

# 基于数据清洗及 LSTM 神经网络的 CVT 故障诊断

武晓冬, 王麟斌, 牛天聪, 赵 锟, 温婉宁

(山西大学电力与建筑学院, 山西 太原 030006)

**摘要:** 电容式电压互感器 (capacitive voltage transformer, CVT) 在电网设备监测工作中有着重要作用。传统的 CVT 异常检测方法存在物理建模复杂, 模型参数和故障数据难以获取等问题。为此提出了一种基于数据清洗及长短期记忆 (long short-term memory neural network, LSTM) 神经网络的 CVT 故障诊断方法。针对容易陷入局部最优和无时间维度等缺点, 引入滑动窗口策略检测序列中的缺失值以及异常点, 提出基于 K-means 聚类算法的清洗流程。根据清洗后电力系统的电压历史数据, 利用 LSTM 得到电压预测值, 通过比较预测值与实际值来判断 CVT 是否故障。通过某变电站实际运行数据对本算法进行验证, 结果表明该方法能够对由于运行状态改变而导致的数据异常进行实时在线监控, 有效提高了 CVT 故障诊断的准确率。

**关键词:** 电容式电压互感器; 数据清洗; K-means 聚类算法; 长短期记忆; 故障诊断

中图分类号: TP277.3; TM451

文献标志码: A

文章编号: 1003-7241(2025)12-0173-04

## CVT fault diagnosis based on data cleaning and LSTM neural network

WU Xiaodong, WANG Linbin, NIU Tiancong, ZHAO Kun, WEN Wanning

(School of Electricity and Architecture, Taiyuan 030006, China)

**Abstract:** Capacitive voltage transformer (CVT) plays an important role in the monitoring of power grid equipment. The traditional CVT anomaly detection method has some problems, such as complicated physical modeling, difficult to obtain model parameters and fault data. Therefore, a fault diagnosis method of CVT based on data cleaning and long short term memory (LSTM) neural network is proposed. Aiming at the shortcomings of local optimization and no time dimension, sliding window strategy is introduced to detect missing values and abnormal points in the sequence, and a cleaning process based on K-means clustering algorithm is proposed. According to the voltage history data of the power system after cleaning, the voltage prediction value is obtained by using LSTM, and the fault of CVT is judged by comparing the predicted value with the actual value. The algorithm is verified by the actual running data of a substation, and the results show that the method can monitor the data anomalies caused by the change of running state in real time, and effectively improve the accuracy of CVT fault diagnosis.

**Keywords:** capacitive voltage transformer; data cleaning; K-means clustering algorithm; long short term memory; fault diagnosis

## 0 引言

“十四五”期间,我国电力设备规模快速增长,已成为我国电网的重要组成部分,与电网的安全稳定运行有着密切的联系。近年来,设备的智能化运维在对其数据的挖掘和利用仍缺少一种行之有效的方法和手段,因此需要对其进行相关的研究与探索<sup>[1-2]</sup>。

目前,500 kV 高压电网中大量采用了电容式电压互感器的特性对故障定位、继电保护、监控等二次设备的安全稳定运行有重要影响。但是在数据传输过程中会因受到外界环境干扰、在线监测系统不稳定以及传感器短时失效等,导致收集到的数据存在缺失和异常等问题<sup>[3]</sup>。文献[4]提出了基于机器学习的配电网异常缺失数据动态清洗方法,解决了异常数据辨识及缺失数据填补精度与速度的平衡。文献[5]提出了一种基于双重回归算法的 CVT 在线监测系统,通过采集 CVT 的运行数据如相位差、介质

损耗等,在时间维度上建立上述数据时间序列回归模型,进一步监测 CVT 的性能参数是否发生异常变化。文献[6]提出了基于循环神经网络的在线辅助 CVT 电压测量方法,通过比较预测值与 CVT 实际值判断 CVT 的状态。但上述文献均需要大量多维数据。

因此,本文通过对 CVT 历史电压数据,利用滑动窗口与 K-means 聚类结合检测数据的异常值,将清洗后的历史电压数据作为 LSTM 的输入值得到电压预测值。将神经网络的输出值与 CVT 的实际值进行阈值判断,对 CVT 内部进行故障诊断。实际数据验证,本文所提方法不需要海量、多维的数据,对 CVT 内部故障具有良好的诊断效果。

## 1 CVT 理论分析及模型建立

CVT 由电容分压器和电磁单元两部分组成。依据一定的比例可将电容分压器分为高压电容  $C_1$  和中压电容  $C_2$  组成,其中高压电容  $C_1$  中的  $C_{11}$ 、 $C_{12}$  和  $C_{13}$  承担来自一次系统的高压。为测量中压电容所分得的电压值,在中压电容抽头处装设中间变压器  $T$ ,中间变压器  $T$ 、补偿电抗器  $L$

\* 基金项目:2023 年山西省研究生教育创新计划支持 (2023AL05)

收稿日期:2024-11-13

与阻尼绕组  $Z$  构成电磁单元,传统物理建模方法需要附加设备或传感器,实现难度较大,因此建立等效模型如图 1 所示。

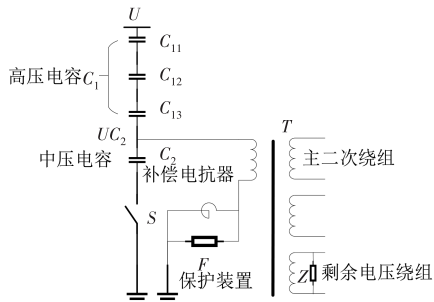


图 1 电容式电压互感器等效模型

当电压加到电容式电压互感器上时,根据电容分压器的原理,可以得出中压电容  $C_2$  的等值电压为

$$U_{c2} = U \frac{C_1}{C_1 + C_2} \quad (1)$$

电压  $U_{c2}$  通过中间变压器  $T$  调压,可得到绕组电压值。

由式(1)可知,在 CVT 的实际工作过程中,  $C_1$  和  $C_2$  中的电容元件击穿,均会造成电压数据的变化。因此可以将其作为判断 CVT 是否发生内部故障的依据<sup>[7]</sup>。

## 2 基于数据清洗及 LSTM 的故障诊断

### 2.1 建立基于滑动窗口的异常数据集

时间序列  $X$  是由采样时刻和采样点组成的元素集合  $X = \{(t_0, x_0), (t_1, x_1), \dots, (t_i, x_i), \dots\}$ 。由于安全限制不能安装传感器而获取故障数据困难,而电压数据随着智能电网调控系统的建设易获取,因此将电容式电压互感器的历史电压数据记为  $X_i = (t_i, x_i)$ ,  $t_i$  表示电压采样时刻,  $x_i$  表示 CVT 电压值。本文基于固定窗口长度  $w$  的滑动时窗对局部 CVT 电压数据段进行数据清洗。

本文定义  $S_w[t-w:t]$  为一个时间间隔为  $w$  的滑动窗口且  $t > w$ 。该滑动窗口内一系列采样点表示为  $S_w(x_i(w)) = \{x_{i-w}, x_{i-w+1}, \dots, x_{i-1}, x_i\}$ , 其中  $w$  是滑动窗口长度,可以依据实际情况及所需聚类数据的数据量取不同的值。具体判断异常数据过程如下。

- 1) 选定一段历史电压数据,从起始采样点  $t$  起,放置长度为  $w$  的滑动窗口  $S_w(x_i(w))$ ;
- 2)  $r$  为窗口单次向正方向滑动的距离,则  $X_{i+r} = (t_{i+r}, x_{i+r})$  为滑动后窗口的节点;
- 3) 为降低算法的计算复杂性,只将滑动窗口内的历史电压数据  $\{x_{i-w}, x_{i-w+1}, \dots, x_{i-1}, x_i\}$  作为 K-means 聚类算法的输入参数。

滑动窗口的示意图如图 2 所示<sup>[8]</sup>。

### 2.2 基于 K-means 聚类的数据清洗流程

基于 CVT 实时监测的数据特点,电力系统电压数据波动性增强,噪声数据与大量波动数据混杂,CVT 故障诊断精度降低,因此采用 K-means 聚类算法进行数据清洗,

由于本文算法在固定滑动时间窗口内所含数据较少,又能有效的处理具有显著差异的样本数据,故聚类簇选择 2 即可。

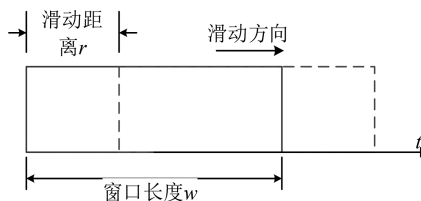


图 2 滑动窗口示意图

- 1) 将 CVT 的历史电压数据集  $\{x_1, x_2, \dots, x_i\}$  输入到 K-means 中进行聚类,输出 2 个聚类簇;
- 2) 用欧式距离作为衡量电压数据相似度的指标,计算每个采样点与簇质心之间的欧式距离,并按照距两个簇质心点距离最近的原则,将所有采样点分配到簇质心所对应的簇中;
- 3) 针对每个簇类,重新计算它的簇质心位置,当达到设置的簇类次数或簇质心位置不再改变时,聚类结果输出<sup>[9]</sup>;
- 4) 按照上述步骤利用 K-means 聚类算法构建数据清洗模型,即可得到电压数据簇质心。若两类簇质心之间的距离大于阈值,可以判断该时刻的电压数据异常,需要进行数据清洗。

### 2.3 基于 LSTM 的 CVT 故障预测

长短期记忆(long short-term memory, LSTM)神经网络是一种改进后的时间循环神经网络,适用于电力系统电压预测<sup>[10-11]</sup>。

本文采用的 LSTM 神经网络电压预测模型由一个输入层,三个隐藏层和一个输出层组成,输入层接收原始电压数据,隐藏层各包含 10 个神经元,输出层为 1 个神经元,输出电压预测值,模型损失函数为均方根误差(root mean square error, RMSE)。将变电站实际母线的电压历史数据输入到 LSTM 模型中,得到电压预测值。基于神经网络的故障预测模型如图 3 所示。

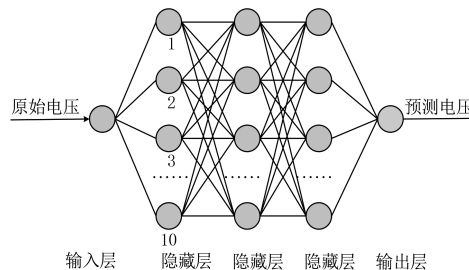


图 3 基于神经网络的故障预测模型

具体算法步骤如下:

- 1) 数据预处理,将样本分为训练集和测试集;
- 2) 搭建 LSTM 模型,配置 LSTM 的隐层数和学习率;
- 3) 对 LSTM 权重梯度进行初始化,利用训练集电压数据对 LSTM 进行训练,更新 LSTM 权重梯度;
- 4) 训练 LSTM 网络参数后,使用该网络对电压数据进行预测。

## 2.4 故障诊断的具体算法实现步骤

由于聚类算法是一种以无序数据集为基础的分类算法,无法体现数据集的时间属性,而 CVT 采集的数据与时间序列数据流特征吻合,采用以数据流滑动时间窗口为基础的聚类算法对 CVT 的历史电压值进行故障诊断,流程图如图 4 所示,具体步骤如下:

1) 滑动窗口截取序列。从时间轴第一个待测数据开始,放置长度为  $w$  的滑动窗口,以  $r$  为滑动距离延时间轴方向进行滑动;

2) 数据聚类。对每次滑动后窗口内的电压数据进行聚类处理得到两类簇质心;

3) 计算簇质心距离。在传感器不稳定或正常的电网电压小幅波动时,通过聚类算法计算出的两类簇质心距离差值变化比较明显,若差值大于阈值,则该数据点为异常值需要清洗,若小于阈值,则该点为正常值需要保留,清洗后数据存储;

4) LSTM 电压预测模型。将清洗后的电压历史值作为 LSTM 模型的输入训练网络得到电压预测值;

5) CVT 故障数据。将电压历史数据输入 CVT 仿真模型中,通过设置电容参数变化,模拟 CVT 在故障状态下的输出特性,从而生成相应的故障数据,用于后续故障检测;

6) 阈值判断。最后,将经过清洗后的 LSTM 电压预测值作为准确值,设置上下浮动的阈值,对 CVT 仿真模型输出的故障数据进行判断,若其超出阈值范围,则视为发生了故障。

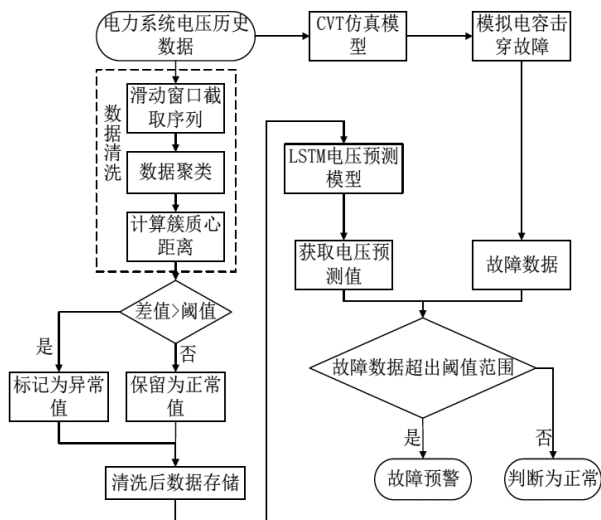


图 4 基于数据清洗和 LSTM 的故障诊断的流程图

## 3 算例分析

### 3.1 数据清洗必要性分析

本文选取 CVT 在线监测数据为例,对其进行故障诊断。被测数据是某站 500 kV 母线 CVT 所测电压数据。该数据集包含了 8 640 个数据点,时间间隔为 5 分钟,如图 5 所示。可以看出,CVT 原始电压数据波动较大,难以有效的在电压数据中发现异常并判断需要清洗的异常数据。

仿真使用的 LSTM 神经网络模型初始学习率为 0.005,训练迭代次数为 250 次,在 125 轮后通过乘以因子 0.2 来降低学习率,历史电压数据前 6 048 点作为训练集,后 2 592 个点作为测试集,LSTM 电压预测结果如图 6 所示。

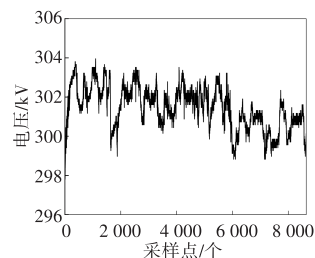


图 5 CVT 原始电压数据

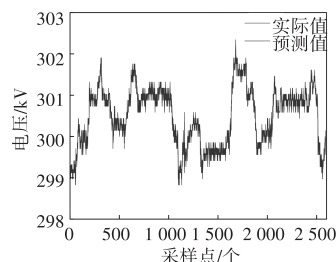


图 6 LSTM 预测结果

预测结果的 RMSE 为 0.949 39,通过计算各采样点的实际值与预测值的偏差计算,得到逐点的电压预测准确率,其中 7 个点的偏差值大于 2,23 个点的电压偏差值大于 1.5,由于其邻近时刻的电压采样点为正常数据,因此可以得出该时刻存在的异常点为数据噪声。该噪声可能是由于某传感器不稳定造成,由于这些点的预测值与实际值偏差较大,若以该预测值作为 CVT 故障诊断的临界值,将会导致诊断误差偏大,因此需要将这些异常点进行清洗。

设定聚类算法有两类簇质心,滑动窗口宽度  $w$  为 5,窗口滑动距离  $r$  为 1,将历史数据通过滑动窗口输入到聚类中,计算 K-means 聚类后的两类簇质心之间的差值,部分 CVT 历史电压数据异常区间内的簇质心如图 7 所示,可明显看出部分区域两类簇质心之间的距离差值明显较大,表明该窗口内电压数据存在异常,有必要进行数据清洗。在经过数据清洗后 RMSE 为 0.384 96,均方根误差明显下降,预测精度明显提高。

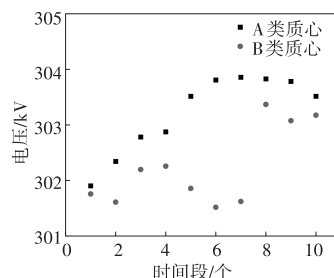


图 7 部分电压数据异常区间内的簇质心

### 3.2 故障诊断性能分析

为保证仿真环境中 CVT 与实际工作环境中一致,将历史电压数据输入式(1)的 CVT 模型中。由于分压电容

的部分击穿会导致该节电容值变化,仿真过程中用电容变化量表征等值电容的击穿程度。为模拟 CVT 内部电容被击穿的情况,选取 CVT 电容  $C_1$  击穿、 $C_2$  击穿状态进行试验,分别设置电容击穿 1%、2%、3%、4%、5%,以及多次累积击穿故障,共 12 种故障模拟击穿后的电压变化。采集电容实验过程中 CVT 的故障数据,电容式电压互感器仿真故障数据如图 8 所示。

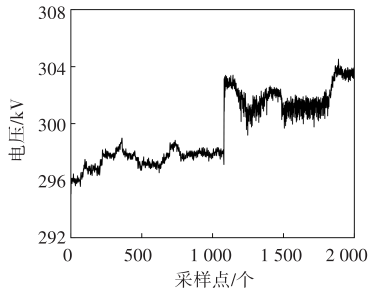


图 8 故障数据

采用准确率指标评价各算法诊断性能,准确率等于正确的正常状态检出个数加正确的故障状态检出个数的和与总采样个数之比的百分数。本文采用了三种算法对 12 组故障测试样本进行 CVT 的故障诊断,其中算法一将滑动时窗 K-means 聚类算法用于直接筛选故障数据中的异常值,但聚类算法只能识别电压跃变数据,无法识别长期故障,因此准确率低。算法二仅用 LSTM 神经网络进行检测,检出准确率偏低,算法三为本文所提算法,相比于其他算法采用了数据清洗与 LSTM 算法,有效避免了因预测偏差而导致的准确率下降,误诊率和漏诊率最低。检测准确率统计曲线如图 9 所示。

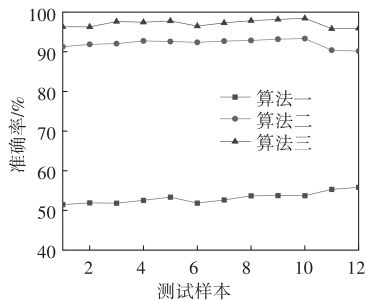


图 9 检测准确率统计曲线

图9中算法一、算法二、算法三的平均诊断准确率分别

为 53.16%、92.13%、97.12%,且算法三的最小诊断准确率为 95.81%。

## 4 结束语

为了实现 CVT 实时故障诊断,本文提出了基于数据清洗和 LSTM 神经网络的故障诊断模型,研究采用基于数据流滑动时间窗口的聚类算法对 CVT 异常数据进行清洗,有助于故障诊断准确率的提高。通过对比 K-means 聚类、LSTM 故障诊断算法的诊断准确率可知,数据清洗和 LSTM 神经网络模型故障诊断最小准确率为 95.81%,相对其他算法均有显著提高。解决了 CVT 故障诊断中实时性问题,证实了该方法的有效性和准确性。

## 参考文献

- [1] 贾骏,杨强,付慧,等. 基于电力设备大数据的预训练语言模型构建和文本语义分析[J]. 中国电机工程学报, 2023, 43(3): 1027-1037.
- [2] 蒲天骄,陈盛,赵琦,等. 能源互联网数字孪生系统框架设计及应用展望[J]. 中国电机工程学报, 2021, 41(6): 2012-2029.
- [3] 高正男,杨帆,胡妹博,等. 面向新能源电力系统状态估计的伪波动数据清洗[J]. 高电压技术, 2022, 48(6): 2366-2377.
- [4] 梅玉杰,李勇,周王峰,等. 基于机器学习的配电网异常缺失数据动态清洗方法[J]. 电力系统保护与控制, 2023, 51(7): 158-169.
- [5] LI J, HE D, QIN J, et al. On-line monitoring system and double regression algorithm of capacitor voltage transformer [C]//2019 IEEE Sustainable Power and Energy Conference (ISPEC), Beijing, China. IEEE, 2019: 2686-2689.
- [6] 陈文中,陈俊杰,许侃,等. 基于循环神经网络的一种在线辅助 CVT 电压误差测量方法[J]. 东华大学学报(自然科学版), 2022, 48(1): 53-57, 65.
- [7] 严英杰,盛戈峰,刘亚东,等. 基于滑动窗口和聚类算法的变压器状态异常检测[J]. 高电压技术, 2016, 42(12): 4020-4025.
- [8] 强文渊,韦家富,刘友波,等. 基于滑动时窗 K-Means 聚类的 CVT 早期故障诊断[J]. 电力电容器与无功补偿, 2019, 40(3): 89-95.
- [9] 金静,方园,费洋,等. 模式聚类与周期分解的能耗监测及异常检测方法[J]. 电测与仪表, 2023, 60(11): 151-157.
- [10] 张明泽,栾文鹏,艾欣,等. 基于边缘计算的台区短期负荷预测方法[J]. 电测与仪表, 2024, 61(4): 93-99.
- [11] 代杰杰,宋辉,盛戈峰,等. 采用 LSTM 网络的电力变压器运行状态预测方法研究[J]. 高电压技术, 2018, 44(4): 1099-1106.

作者简介:武晓冬(1978—),男,博士,副教授,研究方向:电力系统运行与控制。

(上接第 100 页)

[8] 单铨洋,孙战里,曾志刚. RFNet:用于三维点云分类的卷积神经网络[J]. 自动化学报, 2023, 49(11): 2350-2359.

[9] 魏鑫,李晓婷,赵世慧,等. 基于卷积神经网络和随机森林的毁伤效果评估方法[J]. 火力与指挥控制, 2023, 48(3): 185-190.

[10] 冷冰,冷敏,常智敏,等. 基于 Transformer 结构的深度学习模型用于外周血白细胞检测[J]. 仪器仪表学报, 2023, 44(5): 113-120.

[11] 陈毅夫,何敬,刘刚,等. 融合 Swin-Transformer 网络模型的水体高光区域提取[J]. 遥感信息, 2023, 38(4): 129-136.

[12] 李云波. 基于全卷积神经网络的多级阈值图像分割算法[J]. 微型电脑应用, 2023, 39(6): 145-147.

[13] 吉旭瑞,刘静,吉辉,等. 改进全局上下文注意力新冠肺炎 X 光诊断方法[J]. 计算机工程与应用, 2023, 59(21): 222-230.

[14] 薛均晓,武雪程,王世豪,等. 基于改进 YOLOv4 的自然人群口罩佩戴检测方法[J]. 郑州大学学报(工学版), 2022, 43(4): 16-22.

[15] 孔刘玲,刘秀文. 基于改进 YOLOv4 算法的船舶目标检测方法[J]. 船舶工程, 2022, 44(1): 96-147.

作者简介:林嘉鑫(1992—),男,硕士,工程师,研究方向:技术平台、人工智能。