (10)$$

式中: Y_i 为甲烷菌体浓度的预测值, y_i 为甲烷菌体浓度的真实值, N 为测试样本总数量.

3.5 原Informer模型与其他模型对比分析

3.5.1 评价指标分析

使用Python和Pytorch框架分别搭建基于RNN、LSTM、GRU、Informer网络模型^[16-17]的IBB系统甲烷菌体浓度预测模型.表1为基于4种预测模型的MAE、RMSE对比结果.

可以看出:在当前甲烷菌体浓度预测任务中,Informer模型相比于其它三种时间序列预测的模型,取得了最好的预测效果,其MAE与RMSE均显著小于其余三种模型.但是由于Informer其复杂的网络结构和添加的多头注意力机制,使得Informer

表1 模型评价指标横向对比

Tab. 1 Horizontal comparison of model evaluation indexes

预测模型	MAE	RMSE	训练周期	训练时间/s
RNN	0.218	0.356	20	10
LSTM	0.167	0.309	20	20
GRU	0.204	0.343	20	15
Informer	0.119	0.195	15	85

注:加粗字体为当前指标最优项。

模型训练时间显著长于其余三种模型。综上,在利用 Informer 来实现高准确率预测的同时,还需要优化 Informer 模型结构,提高预测效率。

3.5.2 预测值与真实值拟合效果分析

测试集包含 960 组数据,共计 4 个培养周期,为了突出不同模型之间的差异,更清晰地对比预测值与真实值,选择将第一个周期的预测值与真实值进行可视化,可视化对比结果见图 7。由图可知:基于 Informer 的预测模型相比于其他 3 个模型,预测效果最优。得益于 Informer 模型全局和局部注意力机制、长短期建模等特性,在周期后段,Informer 预测值与真实值的拟合效果具有显著优势。

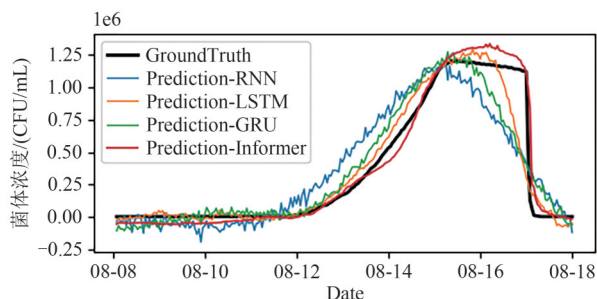


图7 Informer模型与其他模型预测结果可视化对比图

Fig. 7 Visual comparison of prediction results between Informer and other models

3.6 PCA-Informer⁺模型、Informer⁺、原 Informer 模型对比分析

利用 Python 语言和 Pytorch 深度学习框架搭建原 Informer、Informer⁺、PCA-Informer⁺模型。模型训练阶段设置早停机制,连续 3 次验证集 loss 值不再降低则停止训练。表 2 为原 Informer、Informer⁺、PCA-Informer⁺模型的 MAE、RMSE、训练周期、训练时间对比结果。

表2 模型评价指标纵向对比

Tab. 2 Longitudinal comparison of model evaluation indexes

预测模型	MAE	RMSE	训练周期	训练时间/s
原 Informer	0.119	0.195	15	85
Informer ⁺	0.087	0.168	14	80
PCA-Informer ⁺	0.088	0.173	12	69

注:加粗字体为当前指标最优项。

可以看出:通过更换 WeightedMSELoss 损失函数、添加周期时间特征提取方法的 Informer⁺模型相较于原 Informer 模型,在甲烷菌体浓度预测任务上准确率更高;在 Informer⁺模型基础上添加 PCA 降维方法的 PCA-Informer⁺模型,其预测准确率与 Informer⁺几乎相同,但是训练周期与训练时间均有减少,证明了降维后的数据加速了模型的收敛,提高了模型的预测效率。综上,本文提出的 PCA-Informer⁺模型在甲烷菌体浓度预测任务上,相较于原 Informer 模型 MAE、RMSE 分别降低了约 26%、11%,训练速度提高了约 18%。

4 结论

本文以江苏连云港某生物化工有限公司的 IBB 系统培养甲烷菌体作为研究对象,探索了一种利用深度学习来预测甲烷菌体浓度的方法,提出了 PCA-Informer⁺菌体浓度预测模型,并与其他传统时间序列预测模型及原 Informer 模型进行对比实验,得出以下结论:

(1)所有模型均可预测出甲烷菌体浓度的整体变化趋势,证明了在厌氧发酵过程中利用深度学习解决问题的可行性。

(2)经过 Informer 模型与其他传统时间序列预测模型进行横向对比,得出:得益于 Informer 模型通过全局和局部注意力机制、长短期建模等特性,Informer 在长时间序列预测任务中效果显著优于其他模型。

(3)基于 Informer 模型改进的 PCA-Informer⁺模型,进一步提高了甲烷菌体浓度预测的准确率,并提升了模型的效率。

参考文献

- [1] 黄慧玲,明红霞,樊景凤,等.海水浴场中粪便污染指示细菌的浓度预测模型研究进展[J].海洋环境科学,2020,39(6):953-959.
- [2] 赵亭亭,杨世品,李丽娟,等.基于机理和数据驱动建立牛粪厌氧发酵产气预测混合模型[J].生物化学工程,2023,57(3):39-48.
- [3] 高学金,孟令军,高慧慧.基于注意力 LSTM 的多阶段发酵过程集成质量预测[J].控制与决策,2022,37(3):616-624.
- [4] 董浩,孙琳,欧阳峰.基于 Informer 的 PM2.5 浓度预测[J].环境工程,2022,40(6):48-54,62.
- [5] WEI H, WANG W S, KAO X X. A novel approach to

- ultra-short-term wind power prediction based on feature engineering and informer[J]. *Energy Reports*, 2023, 9: 1236-1250.
- [6] WANG Y, DOU Y, PENG C, et al. Multi step prediction method of ship pitch based on Conv-Informer model[C]// *OCEANS 2023*, Limerick: IEEE, 2023: 1-6.
- [7] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(12): 11106-11115.
- [8] HAN K, XIAO A, WU E, et al. Transformer in transformer[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 15908-15919.
- [9] XU Z. A solar irradiance prediction method based on signal decomposition and Informer[C]//2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML). Chengdu: IEEE, 2023: 1132-1136.
- [10] WANG W, WEI F, DONG L, et al. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers[J]. *Advances in Neural Information Processing Systems*. 2020, 33: 5776-5788.
- [11] ZHANG Y, XU B, ZHAO T. Convolutional multi-head self-attention on memory for aspect sentiment classification [J]. *IEEE/CAA Journal of Automatica Sinica*, 2020, 7(4): 1038-1044.
- [12] REHMAN A, KHAN A, ALI M A, et al. Performance analysis of PCA, sparse PCA, kernel PCA and incremental PCA algorithms for heart failure prediction [C]//2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). Istanbul: IEEE, 2020: 1-5.
- [13] MARUKATAT S. Tutorial on PCA and approximate PCA and approximate kernel PCA [J]. *Artificial Intelligence Review*, 2023, 56(6): 5445-5477.
- [14] AGGARWAL V, GUPTA V, SINGH P, et al. Detection of spatial outlier by using improved Z-score test [C]// 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli: IEEE, 2019: 788-790.
- [15] HODSON T O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not [J]. *Geoscientific Model Development*, 2022, 15 (14): 5481-5487.
- [16] SUGIYARTO A W, ABADI A M. Prediction of Indonesian palm oil production using long short-term memory recurrent neural network (LSTM-RNN) [C]// 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS). Ipoh: IEEE, 2019: 53-57.
- [17] YANG S, YU X, ZHOU Y. LSTM and GRU neural network performance comparison study: Taking yelp review dataset as an example [C]//2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAL). Shanghai:IEEE, 2020: 98-101.

(责编&校对 雷建云)