Prediction Method of Dissolved Gas Concentration in Transformer Oil Based on CEEMD-TCN Model*

YANG Haijing, SUN Yunquan^{*}, ZHU Wei, QIAN Yao, JIN Hao

(College of Electrical and Information Engineering, Jiangsu University, Zhenjiang Jiangsu 212013, China)

Abstract: The concentration of dissolved gas in transformer oil is an important indicator for understanding the operating status of the transformer and judging whether the transformer is faulty. In view of the non-linear and non-stationary characteristics of the dissolved gas concentration sequence in the oil, the data training model directly will significantly reduce the prediction accuracy, so a prediction method based on the combination of complete ensemble empirical mode decomposition and temporal convolutional network is proposed. First, the original sequence is decomposed into sub-sequence components of different scales, the temporal convolutional network is trained after preprocessing, and the network hyperparameters are optimized, and the prediction results of each component are superimposed and reconstructed to obtain the final prediction result. Experimental verification shows that the model has small prediction errors and high prediction accuracy.

Key words: complete ensemble empirical mode decomposition; temporal convolutional network; dissolved gas concentration in transformer oil; prediction

EEACC:7210

doi:10.3969/j.issn.1005-9490.2021.04.022

基于 CEEMD-TCN 模型的变压器油中 溶解气体浓度预测方法*

杨海晶,孙运全*,朱 伟,钱 尧,金 浩 (江苏大学电气信息工程学院,江苏镇江 212013)

要:变压器油中溶解气体浓度是了解变压器运行状态、判断变压器是否发生故障的重要指标。针对油中溶解气体浓度 序列非线性、非平稳性的特点、数据直接训练模型会明显降低预测精度,因而提出了一种基于互补集合经验模态分解和时间 卷积网络相结合的预测方法。首先,将原始序列分解成不同尺度的子序列分量,经过预处理后训练时间卷积网络,并优化网 络超参数,各分量的预测结果叠加重构从而获得最终预测结果。通过实验验证表明,该模型预测误差小,预测精度高。 关键词:互补集合经验模态分解:时间卷积网络:变压器油中溶解气体浓度:预测

中图分类号:TM910.6 文献标识码:A

电力变压器是构成现代电力系统关键设备之 一,确保变压器安全稳定的运行是电力设备领域的 重要课题^[1]。在变压器正常运行的进程中,设备会 产生老化或者发生电、热故障,会导致产生多种气 体。这些气体溶解在绝缘油中,其成分含量以及不 同气体间的比例关系可以反映出当前变压器的健康 状况^[2-3]。所以,对溶解在变压器油中的各类气体 浓度进行预测和分析,为变压器的运行状况的评估 和故障的诊断提供了重要依据。

迄今为止,大量国内外学者对变压器油中溶解

文章编号:1005-9490(2021)04-0887-06

气体浓度预测这一课题进行了研究,并取得了许多 重要成果。该课题的研究可以归纳为3个方向:统 计预测、智能预测和组合预测。

统计预测包含模糊预测模型[4]、灰度模型[5]和 时间序列预测模型[6]等,此类模型最终预测结果受 限于实验数据的分布规律。模糊预测对不确定事件 具有良好的预测效果;灰度模型和时间序列模型分 别在处理指数型数据和线性型数据时具有优势。

智能预测是指构建机器学习和深度学习的预测 模型,使用海量的历史数据样本进行训练,最终使其

项目来源:中国博士后面上基金项目(20110491358);江苏大学高级人才项目(13DG054) 收稿日期:2020-09-28 修改日期:2020-11-30

能够较好反映时间序列的发展趋势。其中常见的机 器学习模型有支持向量机、人工神经网络和决策树 等^[7]。由于传统的机器学习模型无法高效地处理 大量训练数据以及提取训练数据中的时序趋势特 征,影响其对油中溶解气体的预测精度。伴随着人 工智能技术的不断发展,深度学习中的循环神经网 络和卷积神经网络可以有效解决这一问题^[8]。循 环神经网络具有循环反馈网络结构,在时序预测问 题中具有较强的适应能力。卷积神经网络具有特殊 的卷积层结构,被广泛用应于图像识别。一些专门 的卷积神经网络模型可以较好地完成时间序列预 测,如时间卷积网络、WaveNet等。

组合预测的研究可以分成两个方向,一个方向 是使用多种不同算法对收集到的数据分别预测,最 终预测结果取决于各算法的预测结果以及各算法的 权重值,该方向的预测结果易受权重分配影响,导致 预测结果失真;另一个研究方向是使用序列处理方 法对油中溶解气体浓度序列进行预处理,再对经过 处理的子序列分量分别建模,将各个模型结果叠加 重构得到原始序列的组合预测结果。常用的序列预 处理方法有小波分解^[9]、经验模态分解^[10]等。

鉴于以上的研究背景,提出了一种基于互补集 合经验模态分解(Complete Ensemble Empirical Mode Decomposition, CEEMD)方法和时间卷积网络(Temporal Convolutional Network, TCN)的组合预测模型。 长期收集的变压器油中溶解气体浓度序列经过 CEEMD分解后得到的多组子分量和一个残余分量, 再通过卷积神经网络(Convolutional Neural Network, CNN)的专门针对序列预测的变种 TCN 分别进行训 练并得到预测结果,再对各组结果进行重构组合,最 终得到完整的序列预测趋势图像。根据实例所预测 的结果表明,相比于传统的各种预测模型,所提出的 CEEMD-TCN 模型预测精度更高。

1 互补集合经验模态分解

1.1 经验模态分解

经验模态分解(Empirical Mode Decomposition, EMD)方法被广泛应用于信号分析领域的非平稳、 非线性信号分析^[11]。在处理时间序列问题时,由于 实际序列受到多重复杂因素干扰,这些因素可能彼 此耦合,难以用一个精准的数学模型表示,因此直接 使用预测模型拟合误差大,预测效果准确度不高。 而 EMD 方法能够将原始序列以不同时间尺度分解 为一系列具有一定规律性的子序列,可以明显降低 对原始序列的拟合难度。 EMD 方法将复杂序列分解为若干固有模态函数(Intrinsic Mode Function, IMF)和一个残余分量,确定 IMF 需满足 2 个条件:(1)在序列中,极值点和 过零点的个数必须相等或者最多相差一个;(2)在 序列中任意时间点,由局部极大值点和极小值点所 形成的上、下包络线平均值为 0。

EMD 方法的步骤如下:

(1)设原始序列 $x(t) = \{x_1, x_2, \dots, x_i\}, 从中取 出局部极大值点和局部极小值点,将所有局部极大 值点和局部极小值点分别以三次样条曲线连接并拟 合成 <math>x(t)$ 的上、下包络线 $U_x(t), V_x(t)$ 。

(2) 求出 $U_x(t)$ 、 $V_x(t)$ 的平均值,记为 $m_1(t)$:

$$m_1(t) = \frac{U_x(t) + V_x(t)}{2}$$
(1)

(3)求取原始序列 x(t)与 $m_1(t)$ 之差 $h_1(t)$:

$$h_1(t) = x(t) - m_1(t)$$
 (2)

(4) 若 $h_1(t)$ 满足 IMF 分量的条件,则记为 x(t)的第一个 IMF 分量;若不满足,将 x(t) 替换成 $h_1(t)$,重复前三步,直到 $h_1(t)$ 满足 IMF 条件,则将 $h_1(t)$ 记为第一个 IMF 分量 $c_1(t)$:

$$c_1(t) = h_1(t)$$
 (3)

(5) 将原始序列 x(t) 减去第一个 IMF 分量 $c_1(t)$,结果记做残余分量 $r_1(t)$:

$$r_1(t) = x(t) - c_1(t)$$
 (4)

将 r₁(t) 作为原始序列, 重复以上(1) 到(5) 步骤, 得到第二个 IMF 分量, 重复 n 次, 得到第 n 个 IMF 分量, 有:

$$r_{n}(t) = r_{n-1}(t) - c_{n}(t)$$
(5)

当最终的残余分量 $r_n(t)$ 为常数或为单调函数 时停止迭代。序列 x(t) 被分解为 $n \land IMF$ 分量和 1 个残余分量 $r_n(t)$ 之和。

原始信号 x(t) 可以表示为:

$$x(t) = \sum_{i=1}^{n} c_i + r_n$$
 (6)

式中: c_i 为第i个 IMF 分量; r_n 为分解完成后的残余分量。

1.2 互补集合经验模态分解

EMD 方法分解序列难以解决组合分量的模态 混叠现象。模态混叠是在某一 IMF 中包含了相互 耦合、难以分解的其他时间尺度的 IMF 分量。为了 改善这一问题, Huang 和 Wu^[12]提出一种噪声辅助 数据分析的集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)。EEMD 是一种辅助信 号加入法,通过在序列分解的过程中多次添加高斯 白噪声并进行 EMD 分解,将多次分解的结果进行平 均,得到最终的 IMF。

然而,EEMD 方法添加的白噪声并不能真正地 完全抵消,会对数据的完备性造成影响。因此, Torres 等^[13]提出了一种互补集合经验模态分解,通 过对原始信号添加大小相同,符号相反的白噪声来 保证数据的完备性。

CCEMD 的步骤如下:

(1) 成对地给原始序列添加大小相同、符号相 反的白噪声:

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} S \\ N \end{bmatrix}$$
(7)

式中: M_1 、 M_2 分别为加入白噪声后的序列,S为原始 序列,N为添加的白噪声。

(2)采用 EMD 分解方法对添加白噪声后的 M₁、M₂序列进行分解,将得到第*i*个序列第*j*个 IMF 分量记 C_{ii}。

(3) 对所有得到 IMF 分量进行平均得到最后分 解结果:

$$IMF_{j} = \frac{1}{2n} \sum_{i=1}^{2n} C_{ij}$$
 (8)

式中:IMF; 表示得到的第 j 组 IMF 分量。

2 时间卷积网络

时间卷积网络是一种新型的时间序列预测算法。 TCN采用一维卷积网络,其架构中包括了膨胀因果卷 积(Dilated Causal Conv)和残差模块(Residual block)^[14]。相较于长短期记忆网络(Long Short-Term Memory,LSTM)和门控循环单元(Gated Recurrent Unit,GRU),TCN有效地避免了梯度消失和梯度爆炸 的问题,具有占用内存低、梯度稳定、并行性良好和感 受野灵活的优点。

TCN 的卷积层结构如图 1 所示,每一层的 t 时 刻的结果都依赖于上一层的 t,t-1,...,0 时刻的值, 而每一层对上一层信息的提取,都是通过间隔采样 的形式,逐层膨胀系数以 2 的指数增长。因此 TCN



可以用较少的层,获得很大的感受野。由于膨胀卷 积的特性,每一层都需要做 Padding 操作来对边缘 进行填充,以此保证不丢失上一层的信息。

由图 1 中的膨胀因果卷积结构可知,通过调整 卷积核大小、模型深度、膨胀系数等参数即可实现对 指定长度序列的整体感知。设输入的一维序列 *x* ∈ *R*ⁿ,卷积核 *f*:{0,…,*k*-1}→*R*,对该序列元素 *s* 经过 膨胀因果卷积运算 *F* 的表达式为:

$$F(s) = (x * {}_{d} f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d-i}$$
(9)

式中:d为膨胀系数;k为卷积核大小;" * $_{d}$ "表示卷积运算。

TCN的残差模块如图 2 所示,输入的数据需要 通过两轮的膨胀因果卷积、权重归一化、激活函数和 Dropout。为了增加 TCN 的准确率,引入了残差卷积 的跳层连接和 1×1 卷积操作。



3 CEEMD-TCN 组合预测模型

提出一种基于 CEEMD-TCN 组合预测模型,其 建模的一般步骤为:

(1)对油中溶解气体浓度数据预处理,剔除异 常数据,使用均值法填充空缺数据,构成完整的原始 时间序列;

(2)对原始序列进行分解,以降低其非线性、非 平稳性对模型所产生的不利影响。通过 CEEMD 方 法处理序列,得到多组 IMF 分量和一组残余分量;

(3)分别对各个子序列数据归一化,采用窗口 滑动的处理方式,以被选定采样长度的时间点数据 为特征,下一个时间点数据为标签,构建 TCN 预测 模型,使用优化算法来调整预测模型的超参数;

(4)使用训练集训练经过优化的 TCN 预测模型, 通过单步预测,即可得到各个子序列的预测结果,将 所有分量的结果叠加重构后得到最终预测结果。

其中,CEEMD-TCN 预测的流程如图 3 所示:



图 3 CEEMD-TCN 预测模型结构图

4 实例分析

4.1 预测模型评价指标

为了更好地评定模型的预测效果,使用平均绝对 百分比误差(Mean Absolute Percentage Error, MAPE) 和均方根误差(Root Mean Square Error, RMSE)作为 评价指标,公式为:

$$e_{\text{MAPE}} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|x_i' - x_i|}{x_i}$$
(10)

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x'_{i} - x_{i})^{2}}$$
(11)

4.2 数据集划分

为了验证 CEEMD-TCN 预测模型的准确性,选 用某 330 kV 变压器油色谱在线监测数据为例,以 2015 年 3 月 16 日至 2016 年 8 月 23 日监测周期为 24 h 的监测数据为数据样本,共计 527 组。将 2015 年 3 月 16 日至 2016 年 7 月 4 日的 477 组监测数据 作为训练集,将 2016 年 7 月 5 日至 2016 年 8 月 23 日的 50 组监测数据作为测试集。

本次实验所使用的电脑配置为 Inter Core i7-6700HQ 处理器,8GB 内存。仿真平台为 MATLAB 2018b 版,训练环境是 Python 3.7.6。 变压器油色谱在线数据由于变压器的稳定性以 及可能存在的外界因素干扰,常常会导致产生异常 数据,因此需要预先剔除异常值,使用均值法填充空 缺数据,以确保预测模型精度。选用 H₂ 浓度监测 序列为原始序列展开分析。经过预处理的原始序列 如图 4 所示:



4.3 数据处理

采用 CEEMD 方法对 H₂ 原始序列进行分解,依 次分离可得到 8 组 IMF 分量和 1 组残余分量。分解 结果如图 5 所示:

为了提高 TCN 预测模型的训练效率,需要针对 各个子序列做归一化处理。数据归一化选用离差标



准化公式:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

式中:*x*_{max},*x*_{min}分别为样本序列中的最大值和最小值,*y*_i为归一化后的数据。

4.4 实验结果与分析

各子序列分量的预测模型可以使用训练集数据 对 TCN 模型进行训练得到。模型的超参数通过 Adam 优化算法更新,学习率 lr 大小设置为 0.001, batch_size 设置为 36,dropout_rate 设置为 0.05,最大 迭代周期数为 100。

使用预测模型对未来 50 d 进行预测,即可获得各 子序列分量的预测结果。将预测结果叠加重构,最终 所得到的数据即为 H,浓度在未来 50 d 的预测结果。

为了验证模型的有效性,使用同一训练集分别训 练 CEEMD-TCN、TCN、CEEMD-BPNN 和 CEEMD-RNN 模型。其中,BPNN 指反向传播神经网络(Back Propagation Neural Network, BPNN),RNN 指循环神经网络 (Recurrent Neural Network,RNN)。神经网络模型采 用 Adam 优化算法更新超参数,设置隐藏层层数为1, 隐藏层神经元个数为 100,学习率 lr 为 0.001,迭代周 期数为 100。实际值和各模型预测值曲线如图 6 所示:



图 6 CEEMD-TCN 模型 H₂ 浓度预测及其与 TCN、 CEEMD-BPNN、CEEMD-RNN 模型的对比

从图 6 中可以看出, CEEMD-TCN 模型预测结 果相较于 CEEMD-BPNN 和 CEEMD-RNN 模型明显 波动较小,具有较好的预测效果。而采用 CEEMD 方法分解的 TCN 模型预测结果对实际值拟合精度 要优于未采用分解的模型。图 7 为 CEEMD-TCN 模 型与其他预测模型的相对误差对比:



从图 7 的预测模型相对误差对比结果可以看出,CEEMD-TCN 模型的相对误差在[-2,2]这一区间内波动,而其他预测模型的相对误差的波动范围则在[-5,5]区间内,这一结果表明 CEEMD-TCN 模型的预测精度明显要高于其他几种模型。

使用 e_{RMSE}和 e_{MAPE}来评估测试样本预测误差指标,评估结果如表1 所示:

表1 CEEMD-TCN 与其他模型预测结果评估

空测程制	误差类型		
顶侧侠型 -	e_{RMSE}	$e_{ m MAPE}/\%$	
CEEMD-TCN	0.584	0.389	
TCN	1.357	0.822	
CEEMD-BPNN	1.591	1.062	
CEEMD-RNN	1.117	0.765	

结合表 1 中数据可知, 在对油中溶解的 H₂ 浓度的预测结果中, 相较于直接使用 TCN 模型预测, 通过采用 CEEMD 分解原始序列的方法, 使得 e_{RMSE} 和 e_{MAPE} 指标分别降低 0.773 和 0.433%。而与 CEEMD-BPNN 和 CEEMND-RNN 相比, CEEMD-TCN 的 e_{RMSE} 分别减小 1.007、0.533, e_{MAPE} 分别减小 0.673%、0.376%。因此, CEEMD-TCN 模型的预测精度更高, 预测数据更符合实际值。

4.5 其余气体浓度预测实验结果

在上文的实验案例中,重点分析了 CEEMD-TCN 模型对 H₂ 浓度预测的结果。为了了解 CEEMD-TCN 模型对变压器中除 H₂ 外其余气体浓度变化趋势的预 测情况,将其预测结果分别与其他 3 种预测模型 对比。

由表 2 可知,相比较于 TCN、CEEMD-BPNN、 CEEMD-RNN 模型,使用 CEEMD-TCN 模型预测其 余油中溶解气体同样具有较高的预测稳定性。

表 2 其余气体浓度预测实验结果对比

编号	复休 计	语素	预测模型类型			
	类型	类型	CEEMD- TCN	TCN	CEEMD- BPNN	CEEMD- RNN
1	CH_4	$e_{ m RMSE}$ $e_{ m MAPE}/\%$	0.192 0.587	0.267 0.901	0.318 1.159	0.420 1.835
2	C_2H_6	$e_{ m RMSE}$ $e_{ m MAPE}/\%$	1.247 1.827	3.258 2.620	4.983 3.127	3.562 2.687
3	C_2H_4	$e_{ m RMSE} \ e_{ m MAPE} / \%$	5.201 0.965	8.166 1.682	9.713 1.781	14.869 2.213
4	C_2H_2	$e_{ m RMSE} \ e_{ m MAPE} / \%$	0.077 1.235	0.162 1.904	0.191 2.119	0.158 1.867
5	CO	$e_{ m RMSE}$ $e_{ m MAPE}/\%$	3.736 0.703	7.820 1.594	6.082 1.319	5.180 1.072
6	CO_2	$e_{ m RMSE}$ $e_{ m MAPE}/\%$	8.821 0.513	11.983 0.762	17.012 1.309	14.929 0.977

5 结论

针对变压器油中溶解气体浓度预测问题展开研究,提出基于 CEEMD-TCN 模型的预测方法,得出结论如下:

(1)采用 CEEMD 方法处理原始序列,可以有效 解决不同时间尺度信息间的相互耦合问题,更深层次 地挖掘原始序列特性,能够极大提高模型预测精度;

(2)采用 TCN 模型预测油中溶解气体浓度各子 序列分量,其结构中的膨胀因果卷积和残差模型能 够有效地捕捉序列内部关联特性,进而高精度地追 踪序列变化趋势;

(3)通过实验分析,采用同一油中溶解气体浓度数据集,CEEMD-TCN 相对于 TCN、CEEMD-BPNN和 CEEMD-RNN 预测模型,其预测误差更低,预测精度更高。

参考文献:

- 李刚,于长海,刘云鹏,等. 电力变压器故障预测与健康管理: 挑战与展望[J]. 电力系统自动化,2017,41(23):156-167.
- [2] 李振杰,卞朝晖,陈学民,等.变压器油色谱在线监测数据有效 性评估[J].高压电器,2018,54(6):158-163,169.
- [3] 杨志越,牛华宁. 基于 DGA 的变压器状态监测与故障诊断技 术研究[J]. 河北电力技术,2018,37(3):11-14.
- [4] 刘君,赵立进,黄良,等.基于改进模糊时间序列的变压器油中
 气体预测方法[J].武汉大学学报(工学版),2017,50(5):
 754-759.
- [5] 肖怀硕,李清泉. 灰色理论-变分模态分解和 NSGA-I 优化的 支持向量机在变压器油中气体预测中的应用[J]. 中国电机工 程学报,2017,37(12):3643-3653.
- [6] 刘慧鑫,张江龙. 基于时间序列模型的变压器油中溶解气体预测[J]. 高压电器,2019,55(12):193-199.
- [7] 徐肖伟,李鹤健. 基于随机森林的变压器油中溶解气体浓度预测[J]. 电子测量技术,2020,43(3):66-70.
- [8] 刘可真,苟家萁,骆钊,等. 基于粒子群优化-长短期记忆网络 模型的变压器油中溶解气体浓度预测方法[J]. 电子测量技 术,2020,43(3):2778-2784.
- [9] 王兴,荣海娜,王健,等. 基于小波分解和长短期记忆网络的变 压器油中溶解气体浓度预测[J]. 电工技术,2020(9):24-33.
- [10] 刘云鹏,许自强. 基于经验模态分解和长短期记忆神经网络的 变压器油中溶解气体浓度预测方法[J]. 中国电机工程学报, 2019,29(39):3998-4007.
- [11] Huang N E, Shen Z, Long S R. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis [J]. Proceedings of the Royal Society of London, 1998(454A):903-995.
- [12] Wu Z H, Huang N E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method [J]. Advances in Adaptive Data Analysis, 2009, 1(1):1-41.

- [13] Torres M E, Colominas M A, Schlotthauer G, et al. A Complete Ensemble Empirical Mode Decomposition with Adaptive noise [C]// IEEE Inter Conf on Acoustics, Speech and Signal Proc. Prague, CzechRepublic, 2011;4144–4147.
- [14] Bai S, Kolter J Z, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[J]. arXiv preprint arXiv:1803.01271,2018.



杨海晶(1995—),男,汉族,江苏盐城人, 硕士研究生,主要研究方向为电力设备故 障诊断和智能算法,y18252039321@ 163.com;



孙运全(1969—),男,汉族,山东莒南 人,江苏大学,教授,博士,主要研究方向 为电力系统电能质量控制,新能源电动 汽车能量管理等领域,sunyunquan99@ 126.com;



朱 伟(1995—),男,汉族,江苏南京 人,硕士,主要研究方向为电气设备故 障诊断和电力系统分析及其自动化, 366165822@qq.com;



钱 尧(1995—),男,汉族,江苏南通 人,硕士,主要研究方向为非接触能量 传输技术,380867458@qq.com;



金 浩(1997—),男,汉族,江苏沭阳 人,硕士研究生,主要研究方向为矿井 人员定位,3070107842@qq.com。