

# 基于改进关联规则挖掘的变压器油 中溶解气体分析模型

邓佳乐<sup>1</sup>, 孙辰昊<sup>2</sup>, 胡博<sup>3</sup>, 岳一石<sup>5</sup>, 易洲楠<sup>4</sup>, 李绍龙<sup>2</sup>

(1. 国网四川省电力公司天府新区供电公司, 四川 成都 610213; 2. 长沙理工大学电气与信息工程学院, 湖南 长沙 410114;  
3. 国网河南省电力公司, 河南 郑州 450000; 4. 国网湖南省电力有限公司长沙供电分公司, 湖南 长沙 410015;  
5. 国网湖南省电力有限公司电力科学研究院, 湖南 长沙 410007)

**摘要:**关联规则挖掘算法常用于基于油中溶解气体分析的变压器故障诊断中。为进一步提升诊断效果,提出一种基于改进关联规则挖掘模型的变压器故障诊断方法。首先,构建可调整的状态重要度评估标准计算方式,能够适应不同输入特征并将其中的罕见高危数据纳入分析,从而有效应对现实应用过程中可能出现的极端状况;其次,直接基于输入特征量导致的故障风险而非特征量的数据占比或出现频率求解相应故障风险权重,能够更加准确地衡量各特征量所带来的影响;最后应用 Relim 算法进行关联规则挖掘,从而改善挖掘效率。实例仿真结果表明,所提出方法相较采用固定重要度评估标准计算方式、传统风险权重求解方法以及 Apriori 关联规则挖掘算法的故障诊断方法,具有更好的诊断准确率、实际可行性以及运算效率。

**关键词:**变压器故障;油中溶解气体分析;加权关联规则挖掘;组件重要度测量;Relim 算法

DOI:10.19781/j.issn.1673-9140.2022.03.020 中图分类号:TM407 文章编号:1673-9140(2022)03-0165-08

## Evaluation model of the power transformer dissolved gas analysis based on the enhanced association rule mining algorithm

DENG Jiale<sup>1</sup>, SUN Chenhao<sup>2</sup>, HU Bo<sup>3</sup>, YUE Yishi<sup>5</sup>, YI Zhounan<sup>4</sup>, LI Shaolong<sup>2</sup>

(1. Tianfu New Area Power Supply Company, State Grid Sichuan Electric Power Company, Chengdu 610213, China; 2. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China; 3. State Grid Henan Electric Power Company, Zhengzhou 450000, China; 4. Changsha Power Supply Company, State Grid Hunan Electric Power Co., Ltd., Changsha 410015, China; 5. Electric Power Science Research Institute, State Grid Hunan Electric Power Co., Ltd., Changsha 410007, China)

**Abstract:** Association rule mining methods are commonly utilized to analyze the dissolved gas which is applied to diagnose the power transformer fault events. For the purpose of improving the performance, this paper proposes a diagnosis method for power transformer fault events based on the enhanced association rule mining algorithm. Firstly, the conditional significance measurements which can be adapted for different input features are established. Thus the rarely distributed but risky data can be incorporated in analysis, and all the potential circumstances in reality can be considered. Next, the corresponding risk weights of input data are generated through their probability of causing a

收稿日期:2021-03-31;修回日期:2021-06-09

基金项目:湖南省重点实验室开放基金(2020ZNDL006)

通信作者:孙辰昊(1991-),男,博士,讲师,主要从事电力大数据及人工智能理论研究;E-mail:chenhaosun@csust.edu.cn

fault rather than their statistical distribution. Therefore, the impact of each input will be measured more precisely. Finally, Relim algorithm is applied to raise the efficiency of mining. The experimental study shows that the proposed method is more pinpoint, realizable and efficient compared with the methods with the fixed significance measurements, the conventional technique to calculate the risk weight, and Apriori algorithm.

**Key words:** power transformer fault event; dissolved gas; weighted association rule mining; component importance measure; Relim algorithm

电力变压器作为输电线路系统中最重要设施之一,其稳定运行至关重要。任何类型的变压器故障都可能导致电力供应的中断,进而对电网造成极大的损失。考虑到变压器故障通常发生在运行期间,因此,依据变压器内部特性进行及时有效的故障诊断,就能够在第一时间实施针对性的应对措施,从而大大减少潜在故障风险。

同电气参数一样,变压器所含绝缘油中溶解气体的含量也能够提供变压器运行状态的有效信息,但不同之处在于其基本不受变压器内电磁环境变化的影响。因此,变压器油中溶解气体分析(dissolved gas analysis, DGA)常被用于变压器的故障诊断<sup>[1]</sup>。

目前,许多研究者提出了大量关于 DGA 的研究思路,也取得了可观的成果。第 1 类是基于优化算法的 DGA 方法。文献[2]采用基于改进量子粒子群优化支持向量机算法实施变压器故障诊断,能够进一步提高方法的全局搜索能力;针对变压器故障诊断,文献[3]基于征兆子集筛选效果设计了一种征兆优选方法,筛选出的征兆子集相较于传统比值方法具有优势;文献[4]通过结合遗传算法和支持向量机,实现针对 DGA 特征量的优选。基于优化算法的 DGA 方法往往能够获取直观的故障概率,但必需统计时段较长的大容量数据,对输入数据的要求较高。第 2 类是采用神经网络的 DGA 方法。文献[5]在变压器振动机理的基础上,利用自组织特征映射神经网络,提高了变压器故障诊断效率;文献[6]针对传统深度信念网络方法中存在的误判,构建组合 DBN 故障诊断方法,提高了故障诊断效果;文献[7]将卷积神经网络应用在变压器故障诊断中;文献[8]基于模糊神经网络和局部统计提出了一种变压器状态评估方法。基于神经网络的 DGA 方法一般拥有较强的稳定性和容错性,但一般需要较大容量的输入数据。第 3 类是基于向量机的 DGA 方

法。文献[9]通过 BP 神经网络为基于支持向量机的变压器故障检测选择输入特征,提高了准确率;文献[10]结合 AdaBoost 算法和二次映射支持向量机,提出了变压器故障诊断模型;文献[11]应用支持向量机实现对变压器故障的分类与判别。尽管基于向量机的 DGA 方法结果一般为全局最优,但较难解决多分类问题。针对以上这些问题,关联规则挖掘(association rule mining, ARM)由于可根据需求选择输入特征或状态,常被应用于 DGA 方法中。文献[12]搭建一种基于关联规则和变权重系数的变压器状态综合评估的模型;文献[13]为诊断变压器故障,提出了一种结合集对分析和 ARM 的集成算法;文献[14]将概率图像模型应用到 ARM 算法之中,进一步提升了效率。尽管上述文献取得了一定进展,但依然存在一些可改进的地方。首先,传统 ARM 算法通常采用固定统一的重要度评估标准,故一些出现频率较低的数据将被直接剔除。但在这些罕见数据中同样存在能够引起故障的高危数据,所以也应该予以分析;其次,输入数据的风险权重应该由其所产生的风险决定,而非出现频率;最后,这些 ARM 方法的运算效率还可以进一步被提升。

针对上述问题,本文提出一种基于加权关联规则挖掘(weighted association rule mining, WARM)模型的 DGA 方法。在该模型中,针对罕见数据改进重要度评估标准计算方式,从而从罕见数据中筛选出罕见高危数据并加入后续分析中,而非直接舍弃。首先基于组件重要度测量(component importance measure, CIM)提出一种基于输入数据自身影响程度的风险权重计算方法;然后应用 Relim 挖掘流程进行运算,比传统 Apriori 算法运算速度快;最后基于某实际系统变压器运行状态数据实施实例仿真,验证所提方法能够同时改善变压器故障诊断的正确率、运行效率以及现实应用中的可信度。

# 1 加权关联规则挖掘模型

## 1.1 关联规则挖掘原理

ARM 算法最早由 Agrawal 等提出<sup>[15]</sup>,其主要目标是挖掘数据库中各个变量之间的隐含关系。

假设  $I$  为一个包含所有物品的集合,  $X$  与  $T$  是  $I$  的子集,  $X$  被称为物品集,  $T$  被称为目标集。假设  $D = \{t_1, t_2, \dots, t_m\}$  为包含所有记录的数据库, 其中每一条记录均由物品集和目标集组成。若  $X$  与  $T$  之间不存在交集, 且挖掘结果显示: 若  $X$  出现时目标  $T$  也会发生, 则一条关联规则可以被表示为  $X \rightarrow T$ 。

通过对输入数据进行判别, 能够获取相应高频物品集, 并进一步构建出关联规则。目前, 通常采用重要度评估标准实现高频物品集和关联规则的筛选。

支持度 (Support)<sup>[16]</sup> 用于挖掘高频物品集, 可被写为包含物品集  $X$  记录的基数在所有记录的基数中所占比例, 即

$$S_U(X) = \frac{|X \subseteq t|}{|D|} \quad (1)$$

置信度 (Confidence)<sup>[16]</sup> 用于验证所挖掘的规则是否为关联规则, 可由数据库中所有同时含有  $X$  与  $T$  记录的基数在含有  $X$  的记录的基数中所占比例来表示, 即

$$C_O(X \rightarrow Y) = \frac{S_U(X \cup Y)}{S_U(X)} \quad (2)$$

关联规则挖掘的过程通常可概括为 2 步: ①使用预设的支持度阈值挖掘数据库中的高频物品集; ②基于选出的高频物品集, 采用预设的置信度阈值筛选出相应的关联规则。

## 1.2 输入数据的预处理

参照行业标准<sup>[17]</sup>, 本文选取 7 种变压器中溶解气体作为输入特征, 其中包括: 氢气 ( $H_2$ )、甲烷 ( $CH_4$ )、乙炔 ( $C_2H_2$ )、乙烯 ( $C_2H_4$ )、乙烷 ( $C_2H_6$ )、一氧化碳 ( $CO$ ) 的相对产气速率及二氧化碳 ( $CO_2$ ) 的相对产气速率。所研究的 7 种变压器状态包括正常工作、低温过热、中温过热、高温过热、低能放电、高能放电及局部放电。

为方便后续数据挖掘, 需要将数据进行预处理

并实现整合。

假设  $\{t_1, t_2, \dots, t_m\}$  为一个包含  $D$  中各条记录编号的向量。假设  $C = \{c_1, c_2, \dots, c_j, \dots, c_7\}$  为一个包含 7 种输入特征的向量, 其中  $c_j$  为其中任意一个特征。对于  $c_j$ , 假设  $\{v_{j,1}, v_{j,2}, \dots, v_{j,k}, \dots, v_{j,l}\}$  为其中包含所有特征量数值的向量。假设  $T = \{T_1, T_2, \dots, T_i, \dots, T_m\}$  为包括所有记录中变压器状态的向量, 其中  $T_i$  为其中任意一条记录中的变压器状态。  $T_i$  可能属于所有 7 种变压器状态中的一种, 即  $T_i = T(o) \in \{T(g_1), T(g_2), \dots, T(g_7)\}$ 。综合以上假设, 能够构建数据的整合空间矩阵为

$$Z = \begin{bmatrix} C & c_1 & \cdots & c_j & \cdots & c_7 & T \\ t_1 & i_{11} & \cdots & i_{1j} & \cdots & i_{17} & T_1 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ t_i & i_{i1} & \cdots & i_{ij} & \cdots & i_{i7} & T_i \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ t_m & i_{m1} & \cdots & i_{mj} & \cdots & i_{m7} & T_m \end{bmatrix} \quad (3)$$

式中  $i_{ij}$  为第  $i$  行记录中的一个物品, 即所在列对应的特征  $c_j$  中的任意一个特征量数值  $v_{j,k}$ 。

## 1.3 重要度评估标准的改进

作为输电线路系统中最重要组件, 变压器的正常稳定运行需要得到首要保障。不可否认的是, 对于一些出现较少的溶解气体数值的分析与诊断将在一定程度上增加运维检修成本, 但能够将现实中所有可能出现的情况设计相应预案, 从而将变压器的潜在故障风险降至最低。因为当这些罕见高危特征量数值导致变压器故障时, 同样将引起严重的损失。因此, 对出现频率较少的罕见数据进行诊断, 能够从中挖掘出真正的罕见高危数据。如此, 一方面能够在将来这些罕见高危数据再次出现时, 保证快速应对, 另一方面也能有效地改善整体诊断准确率。

目前对于所有的输入特征, 传统 ARM 算法中通常采用预设且相同数值的重要度评估标准计算公式计算重要度得分。由文 1.1 节中的背景介绍可知, 对于含有出现罕见数据的记录, 由传统重要度评估标准计算公式所计算的得分一般较低, 容易低于所设置阈值从而被直接剔除。为改进这一不足, 本文对原有重要度评估标准计算公式进行改进, 基于

式(1)、(2)提出了一套可变的狀態重要度評估標準計算公式。通過應用這一套狀態重要度評估標準計算公式,包含各個特征中的罕見數據的記錄將被單獨提出,並由為該特征所專門生成的狀態重要度評估標準計算公式計算相應得分,再與所設閾值相比較。這樣便能夠得出基於該特征中罕見數據的高頻物品集和關聯規則。

$$S_U(X^c + X^s)_j = \frac{|t_i \in \mathbf{Z}(i,1); X^c \subseteq \mathbf{Z}(i,H) \neq 0; \mathbf{Z}(i,j) \in X^s \neq \varphi|}{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) \in X^s \neq \varphi|} \quad (5)$$

$$c_o(X^c + X^s \rightarrow T)_{j,a} = \frac{|t_i \in \mathbf{Z}(i,1); X^c \subseteq \mathbf{Z}(i,H) \neq 0; \mathbf{Z}(i,j) \in X^s \neq \varphi; \mathbf{Z}(i,n+1) = T(o)|}{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) \in X^s \neq \varphi|} \quad (6)$$

式中  $|\cdot|$  為同時滿足其中所有條件的故障記錄的基數;  $T(o)$  為 7 種變壓器狀態中的一種;  $H$  為一個數值區間,取值 2~8。

由式(6)可知,對於置信度,  $a$  表明存在 7 種不同形式的置信度計算公式,對應含有 7 種不同狀態的記錄,即含有不同故障類型的記錄的置信度計算公式也不相同。

#### 1.4 風險权重計算方法的改進

在通過狀態重要度評估標準計算公式並基於各個特征進行高頻物品集和關聯規則的求解後,需要將這些基於不同特征的結果進行彙總。考慮到不同特征量與變壓器故障之間的關聯程度也不相同,故對於各個輸入特征中所有數值的風險权重進行分析。當前大多數研究一般采用特征量在數據庫中的占比或出現頻率來求解权重。但在現實中,特征量與變壓器故障之間的關聯程度與該特征量的數據占比或出現頻率並沒有直接聯系。因此,本文為實現更加精確地計算風險权重,採用 CIM 計算各個特征量數值對於整體故障風險的影響程度作為其風險权重。

首先,本文假設  $D_j^s$  為數據庫  $D$  的一個子集,

$$\theta_{j,k}^s = \sum_{i=2}^{|\mathbf{Z}(i,1)|} \frac{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) = v_{j,k}; \mathbf{Z}(i,j) \in X^r|}{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) \in c_j|} \prod_{k=1}^l \left( \sum_{i=2}^{|\mathbf{Z}(i,1)|} \frac{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) \neq v_{j,k}; \mathbf{Z}(i,j) \in X^r|}{|t_i \in \mathbf{Z}(i,1); \mathbf{Z}(i,j) \in c_j|} \right) \quad (11)$$

式中  $k$  為共計  $l$  個特征量中的任意一個。

#### 1.5 Relim 算法原理及優勢

目前最為常用的 ARM 算法是 Apriori 算法<sup>[15]</sup>。Apriori 算法首先從數據庫中篩選出單項高頻物品,並採用連枝和剪枝的方法將其逐漸擴展為

首先,將所有記錄中的物品集劃分為兩部分,並重新寫為

$$X^c + X^s \rightarrow T \quad (4)$$

式中  $X^c$ 、 $X^s$  分別為含有常見數據的物品集和含有罕見數據的物品集。

由此,對於含有任意特征  $c_j$  中罕見數據的記錄,相應的狀態重要度評估標準計算公式為

包括含有任意特征  $c_j$  中罕見數據的所有故障記錄。則對於該特征  $c_j$  中的一個特征量  $v_{j,k}$ ,其風險权重由兩部分組成,即

$$\theta_{j,k} = \theta_{j,k}^c + \theta_{j,k}^s \quad (7)$$

式中  $\theta_{j,k}^c$ 、 $\theta_{j,k}^s$  分別為常見物品集和罕見物品集部分的風險权重。

當  $v_{j,k}$  為  $c_j$  中的常見數值時,式(7)中的  $\theta_{j,k}^s$  為 0,而  $\theta_{j,k}^c$  可被寫為

$$\theta_{j,k}^c = \sum_{i=2}^{|\mathbf{Z}(i,1)|} \frac{|\mathbf{Z}(i,j) = v_{j,k}|}{m} \quad (8)$$

當  $v_{j,k}$  為  $c_j$  中的罕見數值時,式(7)中的  $\theta_{j,k}^c$  為 0,  $\theta_{j,k}^s$  可通過 CIM 求解。Birnbbaum's 測量<sup>[18]</sup>是一種簡便的 CIM 計算方法,其數學表達為

$$I^B(k) = \frac{\partial h(p)}{\partial p_k} \quad (9)$$

通過 pivotal 分解,式(7)能被線性表示為

$$I^B(k) = h(1_k, p) - h(0_k, p) \quad (10)$$

式中  $h(1_k, p)$ 、 $h(0_k, p)$  分別為當特征量  $v_{j,k}$  確定與變壓器故障相關或無關時的整體故障風險。

通過式(10),  $\theta_{j,k}^s$  可被寫為

更多項的物品集,直到這些物品集滿足相應要求成為高頻物品集。最後通過確定高頻物品集進而求解出相應的關聯規則。

儘管結果準確,但 Apriori 算法運行過程中將產生較多的候選物品集,從而導致生成大量冗余規則,

降低了运行效率。针对这个问题,Relim 算法在运行过程中无需候选物品集,具有结构简单、运行速度快的优点。Relim 算法主要通过建立相应的记录链表组求解相应高频物品集。因此,Relim 算法无需诸如高频模式树等复杂数据结构,其中所有挖掘过程能够在简单的递归函数中完成,加快了挖掘速度。Relim 算法的基本流程如下:

- 1)由数据库中搜索单项高频物品集,按支持度大小排序;
- 2)将转换后的数据库设为记录链表组,其中各记录链表按头元素支持度大小排序;
- 3)按顺序依次对每个记录链表进行搜索,挖掘出高频物品集;然后将该记录链表删除,并构建以该链表中头元素为前缀的新记录链表组;将原记录链表组和新记录链表组合并;
- 4)将所有记录链表挖掘完毕。

## 1.6 基于 WARM 的 DGA 方法的构建

根据上述讨论,本文构建基于 WARM 的 DGA 方法,基本流程如图 1 所示。

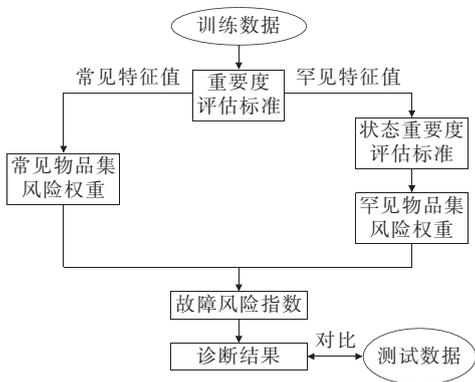


图 1 基于 WARM 的 DGA 方法流程

Figure 1 Flow chart of the WARM-based DGA method

基于 WARM 的 DGA 方法流程具体如下:

- 1)对训练数据库中的第一个输入特征中的所有特征量数值采用预设重要度评估标准阈值挖掘出相应罕见数值;
- 2)将含有该特征中罕见数值的记录集中于数据库子集中,并通过相应特征的状态重要度评估标准得分计算公式挖掘出高频物品集和关联规则;
- 3)将训练数据库中所有特征依次重复进行前两步;
- 4)分别计算常见和罕见物品集的风险权重,并

汇总;

5)基于风险权重,将测试数据库中的记录依据其所含有的各个特征的特征量数值计算相应的故障风险指数;

6)将故障风险指数与记录的真实结果对比。

## 2 算例分析

### 2.1 实验数据

本文采用中部某省区域内高压线路系统中变压器记录进行实验验证。样本数据共计 564 条,涵盖文 1.2 节中所引述的 7 种气体的含量( $H_2$ 、 $CH_4$ 、 $C_2H_2$ 、 $C_2H_4$ 、 $C_2H_6$ )和相对产气速率( $CO$ 、 $CO_2$ ),以及 7 种变压器状态(正常工作、低温过热、中温过热、高温过热、低能放电、高能放电及局部放电)。

### 2.2 实验验证方法

本文采用 3:1 的比例划分实验数据,即 423 条记录作为训练数据,141 条记录作为测试数据。

在将诊断结果与测试数据对比时,本文通过接收者操作特征曲线(receiver operating characteristic, ROC)和精确—召回曲线(precision-recall, PR)共同衡量相应对比结果。在这两类曲线的基础上,采用线下包围面积(area under the curve, AUC)为诊断方法效果的检验参数,其中 AUC 的数值越高则证明诊断越准确。

### 2.3 整体故障诊断结果

首先,本文将输入数据库涵盖的所有类型变压器故障作为整体,并采用基于 WARM 的 DGA 方法对变压器是否故障进行诊断,即变压器的状态仅有故障和正常工作 2 种。

为验证所提出 DGA 方法(WARM(Relim))的有效性,加入 2 种 DGA 方法作为对比。其中,一种是同样采用改进后的重要度评估标准和风险权重计算方式但挖掘时应用传统 Apriori 算法的 DGA 方法(WARM(Apriori)),另一种是采用传统重要度评估标准和风险权重计算方式及 Apriori 算法的 DGA 方法(ARM(Apriori))。3 种 DGA 方法基于整体故障诊断结果的 ROC 曲线及 PR 曲线对比如图 2、3 所示,3 种方法的运行时间如表 1 所示。

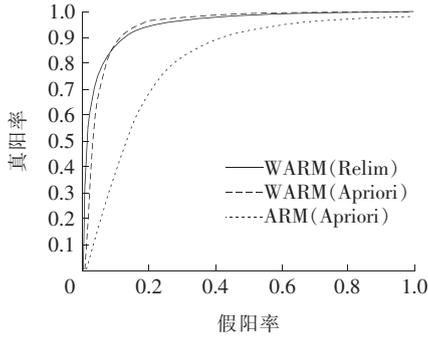


图 2 基于整体故障诊断结果的 ROC 曲线对比

Figure 2 Comparison of the general diagnosis case by the ROC curve

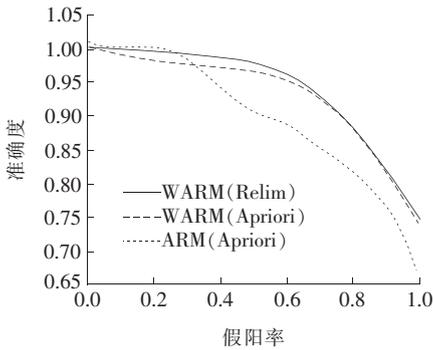


图 3 基于整体故障诊断结果的 PR 曲线对比

Figure 3 Comparison of the general diagnosis case by the PR curve

表 1 整体故障诊断的运行时间对比

Table 1 Comparison of the processing time of the general diagnosis case

方法	运行时间/s
WARM(Relim)	0.030 9
WARM(Apriori)	0.035 6
ARM(Apriori)	0.034 7

由图 2、3 可知,采用基于改进重要度评估标准和风险权重计算方式的 WARM 方法相比传统 ARM 方法,能够实现更为精确的变压器故障诊断。其中,基于 WARM 和 Relim 的 DGA 方法相较于基于 ARM 和 Apriori 的 DGA 方法,分别在 ROC 和 PR 曲线的 AUC 数值上提升了 17.3% 和 13.8%。此外,还可以得出应用 Relim 和 Apriori 算法得到的诊断结果精确性较为接近,即 Relim 算法并不能显著改善诊断精度。但由表 1 可知,应用 Relim 算法的运行时间相较于应用 Apriori 算法减少了 12.3%。因此,本文所提出的基于 WARM 和 Relim 的 DGA 方法能够在减少运行时间的基础上有效地提升诊断精度。

## 2.4 各类型故障诊断结果

本文对所有 7 种变压器状态实施分类诊断,所得出的诊断精度与运行时间对比如图 4~6 所示。

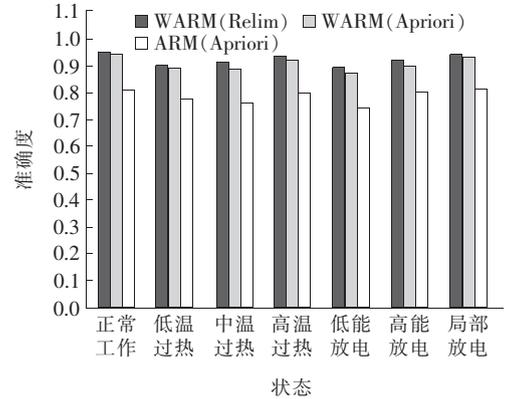


图 4 基于各类型故障诊断的 AUC(ROC) 对比

Figure 4 Comparison of the fault diagnosis case by the AUC (ROC curve)

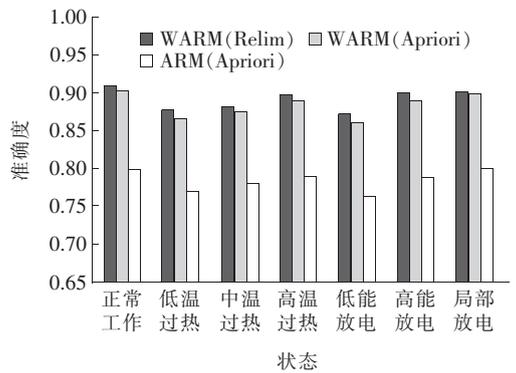


图 5 基于各类型故障诊断的 AUC(PR) 对比

Figure 5 Comparison of the fault diagnosis case by the AUC (PR curve)

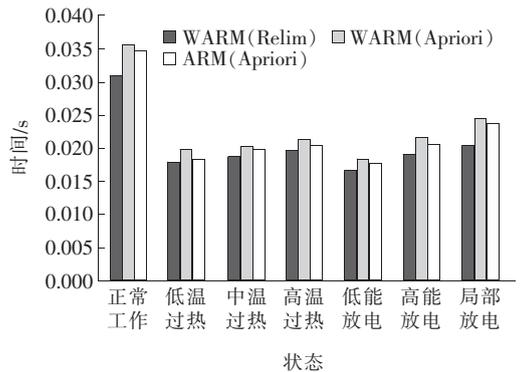


图 6 各类型故障诊断的运行时间对比

Figure 6 Comparison of the processing time of the fault diagnosis case

由图 4、5 可知,对重要度评估标准和风险权重计算方式的改进均能够有效地改善针对所有 7 种变

压器状态诊断的精度。其中,基于 WARM 和 Relim 的 DGA 方法相较于基于 ARM 和 Apriori 的 DGA 方法,分别在 7 组 ROC 和 PR 曲线的 AUC 数值上平均提升了 15.6% 和 12.7%。由图 6 可知,应用 Relim 算法的运行时间相较于应用 Apriori 算法平均减少了 10.7%。因此,对于变压器的不同运行状态,本文所提出的基于 WARM 和 Relim 的 DGA 方法同样能够在减少运行时间的基础上有效地提升针对每一种变压器状态的诊断精度。此外,该方法也能够分析可能出现的罕见高危特征量和变压器故障类型,从而进一步减少变压器出现故障的风险。

### 3 结语

针对目前基于关联规则挖掘算法的 DGA 方法中所存在的直接忽略罕见高危数据、特征量权重计算过于简单以及挖掘所需时间较长,本文提出了一种基于 WARM 的变压器 DGA 诊断方法,主要研究如下:

1) 为将各个特征中的罕见高危数据纳入分析,对重要度评估标准计算公式进行了改进,能够在提升诊断精度的同时涵盖现实中可能出现的极端情况;

2) 基于 CIM 直接求解各个输入特征量导致变压器故障的风险程度,相较基于出现频率能够更加准确地衡量相应的故障风险权重;

3) 采用 Relim 算法进行关联规则挖掘,相较传统 Apriori 算法能够有效地改善挖掘效率。

实验结果表明,本文所提出的变压器故障诊断方法能够同时改善诊断的正确率、实用性和运行效率。

### 参考文献:

[1] 何先华,张远鹏,崔桂兴,等. 基于回归算法的变压器故障检测方法研究[J]. 电力系统保护与控制, 2020, 48(21):132-139.  
HE Xianhua, ZHANG Yuanpeng, CUI Guixing, et al. Research on transformer fault detection method based on a regression algorithm[J]. Power System Protection and Control, 2020, 48(21):132-139.

[2] 党东升,张树永,葛鹏江,等. 基于改进量子粒子群优化支持向量机的变压器故障诊断方法[J]. 电力科学与技术学报, 2019, 34(3):108-113.

DANG Dongsheng, ZHANG Shuyong, GE Pengjiang, et al. Transformer fault diagnosis method based on support vector machine optimized by improved quantum-behaved particle swarm optimization[J]. Journal of Electric Power Science and Technology, 2019, 34(3):108-113.

[3] 张育杰,冯健,李典阳,等. 基于油色谱数据的变压器故障征兆新优选策略[J]. 电网技术, 2021, 45(8):3324-3331.

ZHANG Yujie, FENG Jian, LI Dianyong, et al. New feature selection method for transformer fault diagnosis based on DGA data[J]. Power System Technology, 2021, 45(8):3324-3331.

[4] 王晶,许素安,洪凯星,等. 基于 DGA 特征量优选与 GA-SVM 的变压器故障诊断模型[J]. 变压器, 2020, 57(12):36-40+46.

WANG Jing, XU Suan, HONG Kaixing, et al. Transformer based on DGA feature quantity optimization and GA-SVM fault diagnosis model[J]. Transformer, 2020, 57(12):36-40+46.

[5] 夏玉剑,李敏,向天堂,等. 基于 SOM 的变压器绕组和铁芯故障诊断[J]. 电力科学与技术学报, 2018, 33(2):129-134.

XIA Yujian, LI Min, XIANG Tiantang, et al. Fault diagnosis of transformer winding and core based on SOM[J]. Journal of Electric Power Science and Technology, 2018, 33(2):129-134.

[6] 刘胜军,孙志鹏,沈辰,等. 基于振动频谱分析和总谐波畸变率的电力变压器故障诊断方法研究[J]. 电网与清洁能源, 2021, 37(3):86-91.

LIU Shengjun, SUN Zhipeng, SHEN Chen, et al. Research on fault diagnosis of power transformers based on spectral analysis of vibration signals and total harmonic distortion[J]. Power System and Clean Energy, 2021, 37(3):86-91.

[7] 郝玲玲,朱永利,王永正. 基于 DCAE-KSSELM 的变压器故障诊断方法[J]. 中国电力, 2022, 55(2):125-130.

HAO Lingling, ZHU Yongli, WANG Yongzheng. Transformer fault diagnosis method based on DCAE-KSSELM[J]. Electric Power, 2022, 55(2):125-130.

[8] RIGATOS G, SIANO P. Power transformers' condition monitoring using neural modeling and the local statisti-

- cal approach to fault diagnosis[J]. International Journal of Electrical Power and Energy Systems, 2016, 80(1): 150-159.
- [9] 陈铁,陈卫东,李咸善,等. 基于 EMD 和 GCT 的变压器油中溶解气体预测[J]. 高压电器, 2022, 58(04): 70-79.  
CHEN Tie, CHEN Weidong, LI Xianshan, et al. Dissolved gas prediction in transformer oil based on EMD and GCT[J]. High Voltage Apparatus, 2022, 58(04): 70-79.
- [10] 刘君,赵立进,黄良,等. 基于 Adaboost. MK 和 SM-SVDD 的变压器故障诊断方法[J]. 电力科学与技术学报, 2017, 32(3): 139-144+152.  
LIU Jun, ZHAO Lijin, HUANG Liang, et al. Fault diagnosis method for power transformer based on Adaboost. MK and SM-SVDD[J]. Journal of Electric Power Science and Technology, 2017, 32(3): 139-144 + 152.
- [11] BACHA K, SOUAHLIA S, GOSSA M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine[J]. Electric Power Systems Research, 2012, 83(1): 73-79.
- [12] 谭贵生,曹生现,赵波,等. 基于关联规则与变权重系数的变压器状态综合评估方法[J]. 电力系统保护与控制, 2020, 48(1): 88-95.  
TAN Guisheng, CAO Shengxian, ZHAO Bo, et al. An assessment of power transformers based on association rules and variable weight coefficients[J]. Power System Protection and Control, 2020, 48(1): 88-95.
- [13] LI L, YONG C, JUN X, et al. An integrated method of set pair analysis and association rule for fault diagnosis of power transformers[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2015, 22(4): 2368-2378.
- [14] SHENG G, HOU H, JIANG X, et al. A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model[J]. IEEE Transactions on Smart Grid, 2018, 9(2): 695-702.
- [15] AGRAWAL R, IMIELIŃSKI T, SWAMI A. Mining association rules between sets of items in large databases[J]. Acm Sigmod Record, 1993, 22(2): 207-216.
- [16] HIPPIE J. Algorithms for association rule mining—a general survey and comparison[J]. ACM, 2000, 2(1): 58-64.
- [17] 孔灿,张大宁,田杰,等. 110 kV 变压器油纸绝缘套管不均匀绝缘受潮劣化分析[J]. 智慧电力, 2020, 48(4): 119-124.  
KONG Can, ZHANG Daning, TIAN Jie, et al. Non-uniform moisture degradation analysis of 110 kV transformer oil-paper insulation bushing[J]. Smart Power, 2020, 48(4): 119-124.
- [18] BIRNBAUM Z. On the importance of different components in a multicomponent system[J]. Multivariate Analysis II, 1968, 1(1): 581-592.