

# 基于配用电数据关联的中压配电网断线故障诊断方法

吴素我<sup>1</sup>, 张 焰<sup>1</sup>, 苏 运<sup>2</sup>

(1. 上海交通大学 电气工程系, 上海 200240; 2. 国网上海市电力公司, 上海 200437)

**摘要:** 基于配用电信息系统数据和关联规则算法, 提出一种诊断中压配电网分支线断线不接地故障的方法。通过分析相互关联的配用电信息系统数据, 提出基于数据特征选择的关联规则挖掘方法, 并通过卡方分裂算法将连续型特征量转换为布尔型特征量, 同时采用 MSAPriori 算法解决故障信息中的稀有项问题, 然后在此基础上应用 kule 准则消除冗余规则以形成约简的代表规则家族。以华东某地区配用电信息系统中的历史数据为依据进行实际算例分析, 结果说明所提出的方法能够大量减少无效挖掘, 显著提高效率和准确度, 适用于中压配电网断线故障的在线诊断。

**关键词:** 配电网; 数据关联; 特征选择; 布尔离散化; 断线故障诊断

**中图分类号:** TM 72

**文献标识码:** A

**DOI:** 10.16081/j.issn.1006-6047.2017.07.015

## 0 引言

中压配电网担负着为用户直接供电的重任。一直以来, 人们对各类短路故障进行了较多深入的研究, 在故障诊断及故障快速切除等方面取得了丰硕成果并已有有效地用于实际电网中。但对于中压配电网断线故障诊断, 尤其是分支线断线不接地故障的诊断的研究还较少, 缺乏有效的诊断方法。目前关于断线故障诊断的研究还主要集中在非分支线断线故障选线及定位方面。

文献[1]应用小波奇异性检测功能获取单相断线故障引起的正序电流暂态分量模极大值极性和大小, 以此进行故障选线, 并结合小波神经网络实现模极大值与故障点位置之间的映射关系, 进行故障定位。文献[2]提出了基于负序电压幅值的单相断线判据以及基于负荷监测仪的单相断线故障区域判定方法。文献[3]考虑同杆双回线可能出现两回线路参数不相等的情况, 研究了各种断线故障情况下的线路电压、电流的序量关系, 给出了断线故障后电气量的求解方法和结果, 可以作为双回线非全相运行状态下故障诊断的判据。文献[1-3]提出的诊断方法需要利用故障后电流暂态量和电压电流序分量, 由于暂态量成分复杂, 很难提取出有效信息, 并且目前的配用电信息系统基本不采集这 2 种类型的数据, 因此实用性不强。文献[4]采用梯形模糊数估计配变负荷变化情况, 当断线相电流幅度的减小超过设定值且满足电压条件时, 计算此单相线路电流, 通过分析某节点电流为零的可能性最大, 判断此节点与其

后节点之间发生了断线故障, 该诊断方法需要遍历配电网拓扑结构, 运算量较大、计算时间较长。文献[5]分析了电力线路断开时行波产生的机理以及断线故障初始行波及其传输特征, 探讨了利用行波原理测量断线故障距离的方法和可行性。文献[1]和[5]所提方法用于诊断断线故障的前提是必须有效检测到断线故障的发生, 然后才能进行故障选线或故障定位。而中压配电网馈线上的分支线发生断线不接地故障时, 馈线电流可能变化很小, 不足以让保护装置动作发出告警信号, 运行人员难以及时发现断线故障, 往往是待用户打电话报修才知道, 影响了对用户供电的可靠性。

随着电网信息系统建设的不断深入, 许多地区的供电侧和用电侧信息采集系统全部接入到综合数据平台, 积累了从配电到用电的大量运行数据, 这些数据中可能蕴藏着与断线故障诊断相关的、未知的、潜在的有价值的信息。如果能从海量数据中快速筛选出有用信息并挖掘出它们之间的潜在联系, 实现对线路故障发生时间及位置的准确判断, 则可以为电网运行人员快速抢修提供依据, 达到有效降低故障损失、提高用户供电可靠性的目的。数据挖掘作为一种处理海量数据的方法, 能够挖掘出数据中所包含的有用信息, 已被应用于多个领域, 近年来在配电网故障诊断<sup>[6-7]</sup>和变压器故障诊断<sup>[8-9]</sup>方面也得到了应用。

本文从 10 kV 配电网实际运行出发, 利用配用电信息系统数据, 提出基于数据的条件特征和结果特征; 通过对关键条件特征的选择和数据类型的转换, 达到满足关联规则算法的模型要求; 根据用户用电数据的特点, 采用 MSAPriori 算法对反映线路及设备运行状态的条件特征量进行分析, 得到 10 kV 配

电网分支线断线不接地故障诊断规则集,并对规则集进行约简以帮助电网运行人员对故障做出更为准确的诊断。最后通过实际算例,证明本文提出的方法适用于中压配电网分支线断线不接地故障的在线诊断。

## 1 配用电信息系统及其数据特征

华东某地区配电网目前投运的配用电信息系统及其输出数据类型如下。

**a.** 生产管理系统的输出数据类型:一次设备台账、电网地理拓扑、电网电气拓扑、工作票、停复、缺陷等生产业务,以及日计划、实时负荷、调度记录等数据模块。

**b.** 故障抢修管理系统的输出数据类型:故障区域、设备类型、故障电压、电站线路、故障原因、故障设备、抢修耗时、故障类型、故障归属、报修内容。

**c.** 供电电压监测系统的输出数据类型:电压幅值,采集目标有 B 类 35 kV 和 10 kV 专线用户电压、C 类 10 kV 用户,其中 B 类和 C 类采集点共 1000 多个,D 类用户不采集,并做出 3 类电压统计值。

**d.** 配变负荷监测系统的输出数据类型:10 kV 配变负荷电压、电流、有功、无功数据。

**e.** 调度自动化系统的输出数据类型:10 kV 馈线有功功率、无功功率、电流等潮流信息以及 10 kV 开关站、配电站、环网柜电压。

**f.** 配电自动化系统的输出数据类型:遥信、遥控、遥测、遥调。

**g.** 电能质量系统的输出数据类型:对暂态事件进行查询、显示,对稳态电能质量进行全面监测,包括电压、电流、谐波(25 次)、电压不平衡度、闪变等五大类电能质量指标。

**h.** 用户信息采集系统的输出数据类型:大用户的日分时冻结电量(峰、平、尖、谷),96 点三相有功、电压、电流;低压大用户的日分时冻结电量(峰、平、尖、谷),96 点三相有功、电压、电流;低压小用户、大用户的日分时冻结电量(平、谷)。

**i.** 负荷控制与管理系统的输出数据类型:采集负荷控制与管理系统的覆盖的 1000 多个用户的有功功率、电流、电压。

**j.** 营销业务管理系统的输出数据类型:业扩、电费核算、收费、计量资产、用户档案、窃电信息。

**k.** 95596 客服系统的输出数据类型:反映停电类型、停电线路、停电时间、频度等;停送电管理工作类数据;电网薄弱环节的数据;营业问题类数据;反映营业厅营业时间的数据。

**l.** 可靠性分析系统的输出数据类型:每起停电事件的时户数、用户数、停电性质、停电时间、送电时

间、持续时间、线路名称、设备原因、技术原因、责任原因、备注、事件类型、区域。

**m.** 线损系统的输出数据类型:10 kV 馈线统计线损、理论线损;公变台区统计线损。

配用电信息系统中与断线故障诊断相关的数据类型特征有以下几类。

**a.** 故障发生信号和告警信号:来自调度自动化系统,该类信号是判断故障发生的标志。

**b.** 配电网潮流数据:来自供电电压监测系统、配变负荷监测系统、调度自动化系统和配电自动化系统,包括 10 kV 馈线和配变的有功功率、无功功率、电压、电流数据。

**c.** 用户用电量数据:来自用户信息采集系统和负荷控制与管理系统的,包括中高压用户和低压大用户 96 点三相有功功率、电压、电流数据。

**d.** “四通”信号:来自配电自动化系统,本文以遥测为准,保护动作后,相应的量测量变为 0,则遥信信号正确,若无保护动作信号,视为该信号丢失。

**e.** 重合闸信号:来自配电自动化系统和调度自动化系统,若量测数据降低至 0 后又恢复,则自动重合闸成功,否则自动重合闸失败。

**f.** 故障描述数据:来自可靠性分析系统和故障抢修管理系统,包括每起停电事件的停电时间、送电时间、持续时间、线路名称、事件类型、故障区域等数据。

为提取断线故障诊断规则而需构建数学模型时,首先要确定输入特征变量,即规则的条件特征  $X$  和结果特征  $Y$ 。规则的结果特征  $Y$  为断线故障类型,条件特征  $X$  为断线故障的数据项表征。鉴于分支线断线不接地故障很难被察觉,保护不会动作,因此只考虑量测数据。同时鉴于目前配用电信息系统基本不采集暂态量数据,这里只考虑稳态量特征。显然分支线断线不接地故障会影响用户的正常用电,因此配电变压器和用户的电压、电流、有功功率和无功功率都会产生很大的变化<sup>[10-11]</sup>,同时变电站内 10 kV 馈线首段的电压、电流、有功功率和无功功率也会根据故障位置的远近发生相应的变化<sup>[2]</sup>,根据文献<sup>[12]</sup>可知电压、电流、有功功率和无功功率的变化率也是故障诊断的常用稳态特征。因此从配电网实际运行数据出发,综合已有研究,本文提出的条件特征包括:

$$A(x) = \left\{ U, I, P, Q, \frac{dU}{dt}, \frac{dI}{dt}, \frac{dP}{dt}, \frac{dQ}{dt} \right\} \quad (1)$$

其中,  $U$  为电压;  $I$  为电流;  $P$  和  $Q$  分别为有功和无功功率。

基于配用电信息系统及配用电信息系统输出数据类型,本文将数据分为 3 个层级,即 10 kV 馈线

层级、配电变压器层级和用户负荷层级,各层级均有如式(1)所述的8类特征,每类特征又都有a、b、c三相(即故障相与非故障相),共有72个条件特征量。为了计算各个候选特征权重,得到准确的关键特征选择结果,需要删去数据质量较差的特征量。在目前的配用电信息系统中,有功和无功功率数据的有效率较低且10 kV馈线电流一般只采集一相,因此不考虑有功和无功功率及其变化率的相关特征量,并且10 kV馈线层级的电流特征量只包括单相相电流和单相相电流变化率,最后得到32个候选特征量,包括10 kV馈线相电流、10 kV馈线相电流变化率、10 kV馈线a相母线电压、10 kV馈线a相母线电压变化率等。

## 2 考虑配用电数据特点的关联规则挖掘

断线故障的发生与设备维护水平、运行参数以及网络拓扑之间有着复杂的因果关系,这种关系在配用电信息系统大量的数据中表现为运行参数、负荷等数据项之间的关联,因此关联规则挖掘方法非常适合用于处理断线故障诊断。当采集到的历史运行数据足够覆盖可能出现的各种断线故障事件时,事件的数据项表征和故障类型之间常常表现为有一定概率支持的关联性,当这种关联性在类似的事件出现时,可以被用来判定可能的故障类型。简而言之,进行关联规则挖掘的目的就是对历史运行数据进行规则性的总结并利用规则对故障的快速诊断进行指导。

### 2.1 数据的关键特征选择

常规的关联规则挖掘算法是建立在以下客观前提假设之上的:条件特征中的各项具有近似的性质和作用,即重要性相同或者重要性相近。输入特征变量的选择合适与否,很大程度上决定着最终规则集的准确性,因此,首先要筛选出重要性相似的条件特征<sup>[13]</sup>。

本文采用的是RELIEF算法<sup>[14]</sup>进行特征选择。RELIEF算法是一种基于实例的特征选择方法,其基本思想是如果某个特征为关键特征,那么在同类样本中,该特征的值应相同;而在异类样本中,该特征的值应不同。权重取值的基本思想是从训练集 $D$ 中随机选择一个样本 $R$ ,然后从和 $R$ 同类的样本中寻找最近邻样本 $H$ ,称为Near Hit,从和 $R$ 不同类的样本中寻找最近邻样本 $M$ ,称为Near Miss,然后根据以下规则更新每个特征的权重:如果 $R$ 和Near Hit在某个特征上的距离小于 $R$ 和Near Miss的距离,则说明该特征对区分同类和不同类的最近邻是有益的,则增加该特征的权重;反之,如果 $R$ 和Near Hit在某

个特征的距离大于 $R$ 和Near Miss的距离,说明该特征对区分同类和不同类的最近邻起负面作用,则降低该特征的权重。为了将分支线断线不接地故障和分支线断线接地故障以及正常的负荷波动分离开,本文将分支线断线不接地故障取为同类样本,将分支线断线接地故障样本以及正常负荷波动取为异类样本,并对RELIEF算法的计算公式进行适当改进。

改进后的RELIEF算法执行流程如下。

步骤1 给定样本集合 $E=\{(\mathbf{x}_i, y_i)\} (i=1, 2, \dots, n)$ ,其特征集合为 $A(\mathbf{x})=\{a(j)\} (j=1, 2, \dots, d)$ 。

步骤2 对各特征的值进行归一化处理,设数据集 $X$ 由 $m$ 个特征变量、 $n$ 个样本组成,将 $X$ 中的样本按时间升序进行排列。设 $u_{i+1}$ 为故障后样本, $u_i$ 为故障前样本,则 $u'_i = \frac{u_{i+1} + 1}{u_i} (i=1, 2, \dots, m)$ 。

步骤3 设定各特征的初始权重 $\omega_j=0$ ;设对集合 $(\mathbf{x}, y)$ 的采样次数为 $S$ 。

步骤4 随机从 $E$ 中选取一个样本 $(\mathbf{x}, y)$ 。

步骤5 找到与样本 $(\mathbf{x}, y)$ 欧氏距离最近同类样本 $H(\mathbf{x})$ 和异类样本 $M_1(\mathbf{x})$ 、 $M_2(\mathbf{x})$ 。

步骤6 按式(2)计算权值:

$$\omega_j = \omega_j + \frac{1}{2} (|\mathbf{x}^{(j)} - M_1^{(j)}(\mathbf{x})| - |\mathbf{x}^{(j)} - H^{(j)}(\mathbf{x})| + |\mathbf{x}^{(j)} - M_2^{(j)}(\mathbf{x})| - |\mathbf{x}^{(j)} - H^{(j)}(\mathbf{x})|) \quad (2)$$

综上所述,同类样本与异类样本特征统计量差异越明显,RELIEF算法对该条件特征赋予权值越高,则该条件特征重要性越高。

### 2.2 特征量的布尔离散化

关联规则算法是针对离散型数据的,而条件特征量都是连续型数据,因此必须对各特征量进行离散化处理,将离散化后的连续属性变换成分类属性。此外,关联规则算法还要求数据是二元属性形式,因此为了避免每个特征量转换为二元属性后出现不必要的属性联系,还需要为每一个分类属性引入一个二元属性。

连续数值的离散化方法有很多,例如等宽或等频率法<sup>[15]</sup>、C4.5法<sup>[16]</sup>、熵值法<sup>[17]</sup>等。由于本文用到的数据样本带有类别信息,因此采用卡方分裂算法<sup>[18]</sup>来进行数据离散化。卡方分裂算法是一种基于卡方分布的监督型离散化方法。其采用自上向下的策略,把整个属性值区间当做一个离散的属性值,对数据进行升序排列,并依次在数据中插入一个断点,分别计算插入断点时的卡方值 $\chi^2$ ,然后选择出现最大卡方值的断点对该区间进行划分,把一个区间分为2个相邻区间,每个区间对应一个离散的属性值,该

划分一直进行下去,直到满足停止条件。卡方分裂算法的关键在于划分点的选取,因此置信水平的确定十分重要,检验的置信水平值太高可能导致过分离散化,而置信水平值太低则可能导致离散化不足,通常置信水平值设在 0.01~0.10 范围内。选取合适的置信水平值之后,就可以得到条件特征量的离散区间形如表 1 所示。表中  $I_1, I_2, I_3, \dots, U_{a1}, U_{a2}, U_{a3}, \dots, U_{b1}, U_{b2}, U_{b3}, \dots, U_{c1}, U_{c2}, U_{c3}, \dots$  可能相等,也可能不等。

表 1 条件特征离散区间

Table 1 Discrete intervals of conditional features

序号	特征量名称	区间 1	区间 2	区间 3
1	馈线电流	$(I_1, I_2)$	$(I_3, I_4)$	$(I_5, I_6)$
2	馈线母线 a 相电压	$(U_{a1}, U_{a2})$	$(U_{a3}, U_{a4})$	$(U_{a5}, U_{a6})$
3	馈线母线 b 相电压	$(U_{b1}, U_{b2})$	$(U_{b3}, U_{b4})$	$(U_{b5}, U_{b6})$
4	馈线母线 c 相电压	$(U_{c1}, U_{c2})$	$(U_{c3}, U_{c4})$	$(U_{c5}, U_{c6})$
⋮	⋮	⋮	⋮	⋮

然后再对每一个离散区间引入一个二元属性。经过数据布尔离散化处理后将连续型数值映射到布尔型数值上,满足关联规则挖掘对条件特征数据类型的要求后,开始关联规则的挖掘。

### 2.3 考虑稀有项的关联规则挖掘

利用关联规则可以对数据间的关系(相似性、特征性、差异、依赖、趋势、反常等)进行建模和描述,挖掘数据项中有意义的规律<sup>[5]</sup>。

关联规则挖掘的对象一般是事务数据集。关联规则可定义为:设  $T=\{t_1, t_2, \dots, t_k, \dots, t_m\}$  为事务数据库,  $t_k$  为  $T$  中的第  $k$  件事务,  $I=\{i_1, i_2, \dots, i_k\}$  是二元属性的集合,其中的元素称为项,记  $D$  为样本集合  $T$ ,对任意  $k, t_k \subseteq I; X$  与  $Y$  是  $I$  的子集,  $X \cap Y = \emptyset$ ,在事务数据集中寻找  $X$  与  $Y$  的关联关系,关联关系记为  $X \Rightarrow Y$ <sup>[6]</sup>。

关联规则里有 2 个重要概念:支持度和置信度。支持度表示  $X$  和  $Y$  这一事件组合在总事务记录中出现的概率,置信度是指出现了项集  $X$  的事务中  $X$  和  $Y$  这一事件组合出现的概率。

支持度的计算公式为:

$$S(X \Rightarrow Y) = \frac{|T|XUY \subseteq T, T \in D|}{|D|} \quad (3)$$

置信度的计算公式为:

$$C(X \Rightarrow Y) = \frac{|T|XUY \subseteq T, T \in D|}{|T|X \subseteq T, T \in D|} \quad (4)$$

关联规则挖掘的目的在于挖掘出可信且有意义的规则。最小支持度  $S_{\min}$  和最小置信度  $C_{\min}$  规定了支持度和置信度的阈值<sup>[7]</sup>,只有达到这 2 个阈值时才是强关联规则。

Apriori 算法<sup>[19]</sup>是一种最有影响的挖掘布尔关联规则的频繁项集算法,其核心思想是通过候选集生成和情节的向下封闭检测这 2 个阶段来挖掘频繁项集,然后再由频繁项集产生强关联规则。挖掘频繁项集主要分两步。

**a. 连接步:**首先简单统计所有含一个元素的项集出现的频率,并找出那些不小于最小支持度的项集,即一维频繁项集,然后开始循环处理根据第  $k-1$  步生成的  $k-1$  维最大项集产生  $k$  维候选项集,计算得到候选项集的项集支持度,将其与最小支持度比较,从而找到  $k$  维最大频繁项集。

**b. 剪枝步:**根据情节的向下封闭性,即如果一个项集的子集不是频繁项集,则该项集肯定也不是频繁项集来删除非频繁项集。

鉴于采集设备和通信通道的有限性,并不是所有电力用户的数据都能被存储,目前用户信息采集系统只采集中高压用户和低压大用户的用电数据,所以用户层级的数据项集出现的频率较低,这些不频繁出现在数据集里的项即被称为稀有项。当最小支持度阈值设置得过高时,就不能挖掘到有关这部分数据项集的强关联规则,但当最小支持度阈值设置得太低时,又会出现大量无用的关联规则<sup>[13]</sup>。因此,利用统一的最小支持度进行整个数据集的关联规则挖掘是不合理的。

针对配用电信息系统的实际情况,本文提出采用 MSApriori 算法来解决该问题。在 MSApriori 算法中,定义数据集合里的每一项都有其独有的最小支持度,记为  $MIS(i)$ ,项集的最小支持度由其包含的项的最小支持度表示,记为  $MS(A)$ ,规则的最小支持度由其所包含项集的最小支持度表示<sup>[13]</sup>。在该定义下,MSApriori 算法仍然采用 Apriori 算法逐级迭代的方式寻找频繁项集,只不过为了保持 Apriori 算法的向下封闭性,需要对项集按最小项支持度从小到大进行排序<sup>[17]</sup>。因此,MSApriori 算法可以在配用电信息系统数据采集频率不一致的情况下挖掘到有用的规则。

此外,为挖掘到配电网运行人员感兴趣的规则,本文只筛选结果特征为含有分支线断线不接地故障的规则。

### 2.4 数据挖掘后的冗余规则处理

随着样本数据集的不断增大,所生成的关联规则数量也不断增多,会产生大量无秩序且冗余的规则,而电网运行人员很难对这些规则进行分析和应用。如规则  $R_1$  为  $(a_1, a_2) \rightarrow b_1 (S_{\min}=10\%, C_{\min}=80\%)$ ,规则  $R_2$  为  $(a_1, a_2, a_3) \rightarrow b_1 (S_{\min}=10\%, C_{\min}=82\%)$ ,显然一个规则并没有比前一个规则提供更多的有用信息,只是由于随机波动导致置信度的提高<sup>[19]</sup>,因此

规则  $R_2$  是冗余规则。为了有效保留有用规则,同时便于电网运行人员利用关联规则作出决策,需要对规则集进一步处理,比如合并相似规则以及删除冗余规则。

本文采用 kulc 准则来删除冗余规则。kulc 准则中的 kulc 系数被用来对 2 个置信度做平均处理,即:

$$kulc(A, B) = \frac{1}{2} [\text{confidence}(A \rightarrow B) + \text{confidence}(B \rightarrow A)] \quad (5)$$

其中,  $\text{confidence}(A \rightarrow B)$  表示规则  $A \rightarrow B$  的置信度;  $\text{confidence}(B \rightarrow A)$  表示规则  $B \rightarrow A$  的置信度。

当规则的条件特征增加时,如果 kulc 系数没有增大,则可以认为条件特征和结果特征相关关系没有增强,该规则是冗余规则。在处理后的规则集中,满足最小支持度和最小置信度且规则形式最精简的规则可以被认为是代表规则。随着规则条件特征的增加,规则的置信度得到了有效提高,则认为该规则是代表规则的后裔规则。每条代表规则和其所有后裔规则形成代表规则家族<sup>[21]</sup>,这些代表家族构成完整的诊断规则集。

代表规则家族形式如下:

$$\begin{aligned} (I_1, I_2) &\rightarrow G_1 \quad C = m_1 \\ (I_1, I_2), (U_{a1}, U_{a2}) &\rightarrow G_1 \quad C = m_2 \\ (I_1, I_2), (U_{a1}, U_{a2}), (U_{b2}, U_{b3}) &\rightarrow G_1 \quad C = m_3 \\ &\vdots \end{aligned}$$

其中,  $G_1$  为断线故障类型;  $m_3 > m_2 > m_1$ 。

### 3 基于配用电数据关联的断线故障诊断流程

基于配用电数据的关联规则挖掘过程如图 1 所示。

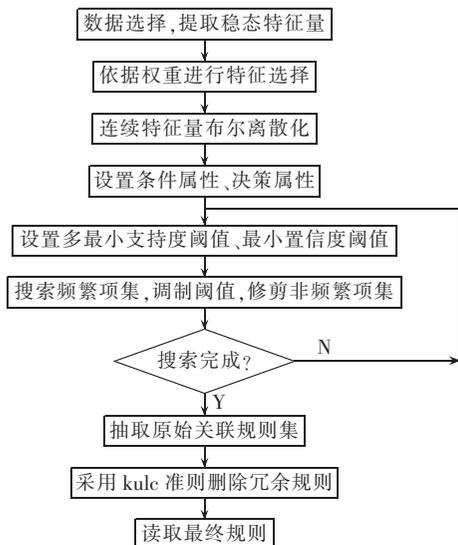


图 1 关联规则挖掘流程图

Fig.1 Flowchart of association rule mining

基于关联规则挖掘的断线故障诊断流程如图 2 所示。

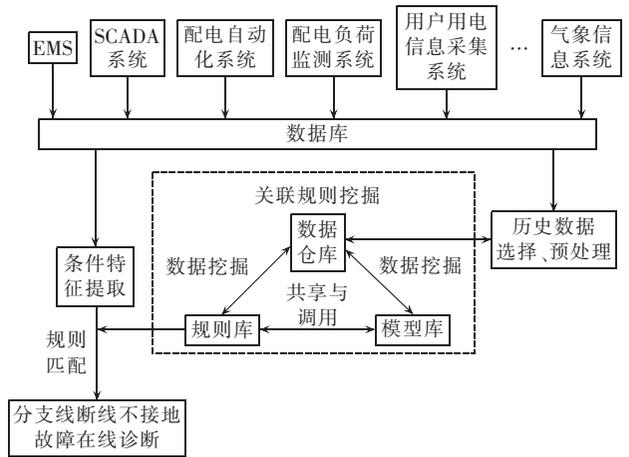


图 2 基于关联规则挖掘的断线故障诊断流程

Fig.2 Flowchart of open-line fault diagnosis based on association rule mining

来自调度 SCADA 系统/EMS 和配电网 DSCADA 系统/DMS 的电网运行实时数据及历史数据、配电自动化系统的遥信、遥控、遥测和遥调数据、用户用电信息采集系统的负荷数据、配变负荷监测系统的杆变和配变数据、气象信息系统数据等通过统一接口接入数据库。数据经过选择和预处理后存入数据仓库。模型库和知识库存关联规则挖掘所得的模型和规则,它们可以相互共享与调用。以候选条件特征量为基础,通过关键条件特征量选择和数据布尔离散化来建立关联规则诊断模型,采用合适的关联规则算法从历史数据中提炼得到分支线断线不接地故障的关联规则集,以此规则集作为分支线断线不接地故障诊断的判据。

根据采集的历史数据,利用离线挖掘出的规则确定当前情况下的故障类型以及规则置信度,用于指导运行人员及时发现故障。图 2 在线诊断模块具体步骤如下。

a. 实时读入 10 kV 馈线和馈线分支线上所有配电变压器以及配电变压器上所连用户的运行和负荷数据。

b. 根据 RELIEF 算法选出的数据关键特征处理原始数据得到关键条件特征。

c. 将步骤 b 结果与历史数据挖掘的关联规则中的条件特征进行匹配。为了缩小步骤 a 中的实时数据量,提高诊断速度,优先匹配馈线电流规则。如果未找到合适的规则进行匹配且出现了新的分支线断线不接地故障情况,则视为稀有故障,按照图 1 方法新增稀有故障的规则。需要注意的是,每次挖掘出新的规则后,需要更新规则集。

## 4 算例分析

### 4.1 算例描述

本文利用华东某地区近两年的配用电信息系统数据进行中压配电网分支线断线不接地故障诊断的算例分析。10 kV 馈线数据来自配电网调度自动化系统和配电自动化系统,10 kV 变压器数据来自配变负荷监测系统,中高压用户和低压大用户数据来自用户信息采集系统。经过筛选后得到近两年该地区的 320 条历史记录,以此建立含 300 条记录的挖掘库和含 20 条记录的测试库。鉴于多相断线故障发生的情况极少,历史故障数据也较少,因此只考虑单相断线故障。挖掘库中包含 100 条分支线断线不接地故障记录、100 条分支线断线接地故障记录和 100 条无故障运行记录。挖掘库和测试库两者结构完全一致,区别在于挖掘库包含故障类型信息,用于通过挖掘来获取规则,而测试库不包含故障类型信息,用于检验规则的有效性。

### 4.2 关键特征选取

以式(1)最后确定的 32 个特征量作为数据的条件候选特征,考虑到发生单相断线故障后,两非故障相变化幅度相似,因此把两相合为一相计算。

利用 RELIEF 算法得到的这些特征的权重如表 2 所示。

表 2 基于 RELIEF 算法的特征权重  
Table 2 Feature weights calculated by RELIEF algorithm

序号	候选特征	权重(均值)
1	配变故障相电流变化率	0.1421
2	配变故障相电压	0.1346
3	配变非故障相电压	0.1208
4	配变故障相电流变化率	0.1152
5	用户非故障相电流变化率	0.1109
6	用户故障相电流变化率	0.1037
7	馈线相电流	0.0988
⋮	⋮	⋮
21	馈线故障相母线电压	0
22	馈线非故障相母线电压	0

根据表 2 所示的结果,选取重要性相似即权重相似的条件特征量作为关键特征量,其中序号 1—7 的候选特征权重都在 0.1 左右,因此选择这 7 个候选特征。

### 4.3 关联规则挖掘

取挖掘库中的一条历史记录为例,具体说明物理含义,有名值样本如表 3 所示。

不断地在 0.01~0.10 之间修改置信水平,以进行交互式挖掘,最终将置信水平值设为 0.08。由于非故障相特征量权重基本相同,数据变化幅度基本相同,因此只需取一相进行数据离散化,最终得到各关键特征量的离散化分段区间如表 4 所示。

表 3 挖掘库中的某有名值样本  
Table 3 An actual sample in mining database

特征量	特征量值		
	分支线断线不接地故障	分支线断线接地故障	正常值
配变故障相电流/A	0.022	0.172	0.346
配变非故障相电流/A	0.116	0.197	0.363
配变故障相电压/V	106.1	354.1	223.6
配变非故障相电压/V	159.5	377.2	221.9
用户故障相电流/A	0.013	0.021	0.166
用户非故障相电流/A	0.155	0.136	0.175
馈线相电流/A	3.000	5.000	8.000

表 4 特征量离散化结果  
Table 4 Results of characteristic variable discretization

特征量	离散化结果		
	分支线断线不接地故障区间	分支线断线接地故障区间	正常区间
配变故障相电流变化率	(0.980,150.175)	(0.980,250.842)	(0,0.980)
配变故障相电压	(0,0.487)	(1.501,1.600)	(0.487,1.501)
配变非故障相电压	(0.434,0.732)	(1.632,1.832)	(0.732,1.632)
配变非故障相电流变化率	(0.826,52.473)	(0.826,15.519)	(0,0.826)
用户故障相电流变化率	(0.809,192.302)	(0.809,71.374)	(0,0.809)
用户非故障相电流变化率	(0.967,79.206)	(0.967,135.623)	(0,0.967)
馈线相电流	(0,0.800)	(0,0.800)	(0.800,1.099)

在挖掘库中增加 50 条不良记录,用于模拟故障记录中的信息丢失、信息错误和冗余信息,检验程序的健壮性。挖掘关联规则的各种算法的输出结果如表 5 所示,其中情况 1 为 300 条良好故障记录,情况 2、3、4 是在情况 1 的基础上分别增加 50 条部分信息丢失故障记录、50 条错误信息故障记录和 50 条冗余信息故障记录。根据各项出现频率及其他属性,设定最小项支持度,通过修改阈值进行交互式挖掘,以确定合适的阈值。在该算例中,确定的适合阈值如下:Apriori 算法中的最小支持度为 6%,最小置信度为 75%;在 MSApriori 算法中设定用户非故障相电流变化率和用户故障相电流变化率最小支持度为 0.3%,其他条件特征最小支持度为 6%,最小置信度为 75%。关联规则挖掘结果如表 5 所示。

从表 5 可以看出,利用 MSApriori 算法与 kulc 准则结合使挖掘结果中的冗余规则大幅减少,提高了挖掘效率和准确度。挖掘结果中的代表规则家族部分如下所示。

表 5 包含不良数据的挖掘结果对比  
Table 5 Comparison of data mining results,  
including bad data

情况	算法	规则数	最小置信度/%
1	Apriori	28	91.58
2		26	89.51
3		31	86.01
4		28	91.58
1	MSApriori	36	79.55
2		33	76.74
3		42	75.76
4		36	79.55
1	MSApriori+ kulc 准则	16	83.69
2		16	83.69
3		16	83.69
4		16	83.69

(1)代表规则家族 1。

a. 当馈线相电流落在区间(0,0.800)时,发生分支线断线不接地故障的置信度为 77.52%。

b. 当馈线相电流落在区间(0,0.800)、用户故障相电流变化率落在(0.809,192.302)时,发生分支线断线不接地故障的置信度为 92.59%。

c. 当馈线相电流落在区间(0,0.800)、用户故障相电流变化率落在区间(0.809,192.302)、配变故障相电压落在区间(0,0.487)时,发生分支线断线不接地故障的置信度为 100%。

(2)代表规则家族 2。

a. 当配变故障相电流变化率落在区间(0.980,150.175)时,发生分支线断线不接地故障的置信度为 83.69%。

b. 当配变故障相电流变化率落在区间(0.980,150.175)、配变非故障相电压落在区间(0.434,0.732)时,发生分支线断线不接地故障的置信度为 87.89%。

c. 当配变故障相电流变化率落在区间(0.980,150.175)、配变非故障相电压落在区间(0.434,0.732)、配变非故障相电流变化率落在区间(0.826,52.473)时,发生分支线断线不接地故障的置信度为 92.59%。

d. 当配变故障相电流变化率落在区间(0.980,150.175)、配变非故障相电压落在区间(0.434,0.732)、配变非故障相电流变化率落在区间(0.826,52.473)、馈线相电流落在区间(0,0.800)时,发生分支线断线不接地故障的置信度为 100%。

#### 4.4 在线故障实验

利用 MSApriori 算法产生的诊断规则对测试库的故障记录进行判断,考虑到测试库中的数据都是目前配用电信息系统能够采集到的数据记录,因此将测试库的数据看作实时数据。在进行关键特征量选取时,故障前后变化较为剧烈的条件特征量认为是故障相特征量,变化较为平缓的认为是非故障相

特征量。诊断结果如下。

(1)诊断结果 1。

a. 配变故障相电流变化率为 66.51667,配变非故障相电压为 0.677。

b. 关联规则:分支线断线不接地故障的置信度为 87.89%。

c. 诊断结果:分支线断线不接地故障,诊断正确。

(2)诊断结果 2。

a. 馈线相电流为 0.391842,用户故障相电流变化率为 24.04971。

b. 关联规则:分支线断线不接地故障的置信度为 92.59%。

c. 诊断结果:分支线断线故障,诊断正确。

(3)诊断结果 3。

a. 配变故障相电流变化率为 110.6467,配变非故障相电流变化率为 51.07875。

b. 关联规则:分支线断线不接地故障的置信度为 7.81%。

c. 诊断结果:分支线断线接地故障,诊断错误。

诊断结果 4—19 省略。

(20)诊断结果 20。

a. 配变故障相电流变化率为 61.74333,配变非故障相电流变化率为 76.01971,配变非故障相电压为 0.44003。

b. 关联规则:分支线断线不接地故障的置信度为 87.89%。

c. 诊断结果:分支线断线不接地故障,诊断正确。

由诊断结果 3 可知,当实时数据采集不全时,故障诊断的准确性受到很大影响,因此在线数据采集的充分性和准确性尤为重要。

在实时数据采集准确的基础上,在线诊断测试结果表明诊断规则的匹配率达到 90% 以上,并且在线诊断过程较为简单,计算速度很快,如果实时数据采集频率高、数据量大,可以采用并行计算的方法提高规则匹配速度。同时如果需要考虑新的断线故障类型,可以在线进行断线故障规则挖掘,该挖掘库的规则挖掘仅用时 2.3 s,满足在线诊断的速度要求。在将挖掘库的数据看作历史数据、测试库的数据看作实时数据的条件下,对测试库的有效诊断表明,本文的方法适用于实时断线故障诊断,可以为电网运行人员及早发现断线故障提供参考。

## 5 结论

在发生断线不接地故障特别是分支线断线故障后,由于电流、电压的变化幅度较小,保护不能及时动作,调度也不能收到告警信号,因此本文提出了一种利用配用电数据之间关联关系进行关联规则挖掘

的分支线断线不接地故障诊断方法,将故障后电流、电压隐藏的变化特征以一种显性的、相对直接的关联规则集的方式表现出来,以此作为调度运行人员诊断此类故障的参考依据。结合关联规则算法和配用电信息系统输出数据的特点,通过以下 3 个环节提高了诊断的准确性:

**a.** 候选条件特征里的弱相关特征在很大程度上影响着最终规则集的准确与否,利用 RELIEF 算法筛选出具有相似重要性的特征量;

**b.** 将连续型特征量进行布尔离散化后再进行关联规则分析,以此得到的规则能更准确地表述出该故障类型与特征量的值的关系;

**c.** 通过采用 MSAPriori 算法充分考虑了用户层级条件特征出现频率低的特点,因此能挖掘到有效的诊断规则。

仿真结果表明,经由上述 3 个环节完善后的断线故障诊断方法精度高且具有较高的容错性。

但是也需要注意以下 2 点。

**a.** 由于断线故障相较于短路故障发生的概率较小,导致历史数据有限。如果挖掘库样本量越大,能够诊断的故障模式越多,则挖掘到的关联规则越准确。

**b.** 数据挖掘方法是一个不断反馈的过程,当网络结构发生变化时,需要重新积累历史数据,挖掘新的规则,只有获取足够多新的挖掘库样本时,才可以进行新的分析。

## 参考文献:

- [1] 朱玲玲,张华中,王正刚,等. 基于小波神经网络单相断线故障选线和定位[J]. 电力系统保护与控制,2011,39(4):12-17.  
ZHU Lingling,ZHANG Huazhong,WANG Zhenggang,et al. Single-phase open-line fault selection and location based on wavelet neural networks[J]. Power System Protection and Control,2011,39(4):12-17.
- [2] 尤毅,刘东,李亮,等. 基于负荷监测仪的 10kV 架空线单相断线不接地故障区域判定[J]. 电力系统保护与控制,2012,40(19):144-149.  
YOU Yi,LIU Dong,LI Liang,et al. Detection method based on load monitoring for 10 kV overhead line single-phase ungrounded open fault area[J]. Power System Protection and Control,2012,40(19):144-149.
- [3] 屈刚,李长凯. 配电网架空线路断线故障定位[J]. 电力自动化设备,2005,25(12):35-38.  
QU Gang,LI Changkai. Fault locating of non-grounding line break in distribution network[J]. Electric Power Automation Equipment,2005,25(12):35-38.
- [4] 俞恩科,薛永端,李世强,等. 断线故障行波测距分析[J]. 电气应用,2015,34(3):121-125.  
YU Enke,XUE Yongduan,LI Shiqiang,et al. Analysis of fault traveling wave fault location[J]. Engineering Design,2015,34(3):121-125.
- [5] 李博通,李永丽,李洪波,等. 同杆双回线断线故障复合序网分析法[J]. 电力系统自动化,2010,34(9):76-80.  
LI Botong,LI Yongli,LI Hongbo,et al. New method for break-conductor fault analysis on double-circuit transmission lines based on interconnection of sequence networks[J]. Automation of Electric Power Systems,2010,34(9):76-80.
- [6] 李再华,白晓民,周子冠,等. 基于特征挖掘的电网故障诊断方法[J]. 中国电机工程学报,2010,30(10):16-22.  
LI Zaihua,BAI Xiaomin,ZHOU Ziguan,et al. Method of power grid fault diagnosis based on feature mining[J]. Proceedings of the CSEE,2010,30(10):16-22.
- [7] 聂倩雯,高玮. 基于关联规则数据挖掘技术的电网故障诊断[J]. 电力系统保护与控制,2009,37(9):8-14,19.  
NIE Qianwen,GAO Wei. A power network fault diagnosis method based on data mining association rules[J]. Power System Protection and Control,2009,37(9):8-14,19.
- [8] 郑元兵,孙才新,李剑,等. 变压器故障特征量可信度的关联规则分析[J]. 高电压技术,2012,38(1):82-88.  
ZHENG Yuanbing,SUN Caixin,LI Jian,et al. Association rule analysis on confidence of features for transformer faults[J]. High Voltage Engineering,2012,38(1):82-88.
- [9] 谢龙君,李黎,程勇,等. 融合集对分析和关联规则的变压器故障诊断方法[J]. 中国电机工程学报,2015,35(2):277-286.  
XIE Longjun,LI Li,CHENG Yong,et al. A fault diagnosis method of power transformers by integrated set pair analysis and association rules[J]. Proceedings of the CSEE,2015,35(2):277-286.
- [10] 马腾. 10 kV 配电线路断线故障检测与定位研究[D]. 济南:山东大学,2013.  
MA Teng. Line-break fault detection and location of 10 kV distribution line[D]. Ji'nan:Shandong University,2013.
- [11] 沈海滨,陈维江,王颂虞,等. 10 kV 架空线路雷击断线故障的模拟试验研究[J]. 电网技术,2011,35(1):117-121.  
SHEN Haibin,CHEN Weijiang,WANG Songyu,et al. Simulation tests on lightning stroke-caused wire-breakage of 10 kV overhead transmission line[J]. Power System Technology,2011,35(1):117-121.
- [12] 张沛超,谭啸风,杨珮鑫. 孤岛检测的关键特征识别及元学习方法[J]. 电力系统自动化,2014,38(18):72-78.  
ZHANG Peichao,TAN Xiaofeng,YANG Peixin. Critical feature identification and meta-learning method for islanding detection[J]. Automation of Electric Power Systems,2014,38(18):72-78.
- [13] 张小利. 多最小支持度关联规则挖掘的研究与应用[D]. 西安:西北大学,2005.  
ZHANG Xiaoli. Researches and applications on association rules mining with multiple minimum supports[D]. Xi'an:Northwest University,2005.
- [14] SUN Y,LI J. Iterative RELIEF for feature weighting[C]//Proceedings of the 23rd International Conference on Machine Learning. [S.l.]:ACM,2006:913-920.
- [15] 梁红旗. 数值属性离散化方法研究[J]. 信息技术,2008,10(5):99-101.  
LIANG Hongqi. Research on the discretization methods for numerical attributes [J]. Information Technology,2008,10(5):99-101.
- [16] 赵静娴,倪春鹏,詹原瑞,等. 一种高效的连续属性离散化算法[J]. 系统工程与电子技术,2009,31(1):195-199.

- ZHAO Jingxian, NI Chunpeng, ZHAN Yuanrui, et al. Efficient discretization algorithm for continuous attributes [J]. *Systems Engineering and Electronics*, 2009, 31(1): 195-199.
- [17] 李建强, 刘吉臻, 张栾英, 等. 基于数据挖掘的电站运行优化应用研究[J]. *中国电机工程学报*, 2006, 26(20): 118-123.
- LI Jianqiang, LIU Jizhen, ZHANG Luanying, et al. Rule mining in power plant operation optimization [J]. *Proceedings of the CSEE*, 2006, 26(20): 118-123.
- [18] 郑元兵, 孙才新, 李剑, 等. 变压器故障特征量可信度的关联规则分析[J]. *高电压技术*, 2012, 38(1): 82-88.
- ZHENG Yuanbing, SUN Caixin, LI Jian, et al. Association rule analysis on confidence of features for transformer faults [J]. *High Voltage Engineering*, 2012, 38(1): 82-88.
- [19] 赵洪英, 蔡乐才, 李先杰. 关联规则挖掘的 Apriori 算法综述[J]. *四川理工学院学报(自然科学版)*, 2011, 24(1): 66-70.
- ZHAO Hongying, CAI Lecai, LI Xianjie. Overview of association rules apriori mining algorithm [J]. *Journal of Sichuan University of Science & Engineering (Natural Science Edition)*, 2011, 24(1): 66-70.
- [20] LIU B, HSU W, MA Y. Mining association rules with multiple minimum supports [C]//*Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM, 1999: 337-341.
- [21] 邹翔, 王清毅, 肖明军, 等. 电力调度数据挖掘后处理方法的研究[J]. *小型微型计算机系统*, 2003, 24(6): 1054-1057.
- ZOU Xiang, WANG Qingyi, XIAO Mingjun, et al. Research on post-processing of data mining for electric power dispatching [J]. *Mini-micro Systems*, 2003, 24(6): 1054-1057.

#### 作者简介:



吴素我

吴素我(1991—),女,安徽黄山人,硕士研究生,主要研究方向为数据挖掘在故障诊断中的应用(E-mail: wusuwo26@163.com);

张焯(1958—),女,安徽合肥人,教授,博士研究生导师,博士,主要研究方向为电力系统规划、电力系统可靠性、大数据分析等;

苏运(1987—),男,山东德州人,工程师,主要研究方向为电力系统仿真、配用电数据分析。

## Open-line fault diagnosis based on data association of MV distribution network

WU Suwo<sup>1</sup>, ZHANG Yan<sup>1</sup>, SU Yun<sup>2</sup>

(1. Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

2. State Grid Shanghai Municipal Electric Power Company, Shanghai 200437, China)

**Abstract:** A method based on the data of its power distribution information system and the association rules is proposed for diagnosing the non-grounding open-line faults of MV distribution network. The correlated data of the electric power distribution information system are analyzed and a method of association rule mining based on the data feature selection is proposed. Chi square splitting algorithm is applied to convert the continuous features into Boolean features, the MSApriori algorithm is applied to solve the rare item problem of fault information, and kule principle is adopted to eliminate the redundant rules and form a reduced rule family. A practical analysis is carried out based on the historical data of a regional power distribution information system in East China, which shows that, since the invalid data mining is greatly reduced and the efficiency and accuracy are significantly enhanced, the proposed method is suitable for the on-line diagnosis of open-line faults of MV distribution network.

**Key words:** distribution network; data association; feature selection; Boolean discretization; open-line fault diagnosis