

加权频繁模式在电网告警频报信号分析中的应用

陈 昕

(中国能源建设集团江苏省电力设计院有限公司,江苏 南京 211102)

摘要:分析电网告警频报信号,可以找到与频报信号有强关联的影响因素,有效地发现可能产生频报信号的设备,因此提出一种基于加权频繁模式的频报信号分析方法。首先采用层次分析法(AHP)计算出频报信号中各影响因素的权重,再对频繁模式挖掘算法进行研究,提出加权改进FP-growth算法,最后将加权频繁模式应用到频报信号分析中,在挖掘出的频繁模式中突出权重大的影响因素。通过实例说明了该方法的可行性。

关键词:频发信号;层次分析法;加权频繁模式;加权改进FP-growth算法

中图分类号:TM73

文献标志码:A

文章编号:2096-3203(2018)02-0138-05

0 引言

随着电网规模的扩大,产生的告警信号也日益增长^[1]。由于远动系统、通信系统和测量设备在运行中的异常以及必要的设备检修,导致在上传的告警信息中还包含大量的告警频报信号^[2-3]。摒弃设备检修引起的的频报信号,剩余频报信号虽然不能反映电网的真实状态,但对其进行分析,可以找到引起告警频报的相关因素,从而发现可能产生频报信号的设备,并密切关注。

文中将数据挖掘领域中的频繁模式挖掘技术^[4]引入频报信号的分析中。由于频报信号影响因素具有不同重要性,对各影响因素赋权,应用加权频繁模式概念,提出结合加权模型的加权改进FP-growth算法,可以挖掘出与频发信号强关联的影响因素,继而可以找到可能产生频报信号的设备并加以密切关注,产生的频繁模式可以为设备检修提供规则知识。

1 电网告警频报信号分析方法

1.1 频报信号影响因素的选取

为了尽可能地找到产生频报信号的设备,需要归纳出告警信号频报的影响因素。对频报信号进行分析,设使用设备类型 X_1 、运行年限 X_2 、大修次数 X_3 和告警频报发生月份 X_4 为影响因素。其中,设备类型分为线路、母线、变压器、互感器、避雷器、断路器、隔离开关、电容器和电抗器等。运行年限和大修次数为离散值,需要进行离散化处理。按照文献[5]规定,设定运行年限小于等于10 a为短,大于10 a且小于20 a为中,大于等于20 a为长。设定大修次数小于等于2次为低,大于2次小于5次为中,

大于等于5次为高。告警频报发生月份即为告警报发时刻所在的月份。由于频报信号各影响因素具有不同的重要性,因此需要对各影响因素进行加权处理,权值越大表示影响因素越重要。

层次分析法^[6]建立在专家咨询基础上,把各影响因素的权重赋值简化为影响因素重要性的两两比较,然后进行数学处理,对各影响因素赋值,具有可信、灵活和实用的特点。考虑到告警频报信号的研究尚处于起步阶段,获得大量有明确结论的样本存在很大困难,需要借鉴专家经验。因此,文中采用层次分析法(AHP)求取影响因素权重。

邀请专家按表1要求给出各影响因素间两两比较的相对重要性,构成评判矩阵。

表1 评判矩阵标度及含义

Tab.1 Evaluation matrix scale and meaning

标度	含义
1	表示因素 u_i 与 u_j 比较,具有同等重要性
3	表示因素 u_i 与 u_j 比较, u_i 与 u_j 比稍微重要
5	表示因素 u_i 与 u_j 比较, u_i 与 u_j 比明显重要
7	表示因素 u_i 与 u_j 比较, u_i 与 u_j 比强烈重要
9	表示因素 u_i 与 u_j 比较, u_i 与 u_j 比极端重要
2,4,6,8	2,4,6,8分别表示相邻判断1-3,3-5,5-7,7-9的中值
倒数	表示因素 u_i 与 u_j 比较得判断 u_{ij} ,则表示因素 u_j 与 u_i 比较得判断 $u_{ji} = 1/u_{ij}$

根据评判矩阵计算权重,求出矩阵的最大特征值及其对应的特征向量,所求的特征向量即为权重分配。由于评估人不可能精确判断出权重的精确度,需要对判断矩阵进行一致性检验,通过一致性检验公式,表示如下:

$$C_R = C_I/R_I \quad (1)$$

其中: C_R 为评判矩阵的随机一致性比率; C_I 为

判断矩阵的一般一致性指标,由下式给出:

$$C_I = \frac{1}{n-1}(\lambda_{\max} - n) \quad (2)$$

R_I 为评判矩阵的一般一致性指标,对于 1-9 阶判断矩阵, R_I 值如表 2 所示。

表 2 n 阶评判矩阵 R_I 值
Tab.2 R_I value of evaluation matrix

n	R_I	n	R_I
1	0	6	1.24
2	0	7	1.32
3	0.58	8	1.41
4	0.9	9	1.45
5	1.12		

当 $C_r < 0.1$ 时,即认为判断矩阵具有满意的一致性,说明权重分配是合理的,否则需要调整判断矩阵,直至取得具有满意的一致性为止。

1.2 加权频繁模式

现有的频繁模式虽然在某种程度上发现了事务数据库中频繁出现的数据项,但这些数据项在数据库中的重要程度被看作是一致的,没有进行区分。这样不能体现各数据项具有不同的重要程度,也会因此挖掘出一些过于“平凡”的频繁模式,具有一定的局限性。针对这个问题,需要引入权重的概念,解决数据项重要程度不一致的问题,由此产生的频繁模式称为“加权频繁模式”。

加权频繁模式挖掘算法^[7]研究最初是围绕商品交易展开的,解决了往往优先考虑利润较高项目而忽略利润较低项目的问题。为了更好地满足用户需求,提出了一些新算法上的尝试^[8-14]。文中的研究对象是告警频报信号数据库,考虑各影响因素的重要性不同,引入加权频繁模式概念,提出将加权模型与改进 FP-growth 算法^[12]相结合来有效挖掘出与频报信号有强关联的影响因素。

$I = \{i_1, i_2, \dots, i_m\}$ 是由 m 个不同项目组成的集合, $D = \{T_1, T_2, \dots, T_n\}$ 是事务数据库,其中每个事务 $T_i (i = 1, 2, \dots, n)$ 包含事务唯一标识 T_{id} 和一个 I 的子集 X 。

定义 1: 项目集 $I = \{i_1, i_2, \dots, i_m\}$ 中每一项目 i_j 有一个权值 $w(i_j)$, 其中 $0 \leq w(i_j) \leq 1$ 。当项目具有权重后,其项目集 X 也具有相应的权重,记作 $W(X)$:

$$W(X) = \prod_{i \in X} w(i_j) / \sum_{i \in X} w(i_j) \quad (3)$$

当 $|X| = 1$ 时, $W(X) = w(i_j)$ 。 X 的加权支持度 $W_s(X)$ 定义为:

$$W_s(X) = W(X) \times S(X) \quad (4)$$

$S(X)$ 表示项目集 X 在事务记录中出现频率,即:

$$S(X) = \frac{X_{\text{count}}}{N} \quad (5)$$

式中: N 为事物总数。若 $W_s(X) \geq W_{s,\min}$, 其中 $W_{s,\min}$ 为最小加权支持度,则称 X 为加权频繁模式。当加权频繁模式 X 含有 k 个项目时,称为加权 k -频繁模式。

在告警频报信号影响因素频繁模式挖掘中,对各影响因素赋予不同权重,采用加权频繁模式挖掘算法挖掘出大量的加权频繁模式。这些频繁模式中只有一些是用户感兴趣的,通过由用户设定加权支持度阈值,可以控制频繁模式数量,其中不满足阈值的频繁模式不作为知识向用户提供。

1.3 基于加权模型的频繁模式挖掘算法

FP-growth 算法^[16-17]在当前挖掘频繁模式算法中应用最广,并且不需要候选集,大大节约了计算空间。但是,该算法也有一些不足。它的主要缺点是建树和挖掘过程都需要占用大量的内存。当数据库很大,或者数据库中的频繁 1-项集的数目很大时,运行速度将大为降低。更有甚者,由于无法构造基于内存的 FP-tree,该算法不能有效地工作。为了克服这些不足,文中使用改进 FP-growth 算法^[12],主要思想是在继承 FP-growth 算法不需要产生候选项集的优点的基础上,将数据库进行频繁 1-项集的项总数次扫描,每次扫描分别得到各个频繁 1-项集的项的数据库子集。然后分别对各项数据库子集使用 FP-growth 算法进行约束频繁项挖掘,得到含有各个频繁 1-项集的项的频繁模式,最后将这些频繁模式合并起来便得到整个数据库的所有频繁模式。

文中将加权模型与改进 FP-growth 算法相结合。加权改进 FP-tree 定义如下:

(1) 定义根节点的标记为 null,其子节点为项前缀子树集合,同时包含项头表。

(2) 子节点由 3 个域组成: item-name, count, node_link。其中 item-name 记录了该节点所代表的项目名字;count 记录了所在路径表中达到此节点的项目个数;node_link 指向下一个具有同样的 item-name 域的节点,要是不存在,就为 null。

(3) 项头表的结构是 item-name, W_s 和 node_link。其中 item-name 是项头表项目名称; W_s 记录了该项目集的加权支持度,按定义 1 计算得到;node_link 指向表中具有与该表项相同 item-name 域的第

一个节点。

文中构造的加权改进 FP-growth 算法如下:

输入:事务数据库 D , $W_{s,\min}$;

输出: D 中的频繁模式。

算法:

(1) 扫描数据库 D 一次,找出候选 1-项集的集合,计算它们的加权支持度。然后,按照加权支持度递减排列候选 1-项集的各项,得到候选 1-项集的集合 F 。将 F 中加权支持度小于 $W_{s,\min}$ 的项删除,得到频繁 1-项集的集合 L 。设 $L = \{I_1, I_2, \dots, I_m\}$, 其中, I_1 的加权支持度最高, I_m 的加权支持度最小。

(2) 再次扫描数据库 D ,将加权支持度小于 $W_{s,\min}$ 的项从各事务中删除,然后按照各项的加权支持度递减地将各事务中的项进行重新排列,得到数据库为 D' 。

(3) 根据频繁 1-项集 L 中的各项的加权支持度计数,按照以下规则由小到大依次构造各项的数据库子集,并利用 FP-growth 算法对其进行约束频繁项挖掘。

(4) 对于 L 中的每个项 $I_i (i = m, m-1, \dots, 1)$ 进行如下处理:

(a) 扫描数据库 D' ,从中提取所有含项 I_i 的事务,然后,删除这些事务中加权支持度小于该项的加权支持度的项,所得事务集合便为项 I_i 的数据库子集 D_i 。

(b) 对数据库子集 D_i ,利用 FP-growth 算法进行包含项 I_i 的约束频繁模式挖掘,其挖掘过程如下:利用数据库子集 D_i ,构造 FP-tree,并创建项头表 HT。构造 FP-tree 时,该数据库子集中各事务的项按照频繁 1-项集 L 中的次序处理。用项头表 HT 中的项 I_i 的加权支持度及其节点链信息,构造该项的条件模式基,然后构造其条件 FP-tree,就能在该条件 FP-tree 上挖掘出包含该项的频繁模式,完成在数据库子集 D_i 上的约束频繁模式 X_i 挖掘。

(5) 当 L 中所有的项的约束频繁模式 X_i 被依次挖掘出来后,合并这些约束频繁模式,即取这些约束频繁模式 X_i 的并集,便可得到数据库 D 的所有频繁模式,结束挖掘过程。

2 实例分析

为了验证文中所提方法的有效性,从福建省网采集了从 2013 年 1 月 1 日到 2013 年 6 月 30 日的告警数据,经过去噪处理后得到噪声数据中的告警频报信号。

(1) 对频报信号影响因素进行整理、形成和编号,形成告警频报信号记录表,存放在告警频报信号数据库中。

频报信号影响因素的整理:告警历史事项表中的告警事件可用,调度运行管理系统(outage management system, OMS)中检修记录和设备投运时间可用。

频报信号影响因素的形成:从告警事件中提取出设备类型和告警报发时间所在的月份,统计和计算得到大修次数和运行年限。

频报信号影响因素的编号:对频报信号影响因素按不同变电站编号,相同变电站的频报信号具有相同编号,对具有相同编号的记录进行频繁项集挖掘。

例如:某条告警频报信号为 500 kV 某条线路某开关,线路投运年限为 9 a,大修次数为 1 次,频报时间发生在 2 月。对于这样一条告警频报信号,预处理后形成告警频报信号记录表如表 3 所示。

表 3 告警频报信号记录表

Tab.3 Alarm frequency reported signal transaction table

编号	设备类型	投运年限	大修次数	告警月份
3	500 kV 线路断路器	短	低	二月

(2) 采用层次分析法(AHP)计算各影响因素(设备类型 X_1 、运行年限 X_2 、大修次数 X_3 和告警频报月份 X_4)的权重。

邀请专家按 AHP 要求给出各影响因素间两两比较的相对重要性,数据如表 4 所示($E_1 \sim E_4$ 分别表示 4 位专家, $X_1 \sim X_4$ 为影响因素代表符号)。

表 4 专家对各影响因素的权重分配表

Tab.4 Weights of every index assigned by experts

专家编号	X_1	X_2	X_3	X_4
E_1	1	1/2	1/3	2
E_2	2	1	1/2	3
E_3	2	1	1	3
E_4	1	1/2	1/3	1

即形成的评判矩阵:

$$W = \begin{bmatrix} 1 & 1/2 & 1/3 & 2 \\ 2 & 1 & 1/2 & 3 \\ 2 & 1 & 1 & 3 \\ 1 & 1/2 & 1/3 & 1 \end{bmatrix}$$

根据评判矩阵 W ,计算出权重并归一化,得到:

$$w = \{0.182\ 0, 0.314\ 1, 0.357\ 4, 0.146\ 5\}$$

利用式(1)、式(2)计算出 $C_R = 0.047\ 1 < 0.1$,即

认为评判矩阵具有满意的一致性,说明权重分配是合理的。

(3) 采用加权改进 FP-growth 算法对告警频报信号数据库进行频繁模式挖掘。

例如,由表 3 形成的项目如表 5 所示。

表 5 项目名称及其权重
Tab.5 Project name and their weights

项目名称	权重
500 kV 线路断路器	0.182 0
投运年限为短	0.314 1
大修次数为低	0.357 4
二月	0.146 5

假设 $W_{s,\min} = 0.5$, 扫描告警频报信号事务表, 得到表 5 中“二月”小于 $W_{s,\min}$, 删除项目“二月”, 由此得到表 3 中事务为 {500 kV 线路断路器, 投运年限为短, 大修次数为低}。由定义 1 和表 5 中各权重计算得到该事项的 $W_s(X) = 0.32 < 0.5$, 因此该事务不是频报信号影响因素的频繁模式。

利用加权改进 FP-growth 算法对告警频报信号数据库进行分析, 设定 $W_{s,\min}$ 为 0.2 时生成频报信号影响因素的部分频繁模式如表 6 所示。

表 6 告警频报信号影响因素的频繁模式(部分)
Tab.6 Frequency item sets of alarm frequency reported signal influence factors (part)

频繁模式	加权支持度
{投运时间为长}	0.81
{大修次数为高}	0.75
{投运年限为中, 大修次数为高}	0.62
{220 kV 线路断路器, 投运年限为长}	0.56
{500 kV 母线 PT, 投运年限为中, 大修次数为中}	0.48
{220 kV 刀闸, 三月}	0.31

通过频繁模式挖掘分析可以得到需要的规则知识, 如 {220 kV 线路断路器, 长} 说明某变电站投运年限超过 20 a 的 220 kV 线路断路器容易产生频报信号; {500 kV 母线 PT, 中, 中} 说明某变电站投运年限在 10~20 a 年之间以及大修次数在 2~5 次之间的 500 kV 母线上的电压互感器容易产生频报信号。这些规则可以指导运行人员根据设备运行检修记录, 找到可能会产生频报信号的设备并加以密切关注, 检查该设备运行状况是否会进一步恶化, 出现故障, 对运行人员的决策起参考作用。

3 结论

加权频繁模式与频繁模式相仿, 只是更进一步

地考虑到各数据项的不同重要程度, 对各数据项进行赋权。文中从数据智能处理的角度出发, 将加权频繁模式引入告警频发信号分析中, 提出加权模型与改进 FP-growth 算法相结合的加权改进 FP-growth 算法。该算法在计及各影响因素权重和继承 FP-growth 算法不需要产生候选项集的优点的基础上, 将整个数据库分解为子数据库进行频繁模式挖掘, 可以大大降低对告警频报信号数据库的搜索开销, 在时间和空间上都有很好的效率。

利用该算法产生的频报信号影响因素频繁模式可以发现与频报信号具有强关联的影响因素, 从而更加清楚地认识频报信号的产生。工程实现中, 频报信号影响因素频繁模式可以指导运行人员发现可能产生频报信号的设备并加以密切关注, 可以和设备检修形结合, 为设备检修提供规则知识。

参考文献:

- [1] 李云鹏. 智能告警专家处理系统在南通电网的应用[J]. 江苏电机工程, 2008, 27(5): 48-50.
LI Yunpeng. Application of intelligent warning expert system to nantong power system [J]. Jiangsu Electrical Engineering, 2008, 27(5): 48-50.
- [2] 欧阳永坚, 郭建, 鲁国刚. 变电站自动化系统通信去抖方法分析[J]. 电网技术, 2006, 30: 47-50.
OUYANG Yongjian, GUO Jian, LU Guogang. Analysis of subtractive dithering method for signals measurement in substation automation system [J]. Power System Technology, 2006, 30: 47-50.
- [3] 陈刚, 蔡泽祥. 变电站远动信息采集的分析和改进[J]. 继电器, 2003, 31(4): 73-75.
CHEN Gang, CAI Zexiang. Analysis and improvement of remote information collection in substations [J]. RELAY, 2006, 30: 47-50.
- [4] 匡向阳, 张玲. 基于 Hadoop 的 FP-Growth 关联规则并行改进算法[J]. 计算机应用研究, 2017, 35.
SHE Xiangyang, ZHANG Ling. Parallel improved algorithm of FP-Growth association rules based on Hadoop [J]. Application Research of Computers, 2017, 35.
- [5] DL/T 573-2010, 电力变压器检修导则[S].
DL/T 573-2010, Maintenance Guide for Power Transformers [S].
- [6] 王毅, 丁力, 侯兴哲, 等. 基于层次分析法的加权力线窃电检测方法[J]. 科学技术与工程, 2017, 17(33): 96-103.
WANG Yi, DING Li, HOU Xingzhe, et al. Weighted LOF stealing detection method based on analytic hierarchy process [J]. Science Technology and Engineering, 2017, 17(33): 96-103.
- [7] 耿汝年. 加权频繁模式挖掘算法研究[D]. 无锡: 江南大学. 2008.
KEN Runian. Research of weighted frequent pattern algorithm

- [D]. Wuxi: Jiangnan University. 2008.
- [8] 万 军. 加权关联规则挖掘算法的研究与改进[D]. 南宁: 广西大学, 2013.
WAN Jun. Research and improvement of weighted association rule mining algorithm[D]. Nanning: Guangxi University, 2013.
- [9] 陈 文. 基于FP树的加权频繁模式挖掘算法[J]. 计算机工程, 2012, 38(6): 63-65.
CHEN Wen. Mining algorithm for weighted frequent pattern based on FP-tree [J]. Computer Engineering [J], 2012, 38(6): 63-65.
- [10] 刘闻超, 施化吉, 马素琴. 加权模糊关联挖掘算法[J]. 计算机工程与设计, 2010, 31(16): 3654-3657.
LIU Wenchao, SHI Huaji, MA Suqin. Algorithm of weight fuzzy association rules[J]. Computer Engineering and Design, 2010, 31(16): 3654-3657.
- [11] 王 艳, 薛海燕, 李玲玲, 等. 一种改进的加权频繁项集挖掘算法[J]. 计算机工程与应用, 2010, 46(23): 135-137.
WANG Yan, XUE Haiyan, LI Lingling, et al. Improved algorithm for mining weighted frequent patterns [J]. Computer Engineering and Application, 2010, 46(23): 135-137.
- [12] 吴 俊, 李 珉, 惠 峻. 基于关联度分析的电网可靠性指标评价[J]. 江苏电机工程, 2015, 34(6): 82-84.
WU Jun, LI Min, HUI Jun. Network reliability evaluation based on association analysis [J]. Jiangsu Electrical Engineering, 2015, 34(6): 82-84.
- [13] 成乐祥, 季 丽. 基于加权K-means聚类遗传算法的变电站规划[J]. 2016, 35(6): 9-12.
CHENG Lexiang, JI Li. Substation planning based on weighted K-means cluster algorithm and genetic algorithm [J]. Jiangsu-Electrical Engineering, 2016, 35(6): 9-12.
- [14] 李彦伟, 戴月明, 王金鑫. 一种挖掘加权频繁项集的改进算法[J]. 计算机工程与应用, 2011, 47(15): 165-167.
LI Yanwei, DAI Yueming, WANG Jinxin. Improved algorithm for mining weighted frequent itemsets [J]. Computer Engineering and Application, 2011, 47(15): 165-167.
- [15] 许延祥, 曹军威, 许杏桃, 等. 基于FP-growth算法的电压事件干扰源定位方法[J]. 华东电力, 2014, 42(7): 1299-1304.
XU Yanxiang, CAO Junwei, XU Xingtiao, et al. A method of locating voltage disturbance sources based on FP-growth algorithm [J]. East China Electric Power, 2014, 42(7): 1299-1304.
- [16] 王新宇, 杜孝平, 谢昆青. FP-growth算法的实现方法研究[J]. 计算机工程与应用, 2004, 40(9): 174-176.
WANG Xinyu, DU Xiaoping, XIE Kunqing. Research on implementation of the FP-growth algorithm [J]. Computer Engineering and Application, 2004, 40(9): 174-176.

作者简介:



陈 昕

陈 昕(1989—), 女, 硕士, 工程师, 主要从事电力系统和新能源方面的工作(E-mail: cx15830207531)。

Application of Weighted Frequent Pattern in Alarm Frequently Reported Signal Analysis of Power Grid

CHEN Xin

(China Energy Engineering Group Jiangsu Power Design Institute Co., Ltd., Nanjing 211102, China)

Abstract: Since alarm frequently reported signal analysis of power grid can effectively find strong associated influence factors of the frequently reported signal and devices that probably generate frequently reported signals, a method of alarm frequently reported signal analysis based on Weighted Frequent Pattern is proposed. Firstly, Analytic Hierarchy Process (AHP) is used to calculate the weight of each influence factor. And then Doing research on frequent pattern mining and weighted improved FP-growth algorithm is presented. Finally, weighted frequent pattern applied into alarm frequently reported signal analysis that can dig out the frequent item which can highlight significant factors. The example illustrates the feasibility of this method.

Key words: frequency signal; AHP; weighted frequent item sets; weighted improved FP-growth algorithm

(编辑 杨卫星)