

DOI: 10.12158/j.2096-3203.2023.02.022

雷电引起的电压暂降严重程度自学习评估方法

王毅¹, 刘书铭¹, 唐钰政¹, 夏中原², 郑晨¹(1. 国网河南省电力公司电力科学研究院, 河南 郑州 450000;
2. 国网河南省电力公司, 河南 郑州 450000)

摘要: 雷击故障是造成电网电压暂降的主要原因之一, 准确评估雷电造成的电压暂降严重程度可以为制定最优治理方案和敏感用户选址提供依据。文中提出一种数据驱动的电压暂降严重程度自学习评估方法。首先, 基于雷电造成的电压暂降机理, 结合雷电定位系统和电能质量监测系统中的监测信息选取参与挖掘的参数; 其次, 减少离散化结果对规则准确性的影响, 使用离散化评价系数确定不同参数的离散区间数目; 然后, 针对电网数据库动态变化时挖掘算法效率过低的问题, 使用基于增量式学习的关联规则挖掘算法持续更新挖掘规则, 从而赋予其自学习的能力; 最后, 提出基于综合赋权法的加权欧氏距离评估实际场景的电压暂降严重程度。通过某地区电网的监测数据和 IEEE 30 节点系统仿真数据进行实证分析, 结果证明文中方法能在实际应用中准确挖掘有价值规则, 实现关注节点的电压暂降严重程度评估。

关键词: 雷击-暂降; 自学习; 电压暂降严重程度; 关联规则; 雷电定位系统; 电能质量监测系统

中图分类号: TM73

文献标志码: A

文章编号: 2096-3203(2023)02-0188-09

0 引言

电压暂降会造成巨大的经济损失, 已引起电网公司、用户和政府重视, 而雷电是引发电压暂降的主要原因之一^[1-2]。准确评估雷击造成的暂降事件严重程度, 可为制定最优治理方案和敏感用户选址提供依据。

传统电压暂降严重程度评估方法^[3-7]主要分为实测统计法^[8]、随机预估法^[9]和状态估计法^[10]。实测统计法是指在系统中选取部分站点进行在线监测, 并对监测数据进行分析计算以确定被测点的暂降水平, 但仅适用于系统拓扑和参数确定的情况, 不仅无法评估全系统的暂降水平, 而且评估准确性依赖于监测时间的长短。随机预估法则以随机建模为基础, 通过概率模型对电压暂降进行随机模拟, 不进行长时间在线监测也能实现全系统的暂降水平评估, 但随机预估法使用的历史统计资料受天气、检修情况等因素影响, 每年变化较大, 且如果系统改建或者扩建, 历史统计资料也将不再适用。状态估计法是实测统计和随机预估法的结合, 能利用有限的监测数据对一段时间内系统的暂降水平进行准确评估, 但监测装置配置方法对结果精度影响较大, 若不能保证全系统暂降可观, 则评估结果会产生较大误差。

收稿日期: 2022-10-17; 修回日期: 2022-12-21

基金项目: 国家电网有限公司科技项目“面向多元用户的电能质量智能感知及增值服务关键技术研究及试点应用”(202024211A)

近年来, 电网监测系统规模不断扩大, 其中存储的海量监测数据为电压暂降评估提供了新的方式。数据驱动的电压暂降评估方法的基本思想为: 基于电压暂降监测数据, 通过数据挖掘获取参数与暂降指标之间的关系, 实现系统电压暂降评估。文献[11]以电压暂降数据为基础, 通过 Apriori 算法挖掘规则并根据规则间的距离预测系统暂降水平; 文献[12]利用改进的 AprioriTid 算法和灰靶理论实现电压暂降事件评估; 文献[13]构建互信息网络模型对电压暂降进行数据挖掘, 并利用连接权提取关联规则构成电压暂降规则库。文献[11-13]不依赖于网络拓扑, 在历史数据库固定的情况下具有较高的实际应用价值。而事实上, 随着电网中数据的积累, 关联规则的时效性也在不断变化。上述算法并不适用于动态数据库, 在数据库变化时须重新构建并扫描数据库, 算法效率过低。

针对上述问题, 文中提出一种数据驱动的电压暂降自学习评估方法。其中“自学习”是指电压暂降评估方法能够在长期的使用过程中, 面对数据库的动态变化, 基于已经挖掘的关联规则实现规则库的更新, 从而提高算法效率和规则的准确性。首先, 基于雷电定位系统(lightning location system, LLS)和电能质量监测系统(power quality monitoring system, PQMS)中的实际数据, 形成雷击-暂降事件数据库; 其次, 基于 K-means 算法离散数据库中的连续数据, 并通过离散化评价系数确定聚类数量; 然后, 基于增量式学习的 AprioriTid 算法挖掘数据库

中参数与电压暂降严重程度之间的关联规则;最后,基于规则库实现系统内任意关注节点的暂降严重程度评估。

1 数据预处理

1.1 挖掘参数选取

当输电线路遭受雷击时,由于雷电流注入,节点电压迅速上升,当绝缘子串两端电压超过其闪络电压时,绝缘子串发生闪络,雷电流沿闪络通道瞬间接地,若建立起稳定电弧持续燃烧,就会导致接地故障,相关节点发生暂降。

雷电与电压暂降事件存在关联,但并非一一映射或简单的因果关系,因此,须对电网中的多元监测平台数据进行分析挖掘。LLS 可提供实时和历史雷击数据^[14],包括雷击位置、时间、峰值电流、极性、回击次数等。PQMS 存储了暂降信息^[15],包括暂降起止时间、位置、电压等级等。在中国,每次雷击跳闸事故都经过了个案分析^[16],通过观察受击杆塔的雷击点或电弧轨迹确定雷击位置。因此,根据线路跳闸记录和暂降事件记录,可以确定造成暂降的雷击参数,并以此获取单次雷击-暂降事件。

PQMS 可以获取暂降幅值和持续时间,结合节点设备类型对应的电压耐受曲线(voltage tolerance curve, VTC),使用 IEEE 1564 标准^[17]计算单次雷击-暂降事件的严重程度:

$$S_e = \frac{1 - V}{1 - V_{\text{curve}}(d)} \quad (1)$$

式中: $V_{\text{curve}}(d)$ 为暂降持续时间为 d 时 VTC 上对应的电压幅值标么值; V 为实际的电压暂降幅值标么值。此外,为评估系统内任何关注节点的暂降严重程度,选取雷击位置与关注节点之间的几何距离作为挖掘参数之一。值得注意的是,实际应用中可根据监测平台记录的参数和雷电造成电压暂降的机理来选取参与挖掘的参数。综上,文中基于 LLS 和 PQMS 选取的参数如表 1 所示。

表 1 选取的挖掘参数

Table 1 Selected parameters for mining

参数	符号	含义	来源
日期	C_1	雷击的日期	LLS
时间	C_2	雷击的时刻	LLS
雷击位置	C_3	雷击的地理位置	LLS
雷电流峰值	C_4	雷击时的最大电流	LLS
几何距离	C_5	雷击位置与关注节点的距离	LLS、PQMS
暂降严重程度	D	事件指标 S_e	PQMS

1.2 数据离散化

为减少数据挖掘过程的不确定性,提高决策规则的精度,有必要进行数据离散化。用更抽象的概念取代原数据层的数据进行数据挖掘,有利于规则的挖掘。

对于海量数据,目前有多种方法可提取数据中有价值的信息,聚类分析就是一种经典分析方法。文中使用 K -means 算法^[18-20]离散连续数据。对于原始数据集, K -means 算法会根据数据点之间的欧式距离将其划分为 k 个表征不同定性区间的簇。然而,该算法的离散区间个数 k 的选取依赖主观经验。为避免人为选取离散区间个数影响后续关联规则挖掘,文中选取 2 种常见离散化评价指标,即 Dunn 有效性系数和轮廓系数(silhouette coefficient, SC)^[21-22]以确定最佳聚类区间个数。

Dunn 系数的计算如下:

$$D(k) = \frac{\min_{a \in I_i, b \in I_j} d(a, b)}{\max_{a, b \in I_k} d(a, b)} \quad (2)$$

利用 K -means 算法将参数 C 的值域分成 k 个区间(I_1, I_2, \dots, I_k), $d(a, b)$ 表示取值 a 与 b 之间的距离。Dunn 系数的分子表示区间之间的离散程度,分母表示区间内部的紧密程度,即 Dunn 系数越大,离散效果越好。

SC 的计算步骤如下:首先,计算 a 与所在区间内其余数据的平均距离 $m(a)$,体现区间内紧密程度;然后,计算 a 到其他区间的距离的最小值 $n(a)$,体现区间之间的离散程度;最后,计算数据点 a 的 SC 值。

$$s(a) = \frac{n(a) - m(a)}{\max(m(a), n(a))} \quad (3)$$

SC 反映了单个数据的离散有效性, $s(a)$ 越大,离散结果越合理。将所有数据的 SC 平均值作为算法的离散评价指标。当 Dunn 系数和 SC 的计算结果出现冲突时,一般来说,出于模型简洁性的要求,离散得到的区间数越少越好。

2 电压暂降严重程度自学习评估方法

2.1 关联规则挖掘算法

AprioriTid 算法^[2]是一种经典的关联规则挖掘算法,其基本思想为:基于用户设定的阈值获取频繁项集并挖掘出关联规则。设 $M = \{t_1, t_2, \dots, t_n\}$ 为事件数据库,其中事件 t_i 是一组项目的集合,项目是指单次事件中单个参数的取值,这些项目的集合为 N ,对于项目集 $T \in N$,如果 T 中有 k 个项目,则又称 T 为 k 项集。对于 k 项集 T ,若其计算出的支

持度高于支持度阈值 S_{\min} , 则可称为频繁 k 项集。

频繁项集代表了一条关联规则, 关联规则是形如 $X \Rightarrow Y$ 的蕴涵式, 其中, $X \subseteq N, Y \subseteq N$, 且 $X \cap Y = \emptyset$ 。为反映 X 与 Y 之间的相关性, 用户设定置信度阈值来判断规则是否为强关联规则, 规则的支持度和置信度计算分别如式(4)和式(5)所示。

$$\text{sup}(X \Rightarrow Y) = \frac{|X \cap Y|}{|U|} \quad (4)$$

$$\text{conf}(X \Rightarrow Y) = \frac{|X \cap Y|}{|X|} \quad (5)$$

式中: $|X|$ 、 $|X \cap Y|$ 分别为参数集合 X 和 $X \cap Y$ 在原始数据库 M 中事件的个数; $|U|$ 为总的事件个数。支持度表征该事件出现在数据库中的频繁性, 置信度表征规则的可信性。

AprioriTid 算法提高了 Apriori 算法的运算效率, 在生成频繁 k 项集时使用 k 阶 TID 表代替原始数据库, 减少了扫描量。由于 k 阶 TID 表存储了生成的所有频繁 k 项集, 因此当 k 较小时, 表中存储的事件数量会高于原始数据库。为进一步提高算法效率, 文中使用决策表代替 TID 表, 每次扫描时只要删除决策表中不含频繁 k 项集的事件所在行即可得到新的决策表, 降低了扫描量。

$$J = \begin{matrix} N_o & V_{c1} & V_{c2} & V_{c3} & V_{c4} & V_{c5} & V_D \\ \begin{bmatrix} 1 & a_1 & b_1 & c_1 & d_1 & e_1 & f_1 \\ 2 & a_2 & b_2 & c_2 & d_2 & e_2 & f_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n & a_n & b_n & c_n & d_n & e_n & f_n \end{bmatrix} \end{matrix} \quad (6)$$

式中: N_o 为事件编号的标识符; V_{c_i} 为参数 C_i 取值的标识符; V_D 为参数 D 取值的标识符。决策表第一列为事件的编号, $a_i, b_i, c_i, d_i, e_i, f_i$ 分别为第 i 个事件中参数 $C_1, C_2, C_3, C_4, C_5, D$ 的取值。

2.2 增量式更新算法

实际电网中的数据是动态变化的, 关联规则的时效性也在不断变化。文中雷击-暂降事件数据库变化主要体现在随着时间的变化, 数据库中添加了新的事件。增量式更新算法就是在数据库发生变化的情况下, 基于已经获取的规则, 挖掘新的关联规则, 删除不满足条件的旧关联规则, 避免决策表的重复扫描和支持度的重复计算, 并实现关联规则库更新。

假设原始数据库为 M , 增加的事件数据库为 M_1 , 则更新的数据库为 $M+M_1$, 新的支持度为:

$$\text{sup}(M + M_1) = \frac{\text{sup}(M) \times |M| + \text{sup}(M_1) \times |M_1|}{|M| + |M_1|} \quad (7)$$

式中: $|M|$ 为原始数据库 M 中的事件个数; $|M_1|$ 为增加的数据库 M_1 中的事件个数。更新数据库时, 设数据库的频繁项集集合分别为 $L(M)$ 和 $L(M_1)$, 分为以下 4 种情况:

(1) 项目集 $T \in L(M)$, 且 $T \in L(M_1)$, 则 T 必是频繁项集。

(2) 项目集 $T \in L(M)$, 且 $T \notin L(M_1)$, 则扫描 M_1 计算 T 的支持度。

(3) 项目集 $T \notin L(M)$, 且 $T \in L(M_1)$, 则扫描 M 计算 T 的支持度。

(4) 项目集 $T \notin L(M)$, 且 $T \notin L(M_1)$, 则 T 不是频繁项集。

计算新增加的数据库 M_1 中各项目集 T 的支持度时, 根据 M_1 的大小计算 $\text{sup}(T)$ 。 T 是 M 中的频繁项集时, 如果 $\text{sup}(T) \geq S_{\min}$, 则 T 是数据库变化之后的频繁项集; 如果 $\text{sup}(T) < S_{\min}$, 则根据式(7)计算判断 T 是否是频繁项集。 T 不是 M 中的频繁项集时, 如果 $\text{sup}(T) \geq S_{\min}$, 则根据式(7)计算判断 T 是否是频繁项集; 如果 $\text{sup}(T) < S_{\min}$, 则 T 不是数据库变化之后的频繁项集。通过基于增量式学习的关联规则挖掘算法获取频繁项集后, 令电压暂降严重程度为 Y , 其余参数为集合 X , 规则形式为 $X \Rightarrow Y$, 其置信度计算公式如下:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{sup}(X \cap Y) \times (|M| + |M_1|)}{\text{sup}_M(X) \times |M| + \text{sup}_{M_1}(X) \times |M_1|} \quad (8)$$

式中: $\text{sup}_M(X)$ 、 $\text{sup}_{M_1}(X)$ 分别为集合 X 在数据库 M 和 M_1 中的支持度。

2.3 关联规则匹配

出于经济性的考虑, 实际电网存在未监测的节点, 在电网拓扑未知时无法评估电压暂降严重程度。为了实现实际情况下关注节点的电压暂降严重程度评估, 使用距离匹配规则获取评估结果。而实际场景的 X 不可能同筛选出的强关联规则完全一致, 须计算实际场景与规则库中规则之间的相似性, 得到相似性最高的规则作为电压暂降严重程度匹配结果。而 Y 受 X 中各个参数的影响程度存在差异, 因此匹配前须赋予 X 中的参数权重。基于主客观赋权法^[23]的优点, 使用层次分析法^[24-26]和熵权法^[27]相结合的综合赋权方法。

层次分析法是根据参数之间的重要程度按照一定标度求得权重。根据专家或用户的要求对参数的重要程度进行两两比较构建标度矩阵 R , 对矩阵 R 进行一致性检测, 根据 R 计算出参数 C_j 的权重 β_j :

$$\beta_j = \sqrt{\frac{\prod_{i=1}^5 r_{ji}}{\sum_{k=1}^5 \sqrt{\prod_{i=1}^5 r_{ki}}} \quad (9)$$

式中: r_{ji} 为矩阵 R 第 j 行第 i 列的元素。

熵权法根据模糊变换理论,得到模糊评判矩阵,从而计算参数 C_j 的熵值 e_j ,熵值越大则重要程度越低,因此使用 $1 - e_j$ 来计算参数的权重 w_j :

$$w_j = (1 - e_j) / \sum_{k=1}^5 (1 - e_k) \quad (10)$$

则参数 c_j 的综合权重为:

$$\alpha_j = \beta_j w_j / \sum_{k=1}^5 \beta_k w_k \quad (11)$$

得到综合权重后,采用加权欧式距离计算实际场景 x 与关联规则 r 之间的距离,距离越小则相似性越高,计算公式如下:

$$d(x, r) = \sqrt{\sum_{j=1}^5 \alpha_j (V_x(C_j) - V_r(C_j))^2} \quad (12)$$

式中: $V_x(C_j)$ 、 $V_r(C_j)$ 分别为实际场景 x 和关联规则 r 中参数 C_j 的取值。电压暂降关联规则挖掘流程如图 1 所示。

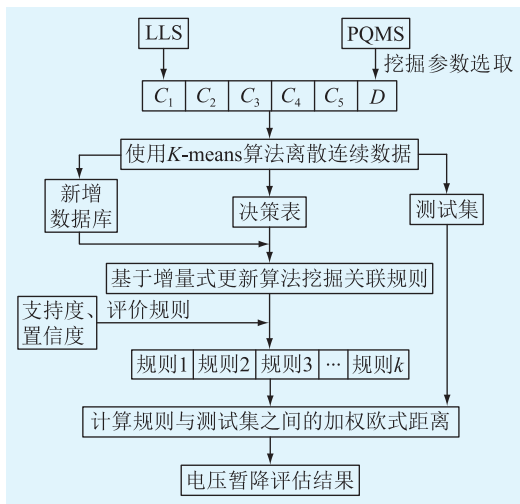


图 1 电压暂降关联规则挖掘流程

Fig.1 Flow chart of voltage sag association rule mining

3 实测算例分析

由于雷击具有季节性,以某地区电网 2016 年—2020 年的 6 月—8 月的 LLS 和 PQMS 实测数据为例,逐年分别匹配出 127、141、139、103、98 条雷击暂降事件。文中将 2016 年的事件作为原始数据库,2017 年—2019 年的事件作为增量训练集,2020 年的 98 个事件作为测试集 1 用于验证研究。为了验证规则在其他年份的数据集上的准确性,分别从 2016 年—2019 年的数据库中逐年抽取 20 条雷击暂降事件,共得到 80 条事件作为测试集 2 进行验证研究。

3.1 历史数据处理

文中选取的挖掘参数中,雷击位置以经纬度表示。按照雷击日期、时间、位置、雷电流峰值、电气距离和暂降严重程度的顺序,一条雷击-暂降事件可以记录为:2019-06-06,00:16:28.779,[102.520 4,28.384 5],-67.9 kA,35.78 km,1.129。雷击日期和时间须转化为数值型数据,文中将雷击日期以年为周期,转化为在 1—92 间变化的数值,将时间以每 10 min 为单位,转化为在 1—144 间变化的连续型数值,则“2019-06-06/00:16:28.779”的日期为 6,时间为 1.6。雷击位置,雷电流峰值、电气距离和暂降严重程度的数据类型为连续数值型数据。

使用 K-means 算法离散连续数据,在不同离散区间个数下,6 个参数的 Dunn 系数如图 2 所示。

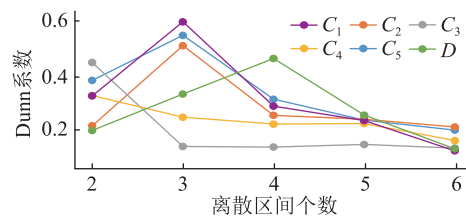


图 2 不同区间个数下参数的 Dunn 系数

Fig.2 Dunn of parameters at different number of intervals

在不同离散区间个数下,6 个参数的 SC 取值如图 3 所示。如图 2 和图 3 所示,基于 Dunn 系数和 SC 选取的离散区间数目一致。参数 C_1 、 C_2 、 C_3 、 C_4 、 C_5 、 D 的离散区间数量分别为 3、3、2、2、3、4。参数的离散结果见表 2。

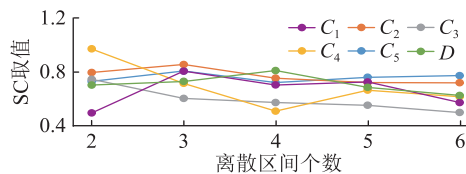


图 3 不同区间个数下参数的 SC 取值

Fig.3 SC values of parameters at different number of intervals

表 2 参数的离散结果

Table 2 Discrete results of parameters

参数	区间
C_1	6 月 ($V_{C1} = 1$); 7 月 ($V_{C1} = 2$); 8 月 ($V_{C1} = 3$)
C_2	00:00—08:00 ($V_{C2} = 1$); 08:00—16:00 ($V_{C2} = 2$); 16:00—24:00 ($V_{C2} = 3$)
C_3	[(26, 33), (99, 105)] ($V_{C3} = 1$, 南); [纬度, 经度] [(30, 33), (105, 108)] ($V_{C3} = 2$, 北)
C_4/kA	[-150, -100] ($V_{C4} = 1$); [-100, 150] ($V_{C4} = 2$)
C_5/km	[0, 35] ($V_{C5} = 1$); [35, 80] ($V_{C5} = 2$); >80 ($V_{C5} = 3$)
D	[0, 0.6] ($V_D = 1$, 优); [0.6, 1.0] ($V_D = 2$, 良); [1.0, 1.6] ($V_D = 3$, 中); [1.6, 2.0] ($V_D = 4$, 差)

3.2 关联规则挖掘

基于原始数据库和离散结果构建初始决策表,部分见表3。

表3 部分事件决策表

Table 3 Partial event decision table

事件	V_{C1}	V_{C2}	V_{C3}	V_{C4}	V_{C5}	V_D
x_1	1	1	1	1	1	1
x_2	1	1	1	1	1	1
x_3	1	1	1	1	2	1
x_4	2	2	2	2	2	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{125}	2	2	1	1	3	2
x_{126}	2	1	1	2	3	3
x_{127}	3	1	1	2	1	4

使用决策表替换 TID 表,通过 AprioriTid 进行关联规则挖掘,各规则支持度和置信度如图4所示。一般情况下,支持度和置信度越高代表规则出现越频繁且越可信。为了获取尽可能多的关联规则,文中结合规则的实际支持度和置信度设置最小支持度和最小置信度分别为 0.01 和 0.5,共获取 13 条关联规则,考虑到篇幅因素,仅给出部分规则,见表4。

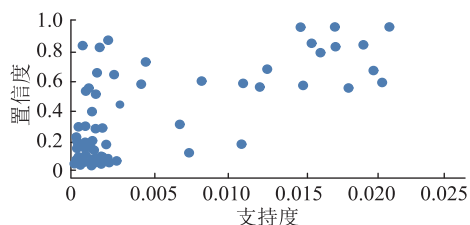


图4 规则的支持度和置信度

Fig.4 Support and confidence of the rules

表4 部分强关联规则

Table 4 Partial strong association rules

编号	规则	置信度
1	$V_{C1} = 1, V_{C2} = 1, V_{C3} = 1, V_{C4} = 1, V_{C5} = 1 \Rightarrow V_D = 1$	1
2	$V_{C1} = 2, V_{C2} = 1, V_{C3} = 1, V_{C4} = 1, V_{C5} = 2 \Rightarrow V_D = 2$	0.5
3	$V_{C1} = 3, V_{C2} = 2, V_{C3} = 1, V_{C4} = 1, V_{C5} = 3 \Rightarrow V_D = 1$	1
4	$V_{C1} = 1, V_{C2} = 1, V_{C3} = 1, V_{C4} = 2, V_{C5} = 3 \Rightarrow V_D = 1$	0.5
5	$V_{C1} = 1, V_{C2} = 1, V_{C3} = 1, V_{C4} = 1, V_{C5} = 1 \Rightarrow V_D = 4$	0.5
\vdots	\vdots	\vdots

对规则进行分析可以得到有价值的信息,如规则1表示:当雷电发生在6月,发生时间为00:00—08:00,位置在南,雷电流峰值为 $[-150, -100)$ kA,几何距离为 $[0, 35)$ km时,关注节点的暂降严重程度为优,置信度为1表明该条规则是完全可信的。将增量训练集按离散结果处理后加入原始数据库,设置最小支持度和最小置信度分别为 0.01 和 0.5,

使用 AprioriTid 算法和文中算法分别挖掘规则,算法消耗时间如表5所示。

表5 算法消耗时间对比

Table 5 Comparison of algorithm consumption time ms

数据库	文中算法	AprioriTid 算法
2016	265	568
2016—2017	354	682
2016—2018	558	1 010
2016—2019	859	1 382

如表5所示,文中使用决策表代替 TID 表,避免了项目集产生过多的问题。在数据库更新时,利用了已挖掘的关联规则,在已经挖掘的规则基础上实现关联规则更新,结果表明文中算法效率要明显优于传统算法。

3.3 关联规则匹配

设置最小支持度和最小置信度分别为 0.01 和 0.5,基于数据库 2016、2016—2017、2016—2018、2016—2019,使用文中算法分别挖掘出 13、22、32、40 条强关联规则。

基于工程经验,使用 1—9 标度的层次分析法计算得到几何距离、雷电流峰值、地理位置、日期、时间的权重分别为 0.39、0.27、0.17、0.11、0.06。而采用熵权法计算得到几何距离、雷电流峰值、地理位置、日期、时间的权重分别为 0.16、0.17、0.2、0.26、0.21。因此基于式(11)计算得到几何距离、雷电流峰值、地理位置、日期、时间的综合权重分别为 0.34、0.25、0.18、0.15、0.08。使用测试集1验证规则的准确性和有效性。由于篇幅限制,图5仅展示20个实际场景与规则匹配结果。

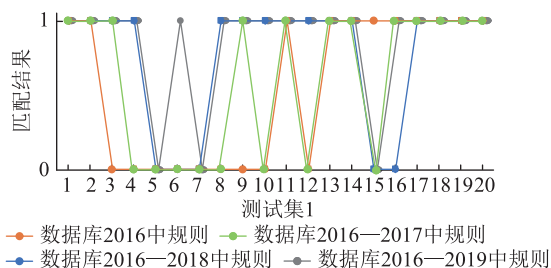


图5 基于测试集1的匹配结果

Fig.5 Matching results based on test set 1

图5中匹配结果为1代表匹配结果正确。在测试集1中,基于数据库2016、2016—2017、2016—2018、2016—2019获取的规则匹配准确率分别为55%、60%、75%和87%。因此,随着规则的增加规则匹配准确性也有所提升,在实际应用中可通过不断扩大数据库提高评估的准确性。

为了验证最终挖掘的规则在各个年份上的准确性,基于数据库 2016—2019 挖掘出的 40 条规则,使用测试集 2 进行验证,匹配结果如图 6 所示。

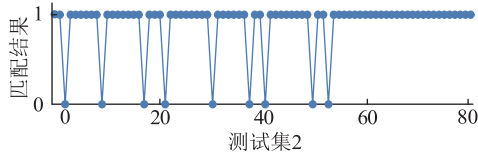


图 6 基于测试集 2 的匹配结果

Fig.6 Matching results based on test set 2

在测试集 2 中,基于数据库 2016—2019 获取的规则匹配准确率为 88.75%。因此,挖掘出的规则适用于评估该区域其他年份雷电-暂降事件的严重程度。

4 仿真算例分析

基于 Matlab 平台,在图 7 所示 IEEE 30 节点测试系统进行故障仿真和算法验证。该网络有 30 个节点,37 条线路和 6 台发电机,有 132 kV 和 33 kV 2 种电压等级;设网络中所有变压器的连接均为星形接地;设所有发电机的正序、负序和零序阻抗分别为 $j0.3$ p.u.、 $j0.2$ p.u.和 $j0.05$ p.u.;设所有发电机中性点通过阻抗 $j0.05$ p.u.;设该系统一年内有 4 种不同的运行方式,即 4 种不同的发电机运行计划,每种运行方式持续一个季度,如表 6 所示。

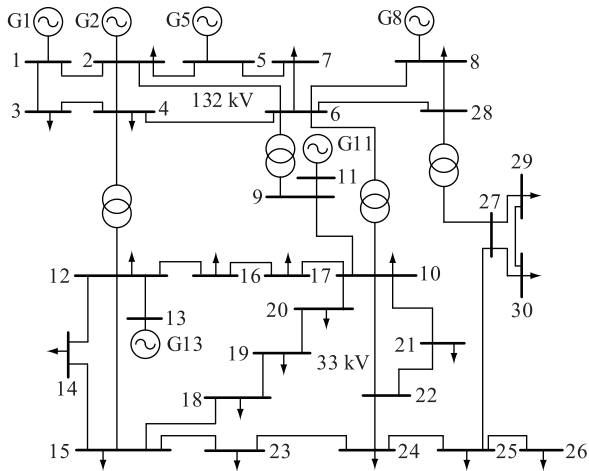


图 7 IEEE 30 节点测试系统

Fig.7 IEEE 30-node test system

4.1 蒙特卡洛模型

采用蒙特卡洛法模拟雷电活动参数,根据各因素变量的分布(如雷电流幅值概率密度分布模型、落雷位置均匀分布模型等)生成随机数,根据落雷位置在系统中随机产生若干短路故障,将其视为电力系统一段运行时间内的历史雷电造成的电压暂降事件,并将其作为算法输入样本。

表 6 一年内系统运行方式

Table 6 System operation mode in one year

发电机	一季度	二季度	三季度	四季度
G1	1	1	1	1
G2	1	1	1	1
G5	0	1	1	0
G8	1	0	1	0
G11	1	1	1	1
G13	1	0	1	0

注:表中“1”代表发电机正常运行;“0”代表发电机未投运。

(1) 雷电活动模拟。目前,雷电流幅值概率分布模型多以累积概率 $P(I)$ 来表示,国内防雷计算中雷电流幅值分布模型主要采用的有对数模型表达式和 IEEE Std 推荐模型^[28]。

$$P = 1/[1 + (I/\alpha)^\beta] \quad (13)$$

式中: α 、 β 为分布参数,体现雷电流幅值分布的差异性。通常 $\alpha = 31$, $\beta = 2.6$ 。通过计算机产生在 $[0, 1]$ 上均匀分布的随机数 t ,采用反函数法获得随机雷电流幅值。

$$I = \alpha(1/t - 1)^{1/\beta} \quad (14)$$

一般认为,雷电落雷位置是随机的。在单次雷击过程中,落雷位置在单位面积内可认为服从均匀分布。计雷击落雷位置为一定区域范围内随机产生的横纵坐标值,则在所研究空间范围内落雷位置 A 服从二维随机均匀概率分布。

(2) 系统故障模拟。选取电压暂降观测点为节点 10。整个网络模拟发生 4 000 起故障,4 种运行方式下各发生 1 000 起故障,得到故障样本 T_1 作为训练集。蒙特卡洛随机变量包括故障电阻 R_f 和故障类型。故障位置服从均匀分布; R_f 的取值从 0、1、2、...、10 Ω 中均匀随机选择;故障类型为单相接地故障,两相相间故障,两相接地故障,三相故障,根据 0.79、0.05、0.12、0.04^[29] 的比例随机产生。

4.2 关联规则挖掘

仿真样本的输入参数为雷击位置 C_3 ,雷电流幅值 C_4 ,几何距离 C_5 ,运行方式 C_6 ,故障类型 C_7 ,输出为电压暂降严重程度 D 。运行方式与季节相关,因此反映了雷电活动发生的时间信息。其中雷击位置 C_3 ,雷电流幅值 C_4 ,几何距离 C_5 和电压暂降严重程度 D 是连续型数据,使用 K -means 算法进行离散化,基于 Dunn 系数和 SC 确定 C_3 、 C_4 、 C_5 、 D 的最佳聚类区间分别为 2、2、4、2。具体的离散结果见表 7。

将 T_1 作为基础数据库,设置最小支持度为 0.01,最小置信度为 0.5,基于 AprioriTid 算法挖掘出 138 条关联规则。由于篇幅限制,仅列出部分规则,如表 8 所示。

表7 具体的离散结果

Table 7 Specific discrete results

参数	区间
C_3	南($V_{C3}=1$);北($V_{C3}=2$)
C_4/kA	$[0,150)$ ($V_{C4}=1$); $[150,400]$ ($V_{C4}=2$)
C_5/km	$[0,20)$ ($V_{C5}=1$); $[20,45)$ ($V_{C5}=2$); $[45,70)$ ($V_{C5}=3$); >70 ($V_{C5}=4$)
C_6	第一季度($V_{C6}=1$);第二季度($V_{C6}=2$); 第三季度($V_{C6}=3$);第四季度($V_{C6}=4$)
C_7	单相接地故障($V_{C7}=1$);两相故障($V_{C7}=2$); 两相接地故障($V_{C7}=3$);三相故障($V_{C7}=4$)
D	$[0,0.8)$ ($V_D=1$,优); $[0.8,2)$ ($V_D=2$,差)

表8 部分规则

Table 8 Partial rules

编号	规则	置信度
1	$V_{C3}=1, V_{C4}=1, V_{C5}=1, V_{C6}=1, V_{C7}=1 \Rightarrow V_D=1$	1
2	$V_{C3}=1, V_{C4}=1, V_{C5}=1, V_{C6}=1, V_{C7}=2 \Rightarrow V_D=2$	0.75
3	$V_{C3}=1, V_{C4}=2, V_{C5}=1, V_{C6}=2, V_{C7}=1 \Rightarrow V_D=1$	1
4	$V_{C3}=2, V_{C4}=1, V_{C5}=2, V_{C6}=2, V_{C7}=1 \Rightarrow V_D=1$	0.8
5	$V_{C3}=2, V_{C4}=1, V_{C5}=3, V_{C6}=3, V_{C7}=4 \Rightarrow V_D=2$	0.5
⋮	⋮	⋮

对规则进行分析可以得到有价值的信息,如规则1表示:当雷电位置在南,雷电流峰值为 $[0,150)$ kA,几何距离为 $[0,20)$ km,运行方式为第一种运行方式,故障类型为单相接地故障时,关注节点的暂降严重程度为优,置信度为1表明该条规则是完全可信的。规则4表示:当雷电位置在北,雷电流峰值为 $[0,150)$ kA,几何距离为 $[20,45)$ km,运行方式为第二种运行方式,故障类型为单相接地故障时,关注节点的暂降严重程度为优,置信度为0.8表明该条规则的可信度为80%。因此,文中方法也同样适用于大量样本及相应参数挖掘。

5 结论

文中充分利用现有监测系统中存储的雷电-暂降数据,以基于增量式学习的关联规则挖掘算法为基础,提出一种电压暂降严重程度自学习评估方法,不用网络拓扑便能获取关注节点的电压暂降严重程度,可以得出以下结论:

(1) 文中方法可以通过不断积累电压暂降事件,对关联规则库不断更新训练,其匹配准确率会随着规则数量的积累而提升,因而该方法具备自学习的能力。

(2) 克服了数据库动态变化和使用TID表计算时AprioriTid算法效率较低的缺点,利用已挖掘规则更新数据库,节省了算法计算时间。

参考文献:

- [1] 孟庆伟,高涵,贾志恒,等.基于上游正序参数比较的电压暂降源定位方法[J].电力系统自动化,2022,46(13):177-186. MENG Qingwei, GAO Han, JIA Zhiheng, et al. Voltage sag source location method based on upstream positive sequence parameter comparison [J]. Power System Automation, 2022, 46 (13): 177-186.
- [2] 刘书铭,王毅,张博,等.基于多元平台监测数据的电压暂降关联规则挖掘方法[J].电瓷避雷器,2021(5):135-143. LIU Shuming, WANG Yi, ZHANG Bo, et al. The mining method of association rules on voltage sag based on multi-platform monitoring database [J]. Insulators and Surge Arresters, 2021 (5): 135-143.
- [3] 栾乐,马智远,莫文雄,等.考虑不同敏感设备耐受特性的用户侧电压暂降严重程度区间评估方法[J].电力系统保护与控制,2021,49(2):140-148. LUAN Le, MA Zhiyuan, MO Wenxiong, et al. Voltage sag severity interval assessment method for user side considering tolerance characteristics of equipment of differing sensitivity [J]. Power System Protection and Control, 2021, 49(2): 140-148.
- [4] 胡文曦,肖先勇,李成鑫.考虑多维特征刻画的电压暂降严重程度评估方法[J].电网技术,2021,45(1):331-338. HU Wenxi, XIAO Xianyong, LI Chengxin. Voltage sag severity assessment method considering multi-dimension characterization [J]. Power System Technology, 2021, 45(1): 331-338.
- [5] 崔鑫.电压暂降严重程度综合评估方法研究[D].北京:华北电力大学,2019. CUI Xin. Research on comprehensive evaluation method of severity of voltage sag [D]. Beijing: North China Electric Power University, 2019.
- [6] 吴国诚,叶樊,梁帅伟,等.基于电压持续曲线的多次电压暂降严重程度评估方法[J].电力自动化设备,2018,38(2):182-191,200. WU Guocheng, YE Fan, LIANG Shuaiwei, et al. Evaluation method of multiple voltage sag severity based on voltage duration curves [J]. Electric Power Automation Equipment, 2018, 38 (2): 182-191, 200.
- [7] 肖先勇,谭亚欧,胡文曦,等.电压暂降系统指标的监测节点数量选择与评估方法[J].电力自动化设备,2020,40(10):8-14. XIAO Xianyong, TAN Yaou, HU Wenxi, et al. Monitoring node number selection and assessment method of voltage sag system index [J]. Electric Power Automation Equipment, 2020, 40 (10): 8-14.
- [8] 曾辉,苏安龙,葛延峰,等.考虑负荷特性的区域电网在线转动惯量快速估计算法[J].电网技术,2023,47(2):423-436. ZENG Hui, SU Anlong, GE Yanfeng, et al. A fast estimation algorithm for online rotational inertia of regional power grids considering load characteristics [J]. Grid Technology, 2023, 47 (2): 423-436.
- [9] 谢伟伦,薛峰,黄志威.基于网络传播特性的配电网电压暂降随机预估方法[J].电力系统保护与控制,2020,48(8):

- 163-171.
- XIE Weilun, XUE Feng, HUANG Zhiwei. Stochastic estimation method of voltage sags for a distribution network based on network propagation property [J]. *Power System Protection and Control*, 2020, 48(8):163-171.
- [10] 付锦, 丁蓝, 苟长松. 基于仿电磁学算法的电压暂降状态估计[J]. *电力系统保护与控制*, 2017, 45(10):98-103.
- FU Jin, DING Lan, GOU Changsong. Voltage sag state estimation based on electromagnetism-like mechanism [J]. *Power System Protection and Control*, 2017, 45(10):98-103.
- [11] 田世明, 卜凡鹏, 齐林海, 等. 电压暂降事件的频繁模式挖掘与知识推理分析[J]. *电力建设*, 2018, 39(5):21-27.
- TIAN Shiming, BU Fanpeng, QI Linhai, et al. Frequent pattern mining and knowledge reasoning of voltage sag events [J]. *Electric Power Construction*, 2018, 39(5):21-27.
- [12] 沈翔, 杨洪耕, 段晨. 基于灰靶理论与云模型的电压暂降事件数据挖掘分析方法[J]. *电网技术*, 2019, 43(2):722-731.
- SHEN Xiang, YANG Honggeng, DUAN Chen. An analytical method of data mining on voltage sag based on gray target theory and cloud model [J]. *Power System Technology*, 2019, 43(2):722-731.
- [13] 浦雨婷, 杨洪耕, 马晓阳. 基于数据挖掘与改进灰靶的电压暂降严重度分析与评估[J]. *电力系统自动化*, 2020, 44(2):198-206.
- PU Yuting, YANG Honggeng, MA Xiaoyang. Analysis and evaluation of voltage sag severity based on data mining and improved grey target theory [J]. *Automation of Electric Power Systems*, 2020, 44(2):198-206.
- [14] 王宇, 谷山强, 孟刚, 等. 雷电定位系统反演地闪回击电流的准确度受回击速度取值的影响[J]. *高电压技术*, 2021, 47(5):1617-1624.
- WANG Yu, GU Shanqiang, MENG Gang, et al. Effect of lightning return stroke speed on accuracy of inversed return stroke peak current using lightning location system [J]. *High Voltage Engineering*, 2021, 47(5):1617-1624.
- [15] 肖先勇, 胡誉蓉, 王杨, 等. 基于非同步电能质量监测系统的谐波状态估计[J]. *中国电机工程学报*, 2021, 41(12):4121-4132.
- XIAO Xianyong, HU Yurong, WANG Yang, et al. Harmonic state estimation based on asynchronous power quality monitoring system [J]. *Proceedings of the CSEE*, 2021, 41(12):4121-4132.
- [16] HE J L, WANG X, YU Z Q, et al. Statistical analysis on lightning performance of transmission lines in several regions of China [J]. *IEEE Transactions on Power Delivery*, 2015, 30(3):1543-1551.
- [17] IEEE guide for voltage sag indices: IEEE 1564-2014[S]. Institute of Electrical and Electronics Engineer, 2014.
- [18] 陈壮, 姜红, 郝丁成, 等. 基于K-means和簇内误差平方和的塑料快递包装袋X射线荧光光谱检验[J]. *激光与光电子学进展*, 2022, 59(11):489-495.
- CHEN Zhuang, JIANG Hong, HAO Dingcheng, et al. X-ray fluorescence spectroscopy inspection of plastic courier bags based on K-means and sum of squared intra-cluster errors [J]. *Advances in Lasers and Optoelectronics*, 2022, 59(11):489-495.
- [19] 雒明雪, 苑迎春, 陈江薇, 等. 基于邻域密度的K-means初始聚类中心优选方法[J]. *重庆理工大学学报(自然科学)*, 2021, 35(10):180-186.
- LUO Mingxue, YUAN Yingchun, CHEN Jiangwei, et al. An optimization method for initial clustering centers of K-means based on neighborhood density [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2021, 35(10):180-186.
- [20] 赵书强, 要金铭, 李志伟. 基于改进K-means聚类 and SBR算法的风电场场景缩减方法研究[J]. *电网技术*, 2021, 45(10):3947-3954.
- ZHAO Shuqiang, YAO Jinming, LI Zhiwei. Wind power scenario reduction based on improved K-means clustering and SBR algorithm [J]. *Power System Technology*, 2021, 45(10):3947-3954.
- [21] 孙林, 刘梦含, 徐久成. 基于优化初始聚类中心和轮廓系数的K-means聚类算法[J]. *模糊系统与数学*, 2022, 36(1):47-65.
- SUN Lin, LIU Menghan, XU Jiucheng. K-means clustering algorithm using optimal initial clustering center and contour coefficient [J]. *Fuzzy Systems and Mathematics*, 2022, 36(1):47-65.
- [22] 尹世庄, 王韬, 谢方方, 等. 基于互信息和轮廓系数的聚类结果评估方法[J]. *兵器装备工程学报*, 2020, 41(8):207-213.
- YIN Shizhuang, WANG Tao, XIE Fangfang, et al. Protocol clustering evaluation method based on mutual information and contour coefficient [J]. *Journal of Ordnance Equipment Engineering*, 2020, 41(8):207-213.
- [23] 戴国华, 戴睿, 张琪瑞, 等. 基于主客观赋权相结合的省级电网发展诊断分析方法与实证研究[J]. *电力系统保护与控制*, 2022, 50(2):110-118.
- DAI Guohua, DAI Rui, ZHANG Qirui, et al. Empirical study and analysis of provincial power grid development diagnosis based on the combination of a subjective and objective weighting method [J]. *Power System Protection and Control*, 2022, 50(2):110-118.
- [24] 张小莲, 李恒聪, 胡淇, 等. 多站融合选址综合评估研究[J]. *电力工程技术*, 2022, 41(2):53-59.
- ZHANG Xiaolian, LI Hengcong, HU Qi, et al. Comprehensive location evaluation of multi-station integration [J]. *Electric Power Engineering Technology*, 2022, 41(2):53-59.
- [25] 彭俊程, 贺英倩, 周华兵, 等. 基于模糊层次分析法的扶贫光伏电站类型选择[J]. *供用电*, 2021, 38(5):70-75.
- PENG Juncheng, HE Yingqian, ZHOU Huabing, et al. Photovoltaic poverty alleviation power stations type selection based on fuzzy analytic hierarchy process [J]. *Distribution & Utilization*, 2021, 38(5):70-75.
- [26] 邓铭, 黄际元, 吴东琳, 等. 基于层次分析法的“源-网-荷-

储”互动方案汇聚潜力评估[J]. 供用电, 2022, 39(9): 83-92.

DENG Ming, HUANG Jiyuan, WU Donglin, et al. Assessment of the convergence potential of the "source-network-load-storage" interactive scheme based on the analytic hierarchy process[J]. Distribution & Utilization, 2022, 39(9): 83-92.

[27] 马纪, 刘希喆. 基于序关系-熵权法的低压配网台区健康状态评估[J]. 电力系统保护与控制, 2017, 45(6): 87-93.

MA Ji, LIU Xizhe. Evaluation of health status of low-voltage distribution network based on order relation-entropy weight method[J]. Power System Protection and Control, 2017, 45(6): 87-93.

[28] 赵淳, 雷梦飞, 王剑, 等. 雷电流幅值累积概率分布曲线拟合方法[J]. 高电压技术, 2018, 44(5): 1598-1604.

ZHAO Chun, LEI Mengfei, WANG Jian, et al. Curve fitting me-

thod of cumulative probability distribution of lightning current magnitude[J]. High Voltage Engineering, 2018, 44(5): 1598-1604.

[29] WANG Y, LI S Y, XIAO X Y. Estimation method of voltage sag frequency considering transformer energization[J]. IEEE Transactions on Power Delivery, 2021, 36(6): 3404-3413.

作者简介:



王毅

王毅(1993), 男, 硕士, 工程师, 从事电能质量治理相关工作(E-mail: wangyi_cole@whu.edu.cn);

刘书铭(1984), 男, 硕士, 高级工程师, 从事电能质量治理相关工作;

唐钰政(1989), 男, 硕士, 高级工程师, 从事电能质量治理与变流器控制相关工作。

Self-learning estimation method for the severity of voltage sags caused by lightning

WANG Yi¹, LIU Shuming¹, TANG Yuzheng¹, XIA Zhongyuan², ZHENG Chen¹

(1. Electric Power Research Institute of State Grid Henan Electric Power Company, Zhengzhou 450000, China;

2. State Grid Henan Electric Power Company, Zhengzhou 450000, China)

Abstract: Lightning is one of the main causes of voltage sags in power grid. Accurate estimation of the severity of voltage sags caused by lightning can provide a basis for developing optimal management plans and siting sensitive users. In this paper, a data-driven self-learning estimation method for the severity of voltage sags is proposed. Firstly, based on the mechanism of voltage sags caused by lightning, the parameters involved in mining are selected by the monitoring information in lightning location system and power quality monitoring system. Secondly, the influence of discretization results on the accuracy of rules is reduced, and the number of discretization intervals for different parameters is determined by using discretization evaluation indexes. Then, to solve the problem of low efficiency of mining algorithm when the grid database changes dynamically, the incremental learning-based association rule mining algorithm to continuously update the mined rules, which gives it the ability of self-learning. Finally, a weighted Euclidean distance based on the integrated assignment method is proposed to evaluate the severity of voltage sags in real scenarios. The results of the empirical analysis by monitoring data of a regional power grid and simulation data of IEEE 30-node prove that the method in this paper can accurately mine valuable rules in reality and realize the severity assessment of voltage sags of the concerned nodes.

Keywords: lightning strike-voltage sag; self-learning; severity of voltage sags; association rule; lightning location system; power quality monitoring system

(编辑 方晶)