

引用格式:张少峰,王佳琳,李润润,等.基于IAR-CI模型的配变重过载预测方法[J].电力科学与技术学报,2024,39(5):67-76.

Citation: Zhang Shaofeng, Wang Jialin, Li Runrun, et al. Overload prediction method of distribution transformer based on IAR-CI model[J]. Journal of Electric Power Science and Technology, 2024, 39(5): 67-76.

基于 IAR-CI 模型的配变重过载预测方法

张少峰¹, 王佳琳^{1,2}, 李润润^{1,2}, 段笑阳^{1,2}, 孙辰昊³, 陈春³

(1. 国网河南省电力公司三门峡供电公司, 河南 三门峡 472000; 2. 国网河南省电力公司卢氏县供电公司, 河南 三门峡 472200;
3. 长沙理工大学电气与信息工程学院, 湖南 长沙 410114)

摘要:随着电力系统的数字化和智能化发展,配变重过载预测成为了实现智能状态检修的关键技术之一。配变过载时空因子在现实场景中通常呈偏置分布。其中,部分高风险罕见(high risk and rare, HRR)因子一旦出现,将对变压器造成无法逆转的伤害。为此,该文提出一种基于提高关联规则关键重要性(improved association rules-criticality importance, IAR-CI)模型的配变重过载预测方法。首先,考虑内部与外部因素,收集多个数据源并建立配变运行状态数据库,且通过ICA识别与配变重过载强关联的罕见高危时段与HRR;其次,基于关键性重要度(criticality importance, CI)度量计算,设计一种因子权重计算方法,准确衡量因子的风险权重;最后,应用TBFP-Growth算法,增强模型的运行效率。采用中国南方某地区电网数据进行算例仿真。研究表明,该方法能够提升配变重过载的预测性能,有助于后续巡检、检测策略的合理统筹和科学规划,可在降低电力设备运维检修成本的同时提高供电的可靠性。

关键词:配变;重过载;关联规则;关键性重要度(CI);高风险罕见因子(HRR)

DOI: 10.19781/j.issn.1673-9140.2024.05.007 中图分类号: TM421 文章编号: 1673-9140(2024)05-0067-10

Overload prediction method of distribution transformer based on IAR-CI model

ZHANG Shaofeng¹, WANG Jialin^{1,2}, LI Runrun^{1,2}, DUAN Xiaoyang^{1,2}, SUN Chenhao³, CHEN Chun³

(1. Sanmenxia Electric Company, State Grid Hebei Electric Power Co., Ltd., Sanmenxia 472000, China; 2. Lushi Electric Company, State Grid Hebei Electric Power Co., Ltd., Sanmenxia 472200, China; 3. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China)

Abstract: With the digitization and intellectualization of power systems, the prediction of distribution transformer overload has become one of the key technologies for realizing intelligent condition-based maintenance. In real-world scenarios, the spatiotemporal factors of distribution transformer overload often exhibit a biased distribution, among which some high-risk and rare (HRR) factors, once occurred, can cause irreversible damage to transformers and should not be ignored. Therefore, this paper proposes a prediction method for distribution transformer overload based on the improved association rules-criticality importance (IAR-CI) model. Firstly, considering both internal and external factors, multiple data sources are collected to establish a database of distribution transformer operating states, and ICA is used to identify rare high-risk periods and HRR factors that are strongly associated with severe transformer overload. Secondly, based on the criticality importance (CI) metric calculation, a factor weighting method is designed to accurately measure the risk weight of each factor. Finally, the TBFP-Growth algorithm is applied to enhance the operational efficiency of the model. Simulation analysis conducted in a region in southern China demonstrates that the proposed method can improve the prediction performance of severe distribution transformer overload, facilitating the reasonable planning and scientific scheduling of subsequent inspection and testing strategies. This reduces the operation

收稿日期: 2023-07-12; 修回日期: 2023-10-11

基金项目: 国网河南省电力公司科技项目资助(521710230002)

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等方面的研究; E-mail: chenchaosun@csust.edu.cn

and maintenance costs of power equipment while enhancing the reliability of power supply.

Key words: distribution transformer, overload, association rules, criticality importance, high-risk and rare factors

变压器作为电力系统的关键设备,在电力传输和分配中扮演着至关重要的角色。变压器的正常运行对于确保电网的稳定性、可靠性和持续供电至关重要^[1-2]。然而,变压器面临着各种潜在的运行问题,其中之一便是重过载。变压器重过载是指变压器运行时所承受的负荷超出其额定容量的情况。变压器过载会导致变压器内部发热加剧,可能引发设备损坏、损失电能转换效率、缩短设备寿命甚至发生火灾等严重后果。此外,过载可能导致电力故障、电压波动等问题。这些潜在的问题不仅直接影响了居民用电的满意度,还对人民的生产、生活产生了威胁;也给电网的安全运行带来了极大的风险^[3-4]。

在当前的电力系统中,提前发现并防止变压器的重过载情况的重过载预测成为了一个关键的研究领域,为此,许多方法和技术被提出并得到了应用。

一种常见的重过载预测方法是基于历史数据的统计分析。这种方法利用历史负荷数据进行分析 and 建模,通过统计方法来推断未来的重过载情况。文献[5-6]分别通过 logistic 回归模型和极限梯度提升(extreme gradient boosting, XGBoost)算法分析历史负荷等数据,对变压器重过载故障进行预测。文献[7]将粗糙集理论与 D-S 证据理论引入多元回归算法,利用粗糙集理论对影响因子进行重要性排序,建立多个多元回归模型对负荷进行预测。文献[8]基于历史运行数据建立配电网源荷随机行为的马尔可夫模型,采用吉布斯算法对该模型进行蒙特卡洛随机抽样,评判载荷安全性。该类方法的不足之处在于当出现新的负荷模式、数据缺失或者异常时,统计模型可能无法有效地适应和预测。另一种常见的方法是基于物理模型的预测。文献[9]通过引入导线动态电热特征量,构建了新的控制模型,挖掘输电线路耐受过负荷的能力。文献[10]通过分析变压器的热源组成与传热过程,提出了基于温升特性的油浸式变压器负荷能力评估模型。文献[11]对配电变压器建立了动态热路模型,精确计算了波动过载工况下的动态传热过程。该类方法建立了与变压器相关的数学模型,考虑了其物理特性和电热参数,通过模拟变压器的运行状

态来预测重过载情况,但通过这类方法建立准确的物理模型可能相对困难且耗时,需要大量的系统参数和拓扑信息,并且模型中的参数和假设可能存在误差,影响预测结果的精度。近年来,机器学习方法在重过载预测中得到了广泛应用。如在支持向量机(support vector machines, SVM)方面,文献[12-13]提出了采用改进灰狼算法与最小二乘支持向量机耦合的方法来诊断变压器故障。在神经网络方面,文献[14-15]提出了一种基于卷积神经网络和门控循环单元组合神经的变压器短期故障预测模型。文献[16]提出了一种基于图神经网络的变压器故障预测模型。文献[17]提出了一种卷积神经网络和长短期记忆网络的混合模型,对变压器短期负荷进行预测。在贝叶斯方面,文献[18]提出了加权秩和比与改进贝叶斯相结合的变压器故障诊断模型。文献[19-20]提出一种基于贝叶斯网络和假设检验的溶解气体分析的新型多分类概率诊断模型。在深度学习算法方面,文献[21]基于深度学习算法,针对网络野火信息不完全的多微电网配电系统,提出了一种新的概率潮流计算方法。文献[22-23]采用深度学习算法的堆叠自编码器提取变压器数据,通过加权贝叶斯算法整理和分析数据,实现了对一台 10 kV 变压器的故障诊断。这些方法使用机器学习算法学习变压器过载与其他输入因子之间的关系并进行预测。然而,这些机器学习方法在计及多源因子的场景时,没有针对现实场景时空数据偏置的环境下做出相关的分析。此外,机器学习模型的调参和优化也可能较为困难。

为此,基于目前重过载预测方法现状的不足,本文提出一种基于 IAR-CI 模型的配变过载预测方法。该方法能直接挖掘大量偏置数据中的潜在规律和联系,且可以基于不完整或短跨度的输入数据生成预测结果,无须进行归一化等预备工作。首先,对 4 种关联指标从时空双维度实施了相应改进,涵盖极端的因子与场景;其次,通过关键性重要度(critically importance, CI)重要度诊断,建立了各个元素的相对风险权重求解方法,能够更加真实有效地衡量它们对于系统整体稳定性的影响程度;最后,整合模型并使用 TBFP-Growth 挖掘算法运行仿真。经实际算例分析,本文所提出的提高关联规则关键重要性(improved association

rules-criticality importance, IAR-CI)模型的有效性及灵活性均得到了验证。

1 考虑内部与外部数据配变状态特征库建立

在全面预测配电变压器过载风险时,需要综合考虑内部和外部因素的影响^[24]。这些内外因素之间的共同作用决定了配电变压器过载的风险水平。为了确保模型的综合性和准确性,本研究在构建配电变压器状态特征库时,综合考虑了不同因素的影响,并在选择输入条件时采取了综合策略,如图 1 所示。

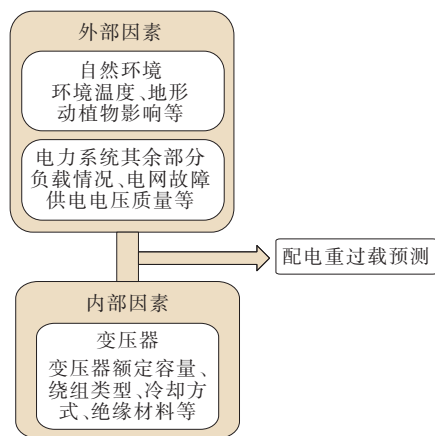


图 1 配电变压器重过载的内外部影响因素
Figure 1 Internal and external influencing factors of distribution transformer overload

为完成对配变重过载的预测,需要对收集的内、外部数据进行整合,并汇总到统一的数据处理空间中。在本文中,假设 $T = \{t_1, t_2, \dots, t_j, \dots, t_n\}$ 为一个包含所有输入特征的集合,其中, t_j 为输入特征变量。每一个输入特征 t_j 都由一组因子 $c_{j,1}, c_{j,2}, \dots, c_{j,k}, \dots, c_{j,i}$ 组成。设 $N = \{n_1, n_2, \dots, n_i, \dots, n_m\}$ 为 m 条记录所对应的标签。设 $Y = \{Y_1, Y_2, \dots, Y_i, \dots, Y_m\}$ 为 m 个目标变量集合, Y_i 表示一个目标变量,在本文中,其代表变压器重过载是否中的一个。设 $P = \{P_1, P_2, \dots, P_i, \dots, P_m\}$ 为 m 个包含所有故障所在季度的集合。

基于这些设定,将各个集合写为矩阵形式: $N = (n_1, n_2, \dots, n_i, \dots, n_m)^T$, $T = (t_1, t_2, \dots, t_j, \dots, t_n)$, $Y = (Y_1, Y_2, \dots, Y_i, \dots, Y_m)^T$, $P = (P_1, P_2, \dots, P_i, \dots, P_m)^T$ 。基于此,本章构建了进行配变重过载预测的处理空间矩阵 D_s ,即

$$D_s = \begin{bmatrix} N & t_1 & \dots & t_j & t_{j+1} & \dots & t_n & Y & P \\ n_1 & m_{11} & \dots & m_{1j} & m_{1j+1} & \dots & m_{1n} & Y_1 & P_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ n_i & m_{i1} & \dots & m_{ij} & m_{ij+1} & \dots & m_{in} & Y_i & P_i \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ n_m & m_{m1} & \dots & m_{mj} & m_{mj+1} & \dots & m_{mn} & Y_m & P_m \end{bmatrix} \quad (1)$$

式中,从第二行开始的每一行都代表一次故障的记录; t_1, \dots, t_j 代表外部因素特征; t_{j+1}, \dots, t_n 代表内部因素特征; m_{ij} 代表记录 n_i 中特征 t_j 的一个环境因子 $c_{j,k}$; Y_i 是多因子共同决定的结果; n_i 是此故障事件的编号标签; P_i 表示一次故障所在的季度,即 $P_i = P(s) \in \{P(1), P(2), P(3), P(4)\}$ 表示 4 个季度中的任意一个。

2 计及低概率高风险因子的改进关联识别方法

2.1 关联规则指标的选取

对于变压器中的重过载记录,令 $I = \{m_1, m_2, m_3, \dots\}$ 为包含其中所有数据库环境因子的集合,假设 X 为 I 的一个子集, Y 为目标变量,则一条规则可以表示为: $X \rightarrow Y$ 。如果该规则通过所有关联状态阈值,那么可以认为当因子 X 出现时,结果 Y 也很可能会出现^[25-26]。

关联规则是为了寻找不同事务项集之间的关联关系。若存在事务前项 X 与事务后项 Y ,两者之间的因果依存关系就是其关联规则。目前,最常用的两个关联重要度诊断指标为支持度和置信度。

支持度 (support) 表示在输入数据库 D_s 中,含有 X 的记录数量占总数的比例,即通过计算特定规则在所有变压器重过载事件中出现的频率来衡量。较高的支持度表示该规则在重过载故障中的出现较为常见。对于一条关联规则 $X \rightarrow Y$, X 的支持度 $u(X)$ 的公式为

$$u(X) = \frac{B_f(n \in D_s, X_n)}{B_f(D_s)} \times 100\% \quad (2)$$

式中, $B_f(D_s)$ 为 D_s 中 i 满足所包含全部条件的故障纪录的基数。

置信度 (confidence) 表示在输入数据库 D_s 中,同时含有 X 和 Y 的项集数量在含有 X 的记录中所占比例。置信度可以用来评估规则的关联性。可以计算规则中描述的过载特征或条件(如过载电流、过载温度等)出现时,发生变压器重过载的概率。较高的置

信度指示该规则与变压器重过载之间的关联较强。一条关联规则 $X \rightarrow Y$ 的置信度 $c(X \rightarrow Y)$ 公式为

$$c(X \rightarrow Y) = \frac{u(X \cap Y)}{u(X)} \times 100\% \quad (3)$$

在变压器重过载故障的关联规则挖掘中,除了上述支持度和置信度之外,还可以考虑使用Kulc度量 K 和不平衡比 R 来选取关联规则的指标。

Kulc度量(Kulczynski)是一种衡量规则质量的指标,它结合了规则的置信度和支持度^[27]。该度量考虑了规则的准确性和频繁性,并可用于判断规则是否具有足够的重要性。一条关联规则 $X \rightarrow Y$ 的Kulc系数 K 的公式为

$$K(X \rightarrow Y) = \frac{c(X \rightarrow Y) + c(Y \rightarrow X)}{2} \quad (4)$$

在Kulc度量的基础上,继续引入不平衡比 R (imbalance ratio)来衡量规则中两个项集之间的不平衡性^[28]。在变压器重过载故障的关联规则挖掘中,可以使用 R 来度量过载因子和故障结果之间的不平衡性。一条关联规则 $X \rightarrow Y$ 的不平衡比 $R(X \rightarrow Y)$ 公式为

$$R(X \rightarrow Y) = \frac{c(X \rightarrow Y)}{c(Y \rightarrow X)} = \frac{B_f(u(X) - u(Y))}{u(X) + u(Y) - u(X \cap Y)} \quad (5)$$

$$\min u_s = \min u_0 \cdot \frac{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P(s))}{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P_{\max}(s))} \quad (6)$$

$$\min cf_s = \min cf_0 \quad (7)$$

$$\min K_s = \min K_0 \cdot \frac{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P(s); D_s(i, n+2) = Y(c))}{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P_{\max}(s); D_s(i, n+2) = Y(c))} \quad (8)$$

$$\min R_{h_i} = \min R_0 \cdot \frac{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P(s); D_s(i, n+2) = Y(c))}{B_f(n_i \in D_s(i, 1); D_s(i, n+3) = P_{\max}(s); D_s(i, n+2) = Y(c))} \quad (9)$$

式(6)~(10)中, $i = 2, 3, \dots, (m+1)$ 代表输入数据库 D_s 中的一行, $B_f()$ 代表 D_s 中同时满足所包含全部条件的故障记录的基数,下标0代表初始设定的关联状态阈值, $P(s)$ 表示春夏秋冬中的任一季度;其中, $s \in \{1, 2, 3, 4\}$, $P_{\max}(s)$ 表示变压器重过载频率最高的一个季度, $Y(c)$ 代表变压器重过载与否。

2.3 挖掘高风险罕见因子的重要度得分计算

由于罕见数据集中的高风险罕见(high risk and rare, HRR)因子一旦发生,将会造成较为严重的后果,故应进一步从罕见因子中挖掘HRR因子。因此,将一条关联规则 $X \rightarrow Y$ 进一步扩展为包含常见

2.2 考虑属性时段重要程度的关联阈值改进

当应用传统的关联规则挖掘(association rule mining, ARM)方法进行变压器重过载故障的预测时,存在一个重要问题,即忽视了罕见时段中过载因子的影响。当前的ARM算法在处理不同时间段出现的因子时,仍然采用相同且固定的关联状态阈值。换言之,无论是在不同季节还是在不同环境状态下,均采用相同的阈值进行评估和分析。这种做法导致了罕见时段中过载因子的重要度得分较低,最终可能低于1a中设定的关联状态阈值,从而使那些因子所在的罕见时段被直接排除在外。

因此,在变压器重过载故障的关联规则挖掘中,为了考虑属性时段的重要程度,需对阈值进行改进处理。将所收集的数据库数据按照合适的时间间隔划分为属性时段,本文定义1a中一个季度作为基准属性时段。再根据每个属性时段的重要程度设置关联状态阈值,只有达到或超过时段的阈值中才会被考虑在内,而低于时段的阈值将被排除。通过引入属性时段重要程度的阈值改进计算,使模型能够更有效地发现与变压器重过载故障相关的规律和趋势。

基于支持度、置信度、Kulc度量及不平衡比这4种重要度诊断标准,依据故障在各个季节中的分布情况,分别设定4种对应的属性时段重要程度阈值。相应的阈值设定表达式为

变量和罕见变量的形式,即

$$X^c + X^h \rightarrow Y \quad (10)$$

式中, X^c 、 X^h 分别为高频关联集和罕见数据集。

在传统的ARM模型中,当预测变压器重过载故障时,存在另一个重要问题:ARM模型在处理出现频率较少的罕见数据集时,仍然采用固定的重要度得分计算方法,这导致在数据挖掘过程中将直接筛选罕见因子。为此,需要依据不同环境特征下罕见因子的分布情况,计算出不同形式的罕见因子的得分,以便能够同时从罕见数据集中进一步挖掘出与目标强关联的HRR因子。本文对支持度 u 、置信度 c 、Kulc系数 k

和不平衡比 R 这 4 种关联指标的标准得分计算方法进行改进。当一条关联规则 $X^c + X^h \rightarrow Y$ 含有某一个

环境特征 t_j 中任意一个罕见环境元素时,该特征中罕见因子的关联指标标准得分计算公式为

$$S_{j(u)} = \frac{B_f(n_i \in D_s(i, 1); X^c D_s(i, M_d) \neq \emptyset; D_s(i, j) \in X^h \neq \emptyset)}{B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset)} \times 100\% \quad (11)$$

$$S_{j,c} = \frac{B_f(n_i \in D_s(i, 1); X^c D_s(i, M_d) \neq \emptyset; D_s(i, j) \in X^h \neq \emptyset; D_s(i, n+2) = Y(c))}{B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset)} \times 100\% \quad (12)$$

$$S_{j,k} = \frac{1}{2} \cdot \left(S_{j,c} + \frac{S_{j,c} \cdot B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset)}{B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset; D_s(i, n+2) = Y(c))} \right) \quad (13)$$

$$S_{j,R} = \frac{B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset; D_s(i, n+2) = Y(c))}{B_f(n_i \in D_s(i, 1); D_s(i, j) \in X^h \neq \emptyset)} \cdot S_{j,u} \quad (14)$$

2.4 TBFP-Growth 算法实施流程

TBFP-Growth 算法相较于 FP-Growth 算法,在提高算法性能和效率方面取得了显著进展。TBFP-Growth 算法采用启发式策略,选择条件模式基,避免了对所有频繁项的逐个扫描,减少了递归调用的次数。这一改进使得算法只需构建条件 FP 树,不仅提高了算法的效率,还降低了内存消耗。尤其对于大规模数据集和资源受限的环境下的挖掘任务而言,该算法表现出了更好的扩展性。通过减少递归构建条件模式基的次数,该算法能有效地提升挖掘过程的速度^[29]。TBFP-Growth 算法的具体流程如图 2 所示。

3 基于 CI 诊断的权重计算方法

为量化不同因子对配变重过载状态的影响程度,必须考虑高频关联集与罕见数据集对系统整体风险分别有不同的影响计算方式,基于 CI 计算方法,设计了一种针对每个因子在不同环境下的权重计算方法^[30]。

一个系统整体的风险通常取决于各个部件的风险。在此背景下,部件代表了不同特征下的因子。例如,在夏季中,温控设备的频繁使用导致负载率显著增加。此外,当环境温度较高时,变压器内部温度也会上升。因此,负载率和温度对于变压器过载风险具有较大的权重。但对于 HRR 因子而言,虽然其出现频率较低,但对系统造成的风险极高。因此,本文根据各个元素对系统的影响程度来确定其相对风险权重。

为区别分析 HRR 的相对风险权重,本文构建了高风险罕见变量子空间 X^{hh} : 包含具有某个特征 t_j 中任意 HRR 因子的事务集合。此外, t_j^c 和 t_j^h 分别为该特征中所有的高频关联因子与 HRR 因子。

本文中,某个因子 $c_{j,k} \in t_j$ 的相对风险权重 $\tau_{c_{j,k}}$ 由两部分组成,可写为

$$\tau_{c_{j,k}} = \tau_{j,k}^c + \tau_{j,k}^h \quad (15)$$

式中, $\tau_{j,k}^c$ 为高频关联因子所产生的风险; $\tau_{j,k}^h$ 为 HRR 因子所产生的风险。

高频关联因子所产生的风险计算公式如下:

$$\tau_{j,k}^c = \begin{cases} 0, & c_{j,k} \in t_j^h \\ \sum_{i=2}^{B_f(n_i \in D_s)} \frac{B_f(D_s(i, j) = c_{j,k})}{B_f(m)}, & c_{j,k} \in t_j^c \end{cases} \quad (16)$$

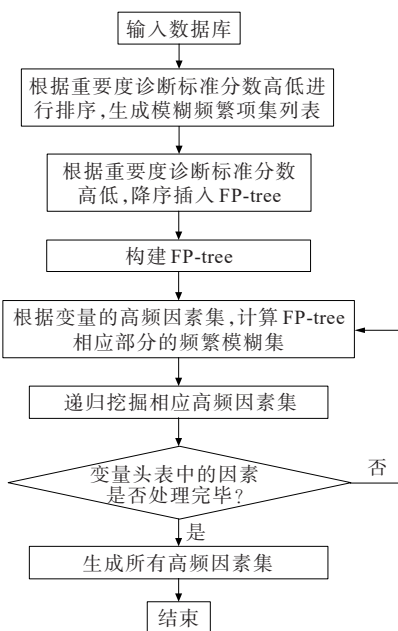


图 2 TBFP-Growth 算法流程

Figure 2 Flowchart of TBFP-Growth algorithm

式中, $B_f(m)$ 代表 D_s 中所有故障记录的基数。

对 HRR 来说, 如何识别和处理最重要的组件是关键。为此, 通过分析组件的 CI 来量化高风险罕见因子对系统潜在的影响。

在组件重要性度量中, CI 是一种衡量组件对系统整体功能和性能的重要程度的指标。其中, $I^{CR}(c_{j,k})$ 度量了当系统发生故障时, 输入因子 $c_{j,k}$ 对系统故障的贡献程度。其表达式为

$$I^{CR}(c_{j,k}) = \frac{I^B(R_i(n))}{h(x_i R(n))} \quad (17)$$

式中, $R_i(n)$ 为部件发生的风险; $h(x_i R(n))$ 为系统整体故障的风险; x_i 为 0-1 度量, $h(1, R(n))$ 为当因

$$h(R(n)) = \prod_{j=2}^{n+1} \left[1 - \prod_{k=1}^l \left(\sum_{i=2}^{B_f(n_i \in X^{th})} \frac{B_f(n_i \in D_s(i, 1); D_s(i, j) = c_{j,k})}{B_f(t_i \in D_s(i, 1); D_s(i, j) \in t_j)} \right) \right] \quad (19)$$

因此, 来自 HRR 因子的风险可以表示为

$$I^{CR}(c_{j,k}) = \frac{(h(1_i, R(n)) - h(0_i, R(n))) \cdot \sum_{i=2}^{B_f(n_i \in X^{th})} \frac{B_f(n_i \in D_s(i, 1); D_s(i, j) = c_{j,k})}{B_f(t_i \in D_s(i, 1); D_s(i, j) \in t_j)}}{\prod_{j=2}^{n+1} \left[1 - \prod_{k=1}^l \left(\sum_{i=2}^{B_f(n_i \in X^{th})} \frac{B_f(n_i \in D_s(i, 1); D_s(i, j) = c_{j,k})}{B_f(t_i \in D_s(i, 1); D_s(i, j) \in t_j)} \right) \right]} \quad (20)$$

4 模型实施流程与框架

通过这些方法, 本文建立 IAR-CI 分类器模型, 其具体实施步骤如下:

1) 根据配变输入特征收集相应的数据, 从外部与内部特征值综合求解不同因子风险值;

2) 根据数据库的训练数据, 针对某个特征 t_j 所含有的全部特征因子通过各个时段的关联规则重要度得分计算, 并比较式(6)~(9)所计算的关联状态阈值挖掘其中的罕见数据集;

3) 含有该特征中任一罕见因子的故障记录被归类于罕见数据集 X_j^h 中, 并通过式(11)~(14)挖掘基于这些罕见数据集的 HRR 因子, 用以刻画对配变状态的潜在影响;

4) 对于训练数据集中的每一个环境特征, 依次重复进行步骤 1)~3);

5) 由文 3 中式(17)求解各个因子对故障发生的相对风险值;

6) 基于步骤 1)~5), 对训练数据集的每一条故障记录所含有的特征因子计算部件风险值并累加获得该条故障记录风险值, 最后进行归一化处理(0→

子 $c_{j,k}$ 确定有关联, 即该部件因子风险为 1 时, 系统整体发生故障的风险; $h(0, p(n))$ 为当因子 $c_{j,k}$ 确定无关联, 即该部件因子风险为 0 时, 系统整体发生故障的风险。

在式(17)中, I^B 为度量重要性的 Birnbaum 重要度, 其描述了组件的可靠性变化对系统可靠性的影响程度^[31], 其数学表达式为

$$\tau_{c_{j,k}} = \tau_{j,k}^c + \tau_{j,k}^h \quad (18)$$

在配电变压器系统中, 一条故障记录的发生由不同特征下的唯一元素组成, 而在已知的故障记录中, 缺少任何一个特征条件, 其故障都不会发生。基于此, 假设每个特征因子之间相互独立, 则系统整体故障的风险可以写为

1: 不可能发生→一定发生);

7) 将预测故障风险度与测试集中相应记录的真实过载记录相比较(0 或 1: 配变重过载发生或未发生), 从而验证本文预测模型的性能。

结合以上步骤, 本章构建的 IAR-CI 分类器模型实施流程如图 3 所示。

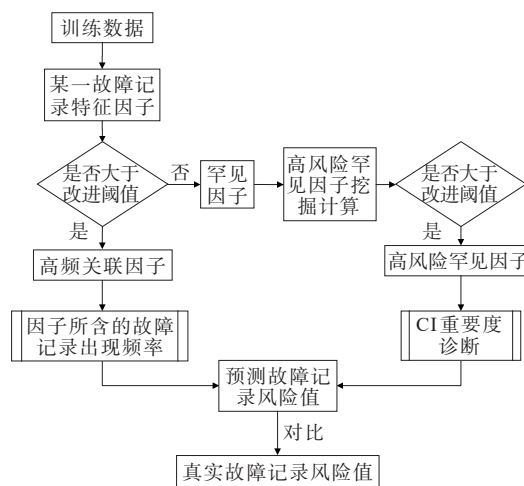


图 3 IAR-CI 分类器模型实施流程

Figure 3 Implementation flowchart of IAR-CI classifier model

5 算例分析

5.1 实验数据及验证方法

本文以中国南方某地区 2018 年 2 月至 2020 年 3 月的 30 511 条实际历史纪录为例,收集总共 21 个特征,对本文所提出的方法进行精度验证。其中,仿真采用十折交叉验证方法,以 70% 历史记录作为训练集,10% 为验证集,20% 为测试集。

为了评估本章中模型的性能,基于混淆矩阵中真正例率、假正例率、查全率等度量指标,绘制 ROC^[32] (receiver operating characteristic curve) 和 PR^[33] 曲线 (precision-recall curve)。一方面,显示模型在不同分类阈值下的整体分类能力,对模型的敏感性和特异性进行权衡;另一方面,描述模型当正负样本的分布不均衡时所诊断的性能。除此之外,在这两类模型评估曲线的基础上,采用线下包围面积 (area under the curve, AUC) 为评判模型优劣的检验指标,某模型 AUC 的 I_{AUC} 数值越高,该模型预测效果越优。

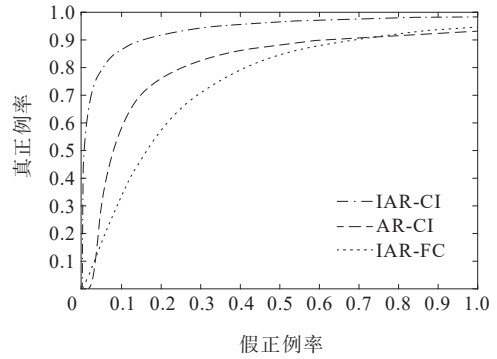
5.2 配变过载诊断内部对比分析

为验证所提出 IAR-CI 分类器模型的准确性,加入 2 种内部模型作为对比。其中,一种是采用传统的关联规则并使用 CI 重要度计算权重的 AR-CI (association rules-criticality importance) 分类器模型,另一种是采用无关联规则和频率计算权重 AR-FC (association rules-frequency calculation) 的分类器模型。这 3 种分类器模型的 PR 曲线及 ROC 曲线对比如图 4 和表 1 所示。

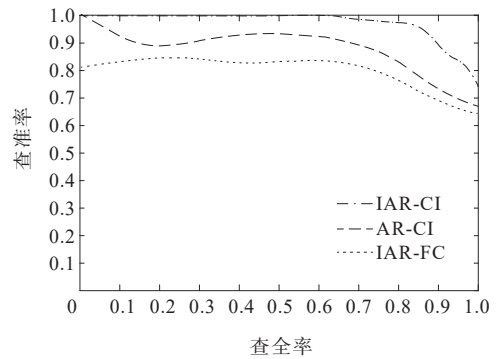
由图 4 和表 1 可知,IAR-CI 分类器模型相较其他 2 个模型分别在 ROC 和 PR 曲线上都有较大 AUC 数值,特别是本方法 PR 下的 AUC 极高。这是因为该模型考虑了数据集中不同的特征在具体应用场景中不同的分布和高风险罕见因子所带来的影响。因此,采用基于改进关联规则和 CI 重要度风险权重计算方法的模型能实现更为精确的配电变压器重过载诊断。

表 1 内部 I_{AUC} 值对比

模型	$I_{AUC}(\text{ROC})$	$I_{AUC}(\text{PR})$
IAR-CI	0.931	0.942
AR-CI	0.834	0.821
AR-FC	0.762	0.735



(a) ROC 曲线



(b) PR 曲线

图 4 内部测试效果对比

Figure 4 Effect comparison of internal test

最后,为了改善模型的运行速度,本文统一采用 IAR-CI 分类器模型,并采用不同的算法来对比,结果如图 5 所示。

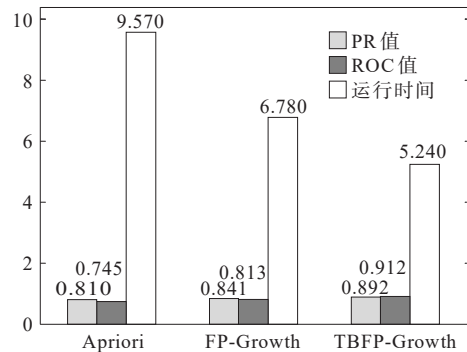


图 5 算法测试对比

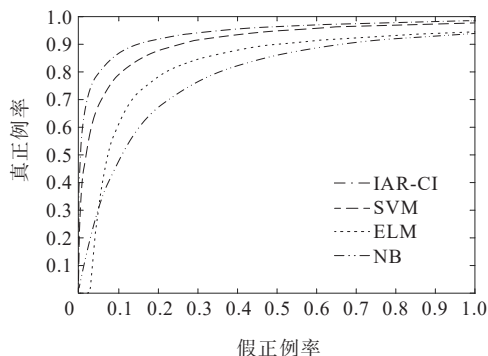
Figure 5 Algorithm test comparison

从图 5 中可看出,在 IAR-CI 分类器模型的基础上,TBFP-Growth 算法的运行时间比 FP-Growth 和 Apriori 算法的分别提升了 22.71% 和 45.24%,且在精度上也有明显的提升。因此,基于 TBFP-Growth 算法的 IAR-CI 分类器模型可以更好地反映输入特征与输出间的映射关系,具有更高的分类准确率和更快的运行时间,并且适用于偏置

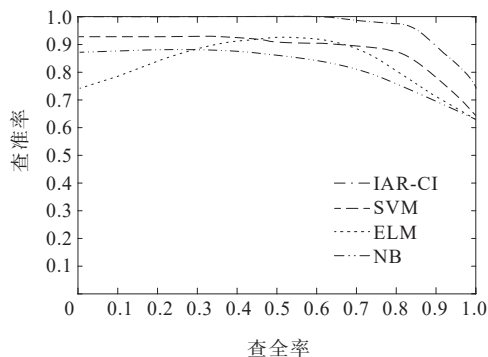
数据环境,能有效地对配变重过载预测起指导作用。

5.3 配变过载诊断外部对比分析

同时,为了验证该方法相较其他机器学习方法的优越性,对比目前常见且用得最多的分类器模型SVM、极限学习机(extreme learning machine, ELM)、朴素贝叶斯算法(naive bayes, NB),并搭建配变重过载模型,采用十折交叉验证方法,绘制各分类器的ROC与PR曲线,并计算各模型AUC值,结果分别如图6与表2所示。



(a) ROC曲线



(b) PR曲线

图6 外部测试效果对比

Figure 6 Effect comparison of external test

表2 外部 I_{AUC} 值对比

Table 2 Comparison of external I_{AUC} values

模型	I_{AUC} (ROC)	I_{AUC} (PR)
IAR-CI	0.931	0.942
SVM	0.834	0.814
ELM	0.812	0.801
NB	0.805	0.782

由图6与表2可知,本文所提出的IAR-CI模型具有最高的故障诊断准确率。其中,ROC与PR的AUC值分别为0.931和0.942,而其他分类器模型判断准确率较低,且PR曲线下的 I_{AUC} 值大致相同。这

从侧面也可以反映出,当面对高维数据且数据分布不均衡的情况下,该方法相较现有普通机器学习方法也能获得更满意的效果。

6 结语

由于目前配电变压器风险预测方法大多未考虑数据分布不平衡所带来的影响,为此本文提出了一种基于IAR-CI模型的配变过载风险诊断方法,对该模型进行训练与测试,结论如下:

1) 根据内部因素与外部因素,建立配变状态特征库,结合关联规则特定指标评价原则,在时间维度上,设置属性时段重要程度的关联状态阈值,以保证高频关联特征因子和罕见特征因子的合理划分;在空间维度上,设置特征因子的重要度得分,从而挖掘识别HRR,全面考虑配电变压器过载事件发生的各种重要因子;

2) 基于CI重要度权重计算方法,解决高风险罕见因子对系统影响程度的量化问题,通过算例分析,本文设计的不同因子划分权重计算方法得到的权重相较基于固定因子出现频率的线性权重具有更高的准确率和合理性;

3) 采用与传统Apriori与FP-Growth算法不同的TBFP-Growth算法进行关联规则挖掘。通过算例分析,该算法能有效地提升挖掘过程的速度并降低内存消耗。

4) 实验结果表明,本文所提出的变压器重过载预测方法能够在数据偏置的环境下准确地监测和预测重过载情况,并及时预警和防范潜在的设备损坏和停电事故,可为相应设备的运维检修提供参考,继而助力系统稳定运行和供电可靠性。

参考文献:

- [1] 王艳,李伟,赵洪山,等.基于油中溶解气体分析的DBN-SSAELM变压器故障诊断方法[J].电力系统保护与控制,2023,51(4):32-42.
WANG Yan, LI Wei, ZHAO Hongshan, et al. Transformer DGA fault diagnosis method based on DBN-SSAELM [J]. Power System Protection and Control, 2023, 51(4): 32-42.
- [2] 刘志坚,何蔚,刘航,等.基于格拉姆角场变换和深度压缩模型的变压器故障识别方法[J].电网技术,2023,47(4):1478-1490.
LIU Zhijian, HE Wei, LIU Hang, et al. Fault identification method for power transformer based on gramian angular field transformation and deep compression model[J].

- Power System Technology,2023,47(4):1478-1490.
- [4] 林少娃,陈奕汝,顾洁,等.基于隐含狄利克雷分布主题模型和特征级异构数据融合的电力故障主动性预警研究[J].电子器件,2022,45(2):432-438.
LIN Shaowa, CHEN Yiru, GU Jie, et al. Proactive warning system based on electronic power user interactive complaint text and multi-source heterogeneous big data analysis[J]. Chinese Journal of Electron Devices, 2022, 45 (2):432-438.
- [5] LI M, ZHOU Q. Distribution transformer mid-term heavy load and overload pre-warning based on logistic regression[C]//2015 IEEE Eindhoven PowerTech. Eindhoven, Netherlands. IEEE, 2015:1-5. [LinkOut]
- [6] WU Q, CHEN Z, SU H T, et al. Heavy overload forecasting of distribution transformers based on XGBoost[C]//2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). Macao, China. IEEE, 2019:1-5.
- [7] 陈毅波,郑玲,姚建刚.基于粗糙集理论与D-S证据理论改进的多元回归负荷预测方法研究[J].电力系统保护与控制,2016,44(6):62-67.
CHEN Yibo, ZHENG Ling, YAO Jiangang. Improved multiple regression load forecasting method based on rough set theory and D-S evidence theory[J]. Power System Protection and Control, 2016, 44(6):62-67.
- [8] 胡剑,王建,熊小伏,等.计及线路动态电热特性的交直流混联电网过载控制策略[J].电力系统保护与控制,2020,48(7):66-75.
HU Jian, WANG Jian, XIONG Xiaofu, et al. An overload control strategy for AC/DC hybrid power grid considering dynamic electro-thermal characteristics of transmission lines[J]. Power System Protection and Control, 2020, 48(7):66-75.
- [9] 郝钰,石华林,范瑞祥,等.基于动态热路模型的配变变压器过载工况下温升研究[J].变压器,2020,57(12):22-26.
HAO Yu, SHI Hualin, FAN Ruixiang, et al. Study on temperature rise of distribution transformer under overload condition based on dynamic thermal circuit model[J]. Transformer, 2020, 57(12):22-26.
- [10] 李元,刘宁,梁钰,等.基于温升特性的油浸式变压器负荷能力评估模型[J].中国电机工程学报,2018,38(22):6737-6746.
LI Yuan, LIU Ning, LIANG Yu, et al. A model of load capacity assessment for oil-immersed transformer by using temperature rise characteristics[J]. Proceedings of the CSEE, 2018, 38(22):6737-6746.
- [11] 杜晓东,赵建利,刘科研,等.基于数字孪生的光伏高比例配电网过载风险预警方法[J].电力系统保护与控制,2022,50(9):136-144.
DU Xiaodong, ZHAO Jianli, LIU Keyan, et al. Digital twin early warning method study for overload risk of distribution network with a high proportion of photovoltaic access[J]. Power System Protection and Control, 2022, 50(9):136-144.
- [12] WU Y H, SUN X B, DAI B F, et al. A transformer fault diagnosis method based on hybrid improved grey wolf optimization and least squares-support vector machine [J]. IET Generation, Transmission & Distribution, 2022, 16 (10):1950-1963.
- [13] 夏正龙,陆良帅,吴启凡,等.改进灰狼算法在含风电的配电网无功优化中的应用[J].智慧电力,2023,51(6):63-70.
XIA Zhenglong, LU Liangshuai, WU Qifan, et al. Application of improved grey wolf in reactive power optimization of distribution networks with wind power integration[J]. Smart Power, 2023, 51(6):63-70.
- [14] 李云溟,咸日常,张海强,等.基于改进灰狼算法与最小二乘支持向量机耦合的电力变压器故障诊断方法[J].电网技术,2023,47(4):1470-1478. LI Yunhao, XIAN Richang, ZHANG Haiqiang, et al. Fault diagnosis for power transformers based on improved grey wolf algorithm coupled with least squares support vector machine[J]. Power System Technology, 2023, 47(4):1470-1478.
- [15] 郭方洪,刘师硕,吴祥,等.基于联邦学习的含不平衡样本数据电力变压器故障诊断[J].电力系统自动化,2023,47(10):145-152.
GUO Fanghong, LIU Shishuo, WU Xiang, et al. Federated learning based fault diagnosis of power transformer with unbalanced sample data[J]. Automation of Electric Power Systems, 2023, 47(10):145-152.
- [16] 周秀,怡恺,李刚,等.基于邻域粗糙集与AMPOS-ELM的变压器DGA故障诊断[J].电力科学与技术学报,2022,37(3):157-164.
ZHOU Xiu, YI Kai, LI Gang, et al. A transformer DGA fault diagnosis approach based on neighborhood rough set and AMPPO-ELM[J]. Journal of Electric Power Science and Technology, 2022, 37(3):157-164.
- [17] LUO S M, RAO Y, CHEN J, et al. Short-term load forecasting model of distribution transformer based on CNN and LSTM[C]//2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE). Beijing, China. IEEE, 2020:1-4.
- [18] 朱保军,咸日常,范慧芳,等.WRSR与改进朴素贝叶斯融合的变压器故障诊断技术研究[J].电力系统保护与控制,2021,49(20):120-128.
ZHU Baojun, XIAN Richang, FAN Huifang, et al. Transformer fault diagnosis technology based on the fusion of WRSR and improved naive Bayes[J]. Power System Protection and Control, 2021, 49(20):120-128.
- [19] 邱海枫,苏宁,田松林.改进支持向量机在电力变压器故

- 障诊断中的应用研究[J]. 电测与仪表, 2022, 59(11): 48-53.
- QIU Haifeng, SU Ning, TIAN Songlin. Research on the application of improved support vector machine in power transformer fault diagnosis[J]. *Electrical Measurement & Instrumentation*, 2022, 59(11): 48-53.
- [20] 何宁辉, 朱洪波, 李秀广, 等. 基于贝叶斯网络和假设检验的变压器故障诊断[J]. *电力科学与技术学报*, 2021, 36(6): 20-27.
- HE Ninghui, ZHU Hongbo, LI Xiuguang, et al. Transformer fault diagnosis based on Bayesian network and hypothesis testing[J]. *Journal of Electric Power Science and Technology*, 2021, 36(6): 20-27.
- [21] XIAO H, PEI W, WU L, et al. A novel deep learning based probabilistic power flow method for Multi-Microgrids distribution system with incomplete network information [J]. *Applied Energy*, 2023, 335: 120716.
- [22] 张朝龙, 何怡刚, 杜博伦, 等. 基于深度学习的电力变压器智能故障诊断方法[J]. *电子测量与仪器学报*, 2020, 34(1): 81-89.
- ZHANG Chaolong, HE Yigang, DU Bolun, et al. Intelligent fault diagnosis method of power transformer using deep learning[J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 34(1): 81-89.
- [23] 马旭聪, 唐文虎, 牛哲文, 等. 非均衡数据集下基于孪生卷积网络的变压器绕组变形故障识别方法[J]. *高压电器*, 2023, 59(10): 120-128.
- MA Xucong, TANG Wenhui, NIU Zhewen, et al. Deformation fault identification method for transformer windings based on twin convolutional network under unbalanced data set[J]. *High Voltage Apparatus*, 2023, 59(10): 120-128.
- [24] 童光华, 董亮, 任永平, 等. 基于DBN和K-means聚类的配变重过载预警方法[J]. *现代电力*, 2021, 38(5): 492-501.
- TONG Guanghua, DONG Liang, REN Yongping, et al. Overload warning for distribution transformer based on DBN and K-means[J]. *Modern Electric Power*, 2021, 38(5): 492-501.
- [25] 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.
- [26] 宋继明, 毛继兵, 马卫华, 等. 基于深度神经网络的特高压变压器滤油注油过程故障诊断技术研究[J]. *电网与清洁能源*, 2023, 39(12): 95-103.
- SONG Jiming, MAO Jibing, MA Weihua, et al. Research on oil filtration and injection process fault diagnosis for ultrahigh voltage transformers based on deep neural networks[J]. *Power System and Clean Energy*, 2023, 39(12): 95-103.
- [27] 余飞娅, 叶文波. 基于FP-Growth算法的计量主站告警分析研究[J]. *电气自动化*, 2021, 43(6): 30-32+35.
- YU Feiya, YE Wenbo. Analysis and research on alarm of master Metering Station based on FP-growth algorithm [J]. *Electrical Automation*, 2021, 43(6): 30-32+35.
- [28] 程江洲, 聂玮瑶, 张赟宁, 等. 基于FP-network关联规则挖掘算法的配电网薄弱点分析研究[J]. *电测与仪表*, 2021, 58(3): 47-53.
- CHENG Jiangzhou, NIE Weiyao, ZHANG Yunning, et al. Analysis and research on weak point of distribution network based on FP-network association rule mining algorithm[J]. *Electrical Measurement & Instrumentation*, 2021, 58(3): 47-53.
- [29] SINTHUJA M, PUVIARASAN N, ARUNA P. Mining frequent itemsets using proposed top-down approach based on linear prefix tree (TD-LP-growth)[M]// *Lecture Notes on Data Engineering and Communications Technologies*. Singapore: Springer Singapore, 2018: 23-32.
- [30] ESPINOZA S, POULOS A, RUDNICK H, et al. Risk and resilience assessment with component criticality ranking of electric power systems subject to earthquakes[J]. *IEEE Systems Journal*, 2020, 14(2): 2837-2848.
- [31] MIZIULA P, NAVARRO J. Birnbaum importance measure for reliability systems with dependent components[J]. *IEEE Transactions on Reliability*, 2019, 68(2): 439-450.
- [32] 张立石, 梁得亮, 刘桦, 等. 基于小波变换与逻辑斯蒂回归的混合式配电变压器故障辨识[J]. *电工技术学报*, 2021, 36(增刊2): 467-476.
- ZHANG Lishi, LIANG Deliang, LIU Hua, et al. Fault identification of hybrid distribution transformer based on wavelet transform and logistic regression[J]. *Transactions of China Electrotechnical Society*, 2021, 36(Sup 2): 467-476.
- [33] 裴小邓, 罗林, 陈帅, 等. 面向电力变压器油中溶解气体的卷积神经网络诊断方法[J]. *辽宁石油化工大学学报*, 2020, 40(5): 79-85.
- PEI Xiaodeng, LUO Lin, CHEN Shuai, et al. A convolutional neural network diagnosis method for dissolved gas in power transformer oil[J]. *Journal of Liaoning Shihua University*, 2020, 40(5): 79-85.