



基于关联规则与重构误差的二次系统故障检测方法

王阳¹, 马伟东¹, 刘泊溟², 王博石², 姚凯³, 韩伟¹, 余娟²

(1. 国网河南省电力公司电力科学研究院, 河南 郑州 450003; 2. 输变电装备技术全国重点实验室(重庆大学), 重庆 400044; 3. 国网河南省电力公司, 河南 郑州 450003)

摘要: 二次系统是否可靠直接关系整个变电站乃至系统能否安全可靠运行。随着高比例新能源并网, 如何有效检测二次系统故障愈发重要。针对现有逻辑回路的故障方法对数据完备性要求较高而难以实际应用, 现有二次设备的故障检测方法难以辨识正常数据和故障数据的微小差异导致计算精度无法保障的问题, 提出基于关联规则与重构误差的二次系统故障检测方法。首先, 利用 Apriori 算法求出故障报警信息与逻辑回路中故障装置的关联规则, 实现逻辑回路故障快速诊断; 然后, 利用正常二次设备的运行数据训练个体判别器, 通过衡量待判别数据的重构误差来判别二次设备运行状态, 并利用集成学习模型量化设备当前故障检测概率; 最后, 对集成学习模型进行集成优化, 以提高二次设备异常预警的可信度。利用河南省某变电站实际运行数据集进行仿真测试, 验证了所提方法的有效性和准确性。

关键词: 二次系统; 故障检测; 数据驱动; 关联规则; 重构误差

DOI: 10.11930/j.issn.1004-9649.202308131

0 引言

二次系统是保障电网安全稳定运行的重要防线^[1]。国家能源局在《电力二次系统安全管理若干规定》中指出电力企业及相关电力用户应建立二次系统安全双重预防体系, 加强二次系统安全风险管理和隐患排查治理^[2]。随着新能源大规模开发、高比例接入电网, 电力二次系统的种类也不断增加, 不同设备、信息系统之间的交互变得复杂, 其安全面临前所未有的挑战^[3]。

解决二次系统故障检测问题的主流方法分为 3 类: 基于电力系统企业的二次系统评价导则的故障检测方法、由一次系统演化而来的二次系统状态评价模型、数据驱动的二次系统安全预警方法^[4]。其中, 电力系统企业的二次系统评价导则分别从投运前资料完备性、历史故障记录、检修记录和实时监测记录等方面选取故障检测信息, 其对设备状态的影响程度均由专家系统判定。该

类方法简单实用, 但准确性不足^[5]。由一次系统演化而来的二次系统状态评价模型在一定程度上避免了人为主观因素对评价结果的影响, 但其所得平均故障时间、元件可用度等指标不能完全反映设备的状态^[6]。数据驱动的二次系统安全预警方法主要通过对二次系统历史数据进行建模, 进而实现二次系统故障检测、异常监测, 但不同设备在历史数据、评估指标上会有较大差异, 需要针对不同设备的特性分别进行特征分析, 因此现有数据驱动方法通用性较差, 难以推广应用^[7]。

二次系统故障检测的研究对象主要是逻辑回路和二次设备。随着智能二次设备的兴起, 大量反映逻辑回路与二次设备状态的信息涌现^[8]。针对逻辑回路故障检测问题, 文献^[9-10]从构建故障树的角度评估逻辑回路当前可靠性并给出底事件故障分析, 该方法须遍历各个设备, 并根据可疑设备的报警信息计算其故障概率, 该类方法对数据的完备性要求较高且计算复杂性高, 难以实际应用。针对基于历史数据分析的二次设备故障检测问题, 文献^[11]采用粗糙集理论建立了智能变电站二次设备的不完备评估体系, 并建立二次设备不同劣化等级下隶属度函数值。文献^[12]使用模糊综合评价法进行智能变电站二次系统状态

收稿日期: 2023-08-30; **修回日期:** 2023-12-19。

基金项目: 国网河南省电力公司电力科学研究院科技项目(基于全寿命周期的变电站二次系统可靠性评价技术研究, SGHADK00DWJS2200241)。



评价。上述研究设定评估等级比较粗糙且对应的隶属度函数完全依赖专家经验，对二次设备可靠性量化的科学性不足，计算精度难以满足工程实际需求。而相比之下，Apriori 算法可以充分利用二次系统报文信息，挖掘逻辑回路和报文告警信息之间的潜在关联关系，提取出频繁出现的关联规则^[13-14]，有效辅助分析人员从复杂告警报文信息中摒弃正常操作的告警，锁定故障二次回路，然而 Apriori 算法仅能筛选出多个可能的故障装置，而如何进一步精准定位故障设备还未见相关报道。基于重构误差的异常检测方法则能充分利用二次设备运行状态信息，精准量化正常设备与故障设备间运行状态的微小差异，实现故障设备的精准可信定位，然而针对每个设备都基于重构误差搭建一个故障检测模型存在计算负担高、数据需求过大的问题，如何针对单个故障检测模型输出具有不确定性导致结果可信性不足的问题也有待研究。

针对上述问题，本文提出基于关联规则与重构误差的二次系统故障检测方法。

1 海量报文信息的关联规则提取方法

基于 Apriori 算法的二次系统逻辑回路故障检测方法本质上是利用大量报文信息数据集，挖掘其中的关联性，描述报文信息和二次系统故障之间的关联关系^[15]。二次系统逻辑回路和二次设备的典型故障主要包括电源插件故障、CPU 插件故障、开入开出插件故障、采样插件故障、通信插件故障、模数转换插件故障、通道接口插件故障、二次回路故障（含采样回路和控制回路故障）等。本文从河南省某智能变电站的报文信息中抽取的部分典型示例数据如表 1 所示。由表 1 可以发现，某一装置发生故障不是只有一个简单的告警信息，而是有多条告警信息共同出现，同时针对同一故障原因其报警信息可能不同，且具有相似报警信息的故障原因也不同。因此，不能简单地通过报警信息查看就确定故障装置或故障原因，需通过有效的数据分析手段挖掘告警信息和故障原因间的关联规则，辅助监控人员从海量报警信息中精准定位告警原因。对于关联分析，通常定义所含元素各不相同的集合为项集，其中

表 1 报文信息数据集示例

Table 1 Example of message information dataset

| 告警信息 | 故障原因 |
|-------------------------------|---------|
| 线路1运行灯熄灭、线路1保护装置失电告警、装置电源空开跳开 | 二次回路短路 |
| 保护装置告警灯无法复归、保护自检异常、参数自检异常 | CPU插件损坏 |
| 保护装置告警断线 | CPU插件损坏 |
| 保护板、采样板异常 | 采样插件损坏 |
| 公用测控稳控装置异常、现场检查装置闭锁、运行灯灭 | 电源插件损坏 |
| 运行灯熄灭、线路保护装置失电告警、液晶面板灭 | 电源插件损坏 |

元素个数称为项集的长度，记作 k 。该项集称为 k 项集。用以关联分析的样本称为样本集，样本集是项集的子集。用来衡量关联分析结果好坏的指标包括支持度、置信度以及提升度。支持度 $s(X \rightarrow Y)$ 表示项集 $\{X, Y\}$ 在总集 $\{I\}$ 出现的概率，可表示为

$$s(X \rightarrow Y) = P(X \cup Y) / P(I) \quad (1)$$

式中： $P(X \cup Y)$ 为 X 和 Y 发生的概率； $P(I)$ 为总集中任意事件发生的概率。

置信度 $c(X \rightarrow Y)$ 是指关联规则的可信程度，表示 Y 在先决条件 X 发生的情况下发生的概率 $P(Y|X)$ ，即

$$c(X \rightarrow Y) = P(Y|X) = P(Y \cup X) / P(Y) \quad (2)$$

式中： $P(Y)$ 为 Y 发生的概率。

提升度为

$$l(X \rightarrow Y) = P(Y|X) / P(Y) \quad (3)$$

频繁项集指支持度、置信度以及提升度达到阈值的关联规则集合，对于二次系统的报文信息逐次通过支持度、置信度以及提升度的阈值进行筛选，直到不能再找出频繁 k 项集为止^[16-17]。第 k 次迭代找出频繁项集 L_k 均需要一次完整的数据库扫描。

上述过程的具体操作如下。1) 算法扫描所有二次系统的故障信息以及同时段（先于故障前）的其他报文信息统计包含各告警信息（亦包含故障装置）出现的频次，形成 C_1 。2) 对每个项最小支持度等指标进行计算，随后依据上述指标删除不满足的告警信息与故障装置，至此获得频繁 L_1 项集。3) 将 L_1 中各项两两组合，此时每个项集中包含两项告警信息（亦包含故障装置）得到



C_2 ，重新计算新项集中各项的统计指标，从 C_2 中删除不满足要求的项，获得频繁 L_2 项集。4) 对 C_1 项集与 C_2 项集重复步骤 2~3 生成 L_k 。

根据其结果选取出一条重要的关联规则，其中包含高支持度、置信度以及提升度的告警信息及故障装置信息。工作的场站人员在得到重要关联规则之后即可作为参考依据，即每当有多条告警信息同时出现时，如果包含在某一条关联规则中，且故障还并未出现，可以提前对关联规则所映射的潜在故障装置进行排查，减小损失。

2 基于重构误差的二次设备故障检测方法

2.1 二次设备故障检测集成框架

为了应对个体判别器输出的不确定性，提高二次设备故障检测结果的稳定性和可信性，基于 bagging 集成学习方法构建了二次设备报警框架，如图 1 所示。使用不同的正常二次装置样本集训练得到多个个体判别器，综合所有个体判别器预测判别结果，得到二次设备故障的可信概率 P 为

$$P = \sum_{k=1}^n y_k / n \quad (4)$$

式中： y_k 为第 k 个判别器对该二次设备的判断结果，取值为 0 或者 1； n 为图 1 框架中个体判别器的总个数。

图 1 中各个体判别器的模型架构相同，但其训练样本集分别是不同的正常二次设备数据。个体判别器参数受输入样本集的影响较大，通过集

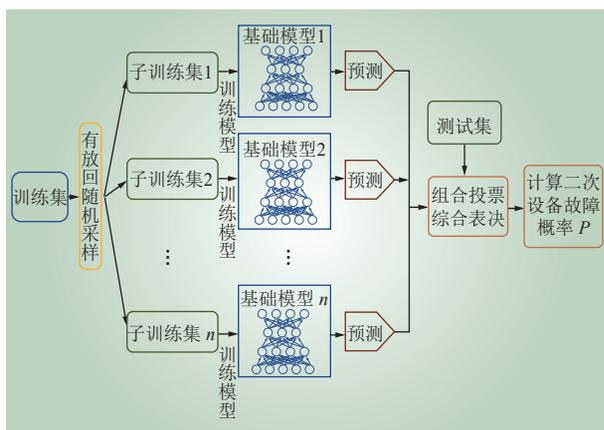


图 1 基于集成学习的二次设备故障检测框架

Fig. 1 Framework for secondary equipment fault detection based on ensemble learning

成学习综合多个个体判别器的判定结果来进行二次设备故障判定，可以提高模型整体的稳定性^[18]。

集成框架的预警精度取决于个体判别器的特征挖掘能力^[19]。因此，为提高集成框架的预警精度，本文基于无监督思想，利用重构误差计算形成个体判别器。无监督学习的核心思想是通过数据进行分类，挖掘无标记数据集中的潜在规律^[20]。其中，重构误差是无监督学习方法进行分类的重要依据，在实际工程中，潜在故障二次设备数据与正常设备数据存在差异，因此其重构误差较大。通过计算原始数据与重构数据的差值，即得重构误差。本文利用重构误差测算二次设备潜在故障风险，其流程如图 2 所示。用正常二次设备运行数据作为训练样本，采用 Adam 优化算法训练个体判别器。正常的二次设备模型输入向量记作 $\mathbf{X}_{in} = [X_1, X_2, \dots, X_n]$ ，包含待检测二次设备的数据特征，具体输入维度由待检测二次设备测量特征参数确定。各向量均包含各时刻下设备的运行状态，是一个时序向量，如 $\mathbf{X}_1 = [x_{1,1}, \dots, x_{1,t}, \dots, x_{1,T}]$ ，其中 $x_{1,t}$ 表示该时间序列中 t 时刻下的某二次设备特征值， T 为总的观测时长。值得注意的是，为了避免出现各特征值之间属性不一、数值范围差异较大等情况，上述二次设备特征变量均为归一化后的数据。输出特征向量的标签值即为输入特征，输出值（即输入特征的重构向量）记作 $\hat{\mathbf{X}}_{out} = [\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n]$ ，个体判别器的输入、输出特征向量的维数一致。模型训练过程优化目标为

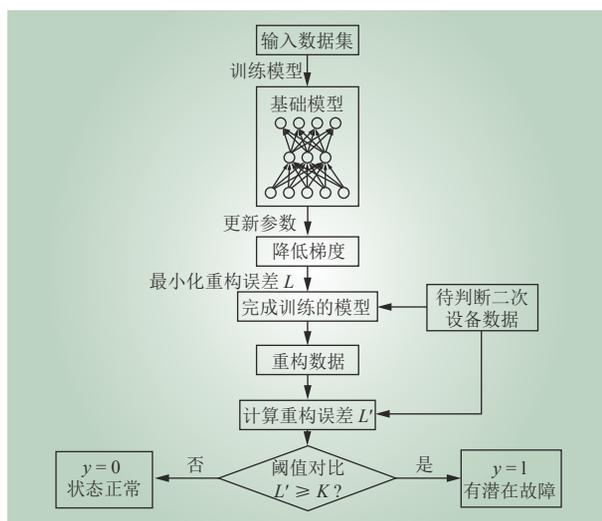


图 2 个体判别器判别流程

Fig. 2 Individual discriminator discrimination process



最小化输入特征向量的重构误差，可表示为

$$\min L(\mathbf{X}_{in}, \tilde{\mathbf{X}}_{out}) = \frac{1}{m} \sum_{t=1}^T \sum_{i=1}^m (x_{in,i,t} - x_{out,i,t})^2 \quad (5)$$

式中： L 为输入 \mathbf{X}_{in} 与输出 $\tilde{\mathbf{X}}_{out}$ 的均方差，即重构误差； m 为 \mathbf{X}_{in} 和 $\tilde{\mathbf{X}}_{out}$ 的维度； $x_{in,i,t}$ 和 $x_{out,i,t}$ 分别为 \mathbf{X}_{in} 、 $\tilde{\mathbf{X}}_{out}$ 中第 i 维度 t 时刻的特征值。

上述个体判别器训练的过程可视为将正常二次设备数据特征内蕴于重构误差模型的过程。待检测二次设备数据输入个体判别器计算后得到新的重构误差，判别规则为

$$y = \begin{cases} 1, & L'(\mathbf{X}'_{in}, \tilde{\mathbf{X}}'_{out}) \geq \sigma \\ 0, & L'(\mathbf{X}'_{in}, \tilde{\mathbf{X}}'_{out}) < \sigma \end{cases} \quad (6)$$

式中： \mathbf{X}'_{in} 表示某一待判断的二次设备的输入； $\tilde{\mathbf{X}}'_{out}$ 表示某个待判定对象二次设备输出； y 表示对应判断二次装置的个体判别器的判断结果； σ 表示潜在故障的判定阈值。通过完成模型计算待判定设备样本数据的重构误差，若该重构误差大于阈值，则该二次设备判断为潜在故障，赋值为1；若该重构误差小于设定的阈值，则该二次设备判断为正常，赋值为0。

2.2 二次设备故障检测集成优化方法

上节所提二次设备故障检测集成框架还不能自适应确定选择哪些个体判别器。为确定组成最佳集成模型的个体判别器，首先要明确个体判别器的准确率、个体判别器多样性与集成模型准确率的关系。文献[21]的误差-分歧理论将集成模型的误差分解到个体判别器的特征上，具体描述为：假设使用一组个体判别器 h_1, \dots, h_M 以形成集成模型，其中， M 为个体判别器数量。通过加权平均获得输出为

$$H(x) = \sum_{i=1}^M w_i h_i(x) \quad (7)$$

式中： $H(x)$ 表示集成模型的输出； w_i 为判别器 h_i 的权重，且由 $w_i \geq 0$ 与 $\sum_{i=1}^M w_i = 1$ 约束。

初始时，将所有个体判别器的权重都设置为 $w_i = 1/M$ ，随着个体判别器数量变化，各个个体判别器的权重相应变化。在后面的优化过程中，部分个体判别器会被剔除，其权重将会变为0。

给定样本 x ，则个体判别器的分歧可以定义为

$$A(h_i|x) = (h_i(x) - H(x))^2 \quad (8)$$

集成模型的分歧可定义为个体判别器分歧的加权平均，即

$$\bar{A}(h|x) = \sum_{i=1}^M w_i (h_i(x) - H(x))^2 \quad (9)$$

假设样本 x 的真实情况为 $f(x)$ ，则个体判别器在整体样本的加权平均误差 $\bar{E}(h|x)$ 可以表示为

$$\bar{E}(h|x) = \sum_{i=1}^M w_i (f(x) - h_i(x))^2 \quad (10)$$

结合式(9)(10)，集成模型的误差 $E(H|x)$ 可表示为

$$E(H|x) = (f(x) - H(x))^2 = \bar{E}(h|x) - \bar{A}(h|x) \quad (11)$$

因误差代表准确率，分歧代表多样性，由式(11)可知误差 $\bar{E}(h|x)$ 越小，个体判别器准确率越高；分歧 $\bar{A}(h|x)$ 越大，个体判别器多样性越大，集成模型误差 $E(H|x)$ 越小。因此，为建立通用化的二次设备故障检测方法，本文将设计一种以集成误差最小为优化目标的优化模型。给定一组个体判别器和一组待判别样本，其中，个体判别器的数量为 M ，待判别样本数量为 m 。用此组判别器中的每一个判别器对待判别样本进行判别，判别结果记录在矩阵 \mathbf{P} 中， \mathbf{P} 中元素 P_{ij} 为

$$P_{ij} = \begin{cases} 0, & \text{个体判别器 } j \text{ 对样本 } i \text{ 判断正确} \\ 1, & \text{个体判别器 } j \text{ 对样本 } i \text{ 判断错误} \end{cases} \quad (12)$$

令 $\mathbf{U} = \mathbf{P}^T \mathbf{P}$ ，则 \mathbf{U} 的对角元 U_{ii} 代表个体判别器 h_i 判断错误的次数，代表了 $\bar{E}(h|x)$ 中的个体判别器的准确率，即对角元为0的值越多，准确率越高。其非对角元 U_{ij} 则表示个体判别器 h_i 与 h_j 同时出错的次数，代表了式(11)中个体判别器的分歧，即非对角元为1的值越多， $\bar{A}(h|x)$ 越大，总分歧越大。对 \mathbf{U} 中元素进行归一化处理得到 $\tilde{\mathbf{U}}$ ， $\tilde{\mathbf{U}}$ 中的元素为

$$\tilde{U}_{ij} = \begin{cases} \frac{U_{ii}}{M}, & i = j \\ \frac{1}{2} \left(\frac{U_{ij}}{U_{ii}} + \frac{U_{ij}}{U_{jj}} \right), & i \neq j \end{cases} \quad (13)$$

按照式(11)的定义，若 $\tilde{\mathbf{U}}$ 中每个元素都最小，则集成模型的效果最好，即 $\tilde{\mathbf{U}}$ 就是集成误差的一个很好的近似。因此，集成模型优化问题可以转化为二次整数规划问题，即



$$\min_x x^T \tilde{U}x \quad (14)$$

$$\text{s.t.} \sum_{i=1}^N x_i = k \quad (15)$$

$$x_i \in \{0, 1\} \quad (16)$$

式中：二值标量 x_i 表示第 i 个体判别器是否被选入集成模型中； k 为优化后的集成模型规模。

二次设备涉及种类繁多，每种二次设备的集成模型规模应由实际情况决定，本节优化方法可以找出给定集成模型规模的最优集成模型。

3 数据驱动的二次系统故障检测算法流程

基于数据驱动的二次系统故障检测方法如下。

1) 数据预处理。首先，对报文信息进行统一格式，删减冗余（见表 1）。然后，为了消除计算重构误差时各特征向量量纲的影响，利用 z-score 标准化方法对特征进行归一化处理。

2) 模型训练。对于报文信息，根据式 (1)~(3) 分别计算各组报文信息的置信度、支持度、提升度，选出符合阈值要求的关联规则。对于二次设备运行状态信息，根据 2.1 节建立二次设备故障检测集成模型。

3) 模型优化。根据式 (12) 记录一组判别器对一组样本的判别信息，并通过式 (13)~(16) 将个体判别器是否加入集成模型问题转化为二次整数规划模型，选出最佳集成模型。可以通过如下指标对模型性能进行评价。

$$C_{AC} = (P_T + N_T) / (P_T + N_T + P_F + N_F) \quad (17)$$

$$P_{RE} = P_T / (P_T + P_F) \quad (18)$$

$$C_{RE} = P_T / (P_T + N_F) \quad (19)$$

$$F_1 = 2P_{RE}C_{RE} / (P_{RE} + C_{RE}) \quad (20)$$

式中： P_T 和 P_F 分别为被模型预测为正类的正样本和负样本； N_F 和 N_T 分别为被模型预测为负类的负样本和正样本； C_{AC} 为准确率，代表判断正确的样例占总数的比例； P_{RE} 为精确率，代表在所有被预测为正的样本中实际为正样本的概率； C_{RE} 为召回率，代表预测为正类的正样本占全部正样本的比例； F_1 为精确率和召回率的调和平均

值，能兼顾判断模型的精确率和召回率。

4) 综合判断。根据监控设备实时采集到的二次系统监测数据，分别将报文信息数据集（报文告警信息）和二次设备数据集（二次设备运行状态监测信息）输入步骤 2 提取的各类关联规则与步骤 3 训练完备的最优集成模型中，如果告警信息包含在所提挖掘的关联规则中，则可提前排查出潜在故障的二次设备，针对这些潜在的二次设备进一步观察其在步骤 3 集成优化模型中的重构误差，如果超过设定的重构误差阈值，则可综合判断该设备发生故障，否则无法直接确定该设备为异常，需要专家进行人工核对。

4 算例分析与讨论

4.1 模型设置与指标介绍

本文所提方法主要针对二次系统的报文信息数据以及二次设备数据，通过河南某变电站的二次系统报文信息数据以及 2 种二次设备数据对所提方法进行验证。二次系统报文信息共 400 条，该 400 条数据是从海量报文告警信息中经过专家筛选且确定该条报警信息对应有装置故障的数据，全部用于关联规则提取，具体格式如表 1 所述。值得注意的是，虽然每天告警信息达上万条，但大部分是无效报警信息（如由于正常操作引起的报警），由于其未含有告警信息和故障原因的关联关系，不能加入样本集进行关联规则提取。二次设备运行数据包括河南某变电站线路保护 20 套，含保护装置温度、工作电压、GOOSE 插件发出-接受光强、插件发出-接受光强、纵联通道发出-接受光强数据，其中有 1 套保护出现严重老化。本文从正常线路保护的样本数据中随机选择 8 组作为训练集，剩余的 12 组线路保护数据（含一组老化设备数据）作为测试集数据。本文所有算例仿真均在 Intel(R) Core(TM) i7-9750H CPU@2.60 GHz、24 GB RAM 的硬件环境下测试。

对基于关联规则的二次系统报文信息故障检测分析方法算例指标设置为：支持度在 0.017 以上、置信度在 0.5 以上并且提升度在 3 以上。

算例模型对比设置如下。

M1：基于重构误差计算的二次设备个体判别器模型（作为基础模型），模型选用 Bi-LSTM，



隐含层数为 2 层，每层神经元数为 32，初始学习率选为 0.001，优化器选用 Adam (adaptive method)。

M2：以 M1 作为基础模型，并采用基于集成学习的二次设备故障检测模型。

M3：本文所提方法，与 M2 采用相同的集成模型，但采用 2.2 节中优化方法进行优化^[22]。

4.2 关联规则方法结果展示

首先将 400 条二次系统报文信息进行解析整理，将统一格式后的数据集作为输入，在使用 Apriori 算法进行告警信息与故障装置的关联规则挖掘后，计算出重要关联规则所对应的支持度、置信度以及提升度的具体数据信息（见表 2）。针对保护电源插件损坏、通道故障以及电源插件损坏，分别有 2 类告警信息，可以为工作人员提供参考。1) 有效避免集控站的监控分析人员从海量报警信息中筛除无用的报警信息，即满足所挖掘出的关联规则的报文信息才可能是有效告警信息；2) 对二次系统发出的有用告警信号进行查询确定故障原因，对潜在的故障装置进行检查，为检修人员提供检修依据。本文所提方法具有扩展性，随着二次系统报文信息增加，所提方法可提取出更多有价值的关联规则。

表 2 有效关联规则
Table 2 Effective association rules

| 告警信息 | 故障装置 |
|----------------------------------|----------|
| 对侧保护通道告警，线路1保护装置失电告警 | 保护电源插件损坏 |
| 线路1运行灯熄灭，线路1保护装置失电告警 | 保护电源插件损坏 |
| 现场“通道异常灯”亮，后台报“**线光距保护装置闭锁或异常”信号 | 通道故障 |
| 线路1保护2告警，线路1保护1告警 | 通道故障 |
| 装置运行灯不亮，装置失电告警 | 电源插件损坏 |
| 液晶面板灭，运行灯熄灭，线路1保护装置失电告警 | 电源插件损坏 |

4.3 故障检测集成模型稳定性验证

对线路保护算例对比数据设置如下：M1 在 20 组线路保护数据中随机选择 8 组正常线路保护的样本数据作为训练集，剩余的 12 组线路保护数据（含一组老化设备数据）作为测试集数据；M2 共包括 8 个基础模型，每个基础模型分别采用一组正常线路保护的数据来训练，但是训练集和测试集与 M1 相同。随机选用不同的 8 组正常线路保护数据作为训练集，剩余 12 组设备（含一组

老化设备）作为测试集，对 M1、M2 分别进行 3 次测试，结果如表 3 所示。

表 3 M1 与 M2 预警结果指标比较
Table 3 Comparison of early warning results for line protection data between M1 and M2

| 模型 | 实验次数 | $C_{AC}/\%$ | $P_{RE}/\%$ | $C_{RE}/\%$ | $F_1/\%$ | 排序 | 总时长/s |
|----|------|-------------|-------------|-------------|----------|----|-------|
| M1 | 1 | 90 | 50.0 | 100 | 66.7 | 1 | 4.7 |
| | 2 | 80 | 33.3 | 100 | 50.0 | 2 | 5.2 |
| | 3 | 80 | 33.3 | 100 | 50.0 | 1 | 6.4 |
| M2 | 1 | 100 | 100.0 | 100 | 100.0 | 1 | 69.1 |
| | 2 | 100 | 100.0 | 100 | 100.0 | 1 | 60.4 |
| | 3 | 100 | 100.0 | 100 | 100.0 | 1 | 71.9 |

由表 3 对比可知，M1 在 3 次测试中评估指标的波动较大，且相应的性能指标较低；M2 在 3 次测试中的评估指标基本不变，性能指标达到了 100%。上述结果表明：M2 相比较于 M1 具有更高的稳定性；M2 的训练测试总时长多于 M1，但不会影响故障检测。综上，本节算例验证了集成模型相较于基础模型具有更好的稳定性和精度。

4.4 集成优化模型精度提升验证

基于集成优化的二次设备故障检测方法旨在给定集成模型规模的情况下，给出基于误差-分歧理论的最优集成模型。为了体现该方法对集成模型精度的提升效果，对线路保护数据设置算例如下：M2 的训练集和测试集设置同 4.3 节。将优化后的集成模型规模 k 设定为 5，在 8 组训练集中随机选择 3 组集成模型规模为 5 的模型作为对照。测试集相同，测试结果如表 4 和图 3 所示。由表 4 知，对于线路保护数据而言，相比于原来集成模型 M2，优化后模型 M3 以更小的集成规模实现更高的准确率。图 3 展示了各个模型在测试集数据中高预警概率设备的预警概率值。对于线路保护

表 4 优化前后的集成模型性能比较
Table 4 Performance comparison of integrated models before and after optimization

| 模型 | 模型个数 | $C_{AC}/\%$ | $C_{RE}/\%$ | 是否漏判 |
|-----------|------|-------------|-------------|------|
| 优化后(M3) | 5 | 100 | 100 | 否 |
| 随机集成1(M2) | 5 | 83 | 100 | 否 |
| 随机集成2(M2) | 5 | 83 | 100 | 否 |
| 随机集成3(M2) | 5 | 83 | 100 | 否 |
| 原集成模型(M2) | 8 | 92 | 100 | 否 |

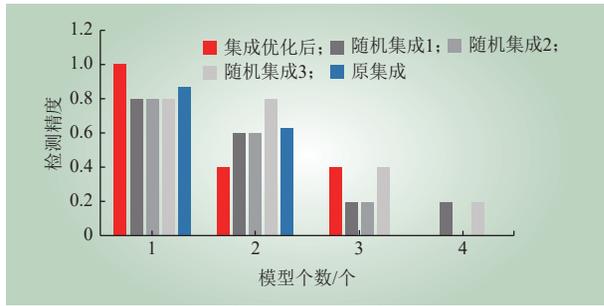


图 3 优化前后故障检测模型效果对比

Fig. 3 Effect comparison of fault early warning models before and after optimization

数据而言，优化后的模型判定老化线路保护的故障检测概率值为 100%，并且对于未老化的线路保护误判率显著低于原模型和随机模型。综上，集成优化模型提高了原模型的精度，并且相较于随机集成模型精度更高。

4.5 二次设备典型故障判别有效性验证

利用 M3 模型进一步测试其对 GOOSE 通道中断与通风不畅 2 种典型二次保护设备故障的判别效果（见表 5），并利用 M3 中间过程的重构误差计算各个典型故障的重构误差占比分布，直接利用重构误差占比分布定位故障源类型（见图 4、5）。由表 5 可知，M3 对 2 种故障判别的精确率与准确率均为 100%。

表 5 M3 在二次设备典型故障中的判别效果

Table 5 The discrimination effect of M3 in typical faults of secondary equipment

| 故障 | 模型个数 | $C_{AC}/\%$ | $C_{RE}/\%$ | 是否漏判 |
|--------|------|-------------|-------------|------|
| 纵联通道中断 | 10 | 100 | 100 | 否 |
| 通风不畅 | 10 | 100 | 100 | 否 |

由图 4 可知，纵联通道中断可直接引起该套保护对应光口纵联光强重构误差占比变大，M3 判别出故障后，可直接通过重构误差占比分布定位通道中断故障。

由图 5 可知，通风不畅可直接引起该套保护装置温度的重构误差占比变大，M3 判别出故障后，可直接通过重构误差占比分布定位通风不畅故障。

5 结语

针对现有二次系统故障检测方法难以综合对

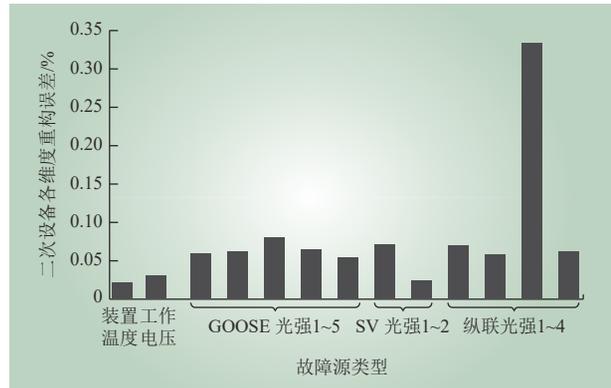


图 4 纵联通道中断故障重构误差占比分布

Fig. 4 Distribution of reconstruction error proportion of longitudinal channel interruption faults

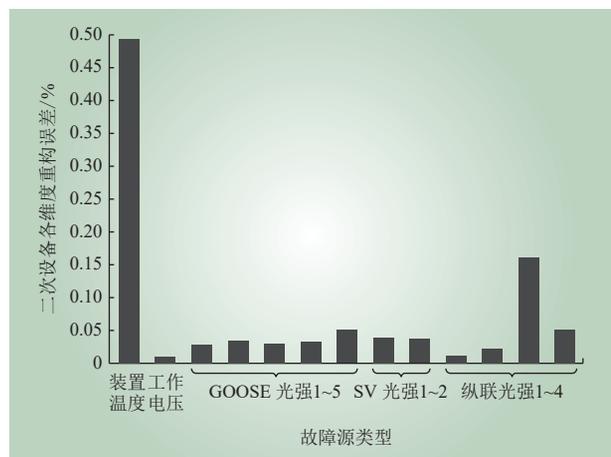


图 5 二次设备通风不畅重构误差占比分布

Fig. 5 Distribution of reconstruction error proportion due to poor ventilation of secondary equipment

二次回路以及二次设备进行故障检测定位，并且针对单个设备检测难以适用于正常数据和异常数据变化差异微小的场景，本文提出基于关联规则与重构误差计算的二次系统故障检测方法。利用河南某变电站报文信息数据集与二次设备数据集，验证了本文所提方法的有效性。囿于数据原因，本文所提方法对于更多不同二次设备故障识别以及故障源定位还未充分验证，未来将结合二次设备的特点利用本文所提方法进行探索。

参考文献：

[1] 北极星输配电网. 2020 南瑞集团调研 [EB/OL]. (2022-07-04)[2022-12-13]. <https://news.bjx.com.cn/html/20220704/1238071.shtml>.
 [2] 国家能源局. 《电力二次系统安全管理若干规定》政策解读



- 读 [EB/OL]. (2022-11-23)[2022-12-13]. http://www.nea.gov.cn/2022-11/23/c_1310679041.htm.
- [3] 唐志军, 李泽科, 陈建洪, 等. 考虑设备相关性的智能变电站二次系统可靠性分析 [J]. 福州大学学报 (自然科学版), 2021, 49(6): 782-789.
- TANG Zhijun, LI Zeke, CHEN Jianhong, *et al.* Reliability analysis for intelligent substation secondary system considering equipment correlation[J]. Journal of Fuzhou University (Natural Science Edition), 2021, 49(6): 782-789.
- [4] 赵红. 智能变电站二次设备状态评估方法研究 [D]. 重庆: 重庆大学, 2017.
- ZHAO Hong. Research on status evaluation method for smart substation secondary equipments[D]. Chongqing: Chongqing University, 2017.
- [5] 胡宇薇. 神经网络专家系统在电力系统输电设备保护中的应用研究 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
- HU Yuwei. Application and research on neural networks and expert system in power system transmission equipment protection[D]. Harbin: Harbin Institute of Technology, 2018.
- [6] 胡凡. 智能变电站二次系统可靠性评估 [D]. 济南: 山东大学, 2015.
- HU Fan. Reliability evaluation of secondary system of intelligent substation[D]. Jinan: Shandong University, 2015.
- [7] 唐楚雪, 陈德明. 基于模糊数学和变权理论的智能变电站二次设备状态估计 [J]. 电气自动化, 2018, 40(6): 71-73.
- TANG Chuxue, CHEN Deming. State evaluation for secondary equipment of intelligent substations based on fuzzy mathematics and variable weight theory[J]. *Electrical Automation*, 2018, 40(6): 71-73.
- [8] 滕宇行. 变电站二次设备智能巡检系统设计与实现 [D]. 大连: 大连理工大学, 2023.
- TENG Yuhang. Design and implementation of intelligent inspection system for secondary equipment of substation[D]. Dalian: Dalian University of Technology, 2023.
- [9] 吴文传, 吕颖, 张伯明. 继电保护隐患的运行风险在线评估 [J]. 中国电机工程学报, 2009, 29(7): 78-83.
- WU Wenchuan, LÜ Ying, ZHANG Boming. On-line operating risk assessment of hidden failures in protection system[J]. *Proceedings of the CSEE*, 2009, 29(7): 78-83.
- [10] 南东亮, 王维庆, 张陵, 等. 基于关联规则挖掘与组合赋权-云模型的电网二次设备运行状态风险评估 [J]. 电力系统保护与控制, 2021, 49(10): 67-76.
- NAN Dongliang, WANG Weiqing, ZHANG Ling, *et al.* Risk assessment of the operation state of power grid secondary equipment based on association rule mining and combination weighting-cloud model[J]. *Power System Protection and Control*, 2021, 49(10): 67-76.
- [11] 鄢小坤, 赵武智, 牛静, 等. 一种智能变电站二次设备状态评价方法 [J]. 电子器件, 2021, 44(3): 664-669.
- WU Xiaokun, ZHAO Wuzhi, NIU Jing, *et al.* A state assessment method for secondary equipment in intelligent substation[J]. *Chinese Journal of Electron Devices*, 2021, 44(3): 664-669.
- [12] 李龙, 陈乾, 杨瑞, 等. 基于 Hamacher 算子的变电站自动化二次设备状态模糊综合评估方法 [J]. 湖北电力, 2022, 46(3): 45-49.
- LI Long, CHEN Qian, YANG Rui, *et al.* Fuzzy comprehensive evaluation method for substation automatic secondary equipment status based on Hamacher operator[J]. *Hubei Electric Power*, 2022, 46(3): 45-49.
- [13] 陈勇, 李胜男, 张丽, 等. 基于改进 Apriori 算法的智能变电站二次设备缺陷关联性分析 [J]. 电力系统保护与控制, 2019, 47(20): 135-141.
- CHEN Yong, LI Shengnan, ZHANG Li, *et al.* Association analysis for defect data of secondary device in smart substations based on improved Apriori algorithm[J]. *Power System Protection and Control*, 2019, 47(20): 135-141.
- [14] 张宇泽, 张日新. 基于 Apriori 算法的配电线路故障关联分析 [J]. 电工技术, 2022(5): 138-140.
- ZHANG Yuze, ZHANG Rixin. Association analysis of distribution line fault based on Apriori algorithm[J]. *Electric Engineering*, 2022(5): 138-140.
- [15] 王鸣誉, 李铁成, 任江波, 等. 利用 Apriori 算法实现变电站二次系统故障诊断 [J]. 电力系统及其自动化学报, 2021, 33(11): 145-150.
- WANG Mingyu, LI Tiecheng, REN Jiangbo, *et al.* Realization of fault diagnosis of substation secondary system using Apriori algorithm[J]. *Proceedings of the CSU-EPSA*, 2021, 33(11): 145-150.
- [16] REN Y W, LIU L, WANG Z H, *et al.* A fault analysis and prediction of aircraft based on association rules and weibull distribution[C]// 2019 10th International Conference on Information Technology in Medicine and Education (ITME). Qingdao, China. IEEE, 2019: 570-578.
- [17] 白浩, 袁智勇, 孙睿, 等. 基于 Apriori 算法和卷积神经网络的配电设备运行效率主要影响因素挖掘 [J]. 电力建设, 2020, 41(3): 31-38.
- BAI Hao, YUAN Zhiyong, SUN Rui, *et al.* Method based on Apriori algorithm and convolution neural network for mining main influencing factors of distribution equipment operation efficiency[J].



- Electric Power Construction, 2020, 41(3): 31–38.
- [18] 史佳琪, 马丽雅, 李晨晨, 等. 基于串行-并行集成学习的高峰负荷预测方法 [J]. 中国电机工程学报, 2020, 40(14): 4463–4472, 4726.
- SHI Jiaqi, MA Liya, LI Chenchen, *et al.* Daily peak load forecasting based on sequential-parallel ensemble learning[J]. Proceedings of the CSEE, 2020, 40(14): 4463–4472, 4726.
- [19] MEHMOOD K, UL HASSAN H T, RAZA A, *et al.* Optimal power generation in energy-deficient scenarios using bagging ensembles[J]. IEEE Access, 2019, 7: 155917–155929.
- [20] 郑元兵. 变压器故障诊断与预测集成学习方法及维修决策模型研究 [D]. 重庆: 重庆大学, 2011.
- ZHENG Yuanbing. Research on ensemble learning of fault diagnosis and prediction and maintenance decision-making models for transformers[D]. Chongqing: Chongqing University, 2011.
- [21] ZHANG Y, BURER S, STREET W N. Ensemble pruning via semi-definite programming[J]. Journal of Machine Learning Research, 2006, 7: 1315–1338.
- [22] HU R H, ZHOU S B, LIU Y S, *et al.* Margin-based Pareto ensemble pruning: an ensemble pruning algorithm that learns to search optimized ensembles[J]. Computational Intelligence and Neuroscience, 2019, 2019: 7560872.

作者简介:

王阳 (1986—), 女, 高级工程师, 从事继电保护等研究, E-mail: wangyang_0201@qq.com;

余娟 (1980—), 女, 通信作者, 教授, 从事电网分析、人工智能等研究, E-mail: yujuancqu@qq.com。

(责任编辑 李博)

Secondary System Fault Detection Method Based on Association Rules and Reconstruction Error

WANG Yang¹, MA Weidong¹, LIU Jiming², WANG Boshi², YAO Kai³, HAN Wei¹, YU Juan²

(1. State Grid Henan Electric Power Research Institute, Zhengzhou 450003, China; 2. State Key Laboratory of Power Transmission Equipment Technology (Chongqing University), Chongqing 400044, China; 3. State Grid Henan Electric Power Co., Ltd., Zhengzhou 450003, China)

Abstract: The reliability of the secondary system directly affects the safe and reliable operation of the entire substation and even the power grid. With the high proportion of renewable energy sources integrated into power systems, it is increasingly important to effectively detect the secondary system faults. However, existing research faces two challenges: on the one hand, the existing fault detection methods for logic circuits have high requirements for data completeness and are difficult to apply in practice; on the other hand, the existing fault detection methods for secondary equipment are difficult to identify the small differences between normal and fault data, and the computational accuracy is difficult to guarantee. Therefore this paper proposes a secondary system fault detection method based on association rules and reconstruction errors. Firstly, the Apriori algorithm is used to derive the association rules between fault alarm information and fault devices in the logic circuit, achieving rapid diagnosis of logic circuit faults. Then, the individual discriminator is trained using the operational data of normal secondary equipment, and the operating status of the secondary equipment is determined by measuring the reconstruction error of the data to be discriminated. The ensemble learning model is used to quantify the current fault detection probability of the equipment. Finally, the ensemble learning model is optimized to improve the accuracy of secondary equipment anomaly warning. The effectiveness and accuracy of the proposed method was verified by simulating the dataset from a substation in Henan Province.

This work is supported by the Science and Technology Project of Electric Power Research Institute of State Grid Henan Electric Power Company (Research on Reliability Evaluation Technology of Substation Secondary System Based on Life Cycle, No.SGHADK 00DWJS2200241).

Keywords: secondary system; fault detection; data driven; association rules; reconstruction error