

基于改进 FCM 的变压器故障模糊分类方法

代子阔¹, 赵庆源², 李飞², 翟兴², 赵帅科², 谭玉华³

(1. 国网辽宁省电力有限公司, 沈阳 110000; 2. 国网本溪供电公司, 辽宁 本溪 117020;
3. 上海理工大学 机械工程学院, 上海 200093)

摘要: 变压器油中溶解气体分析(DGA)是识别变压器的故障类型的一项重要技术, 模糊聚类是一种有效的分析手段。但传统模糊聚类算法存在对随机初始化的聚类中心敏感、隶属度函数的有效度量范围较小使其容易陷入局部极值点的问题, 因而实际分类效果不佳。针对传统 FCM 的不足, 首先采用 Canopy 算法对 DGA 数据进行粗聚类, 将其结果作为后续 FCM 聚类的初始聚类中心和最佳聚类数, 降低了人为和随机初始化参数的主观性; 然后通过引入负指数函数形式的相似度指标重构了 FCM 隶属度的迭代函数, 降低了算法陷入局部极值点的可能性; 最后通过对故障气体数据进行实例分析, 验证了改进后的算法在识别变压器故障类别上的有效性和实用性。

关键词: 溶解气体分析; 故障模糊分类; 变压器; fuzzy C-means; 局部极值

DOI: 10.15938/j.jhust.2023.04.014

中图分类号: TM72 **文献标志码:** A **文章编号:** 1007-2683(2023)04-0118-09

Fuzzy Classification Method for Transformer Faults Based on Improved FCM

DAI Zikuo¹, ZHAO Qingyuan², LI Fei², ZHAI Xing², ZHAO Shuaike², TAN Yuhua³

(1. State Grid Liaoning Electric Power Company Limited, Shenyang 110000, China;

2. State Grid Benxi Power Supply Company, Benxi 117020, China;

3. School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Dissolved gas analysis (DGA) in transformer oil is an important technology to identify transformer fault types, and fuzzy clustering is an effective analysis method. However, the traditional fuzzy clustering algorithm is sensitive to the randomly initialized cluster center and the effective measurement range of membership function is small, which makes it easy to fall into local extremum points, so the actual classification effect is not good. To address the shortcomings of the traditional FCM. Canopy algorithm was first used to conduct rough clustering of DGA data, and the results were used as the initial cluster center and optimal cluster number of subsequent FCM clustering, which reduced the subjectivity of artificial and random initialization parameters. Then, the iterative function of FCM membership was reconstructed by introducing the similarity index in the form of negative exponential function, which reduced the possibility of the algorithm falling into local extremum points. Finally, the effectiveness and practicability of the improved algorithm in transformer fault classification are verified by analyzing the fault gas data.

Keywords: dissolved gas analysis; fuzzy fault classification; transformer; fuzzy C-means; local extreme value

收稿日期: 2022-02-15

基金项目: 国家自然科学基金(51777126); 国网公司科技项目(2021YF-19).

作者简介: 代子阔(1985—), 男, 高级工程师;

赵庆源(1981—), 男, 高级工程师.

通信作者: 谭玉华(1989—), 女, 博士, 讲师, E-mail: tan.yuhua@usst.edu.cn.

0 引 言

电力变压器是电力系统中最重要设备之一,其安全可靠运行是保证配电网可靠性的重要条件,变压器的故障类型识别对提升运检效率,保证其可靠运行有着重要的作用。变压器正常运行时,变压器内部的绝缘材料受到电、热、水等的共同作用,在一些金属材料的催化下,变压器不断老化,这个过程中产生了一些气体。当变压器内部逐渐形成故障时,气体产生速率会不断变大并在油中积累。不同类型的气体往往代表着不同的故障类型,如一些放电故障会大量出现的气体,会在一些过热性故障时大量出现。因此,变压器油中溶解气体分析 (dissolved gas analysis, DGA) 是识别故障类型的有效手段^[1-5]。

模糊数学、神经网络、传统 DGA 分析法等方法都在变压器故障分类上取得了较大进展。传统方法如特征气体法、三比值法、大卫三角形法 etc 都在变压器故障类型识别上得到了广泛应用^[6-8]。相比而言,模糊聚类分析方法引入了隶属度的概念,分析隶属于各种故障的程度,而不是非此即彼的判断,更符合实际情况。模糊 C 均值 (Fuzzy C-means, FCM) 是最典型的模糊聚类算法,也被广泛应用到变压器的故障类型识别中,有诸多文献对该问题进行了研究^[9-11]。但是,传统 FCM 在识别变压器的故障类型时,由于算法本身的缺陷往往难以获得较好的分类效果:初始聚类中心是随机产生的、最佳聚类数也是人为确定的,另外 FCM 采用距离倒数的幂作为相似度指标使得样本在聚类中心较小范围内的聚类效果较好,而 FCM 的分类结果对初始值较敏感,若随机选择的初始值质量欠佳,传统 FCM 在迭代过程中就容易陷入局部最优,最终使得变压器的模糊分类效果不佳。

针对上述问题,国内外学者开展了相关研究。文[12]利用遗传算法和模拟退火算法对传统 FCM 的搜索性能进行改进,利用油中溶解气体数据实现变压器故障类型的诊断分类;文[13]提出了一种改进的概率密度函数初始化聚类中心的模糊 C 均值聚类算法,并与三比值法和传统 FCM 算法相比较,验证了其有效性。上述文献虽然有助于提高正确识别故障类型,但聚类数仍是人为确定的。文[14]利用 Canopy 算法无需手动指定聚类数的优点,结合 K-Means 算法对电力客户进行划分。文[15]将 Canopy 算法与传统 FCM 算法结合,对三峡水库的

水质样本数据进行分类,判断出水质状况,提高聚类效率。Canopy 算法可用于自行确定分类数,但是该算法尚未应用于变压器油中溶解特征气体的故障分类中。

综上所述,本文首先通过 Canopy 算法对数据进行粗聚类,得到最佳聚类数和聚类中心,然后将其作为 FCM 聚类分析的初始条件进行二次聚类。最后通过重构负指数形式的隶属度迭代函数,在一定程度上降低了算法陷入局部最优的可能性,并利用实际 DGA 数据进行验证分析。

1 FCM 算法分析 DGA 的原理及存在的问题

1.1 FCM 分析 DGA 数据的原理

国际电工委员会 (international electrotechnical commission, IEC) 对充油变压器的故障进行了分类,其关系如图 1 所示。从物理本质上将其分为放电型故障、局部放电故障和过热型故障三大类^[14-18]。放电型故障的本质是变压器受到放电能量后其绝缘介质被击穿而导致的故障,根据其受到的能量密度的高低将其再细分为电弧放电故障和火花放电故障;将变压器温度在 300 ~ 700℃ 之间的故障称为过热型故障,又根据温度大小细分为高温、中温和低温过热故障^[19-21]。

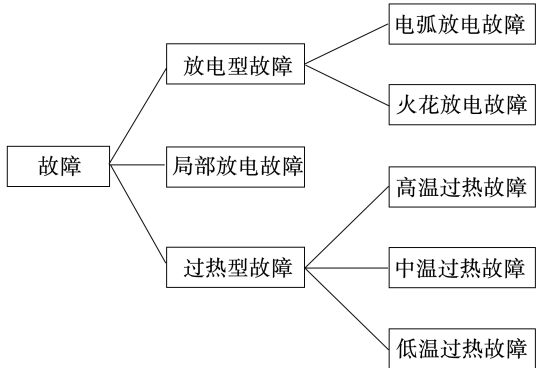


图 1 充油变压器故障类型

Fig. 1 Oil-filled transformer fault types

变压器绝缘油主要是由 C、H 元素组成的化合物,当其受到热能和电能的作用时,会生成一些不稳定的游离基,它们的重新组合就会生成不同的故障气体。若变压器出现过热型故障,碳氢化合物分子的化学键受热断裂,当温度较低时,主要分解物为氢气和低分子烃类,其中低分子烃类以 CH₄ 为主;随着温度的不断升高,生成的低分子烃类会继续进行分解,当温度大于 500℃ 时通过不断地分解组合生成较多的 C₂H₄ 和 H₂,而温度大于 800℃ 时则会产

生少量 C_2H_2 。若变压器出现放电型故障,其产生的能量同样会破坏绝缘油的化学键,当生成的能量较少时,主要通过离子反应破坏 C-H 键,而 C-C 键的破坏则需要更高的能量,因此,局部放电时的特征气体以 H_2 、 CH_4 为主,而火花放电和电弧放电时又会产生 C_2H_2 、 C_2H_4 、 C_2H_6 等气体。

这 5 种气体的量与不同的故障类型有着非常紧密的关系,变压器发生同一种类型的故障时分解产生的气体成分具有相似性,因此,可以通过判断变压器油中这些特征气体的含量进行故障类别的判断。但对同属于过热型或放电型的故障而言,实际中并无绝对划分界限,如并没有绝对的温度值来区分低温、中温、高温故障,其本质上是 C-H 键和 C-C 键在热量或能量的作用下的断裂和重组,只是诸如破坏 C-C 键可能需要更多的能量,而 FCM 算法就放弃了非此即彼的硬划分方法,采用隶属度描述特征气体数据属于某一故障类型的程度,也更符合变压器故障生成特征气体的实际情况。通过 FCM 聚类算法对变压器内部监测到的气体进行分类处理,使得每一类中各特征气体的含量占比尽可能相似,表示同一种故障类型,实现变压器的故障诊断分类。

FCM 利用模糊数学思想求出隶属于各类别的程度,在工程实际问题中得到了广泛的应用。假设待分类的特征气体的数据集 $X = \{x_1, x_2, \dots, x_n\}$, 每组数据都是 5 个维度,分别表示 5 种特征气体, $x_i = (x_{i1}, x_{i2}, \dots, x_{i5})$ 。定义欧氏距离 d_{ij} 作为样本 x_i 与聚类中心 c_j 的相似性度量, $d_{ij} = \|x_i - c_j\|$ 。定义隶属度 μ_{ij} 为样本 x_i 对聚类中心 c_j 的归属程度,且样本 x_i 对所有聚类中心的隶属度之和为 1, 即 $\sum_{j=1}^c \mu_{ij} = 1$ 。FCM 算法的分类原理是,在满足上述约束条件的基础上,通过将类内的加权误差平方和最小化来获得最优分类:

$$\min J(U, V) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m d_{ij}^2 \quad (1)$$

$$s. t. \sum_{j=1}^c \mu_{ij} = 1$$

式中: m 是一个表示模糊程度的权重指数, $m \in [1, +\infty)$, 一般取 2。

通过拉格朗日乘数法获得 FCM 算法中 μ_{ij} 和 c_j 的迭代公式:

$$\left. \begin{aligned} \mu_{ij} &= \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}}, \\ c_j &= \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m}, \end{aligned} \right\} 1 \leq j \leq c, 1 \leq i \leq n \quad (2)$$

FCM 算法对变压器内特征气体的分类步骤为:

1) 特征气体数据的预处理。在分类之前,需要对原始数据进行标准化,以消除不同气体之间数量级上的差异。

2) 初始化。随机初始化聚类中心 $C(0)$ 和隶属度矩阵 $U(0)$, 迭代次数 $k=0$, 收敛阈值 $\varepsilon = 10^{-3[9]}$ 。

3) 由式(2)迭代求取新的聚类中心 $C(k)$ 和隶属度矩阵 $U(k)$, $k = k + 1$ 。

4) 不断重复步骤 3), 直到满足收敛条件。收敛条件为最后两次迭代的隶属度之差小于收敛阈值 ε 。

理论上,FCM 算法结束后,每一类内 5 种特征气体含量占比最为相似,每一类之间最不相似,每一个类别都对应着一种故障类别。但是,由于传统 FCM 算法存在着诸多不足,导致在实际的特征气体数据聚类时,会出现聚类效果不佳等情况。

1.2 FCM 在分析 DGA 中存在的问题

初始聚类中心。在传统 FCM 算法中,初始聚类中心和聚类数都是通过随机初始化或人为设定的,聚类结果的好坏依赖于初始值的选取。在 DGA 数据分类的应用中,有可能出现聚类效果不佳,故障类型识别不准确等情况^[22-23]。

局部极值问题。由式(2)可知,传统 FCM 用 d_{ij} 倒数的幂来描述样本与聚类中心的相似度,再对其标准化得到隶属度。当样本离聚类中心较近时,相似度迅速增大;稍远时相似度便迅速减小。换句话说,传统 FCM 算法仅对聚类中心附近的数据具有较好的分类效果,有效的刻画区间比较小,容易陷入局部极值的问题。

2 基于改进 FCM 的变压器 DGA 故障分类

基于第 1 节中对传统 FCM 算法的原理和缺陷的分析,本节对其做针对性的改进,并将改进后的 FCM 算法应用到变压器 DGA 故障分类中。

2.1 聚类数和聚类中心的确定

Canopy 算法是一种无需事前指定聚类个数的聚类算法,具有较大的实际应用价值^[24]。但由于其聚类的精度较低,本文考虑在进行 FCM 聚类之前使用 Canopy 对预处理后的特征气体数据进行“粗”聚类,将 Canopy 的聚类结果作为后续 FCM 聚类算法的初始输入参数。FCM 的初始聚类中心和聚类个数已经由 Canopy 算法所确定,无需再认为主观设定。Canopy 算法的聚类流程如下:

步骤 1: 设置两个距离划分阈值 T_1 和 T_2 , 且 $T_1 > T_2$ (T_1 和 T_2 的值可以通过交叉验证来确定, 一般经验值为 $T_1 = 2d_{ij, avg}$, $T_2 = 2T_1$)^[25-26]。

步骤 2: 从预处理后的数据集中任取一组特征气体样本点 X , 作为第 1 个 Canopy 并将 X 从数据集中删去。

步骤 3: 计算数据集中所有点与 Canopy 点的欧氏距离 d 。如果 $d < T_2$, 则将相应的样本点归入该 Canopy 集中; 如果样本点与所有 Canopy 点的欧氏距离都满足 $d > T_1$, 则可将该点视为一个新的 Canopy 集; 如果样本点与某个 Canopy 点的欧氏距离都满足 $d < T_1$, 但与其他 Canopy 点的欧氏距离都不满足 $d < T_2$ 时, 同样将其视为一个新的 Canopy 集。

步骤 4: 重复步骤 3, 直到数据集为空集。

采用 Canopy 算法的好处是: FCM 聚类的初始聚类中心和聚类数完全可以由 Canopy 算法确定, 避免了人为确定的主观性; FCM 算法对噪声抗干扰较弱, 通过 Canopy 对比较小的 NewPoint 的簇直接去掉, 有利于抗干扰。

2.2 隶属度迭代函数的重构

针对传统 FCM 算法中隶属度仅对聚类中心非常近的数据有较好的刻画效果, 容易陷入局部极值点的问题, 本文对 FCM 的隶属度迭代函数进行了重新构建。

首先, 构建了基于负指数函数的相似度指标:

$$\eta_{ij} = e^{-ad_{ij}} \quad (3)$$

式中: a 为灵敏系数, 用于调整相似度函数的灵敏程度; d_{ij} 为两点之间的欧氏距离; η_{ij} 为两点的相似程度。

然后, 对某样本对所有聚类中心的相似指标进行标准化, 构建隶属度函数:

$$\mu_{ij} = \frac{\eta_{ij}}{\sum_{k=1}^c \eta_{ik}} = \frac{e^{-ad_{ij}}}{\sum_{k=1}^c e^{-ad_{ik}}} \quad (4)$$

通过重构后的迭代函数可知两点之间的相似度控制在 0 到 1 之间, 较原有模型变得较为平缓, 度量的有效范围变宽, 在一定程度上减小了陷入局部极值点的可能。样本对聚类中心的隶属度随着对聚类中心的远离单调递减。

不同的灵敏度取值也会影响算法陷入局部的概率, 对于标准化数据, 假设任意两组 DGA 数据欧式距离的最大值为 1, 如图 2 所示。当 $x = d_{ij}$ 时, a 取 1, 4, 10 时的相似度度量范围分别为: $\eta_1 = [0.3679, 1]$, $\eta_4 = [0.0183, 1]$, $\eta_{10} = [0.000045, 1]$ 。其中,

当 $a=1$ 时, 虽然相似度函数最为平缓但其取值范围相比于 $[0, 1]$ 误差较大; 当 $a=4$ 和 $a=10$ 时, 虽然其取值区间相较于 $[0, 1]$ 而言都能满足一定的精度, 但当 $a=10$ 时, 在 $d_{ij} \in [0, 0.4]$ 甚至 $d_{ij} \in [0, 0.6]$ 时的刻画效果更好, 而距离在 $[0.6, 1]$ 之间的样本就会面临如传统 FCM 同样的问题, 即在原理聚类中心时相似度急剧减小。当灵敏度系数取一个足够大的正数时, 相似度的有效度量范围就会像传统 FCM 算法一样较小。相比而言, 当 $a=4$ 时, 既能满足相似度取值范围精度, 相似度函数又相对平缓, 在距离的取值区间内, 更为平缓地获得样本的 $[0, 1]$ 相似度, 进而标准化得到隶属度。因此, 当 $x = d_{ij}$ 时, 相似度 $\eta = e^{-ax}$ 在满足取值区间在 0 到 1 之间的情况下, 灵敏度系数 a 的取值越小越好。

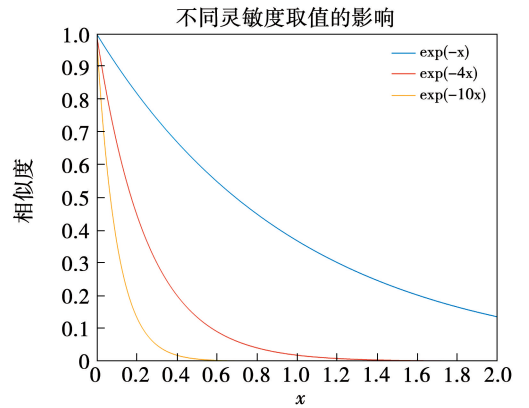


图 2 不同灵敏度系数的影响

Fig. 2 Influence of different sensitivity coefficients

2.3 聚类有效性指标

在得到变压器的故障分类结果后, 需要对本文所改进的方法和传统 FCM 算法的分类结果的有效性、优劣性进行对比分析。变压器故障分类的目的是通过分析 5 种特征气体的含量来对其故障类型进行区分, 这就要求算法得到的属于同一类故障数据之间的相似度越高越好, 而不同类型故障数据间的差异性越大越好。因此, 本文采用 DBI (davies-bouldin index) 指标同时考虑聚类结果的类内和类间的聚类效果, 其表达式为

$$I_{DBI} = \frac{1}{c} \sum_{k=1}^c \max_{k \neq t} \left(\frac{D_k + D_t}{D_{kt}} \right) \quad (5)$$

式中: c 为聚类个数; D_k 为第 k 个类中数据集与该类聚类中心的平均欧式距离, 用于反映第 k 个类中数据集的聚拢程度; D_{kt} 为第 k 类中心与第 t 类中心的欧氏距离, 用于反映类间第 k 类与第 t 类数据的分散程度; I_{DBI} 为类内距离之和与类外距离的比值, 其

计算公式简单且指标值变化范围小,能直观反映聚类质量,DBI 指标适用于评定变压器故障模糊分类结果的有效性和优劣程度: I_{DBI} 越小聚类效果越佳。另外,考虑到传统 FCM 容易陷入局部极值的问题,有时难以得到有效的变压器故障分类结果,存在对故障类型的误判。因此,本文建立了 EI (effectiveness index) 指标进一步描述聚类算法总体的有效性,其表达式为:

$$I_{\text{EI}} = \frac{\eta}{I_{\text{DBI}}} \quad (6)$$

式中: η 为算法得到的故障分类结果的有效率。 EI 指标用于评定变压器故障分类结果的总体有效性和优劣性: I_{EI} 越大聚类效果越佳。

2.4 变压器 DGA 数据的故障模糊分类

根据收集到的变压器特征气体数据集 $X = \{x_1, x_2, \dots, x_n\}$ 对其进行聚类分析。其中, $x_i = (x_{i1}, x_{i2}, \dots, x_{i5})$, x_i 中的元素分别表示 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 和 C_2H_6 5 种特征气体。通过对特征气体含量的相似特性,将相似的数据归为一类,并认为该类是由同一种类型的故障产生的。基于改进 FCM 的变压器 DGA 故障分类步骤流程图如图 3 所示。具体步骤为:

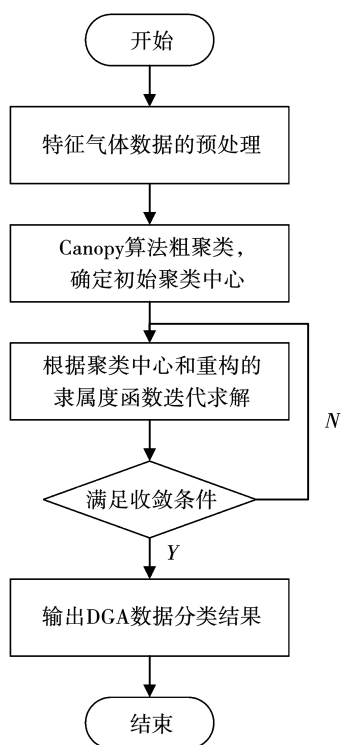


图 3 变压器故障模糊分类流程图

Fig. 3 Flow chart of fuzzy classification of transformer faults

步骤 1:特征气体数据的预处理。在分类之前,需要对原始数据进行处理,以消除不同气体之间数量级上的差异。

步骤 2:对预处理之后的数据集采用 Canopy 算法进行粗聚类,确定聚类数目 c 和聚类中心集 $C = [C_1, C_2, \dots, C_c]$ 。

步骤 3:初始化迭代次数 $k=0$,收敛阈值 ε 。

步骤 4:由式(3)、式(4)得到新的隶属度矩阵 $U(k)$,再根据式(2)得到新的聚类中心, $k=k+1$ 。

步骤 5:不断重复步骤 4,直到满足收敛条件。收敛条件为最后两次迭代的隶属度之差小于收敛阈值 ε 。

3 算例分析

变压器中的绝缘油是由多种碳氢化合物组成的混合物,变压器在正常工作时产生的热量不足以对这些化合物的化学结构造成破坏,但是当发生一些电力故障或热故障时产生的能量会导致这些碳氢化合物分子内部的部分化学键断裂,形成氢气和低分子烃类气体。大量实践研究结果显示,发生不同类型的故障会断裂碳氢化合物的不同化学键,产生的气体类型也会不同。变压器故障时产生特征气体的速率按从大到小依次为 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 。

3.1 验证过程

本文根据文[13]中收集到变压器油色谱数据信息,选择其中 90 条故障数据开展变压器的故障分类,并对比改进前和改进后的 FCM 算法对 DGA 数据进行有效分类的能力。验证过程如下所示:

1)数据初始化处理。5 种特征气体中,各种气体含量差别较大,直接采用气体含量的原始数据将会湮灭掉含量较少的气体所代表的故障类型。含有 H 元素的 5 种气体被称为氢烃,含有 C 、 H_2 种元素的 4 种气体被称为总烃。根据 5 种特征气体所占比重进行初始化: H_2 占氢烃的百分比, CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 占总烃的百分比。

2)利用 Canopy 算法对处理后的特征气体数据进行聚类分析,算法运行 100 次,得到若干种不同聚类结果,并从其中获得最佳分类数和相应的聚类中心。

3)利用改进隶属度迭代方法的 FCM 进行二次聚类。

4)判断聚类结果中每种类型所表示的故障类型。统计每一种类别中每种故障的数目,数量最大者即为该类的故障。

5)分类的正确性、有效性判定。如果一个类别中包含两种及以上的故障类型或两个类别代表同一种故障,则认为该分类结果是无效的;若每个类别都代表者不同的故障类型,则认为该分类过程有效。

6)分析对比改进前后算法的性能。

3.2 改进 FCM 算法的聚类结果

基于 Canopy 算法的 100 次聚类,共得到了 3 种不同的聚类结果,具体如表 1 所示。

表 1 Canopy 聚类结果

Tab.1 Canopyclustering results

聚类数	次数
3	0
4	1
5	3
6	96
7	0

在 100 次聚类中,有 96 次的最佳聚类数是 6 类,而其他数目的聚类数出现的次数只有 4 次。由本文所筛选的样本数据可知,把故障类型分为 6 类更符合实际情况。在此基础上,将这 96 次结果中对应的聚类中心输入到改进的 FCM 算法中,取灵敏度系数为 10,共得到 3 种模式。根据 3.1 节中分类结果的判断准则可知,只有第一种结果是有效的,剩余两种结果是不合理的。得到的 3 种结果的聚类中心如表 2~4 所示。其中,5 种特征气体的数据都是经过初始化处理的数值。

表 2 模式 1

Tab.2 Mode 1

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂
低温过热	0.323 7	0.507 7	0.178 9	0.312 8	0.000 6
中温过热	0.092 9	0.364 2	0.207 1	0.426 5	0.002 2
高温过热	0.151 3	0.276 3	0.126 3	0.594 3	0.004 7
局部放电	0.909 2	0.504 8	0.287 9	0.191 2	0.016 1
低能放电	0.435 6	0.252 5	0.056 1	0.147 8	0.543 6
电弧放电	0.468 3	0.253 3	0.045 6	0.320 9	0.380 2

表 3 模式 2

Tab.3 Mode 2

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂
低温过热	0.325 1	0.508 1	0.176 9	0.313 2	0.001 8
中温过热	0.093 8	0.363 8	0.206 8	0.426 5	0.002 9
高温过热	0.151 4	0.276 3	0.124 2	0.594 3	0.005 2
局部放电	0.908 4	0.505 2	0.288 5	0.191 7	0.014 7
局部放电#	0.908 4	0.505 2	0.288 5	0.191 7	0.014 7
无效分类#	0.455 2	0.250 3	0.051 4	0.231 5	0.466 7

表 4 模式 3

Tab.4 Mode 3

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂
低温过热	0.326 2	0.508 1	0.176 9	0.313 2	0.001 8
中温过热	0.094 4	0.363 8	0.206 8	0.426 5	0.002 9
高温过热	0.149 6	0.283 7	0.121 8	0.587 9	0.006 6
局部放电	0.907 9	0.505 5	0.288 4	0.191 5	0.014 6
局部放电#	0.894 5	0.523 8	0.277 3	0.186 0	0.012 9
无效分类#	0.148 8	0.239 1	0.165 8	0.589 0	0.006 1

在 3 种聚类结果中,只有第 1 个模式的分类结果是有效的,其中有 89 次得到模式 1 的聚类结果,其余 7 次得到了模式 2、3 的聚类结果。在模式 2、3 中,无“#”的分类为算法正确识别故障类型的数据,而带“#”的分类结果都是存在一定问题的。例如,在模式 2 中,算法成功将高、中、低温过热和局部放电故障进行分类,但是却将低能放电和电弧放电的数据归为一类,并且有两个类别的聚类中心重合,都是由局部放电造成的故障。

进一步地,在分类得到模式一的情况下,90 条数据的分类正确率如表 5 所示,总计正确率为 92.2%。

表 5 改进 FCM 算法的正确判别率

Tab.5 Correct discrimination rate ofthe improved FCM algorithm

故障类型	实际样本数	正确分类数	正确率
低温过热	10	9	90%
中温过热	15	14	93.3%
高温过热	22	20	90.9%
局部放电	15	14	93.3%
低能放电	10	9	90%
电弧放电	18	17	94.4%
合计	90	83	92.2%

在本文中,灵敏度系数被用来调整相似度的有效刻画区间,降低隶属度陷入局部极值的概率。为了体现灵敏度系数的不同取值对识别结果的影响,分别选取 $a \in [5,15]$ 时,在运行 100 次改进 FCM 算法后,得到有效结果模式 1 的概率如图 4 所示,得到模式 1 前提下的正确识别率如图 5 所示。

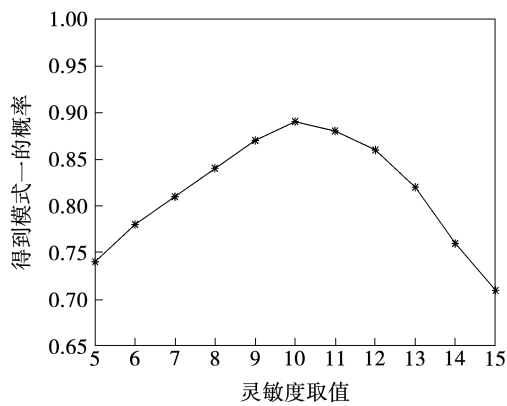


图 4 不同灵敏度系数得到模式 1 的概率

Fig. 4 The probability of getting mode one with different sensitivity coefficients

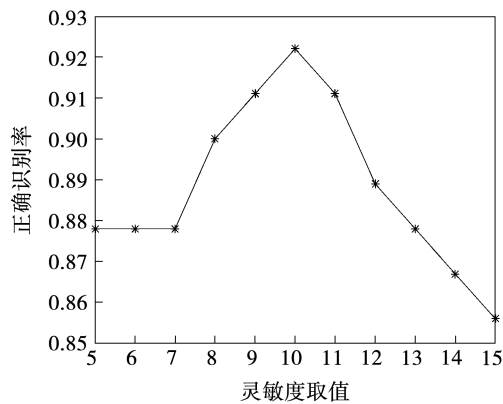


图 5 不同灵敏度系数的正确识别率

Fig. 5 Correct recognition rate of different sensitivity coefficients

由图可知,当灵敏度取 10 时,得到有效分类结果模式一的次数最多为 89 次,且在得到模式一的基础上正确识别故障类型的概率最高为 92.2%。当灵敏度大于或小于 10 时,虽然都相比于传统 FCM 算法的结果更优,但均低于灵敏度系数为 10 时的聚类结果。

3.3 对比分析

采用传统 FCM 算法对故障样本数据进行 100 次分类,得到的聚类结果与采用 Canopy + 传统 FCM 和本文改进 FCM 法(灵敏度取 10)得到的聚类结果进行对比,结果如表 6 所示。

表 6 聚类结果对比
Tab. 6 Comparison of clustering results

方法	传统 FCM	Canopy + FCM	本文改进的 FCM
分类结果有效次数	68	76	89
故障正确判别率	84.4%	87.8%	92.2%

传统 FCM 容易陷入局部极值点,很难得到有效的变压器故障分类结果,由于传统算法本身的局限性,在 100 次随机初始化聚类中心得到的聚类结果中,只有 68 次能够得到有效的聚类结果。而采用 Canopy + 传统 FCM 算法和本文改进的 FCM 算法得到的聚类结果中,分别有 76 次和 89 次能够得到有效分类模式。相较于传统 FCM 而言,聚类结果的有效性分别提升了 11.8% 和 30.8%。在聚类结果是有效的基础上,传统 FCM 算法对故障类型的正确判别率为 84.4%,而采用 Canopy + 传统 FCM 算法和本文改进的 FCM 算法的正确识别率可达 87.8% 和 92.2%。改进后的算法在一定程度上提升了识别故障类型的正确率。

传统 FCM、Canopy + FCM 和本文改进 FCM 算法得到的收敛曲线如图 6 所示。其中,传统 FCM 算法耗时 0.759 4 s,经过 54 次迭代最终收敛值为 11.032;Canopy + FCM 算法耗时 0.697 2,经过 47 次迭代最终收敛到 10.537,由于 Canopy 算法的高效性和更少的迭代次数,使得算法耗时略有减少;而本文改进的 FCM 算法耗时 0.572 5 s,且只经过 38 次迭代目标函数就收敛到 9.402。

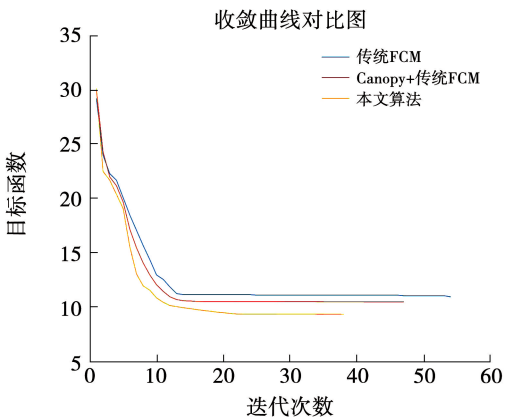


图 6 收敛曲线对比图

Fig. 6 Comparison diagram of convergence curves

3 种算法聚类结果的有效性指标对比如表 7 所示。基于 Canopy + FCM 和本文改进算法的 DBI 指

标相比于传统 FCM 分别小了 7.05% 和 18.75% 左右,表示通过本文的改进,使得分类结果中属于同一类型的故障数据之间更加聚拢,而属于不同故障类型的数据之间更加分散;EI 指标分别提高了 20.24% 和 61.08% 左右,表示综合考虑有效率和聚类效果 DBI 时,改进后的算法在总体上对变压器故障分类的有效性更高。

表 7 聚类指标对比			
Tab. 7 Comparison of clustering index			
方法	传统 FCM	Canopy + FCM	本文改进的 FCM
I_{DBI}	0.277 9	0.258 3	0.225 8
I_{EI}	2.446 9	2.942 3	3.941 5

4 结 论

FCM 是一种典型的模糊分类算法,但在变压器的故障类型识别应用中,由于算法本身的局限性,目标函数易收敛到局部极值点。本文针对其不足做了两点改进,主要得出如下结论:

- 1)利用 Canopy 粗聚类确定了 FCM 的初始聚类中心和最佳聚类数,在一定程度上保证了输入数据的质量,降低了算法陷入局部极值的可靠性,避免了人为确定的主观性。
 - 2)通过引入负指数函数重构了 FCM 算法的隶属度迭代方式,通过引入灵敏系数将隶属度取值平滑地控制在[0,1]之间,拓宽了算法的有效度量范围,也在一定程度上降低了陷入局部极值的可能性。
- 无论是从算法收敛的目标值、耗时、收敛次数还是聚类效果上,本文对传统算法进行的 2 个方面改进都有一定的优化效果,使得分类结果更加符合工程实际。在本文的研究中,虽然描述了重构后隶属度函数中灵敏系数的取值原则:在保证相似度区间精度的同时曲线越平缓越好。本文根据实际经验进行了灵敏度的选择整数值,后续工作可以考虑建立灵敏系数的优化指标,使其取值在数学上最优。

参 考 文 献:

[1] DGA Guide Working Group. IEEE std C57.104 – 2008, IEEE Guide for the Interpretation of Gases Generated in Oil-immersed Transformers[S]. New York, USA: IEEE, 2008.

[2] BAKAR N, ABU-SI Ada A, ISLAM S. A Review of Dis-

solved Gas Analysis Measurement and Interpretation Techniques[J]. IEEE Electrical Insulation Magazine, 2014, 30(3):39.

[3] 王子建. 模糊聚类分析及其在变压器油色谱数据分析中的应用[D]. 武汉: 华中科技大学, 2006.

[4] 魏云冰,王东晖,邵秋华,等. 一种电力变压器油中溶解气体的标定新方法[J]. 电测与仪表, 2016, 53(1):122.

WEI Yunbing, WANG Donghui, SHAO Qihua, et al. A New Calibration Method for Dissolved Gases in Power Transformer Based on Levenberg-Marquardt Algorithm[J]. Electrical Measurement & Instrumentation, 2016, 53(1):122.

[5] 吴建辉,石军,周云峰,等. 变压器油光谱远程在线分析系统[J]. 电测与仪表, 2022, 59(5):166.

WU Jianhui, SHI Jun, ZHOU Yunfeng, et al. On-line Chromatographic Analyzing System for Transformer Oil[J]. Electrical Measurement & Instrumentation, 2022, 59(5):166.

[6] DUVAL M, LAMARRE L. The Duval Pentagona New Complementary Tool for the Interpretation of Dissolved Gas Analysis in Transformers[J]. Electrical Insulation Magazine, IEEE, 2014, 30(6):9.

[7] MANSOUR D E A. Development of a New Graphical Technique for Dissolved Gas Analysis in Power Transformers Based on the Five Combustible Gases[J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2015, 22(5):2507.

[8] SOLEIMANI, MILAD, FAIZ, et al. Dissolved Gas Analysis Evaluation in Electric Power Transformers using Conventional Methods a Review[J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2017, 24(2): 1239.

[9] 李恩文,王力农,宋斌,等. 基于改进模糊聚类算法的变压器油色谱分析[J]. 电工技术学报, 2018, 33(19):4594.

LI Enwen, WANG Linong, SONG Bin, et al. Analysis of Transformer Oil Chromatography Based on Improved Fuzzy Clustering Algorithm[J]. Journal of Electrical Engineering Technology, 2018, 33(19):4594.

[10] 李恩文,王力农,宋斌,等. 基于混沌序列的变压器油色谱数据并行聚类分析[J]. 电工技术学报, 2019, 34(24):5104.

LI Enwen, WANG Linong, SONG Bin, et al. Parallel Clustering Analysis of Dissolved Gas Analysis Data Based on Chaotic Sequences[J]. Journal of Electrical Engineering Technology, 2019, 34(24):5104.

[11] 张贵,丁云飞. 基于加权中智 C 均值算法的变压器故障诊断模型[J]. 水电能源科学, 2020, 38(5):185.

- ZHANG Gui, DING Yunfei. Fault Diagnosis of Transformer Based on Feature-Weighted Neutrosophic C-Mean Algorithm [J]. Hydropower Energy Science, 2020, 38 (5):185.
- [12] 蔡陈. 油浸式变压器 DGA 数据分析及其故障诊断研究[D]. 昆明:昆明理工大学,2016.
- [13] 肖乾望. 基于改进的模糊 C 均值聚类算法的变压器故障诊断研究[D]. 东南大学,2017.
- [14] 许元斌,李国辉,郭昆,等. 基于改进的并行 K-Means 算法的电力负荷聚类研究[J]. 计算机工程与应用, 2017,53(17):260.
- XU Yuanbin, LI Guohui, GUO Kun, et al. Research on Electric Load Clustering Based on Improved Parallel K-Means Algorithm[J]. Computer Engineering and Applications, 2017,53(17):260.
- [15] 邢绍璟. 面向三峡库区水环境监测的数据分析算法研究[D]. 重庆:重庆大学,2016.
- [16] 薛盛炜,李川,李英娜. 改进模糊聚类与主成分分析下的变压器故障识别[J]. 河南科技大学学报(自然科学版), 2020,41(6):39.
- XUE Shengwei, LI Chuan, LI Yingna. Improved Transformer Fault Identification Under Fuzzy Clustering and Principal Component Analysis[J]. Journal of Henan University of Science and Technology (Natural Science Edition), 2020,41(6):39.
- [17] 葛磊蛟,廖文龙,王煜森,等. 数据不足条件下基于改进自动编码器的变压器故障数据增强方法[J]. 电工技术学报,2021,36(S1):84.
- GE Leijiao, LIAO Wenlong, WANG Yusen, et al. A Transformer Fault Data Enhancement Method Based on Improved Autoencoder Under Data Deficiency Conditions [J]. Journal of Electrical Engineering Technology, 2021, 36(S1):84.
- [18] MOHAMMED El Amine Senoussaoui, MOSTEFA Brahmi, ISSOUF Fofana. Combining and Comparing Various Machine-learning Algorithms to Improve Dissolved Gas Analysis Interpretation[J]. IET Generation, Transmission & Distribution, 2018, 12(15):3673.
- [19] 李俭,孙才新,陈伟根,等. 灰色聚类与模糊聚类集成诊断变压器内部故障的方法研究[J]. 中国电机工程学报,2003(2):116.
- LI Jian, SUN Caixin, CHEN Weigen, et al. Study on the Method of Integrating Gray Clustering and Fuzzy Clustering to Diagnose the Internal Faults of Transformers[J]. Chinese Journal of Electrical Engineering, 2003(2):116.
- [20] 张宇婷. 基于 CVA 的变压器故障诊断及预测研究[D]. 长春:长春工业大学,2021.
- [21] 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.
- LIAO Weihang, GUO Innovation, JIN Yu, et al. A Fault Diagnosis Method for Oil-immersed Transformers Based on Four-stage Preprocessing and GBDT [J]. Power Grid Technology, 2019, 43(6):2195.
- [22] 覃煜,黄慧红,方健,等. 基于 FCM 和改进 PCA 的变压器故障诊断[J]. 高压电器, 2018, 54(12):262.
- QIN Yu, HUANG Huihong, FANG Jian, et al. Transformer Fault Diagnosis Based on FCM and Improved PCA[J]. High Voltage Electrical, 2018, 54(12):262.
- [23] 宋清昆,郝敏. 一种改进的模糊 C 均值聚类算法[J]. 哈尔滨理工大学学报, 2007(4):8.
- SONG Qingkun, HAO Min. An Improved Fuzzy C-means Clustering Algorithm [J]. Journal of Harbin University of Science and Technology, 2007(4):8.
- [24] 钱宇骋,甄超,季坤,等. 变压器在线监测数据异常值检测与清洗[J]. 哈尔滨理工大学学报, 2020, 25(5):15.
- QIAN Yucheng, ZHEN Chao, JI Kun, et al. Transformer On-line Monitoring Data Outlier Detection and Cleaning [J]. Journal of Harbin University of Science and Technology, 2020, 25(5):15.
- [25] 宋军英,崔益伟,李欣然,等. 改进分段线性表示与动态时间弯曲相结合的负荷曲线聚类方法[J]. 电力系统自动化, 2021, 45(2):89.
- SONG Junying, CUI Yiwei, LI Xinran, et al. Improved Segmented Linear Representation Combined with Dynamic Time Bending for Load Curve Clustering[J]. Power System Automation, 2021, 45(2):89.
- [26] 贾海峰,刘骥,张明泽,等. 微水含量对老化后变压器油介电性能影响分析[J]. 哈尔滨理工大学学报, 2020, 25(6):70.
- JIA Haifeng, LIU Ji, ZHANG Mingze, et al. Micro Water Content Analysis on the Influence of Transformer Oil Dielectric Properties after Aging[J]. Journal of Harbin University of Science and Technology, 2020, 25(6):70.

(编辑:温泽宇)