

DOI: 10.7667/PSPC180147

## 基于聚类和时间序列分析的变压器状态评价方法

辛建波<sup>1</sup>, 康琛<sup>1</sup>, 翁新林<sup>2</sup>, 陈田<sup>1</sup>, 谢斌<sup>3</sup>, 郭创新<sup>4</sup>

(1. 国网江西省电力有限公司电力科学研究院, 江西 南昌 330000; 2. 国网江西省电力有限公司检修分公司, 江西 南昌 330000; 3. 国网江西省电力有限公司, 江西 南昌 330000; 4. 浙江大学电气工程学院, 浙江 杭州 310027)

**摘要:** 传统的电力变压器 DGA 故障诊断方法, 仅能二值化地判断设备处于健康或故障状态, 无法表征变压器的潜在故障情况, 也无法确定变压器向故障状态转化的趋势。对此, 提出了一种基于聚类和时间序列分析的变压器状态评价方法。首先, 基于点密度判据进行数据预处理, 消除噪声影响。其次, 基于大数据聚类思想, 计算采样数据和历史故障数据簇的相对邻近度, 根据计算结果将设备状态划分为健康、潜伏故障或故障。在此基础上判断故障设备的故障类型, 基于故障类型关联权重计算健康设备的健康得分, 通过时间序列相似性分析方法获取潜伏故障设备的预测故障发展时间。算例分析验证了该方法的可行性与有效性。

**关键词:** DGA; 变压器; 聚类; 大数据; 时间序列分析

### Evaluation method of transformer state based on clustering and time series analysis

XIN Jianbo<sup>1</sup>, KANG Chen<sup>1</sup>, WENG Xinlin<sup>2</sup>, CHEN Tian<sup>1</sup>, XIE Bin<sup>3</sup>, GUO Chuangxin<sup>4</sup>

(1. Electric Power Research Institute, State Grid Jiangxi Electric Power Co., Ltd, Nanchang 330000, China;  
2. Maintenance Branch, State Grid Jiangxi Electric Power Co., Ltd, Nanchang 330000, China;  
3. State Grid Jiangxi Electric Power Co., Ltd, Nanchang 330000, China;  
4. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

**Abstract:** Conventional DGA fault diagnosis methods of power transformer can only judge whether the equipment is in normal or fault condition, but they can neither characterize the potential failure of the normal transformer nor identify the trend of transformer converted into fault state. In order to solve this problem, this paper proposes a transformer state evaluation method based on clustering and time series analysis method. Firstly, the data preprocessing method is carried out to prevent the influence of noise. Then, based on big data clustering method, the proximity of sampled data and historical fault data clusters is calculated. Based on the result, the equipment condition is sorted into healthy, incipient faulty or faulty. For equipment in faulty state, the fault type is identified. For healthy equipment, the health score is calculated based on related weights of fault types. For incipient faulty equipment, the time series similarity analysis method is used to further predict the time span for the equipment changing from current state to fault state. The example analysis verifies the feasibility and effectiveness of the proposed method.

This work is supported by Key Program of National Natural Science Foundation of China (No. 51537010) and Science and Technology Project of State Grid Corporation of China (No. 52182016001J).

**Key words:** DGA; transformers; clustering; big data; time series analysis

## 0 引言

在输变电系统中, 电力变压器是电能转换的核

心<sup>[1]</sup>, 其安全可靠地运行决定了电力系统能否正常运作。因此, 有必要建立合适的模型评估油浸式变压器运行过程当中健康状况, 帮助操作人员及时发现存在较高故障风险的设备。

油浸式变压器绝缘氧化、裂解或碳化过程中会产生多种可溶于变压器油的气体, 对于气体组分、含量进行分析的油中溶解气体分析(Dissolved Gas Analysis, DGA)方法是最为广泛使用的变压器状态

**基金项目:** 国家自然科学基金重点项目资助(51537010)“多重不确定因素下的智能电网风险调度理论与方法研究”; 国网科技项目资助(52182016001J)“输变电设备健康诊断与故障预警云服务平台研究与应用项目”

检测和故障诊断方法之一。通过离线试验或在线监测装置<sup>[2]</sup>, 能够获取变压器油中的气体含量。IEC/IEEE 标准提出的三比值法, 根据气体体积分数的比值进行故障诊断。三比值法简单易用, 但其状态编码不连续<sup>[3]</sup>, 且故障诊断判据基于现场经验基础之上<sup>[4]</sup>, 在实际应用当中存在局限性。文献[5]利用模糊理论和 TOPSIS 方法增强了评价的客观性。文献[6]结合改良三比值法、大卫三角形法和立方体图示法三种手段, 进行综合性变压器故障诊断。文献[7]基于绝对阈值统计了变压器油中含气量超标的案例, 并对原因与对策进行探究。

近年来, 不少学者将各种智能算法应用于变压器故障诊断领域中, 取得了较好的效果。有研究采用 BP 神经网络进行变压器故障诊断, 提升了诊断的准确性<sup>[8]</sup>。为解决 BP 神经网络存在不易收敛的问题, 有研究将粒子群算法<sup>[9]</sup>、遗传算法<sup>[10]</sup>与神经网络结合, 加快了收敛速度。文献[11]采用深度信念网络取代传统 BP 神经网络结构, 提升了神经网络的判别性能。支持向量机方法具有较好的泛化能力, 能够克服传统智能算法非线性、局部极小等问题, 也被广泛应用于变压器故障诊断当中<sup>[12-14]</sup>。聚类是一种无监督学习的大数据处理方法, 无需先验知识即可将故障进行分类, 相较于其他智能算法更加简单, 并同样拥有较高的诊断准确性。文献[15]提出了模糊 C 均值聚类的方法, 但其存在容易陷入局部最小、受样本分布影响大的特点。文献[16]提出使用核可能性聚类算法对上述问题进行改进, 加快了算法的收敛速度。文献[17]在此基础上用点密度取代可能性函数, 消除了超球面的局限, 进一步提升了诊断准确性。

但上述研究中提出的方法存在两个局限: (1) 未进行数据预处理, 采样噪声对诊断结果有较大影响; (2) 仅能对当前设备所属的状态进行健康或故障的二值化判断, 无法反映出变压器的潜在故障或缺陷, 也无法表征设备向故障状态转换的速度与概率, 难以为设备的运行维护检修提供指导。

基于上述问题, 本文提出了一种变压器健康状态评价方法。首先基于点密度判据对采集的 DGA 数据进行预处理, 剔除噪声点并补全数据; 然后基于大数据聚类思想, 计算待测数据点和不同故障类别下的历史故障数据簇之间的相对邻近度, 将设备状态划分为健康、潜伏故障和故障三类, 判别处于潜伏故障及故障状态设备的故障类型。进一步地, 通过时间序列相似性分析预测处于潜伏故障状态的设备转化为故障状态所需要的时间, 同时计算处于健康状态设备的健康得分, 可应用于变电站的集成

保护系统<sup>[18]</sup>, 并为状态检修提供指导。

## 1 基于点密度判据的数据预处理

为避免噪声数据对状态评价的影响, 首先需要对采样数据序列进行预处理, 剔除噪声点并补全数据。设定计算周期为  $T_c$ , 数据采样周期为  $T_s$ , 可获得数据预处理窗口长度  $n$ :

$$n = \left\lceil \frac{T_c}{T_s} \right\rceil \quad (1)$$

由于计算周期  $T_c$  较短, 窗口长度  $n$  较小。因此在一个窗口内的数据点比较集中, 正常数据点的周围会有较多邻近数据点, 即这一点邻近区域具有较高的密度。异常点一般由采样噪声引起, 分布较为分散, 即处于密度较低的区域。通过计算数据点的密度能够反映其离群度, 从而为异常点的检测提供依据。数据点的密度可以理解为数据点与周围点的邻近度, 而邻近度可以定义为到  $k$  个最接近数据点(最近邻点)平均距离的倒数。数据点  $x$  和它最近邻的  $k$  个点的邻近度可由式(2)表示。

$$l(x, k) = \left( \frac{\sum_{y \in N(x, k)} d(x, y)}{k} \right)^{-1} \quad (2)$$

式中:  $N(x, k)$  是包含数据点  $x$  的  $k$  个最近邻点的集合;  $y$  是一个最近邻点;  $d(x, y)$  是  $x$  和  $y$  之间的距离。数据点之间的距离计算方法较多, 常用的有欧式距离<sup>[19]</sup>、余弦距离、曼哈顿距离等。其中, 欧式距离能够直观反映高维度数据之间的直线距离, 因此本文采用欧式距离进行计算。第  $l$  个数据预处理窗口  $N_l$  当中两个数据点  $x_l$  和  $y_l$  之间的距离  $d(x_l, y_l)$  可用式(3)表示。

$$d(x_l, y_l) = \sqrt{\sum_{i=1}^m \lambda_i (x_{li} - y_{li})^2} \quad (3)$$

式中:  $m$  为数据维数;  $x_{li}$  代表数据点  $x_l$  在维度  $i$  上的坐标;  $y_{li}$  代表数据点  $y_l$  在维度  $i$  上的坐标。为解决数据不同维度上距离度量标准不同的问题, 引入归一系数  $\lambda$ 。归一系数  $\lambda$  需要能够反映数据在此维度下的距离尺度。首先定义在第  $l$  个数据预处理窗口  $N_l$  中所有正常数据点在维度  $i$  上的总距离为  $D_{li}$ 。

$$D_{li} = \sum_{x_{li} \in N_l, y_{li} \in N_l} |x_{li} - y_{li}| \quad (4)$$

第  $l$  个数据预处理窗口中维度  $i$  上的归一系数  $\lambda_{li}$ , 通过上一个数据预处理窗口中正常数据点在维度  $i$  上的总距离占有所有维度距离之和的比值获得。

$$\lambda_{li} = \frac{D_{(l-1)i}}{\sum_{j=1}^n D_{(l-1)j}} \quad (5)$$

对于数据预处理窗口当中的数据点  $x$ ，求取  $k$  点邻近度  $l(x, k)$ 。当  $l(x, k)$  小于阈值  $l_{th}$ ，标记此数据点为噪声数据，否则标记此数据点为正常数据。当窗口中所有数据点都标记完成后，对噪声数据点进行替换。若噪声数据点相邻的前后两个数据点均为正常数据，则将噪声数据替换为前后两个数据点的平均值。否则，将噪声数据点替换为与其采集时间最接近的正常数据点。

## 2 基于大数据聚类思想的设备状态分析

根据变压器常见故障类型，将故障类别划分为：低温过热(F1)、中温过热(F2)、高温过热(F3)、局部放电(F4)、低能放电(F5)、高能放电(F6)。收集设备故障历史数据样本。将包含每一种设备故障类型下的 DGA 数据集合。每个样本点都含有五个维度的参数，即 H<sub>2</sub>、CH<sub>4</sub>、C<sub>2</sub>H<sub>6</sub>、C<sub>2</sub>H<sub>4</sub>、C<sub>2</sub>H<sub>2</sub> 的气体体积浓度。将设备属于某一故障类型的所有历史数据点称为一簇，按照故障类型划分，可将属于 F<sub>1</sub>—F<sub>6</sub> 的历史数据点划分为 C<sub>1</sub>—C<sub>6</sub> 共计六个故障数据簇。

记故障类型编号为  $j$  的簇为  $C_j$ 。采用 DBSCAN 聚类算法的思想，通过计算数据点  $x$  和  $k$  个最近邻的数据点的距离平均值倒数，可以得到反映数据点  $x$  与数据簇  $C_j$  的邻近度  $l_j(x, k)$ 。

$$l_j(x, k) = \left( \frac{\sum_{y \in C_j(x, k)} d(x, y)}{k} \right)^{-1} \quad (6)$$

式中： $C_j(x, k)$  是数据点  $x$  在簇  $C_j$  当中  $k$  个最近邻点的集合； $y$  是簇中的一个  $x$  的最近邻点； $d(x, y)$  是  $x$  和  $y$  之间的欧式距离，可由式(7)计算。

$$d(x, y) = \sqrt{\sum_{i=1}^m \lambda_i (x_i - y_i)^2} \quad (7)$$

式中： $m$  为数据维数； $x_i$  代表数据点  $x$  在维度  $i$  上的坐标； $y_i$  代表数据点  $y$  在维度  $i$  上的坐标； $\lambda_i$  为维度  $i$  上的归一系数，在数据预处理阶段获得。

将  $C_j$  内部数据点作为待测数据点，代入式(6)中进行计算，得到数据点相对于本簇的邻近度，称为簇内邻近度。对所有内部数据点的簇内邻近度取平均值，得到  $C_j$  的基准邻近度  $l_{jref}$ 。

对于待计算数据点  $x$  而言，它与簇的邻近度越大，则表明它越接近于簇，即更加可能发生此类故障。数据点  $x$  和簇  $C_j$  的邻近程度由相对邻近度  $l_{cj}(x, k)$  表示。

$$l_{cj}(x, k) = \frac{l_j(x, k)}{l_{jref}} \quad (8)$$

## 3 基于时间序列相似性分析的趋势预测

设备发生故障前油气信息的变化具有时序特征，对于具有潜伏故障风险的设备，可以基于时间序列相似性分析判断设备从当前状态向故障状态转化的速度与趋势。

### 3.1 确定待测数据时间序列

当待评价设备的数据点  $x$  下和某一故障数据簇  $C_j$  间相对邻近度处于区间[0.6, 0.8]时，截取数据点  $x$  在采集时刻前的多个数据点，作为待测数据时间序列  $X$ ，其步骤如下。

- 1) 设定序列  $X$  的长度  $N$ 。
- 2) 确定末数据点时标  $t_{xN}$ ：将数据点  $x$  的采集时刻设定为  $t_{xN}$ 。
- 3) 计算相对邻近度：设定序列  $X$  的最大搜寻时间  $t_{x\max}$ 。根据式(8)，计算  $t \in [t_{xN} - t_{x\max}, t_{x\max}]$  范围内的数据点  $x(t)$  和故障数据簇  $C_j$  的相对邻近度  $l_{cj}(x(t), k)$ 。
- 4) 确定初数据点时标  $t_{x1}$ ：从  $t_{xN}$  开始向前搜索第一个相对邻近度  $l_{cj}(x(t), k) = 0.5 \pm \varepsilon_x$  的数据点  $x(t)$ ， $\varepsilon_x$  为搜索允许误差。如果有符合要求的数据点，将对应数据采集时刻设定为  $t_{x1}$ 。若无符合要求的数据点，将  $t_{xN} - t_{x\max}$  设定为  $t_{x1}$ 。
- 5) 获取时间序列：以  $d_{tx} = \Delta t_x / (N - 2)$  为抽样间隔，从  $t_{x1}$  开始按照时间次序抽取  $N$  个数据点，得到最终的时间序列  $X = \{x(t_{x1}), x(t_{x2}), \dots, x(t_{xN})\}$ 。

### 3.2 确定故障历史数据时间序列

以故障簇  $C_j$  中某一和数据点  $x$  邻近的数据点  $y$  为例，选取数据点  $y$  采集前的时间序列  $Y$  作为对比序列。其流程和获取待测数据时间序列  $X$  基本一致，仅确定初末数据点时标方式不同，区别如下。

- 1) 确定初数据点时标  $t_{y1}$ ：记故障数据点  $y$  的采集时刻为  $t_{yb}$ ，数据点  $y$  也记作  $y(t_{yb})$ 。设置序列  $Y$  的最大搜寻时间  $t_{y\max}$ 。计算  $t \in [t_{yb} - t_{y\max}, t_{yb}]$  范围内的数据  $y(t)$  和故障数据簇  $C_j$  的相对邻近度  $l_{cj}(y(t), k)$  从  $t_{yb}$  时刻向前搜索满足  $l_{cj}(y(t), k) = l_{cj}(x(t_{x1}), k) \pm \varepsilon_{y1}$  的数据点， $\varepsilon_{y1}$  为初次搜索允许误差。若在  $t_{yb} - t_{y\max}$  时刻前找到满足要求的数据点，将对应的采集时间设定为  $t_{y1}$ 。否则，结束所有步骤，视为寻找失败。

- 2) 确定末数据点时标  $t_{yN}$ ：在  $t_{y1}$  到  $t_{yb}$  区间内进行第二次搜寻，从  $t_{y0}$  开始向后寻找第一个数据点，使得  $l_{cj}(y(t), k) = l_{cj}(x(t_{xN}), k) \pm \varepsilon_{y2}$ ， $\varepsilon_{y2}$  为二次搜索允许误差。若在  $t_{yb}$  时刻前找到符合要求的数据点，将对应的采集时间设定为  $t_{yN}$ 。否则，结束所有步骤，视为寻找失败。

### 3.3 时间序列相似性计算

两个时间序列之间的距离可以反映其相似性特

征, 距离越小, 两个时间序列的相似性就越高。常用的时间序列距离包括欧式距离<sup>[19]</sup>、编辑距离<sup>[21]</sup>、动态时间弯曲距离<sup>[22]</sup>等。其中, 编辑距离用于计算两字符串序列间转化所需要的编辑步数, 欧式距离体现两个时间序列当中下标对应的数据点间距离之和, 动态时间弯曲距离允许不同步点对应计算, 其本质是通过寻找两时间序列间的最小路径来计算它们间的距离。由于本文中的时间序列数据元素为电力变压器的油气信息, 且需要防止部分数据错位造成距离评估结果偏大的现象发生, 所以采用动态时间弯曲距离来衡量两时间序列的相似性。时间序列  $X = \{x(t_{x1}), x(t_{x2}), \dots, x(t_{xN})\}$  和时间序列  $Y = \{y(t_{y1}), y(t_{y2}), \dots, y(t_{yN})\}$  之间的动态时间弯曲距离可以通过递归计算获得<sup>[20]</sup>。

$$D_{\text{dtw}}(X, Y) = \begin{cases} 0 & X = \{\emptyset\} \text{ 且 } Y = \{\emptyset\} \\ \infty & X, Y \text{ 有且只有一个为 } \{\emptyset\} \\ d(x(t_{x1}), y(t_{y1})) + \min(D_{\text{dtw}}(X, R(Y)), \\ D_{\text{dtw}}(R(X), Y), D_{\text{dtw}}(R(X), R(Y))) & X \neq \{\emptyset\} \text{ 且 } Y \neq \{\emptyset\} \end{cases} \quad (9)$$

式中:  $d(x(t_{x1}), y(t_{y1})) = \|x(t_{x1}) - y(t_{y1})\|$ ;  $R(X) = \{x(t_{x2}), x(t_{x3}), \dots, x(t_{xN})\}$ ;  $R(Y) = \{y(t_{y2}), y(t_{y3}), \dots, y(t_{yN})\}$ 。

为方便衡量不同长度序列之间的距离, 将动态时间弯曲距离计算结果归一化, 得到最终用于评价序列相似性的归一距离为

$$D(X, Y) = D_{\text{dtw}}(X, Y) / N \quad (10)$$

式中,  $N$  为序列  $X$  和序列  $Y$  的长度。

## 4 变压器状态综合评价

### 4.1 状态评价方法流程

变压器状态综合评价方法分为三个阶段, 其流程如图 1 所示。

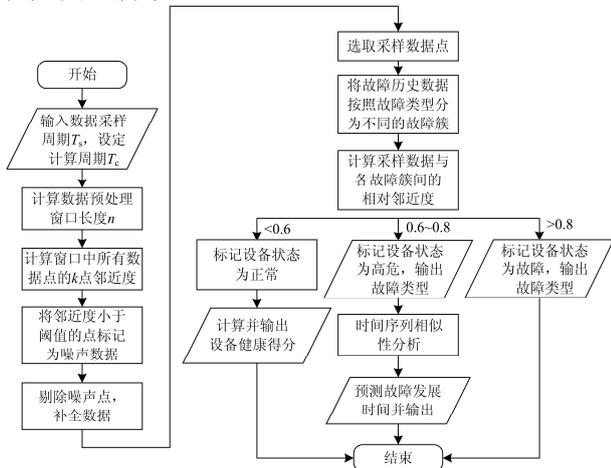


图 1 方法流程图

Fig. 1 Flow chart of method

1) 首先基于点密度判据进行数据预处理, 标记并剔除噪声点, 补全数据。

2) 其次选取采样数据点  $x$ , 计算其与 6 个历史故障数据簇  $C_1$ — $C_6$  之间的相对邻近度。

3) 最后根据相对邻近度计算结果, 对设备状态进行标记并分类处理, 分为三个部分: ① 数据点与某一故障数据簇相对邻近度大于 0.8, 表征设备处于很高的故障风险, 将设备标记为故障, 设备故障类别为此邻近的故障数据簇的故障类别; ② 而相对邻近度处于区间[0.6, 0.8]时, 表明设备处于较高的风险, 将设备状态标记为潜伏故障, 通过时间序列分析方法进一步判断设备向故障状态转化的速度和趋势; ③ 当数据点和所有数据簇的相对邻近度都小于 0.6 时, 将设备状态标记为健康, 计算设备的健康得分。

### 4.2 计算设备健康得分

当相对邻近度计算结果小于 0.6 时, 计算设备的健康得分作为设备状态的综合评价指标。

《油浸式变压器(电抗器)状态评价导则》<sup>[20]</sup> 中提出, 状态量应扣分值由状态量劣化程度及权重共同决定。参考导则当中的状态评价标准, 将过热故障和放电故障作为两个故障大类, 拥有相同的基准权重系数。依照每个大类当中的劣化程度, 可将低温过热、局部放电列为 I 级劣化; 将中温过热、低能放电列为 II 级劣化; 将高温过热、高能放电列为 III 级劣化。以基准权重系数乘以劣化程度得到六种故障类型最终的扣分权重, 如表 1 所示。

表 1 故障类型对应的扣分权重

Table 1 Deduction weight corresponding to fault types

| 故障类型 | 编号 $j$ | 权重 $P_j/\%$ |
|------|--------|-------------|
| 低温过热 | 1      | 8.3         |
| 中温过热 | 2      | 16.7        |
| 高温过热 | 3      | 25.0        |
| 局部放电 | 4      | 8.3         |
| 低能放电 | 5      | 16.7        |
| 高能放电 | 6      | 25.0        |

对于每一类确定故障类型  $j$ , 已知数据点  $x$  与其邻近度  $l_{cj}(x, k)$ , 可以将邻近度转化为设备与这一故障类型相关的健康得分。设定 100 为满分, 则设备在某个故障类别下的健康得分  $F_j(t)$  的表达式为

$$F_j(t) = 100 \times (1 - \max(1, l_{cj}(x, k))) \quad (11)$$

对设备在所有故障类别下的故障得分进行加权, 可以得到设备的最终健康状态得分为

$$F(x) = \sum_{j=1}^6 P_j \cdot F_j(x) \quad (12)$$

式中,  $P_j$  为  $j$  类故障对应的权重, 在表 1 中列出。

### 4.3 预测故障发展时间

当相对邻近度计算结果处于区间[0.6, 0.8]时, 利用时间序列相似性分析预测设备从潜伏故障状态发展到故障状态所需的时间。

基于 3.1 节中提供的方法, 确定待测数据时间序列  $X$ 。根据 3.2 节中提供的方法, 对故障数据簇  $C_j$  中和数据点  $x$  最近邻的  $k$  个故障数据点  $y_i (i=1, 2, \dots, k)$  提取对应的时间序列, 标记为  $Y_i (i=1, 2, \dots, k)$ , 利用式(10)分别计算序列  $X$  和序列  $Y_1, Y_2, \dots, Y_k$  的归一距离为  $D_1, D_2, \dots, D_k$ 。对于符合序列距离条件, 即和序列  $X$  之间的距离  $D_i$  小于阈值  $T_D$  的序列  $Y_i$ , 计算故障发展时间  $\Delta t_{bi}$ 。

$$\Delta t_{bi} = t_{y_{bi}} - t_{y_{Ni}} \quad (13)$$

将符合条件的所有故障发展时间归算到与序列  $X$  相同的时间尺度下再求取平均值, 可以得到设备的预测故障发展时间  $\Delta t_b$ 。

$$\Delta t_b = \sum_{\Delta t_{bi} \in C} \frac{\Delta t_{bi} \cdot \frac{dt_{y_i}}{dt_x}}{N_C} \quad (14)$$

式中:  $C$  为所有距离条件的序列  $Y_i$  的故障发展时间  $\Delta t_{bi}$  的集合;  $N_C$  为  $C$  中的元素个数;  $dt_{y_i}$  为序列  $Y_i$  的采样时间间隔;  $dt_x$  为序列  $X$  的采样时间间隔。预测故障发展时间  $\Delta t_b$  反映了设备从当前高危状态转化为故障状态的预期时间长度。

## 5 算例分析

采用某地区电网 110~220 kV 变压器 2008 年至 2015 年的油气信息数据验证方法有效性。在所有油气信息完整记录的变压器中, 存在故障历史的变压器共 116 台, 选取其中 76 台在故障时刻采样的油气信息作为历史故障数据, 如表 2 所示。

表 2 历史故障数据

Table 2 Historical fault data

|      | $C_1$    | $C_2$    | $C_3$    | $C_4$    | $C_5$    | $C_6$    |
|------|----------|----------|----------|----------|----------|----------|
| 故障类型 | 低温<br>过热 | 中温<br>过热 | 高温<br>过热 | 局部<br>放电 | 低能<br>放电 | 高能<br>放电 |
| 数据数  | 11       | 17       | 13       | 9        | 12       | 14       |

首先验证本文方法评价健康设备及发现潜伏故障的有效性。变压器 A 在 2011 年 11 月 13 日有低温过热故障记录, 且其故障时刻采样油气信息未被划定到历史故障数据集中。截取变压器 A 在 2011 年 5 月 3 日到 11 月 13 日区间内共 17 次采样的油气信息数据进行分析。对每个数据点, 计算其和历史故障数据簇  $C_1$ — $C_6$  的相对邻近度, 并以最大相对邻

近度  $l_{\max}$  为依据, 划分变压器的状态, 如图 2 所示。

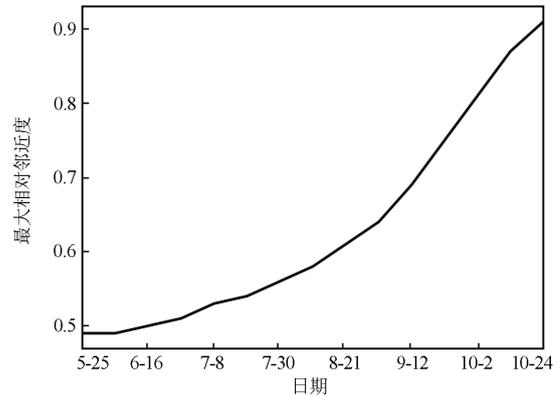


图 2 最大相对邻近度

Fig. 2 Maximum relative proximity

根据设备状态综合评价规则, 在 5 月 3 日到 8 月 21 日间采样的数据, 最大相对邻近度在[0.43, 0.55]区间内, 标记设备状态为健康。在 8 月 21 日到 10 月 15 日间采样的数据, 最大相对邻近度处于区间[0.61, 0.79], 标记设备状态为潜伏故障, 由于  $l_{\max}$  始终等于  $l_{c2}$ , 潜伏故障类型为低温过热。在 10 月 26 日到 11 月 28 日间采样的数据, 最大相对邻近度处于区间[0.81, 0.91], 标记设备状态为故障, 由于  $l_{\max}$  始终等于  $l_{c2}$ , 故障类型为低温过热。

对设备状态为健康区间内的数据点进行健康得分计算, 结果如表 3 所示。

表 3 健康得分

Table 3 Health score

| 日期    | $l_{c1}$ | $l_{c2}$ | $l_{c3}$ | $l_{c4}$ | $l_{c5}$ | $l_{c6}$ | 健康<br>得分 |
|-------|----------|----------|----------|----------|----------|----------|----------|
| 5月25日 | 0.49     | 0.31     | 0.21     | 0.09     | 0.21     | 0.18     | 76.752   |
| 6月5日  | 0.49     | 0.31     | 0.21     | 0.08     | 0.22     | 0.18     | 76.668   |
| 6月16日 | 0.5      | 0.33     | 0.21     | 0.09     | 0.21     | 0.19     | 76.085   |
| 6月27日 | 0.51     | 0.34     | 0.22     | 0.07     | 0.23     | 0.18     | 75.667   |
| 7月8日  | 0.53     | 0.33     | 0.22     | 0.08     | 0.23     | 0.19     | 75.335   |
| 7月19日 | 0.54     | 0.36     | 0.22     | 0.09     | 0.23     | 0.18     | 74.918   |
| 7月30日 | 0.56     | 0.35     | 0.23     | 0.08     | 0.22     | 0.18     | 74.919   |
| 8月10日 | 0.58     | 0.35     | 0.24     | 0.07     | 0.22     | 0.18     | 74.586   |

8 月 21 日采样数据的最大相对邻近度达到 0.61, 表征设备开始进入潜伏故障状态。根据 4.1 节当中的方法, 设定最大搜索时间为 120 天, 搜索允许误差为 0.01, 向前搜寻相对邻近度为  $0.5 \pm 0.01$  的数据, 最终选取 6 月 27 日至 8 月 21 日的所有数据点作为待测数据时间序列  $X$ 。根据 4.2 节中的方法, 设定最大搜索时间为 200 天, 初次和二次搜索允许误差均为 0.02, 从  $C_1$  簇对应的变压器历史数据当中

截取历史故障数据时间序列, 成功截取的历史故障数据有 8 组, 记为  $Y_1$ — $Y_8$ 。计算序列  $X$  和序列  $Y_1, Y_2, \dots, Y_8$  的归一距离分别为  $D_1, D_2, \dots, D_8$ , 结果如表 4 所示。

表 4 时间序列归一距离

Table 4 Unitized distance of time serials

| $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ | $D_6$ | $D_7$ | $D_8$ |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.34  | 0.08  | 0.21  | 0.13  | 0.24  | 0.05  | 0.03  | 0.19  |

设定时间序列相似性分析阈值为 0.1, 8 个序列中仅有  $Y_2$ 、 $Y_6$ 、 $Y_7$  与  $X$  的距离小于阈值, 其对应设备的故障发展时间分别为 48 天、31 天、39 天, 利用式(14)计算变压器 A 的故障预测发展时间, 所得结果为 37.3 天。实际情况为变压器 A 从 8 月 21 日开始进入潜伏故障, 至 10 月 2 日进入故障状态, 共计 42 天。预测故障发展时间的相对误差为 6.3%, 预测结果具有一定的参考价值。

最后验证本文方法对故障状态设备的故障类别判断准确性。将存在故障历史但未用于构建历史故障数据的 50 台变压器在故障时刻采样的油气信息作为待测数据, 判断设备状态及故障类型, 并与三比值法判断结果进行比较, 结果如表 5 所示。

表 5 三比值法与本文方法比较结果对照

Table 5 Comparison among three-ratio method and the proposed algorithm

| 故障类别 | 三比值法 |       | 本文方法 |       |
|------|------|-------|------|-------|
|      | 判正数  | 正确率/% | 判正数  | 正确率/% |
| 低温过热 | 6    | 75.0  | 7    | 87.5  |
| 中温过热 | 8    | 80.0  | 9    | 90.0  |
| 高温过热 | 6    | 66.7  | 9    | 100   |
| 局部放电 | 4    | 57.1  | 5    | 71.4  |
| 低能放电 | 5    | 71.4  | 6    | 85.7  |
| 高能放电 | 7    | 77.7  | 8    | 88.9  |
| 合计   | 36   | 72.0  | 44   | 88.0  |

## 6 结语

本文提出了一种基于聚类和时间序列分析的变压器状态评价方法, 得到以下结论:

1) 该方法基于点密度判据进行数据预处理, 消除了噪声数据的影响。基于聚类方法, 将设备状态划分为健康、潜伏故障和故障三个类别, 判断潜伏故障和故障状态设备的故障类型, 相较于三比值等传统故障诊断算法有更高的判断精度。

2) 基于故障类型加权设定了设备健康得分评价指标, 能够直观地反映设备的健康状况, 为设备的运维、检修安排提供参考。

3) 基于时间序列相似性分析方法, 对状态类别为潜伏故障的设备, 能够较准确地预测其向故障状态转化的时间, 可在故障发生前发现设备的隐藏风险, 有利于及早发现和解决设备隐患。

## 参考文献

- [1] 郭创新, 何迪, 张金江, 等. 考虑热老化的油浸式变压器比例风险模型[J]. 电力科学与技术学报, 2014, 29(2): 14-20.  
GUO Chuangxin, HE Di, ZHANG Jinjiang, et al. A proportional hazard model for oil-immersed transformer concerning thermal aging[J]. Journal of Electric Power Science & Technology, 2014, 29(2): 14-20.
- [2] 郭学梅. 基于 ZigBee 的变压器在线监测系统[J]. 河南师范大学学报(自然科学版), 2010, 38(1): 167-169.  
GUO Xuemei. Design of transformer online monitoring system based on ZigBee[J]. Journal of Henan Normal University (Natural Science), 2010, 38(1): 167-169.
- [3] 宋斌, 刘志雄, 李恩文, 等. DGA 中三比值缺码研究初探[J]. 电力自动化设备, 2015, 35(12): 60-65.  
SONG Bin, LIU Zhixiong, LI Enwen, et al. Preliminary study of undefined three-ratio code in DGA[J]. Electric Power Automation Equipment, 2015, 35(12): 60-65.
- [4] 汪可, 李金忠, 张书琦, 等. 变压器故障诊断用油中溶解气体新特征参量[J]. 中国电机工程学报, 2016, 36(23): 6570-6578.  
WANG Ke, LI Jinzhong, ZHANG Shuqi, et al. New features derived from dissolved gas analysis for fault diagnosis of power transformers[J]. Proceedings of the CSEE, 2016, 36(23): 6570-6578.
- [5] 潘超, 马成廉, 郑玲峰, 等. 一种结合模糊 TOPSIS 法和 BP 神经网络的变压器故障诊断方法[J]. 电力系统保护与控制, 2009, 37(9): 20-24, 29.  
PAN Chao, MA Chenglian, ZHENG Lingfeng, et al. A new method based on fuzzy TOPSIS and BP neural network for power transformer fault diagnosis[J]. Power System Protection and Control, 2009, 37(9): 20-24, 29.
- [6] 吕守国, 崔玉新, 冯迎春, 等. 基于综合分析方法的变压器故障诊断技术与软件开发[J]. 电力系统保护与控制, 2017, 45(10): 123-128.  
LÜ Shouguo, CUI Yuxin, FENG Yingchun, et al. Transformer fault diagnosis method based on comprehensive analysis and its software development[J]. Power System Protection and Control, 2017, 45(10): 123-128.
- [7] 曹松彦, 杨俊, 柯于进. 变压器油中含气量超标原因分析及对策[J]. 热力发电, 2018, 47(8): 138-142.  
CAO Songyan, YANG Jun, KE Yujin. Cause analysis and countermeasures of excessive gas content in transformer oil[J]. Thermal Power Generation, 2018, 47(8): 138-142.
- [8] 王少芳, 蔡金锭. GA-BP 混合算法在变压器色谱诊断法中的应用[J]. 高电压技术, 2003, 29(7): 3-6.

- WANG Shaofang, CAI Jinding. Application of hybrid algorithm based on GA-BP in transformer diagnosis using GAS chromatographic method[J]. High Voltage Engineering, 2003, 29(7): 3-6.
- [9] 施恂山, 马宏忠, 张琳, 等. PSO 改进 RBPNN 在变压器故障诊断中的应用[J]. 电力系统保护与控制, 2016, 44(17): 39-44.
- SHI Xunshan, MA Hongzhong, ZHANG Lin, et al. Application of RBPNN improved by PSO in fault diagnosis of transformers[J]. Power System Protection and Control, 2016, 44(17): 39-44.
- [10] 潘翀, 陈伟根, 云玉新, 等. 基于遗传算法进化小波神经网络的电力变压器故障诊断[J]. 电力系统自动化, 2007, 31(13): 88-92.
- PAN Chong, CHEN Weigen, YUN Yuxin, et al. Fault diagnosis of power transformers based on genetic algorithm evolving wavelet neural network[J]. Automation of Electric Power Systems, 2007, 31(13): 88-92.
- [11] 石鑫, 朱永利, 王刘旺, 等. 基于深度信念网络的电力变压器故障分类建模[J]. 电力系统保护与控制, 2016, 44(1): 71-76.
- SHI Xin, ZHU Yongli, WANG Liuwang, et al. Power transformer fault classifying model based on deep belief network[J]. Power System Protection and Control, 2016, 44(1): 71-76.
- [12] 尹玉娟, 王媚, 张金江, 等. 一种自主核优化的二值粒子群优化-多核学习支持向量机变压器故障诊断方法[J]. 电网技术, 2012, 36(7): 249-254.
- YIN Yujuan, WANG Mei, ZHANG Jinjiang, et al. An autonomic kernel optimization method to diagnose transformer faults by multi-kernel learning support vector classifier based on binary particle swarm optimization[J]. Power System Technology, 2012, 36(7): 249-254.
- [13] 杨洪, 古世甫, 陶加贵, 等. 自适应分级多分类支持向量机在变压器故障诊断中的应用[J]. 高压电器, 2010, 46(5): 49-52.
- YANG Hong, GU Shifu, TAO Jiagui, et al. Application of adaptive hierarchical multi-class SVM to transformer fault diagnosis[J]. High Voltage Apparatus, 2010, 46(5): 49-52.
- [14] 郭创新, 朱承治, 张琳, 等. 应用多分类多核学习支持向量机的变压器故障诊断方法[J]. 中国电机工程学报, 2010, 30(13): 128-134.
- GUO Chuangxin, ZHU Chengzhi, ZHANG Lin, et al. A fault diagnosis method for power transformer based on multiclass multiple-kernel learning support vector machine[J]. Proceedings of the CSEE, 2010, 30(13): 128-134.
- [15] 杜正聪, 牛高远. 基于加权模糊聚类算法的变压器故障诊断方法[J]. 高压电器, 2014, 50(4): 42-48.
- DU Zhengcong, NIU Gaoyuan. Transformer fault diagnosis based on weighted fuzzy clustering algorithm[J]. High Voltage Apparatus, 2014, 50(4): 42-48.
- [16] 熊浩, 孙才新, 廖瑞金, 等. 基于核可能性聚类算法和油中溶解气体分析的电力变压器故障诊断研究[J]. 中国电机工程学报, 2005, 25(20): 162-166.
- XIONG Hao, SUN Caixin, LIAO Ruijin, et al. Study on kernel-based possibilistic clustering and dissolved gas analysis for fault diagnosis of power transformer[J]. Proceedings of the CSEE, 2005, 25(20): 162-166.
- [17] 刘卫华, 廖瑞金, 杨丽君. 基于点密度加权核模糊聚类的变压器故障诊断方法[J]. 电力自动化设备, 2012, 32(6): 66-69, 79.
- LIU Weihua, LIAO Ruijin, YANG Lijun. Power transformer fault diagnosis based on dot density weighted fuzzy kernel clustering[J]. Electric Power Automation Equipment, 2012, 32(6): 66-69, 79.
- [18] HE J, LIU L, LI W, et al. Development and research on integrated protection system based on redundant information analysis[J]. Protection and Control of Modern Power Systems, 2017, 2(2): 102-110. DOI: 10.1186/s41601-017-0043-3.
- [19] MENG J, ZHANG W. Volume measure in 2DPCA-based face recognition[J]. Pattern Recognition Letters, 2007, 28(10): 1203-1208.
- [20] 国家电网公司生产技术部. 油浸式变压器(电抗器)状态评价导则: Q/GDW 169—2008[S]. 北京: 中国电力出版社, 2008.
- Production and Technology Department of State Grid Corporation of China. Guide for condition evaluation of oil-immersed power transformers (reactors): Q/GDW 169—2008[S]. Beijing China Electric Power Press, 2008.
- [21] RISTAD E S, YIANILOS P N. Learning string-edit distance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(5): 522-532.
- [22] 钟锦源, 张岩, 文福拴, 等. 基于时间序列相似性匹配的输电系统故障诊断方法[J]. 电力系统自动化, 2015, 39(6): 60-67.
- ZHONG Jinyuan, ZHANG Yan, WEN Fushuan, et al. A fault diagnosing method in power transmission systems based on time series similarity matching[J]. Automation of Electric Power Systems, 2015, 39(6): 60-67.

收稿日期: 2018-02-01; 修回日期: 2018-04-14

作者简介:

辛建波(1970—), 男, 博士, 江西省电力科学研究院总工程师, 从事开放式电能质量监测工作; E-mail: 449057978@qq.com

康琛(1987—), 男, 硕士, 从事变电设备状态评价工作; E-mail: kangchen-123@163.com

翁新林(1963—), 男, 硕士, 高级工程师, 研究方向为高电压及超高压输变电管理. E-mail: wengxl@jx.sgcc.com.cn

(编辑 魏小丽)