

# 数据驱动的配电开关设备交互式诊断平台

陈国炎<sup>1</sup>, 李俊均<sup>2</sup>, 陈颖<sup>2</sup>, 梅生伟<sup>2</sup>

(1. 中国南方电网广州供电局电力试验研究院, 广东 广州 100142;  
2. 清华大学电机系, 北京 100084)

**摘要:** 高效和准确的设备缺陷诊断有助于提升配电网运行可靠性和安全性。受益于信息化建设, 配电设备到货抽检、型式试验、在线监测等环节积累了大量检测大数据, 奠定了数据挖掘和智能诊断的基础。为了满足配电设备智能运维需求, 本文提出了一种大数据驱动的配电开关设备故障交互式诊断方法。利用设备的物理模型、历史故障诊断结果、在线监测数据以及仿真结果, 所提方法挖掘故障表征与潜在缺陷间的关联关系, 并训练贝叶斯网络, 实现设备缺陷推理。进一步, 根据试验人员的反馈调整诊断结果, 提出设备故障的交互式诊断方法, 通过动态修正推理模型, 提升检测准确性。实际案例测试说明所提方法的有效性。

**关键词:** 配电开关; 贝叶斯网络; 关联规则

**DOI:** 10.12067/ATEEE1810004

**文章编号:** 1003-3076(2019)03-0010-08

**中图分类号:** TM769

## 1 引言

配电开关设备是配电网中的常见装备, 其安全可靠运行是提升配电网供电可靠性的前提条件。开关设备检测和运维是配电网运行维护的重要环节。传统设备故障诊断方法依赖人工经验, 成本高、效率低、准确性较低, 无法满足配电网快速发展需求。研究智能、高效、准确的配电开关设备故障诊断方法, 具有显著的工程价值。

已有的设备缺陷诊断方法主要包括数据驱动与模型驱动两大类。模型驱动方法中, 专家系统发展最早也最成熟。文献[1]将继电保护方面的专家知识转化为 IF-THEN 类型的故障诊断规则。文献[2]提出了一种基于“与-或”逻辑的设备故障分析方法。此类专家系统容错能力较差, 推理计算耗时较长, 难以适应复杂配电设备诊断任务。另一方面, 基于人工神经网络、优化技术及模糊集、粗糙集等方法, 数据驱动的设备故障诊断技术正在快速发展。文献[3]综合利用多层感知器和广义神经网络描述电网继电保护的行为规则, 根据输出的模式分类给出诊断结论; 文献[4]采用概率方法为开关保护动作的

因果关系建立概率因果矩阵, 并采用最优化算法将诊断转换为整数规划问题; 文献[5]考虑只有故障与保护之间、保护和断路器之间的关联关系的不确定性, 通过模糊隶属度求取潜在的故障元件。数据驱动算法的优势在于能够充分利用历史诊断试验所积累的大量数据, 弥补人工总结经验的局限性, 更加便于描述不确定与不精确的知识, 为设备诊断带来灵活性。但是, 纯粹依靠数据驱动算法难以考虑设备的物理结构, 无法从故障机理的角度深入分析故障背后的潜在缺陷, 分析结果缺乏可解释性。

为了弥补上述的不足, 本文提出了一种大数据驱动的配电开关设备故障交互式诊断平台, 综合利用设备的物理模型、历史故障诊断结果以及仿真结果, 挖掘故障表征与潜在缺陷间的关联关系, 构建贝叶斯网络拓扑, 并根据试验人员的反馈调整诊断结果, 实现设备故障的交互式诊断。

## 2 基于贝叶斯网络的配电设备故障诊断

配电设备故障诊断是一个典型的分类问题。如式 1 所示, 对于一个待诊断设备, 其所有可能的设备缺陷为  $N_{\text{defect}} = \{y_1, y_2, \dots, y_k\}$ , 可观测诊断证据集

收稿日期: 2018-10-11

基金项目: 国家自然科学基金项目(51477081)、广州供电局有限公司科技项目(080037KK52160010)

作者简介: 陈国炎(1980-), 男, 福建籍, 高级工程师, 博士, 研究方向为电力系统继电保护、配电自动化等;  
李俊均(1994-), 男, 广东籍, 硕士研究生, 研究方向为配电设备的缺陷诊断等。

为  $E = \{e_1, e_2, \dots, e_h\}$ , 则故障诊断模型  $M$  可视为一个分类模型,即基于诊断证据从候选缺陷中找到最大似然的设备缺陷  $y_d$ 。

$$y_d = \operatorname{argmax} P(y | E, M) \quad (1)$$

诊断模型为概率模型,反映了设备缺陷与诊断证据的相关关系,是诊断平台的核心。考虑到贝叶斯网络可以在不完全信息下做推断,并且可以方便地整合先验知识,采用贝叶斯网络构建诊断模型。

贝叶斯网络可以被记为  $B = \langle G_B, \Theta_B \rangle$ , 其中  $G_B$  表示贝叶斯网络的拓扑结构,  $\Theta_B$  表示贝叶斯网络的参数,即变量的条件概率表。相应地,式(1)中的诊断过程可以被转换为:

$$y_d = \operatorname{argmax} P(y | E, G_B, \Theta_B) \quad (2)$$

基于贝叶斯网络的设备诊断模型在实际应用中,需要解决如下两个关键挑战:①贝叶斯网络的拓扑结构的确定是一个 NP 难问题,为了缩小网络拓扑的搜索范围,本文提出以设备的物理-故障模型作为贝叶斯网络的基本拓扑,并将关联规则以边的形式添加到贝叶斯网络中,构成网络的完整拓扑;②由于实际检测中,提供给模型的诊断证据往往是不充分的,因而诊断模型不能确保在单轮诊断中就能给出可靠的结果,因此,本文提出一种交互式的诊断机制,每一轮诊断中都以试验人员的反馈作为附加诊断证据,用以更新诊断结果。

### 3 贝叶斯网络构建与训练

典型贝叶斯网络的拓扑结构如图 1 所示。贝叶斯网络是一种使用有向无环图 (DAG) 来描述变量之间相关关系的概率图模型。对于给定贝叶斯网络  $B = \langle G_B, \Theta_B \rangle$ , 网络变量的联合概率密度函数为:

$$P(n_1, n_2, \dots, n_{|N|}) = \prod_{n_i \in N} P(n_i | G_B, \Theta_B, \pi_{n_i}) \quad (3)$$

式中,  $\pi_{n_i}$  为节点  $n_i$  的父节点集合;  $N = \{n_i\}_{i=1}^{|N|}$  为网络的节点集合。

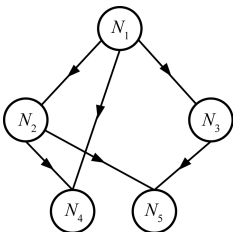


图 1 典型贝叶斯网络的拓扑结构

Fig. 1 Typical topology of a Bayesian network

类问题:给定诊断证据  $E = \{e_1, e_2, \dots, e_h\}$ , 贝叶斯网络拓扑  $G_B$  以及网络参数  $\Theta_B$ , 求解最大似然的设备缺陷  $y_d$ 。因此,合理确定贝叶斯网络拓扑和网络参数,直接决定诊断结果的有效性。

对于含有  $|N|$  个节点的贝叶斯网络,可能的拓扑结构共有  $2^{|N|(|N|-1)/2}$  种<sup>[6]</sup>,通过枚举的方式找到最优网络拓扑是 NP 难问题。

本文基于如下两个假设来缩小网络拓扑的搜索空间:

(1)假设 1:如果两个设备元件在物理上直接关联,那么对应的设备缺陷间有相关关系。

(2)假设 2:如果设备的某种故障表征以及设备缺陷在历史诊断记录中频繁地同时出现,那么它们在概率分布上也是相关的。

因此,基于以上的假设,本文提出如下三步以确定网络拓扑:

(1)建立设备的物理-故障混合模型作为贝叶斯网络基本拓扑,反映空间的相关关系。

(2)对训练数据集进行关联规则挖掘,发现一系列(缺陷-故障)频繁项集,并以横向连接方式添加到贝叶斯网络的基本拓扑中,形成完整的贝叶斯网络。

(3)根据训练的结果对贝叶斯网络实施剪枝操作,减少冗余的连接关系,提升泛化能力。

#### 3.1 基于设备物理-故障模型构建贝叶斯网络基本拓扑

配电开关的物理-故障模型为设备物理模型与历史诊断记录的结合。其中,设备物理模型由设备厂商提供,以树的形式反映了设备元件之间的空间连接关系。以环网柜 (Ring Main Unit, RMU) 为例,设备物理模型如图 2 所示。记物理模型的节点和边分布为  $N_{\text{physic}}$  和  $L_{\text{physic}}$ 。

设备的历史诊断记录为设备全生命周期中的历次诊断信息,包括诊断时间、设备类型、关联部件、故障类型以及相应的设备缺陷,由设备检测机构提供。典型的设备故障诊断记录如附表 1 所示。

记历史诊断记录中的所有故障的集合为  $N_{\text{fault}} = \{f_1, f_2, \dots, f_k\}$ , 发生故障的设备部件为  $N_{\text{component}} = \{c_1, c_2, \dots, c_k\}$ , 建立设备故障与故障部件的连接边集合  $L_{\text{link}} = \{(f_i, c_i)\}_{i=1}^k$ , 则设备的物理-故障模型为  $G_0(N, L_0)$ , 其中  $N = N_{\text{physic}} \cup N_{\text{fault}}$ ,  $L_0 = L_{\text{physic}} \cup L_{\text{link}}$ 。环网柜的物理-故障模型如图 3 所示,将其作为贝叶斯网络的基本拓扑。

设备故障诊断可以建模为一种典型的贝叶斯分

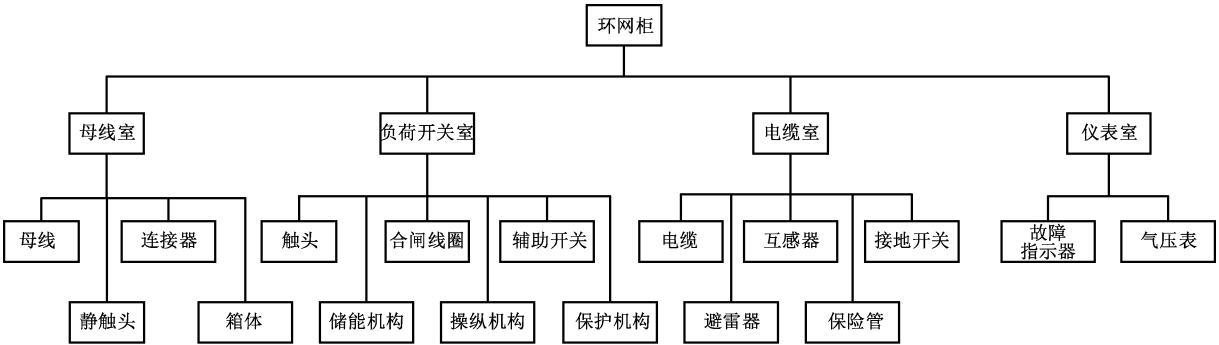


图 2 环网柜物理模型

Fig. 2 Physical model of RMU

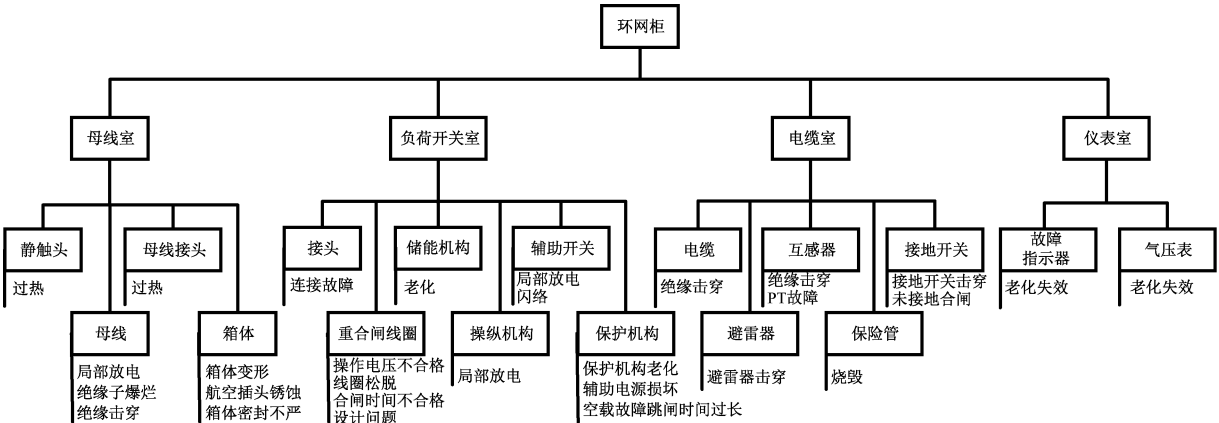


图 3 环网柜的物理-故障混合模型

Fig. 3 Physical-fault model for RMU

3.2 通过关联规则挖掘构建贝叶斯网络完整拓扑

在机器学习领域,关联规则意味着具有明确语义、能够描述数据分布所隐含的客观规律和概念,可以写成“若  $A$  成立,则  $B$  成立”的形式的逻辑规则。支持度和置信度是关联规则的两个重要评价指标。对于规则  $F:(x^{(1)}=v_1) \wedge (x^{(2)}=v_2) \wedge \cdots \wedge (x^{(k)}=v_k) \rightarrow (y=c_f)$ , 在数据集  $(X,Y)=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$  中,支持度  $S_F$  表示满足规则的样本数在数据集中的比例<sup>[7]</sup>,即

$$S_F = \frac{\sum_{i \in N} I(x_i^{(1)}=v_1) \cdots I(x_i^{(k)}=v_k)}{N} \tag{4}$$

式中,  $I(\cdot)$  为示性函数。

规则  $F$  的置信度  $C_F$  表示满足规则  $(x^{(1)}=v_1) \wedge (x^{(2)}=v_2) \wedge \cdots \wedge (x^{(k)}=v_k)$  的样本中,标签  $y=c_f$  的比例<sup>[7]</sup>,即

$$C_F = \frac{\sum_{i \in N} I(x_i^{(1)}=v_1) \cdots I(x_i^{(k)}=v_k) I(y_i=c_f)}{\sum_{i \in N} I(x_i^{(1)}=v_1) \cdots I(x_i^{(k)}=v_k)} \tag{5}$$

在设备故障诊断领域,关联规则挖掘就是要从输入的历史诊断数据中,寻找满足最小支持度与置信度的规则集合。2000 年,Han 等人提出了 FP-growth 算法,该算法采用频繁模式增长的方式,无需遍历大量候选项,大幅提升了计算效率<sup>[8]</sup>。本文将采用该算法进行关联规则的挖掘。FP-growth 算法的主要步骤见文献[9]。该算法输出的所有二元频繁项集合记为  $L_1 = \{(N'_{\text{from}}, N'_{\text{to}}) | N'_{\text{from}}, N'_{\text{to}} \in N\}$ , 令  $L = L_0 \cup L_1$ , 则结合了关联规则的贝叶斯网络拓扑为:

$$G_B = (N, L) \tag{6}$$

3.3 贝叶斯网络剪枝策略

试想在所建贝叶斯网络中,如果添加边使其成为一个全连接网络,则理论上可以学习任何一种概率依赖关系,但实际训练会因为参数过多而导致训练效率低下,并陷入严重的过拟合中。因此,设计合理的剪枝策略,将模型复杂度保持在合理范围内,对提升分类模型的泛化能力与稳定性有着关键意义。

为了综合评估模型的分类精度与泛化能力,本



文基于经典的 K2 算法,提出一种基于交叉熵与模型正则项的网络评分指标:

$$s(G_B | D_T) = \frac{\ln N}{2} |G_B| - LL(G_B | D_T) \quad (7)$$

式中,  $N$  为数据样本数量;  $|G_B|$  为贝叶斯网络参数个数;  $\frac{\ln N}{2} |G_B|$  表征了贝叶斯网络的结构复杂性(倾向于选择更简单的网络模型);为便于验证贝叶斯网络的泛化能力,将设备的历史诊断数据  $D$  先随机抽取 20000 条记录作为测试集  $D_V$ ,再在余下的数据中随机抽取 20000 条记录作为网络的训练集  $D_T$ ;  $LL(G_B | D)$  计算交叉熵,衡量模型分类效果与训练数据的贴合程度:

$$LL(G_B | D_T) = \sum_{x_i \in D_T} P(x_i) \log_2 [P(x_i)] \quad (8)$$

因此,在该评分指标下,最优的模型应该是在最少的参数个数下实现最高的分类精度。基于此,提出如下的模型剪枝算法:

(1) 结合物理-故障混合模型与关联规则挖掘结果,建立贝叶斯网络拓扑  $G_B(N, L)$ 。

(2) 使用设备历史故障数据训练贝叶斯网络  $G_B$ , 求取当前拓扑下的评分结果  $s_0 = s(G_B | D_T)$  作为初始分值,将  $G_B$  中所有边标记为未访问。

(3) 随机在网络中移除一条标记为未访问边  $e_i \in E$ , 得到新的网络拓扑  $G_{Bi}(N, E - E_i)$ 。

(4) 以历史故障数据训练  $G_{Bi}$ , 得到拓扑  $G_{Bi}$  的条件概率表  $\theta_i$ , 并求出当前拓扑的评分结果  $s_i = s(G_{Bi} | D_T)$ 。

(5) 如果  $s_i > s_0$ , 则  $s_0 = s_i$ ,  $G_B = G_{Bi}$ ; 否则将边  $e_i$  加回拓扑中并标记为已访问。

(6) 返回步骤(3)

(7) 输出评分最高的贝叶斯网络拓扑及参数组合。

在得到完整拓扑后,贝叶斯网络参数的求解可以建模为如式(9)所示的优化问题,即条件于当前的训练数据集  $D_T$  与贝叶斯网络拓扑  $G_B$ , 求解最优的条件概率表  $\theta_B$  使得后验概率最大化<sup>[10]</sup>。式(9)可以通过梯度下降法求解。在学习率为  $\eta \in (0, 1)$ 、初始参数为  $\theta_{ij}^{(0)}$  的条件下,参数训练的迭代公式如式(10)所示,迭代结果将会收敛到关于训练数据的后验最大似然(MLE)。

$$\theta_B = \operatorname{argmax}_{\theta} P(D_T | G_B, \theta_B) \quad (9)$$

$$\theta_{i,j}^{(k+1)} = \theta_{i,j}^{(k)} + \eta \sum_{d \in D} \frac{P(n_i, \pi_i | d, G_B)}{\theta_{i,j}^{(k)}} \quad (10)$$

## 4 交互式故障诊断算法与流程

### 4.1 单次诊断算法优化

在得到训练好的贝叶斯诊断模型后,一种可行的诊断方法是给定诊断证据,枚举该设备的每种潜在缺陷  $y_i \in N_{\text{defect}}$ , 选择条件概率最大的缺陷作为诊断结果,如式(11)所示:

$$y_d = \operatorname{argmax}_{y_i \in N_{\text{defect}}} P_B(y_i = 1) \prod_{e_i \in E} P_B(e_i) \quad (11)$$

式中,  $P_B(x)$  代表  $P(x | G_B, \theta_B, \pi_x)$ 。

设备故障诊断是一种典型的贝叶斯推断问题。在介绍具体的诊断算法之前,不妨将贝叶斯网络中的节点分为三类:  $N = \{N_{\text{fault}}, N_{\text{defect}}, N_{\text{link}}\}$ , 其中  $N_{\text{fault}}$  为可观测的故障表征集合,可以用来作为缺陷诊断的证据;  $N_{\text{defect}}$  为设备缺陷,也即缺陷诊断的候选结果集;  $N_{\text{link}}$  为网络中表征空间连接关系的节点,对应着具体的设备部件如保护机构、电缆室等。因此,设备故障诊断可以建模为给定贝叶斯网络拓扑  $G$ 、参数  $\theta$  的条件下,根据观测的故障表征  $E = \{e_1, e_2, \dots, e_h\} \subseteq N_{\text{fault}}$ , 求出最大似然的设备缺陷  $y_d$ , 即

$$y_d = \operatorname{argmax}_{y_i \in N_{\text{defect}}} P(y_i = 1 | E, G_B, \theta_B) \quad (12)$$

但在模型比较复杂的情况下,式(12)的联合条件概率分布计算繁琐,而且会牵涉到大量无关的节点,影响了计算效率。因此,应该结合设备故障诊断自身的特点对算法进行优化。

由人工故障检测的经验常识可知,当设备中某个部件发生缺陷时,首先会影响到与其直接相连的部件以及与之存在关联关系的部件,而其他部件的影响较弱。反之,当某个部件中发现了故障表征,也首先应该从该部件的相邻部件或者关联部件上寻找原因。这意味着,可以利用贝叶斯网络节点的局部结构,缩小诊断过程的搜索范围,减少计算负担。

文献[11]讨论了一种使用马尔科夫毯去除贝叶斯网络中不相关或冗余变量的优化方法,能够提升贝叶斯分类器的性能。对于贝叶斯网络  $G_B(N, E)$  中的任意节点  $Y_i$ , 如果节点集合  $A_i \subset N$  且  $Y_i \notin A_i$ , 满足  $A_i$  d-separate  $Y_i$  与  $N - A_i$ , 也即如果将节点集合  $A_i$  移除,则节点  $Y_i$  与剩余节点没有连接关系,那就称集合  $A_i$  为节点  $Y_i$  的一个马尔科夫毯。显然,节点的马尔科夫毯可以有多种可能的选择。贝叶斯网络部分结构图如图4所示,节点  $Y_i$  的一个马尔科夫毯为  $\{X_3, X_4, X_5, X_6, X_7, X_8\}$ 。

当马尔科夫毯中的每个属性变量给定时,其他

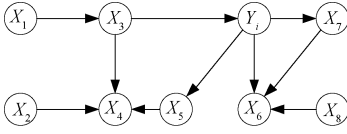


图 4 贝叶斯网络部分结构图

Fig. 4 Partial structure of Bayesian network

节点所提供的信息将被马尔科夫毯所屏蔽,即

$$\begin{aligned} P(y_i = c \mid G_B, \Theta_B) &= P_B(y_i = c) \prod_{x_i \in N} P_B(x_i) \\ &= P_B(y_i = c) \prod_{x_i \in A_i} P_B(x_i) \prod_{x_i \notin A_i} P_B(x_i) \\ &= P_B(y_i = c) \prod_{x_i \in A_i} P_B(x_i) \quad (13) \end{aligned}$$

当诊断证据  $E = \{e_1, e_2, \dots, e_h\}$  给定时,根据元素是否在  $y_i$  的马尔科夫毯中,划分为两个子集:  $E = \{E_{A_i}^+ \cup E_{A_i}^-\}$ , 其中,  $E_{A_i}^+ \subset A_i, E_{A_i}^- \cap A_i = \emptyset$ , 因此有:

$$\begin{aligned} P_B(y_i = c) \prod_{\substack{x_i \in A_i \\ x_i \notin A_i^+}} P_B(x_i) \\ P(y_i = c \mid E, G_B, \Theta_B) = \frac{\prod_{e_i^+ \in E_{A_i}^+} P_B(e_i^+)}{\prod_{e_i^+ \in E_{A_i}^+} P_B(e_i^+)} \quad (14) \end{aligned}$$

相应地,诊断过程表述为:

$$\begin{aligned} y_d = \operatorname{argmax}_{y_i \in N_{\text{defect}}} P(y_i = 1 \mid E, G_B, \Theta_B) \\ = \operatorname{argmax}_{y_i \in N_{\text{defect}}} \frac{P_B(y_i = 1) \prod_{\substack{x_i \in A_i \\ x_i \notin E_{A_i}^+}} P_B(x_i)}{\prod_{e_i^+ \in E_{A_i}^+} P(e_i^+ \mid G_B, \Theta_B)} \quad (15) \end{aligned}$$

与式(12)对比,式(15)的诊断过程借助马尔科夫毯排除了在诊断证据给定下与设备候选缺陷条件独立的变量,缩小了诊断过程的计算规模,提升了诊断效率。

#### 4.2 交互式诊断结果更新流程

一般而言,工程人员提供的诊断证据越充分,则诊断的结果将越可靠。但是,在单次诊断中,工程人员往往只能提供少量外在可观测的诊断证据(如对地短路、保险丝熔断等),与之对应的潜在设备缺陷有多种可能,因而单次诊断不能保证能够得到可靠的结果。交互式诊断允许工程人员对每一轮次的诊断结果提出反馈,补充相关的诊断证据,克服了单轮次诊断的局限性。交互式诊断的流程详述如下:

(1) 设  $E^{(m)} = \{e_i^{(m)}\}_{i=1}^h$  为第  $m$  轮诊断中提供的

诊断证据,由式(15)得第  $m$  轮诊断结果为:

$$y_d^{(m)} = \operatorname{argmax}_{y_i \in N_{\text{defect}}} P(y_i = 1 \mid E^{(m)}, G_B, \Theta_B) \quad (16)$$

工程人员取得本轮诊断结果后,查验设备是否存在缺陷  $y_d^{(m)}$ , 如果缺陷属实,则诊断完成;否则将缺陷  $y_d^{(m)}$  不存在作为补充证据反馈给诊断系统,也即第  $(m+1)$  轮诊断的证据  $E^{(m+1)} = \{e_i^{(m+1)}\}_{i=1}^{h+1}$  为:

$$e_i^{(m+1)} = \begin{cases} e_i^{(m)} & 1 \leq m \leq h \\ \neg y_d^{(m)} & m = h + 1 \end{cases} \quad (17)$$

由式(17)得第  $(m+1)$  轮诊断结果为:

$$y_d^{(m+1)} = \operatorname{argmax}_{y_i \in N_{\text{defect}}} P(y_i = 1 \mid E^{(m+1)}, G_B, \Theta_B) \quad (18)$$

## 5 案例分析

### 5.1 故障诊断平台的实现

基于第 4 节的故障诊断算法,配电开关故障诊断平台如图 4 所示。该平台由三层组成:数据层、诊断层和交互层。数据层包含三个数据源,分别为:物理模型数据库、历史诊断数据库和仿真结果库,向诊断层提供设备物理模型和训练数据集。诊断层在训练数据集中挖掘关联规则,并根据关联规则和设备物理模型构建贝叶斯网络。交互层向检测人员提供友好的人机交互界面(HMI),并可以接收用户对诊断结果正确与否的反馈,作为新的诊断证据调整诊断结果,实现交互式诊断。下面用两个案例说明该系统的具体诊断流程。

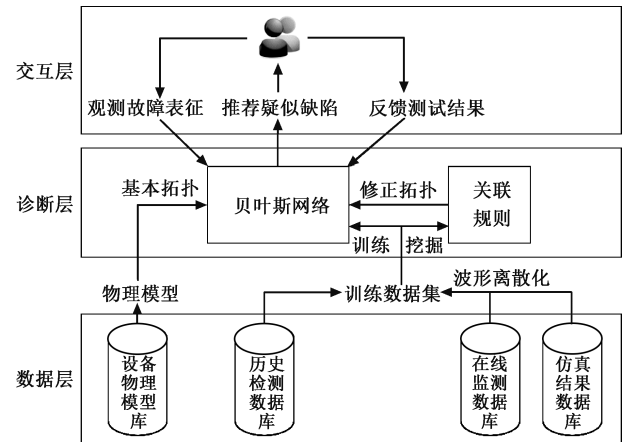


图 5 设备缺陷诊断平台架构图

Fig. 5 Interactive diagnosis platform architecture

### 5.2 诊断案例一

2017 年 3 月,某区局在巡检中发现某 110kV 投运中的环网柜柜门发黑,开柜检查时发现电缆室内肘型头处有烧黑痕迹,表明设备缺陷出在电缆室和

电缆上,除此之外没有其他不良表征。因此,在第 1 轮诊断中,将诊断证据  $E^{(0)} = \{ \text{电缆室} = 1, \text{电缆} = 1 \}$  输入诊断系统中,返回排名靠前的 5 个可疑设备缺陷,其条件概率如附表 2 所示。可见电缆绝缘层的局部放电是最大似然的缺陷  $P(\text{局部放电} = 1 | E^{(0)}, G_B, \Theta_B) = 0.25$ 。将环网柜解体送检,测量其局放电量,发现电缆绝缘层局放电量果然大幅超出规定。更换环网柜电缆后,以上缺陷表征不再出现,表明诊断结果准确。

在本案例中,虽然用户向诊断系统提供的诊断证据比较有限,但贝叶斯网络能够从历史诊断记录中总结出,只要是与环网柜电缆有关的故障,有 1/4 的几率都是由电缆绝缘层局部放电引起的,因而在信息不足的条件下推断设备缺陷为局部放电。

5.3 诊断案例二

2017 年 12 月,运维人员发现某地环网柜开关速断保护动作跳闸,间隔电缆肘型头发生沿面放电,表明此时故障位置在电缆室的电缆中,故障类型为沿面放电。因此,第 1 轮诊断中,诊断证据  $E^{(0)} = \{ \text{电缆室} = 1, \text{电缆} = 1, \text{沿面放电} = 1 \}$ 。系统返回排名靠前的 5 个可疑设备缺陷,其条件概率如附表 3 所示。

可见当前可能性最大的设备缺陷为“电缆局部放电”与“电缆绝缘击穿”,其条件概率均为 0.194,表明系统初步推断当前系统缺陷为环网柜绝缘问题。此后工程人员将环网柜电缆拆下送检,试验结果表明电缆的绝缘没有问题。因此,第 1 轮诊断中将诊断证据修改为:  $E^{(1)} = \{ \text{电缆室} = 1, \text{电缆} = 1, \text{沿面放电} = 1, \text{局部放电} = 0, \text{电缆绝缘击穿} = 0 \}$ 。由此系统更新诊断结果,如附表 4 所示。

可见系统推断的条件概率最大的故障为“保护机构老化失效”,即当配电网因为故障导致过电流或者遭遇雷击时,由于环网柜保护机构老化未能及时切除故障电流,导致柜体内部电缆沿面放电。但经过离线排查,设备保护机构并未出现相关的老化

问题,因此需要在下一轮诊断证据中排除该项缺陷。

$E^{(2)} = \{ \text{电缆室} = 1, \text{电缆} = 0, \text{沿面放电} = 1, \text{局部放电} = 0, \text{电缆绝缘击穿} = 0, \text{保护机构老化失效} = 0 \}$ ,相应的第 3 轮诊断结果如附表 5 所示。

根据最新的诊断结果,工程人员排查电缆头的安装问题,发现该环网柜电缆头距离箱体底板过近,绝缘皮包裹长度不足,在潮湿天气中产生沿面放电。重新调整电缆头安装位置后,未再观察到沿面放电现象。

5.4 诊断性能测试

为了进一步测试本文提出的诊断算法的诊断性能,在包含 20000 条历史诊断记录的测试集  $D_v$  上测试,相关的测试结果统计如附表 6 所示。可见该诊断系统在 20000 条历史记录中,仅通过 1 轮诊断即可得出准确缺陷的有 9355 条,占比 47%;有 19335 条记录能在 3 轮诊断之内得出正确的缺陷类型和缺陷位置,占比 97%;所有的记录均在 6 轮之内获得正确的诊断。

6 结论

本文提出了一种基于贝叶斯网络的配电开关设备缺陷诊断方法。贝叶斯网络以设备的物理模型为基本拓扑,并与采用数据驱动的手段在设备历史诊断记录中挖掘关联规则结合构建完整拓扑。针对诊断证据不足的问题,在单轮诊断算法的基础上提出了交互式多轮诊断流程,以用户的反馈补充诊断证据,更新诊断结果。最后通过两个实际案例展示了该诊断方法的有效性,并使用 20000 条历史诊断记录测试了诊断方法的性能,表明 97% 的故障场景均能在 3 轮以内得到准确的诊断结果,所有的记录均能在 6 轮之内得到准确的结果,验证了诊断方法的准确性及可靠性。

致谢:本文得到中国南方电网广州供电局电力试验研究院的资助,谨此致谢。

附录

附表 1 历史诊断记录样例

App. Tab. 1 Historical diagnosis results sample

诊断时间	设备型号	关联部件	故障类型	设备缺陷
2017/08/02	ABB 11kV 环网柜	电缆	电缆肘型头击穿	电缆绝缘损坏

附表 2 案例一第 1 轮诊断结果

App. Tab. 2 Results of first round diagnosis in case 1

设备缺陷 $y_d^{(0)}$	电缆局部放电	高压短路	电缆绝缘击穿	安装调试不当	未接地合闸
条件概率 $P(y_d^{(0)}   E^{(0)}, G_B, \Theta_B)$	0.250	0.175	0.112	0.04	0.03

附表 3 案例二第 1 轮诊断结果

App. Tab. 3 Results of first round diagnosis in case 2

设备缺陷 $y_d^{(0)}$	电缆局部放电	电缆绝缘击穿	高压短路	电缆室过热	母线绝缘子爆烂
条件概率 $P(y_d^{(0)}   E^{(0)}, G_B, \Theta_B)$	0.194	0.194	0.135	0.097	0.077

附表 4 案例二第 2 轮诊断结果

App. Tab. 4 Results of second round diagnosis in case 2

设备缺陷 $y_d^{(1)}$	保护机构 老化失效	电缆头安装 调试不良	柜体未接地合闸	电缆室过热	电缆连接故障
条件概率 $P(y_d^{(1)}   E^{(1)}, G_B, \Theta_B)$	0.158	0.131	0.099	0.048	0.046

附表 5 案例二第 3 轮诊断结果

App. Tab. 5 Results of third round diagnosis in case 2

设备缺陷 $y_d^{(2)}$	电缆头安装 调试不良	柜体未接地 合闸	电缆室过热	电缆连接故障	保护机构 老化失效
条件概率 $P(y_d^{(2)}   E^{(2)}, G_B, \Theta_B)$	0.263	0.158	0.061	0.017	0.017

附表 6 20000 条历史人工诊断记录回测结果

App. Tab. 6 Results of diagnosis performance test over 20000 records

诊断轮数	1	2	3	4	5	6
诊断正确	9355	16810	19335	19444	19732	20000
诊断正确率(%)	47	84	97	97	99	100
缺陷位置正确,类型错误	10645	3034	412	412	0	0
缺陷类型正确,位置错误	0	0	0	144	0	0
缺陷类型及位置均错误	0	156	253	0	268	0

参考文献 (References):

[ 1 ] Jadid S B, Jeyasurya B, Khaparde S A. Power system fault diagnosis expert system using PROLOG [ A ]. Fourth IEEE Region 10 International Conference TENCON [ C ]. Bombay, India, 1989. 778-781.

[ 2 ] Park Young Moon, Kim Gwang-Won, Sohn Jin-Man. A logic based expert system (LBES) for fault diagnosis of power system [ J ]. IEEE Transactions on Power Systems, 1997, 12 ( 1 ): 363-369.

[ 3 ] Cardoso G, Rolim J G, Zurn H H. Application of neural net-work modules to electric power system fault section estimation [ J ]. IEEE Transactions on Power Delivery, 2004, 19 ( 3 ): 1034-1041.

[ 4 ] Wen F S, Chang C S. Probabilistic approach for fault-section estimation in power systems based on a refined genetic algorithm [ J ]. IEE Proceedings - Generation, Transmission and Distribution, 1997, 144 ( 2 ): 160-168.

[ 5 ] Jong C G, Leu S S. Bayesian-network-based hydro-power fault diagnosis system development by fault tree transformation [ J ]. Journal of Marine Science & Technology, 2013, 21 ( 4 ): 367-379.

[ 6 ] Lo C H, Wong Y K, Rad A B. Bayesian network for fault diagnosis [ A ]. 2003 European Control Conference (ECC) [ C ]. Cambridge, UK, 2003. 1381-1386.

[ 7 ] Dasseni E, Verykios V S, Elmagarmid A K, et al. Hiding association rules by using confidence and support



- [ A ]. International Workshop on Information Hiding [ C ]. 2001. 369-383.
- [ 8 ] Li H, Wang Y, Zhang D, et al. Pfp: Parallel fp-growth for query recommendation [ A ]. ACM Conference on Recommender Systems [ C ]. 2008. 107-114.
- [ 9 ] Chi Y, Wang H, Yu P S, et al. Moment: Maintaining closed frequent itemsets over a stream sliding window [ A ]. Fourth IEEE International Conference on Data Mining (ICDM'04) [ C ]. Brighton, UK, 2004. 59-66.
- [10] Cheng J, Greiner R, Kelly J, et al. Learning Bayesian networks from data: An information-theory based approach [ J ]. Artificial Intelligence, 2002, 137 (1-2): 43-90.
- [11] Aliferis C F, Tsamardinos I, Statnikov A. HITON: A novel Markov Blanket algorithm for optimal variable selection [ A ]. AMIA 2003 Symposium Proceedings [ C ]. 2003. 21-25.

## Data-driven and interactive fault diagnosis of distribution switches

CHEN Guo-yan<sup>1</sup>, LI Jun-jun<sup>2</sup>, CHEN Ying<sup>2</sup>, MEI Sheng-wei<sup>2</sup>

- (1. Power Testing and Research Institute, Guangzhou Power Supply Bureau Co. Ltd. ,  
China Southern Power Grid, Guangzhou 100142, China;  
2. Department of Electrical Engineering, Tsinghua University, Beijing 100084, China)

**Abstract:** Efficient and accurate device fault diagnosis helps to improve the reliability and safety of the distribution network. Benefiting from the development of information technology, a large amount of data has been accumulated in sampling inspection, type test, and online monitoring, laying the foundation of data mining and intelligent diagnosis. To meet the requirement of intelligent operating and maintenance of the distribution network, a data-driven and interactive diagnosis method of distribution switches is proposed. Using the physical model, historical fault diagnosis results, on-line monitoring data and simulation results, the proposed method excavates the relationship between fault representation and potential defects and trains Bayesian network to realize equipment defect reasoning. Furthermore, the diagnosis results are adjusted according to the feedback of the test personnel, and an interactive diagnosis method of equipment fault is proposed, which improves the detection accuracy by dynamically modifying the reasoning model. Practical case tests illustrate the effectiveness of the proposed method.

**Key words:** distribution device; Bayes network; association rules