

文章编号 2095-0020(2019)04-0198-06

改进加权模糊 C 均值算法的风电变压器故障诊断

张 贵, 丁云飞

(上海电机学院 电气学院, 上海 201306)

摘 要 风电变压器作为风力发电系统中电能传输和转换的枢纽设备之一,其安全稳定运行是电网可靠性的重要保障。为提高变压器故障诊断准确度,提出了基于改进加权模糊 C 均值(FCM)算法的风电变压器故障诊断方法。相比于一般的加权 FCM 算法,特征加权 FCM 算法使用固定的特征权重,该方法通过在训练阶段动态更新特征权重,使得不同比重的特征能够体现样本各维度在聚类效果中所起作用的大小。建立基于改进加权 FCM 算法的风电变压器故障诊断模型,充分考虑了不同比重特征对聚类结果的不同影响,能有效改善复杂数据集的聚类性能。实例研究结果表明:该方法有效地提高了故障诊断的准确率,弥补了传统 FCM 固定权重分配的不足。

关键词 风电变压器; 特征加权; 动态更新; 模糊 C 均值算法; 故障诊断

中图分类号

文献标志码 A

Fault diagnosis of wind power transformer based on improved feature-weighted fuzzy C-means algorithm

ZHANG Gui, DING Yunfei

(School of Electrical Engineering, Shanghai Dianji University, Shanghai 201306, China)

Abstract As the main devices for power transmission and conversion in wind power generation systems, the wind power transformers supply safe and stable running, which is the important guarantee for grid reliability. To improve the accuracy of transformer fault diagnosis, a fault diagnosis method for wind power transformers is proposed based on the improved feature-weighted fuzzy C-means(IWFCM) algorithm. Compared with the general feature-weighted fuzzy C-means(WFCM) algorithm, which uses the fixed feature weights, the present method dynamically updates the feature weights during the training phase. Thus, the characteristics of different specific gravities can reflect the role that each dimension of the sample plays in the clustering effect. The fault diagnosis model for wind power transformers based on the IWFCM fully considers the different effects of different features on the clustering results. It can effectively improve the clustering performance of the complex data sets.

收稿日期: 2018-04-02

基金项目: 国家自然科学基金资助项目(11302123);上海市浦江人才计划资助项目(15PJ1402500)

作者简介: 张 贵(1992—),男,硕士生,主要研究方向为故障诊断与智能控制, E-mail:1394368925@qq.com

指导教师: 丁云飞(1976—),女,教授,博士,主要研究方向为模式识别、智能控制、故障诊断、数据挖掘,

E-mail:dingyf@sdju.edu.cn

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

The case shows that the method can effectively improve the accuracy of fault diagnosis and make up for the deficiency of the traditional FCM fixed weight distribution.

Key words wind power transformers; feature weighting; dynamic update; fuzzy C-means (FCM) algorithm; fault diagnosis

风能是我国大力发展的清洁能源,近几年,我国风力发电规模不断扩大,风力发电技术也愈发成熟,并且有着广阔的发展前景^[1]。当风力不稳定的时候,风电系统的输出电压也会不稳定,这会导致风电变压器出现各种故障,甚至有可能发生重大事故。因此,在故障初期精准地发现变压器的故障并对典型故障进行分析,是保障风电变压器安全可靠运行的重要举措^[2]。

模糊C均值(Fuzzy C-Means, FCM)算法是在K-means的基础上加入了模糊的思想,赋予每一个样本数据隶属度,从而实现模糊划分,聚类性能得以加强。FCM算法因其良好的效果在分类和故障诊断等方面得到大量应用。自从1974年Dunn等^[3-4]提出FCM算法后,该算法被科研工作者研究改进,聚类性能和效果进一步提高。如刘长良等^[5]采用变分模态分解方法提取特征向量,通过FCM聚类方法对滚动轴承故障进行故障识别,能快速得到聚类中心,且诊断精度为100%。唐燕雯^[6]将模糊核聚类算法应用于船舶故障诊断,得到了收敛速度快、鲁棒性强的故障诊断模型。代宪亚等^[7]针对电厂的复杂性,将FCM算法和神经网络结合在一起,从确定隶属度到处理参数,再到神经网络诊断,最后将结果融合,从而得出高准确度的诊断结果。李赢等^[8]提出基于FCM和改进相关向量机的模型,该模型采用完全二叉树改进相关向量机,对故障诊断精度有明显提升,也为实时在线检测发现故障并及时做出诊断提供了实用价值。王军辉等^[9]利用集合经验模式分解得到特征矩阵,特征矩阵处理后进行模糊聚类分析,并将其应用到齿轮箱故障诊断中,实验表明,该方法更能有效进行故障分类。

由于FCM算法没有考虑样本向量中每个维度特征的差异性情况,特征加权FCM(Feature-Weighted C-Means, WFCM)算法被提出。文献^[10]

将特征权重、核函数与模糊聚类方法结合,将核函数从低维映射到高维,并对样本各维度加权,对于维度较高的样本有较好的处理效果,该方法在汽轮发电机组故障诊断中得到了有效的验证。文献^[11]改进了人工鱼群优化,并与WFCM算法结合,能够使算法收敛更快并解决FCM受初始值影响较大的问题。文献^[12]对模糊C有序均值算法通过极差正规化的手段特征加权,既保证了算法的健壮性,又能优化样本数据的分布。

本文在WFCM算法的启发下,提出基于改进加权FCM(Improved Feature-Weighted Fuzzy C-Means, IWFCM)算法的故障诊断方法,并将其应用到变压器故障诊断中。该方法考虑到聚类过程中聚类中心的变化会带来样本分布的变化,因此,权重也应动态更新来适应样本的重构。通过采用动态更新特征权重的方法,使得样本的分布更趋于合理,诊断准确度得以提高,从而为变压器的故障诊断技术提供一种新的研究方法。

1 基本算法理论

1.1 FCM算法

FCM算法通过对每一个样本数据赋予隶属度,使得聚类过程中聚到同一类的样本有最大的相似度,从而实现了样本数据的模糊划分。FCM算法以较为简单的结构和较强的聚类性能被应用于众多领域,是最受科研工作者们喜爱的算法之一。

有数据集 $D = \{x_i\} \in R^d$,其中 $x_i \in R^d$, d 为特征维数, C 为分类数。聚类目标函数表示为

$$J(U, C; D) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2 \quad (1)$$

式中: x_i 为样本数据; c_j 为聚类中心; m 为模糊度;

u_{ij} 为隶属度。在特征条件 $\sum_{j=1}^C u_{ij} = 1$ 下,求解式(1)最小值,得到 u_{ij} 和 c_j 的公式为

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

$$c_j = \left(\sum_{i=1}^N u_{ij}^m x_i \right) / \left(\sum_{i=1}^N u_{ij}^m \right) \quad (3)$$

1.2 IWFCM 算法

本文所提到的 IWFCM 算法,通过动态更新特征权重,使得不同比重的权重值能够体现各维度特征在聚类过程中所起作用的大小,这对于特征的细分优化有非常大的提升。IWFCM 算法的目标函数为

$$J(U, C, W; D) = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m [d_{ij}^{(w)}]^2 \quad (4)$$

式中

$$d_{ij}^{(w)} = \|\text{diag}(w)(x_i - c_j)\|$$

$$\text{diag}(w) = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_d \end{bmatrix}$$

在两个特征条件 $\sum_{j=1}^c u_{ij} = 1$ 和 $\sum_{q=1}^d w_q = 1$ 下,求解式(4)最小值,分别得到 u_{ij} 、 c_j 和 w_q 的公式为

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{[d_{ij}^{(w)}]^2}{[d_{ik}^{(w)}]^2} \right)^{\frac{1}{m-1}}} \quad (5)$$

$$c_j = \left(\sum_{i=1}^N u_{ij}^m x_i \right) / \left(\sum_{i=1}^N u_{ij}^m \right) \quad (6)$$

$$w_q = \sum_{l=1}^d \left[\frac{\sum_{i=1}^N \sum_{j=1}^c u_{ij}^m (x_{iq} - c_{jq})^2}{\sum_{i=1}^N \sum_{j=1}^c u_{ij}^m (x_{il} - c_{jl})^2} \right]^{-1} \quad (7)$$

IWFCM 算法是在经典的 FCM 算法上动态更新特征权重,使得权重得以重新分配,某些布局不均衡的样本分布也变得更合理,样本精度得以提高。IWFCM 算法的具体步骤如下:

步骤 1 采集风电变压器故障样本,对样本做预处理。

步骤 2 设置参数,包括分类数 C ,模糊度 m ,隶属度 u_{ij} 。

步骤 3 初始化隶属度 u_{ij} 和特征权重 w_d 。

步骤 4 按照式(5)~式(7)更新隶属度、聚类中心和特征权重。

步骤 5 判断迭代终止条件是否满足,若不满足,则转到步骤 4;若满足,则停止迭代过程。

步骤 6 依据特征权重值的大小来判断各维度在聚类过程中所起的主次作用,通过隶属度的大小来判断故障类型。

2 风电变压器故障诊断模型

2.1 选取特征气体和预处理数据

风电变压器在工作过程中处于故障状态时,相比于正常工作状态,产生的 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 等含量会大幅增加。而且对于不同类型的故障,原始样本中气体含量会有较大的数量级差异。因此,以 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 这 5 种气体的体积分数作为特征量,为了减少计算误差,需要经过归一化处理后再加载到故障诊断模型。归一化公式为(处理后的样本数据值介于 0 到 1 之间)

$$x_{ij} = x_{ij} / \left(\sum_{k=1}^5 x_{ik} \right) \quad (8)$$

式中: x_{ij} 为原始数据; x_{ij} 为已经处理后的数据; i 为样本序号; j 为属性序号,用 1~5 依次表示 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 5 种气体。

2.2 风电变压器故障诊断的具体实现

基于 IWFCM 算法的模型充分考虑了数据各维度之间权重的差异,动态更新特征权重可以在训练阶段实时体现聚类信息的变化。基于 IWFCM 算法的风电变压器故障诊断的具体实现过程如下:

(1) 采集样本数据。采集风电变压器 5 种油中溶解气体的气体含量数据,构成原始数据集。

(2) 数据预处理。利用式(8)对原始数据做归一化处理,并将处理后的样本集作为输入样本。

(3) 建立故障诊断模型。确定故障样本的聚类中心和隶属度,利用式(7)动态更新数据集特征权重,使得不同的权重能够对样本的分布有更合理的重构,建立起动态加权的故障诊断模型。

(4) 故障结果分类。将样本数据输入诊断模型,聚类后通过隶属度矩阵的大小得出诊断模型聚类后的故障类别,通过特征权重矩阵的大小找出对聚类结果影响较大的几维特征,从而得出 5 种气体中哪几种气体的含量对聚类的效果影响较大。

故障诊断流程如图1所示。

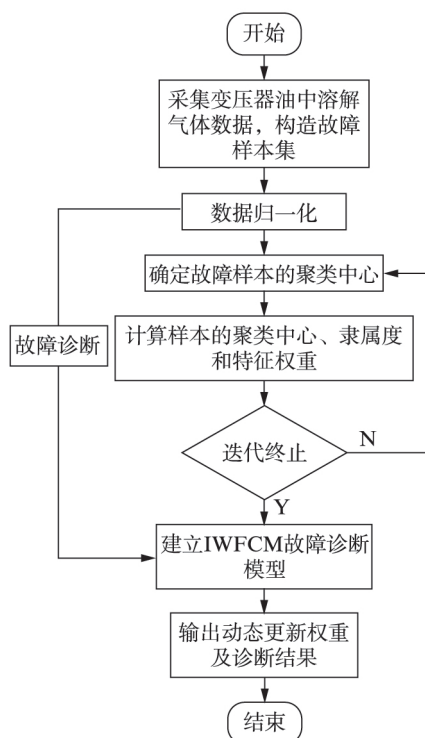


图1 故障诊断流程图

3 实验与结果分析

3.1 聚类样本的获取与选择

本文反映变压器故障类型的油中溶解气体数据来源于风电公司,数据共有246组。在本文中,将风电变压器的故障类型分为6种,具体样本个数分布如表1所示。

表1 故障样本集组成

故障类型	样本个数
低温过热	25
中温过热	42
高温过热	70
局部放电	36
低能放电	20
高能放电	53
合计	246

3.2 故障诊断效果及对比

为验证所提方法的可行性,分别利用 K-means

方法、FCM 和 IWFCM 诊断方法来识别故障并比较实验结果。设置以下参数: $m=2, C=6$ 。

聚类效果如图2~4所示。图中,横向是实际6种聚类结果,纵向是聚类后的情况。混淆矩阵对角线上的数字代表聚类结果与实际情况相符的聚类个数。

低温过热	16.00	8.00	1.00	0.00	0.00	0.00
中温过热	1.00	12.00	28.00	0.00	1.00	0.00
高温过热	1.00	6.00	63.00	0.00	0.00	0.00
局部放电	0.00	1.00	0.00	33.00	0.00	2.00
低能放电	0.00	0.00	0.00	2.00	6.00	12.00
高能放电	0.00	0.00	4.00	2.00	0.00	47.00
	低温过热	中温过热	高温过热	局部放电	低能放电	高能放电

图2 K-means 聚类结果与实际情况的混淆矩阵

低温过热	17.00	6.00	0.00	0.00	0.00	2.00
中温过热	1.00	33.00	7.00	0.00	0.00	1.00
高温过热	1.00	25.00	44.00	0.00	0.00	0.00
局部放电	0.00	0.00	0.00	32.00	2.00	2.00
低能放电	0.00	0.00	0.00	2.00	11.00	7.00
高能放电	0.00	0.00	1.00	2.00	29.00	21.00
	低温过热	中温过热	高温过热	局部放电	低能放电	高能放电

图3 FCM 聚类结果与实际情况的混淆矩阵

低温过热	16.00	5.00	4.00	0.00	0.00	0.00
中温过热	1.00	40.00	7.00	0.00	1.00	0.00
高温过热	1.00	1.00	68.00	0.00	0.00	0.00
局部放电	0.00	3.00	1.00	32.00	0.00	0.00
低能放电	0.00	0.00	0.00	2.00	16.00	2.00
高能放电	0.00	0.00	2.00	6.00	2.00	43.00
	低温过热	中温过热	高温过热	局部放电	低能放电	高能放电

图4 IWFCM 聚类结果与实际情况的混淆矩阵

图2中聚类后的低温过热中有16个样本与实际情况相符,8个被分到了中温过热,1个被分到了高温过热。IWFCM诊断方法在中、高温过热和低、高能放电上,相较于K-means和FCM诊断方法,故障识别精度分别有了大幅度的提升。其中,K-means、FCM、IWFCM诊断方法在中温过热上准确率分别为28.57%、78.57%、95.24%;在高温

过热上准确率分别为 90%、62.86%、97.14%；在低能放电上准确率分别为 30%、55%、80%；在高能放电上准确率分别为 88.68%、39.62%、81.13%。在低温过热和局部放电上,3 种诊断方法的准确率无明显变化。虽然在高能放电上 K-means 诊断方法的精度最高,但其比 IWFCM 方法高的并不多,而且从整体上看,IWFCM 诊断方法的精度是最高的,为 87.40%。

K-means 和 FCM 中各维度的特征权重都是相同的,对于维度较高的数据集,K-means 和 FCM 在聚类结果上会受到不重要维度和噪声特征的干扰。因此,在 FCM 中加入了动态更新的特征权重,使得贡献权重的分配更趋于合理,得到的故障识别率更加精确。

3.3 故障诊断结果分析

相较于 K-means 和 FCM,IWFCM 由于加入了动态更新的特征权重,使得样本数据集的每一维特征都有着不同比重的效果,且该特征权重在训练过程中是实时更新的,因此,对于高维度的数据而言,可以突出某些维度特征的主要作用,降低冗余特征的干扰。在 IWFCM 聚类所得到的结果中,由权重所得到的特征权重直方图如图 5 所示。由图 5 可知,第 3、第 5 维在聚类过程中所占的比重较大,即 C_2H_6 和 C_2H_2 是主导特征。这与实际情况是相符的,因为当变压器绝缘油分解产气只有热源输出时,主要产生低分子烃类气体,其中 CH_4 和 C_2H_6 的含量占总烃的 80% 以上,而当火花放电时,产生的主要气体是 C_2H_2 。由于动态更新特征权重能够优化样本的分布,因此,可以得到最优的聚类中心。最优聚类中心如表 2 所示。

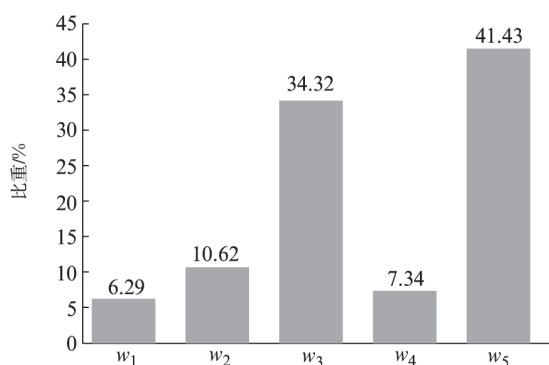


图 5 5 种气体的特征权重直方图

表 2 最优聚类中心

类别	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
低温过热	0.108 3	0.315 7	0.182 2	0.390 1	0.003 7
中温过热	0.458 3	0.134 3	0.022 6	0.185 9	0.199 0
高温过热	0.870 2	0.065 4	0.028 3	0.028 0	0.008 1
局部放电	0.102 0	0.271 0	0.536 3	0.087 7	0.003 0
低能放电	0.131 6	0.276 7	0.083 4	0.500 2	0.008 2
高能放电	0.409 1	0.094 7	0.042 8	0.092 8	0.360 6

4 结 论

风电变压器发生不同故障时,其内部产生的气体含量是不同的,各气体在聚类过程中所起的作用也是不同的。针对一般的聚类方法未考虑到数据各维度之间权重的差异,本文提出了 IWFCM 算法的风电变压器故障诊断方法,该方法在传统的 FCM 算法中通过动态更新特征权重,使得样本数据集的每一维都有着主次划分,突出了某些特征对聚类的主要影响,降低了冗余特征的干扰。结果表明:此方法可以优化数据的分布,极大地提高聚类准确率,特征权重矩阵能够反映出聚类过程中数据集各维度所起作用的不同,对于变压器的故障诊断有着重要的参考价值。

参 考 文 献

- [1] 李军军,吴政球,谭勋琼,等.风力发电及其技术发展综述[J].电力建设,2011,32(8):64-72.
- [2] 刘军,张安红,李基成.风电变压器可靠性、安全性及环境问题综述(1)[J].变压器,2013,50(10):48-50.
- [3] DUNN J C. Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems [J]. Journal of Cybernetics,1974,4:1-15.
- [4] JAMES C, BEZDEK J. Cluster validity with fuzzy sets[J].Cybernetics and Systems,1974,3(3):58-74.
- [5] 刘长良,武英杰,甄成刚.基于变分模态分解和模糊 C 均值聚类的滚动轴承故障诊断[J].中国电机工程学报,2015,35(13):3358-3365.
- [6] 唐燕雯.模糊核聚类在船舶故障诊断中的应用[J].舰船科学,2016,38(24):151-153.

- [7] 代宪亚,茅大钧.基于模糊理论与数据融合的电厂故障诊断技术[J].上海电机学院学报,2016,19(4):201-206.
- [8] 李赢,舒乃秋.基于模糊聚类 and 完全二叉树支持向量的变压器故障诊断[J].电工技术学报,2016,31(4):64-70.
- [9] 王军辉,贾嵘,谭泊.基于 EEMD 和模糊 C 均值聚类的风电机组齿轮箱故障诊断[J].太阳能学报,2015,36(2):319-324.
- [10] 李超顺,周建中,安学利,等.基于加权模糊核聚类的发电机组振动故障诊断[J].中国电机工程学报,2008,28(35):79-83.
- [11] 史丽萍,宋朝鹏,李明泽,等.基于 SAAFS 优化加权模糊聚类算法的变压器故障诊断[J].电测与仪表,2018,55(11):12-18.
- [12] 刘永利,王恒达,刘静,等.特征加权的模糊 C 有序均值聚类算法[J].河南理工大学学报(自然科学版),2019(3):123-13
- [13] XING H J, HA M H. Further improvements in feature-weighted fuzzy c-means [J]. Information Sciences,2014,267:1-15.
- [14] 吉书瑶,吕红芳.无线传感器节点多特征组合加权 K-means 聚类算法[J].上海电机学院学报,2017,20(4):226-231.
- [15] ZHANG J. Fault diagnosis model based on fuzzy support vector machine combined with weighted fuzzy clustering[J].Transactions of Tianjin University,2013,19(3):174-181.
- [16] 张秦棣,宋辉,姜勇,等.基于 OS-ELM 的变压器局部放电模式识别[J].高电压技术,2018,44(4):1122-1130.
- [17] 陈欢,彭辉,舒乃秋,等.基于蝙蝠算法优化最小二乘双支持向量的变压器故障诊断[J].高电压技术,2018,44(11):3664-3671.