DOI: 10.13382/j.jemi.B2306955

基于 IKNN 和 LOF 的变压器回复电压数据清洗方法研究*

陈啸轩 邹 阳 翁祖辰 林锦茹 林昕亮 张云霄
(福州大学电气工程与自动化学院 福州 350108)

摘 要:基于回复电压极化谱提取特征参量是目前广泛应用的变压器油纸绝缘状态评估方法,但极化谱易受工况干扰、人工失误等因素影响而出现特征数据异常的情况,严重降低评估准确性。针对上述问题,该文提出了一种基于局部离群因子(LOF)和改进 K 最近邻(IKNN)的回复电压数据清洗方法。首先,选取回复电压极化谱的回复电压极大值 U_{rmax}、初始斜率 S_r 与主时间常数 t_{cdom} 作为老化特征参量,并基于 LOF 算法对非标准极化谱中的异常特征量数据进行识别与筛除。其次,利用模糊 C 均值(FCM)聚类算法减小噪声点对 KNN 算法的干扰,并通过加权欧氏距离标度突出各特征量间的关联性,进而构建出基于 IKNN 的数据填补模型架构以实现特征缺失数据的填补。最后,代入多组实测数据验证所提数据清洗方法的实效性。结果表明,数据清洗后的状态评估准确率相较于原有数据上升了 50%左右,有效提高了变压器回复电压数据质量,为准确感知变压器运行状况奠定坚实的基础。

关键词: 油纸绝缘;特征数据清洗;局部离群因子算法;回复电压极化谱;改进 K 最近邻算法
中图分类号: TM411 **文献标识码**: A **国家标准学科分类代码**: 470.4037

Recovery voltage data cleaning method for transformer based on IKNN and LOF

Chen Xiaoxuan Zou Yang Weng Zuchen Lin Jinjia Lin Xinliang Zhang Yunxiao
(School of Electrical Engineering and Automation of Fuzhou University, Fuzhou 350108, China)

Abstract: Extracting feature parameters from the recovery voltage polarization spectrum is currently a widely adopted method for evaluating the status of transformer oil-paper insulation. However, the polarization spectrum is prone to anomalous feature data due to factors such as working condition interference and artificial errors, which seriously reduces the accuracy of the evaluation. In response to the above issues, this paper proposed a recovery voltage data cleaning method based on local outlier factor (LOF) and improved K-nearest neighbor (IKNN). Firstly, Maximum recovery voltage U_{rmax}, the initial slope S_r and dominant time constant t_{cdom} of the recovery voltage polarization spectrum were selected as aging feature parameters, and anomalous feature data in the non-standard polarization spectrum were identified and filtered out based on the LOF algorithm. Secondly, the Fuzzy C-means (FCM) clustering algorithm was used to reduce the interference of noise points on the KNN algorithm, and the correlations between various features were highlighted by weighted Euclidean distance scale. Then, a data filling model architecture based on IKNN was constructed to fill in missing feature data. Finally, multiple sets of measured data were incorporated to validate the effectiveness of the proposed data cleaning method. The results indicate that the accuracy of status evaluation after data cleaning has increased by about 50% compared to the original data, which effectively improves the quality of transformer recovery voltage data and lays a solid foundation for accurate perception of transformer operation status.

Keywords: oil-paper insulation; feature data cleaning; local outlier factor algorithm; recovery voltage polarization spectrum; improved K-nearest neighbor algorithm

收稿日期:2023-10-13 Received Date: 2023-10-13
* 基金项目:国家自然科学基金重大研究计划培育项目(92266110)、福建省自然科学基金项目(2019J01248)资助

0　引　言

随着“十四五”规划的顺利推进,实现“数字化电网”建设以及“碳达峰、碳中和”成为未来电力行业发展的热点与重点。新形势下电力系统将呈现规模化新能源分布式接入、交直流多电压等级互联等特征,对电力系统装备的稳定性提出了更高的要求。油浸式变压器作为输配电环节的关键枢纽,其油纸绝缘性能的优劣直接影响着电网的安全运行[1-4]。因此,油纸绝缘状态的准确评估显得尤为重要。基于介质时域响应理论的回复电压法(recovery voltage method, RVM)是当前变压器油纸绝缘状态评估最常用的方法之一,该方法携带绝缘信息量丰富,且测试过程对绝缘无损,上述优点使该方法逐渐成为国内外学者的研究热点[5]。

目前,基于 RVM 的变压器油纸绝缘状态评估方法已取得了一定的研究成果,邹阳等融合灰色关联分析和聚类云模型构建变压器时域数据状态评估体系[6],林智勇等[7]通过回复电压微分解谱法实现变压器油纸绝缘等效电路支路数的准确判定,张宁等[8]通过结合层次分析和逼近理想解法构建油纸绝缘状态评估模型。但上述方法均建立在特征量数据完整且无异常的前提下方能取得良好的评估结果,缺少对特征量异常情况的考虑。而变压器的现场运行环境复杂多变,其周围分布的交直流叠加电场、热应力场会使回复电压测量设备受到电磁波、环境温度、湿度等多重因素的干扰。且随着实验次数的增多,诸如放电不充分、充电电压过低,充放电时间比选择不当等实验误操作也时常发生。上述情况会使基于回复电压极化谱曲线所采集的特征量数据出现异常,进而可能导致油纸绝缘状态评估结果与实际情况相悖,危害电力系统的稳定运行。同时考虑到现场检修时间的紧迫性,实测 RVM 数据有限,如何充分挖掘有限的 RVM 测试数据所携带的绝缘信息就显得尤为重要。因此,对各特征量中的异常数据进行准确清洗,提升绝缘特征信息质量是后续研究的关键。

近年来,基于人工智能技术的数据清洗方法已取得了一定的进展。在异常数据识别方面,孤立森林(isolation forest, IF)、局部离群因子(local outlier factor, LOF)、基于密度的带噪声空间聚类(density-based spatial clustering of applications with noise, DBSCAN)均是目前的研究热点[9-11]。其中局部离群因子算法由于其检测范围广,检测精度高且对训练数据需求量不大等优点而得到广泛的应用。在缺失数据修复方面,基于 K 最近邻(K-nearest neighbor, KNN)理论的缺失数据修复算法是一种准确有效的方法,该方法通过寻找并融合具有相似数据模式的近邻样本信息以实现缺失数据修复目的,在房地产、医学等领域都已有成功的修复案例[12-13]。但传统 KNN 算法存在计算时间过长,内存需求量大,对噪声数据敏感等问题[14],而电力检修时间紧且噪声数据多,传统 KNN 算法易受干扰而无法准确高效地修复特征数据。

针对上述问题,本文提出了一种结合改进 KNN(improved KNN, IKNN)算法和 LOF 算法的回复电压特征量数据清洗方法。该方法基于 LOF 算法对非标准极化谱中的异常特征量数据进行识别,消除了异常数据对后续油纸绝缘状态评估的不良影响;同时,将模糊 C 均值(fuzzy C-means, FCM)聚类算法与加权欧式距离标度同传统 KNN 修复算法进行结合,减小噪声点对数据修复结果的干扰并突出各特征量间的相关性,提升数据修复的高效性和准确性。最后,通过多组实测 RVM 数据进行比较验证了本文所提方法的实效性。

1　油纸绝缘老化特征量的提取

回复电压法通过对绝缘介质外施直流电压以研究绝缘介质的缓慢弛豫极化过程。单次回复电压测试曲线如图 1 所示。以此为基础,通过不断改变外施直流电压的持续时间获得标准回复电压极化谱曲线,如图 2 所示。目前,已有研究从回复电压极化谱曲线中提取一系列特征量用于油纸绝缘状态评估[15-16]。其中,回复电压极大值 U_{rmax} 与回复电压初始斜率 S_{r} 会随油纸绝缘老化状态的加深而逐渐增大,主时间常数 t_{cdom} 则随油纸绝缘老化状态的加深而逐渐减小,上述 3 个特征量与油纸绝缘的老化状态均存在显著联系[16-17]。因此,本文选取 U_{rmax}、t_{cdom} 与 S_{r} 这 3 个特征量用于油纸绝缘状态评估。

但是,回复电压极化谱易受复杂测试环境或人为因素的干扰,使谱线出现峰值异常、陡增、多峰值点等异常现象,导致极化谱曲线非标准化[18],如图 3 所示。非标准极化谱曲线是标准极化谱曲线的畸变产物,基于该曲线所采集的 U_{rmax}、t_{cdom}、S_{r} 等特征量数据无法准确表征绝缘状况,从而污染整个特征量数据库,影响油纸绝缘状态的准确评估。

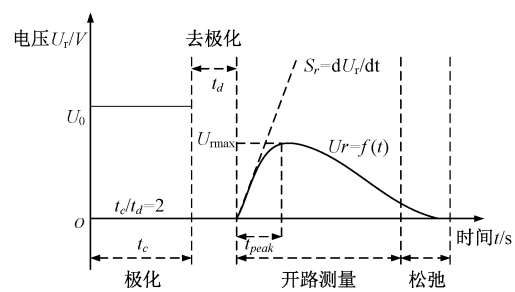


图 1　单次回复电压测试曲线

Fig. 1　Test curve of single recovery voltage

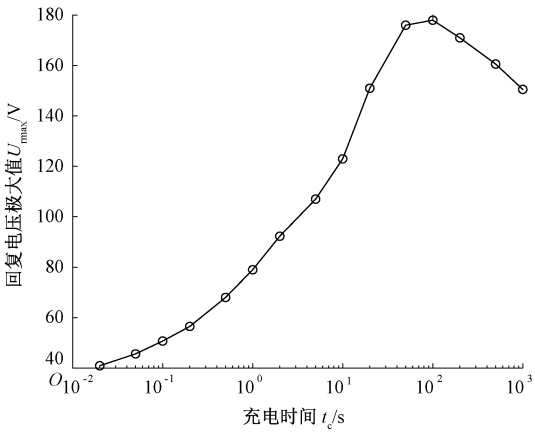
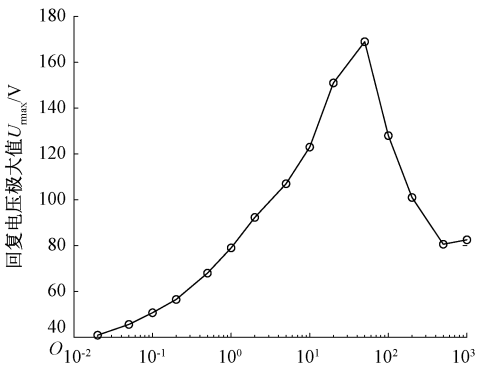


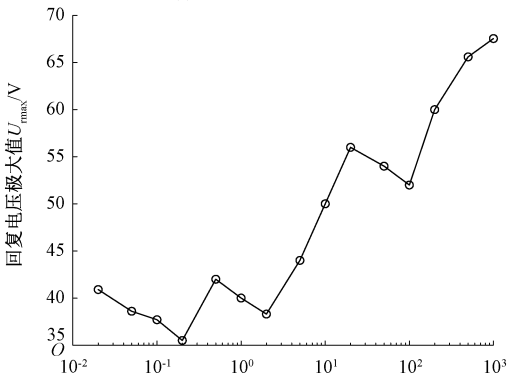
图 2 标准回复电压极化谱曲线

Fig. 2 Standard recovery voltage polarization spectrum curve



(a) 曲线陡增

(a) Sharp increase in curve



(b) 多峰值点

(b) Multiple peak points

图 3 非标准极化谱曲线

Fig. 3 Nonstandard polarization spectrum curve

2 基于 LOF 的异常数据识别算法

局部离群因子算法是一种基于密度的异常值识别算法[19]。该算法具有计算速度快,检测精度高,搜索能力强等优点,其主要识别流程如下:

设 n 维空间中点 O 和点 P 的欧氏距离为 $d(O,P)$,将点 O 领域内的点到点 O 的距离排序得到(O_1,O_2,…,O_k),则为 $d_k(O)=d(O,O_k)$。第 k 距离内所有点的几何称为点 O 的第 k 领域 $N_k(O)$。

以点 O 为中心,任意一点 P 到点 O 的第 k 可达距离定义为点 O 的第 k 距离与点 O,P 的实际欧式距离之间的较大值,如式(1)所示。

$$d_k(O,P)=\max\{d_k(O),d(O,P)\} \tag{1}$$

接着计算点 O 的局部可达密度,其代表数据点 O 的第 k 邻域内所有数据点到数据点 O 的第 k 可达距离平均值的倒数,如式(2)所示。

$$\rho_k(O)=\frac{|N_k(O)|}{\sum_{O\in N_k(P)} d_k(O,P)} \tag{2}$$

最后,计算数据点 O 的局部离群因子,其表示数据点 O 的领域 $N_k(O)$ 内其他点的局部可达密度与点 O 的局部可达密度之比的平均数,如式(3)所示。

$$LOF_k(O)=\frac{\sum_{O\in N_k(P)}\frac{\rho_k(P)}{\rho_k(O)}}{|N_k(O)|} \tag{3}$$

LOF 值小于或接近于 1,说明数据点 O 密度比其领域点密度大或者相近,可判断点 O 为密集点或与其领域点为同一簇,由此可判断点 O 为正常数据点;若 LOF 值大于 1,说明数据点 O 密度比其领域点密度小,由此可判断数据点 O 为异常数据点。

3 基于 IKNN 的缺失数据修复算法

基于上述方法可以有效识别并筛除异常数据,但为充分挖掘回复电压极化谱所携带的绝缘信息,避免部分数据筛除造成的整体数据浪费,本文构建了一种基于改进 KNN 算法的缺失数据修复方法。KNN 算法是一种基于欧式距离标度的缺失数据修复算法,其具有精度高,理论成熟等优点。但传统的 KNN 修复算法存在以下 3 个问题:1)忽略了最近邻的 k 个样本中存在数据噪声的情况,只是笼统地综合最近邻样本的信息,导致最终修复结果产生偏差;2)油纸绝缘各特征量间的数量级相差较大,利用传统 KNN 修复算法会导致数量级小的特征量数据被数量级大的特征量数据所覆盖,导致修复过程不全面;3)油纸绝缘各特征量间存在一定的相关性,而传统欧式距离标度忽略了该相关性,导致最终的修复结果不够精确。针对以上问题,本文做出如下改进。

3.1 结合 FCM 聚类算法排除噪声点

FCM 聚类算法是一种无监督机器学习方法。通过隶属度函数与聚类中心函数的互相迭代更新,最终根据

每个数据与各簇之间隶属度的大小将各数据分到不同的簇中[20]。隶属度函数与聚类中心函数式(4)、式(5)所示：

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}},\forall j\neq i \tag{4}$$

$$p_i=\frac{\sum_{k=1}^{n}(u_{ik})^m x_k}{\sum_{k=1}^{n}(u_{ik})^m} \tag{5}$$

式中：u_{ik} 表示样本 k 属于第 i 个簇的隶属度；c 表示聚类数目；n 表示数据样本个数；m 为模糊化程度；d_{ik} 表示第 k 个样本到第 i 个聚类中心距离；d_{jk} 表示第 k 个样本到第 j 个聚类中心距离；P_i 表示聚类中心值。

如图 4 所示，X_0 是待修复数据点。X_1，X_2，X_3，X_4，X_5，X_6 为 X_0 的 6 个最近邻点。由图 4 可直观看出，X_6 与目标数据点 X_0 的联系并不紧密，即 X_6 是 X_0 的噪声近邻点，通过 FCM 聚类算法可将点 X_0 与 X_6 分到不同的簇中，此时只要对点 X_0 进行簇内修复，就可以避免近邻噪声点的干扰。

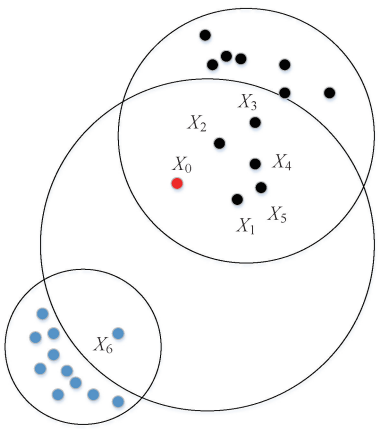


图 4　目标数据点近邻关系图

Fig. 4　Neighborhood graph of target data points

3.2　归一化处理数量级差异

设 $A=(a_{ij})_{M\times N}$ 为变压器回复电压特征量数据矩阵，a_{ij} 表示第 i 台变压器的第 j 个特征量值，M 表示变压器台数，N 表示特征量个数。对所有特征量值采用式(6)进行归一化处理。

$$a'_{ij}=\frac{a_{ij}-\min\limits_i\{a_{ij}\}}{\max\limits_i\{a_{ij}\}-\min\limits_i\{a_{ij}\}} \tag{6}$$

式中：$\max\limits_i\{a_{ij}\}$、$\min\limits_i\{a_{ij}\}$ 分别表示第 i 台变压器第 j 个特征量的最大值和最小值；a'_{ij} 表示第 i 台变压器第 j 个特征量归一化后的规范值。

3.3　相关系数加权的欧式距离标度

设 M 组变压器回复电压特征量数据为 $A_i=(a_{i1},a_{i2},a_{ip},\cdots,a_{iq},\cdots,a_{iN})(1\leqslant i\leqslant M)$，每组样本具有 N 个回复电压特征量，则第 i 组数据的第 p 个特征量与第 q 个特征量的相关系数 γ_{pq} 可表示为：

$$\gamma_{pq}=\frac{\sum_{i=1}^{M}(a_{ip}-u_p)(a_{iq}-u_q)}{\sqrt{\sum_{i=1}^{M}(a_{ip}-u_p)^2}\sqrt{\sum_{i=1}^{M}(a_{iq}-u_q)^2}} \tag{7}$$

式中：u 为相应特征量值的平均数，相关系数的取值为[-1,1]，-1 表示两个特征量强负相关，1 表示两个特征量强正相关，0 表示两个特征量不相关。

假设归一化后 M 组变压器回复电压特征量数据为 $A'_i=(a'_{i1},a'_{i2},\cdots a'_{iN})(1\leqslant i\leqslant M)$，且样本 i 缺少第 p $(1\leqslant p\leqslant N)$ 个特征量值，则样本 i 与样本 j 以相关系数的指数为权值的加权欧氏距离 $d_{ij\mid p}$ 可表示为：

$$d_{ij\mid p}=\sum_{q=1}^{N}e^{\mid\gamma_{pq}\mid}\sqrt{(a'_{iq}-a'_{jq})},q\neq p \tag{8}$$

通过加权欧氏距离可突出与待修复特征量相关性大的特征量的贡献度，使修复结果更贴近真实值。

3.4　特征量数据清洗流程

综上，本文所提数据清洗方法的具体步骤如下：

1)首先基于 LOF 算法对初始数据库进行异常数据识别，并将异常数据筛除。从而将初始数据库分为 A、B 两个数据库(A 数据库中各组数据无特征量值缺失，B 数据库中各组数据的部分特征量值缺失)；

2)采用 FCM 聚类算法对 A 数据库进行模糊划分，获取各聚类中心值；

3)计算 B 数据库中各组数据与各聚类中心的相似度，并将其归到最相似的簇中；

4)在各组数据的相应簇中利用 IKNN 算法进行缺失数据修复；

5)利用 LOF 算法对修复后的各组数据进行异常数据识别，若仍有异常特征量值则筛除并归入 B 数据库，若无则将该组数据归入无异常数据库。

数据清洗整体流程图如图 5 所示。

4　实例验证

为验证本文所提数据清洗方法的实效性，现基于本课题组近十年来利用如图 6 所示的 RVM5461 回复电压测试仪对相同电压等级但不同绝缘状态的油浸式变压器进行回复电压测试所获得的数据，选取其中 120 组变压器实测数据用于构造初始数据库。该数据库中有 15 组数据是基于非标准极化谱曲线所提取的异常特征量值，剩余 105 组数据是基于标准极化谱曲线所采集到的正常数据。

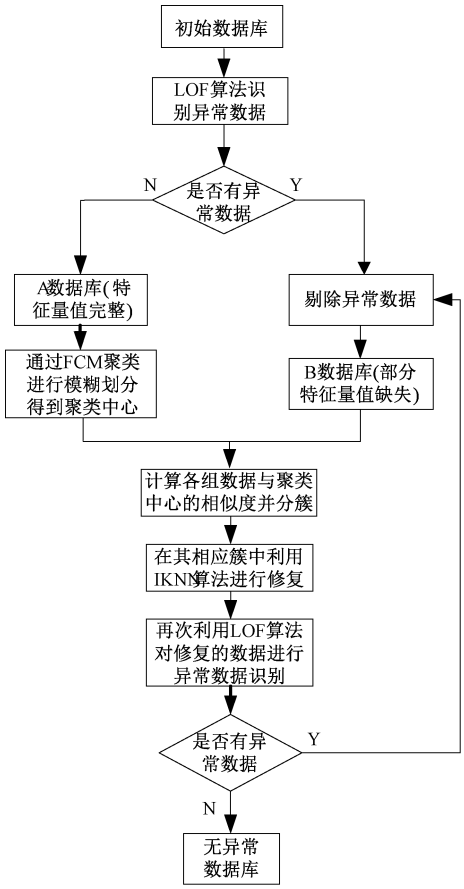


图 5 特征量数据清洗流程

Fig. 5 Feature data cleaning flow chart

图 6 RVM5461 回复电压测试仪

Fig. 6 RVM5461 recovery voltage tester

4.1 异常数据识别实验分析

首先基于 LOF 算法对上述数据库进行异常数据识别。为便于区分与说明,采用圆点代表正常数据,同时结合实际测试工况将识别出的异常数据分为Ⅰ、Ⅱ、Ⅲ组,分别采用十字、星形、三角点表示,最终生成散点图,如图 7 所示。现结合实际测试工况对识别结果进行具体说明:

1)圆点所代表的 105 组数据是基于在 20 ℃ ~ 30 ℃,干燥且电磁干扰较小的环境下测试所得的标准极化谱曲线采集所得,故各特征量值正常。

2)十字点所代表的Ⅰ组数据是由于测试前放电不充分,故极化谱曲线初始部分受到残余电荷干扰,造成回复电压峰值点较早出现[7],故这 5 组数据中主时间常数偏小。

3)星形点所代表的Ⅱ组数据是在雾雨天气下测试所得。潮湿的环境会使绝缘表面产生泄漏电流[20],导致极化谱曲线未至最高点就发生陡降,故这 3 组数据的回复电压极大值偏小。

4)三角点所代表的Ⅲ组数据是在电磁干扰较大的环境中测试所得。此时极化谱曲线出现异常波动,出现了多个局部峰值点[18],由于不同极化谱曲线波动情况不同,故这 7 组数据的各特征量异常情况也不同。

由上述分析可知,造成特征量数据异常变化的原因种类繁多,异常数据的分布范围广且不均匀,为识别过程带来困难,而 LOF 算法由于其高效的搜索能力可以准确识别出各组中存在的部分异常数据,为后续油纸绝缘状态评估创造有利条件。

4.2 缺失数据修复试验分析

筛除上述部分异常数据后,现利用 IKNN 算法对筛除后缺失的数据进行修复试验。参考文献[21]的油纸绝缘状态分级标准,利用 FCM 聚类算法对 105 组正常数据进行聚类分析,获得 4 个聚类中心(Ⅰ ~ Ⅳ),以此建立 4 级不同绝缘状态的油纸绝缘状态标准向量表,如表 1 所示。4 个不同绝缘等级分别为:绝缘状态良好(G)、绝缘状态中等(M)、绝缘状态一般(P)、绝缘状态老化严重(S)。

表 1 油纸绝缘状态标准向量表

Table 1 Standard vector table of oil-paper insulation status

FCM	U_{rmax}/V	t_{cdom}/s	S_r/(V/s)	绝缘状态
Ⅰ	171.88	2 339.64	45.52	G
Ⅱ	226.69	1 555.27	70.91	M
Ⅲ	285.47	1 005.59	93.58	P
Ⅳ	321.33	607.59	112.90	S

接着根据式(7)、(8)计算各组数据与各聚类中心的加权欧式距离值 $d_{ij|p}$,该值越小说明越相似,从而将 15 组缺失数据分别分类至与其最相似的聚类中心中,避免不同聚类中心中的近邻噪声点对修复结果造成影响。

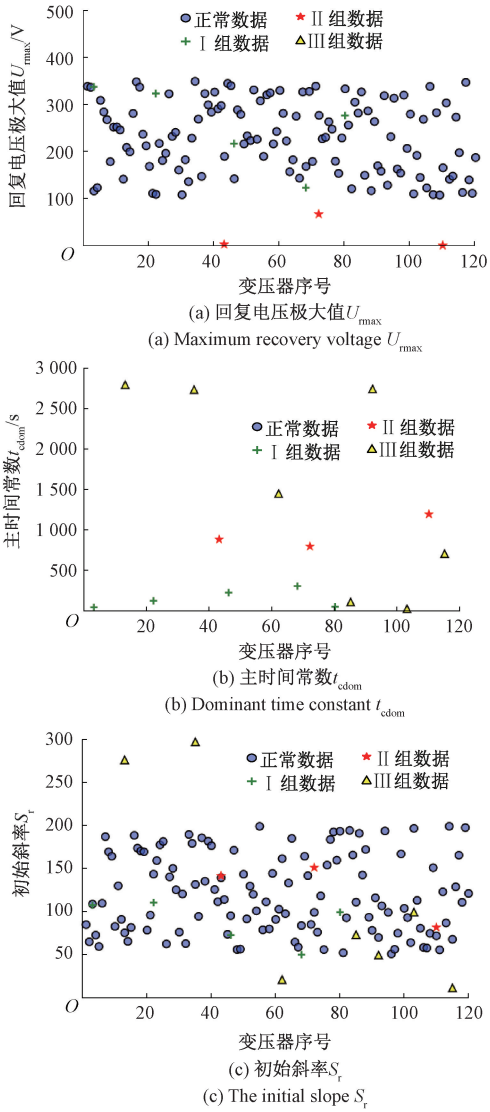


图 7　各特征量数据散点图

Fig. 7　Scatter diagram of each characteristic quantity data

分类后,在各组特征量数据的相应聚类中心中利用 KNN 算法对缺失的特征量值进行修复。修复情况的优劣采用均方根误差(root mean square error, RMSE)来衡量,该值越小说明目标修复效果越好,该值越大则修复效果越差。为综合衡量各特征量值的修复情况,本文取各特征量 RMSE 的均值作为最终衡量标准,计算公式如式(9)所示。

$$RMSE' = \frac{\sum_{j=1}^{n}\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y'_{ij}-y_{ij})}}{n} \tag{9}$$

式中:m 表示数据个数,n 表示特征量个数,y'_{ij} 表示第 i 台变压器第 j 个特征量的修复值,y_{ij} 表示第 i 台变压器第 j 个特征量的真实值。

在 KNN 算法取不同最近邻个数的情况下,采用不同方法修复的平均均方根误差如图 8 所示。

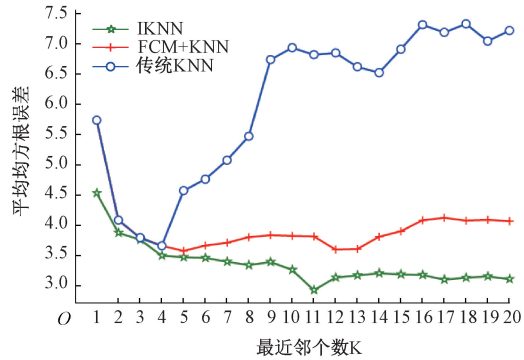


图 8　平均均方根误差

Fig. 8　Mean root mean square error

由图 8 可以看出,当最近邻个数 K<5 时,传统 KNN 算法、FCM-KNN 算法的平均均方根误差相同。因为此时最近邻个数取值较小,未出现近邻噪声点。当 K=5 时,开始出现近邻噪声点,使传统 KNN 算法的平均均方根误差增大,且随着 K 值增大,近邻噪声点增多,平均均方根误差呈增大趋势。当 K>5 时,由于 FCM-KNN 算法与本文所提的 IKNN 算法结合了聚类算法,故未受到近邻噪声点的影响,从而平均均方根误差没有呈持续增大的趋势。同时, IKNN 算法还通过加权欧式距离考虑了各特征量与聚类中心的相关性,故较 FCM-KNN 算法有更低的平均均方根误差,使修复结果更加接近于真实情况。

为避免偶然性,本文还通过人为剔除数据的方式,进行了当缺失数据个数为 25,40 时的缺失数据修复实验,对应平均均方根误差如图 9 所示。

结合图 8、9 的结果可以看出,在修复样本个数不同情况下,本文所提的 IKNN 算法均具有更低的平均均方根误差,此外,由于聚类算法的引入, IKNN 算法的搜索空间由全局样本空间变为各个簇空间,这相当于缩小了搜索范围,因此 IKNN 算法较传统 KNN 修复算法耗时更短,在样本个数为 25 的情况下, IKNN 算法耗时 0.556 4 s,而传统 KNN 算法耗时 2.128 8 s。综上,IKNN 算法在精确度和计算速度上都优于传统 KNN 算法,更适用于修复缺失数据。

4.3　油纸绝缘老化状态评估对比分析

为充分验证本文所提数据清洗方法的实效性,现利用清洗前后的回复电压特征量数据进行油纸绝缘老化状态评估。由于本文数据库样本数量较少,故评估方法采用能较好解决小样本问题的支持向量机(support vector machine, SVM)算法[22]。将 105 组正常数据作为训练集,对比不同测试样本个数状况下的评估正确率,如表 2 所示。

分析表 2 的评估结果可得:

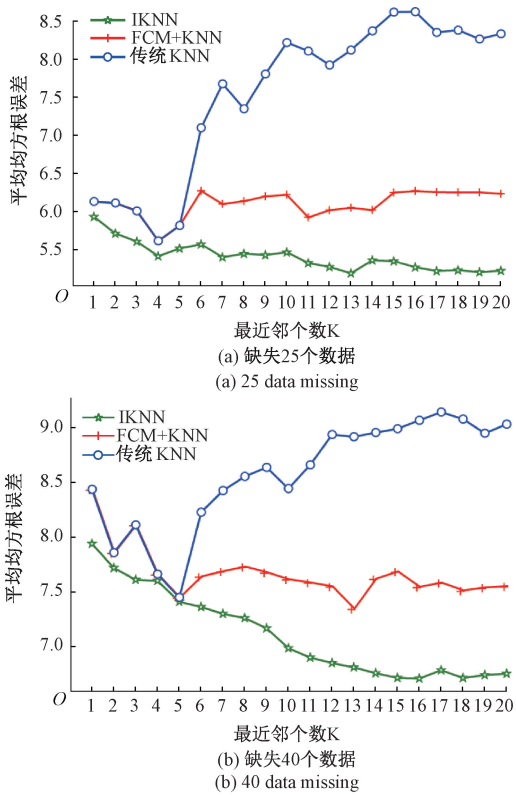


图 9 不同样本数的平均均方根误差

Fig. 9 Mean root mean square error of different sample numbers

1)异常特征量数据使得各组测试样本的评估正确率均下降 50%以上,评估结果大大偏离了实际情况。说明劣质特征量数据对油纸绝缘状态的准确评估有着极大的不良影响。

2)经 KNN 算法和 IKNN 算法修复后,各组测试样本的评估正确率均有明显的提升,且 IKNN 算法比 KNN 算法多提升达 24%左右,充分说明 IKNN 算法具备更良好的缺失数据修复能力。

3)经本文所提方法清洗后,各组测试样本的评估正确率均回升至 85%以上,较异常数据提升了 50%左右的准确率。同无异常数据的正常情况相比,评估正确率下降幅度仅在 8%左右。说明本文所提数据清洗方法可以有效辅助提升绝缘信息质量,实现变压器油纸绝缘状态的精准评估。

5 结 论

本文提出了一种结合 IKNN 算法和 LOF 算法的回复电压数据清洗策略。针对回复电压异常数据来源广、干扰性强等问题,LOF 算法可以准确识别筛除放电不充分、电磁干扰、雾雨潮湿等工况因素产生的异常特征数据,从而避免异常数据对状态评估产生干扰,同时也为后续数据修复奠定夯实的基础。而 IKNN 特征数据修复算法通过 FCM 聚类算法改进了 KNN 算法抗噪能力弱的缺点,同时基于数据归一化和加权欧式距离标度消除数量级差异并突出特征量间关联性,有效提升了算法性能。经分析结果表明,IKNN 算法相较于 FCM-KNN 算法、KNN 算法在不同最近邻个数下均具有更低的平均均方根误差,IKNN 修复后的特征数据比异常数据提高了 50%左右的评估准确率,同时计算效率也获得显著提升,具备良好的工程应用前景。

表 2 不同数据情况下的油纸绝缘老化状态评估正确率

Table 2 Diagnosis results under different data conditions

样本个数	无异常/%	异常/%	KNN 修复/%	FCM-KNN 修复/%	IKNN 修复/%
15	93.33	33.33	53.33	66.67	86.67
25	96	40	56	68	90
35	94.28	38	62.86	71.43	88.57
40	91.11	26.67	51.11	68.89	88.89

本文主要围绕变压器油纸绝缘回复电压数据清洗方法的相关研究展开,而随着电力监测技术的发展,绝缘状态特征数据逐渐呈多源化、异构化的趋势。因此在后续研究中,作者将进一步深入探索油纸绝缘多源特征数据清洗方法的研究,为全面提升电力数据质量提供新的思路。

参考文献

[1] 周亚中,何怡刚,邢致恺,等. 基于 IDBO-ARIMA 的电力变压器振动信号预测[J]. 电子测量与仪器学报, 2023, 37(8):11-20.
ZHOU Y ZH, HE Y G, XING ZH K, et al. Power transformer vibration signal prediction based on IDBO-ARIMA[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(8):11-20.

[2] 邹阳,林超群,叶荣. 油纸绝缘特性研究及微水含量评估[J]. 仪器仪表学报,2020,41(7):117-125.
ZOU Y, LIN CH Q, YE R. Research on oil-paper insulation characteristics and evaluation of its micro-water content[J]. Chinese Journal of Scientific Instrument, 2020, 41(7):117-125.

[3] 殷作洋,吴肖锋,仲伟坤. 基于随机森林的变压器故障识别与实例分析[J]. 电子测量技术,2020,43(23):63-67.
YIN Z Y, WU X F, ZHONG W K. Transformer fault recognition based on random forest and case analysis[J]. Electronic Measurement Technology, 2020, 43(23):63-67.

[4] 袁玉宝,夏经德,刘欢庆,等. 一种考虑电流互感器饱和影响的变压器保护算法[J]. 国外电子测量技术, 2019,38(7):106-111.

YUAN Y B, XIA J D, LIU H Q, et al. Transformer protection algorithm considering the saturation effect of current transformer[J]. Foreign Electronic Measurement Technology, 2019, 38(7):106-111.

[5] 叶荣,蔡金锭. 油纸绝缘极化等效电路的时域介电谱三次微分解析法[J]. 仪器仪表学报,2018,39(6):112-119.
YE R,CAI J D. Analytic method of cubic differential in time domain dielectric spectroscopy for oil-paper insulation polarization equivalent circuit[J]. Chinese Journal of Scientific Instrument, 2018, 39(6):112-119.

[6] 邹阳,林锦茄,李安娜,等. 基于灰色关联分析和聚类云模型的变压器油纸绝缘状态评估[J]. 电力系统保护与控制,2023,51(21):35-43.
ZOU Y, LIN J J, LI AN N, et al. Evaluation of transformer oil-paper insulation status based on grey relational analysis and a cluster cloud model[J]. Power System Protection and Control, 2023, 51(21):35-43.

[7] 林智勇,张达敏,黄国泰,等. 回复电压微分谱线特性的变压器绝缘老化研究[J]. 仪器仪表学报,2017,38(8):1954-1960.
LIN ZH Y,ZHANG D M,HUANG G T, et al. Insulation aging of transformer based on recovery voltage differential spectrum line characteristic[J]. Chinese Journal of Scientific Instrument, 2017, 38(8):1954-1960.

[8] 张宁,蔡金锭. 基于层次分析和逼近理想解法的绝缘状态评估[J]. 仪器仪表学报,2018,39(11):35-42.
ZHANG N, CAI J D. Evaluation of insulation state based on the combination of analytical hierarchy process and TOPSIS[J]. Chinese Journal of Scientific Instrument, 2018, 39(11):35-42.

[9] 彭永志,肖靖,毛建旭,等. 一种基于 DBSCAN 随机圆检测的多瓶口定位算法[J]. 电子测量与仪器学报,2021,35(6):43-52.
PENG Y ZH, XIAO J, MAO J X, et al. Multi bottle mouth positioning method based on DBSCAN random circle detection[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6):43-52.

[10] 傅世元,高欣,张浩,等. 基于元学习动态选择集成的电力调度数据异常检测方法[J]. 电网技术,2022,46(8):3248-3261.
FU SH Y, GAO X, ZHANG H, et al. Anomaly detection for power dispatching data based on meta-learning dynamic ensemble selection[J]. Power System Technology, 2022, 46(8):3248-3261.

[11] 廖伟涵,郭创新,金宇, 等. 基于四阶段预处理与 GBDT 的油浸式变压器故障诊断方法[J]. 电网技术,2019,43(6):2195-2203.
LIAO W H, GUO CH X, JIN Y, et al. Oil-immersed transformer fault diagnosis method based on four-stage preprocessing and GBDT[J]. Power System Technology, 2019, 43(6):2195-2203.

[12] SANJAR K, BEKHZOD O, KIM J, et al. Missing data imputation for geolocation-based price prediction using KNN-MCF method[J]. ISPRS International Journal of Geo-Information, 2020, 9(4):227.

[13] BANIA R K, HALDER A R. R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with KNN imputation for classification of medical data[J]. Computer Methods and Programs in Biomedicine, 2019, 184(4):105122.

[14] ZHAO S, LI J. A fast parameter-free edition algorithm with natural neighbors-based local sets for K nearest neighbor[J]. IEEE Access, 2020,8: 123773-123782.

[15] 陈汉城,蔡金锭. 基于多时域特征参量的变压器油纸绝缘状态综合评估[J]. 电力自动化设备,2017,37(7):184-190.
CHEN H CH, CAI J D. Synthetic insulation state evaluation based on multiple time-domain characteristic parameters for transformer oil-paper[J]. Electric Power Automation Equipment, 2017, 37(7):184-190.

[16] 邹阳,蔡金锭. 变压器极化谱特征量与绝缘状态关系研究[J]. 仪器仪表学报,2015,36(3):608-614.
ZOU Y, CAI J D. Study on the relationship between polarization spectrum characteristic quantity and insulation condition of oil-paper transformer[J]. Chinese Journal of Scientific Instrument, 2015, 36(3):608-614.

[17] 蔡金锭,叶荣,刘庆珍. 基于改进 TOPSIS 和时域特征量的油纸绝缘状态分类分级评估[J]. 电机与控制学报,2020,24(1):86-94.
CAI J D, YE R, LIU Q ZH. Oil-paper insulation classification and grading assessment based on improved TOPSIS and time-domain characteristic parameters[J]. Electric Machines and Control, 2020, 24(1):86-94.

[18] 蔡金锭,王凯. 油纸绝缘变压器非标准极化谱的仿真研究[J]. 电机与控制学报,2014,18(9):49-53.
CAI J D, WANG K. Research on non-standard polarization spectrum of oil-paper insulation transformers[J]. Electric Machines and Control, 2014, 18(9):49-53.

[19] 董泽,贾昊. 基于 EWT-LOF 的热工过程数据异常值检测方法[J]. 仪器仪表学报,2020,41(2):126-134.
DONG Z, JIA H. Outlier detection method for thermal

process data based on EWT-LOF[J]. Chinese Journal of Scientific Instrument, 2020, 41(2):126-134.

[20] 刘仲民,翟玉晓,张鑫, 等. 基于 DBN-IFCM 的变压器故障诊断方法[J]. 高电压技术, 2020, 46(12): 4258-4265.
LIU ZH M, ZHAI Y X, ZHANG X, et al. Transformer fault diagnosis method based on DBN-IFCM[J]. High Voltage Engineering, 2020, 46(12):4258-4265.

[21] 邹阳. 变压器油纸绝缘弛豫响应特性建模及老化诊断研究[D]. 福州: 福州大学,2017.
ZOU Y. Establish model of dielectric relaxation response of transformer's oil-paper insulation and study on diagnosis of aging [D]. Fuzhou: Fuzhou University, 2017.

[22] 陈赛赛,杨晨曦,陈超, 等. 基于小波核扩散与双阶段 SVM 的轴承复合故障分类方法[J]. 仪器仪表学报, 2023,44(10):179-188.
CHEN S S, YANG CH X, CHEN CH, et al. Bearing compound fault classification method based on wavelet kernel diffusion and two-stage SVM[J]. Chinese Journal of Scientific Instrument, 2023,44(10):179-188.

作者简介

陈啸轩,2021 年于西南交通大学获得学士学位,现为福州大学硕士研究生,主要研究方向为电力设备绝缘状态诊断。
E-mail: 646194991@qq.com

Chen Xiaoxuan received a B. Sc. degree from Southwest Jiaotong University in 2021. Now he is a M. Sc. candidate in Fuzhou University. His main research interests include diagnosis of insulation status of power equipment.

邹阳(通信作者),2002 年于福州大学获得学士学位,2012 年于福州大学获得硕士学位,2017 年福州大学获得博士学位,现为现为福州大学副教授,主要研究方向为电气系统智能化故障诊断。
E-mail: 24001744@qq.com

Zou Yang (Corresponding author) received a B. Sc. degree from Fuzhou University in 2002, a M. Sc. degree from Fuzhou University in 2012 and a Ph. D. degree from Fuzhou University in 2017, respectively. Now he is an associate professor in Fuzhou University. His main research interest includes intelligent fault diagnosis of electrical systems.