基于改进 K-means 聚类的变压器异常状态识别模型

谢荣斌,马春雷,张丽娟,靳 斌
(贵州电网有限责任公司贵阳供电局,贵州 贵阳 550001)

摘 要:为有效利用变压器历史正常数据识别变压器是否异常,文章提出了基于改进 K-means 聚类的变压器异常状态识别模型。针对变压器绝大部分运行数据为正常数据,且正常数据逐渐按一定的趋势变化以及异常状态数据变化急剧等特点,基于历史正常数据与 K-means 算法建立变压器异常状态识别模型。根据对正常数据聚类的结果确定用于识别新数据的各个阈值,通过计算新数据到各聚类中心的距离并与各阈值对比确认变压器是否异常。同时针对传统 K-means 算法的缺点,对 K-means 算法进行基于密度与距离选择 K 值与初始聚类中心的改进,使 K-means 算法有稳定的 K 值与聚类中心,聚类过程更加快速、稳定、有效,从而使识别模型计算得到的阈值更可靠。实例分析表明,模型能有效对变压器的异常状态进行快速、准确的识别,为变压器状态评估提供一种新思路。

关键词:变压器;K-means;状态评估;油中溶解气体;异常状态识别
文章编号:2096 - 4633(2018)05 - 0024 - 07 **中图分类号**:TM42 **文献标志码**:B

作为电力系统中关键设备,电力变压器运作是否正常对整个电力系统能否安全、稳定运行具有直接影响。处于异常状态的变压器,会降低电力系统的供电可靠性甚至造成社会经济损失。因此,在运行过程中能提前识别变压器的异常状态,及时对变压器进行状态检修,可减少事故发生的可能性[1]。

目前电力变压器的状态评价方法有层次分析法、贝叶斯网络、支持向量机、人工神经网络等方法[2]。文献[3]使用熵权法改进层次分析法权重确认,既避免主观赋权的相悖性,又考虑专家经验的重要性。文献[4]使用了模糊层次分析法对基于国家电网变压器状态评价规程进行改进,避免了规程中的各状态量权重的主观性。文献[5]基于云理论对状态评估状态量的权重进行最优化计算,文献[6]使用云理论将评估指标的区间模糊化并计算数据与状态等级的关联度作为权重,加权计算变压器的状态等级,两篇文献皆根据获取的数据计算指标权重,使得计算结果较为合理。文献[7]基于支持向量机回归确认各指标的评分权重,文献[8 - 9]基于关联规则挖掘与变权重确认各指标的权重,但以上三篇文献仍参考的固定阈值进行评分。文献[10]综合应用人工神经网络和证据理论构建多信息融合的变压器状态评估模型,将在线监测数据与部分参数的变化趋势结合起来进行计算,取得了较好的效果。文献[11]基于油中溶解气体与支持向量机,使用油中溶解气体的体积分数和产气速率训练分类器,可有效判断变压器状态。文献[12]基于多维度大量的变压器状态数据与分布模型,计算了不同等级变压器的注意值与预警值,相对于传统的阈值更加有效。然而,基于层次分析法与基于固定阈值的状态评估缺少对设备个体化考虑。基于故障样本的神经网络与支持向量机等模型在故障样本足够的情况下效果较好,但对于故障样本较少的电力变压器或者投运时间不长的变压器而言,智能算法难以奏效。同时,目前少有算法利用设备运行数据中极大部分的正常数据对变压器的状态进行分析。

为了解决上述问题,本文将针对投运时间不长或少有发生故障的变压器,建立基于正常状态数据与 K-means 聚类的异常状态识别模型。根据对正常数据聚类的结果确定用于识别新数据的各个阈值,计算新数据到各聚类中心的距离并与各阈值对比,若各个距离都比相应的阈值大,则认为变压器状态异常,若某个距离比相应的阈值小,则认为变压器正

常。为了有效进行正常数据的聚类,对传统 K-means 算法进行基于密度与距离的改进,使得聚类结果更加稳定、可靠,模型计算速度更快。

1　变压器异常状态识别模型

某变压器 2016 年 3 月至 6 月正常运行的部分油中溶解气体数据原始分布如图 1 所示。由图 1 可知,该变压器正常运行时的气体数据是逐渐变化的,不存在数据的突变。而一般的变压器故障会引起气体数据发生急剧的变化。而且,变压器寿命周期内的运行数据绝大部分是正常数据。针对投运不久的变压器,可使用聚类方法对正常数据进行聚类,根据聚类之后数据距离的远近区分正常数据与异常状态数据。

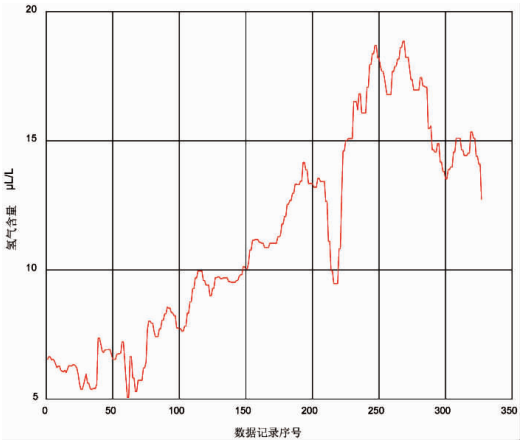


图 1　油中溶解 H_2 含量折线图

Fig. 1　Line graph about content of H_2 dissolved in oil

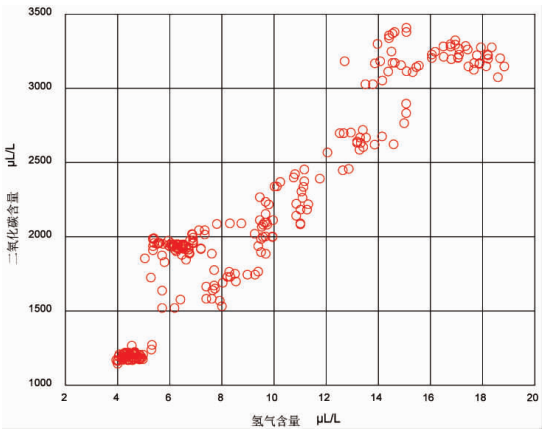


图 2　油中溶解 H_2 与 CO_2 含量分布图

Fig. 2　Distribution diagram about content of H_2 and CO_2 dissolved in oil

如图 2 所示,虽然数据具有明显的变化趋势,但也具有一定的波动性。因此可根据聚类结果计算各个聚类簇内数据距离中心的最大距离,根据距离设定判断阈值,并认为在阈值内波动的数据都是正常的。考虑到数据的变化趋势与波动性,该判断阈值应该比最大距离更大,以防止将正常变化的数据误认为异常数据。根据阈值判断新输入的数据是否为异常状态数据,若是异常状态数据,则需要去检查变压器的状态,反之,则数据应该加入历史正常数据库,同时对聚类结果进行更新。但如果每得到一个正常数据即更新聚类结果,会造成模型计算量较大,而且少量的新数据对聚类结果影响不大,因此可设定数据库更新率达到某个值后,再重新进行聚类。

综上分析,可对变压器的正常数据使用 K-means 聚类,根据聚类结果计算各聚类簇内数据距离中心的最远距离,以该最远距离作为基础阈值,将基础阈值乘一常数 A(为了识别某些波动较大的正常数据,A 应大于 1)得到识别阈值,最后将新输入数据到各聚类中心的距离与各识别阈值比较,若某个距离小于相应的识别阈值,则认为该数据是正常数据并将其加入历史正常数据库,若各个距离都大于各个阈值,则认为该数据是异常状态数据。

由于传统的 K-means 算法具有一定的缺点,因此本文还将对算法进行基于密度与距离的改进,以得到稳定、可靠的聚类结果。

该模型的计算流程如图 3 所示,具体步骤如下。

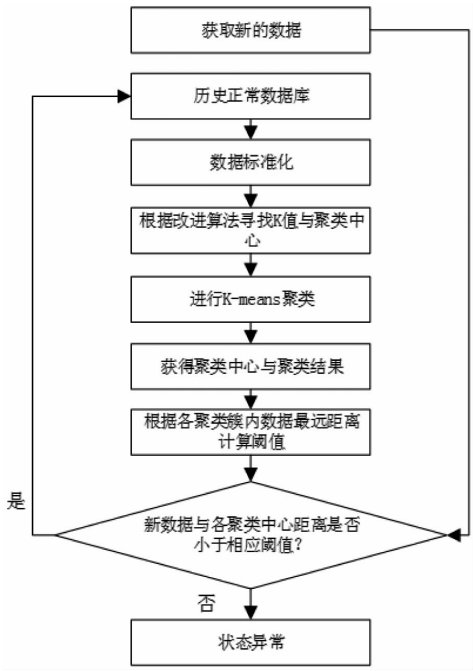


图 3　变压器异常状态识别模型流程图

Fig. 3　Flow chart of power transformer abnormal state recognition model

(1)收集电力变压器的多维度、大量正常运行的历史数据;

(2)对历史数据进行 z – score 标准化;

(3)根据改进的算法寻找 K 值与聚类中心;

(4)使用 K-means 方法进行聚类,并获得聚类中心与聚类结果,并根据各聚类簇内数据与聚类中心的最远距离计算判断阈值;

(5)获取新的数据后,计算数据与各聚类中心的距离并与各个识别阈值相对比;

(6)若某个距离小于相应的识别阈值,则认为该数据是正常数据,若各个距离都大于各个阈值,则认为该数据是异常状态数据。

该基于 K-means 聚类的变压器异常状态识别模型的绝大部分计算量集中在 K 值、初始聚类中心的计算以及聚类过程,同时对新数据的识别过程仅需要判断距离与识别阈值的关系,因此该模型具有稳定性与快速性。

2 K-means 聚类算法原理

聚类分析是数据挖掘领域的一个重要分支,聚类算法可分为基于划分的、基于层次的、基于密度的、基于模型的,其中 K-means 算法是一种基于划分的聚类算法,适合对数值型数据聚类,具有收敛速度快,能有效处理大型数据集[13]。

K-means 算法由 Steinhaus 在 1955 年、Lloyd 在 1957 年、Ball&Hall 在 1965 年、MeQueen 在 1967 年,分别在各自的研究领域中独立提出的一种非监督实时聚类算法,在提出之后该算法得到广泛的研究与应用,并发展存在大量的改进算法[14]。虽然该算法已经被提出超过 50 年了,但目前仍然是应用最广泛的基于划分的聚类算法之一。原理简单、高效是 K-means 聚类算法沿用至今的原因。

K-means 算法的原理为将样本空间 X 的样本分成 K 类,聚类中心为,用表示样本 x_i 与其对应中心 c_j 之间的欧氏距离,样本空间内所有数据点与所属聚类中心距离平方和用目标函数 J 来表示,其数学表达为:

$$J = \sum_{j=1}^{K} J_i = \sum_{j=1}^{K} \sum_{x_i \in c_j} d_{ij} (x_i, c_j)^2 \qquad (1)$$

式中 x_i 为被分配到聚类中心 c_j 的数据,目标函数 J 直接反映聚类效果的好坏,其值越小,则表示该聚类越紧凑、越独立。

聚类过程为:

(1)根据设定的 K 值,随机寻找 K 个数据作为初始聚类中心;

(2)将各数据按最近距离分配给各聚类中心;

(3)根据聚类结果,将各聚类簇的平均位置作为新的聚类中心,再次进行聚类,直至满足收敛条件为止。该算法收敛的条件是当 J 取得极小值、前后两次迭代 J 的差符合精度要求或者达到迭代次数上限。因此可以通过算法的多次迭代,减小目标函数 J 的值来改善与优化聚类结果,当 J 取极小值时的聚类,即为最优聚类方案。

3 K-means 算法的改进

然而 K-means 聚类算法存在对噪声及初值敏感、簇个数需要预先确定及易于陷入局部最优等固有的缺点也导致它们在处理数据时性能下降[15]。针对 K-means 算法随机选择初始聚类中心以及 K 个数需要人为确定的问题,本文对 K 聚类算法进行最佳聚类个数、最佳聚类中心选择的改进。

3.1 传统 K-means 算法的不足

K-means 聚类算法是解决聚类问题的一种经典算法,这种算法具有简单且快速的优点,能够高效处理大数据集。当数据集相对密集、类与类之间距离明显时,K-means 聚类算法能够获得较好的聚类效果。但是,K-means 聚类算法仍存在一些缺陷,在一定程度上影响了它的聚类效果。主要的局限如下:

(1)需要预先定义聚类数目即 K 值。K-means 聚类算法在聚类前需要预先确定参数 K,只有在 K 值已知的前提下算法才能继续执行,但在实际应用中,事先并不知道将数据集分成多少个聚类簇才能获得最优的效果。甚至可以说,需要预先确定 K 值与聚类算法的原始思想是矛盾的。

(2)聚类结果依赖于初始聚类中心的选择,易陷入局部最优解。K-means 算法在确定了聚类数目后,需要随机选取出个初始中心点,然后进入迭代运算,直到目标函数不再变化为止。但是初始聚类中心点完全是随机选取的,初始中心点不同,聚类结果也不同,这导致聚类结果波动范围大,稳定性差。通常,该算法采用的准则函数是最小误差平方和函数,同时存在很多局部最优解,但全局最优解只有一个。聚类算法是基于最小误差平方和函数,这个方法是一种梯度下降算法,所以随机

产生的一种初始中心点就代表了一种搜索方式。沿着不同的搜索方向进行搜索得到的聚类结果往往会是局部最优解。这样导致了该算法严重依赖初始中心点的选取。

(3)易受噪音点和孤立点影响。从聚类算法的迭代过程可以知道,每一次更新的簇类中心是由计算该簇类内所有数据点的平均值获得的。如果该类内的数据相似性较高,则数据较为接近,对更新的中心点影响不大;而如果该类内含有噪音点和孤立点,它们远离数据样本空间,则在计算新的簇类中心时会产生较大波动,会对均值计算产生重大影响,甚至有可能使得新的簇类中心点严重偏离簇类数据样本的密集区域,导致聚类结果有较大偏差。

3.2　对选择 K 值与初始聚类中心的改进

针对上述传统 K-means 算法的问题,本文在分析已有改进算法的基础上提出一种兼顾密度与距离的改进算法[16]。文献[16]中的改进算法以挑选最高密度的两个点作为初步聚类中心,并挑选距离已选取的聚类中心最远的点作为下一个聚类中心,并计算一次迭代划分后的聚类簇相似度确认 K 值,但该改进算法不适合变压器异常状态识别,因为只有高密度点才能作为正常状态数据的聚类中心,聚类后得到的阈值才能有效识别异常状态。本文中的改进算法首先计算数据的平均距离,并计算数据平均距离内的点密度,选取高密度的点为初始聚类中心,最后通过比较各个 K 值下的聚类簇相似度可以确认 K 值的个数。这样做可以使高密度的点优先被选取为聚类中心,可以保证同一个类内的点相对密集,具有高相似性。对于同一个数据集,通过该算法计算得到的聚类中心与 K 值是固定的,这证明了改进后的 K-means 算法具有稳定性,而稳定的聚类中心可使噪音点与孤立点对聚类结果的影响降低。

算法改进的相关定义如下:

(1)数据平均距离 $D_a(M)$:对于含有 n 个数据的数据集 M,对所有数据两两之间的距离相加后求平均,其中数据的欧氏距离用 d 表示。

$$Da(M) = \frac{1}{C_{n\,x_i,x_j\in M}^{2}} \sum d(x_i,x_j) \tag{2}$$

式中 C_2^n 是数据集 M 中任取两个点的组合数。

(2)点密度 $D_e(x,r)$:数据集 M 内与点 x 的距离小于等于 r 的点的数量。

$$De(x,r) = \{p \in M \mid d(x,p) \leqslant r\} \tag{3}$$

(3)聚类簇内平均距离 $D_c(c_i)$:以 c_i 为聚类中心的聚类簇中的 n 个数据到聚类中心的平均距离。

$$Dc(c_i) = \frac{1}{n} \sum_{x\in c_i} d(x,c_i) \tag{4}$$

(4)聚类簇相似度 S:对每一个聚类簇与其他聚类簇之间最大相似度的和求均值。

$$\begin{aligned} S &= \frac{1}{k} \sum_{i=1}^{k} \max\left\{\frac{D_c(c_i) + D_c(c_j)}{d(c_i,c_j)}\right\} \\ &= \frac{1}{k}\left[\max\left\{\frac{D_c(c_1) + D_c(c_j)}{d(c_1,c_j)}\right\}\right. \\ &\quad + \max\left\{\frac{D_c(c_2) + D_c(c_j)}{d(c_3,c_j)}\right\} + \cdots\cdots \\ &\quad \left. + \max\left\{\frac{D_c(c_k) + D_c(c_j)}{d(c_k,c_j)}\right\}\right] \end{aligned} \tag{5}$$

当 S 取最小值时,说明聚类效果最佳,此时的 K 就是最佳的聚类个数。

改进后 K 值与初始聚类中心选择的流程如下:

输入:数据集 M,算法终止条件。

输出:K 值,K 个聚类中心,S 值,目标函数 J。

(1)首先计算数据集 M 中数据两两之间的距离,并计算数据平均距离 $D_a(M)$,然后计算数据的点密度 $D_e(x,R\cdot D_a(M))$(R 为距离系数,过大的距离会将部分孤立点、偏远点算入该点的密度值中,本文中 R 取 0.25),最后将点密度数据组成集合 N。

(2)从集合 N 中选取最大者作为第一个聚类中心,并且将第一个聚类中心和与其距离小于 $R\cdot D_a(M)$ 的点的密度数据从集合 N 中删除。重复在集合 N 中寻找聚类中心直至找到 K 个聚类中心为止。

(3)将数据集 M 中的数据根据找到的 K 个聚类中心进行一次划分,并计算 S 值。

(4)为避免算法的局部最优的缺陷,如果存在两次获得的 S 值大于上一次计算的值,则以 S 值最小时所对应的聚类中心作为 K-means 算法的初始聚类中心,转至步骤(5)。否则,算法继续进行,将 K 值增加 1,转至步骤(2)。

(5)进行 K-means 聚类。

该改进 K-means 算法最后所寻找到的聚类中心是通过计算 K-1 次 S 值所得到的。K 值与聚类中心不再是人为确定或者随机选择,而是通过计算数据分布的密度而产生的,计算结果是稳定的,能有效避免聚类结果的随机性与波动性,因此可提高聚类

结果的准确性。

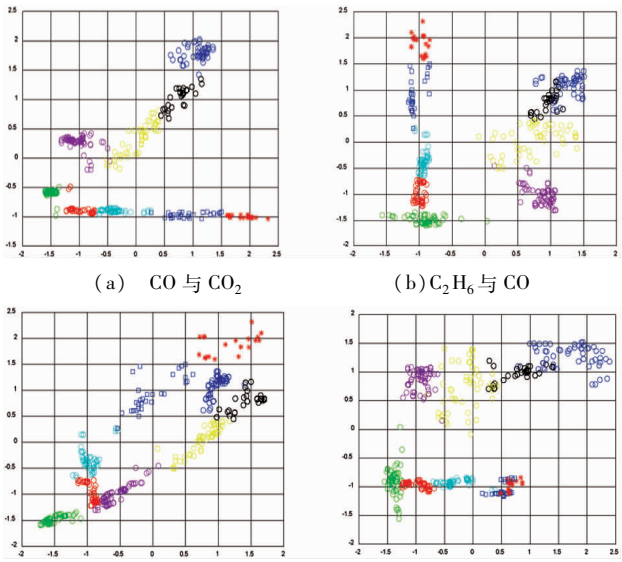
4　实例分析

对某变压器 2016 年 1 月至 8 月的 654 组包含 H_2、CH_4、C_2H_4、C_2H_6、CO、CO_2 等 6 项属性的油中溶解气体数据进行 z－score 标准化后使用改进的 K-means 算法进行 K 值与聚类中心计算。K 值、S 值与聚类一次的 J 值变化结果如表 1 所示，由于 K＝7 与 K＝10 时 S 值比上一个 K 计算的要大，即出现两次 S 值比上一次的大，此时选择 S 最小时的 K，因此最佳的 K 值为 9。由目标函数 J 的值也可以看出来，K＝9 为 J 变化的一个拐点。聚类效果如图 3 所示，图中点的坐标为归一化后数值，因此没有坐标。

表 1　K 值与聚类中心计算结果

Tab. 1　Results of finding K value and cluster centers

K	S	J
2	3. 598 5	6 570
3	1. 859 6	2 308
4	1. 664 8	1 258
5	1. 642 9	1 046
6	1. 407 1	704
7	1. 488 4	665
8	0. 932 7	289
9	0. 904 6	206
10	0. 920 5	177



（a）　CO 与 CO_2　　（b）C_2H_6 与 CO

（c）　CH_4 与 C_2H_4　　（d）H_2 与 CH_4

图 4　变压器油中溶解气体聚类效果图

Fig. 4　Clustering effect diagram of dissolved gas in transformer oil

对该变压器 9 月至 10 月 180 组数据进行识别，其中 32 组异常数据，148 组正常数据。使用改进前后 K-means 聚类算法（K＝9）的识别效果如表 2 所示。

通过对比可知，改进后算法具有更稳定的初始聚类中心，因而能更快迭代达到 J 值最小处，具有更快的收敛时间，同时也能获得更稳定与更小的 J 值。由于改进后的算法选用高密度点作为聚类中心，聚类后计算得到的识别阈值比改进前更小，对异常状态的识别效果更加明显，而对于波动较大的正常数据识别效果稍差于传统算法。

表 2　改进前后的算法时间以及准确率对比

Tab. 2　Comparison of the time and accuracy between traditional K-means and improved algorithm

算法类型	异常状态识别准确率/%	正常状态识别准确率/%	聚类时间/s	迭代次数	J 值
传统	78. 1	97. 8	平均 23. 859	平均 548	平均 672
改进	93. 8	94. 6	0. 566	13	206

6　结论

本文建立变压器异常状态识别模型，并利用改进的 K-means 算法使得聚类结果稳定可靠，通过实例分析进行了验证分析，结果表明：

（1）针对 K-means 算法的需要预先定义聚类数目即 K 值、聚类结果依赖于初始聚类中心的选择、易受噪音点和孤立点影响等缺点，对 K-means 算法进行基于密度与距离选择 K 值与初始聚类中心的改进，使 K-means 算法有稳定的 K 值与聚类中心，算法的迭代次数大大减少，聚类过程更加快速、稳定、有效。

（2）针对变压器绝大部分运行数据为正常数据、正常数据逐渐按一定的趋势变化以及异常状态

数据变化急剧等特点,基于历史正常数据与改进的 K-means 算法建立变压器异常状态识别模型。

参考文献:

[1] 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(23):156 – 167.
LI Gang, YU Changhai, LIU Yunpeng, et al. Challenges and prospects of fault prognostic and health management for power transformer[J]. Automation of Electric Power Systems,2017(23):156 – 167.

[2] 许修乐,李金忠,王健一,等. 变压器可靠性评价及状态评估方法综述[J]. 高压电器,2017,53(08):211 – 216.
XU Xiule,LI Jinzhong,WANG Jianyi,et al. Summary of reliability and state evaluation methods for transformer[J]. High Voltage Apparatus,2017,53(08):211 – 216.

[3] 阳东升,范帅,刘子兴,等. 基于熵值赋权法的配电变压器状态评估方法研究[J]. 南方电网技术,2014,8(04):116 – 121.
YANG Dongsheng, FAN Shuai, LIU Zixing, et al. Study on condition assessment method of distribution transformer based on entropy weight method[J]. Southern Power System Technology, 2014,8(04):116 – 121.

[4] 张晶晶,许修乐,丁明,等. 基于模糊层次分析法的变压器状态评估[J]. 电力系统保护与控制,2017,45(03):75 – 81.
ZHANG Jingjing, XU Xiule, DING Ming, et al. A condition assessment method of power transformers based on fuzzy analytic hierarchy process[J]. Power System Protection and Control,2017, 45(03):75 – 81.

[5] 李孟东,李帅兵,鲜秀明,等. 基于直觉正态云模型和最优变权的变压器绝缘状态评估[J]. 电测与仪表,2016,53(06):42 – 50.
LI Mengdong, LI Shuaibing, XIAN Xiuming, et al. Insulation condition assessment for power transformer based on intuitionistic normal cloud model and optimal variable weights[J]. Electrical Measurement&Instrumentation,2016,53(06):42 – 50.

[6] 杨杰明,董玉坤,曲朝阳,等. 基于区间权重和改进云模型的变压器状态评估[J]. 电力系统保护与控制,2016,44(23):102 – 109.
YANG Jieming, DONG Yukun, QU Zhaoyang, et al. Condition assessment for transformer based on interval weight and improvedcloud model[J]. Power System Protection and Control,2016,44(23):102 – 109.

[7] 张哲,赵文清,朱永利,等. 基于支持向量同归的电力变压器状态评估[J]. 电力自动化设备,2010,30(04):81 – 84.
ZHANG Zhe,ZHAO Wenqing,ZHUYongli,et al. Power transformer condition evaluation based on support vector regression [J]. Electric Power Automation Equipment,2010,30(04):81 – 84.

[8] 谢荣斌,张登,林福昌,等. 基于关联规则与变权重的变压器状态评估方法[J]. 高压电器,2014,50(01):133 – 138.
XIE Rongbin, ZHANG Deng, LIN Fuchang, et al. Transformer condition assessment using association rules and variable weight [J]. High Voltage Apparatus,2014,50(01):133 – 138.

[9] 李黎,张登,谢龙君,等. 采用关联规则综合分析和变权重系数的电力变压器状态评估方法[J]. 中国电机工程学报,2013,33(24):152 – 159.
LI Li, ZHANG Deng, XIE Longjun, et al. A condition assessment method of power transformers based on association rules and variable weight coefficients[J]. Proceedings of the CSEE,2013,33 (24):152 – 159.

[10] 阮羚,谢齐家,高胜友,等. 人工神经网络和信息融合技术在变压器状态评估中的应用[J]. 高电压技术,2014,40(03):822 – 828.
RUAN Ling,XIE Qijia,GAO Shengyou,et al. Application of artificial neural network and information fusion technology in power transformer condition assessment[J]. High Voltage Engineering, 2014,40(03):822 – 828.

[11] 朱永利,申涛,李强. 基于支持向量机和 DGA 的变压器状态评估方法[J]. 电力系统及其自动化学报,2008,24(06):47 – 50.
ZHU Yongli,SHEN Tao,LI Qiang. Transformer condition assessment based on support vector machine and DGA [J]. Proceedings of the CSU – EPSA,2008,24(06):47 – 50.

[12] 齐波,张鹏,徐茹枝,等. 基于分布模型的变压器差异化预警值计算方法[J]. 高电压技术,2016,42(07):2290 – 2298.
QI Bo. ZHANG Peng, XU Ruzhi, et al. Calculation method on differentiated warning value of power transformer based on distribution model[J]. High Voltage Engineering,2016,42(07): 2290 – 2298.

[13] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(01):48 – 61.
SUN Jigui,LIU Jie,ZHAO Lianyu. Clustering algorithms research [J]. Journal of Software,2008,19(01):48 – 61.

[14] 王千,王成,冯振元,等. K-means 聚类算法研究综述[J]. 电子设计工程,2012,20(07):21 – 24.
WANG Qian,WANG Cheng,FENG Zhenyuan,et al. Review of K-means clustering algorithm[J]. Electronic Design Engineering, 2012,20(07):21 – 24.

[15] 陈建娇. 高维数据的 K – harmonic Means 聚类方法及其应用研究[D]. 上海:上海大学,2012.

[16] 程艳云,周鹏. 动态分配聚类中心的改进 K 均值聚类算法[J]. 计算机技术与发展,2017,27(02):33 – 36.
CHENG Yanyun, ZHOU Peng. Improved K-means clustering algorithm for dynamic allocation cluster center[J]. Computer Technology and Development,2017,27(02):33 – 36.

收稿日期:2018 – 02 – 20

作者简介:

谢荣斌(1970),男,高级工程师,硕士,从事高电压技术研究与管理工作。

(本文责任编辑:龙海丽)

Power transformer abnormal state recognition model based on improved K-means clustering

XIE Rongbin, MA Chunlei, ZHANG Lijuan, JIN Bin

(Guiyang Power Supply Bureau of Guizhou Power Grid Co., Ltd., Guiyang 550001 Guizhou, China)

Abstract: For the situation that most of the data of transformers with low running time are normal data, in order to utilize the historical normal state data to identify efficiently whether the new data is abnormal, a power transformer abnormal state recognition model based on improved K-means clustering, is proposed in this paper. Now that most of the power transformer operation data are normal state data, and the normal state data gradually changes according to a certain trend while the abnormal state data changes rapidly, based on the historical normal data and K-means clustering, a recognition model for power transformer is established. According to the clustering results of normal data, thresholds can be acquired to identify new data. By calculating the distances from the new data to the cluster centers and comparing the distances with the thresholds to determine whether the transformer is abnormal. To solve the problem of traditional K-means algorithm, an improvement about choosing K value and initial cluster center based on data density and distance is proposed, which can give K-means clustering stable cluster centers and K value to make the procession of clustering quicker, more stable and efficient. The example analysis shows that the recognition model can effectively identify the abnormal state of the power transformer quickly and accurately and that the model provides a new idea for the state assessment of transformers.

Key words: transformer; K-means; state assessment; dissolved gas in oil; abnormal state recognition