基于关联规则挖掘的电力生产安全事故事件关键诱因筛选

陈碧云[1],丁 晋[1],陈绍南[2]
(1. 广西大学 广西电力系统最优化与节能技术重点实验室,广西 南宁 530004;
2. 广西电网有限责任公司电力科学研究院,广西 南宁 530023)

摘要:随着智能电网、通信网络及电力生产安全事故事件分析水平的提高和发展,电力生产安全事故事件数据量快速增长、复杂性不断增大,逐步构成了电力生产安全事故事件大数据。为在先验事故事件大数据的基础上高效、可靠地对事故诱因进行分类和识别,基于关联规则挖掘进行电力生产安全事故事件关键诱因筛选。根据事故事件的特点,建立电力生产安全事故诱因分析体系,对不同类型的事故进行布尔离散化,并基于关联规则挖掘提出事故诱因的诱发度计算方法,运用 Apriori 算法进行深度关联规则挖掘,并根据强关联规则对关键诱因进行筛选和分析。以某区域近 5 年的事故实例分析验证了该方法的有效性。

关键词:电力生产安全事故事件;诱发度;关联规则;Apriori 算法
中图分类号:TM 732 **文献标识码**:A DOI:10.16081/j.issn.1006-6047.2018.04.010

0 引言

随着能源互联网及电网智能化的发展,电力生产过程状态所涉及的信息具有来源丰富、信息量大、信息种类多、结构复杂等典型的大数据特征[1]。传统的以事故学理论[2]为基础的电力生产事故事件分析法,具有较强的经验性和主观性,已难以适应大数据的安全控制决策的要求。因此,由大数据驱动的电力生产安全风险态势感知将是未来电力生产安全预警和控制决策的关键技术之一。

影响电力生产安全风险的影响因素众多,包括设备、环境、人员、气象等,各类影响因素经过诱发、转化、耦合形成具有强联系性的故障诱因集,并在生产过程中演化为大量的多诱因事故场景,造成不同严重程度的事故后果。因而,从海量历史电力生产安全事故报告中筛选出关键事故诱因,是态势感知的信息处理第一阶段,也是电力生产安全风险状态和发展趋势预测及控制决策的重要基础。

电力系统对信息技术的依赖日益增强,从大数据中提取隐含的、未知的、具有非平凡及潜在应用价值的信息或模式的数据挖掘技术在电力系统尤其是在典型的大数据场景中的应用也日益广泛。随着智能用电海量数据的涌现,必须寻找新的分析方法以满足用电大数据分析和预测的要求。文献[3-4]分别对海量数据背景下短期负荷预测方法和负荷特性特征进行了分析;文献[5]对用户侧的大数据来源进行了分析,并提出了一种电力用户侧大数据分析处理平台;文献[6-7]对用电用户进行分类,并在此基础上用数据挖掘的方法对用户行为进行分析;文献[8]构建了家庭用户用电信息的数据维度模型,并进行了挖掘分析。此外,大数据技术除了能对对象进行有效分类及模式识别外,还可挖掘潜在且有用的新知识。文献[9]在数据挖掘技术的框架下,研究了区域暂态电压稳定评估的体系和方法;文献[10]构建了含分布式电源(DG)系统的精确运行模型,采用数据挖掘技术进行在线孤岛检测。在众多数据挖掘的方法中[11],关联规则挖掘能够客观分析影响因素与发生行为之间的隐藏关系,因而更加符合电力系统运行、生产的实际,相较其他数据挖掘方法,其应用更广泛,常用于进行因果关系的分析。文献[12-13]总结分析了变压器状态的影响变量,以关联规则挖掘为基础,研究了对变压器状态进行客观评价的方法;文献[14-15]研究了关联规则挖掘方法在电网故障诊断中的应用。此外,关联规则挖掘还应用于发电厂站内设备的运行状态分析中[16-19]。

但当前数据挖掘技术及关联规则在电力系统中的应用仅局限于典型的大数据场景,关联规则挖掘仍主要面向实际元件,而忽略了潜在的大数据对象。事实上,影响电力生产安全的事故诱因无序、复杂且多维,是潜在的大数据数据源,并且事故诱因与事故的致因关系恰好是典型的因果关系,对事故诱因进行数据挖掘与关联规则分析,对保障电力生产安全、进行生产风险点预控具有重要意义。为此,本文通过关联规则挖掘,从历史事故报告中提取对电力生产安全性造成威胁的关键诱因,为电力生产的安全风险感知及防控提供依据。

1 电力生产安全事故诱因定义及分类

根据中国南方电网公司颁布的《中国南方电网有限责任公司电力事故事件调查规程(Q/CSG 210020—2014)》对已发事故进行的分析和总结,可将电力生产安全事故事件分为特别重大事故、重大

收稿日期:2017-04-12;**修回日期**:2017-12-30
基金项目:国家自然科学基金资助项目(51767002)
Project supported by the National Natural Science Foundation of China(51767002)

事故、较大事故、一般事故、一级事件、二级事件、三级事件、四级事件、五级事件 9 类。从历史经验中可发现,电力生产安全事故事件的诱因多达数百项,通过对具体诱因的总结和归类,主要可将其分为生理心理性人为因素、行为性人为因素、物理性物理因素、化学性物理因素、环境因素和管理因素六大类(一级诱因),共 54 个子类(二级诱因)(如图 1 所示),涉及三级诱因总共 295 种。

但在上述事故诱因中,有些诱因虽在各类电力生产安全事故事件中频繁出现,但其并非事故事件的主要原因,这些诱因属于与事故事件具有弱关联性的因素。因此,这一类诱因在电力生产安全感知和防控中并不作为关键的感知和防控对象。为降低电力生产安全风险防控的成本和提高电力生产安全风险防控的效率,有必要从大量的事故诱因中提取关联性最强的关键诱因。

定义某一事故诱因对诱发某类电力安全事故事件的作用大小为“诱发度”,并用于提取电力生产安全事故事件的关键诱因。

2　基于关联规则的事故诱因诱发度计算

2.1　关联规则的定义

关联规则是一种用于寻找数据项之间隐藏关系、描述数据之间密切程度的搜索式方法,其挖掘的对象一般是事务数据集。设有 N 个事务集构成的事务数据集 $D=\{T_1,T_2,\cdots,T_N\}$,其中 $T_j(j=1,2,\cdots,N)$ 称为事务数据集 D 的事务集,构成事务集 T_j 的元素称为事务集的项 $I_k(k=1,2,\cdots,p,p$ 为该事务集中项的数量)。D 中所有项的集合为项集 $I=\{I_1,I_2,\cdots,I_m\}$,m 为项的总数。若 I 中包含 K 个项,则 I 为 K 项集,显然 $T\subseteq I$。

概念 1　项集与频繁项集。设事务集 $M=\{i_1,i_2,\cdots,i_t\}$,则 M 为事务数据集 D 的一个 t 项集,其支持度 $V_{\mathrm{Sup}}(M)$ 定义为 D 中包含 M 的事务数所占的百分比,在概率学中可表示为:

$$V_{\mathrm{Sup}}(M)=\frac{|\{M\subseteq T,T\subseteq D\}|}{|D|} \tag{1}$$

若 $V_{\mathrm{Sup}}(M)$ 不小于给定的最小支持度阈值 ε_0,则称 M 为频繁项集。

概念 2　定义在 I 和 D 上的关联规则可表示为 $X\Rightarrow Y$ 的蕴含式,X 为前提,Y 为结论,并且满足 $X\subseteq I,Y\subseteq I,X\cup Y\neq\varnothing$,其成立条件是符合支持度和置信度的要求。

概念 3　关联规则的支持度 $V_{\mathrm{Sup}}(X\Rightarrow Y)$ 用于衡量关联规则在整个数据集的统计重要性,定义为项集 $(X\cup Y)$ 的支持度,即:

$$V_{\mathrm{Sup}}(X\Rightarrow Y)=V_{\mathrm{Sup}}(X\cup Y)=\frac{|\{(X\cup Y)\subseteq T,T\subseteq D\}|}{|D|} \tag{2}$$

概念 4　关联规则的置信度 $V_{\mathrm{Con}}(X\Rightarrow Y)$ 用于衡量关联规则的可信程度,定义为包含 X 和 Y 的事务数与包含 X 的事务数之比,即:

$$V_{\mathrm{Con}}(X\Rightarrow Y)=\frac{V_{\mathrm{Sup}}(X\cup Y)}{V_{\mathrm{Sup}}(X)}=\frac{|\{(X\cup Y)\subseteq T,T\subseteq D\}|}{|\{X\subseteq T,T\subseteq D\}|} \tag{3}$$

概念 5　关联规则的相关度 $V_{\mathrm{Cor}}(X\Rightarrow Y)$ 用于判断强关联规则是否具有偶然性或抑制性,以剔除误导规则。若某强关联规则的相关度为 1,则该规则所反映的频繁项集具有独立性,缺乏必然联系;若某强关联规则的相关度小于 1,则该规则所包含的频

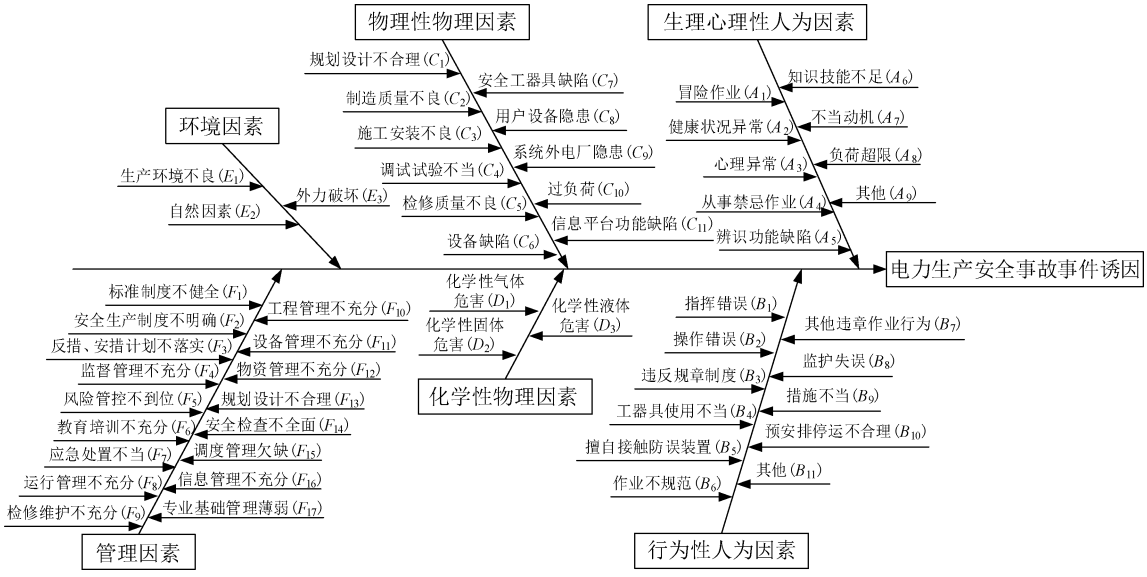

图 1　电力生产安全事故事件诱因结构

Fig.1　Incentive composition of power production safety accidents

繁项集具有抑制性。相关度的具体定义为：

$$V_{\mathrm{Cor}}(X\Rightarrow Y)=\frac{V_{\mathrm{Sup}}(X\cup Y)}{V_{\mathrm{Sup}}(X)\times V_{\mathrm{Sup}}(Y)} \tag{4}$$

2.2 电力生产安全事故事件数据的布尔离散化

关联规则分析均是针对离散的布尔型数据，且数据内需要确定先导和后继关系，因此，必须对事故事件数据进行预处理，即生成包含“事故类型”与“事故诱因”的布尔型离散事故数据。目前，电力生产安全事故事件基础数据主要来源于专家调查组撰写的事故事件分析报告，事故类型和事故诱因均隐含在事故报告中，需要对不同事故的信息进行抽象综合。从历史电力生产安全事故事件报告中，可以甄别出某一事故事件的事故类型和诱因，将布尔矩阵中对应项置1，反之置0，对大量的历史事故事件报告进行挖掘，即可得到映射后的布尔型事故事件信息矩阵 $\boldsymbol{T}^*$，如图2所示。

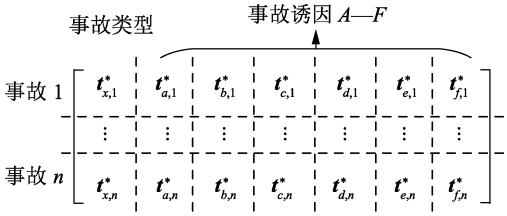

图2 电力生产安全事故事件信息矩阵

Fig.2 Information matrix of power production safety accident

图中，$\boldsymbol{T}^*$ 的元素 $\boldsymbol{t}^*_{i,j}$ 为第 j 个事故事件对应第 i 个事故信息的布尔值，需要注意的是，对于每一个事故事件的事故信息（包括事故类型和事故诱因），$\boldsymbol{t}^*_{i,j}$ 为由该类事故信息的所有类别组成的一个行向量。

2.3 事故诱因诱发度计算

在进行电力生产安全事故诱因诱发度计算前，首先要确定关联规则中各类集合的涵义，根据电力生产安全事故事件的特点，可做如下定义：

事务数据集 $D=\{$电力生产安全事故事件$\}$

事务集 $X_i=\{$第 i 类一级事故诱因作用下诱发的电力生产安全事故事件$\}$

项集 $X_{i,j}=\{$第 i 类一级事故诱因中第 j 个二级诱因作用下诱发的电力生产安全事故事件$\}$

项集 $Y_k=\{$第 k 类电力生产安全事故事件$\}$

电力生产安全事故中某关联规则 $X_{i,j}\Rightarrow Y_k$ 的支持度、置信度、相关度可计算得到：

$$V_{\mathrm{Sup}}(X_{i,j}\Rightarrow Y_k)=\frac{\sigma(X_{i,j}\cup Y_k)}{\sigma(D)} \tag{5}$$

$$V_{\mathrm{Con}}(X_{i,j}\Rightarrow Y_k)=\frac{\sigma(X_{i,j}\cup Y_k)}{\sigma(X_{i,j})} \tag{6}$$

$$V_{\mathrm{Cor}}(X_{i,j}\Rightarrow Y_k)=\frac{\sigma(X_{i,j}\cup Y_k)\times\sigma(D)}{\sigma(X_{i,j})\times\sigma(Y_k)} \tag{7}$$

其中，$\sigma(\cdot)$ 为计数函数。

对于二级事故诱因，其关联规则的置信度可直接反映该因子对各类电力生产安全事故事件的影响程度，因此二级事故诱因的诱发度即为其对应的置信度。而在电力生产安全事故事件中，事故的发生一定存在一级诱因和二级诱因的致因作用，X_i 和 $X_{i,j}$ 与 D 存在唯一的层级关系，通过对某一一级诱因下的各二级诱因诱发度进行加权求和，即可得到该一级诱因的诱发度：

$$w_i=\frac{C_{i,1}+C_{i,2}+\cdots+C_{i,j}+\cdots}{\sigma(C_{i,j})} \tag{8}$$

其中，w_i 为第 i 个一级诱因的诱发度；$C_{i,j}$ 为第 i 个一级诱因中第 j 个二级诱因；$\sigma(C_{i,j})$ 为第 i 个一级诱因中二级诱因的个数。

随着事故分析水平、体系的不断提高和完善，可获得的事故诱因将越来越多，势必会造成关联规则分析及事故机理揭示的复杂度大幅提升，甚至会因为算法复杂度过高而无法计算。因此在关联规则挖掘前，首先要根据诱发度对事故诱因进行筛选，以降低事故诱因维度，进而提高关联规则挖掘的准确性。

3 Apriori 算法

Apriori 算法是挖掘布尔关联规则频繁项集的重要算法，基于两阶段频繁项集思想的递推算法，利用逐层搜索的迭代方法来获得频繁项集，首先找到频繁1项集的集合，记为 L_1，L_1 用于搜寻频繁2项集的集合 L_2，依此类推，直至搜索完所有频繁项集。其中包括如下2个重要概念[20]。

概念6 k 阶频繁项集 L_k。其为由 k 个支持度大于最小支持度阈值 ε_0 的项组成的集合，如频繁2项集形式为 $L_2=\{\{i_1,i_2\},\{i_1,i_3\},\{i_2,i_4\},\cdots\}$。

概念7 k 阶候选项集 G_k。其由 $k-1$ 阶频繁项集 L_{k-1} 和 L_1 通过“并”运算形成。

Apriori 算法的核心步骤如下。

a. 扫描数据库 D，计算各个单项集的支持度，得到频繁1项集的集合。

b. 连接。通过将 L_{k-1} 中的每个元素执行连接，产生 L_k 的候选 k 项集的集合 G_k。

c. 剪枝。由于 G_k 是 L_k 的超集，其元素中也有非频繁的，利用 Apriori 性质，频繁项集的所有非空子集均是频繁的，缩小 G_k 的范围，确定 L_k 频繁项集的挖掘。

d. 再次单趟扫描数据库 D，计算各个项集的支持度，去掉不满足支持度的项集。

重复上述步骤直到搜索完所有频繁项集。最后，再计算 L_k 中项集的可信度，若该项集的可信度

大于给定的最小置信度阈值 δ_0，则为符合条件的关联规则。

需要注意的是，在关联规则挖掘过程中，若支持度的阈值过小，将导致产生大量弱关联规则并延长挖掘时间；反之，则可能遗漏有价值的关联规则。因此，需要通过修改阈值进行交互式挖掘，以确定最佳支持度阈值。此外，在电力生产安全事故事件关键诱因分析中，单一项集都必须包含所有的事故诱因，即 A_1—F_{17}。

4　电力生产安全事故事件关键诱因筛选步骤

根据电力生产安全事故事件的实际特点，可构建相应的关键诱因筛选流程图如图 3 所示。

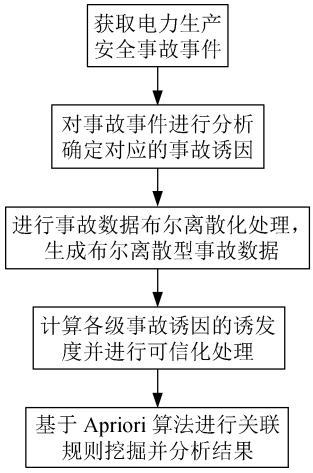

图 3　电力生产安全事故事件关键诱因筛选流程

Fig.3　Key incentive selection process of power production safety accident

具体过程由下列步骤构成。

a. 获取电力生产安全事故事件并进行分析，确定某一事故对应的事故类型和事故诱因。其中事故事件数据来源于评估时间区间内辖区电网所上报的电力系统事故事件报告；事故类型和事故诱因可由事故事件信息管理系统直接获取，也可由专家根据具体事故情况进行人工确定和修正。

b. 进行离散化处理。根据事故对应的事故类型和事故状态因素，对具体事故进行泛化，建立如图 2 所示的电力生产安全事故事件信息矩阵。

c. 计算事故状态因素诱发度并进行可信化筛选。根据式(5)—(8)，分别计算一级诱因和二级诱因的诱发度，并以人为设定的诱发度阈值对小于阈值的事故因素进行筛选，最终得到满足本次关联规则挖掘的完备且可信的关键诱因。

d. 基于 Apriori 算法进行关联规则挖掘并分析结果。设定最小支持度阈值 ε_0、最小置信度阈值 δ_0，根据式(2)—(4)及 Apriori 算法，实现对事故关键诱因的关联规则分析，并进行结果分析。

5　案例分析

以某区域电网近 5 年发生的 469 起电力生产安全事故事件为例，根据上述事故类型及事故诱因模型，映射出如图 2 所示结构的对应布尔型数据表，由于该区域近 5 年未发生特别重大事故和重大事故，本文仅对其中的较大事故、一般事故、一级事件、二级事件、三级事件、四级事件、五级事件 7 类事件(分别以 X_3—X_9表示)进行诱发度计算，结果如图 4 所示。

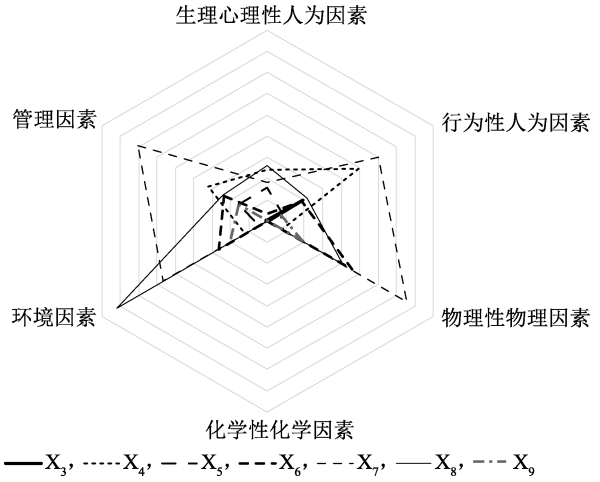

图 4　一级诱因诱发度计算结果

Fig.4　Calculation results of first grade incentives

可见，该区域内所发生的电力系统事故中生理心理性人为因素对较大事故、四级事件的诱发度较大，行为性人为因素对一般事故、三级事件的诱发度较大，物理性物理因素对三级事件的诱发度最大，环境因素对四级事件的诱发度最大，管理因素对三级事件的诱发度最大。但诱发度的计算并未计及各事故诱因对于电力系统事故的关联概率影响，仅反映了各事故诱因对各类电力事故的数值特征，为了更真实、准确地反映事故诱因与事故事件的关联关系，进行关联规则挖掘。

分别以三级事件和四级事件为例，取诱发度阈值 $\xi_0 = 15\%$，筛选出 2 类事故的关键诱因，略去各类事故中非可信的事故状态因素，设最小支持度 $\varepsilon_0 = 5\%$、最小置信度 $\delta_0 = 30\%$，以二级诱因为先导，事故类型为后继，采用 Apriori 算法进行关联规则挖掘，得到 2 类事故的诱因关联特征如图 5 所示。

由图 5 可见，知识技能不足(A_6)、违反规章制度(B_3)、规划设计不合理(C_1)、监督管理不充分(F_4)、风险管控不到位(F_5)、检修维护不充分(F_9)与三级事件 X_7的关联最密切，说明三级事件的发生主要与运维人员自身行为不当相关，因此三级事件的预控要从电力系统运行维护人员自身出发，进行规范性、严格性的管控；而四级事件 X_8的关键诱因中自然因素(E_2)和外力破坏(E_3)具有较高的支持度和置信

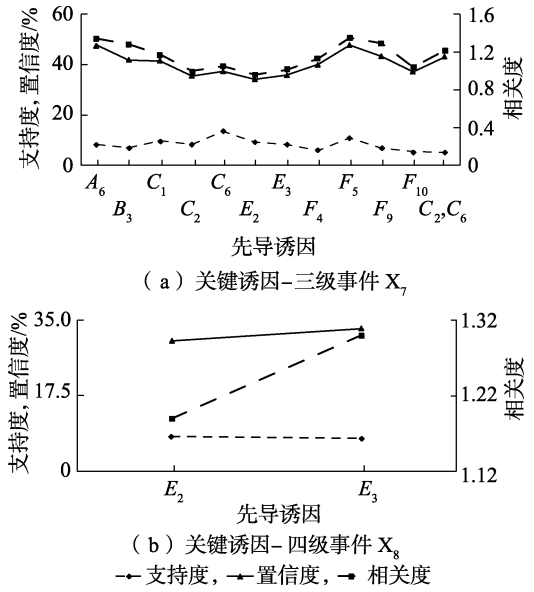

图5 关联规则挖掘结果

Fig.5 Results of association rule mining

度,且 E_3 的关联程度最大,说明四级事件的防控应首先从防外力破坏入手。这些产生的关联规则结果符合实际的电力系统事故情况,挖掘结果准确可靠,描述了电力生产安全事故事件关键诱因与造成的事故事件二者之间的因果联系[21]。

为研究本文所提筛选方法对电力生产安全事故事件诱因的筛选作用,验证所筛选出的电力生产安全事故事件关键诱因的强致因性,保持 ε_0、δ_0 不变,计算不同诱发度阈值下事故 X_7 和 X_8 的诱因筛选结果和关联规则结果如图6所示。

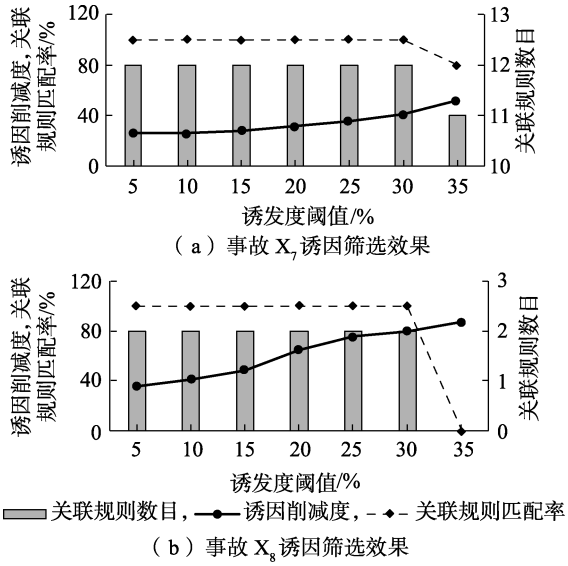

图6 不同诱发度阈值下事故 X_7 和 X_8 诱因筛选效果

Fig.6 Effect of incentive selection of X_7 and X_8 for different incentive thresholds

由图6可见,诱发度阈值在关键诱因筛选中起着"过滤"的作用,其中诱因筛选度是指筛选的诱因数量占全部诱因数量的百分比;关联规则匹配率是在不同诱发度阈值下挖掘出的关联规则与标准关联规则的相同率(本文以 $\xi_0=15\%$ 时的挖掘结果为标准关联规则)。随着诱发度阈值的提高,事故 X_7、X_8 的诱因筛选度均上升,但 X_8 的诱因筛选率明显高于 X_7,这是由于 X_8 的关键事故事件诱因更为单一,提高诱发度阈值将显著筛选掉非关键事故事件诱因;在一定诱发度阈值范围内,诱因的减少并不会引起关联规则结果的变化,说明关键诱因在电力生产安全事故事件的发展演化中起着至关重要的决定性作用,当诱发度阈值过大(如35%),筛选掉关键诱因后,关联规则挖掘结果将出现缺失。在实际中需要合理确定诱发度阈值,保留电力生产安全事故事件的关键诱因。

此外由图5(a)可知,事故 X_7 的发生是在多个关键诱因的致因作用下不断演化、转变形成。为分析各关键诱因之间的演变关系,帮助电网运行维护人员防控风险点,给关键诱因的预控措施的制定提供针对性的辅助策略,以相关度较高的关键诱因 A_6、B_3、C_1、F_4、F_5、F_9 为基础,通过 Apriori 算法计算全2阶频繁项集的发生概率及对应1阶频繁项集的发生概率,利用条件概率公式计算单项事故关键诱因间的演化关系,如表1所示。

表1 X_7 相关度较高关键诱因演化关系

Table 1 Evolution of key incentives highly correlated with X_7

先导关键诱因	后继关键诱因发生概率/%					
	A_6	B_3	C_1	F_4	F_5	F_9
A_6	—	52.4	21.7	37.3	47.0	21.7
B_3	54.4	—	21.5	49.4	48.1	16.5
C_1	16.2	15.3	—	14.4	31.5	19.8
F_4	44.3	55.7	22.9	—	54.3	18.6
F_5	33.6	35.5	32.7	35.5	—	25.2
F_9	24.3	17.6	29.7	17.6	36.5	—

易得,当先导关键诱因变化时,后继关键诱因的发生概率也随之变化。例如,当在违反规章制度(B_3)发生的前提下,造成风险管控不到位(F_5)的概率为48.1%;而在风险管控不到位(F_5)的情况下,造成违反规章制度(B_3)的概率为35.5%,为降低这一组耦合关键诱因集对电力系统安全稳定运行的影响,应优先规范相关规章制度,提高运行维护人员的规范意识,通过其关联作用提高风险管控水平。

6 结论

本文提出利用关联规则挖掘对电力生产安全事故事件进行关键诱因筛选及分析。建立了电力生产安全事故事件抽象复现的方法,采用关联规则挖掘和 Apriori 算法得到了客观的"关键诱因-事故事件"推理规则集,确保电力生产的安全稳定进行。通过

算例证明本文提出的方法在准确挖掘出可靠关联规则的同时,能对诱因数进行有效的筛选,对于辅助电力生产安全态势感知及控制决策的提出具有重要意义。

参考文献:

[1] 邓建玲. 能源互联网的概念及发展模式[J]. 电力自动化设备,2016,36(3):1-5.
DENG Jianling. Concept of energy internet and its development modes[J]. Electric Power Automation Equipment,2016,36(3):1-5.

[2] 孙连捷,张梦新. 安全科学技术百科全书[M]. 北京:中国劳动社会保障出版社,2006:45-47.

[3] 张素香,赵丙镇,王风雨,等. 海量数据下的电力负荷短期预测[J]. 中国电机工程学报,2015,35(1):37-42.
ZHANG Suxiang,ZHAO Bingzhen,WANG Fengyu,et al. Short-term power load forecasting based on big data[J]. Proceedings of the CSEE,2015,35(1):37-42.

[4] 马瑞,周谢,彭舟,等. 考虑气温因素的负荷特性统计指标关联特征数据挖掘[J]. 中国电机工程学报,2015,35(1):43-51.
MA Rui,ZHOU Xie,PENG Zhou,et al. Data mining on correlation feature of load characteristics statistical indexes considering temperature[J]. Proceedings of the CSEE,2015,35(1):43-51.

[5] 王德文,孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中国电机工程学报,2015,35(3):527-537.
WANG Dewen,SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE,2015,35(3):527-537.

[6] 张素香,刘建明,赵丙镇,等. 基于云计算的居民用电行为分析模型研究[J]. 电网技术,2013,37(6):1542-1546.
ZHANG Suxiang,LIU Jianming,ZHAO Bingzhen,et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology,2013,37(6):1542-1546.

[7] 李培强,李欣然,陈辉华,等. 基于模糊聚类的电力负荷特性的分类与综合[J]. 中国电机工程学报,2005,25(24):73-78.
LI Peiqiang,LI Xinran,CHEN Huihua,et al. The characteristics classification and synthesis of power load based on fuzzy clustering[J]. Proceedings of the CSEE,2005,25(24):73-78.

[8] 赵莉,候兴哲,胡君,等. 基于改进 k-means 算法的海量智能用电数据分析[J]. 电网技术,2014,38(10):2715-2720.
ZHAO Li,HOU Xingzhe,HU Jun,et al. Improved k-means algorithm based analysis on massive data of intelligent power utilization[J]. Power System Technology,2014,38(10):2715-2720.

[9] 朱利鹏,陆超,孙元章,等. 基于数据挖掘的区域暂态电压稳定评估[J]. 电网技术,2015,39(4):1026-1032.
ZHU Lipeng,LU Chao,SUN Yuanzhang,et al. Data mining based regional transient voltage stability assessment[J]. Power System Technology,2015,39(4):1026-1032.

[10] 林霞,董晓峰,陆于平,等. 基于数据挖掘的分布式系统孤岛检测方法[J]. 电力系统及其自动化学报,2011,23(2):38-44.
LIN Xia,DONG Xiaofeng,LU Yuping,et al. Application of data mining in island detection of distributed generation of distributed generation[J]. Proceedings of the CSU-EPSA,2011,23(2):38-44.

[11] HAN Jiawei,KAMBER M. 数据挖掘概念与技术(计算机科学丛书)[M]. 2 版. 范明,孟小峰,译. 北京:机械工业出版社,2008.

[12] 李黎,张登,谢龙君,等. 采用关联规则综合分析和变权重系数的电力变压器状态评估方法[J]. 中国电机工程学报,2013,33(24):152-159.
LI Li,ZHANG Deng,XIE Longjun,et al. A condition assessment method of power transformers based on association rules and variable weight coefficients[J]. Proceedings of the CSEE,2013,33(24):152-159.

[13] 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.

[14] 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.

[15] 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.
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.

[16] 万祥,胡念苏,韩鹏飞,等. 大数据挖掘技术应用于汽轮机组运行性能优化的研究[J]. 中国电机工程学报,2016,36(2):459-467.
WAN Xiang,HU Niansu,HAN Pengfei,et al. Research on application of big data mining technology in performance optimization of steam turbines[J]. Proceedings of the CSEE,2016,36(2):459-467.

[17] 隋丽颖,黄孝彬,谭文,等. 基于数据挖掘的火电机组运行模式建立方法研究[J]. 现代电力,2010,27(2):74-77.
SUI Liying,HUANG Xiaobin,TAN Wen,et al. Research on establishment method of thermal power units operation mode based on data mining[J]. Modern Electric Power,2010,27(2):74-77.

[18] 邱凤翔,司风琪,徐治皋. 电站关联规则的主元分析挖掘方法及传感器故障检测[J]. 中国电机工程学报,2009,29(5):97-102.
QIU Fengxiang,SI Fengqi,XU Zhigao. Association rules mining based on principal component analysis and sensor fault detection of power plant[J]. Proceedings of the CSEE,2009,29(5):97-102.

[19] 梁志瑞,陈鹏,苏海锋,等. 关联规则挖掘在电厂设备故障监测中应用[J]. 电力自动化设备,2006,26(6):17-19.
LIANG Zhirui,CHEN Peng,SUN Haifeng,et al. Association rule mining in fault monitoring of power plant equipment[J]. Electric Power Automation Equipment,2006,26(6):17-19.

[20] 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.

[21] 杨宜东,孙志挥. 基于时间序列的电力系统故障诊断规则挖掘[J]. 计算机工程与应用,2003,39(17):223-224.
YANG Yidong,SUN Zhihui. Mining rules based on time sequences for fault diagnosis in electric power systems[J]. Computer Engineering and Applications,2003,39(17):223-224.

作者简介：

陈碧云(1978—)，女，广西北海人，副教授，博士，主要研究方向为电力系统可靠性分析与规划(**E-mail**：chenbiyu@gxu.edu.cn)；

丁　晋(1992—)，男，四川泸州人，硕士研究生，主要研究方向为电力系统可靠性分析与规划(**E-mail**：ding_jin1992@foxmail.com)；

陈碧云

陈绍南(1987—)，男，广西桂平人，博士，主要研究方向为电力系统可靠性分析与规划以及风电相关性分析(**E-mail**：451310528@qq.com)。

Selection of key incentives for power production safety accidents based on association rule mining

CHEN Biyun[1], DING Jin[1], CHEN Shaonan[2]

(1. Guangxi key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning 530004, China;
2. Guangxi Power Grid Electric Power Research Institute, Nanning 530023, China)

Abstract: With the development of smart grid, communication network and power production safety accident analysis level, the data of power production safety accident is growing exponentially with more complexity, which gradually forms the big data of accidents in electric power production. In order to classify and identify the causes of accidents effectively and reliably on the basis of a large amount of prior accident data, the association rule mining is adopted to select the key incentives of power production safety accident. The incentive analysis system of power production safety accident is established according to the characteristics of accidents. Boolean discretization is carried out to different type of accidents, the calculation method for incentive degree of accident causes is proposed based on association rule mining, the Apriori algorithm is adopted for deep association rules mining, and the key causes are selected and analyzed based on the strong association rules. The validity of the proposed method is verified by the analysis of 5-year regional accident cases.

Key words: power production safety accidents; incentive degree; association rules; Apriori algorithm

(上接第67页 continued from page 67)

Fast fault clearance and system recovery technology of MMC based on bridge-arm damping valve groups

SI Zhilei[1,2], LU Yi[3], HAN Kun[1,2], FAN Caiyun[1,2], XIA Kepeng[1,2], HU Siquan[1,2]

(1. XJ Electric Co., Ltd., Xuchang 461000, China; 2. XJ Group Corporation, Xuchang 461000, China;
3. State Grid Zhejiang Electric Power Corporation, Hangzhou 310000, China)

Abstract: The scheme based on the bridge-arm damping valve groups is introduced to solve the technical problems of flexible HVDC fast fault clearance and system recovery. Based on the mechanism analysis of flexible HVDC pole-to-pole short circuit fault and bridge-arm short circuit fault, the specific electrical topology of the bridge-arm damping valve groups available for engineering, the control and protection strategy of damping valve groups, the selection principle of bridge-arm damping parameters and the design process of engineering parameters are proposed. A multi-terminal flexible-HVDC project is taken for example, and its simulative results show that, when the flexible HVDC pole-to-pole short circuit fault occurs, the proposed scheme can accelerate the attenuation of the DC short circuit current and realize fast restart of system after the fault clearance, and when the bridge-arm short circuit fault occurs, the proposed scheme can accelerate the attenuation of the DC component in the valve-side AC current, accelerate the breaking process of AC switches and reduce the failure risk of converter valves. The effectiveness and feasibility of the proposed scheme are verified.

Key words: modular multilevel converter; bridge-arm damping valve groups; DC faults; bridge-arm short circuit fault; fault clearance; fast recovery; restart