

DOI:10.19651/j.cnki.emt.2108003

基于 ACT-Apriori 算法的电网故障诊断方法研究*

程江洲 闫丹阳 冯梦婷 冯馨以
(三峡大学 电气与新能源学院 宜昌 443002)

摘要: 电网拓扑结构愈加复杂,故障后难以快速从海量数据中挖掘有效故障信息且具有较高的计算复杂度;当故障数据不完整或不确定时,无法得到正确的诊断结果。针对此问题,将 Apriori 算法与自编码算法结合改进,形成改进的自编码关联规则挖掘算法(ACT-Apriori)并引入电网故障诊断之中。以保护和断路器动作数据作为条件属性,故障线路为决策属性,建立故障初始决策表;然后利用 ACT-Apriori 算法进行核属性提取并利用动态阈值交互式挖掘技术确定最佳阈值;最后形成最简故障决策表,实现故障信息的诊断推理。采用四母线配电系统作为仿真对象,与传统的 Apriori 算法和 FP-growth 算法及目前最新的 FP-Network 算法进行对比分析,算例结果表明:改进的算法相较于传统关联规则算法,运行时间分别缩减了 90.69% 和 83.55%,内存占用分别缩减了 21.43% 和 15.38%,相较于 FP-Network 算法,在时间复杂度和空间复杂度上均有一定程度优化;且本文算法对故障数据不完备情况下的单重、双重、稀有故障诊断的容错性较高,准确率达到 95.24%,可以有效实现故障的快速诊断。

关键词: 故障诊断;数据挖掘;关联规则;属性约简;ACT-Apriori

中图分类号: TM73 **文献标识码:** A **国家标准学科分类代码:** 510.4

Power network fault diagnosis based on ACT-Apriori algorithm

Cheng Jiangzhou Yan Ranyang Feng Mengting Feng Xinyi

(College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002)

Abstract: In view of the increasingly complex topology structure of power grid, it is difficult to quickly mine effective fault information from massive data after a fault and has high computational complexity, and the fault data is incomplete and uncertain, leading to the failure to get correct diagnosis results. To solve this problem, this paper introduces the self-coding association rule mining algorithm (ACT-Apriori) into power grid fault diagnosis. The initial fault decision table was established by taking the protection and circuit breaker action data as the condition attribute and the fault line as the decision attribute. Then a self-coding association rule mining algorithm is used to extract kernel attributes and the optimal threshold is determined by dynamic threshold interactive mining technology. Finally, the simplest fault decision table is formed, and the fault information of each case is diagnosed and reasoned. In this paper, the four-bus distribution system is used as the simulation object, and compared with the traditional Apriori algorithm, FP-growth algorithm and the latest FP-Network algorithm, the calculation results show that: Compared with the traditional association rule algorithm, the running time of the improved algorithm is reduced by 90.69% and 83.55%, and the memory footprint is reduced by 21.43% and 15.38%, respectively. Compared with the FP-Network algorithm, the time complexity and space complexity are optimized to a certain extent. In addition, the proposed algorithm has high fault tolerance for single, double and rare faults with incomplete fault data, and the accuracy is 95.24%, which can effectively achieve rapid fault diagnosis.

Keywords: fault diagnosis; data mining; associative rules; attribute reduction; ACT-Apriori

0 引言

随着配电网自动化设备的完善,各电网之间的联系日

益密切,在电网发生故障时,海量的动作信息汇总至集控中心,由此单纯依靠人工完成故障诊断愈加困难^[1-2]。另外,海量动作信息在传输过程中会因抖动和延迟出现数据丢包

收稿日期:2021-10-03

* 基金项目:国家自然科学基金面上项目(61876097)、湖北省科技计划项目技术创新重大项目(2016AAA040)资助

和畸变,使得电力系统的响应更为复杂^[3];而且配电网故障也由单重故障向复杂的多重故障演变,加之存在一些稀有故障,极易出现故障漏诊的情况^[4-5]。因此,如何从海量数据中快速提取核属性故障信息,对数据进行降维实现快速故障诊断;如何提高数据在传输过程当中出现的抖动等不确定因素时的容错性;如何防止稀有故障的漏诊,保障各类故障诊断的准确性是目前亟待解决的问题。

近年来,随着数据挖掘技术的发展,先后有学者提出了更具针对性的方法,如 Petri 网络^[6]、粗糙集理论^[7]、贝叶斯理论^[8]、分类决策树算法^[9]。孙铁军等人通过拟合各种保护与断路器之间的逻辑关系,提出了基于信息优化的动态建模模糊 Petri 网的电网故障诊断方法,但当拓扑节点较多时,容易出现组合爆炸的问题^[10]。夏昌浩等^[11]在贝叶斯故障诊断的基础上引入了粗糙集理论,简化了电网拓扑结构,但诊断双重故障时,各个事件的条件概率和先验概率计算繁琐,计算复杂度仍然较高。王磊等^[12]运用反向追踪决策树技术,通过监控保护和断路器的误动技术给出故障诱因,实现诊断技术的大幅度优化,但其对稀有故障的诊断效果欠佳。

数据挖掘中的关联规则挖掘技术可以实现对海量集成数据的有效降维,发现各种设备状态监测量与设备缺陷及故障之间内在的关联关系^[13]。但现有的大数据关联规则算法需要多次扫描全局数据库,计算时间长、算法复杂度高, I/O 占用率较大,不利于快速准确的实现配电网故障诊断^[14]。

基于此,文中提出一种基于自编码关联规则挖掘技术的电网故障诊断方法,实现算法时间复杂度和空间复杂度的双重优化;利用动态阈值交互式挖掘技术,保障多种故障类型下的准确研判,在原有 Apriori 算法应用于电网故障诊断优势的基础上,实现故障诊断准确性和容错性的双重提高。

1 关联规则挖掘技术

1.1 关联规则基本原理

设存在一个由 n 个项目和 m 个事务组成事务数据库 (D),其中 $I = \{i_1, i_2, \dots, i_n\}$ 是 n 个项目的有限集合, $D = \{t_1, t_2, \dots, t_m\}$ 是 m 个事务的有限集合。交易 t 是集合 $I (t \subseteq I)$ 中的一组项目。基于支持度、置信度与提升度提取关联规则 $X \rightarrow Y$ 。

规则 $X \rightarrow Y$ 的支持度,用 $S(X \rightarrow Y)$ 表示,表示同时包含 X 和 Y 事务的百分比,代表了 $X \cup Y$ 的先验概率,其概率表达式如式(1)所示。

$$S(X \rightarrow Y) = P(X \cup Y) \quad (1)$$

置信度 $C(X \rightarrow Y)$ 表示在 Y 发生的条件下 X 发生的可能性,其概率表达式如式(2)所示。其中 $S(X, Y)$ 和 $S(X)$ 分别是关联规则和项集 X 的支持计数。

$$C(X \rightarrow Y) = \frac{S(X \cup Y)}{S(X)} \quad (2)$$

当挖掘出一定关联规则后,采用提升度 $Lift(X \rightarrow Y)$

来删除冗余规则,使得挖掘结果更加精炼。提升度代表 $X \rightarrow Y$ 前因后果的相关程度,其概率表达式如式(3)所示。

$$Lift(X \rightarrow Y) = \frac{C(X \rightarrow Y)}{S(Y)} = \frac{P(Y | X)}{P(Y)} = \frac{P(XY)}{P(X)P(Y)} \quad (3)$$

1.2 基于自编码关联规则挖掘算法 (ACT-Apriori)

首先对原始数据库 (D) 进行预处理,并用自编码位向量代替高频项目形成简化数据库 (RBD),使得数据维度大幅度降低。将数据记录简化后全部读入内存,通过改进频繁项集连接、剪枝过程,从而直接生成候选项集,并扫描数据库计算支持度。由于候选项集与简化数据库 (RBD) 均已排序,一旦搜索到大于候选项的值时,即可停止该事务的搜索。算法整体流程图如图 1 所示。

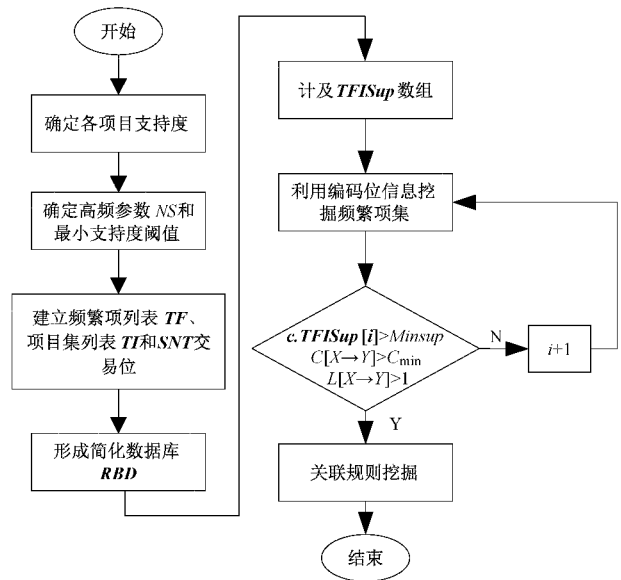


图 1 ACT-Apriori 算法流程图

1) 数据预处理阶段

在数据预处理阶段首先计算事务数据库 (D) 中各项目的支持度计数并以降序排列,其次设定项目的高频参数 (NS),在事务数据库 (D) 中通过选定的 NS 参数忽略部分频繁项并将其用自编码位向量 (SNT) 表示,同时建立频繁项列表 (TF) 和频繁项目集列表 (TI),最终形成简化数据库 (RBD)。相关参数定义如表 1 所示。

表 1 ACT-Apriori 参数表

名称	属性
高频参数 (NS)	较高支持度计数项目
频繁项列表 (TF)	所有 NS 创建的列表每一项称为 TFI TF 中所有可能的排列组合
项目集列表 (TI)	长度介于 $0 \sim NS$ 之间 每一项称为 $TFIT$ 数量为 2^{NS}
自编码位向量 (SNT)	每个 NS 在原始数据库中位置
简化数据库 (RBD)	SNT 代替 TFI 形成的新数据库

2) 频繁项集生成阶段

引理 1: 如果事务 l 包含项集 A 和项集 B , 则 l 包含项集 $\{A+B\}$; 其中, $\{A+B\}$ 表示项集 A 和 B 的构成。

ACT-Apriori 通过使用引理 1 寻找所有的频繁项集, 在保证候选项集 c 包含在事务 T 中并且事务 T 还包含 $TI[i]$ 之后, 可以得出项目集 $\{c+TI[i]\}$ 也将包含在事务 T 中的结论。

在 ACT-Apriori 算法的第二阶段, 用 RDB 代替原始数据库, 把 $TFIs$ 从其余项目中分离出来, 将对候选项生成和支持计数阶段产生影响。在候选项生成中, 于 $Tree$ 数据结构的每个节点中添加了一个大小为 2^{NS} 的支持计数器列表 ($TFISup$) 用以跟踪 $TFIs$, 每个 $TFISup$ 代表相关项集和其中一个 $TFIs$ 的组合。如图 2 所示, 以一个最简单的实例表明了 $TFISup$ 数组的使用。 $TFISup$ 数组的每个条目都将指示项集 $\{c+TFITs\}$ 的支持计数。使用“与”位向量运算可以很容易地确定 $TFISup$ 数组的哪个条目属于哪个项集。所以 $c.TFISup[0]$ 存储项集 c 的支持计数, $c.TFISup[2^{NS}]$ 存储项集 $\{c+TI[2^{NS}]\}$ 的支持计数。在生成频繁项集时, 对于每个项集 c , 如果 $c.TFISup[i] > MinSup$, 算法将输出项集 $\{c+TI[i]\}$, 完成第 k 遍后, 算法将 k 增加 1, 并再次迭代其过程, 直到 $L[k]$ 中没有频繁项集为止。

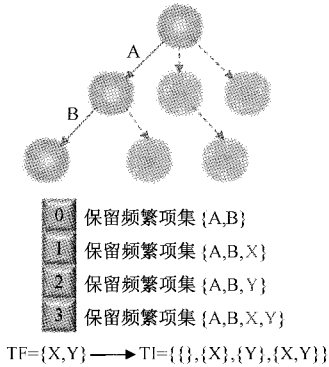


图 2 TFISup 结构示例图

2 基于自编码关联规则数据挖掘配电网故障诊断技术

本文所述挖掘方法流程如图 3 所示, 技术路线图如图 4 所示。整体分为故障采集与预处理模块、ACT-Apriori 算法应用模块、配电网线路故障诊断模块。首先于配电网故障录波器监控主机终端导出保护和断路器的动作信号数据, 将该离散数据进行整合归纳输入 PC 端形成故障初始决策表, 其中以保护和断路器的动作数据作为条件属性, 故障线路作为决策属性; 然后利用 ACT-Apriori 算法进行属性约简形成最简核属性故障决策表。先在小范围数据集中动态调整阈值进行反复挖掘, 确定阈值后再应用于大型数据集的关联规则挖掘之中, 以防止遗漏稀有故障所对应的关联规则; 其次采用提升度量筛选出强关联规则用于配电网故障诊断; 最后引入置信度排序技术, 为检修人员提供辅助决策方法。

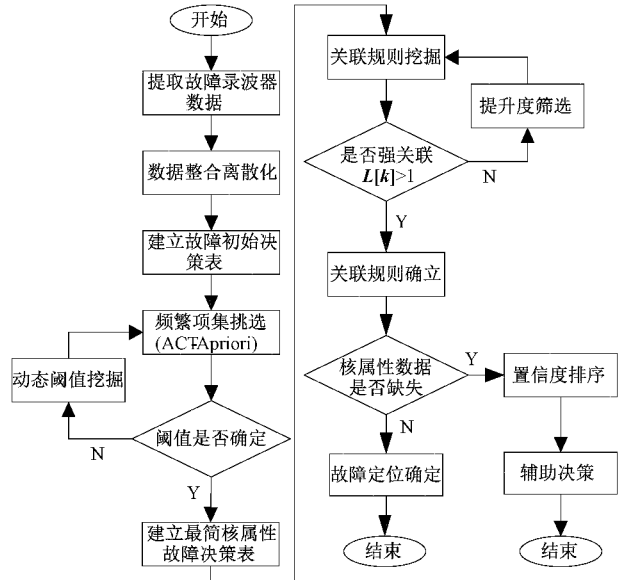


图 3 基于关联规则的配电网故障诊断流程

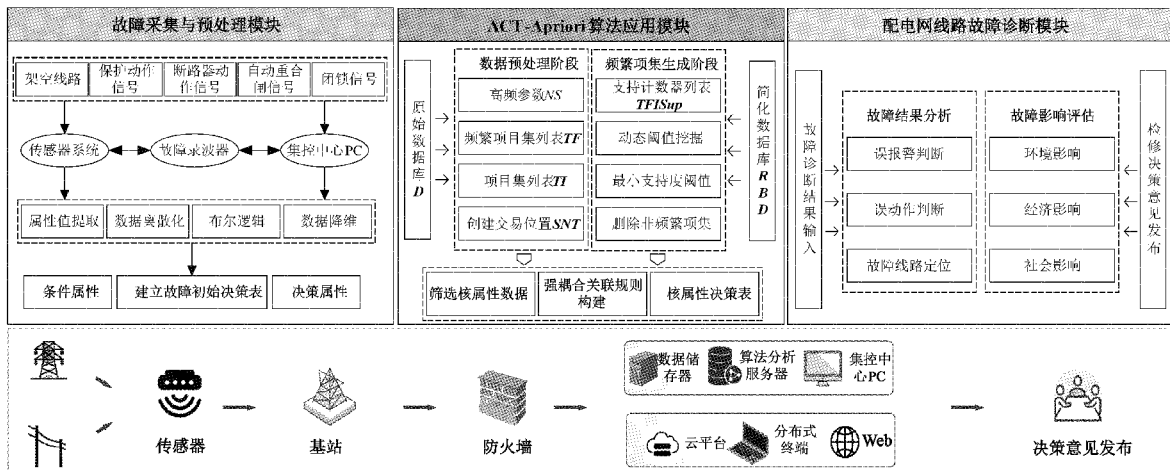


图 4 技术路线图

在实际的配电网系统中,为详细说明诊断流程,以图5简单配电网拓扑结构为例。

该网络拓扑结构由 $Sec_i (i=1,2,3,\dots,5)$ 共5个区域构成,分别配有过流保护 $CO_i (i=1,2,3,\dots,5)$ 和距离保护 $RR_i (i=1,2)$, $QF_i (i=1,2,3,\dots,5)$ 是各区域线路上的断路器。条件属性取值为“1”表示断路器由闭合变位为断开或保护动作,“0”表示断路器未变位或保护未动作,决策属性为故障区域线路。将故障录波器中的数据进行布尔离散化之后形成故障初始决策表,如表2所示。表2中条件属

性共12项,决策属性1项,共形成14条事务。

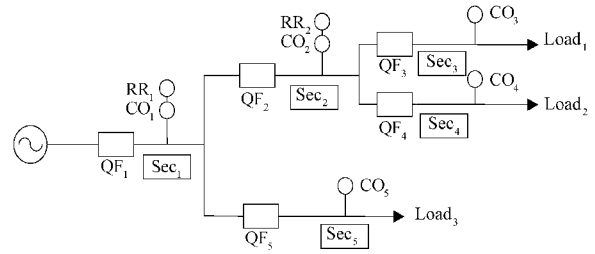


图5 简单配电网拓扑结构图

表2 故障初始决策表

样本	条件属性												决策属性
	QF ₁	QF ₂	QF ₃	QF ₄	QF ₅	CO ₁	CO ₂	CO ₃	CO ₄	CO ₅	RR ₁	RR ₂	
1	1	0	0	0	0	1	0	0	0	0	0	0	Sec ₁
2	0	1	0	0	0	0	1	0	0	0	0	0	Sec ₂
3	0	0	1	0	0	0	0	1	0	0	0	0	Sec ₃
4	0	0	0	1	0	0	0	0	1	0	0	0	Sec ₄
5	0	0	0	0	1	0	0	0	0	1	0	0	Sec ₅
6	0	1	0	0	0	0	0	1	0	0	0	1	Sec ₃
7	0	1	0	0	0	0	0	0	1	0	0	1	Sec ₄
8	1	0	0	0	0	0	1	0	0	0	1	0	Sec ₂
9	1	0	0	0	0	0	0	1	0	0	1	0	Sec ₃
10	1	0	0	0	0	0	0	0	1	0	1	0	Sec ₁
11	1	0	0	0	0	0	0	0	0	1	1	0	Sec ₅
12	0	0	0	0	0	0	0	0	0	0	0	0	None
13	0	1	0	0	0	0	0	0	0	0	0	1	Sec _{3/4}
14	1	0	0	0	0	0	0	0	0	0	0	0	Sec _{2/5}

为了保证置信度设置的合理性,确保稀有故障的准确挖掘以及减少弱关联规则的产出,实际应用中应先选中整体数据库的15%~18%^[15],在该小规模数据库中进行动态交互式挖掘,在权衡二者的情况下,满足时间和内存占用的最优化,进而确定合适的阈值,最终应用于全局数据库,并辅之置信度排序技术,帮助检修人员实现故障的快速诊断。

本算例中所用阈值为 $S(X \rightarrow Y)_{\min} \geq 31.6\%$, $C(X \rightarrow$

$Y)_{\min} \geq 55\%$, $L(X \rightarrow Y)_{\min} \geq 1$,所提取的关联规则置信度如表3所示。进而可以构成更为精简的电网故障决策表,如表4、5所示。此外,通过容错性诊断方法验证该算法得出结果的有效性和容错性,如表6所示。表6中的第1组信息为正确的故障信息,第2组包含了错误的故障信息(CO₃本来为“0”,但错误给出了“1”),样本3、4、5为模拟网络延迟和抖动所导致的数据丢包或畸变产生的不完备故障信息,其中“*”代表未知故障信息^[16]。

表3 关联规则提取表

序号	关联规则	置信度/%	序号	关联规则	置信度/%
0	{QF ₁ , CO ₁ → Sec ₁ }	100	7	{RR ₂ → Sec _{3/4} }	64.2
1	{QF ₂ , CO ₂ → Sec ₂ }	100	8	{RR ₁ → Sec _{2/5} }	73.6
2	{QF ₃ , CO ₃ → Sec ₃ }	100	9	{QF ₃ → Sec ₃ }	100
3	{QF ₄ , CO ₄ → Sec ₄ }	100	10	{QF ₂ → Sec ₂ }	55.1
4	{QF ₅ , CO ₅ → Sec ₅ }	100	11	{QF ₅ → Sec ₅ }	100
5	{RR ₁ , CO ₃ → Sec ₃ }	100	12	{QF ₁ → Sec ₁ }	68.2
6	{RR ₂ , CO ₂ → Sec ₁ }	100	13	{QF ₁ , RR ₁ → Sec _{2/3} }	79.4

表 4 电网故障样本决策表

样本	故障信息(条件属性)					故障区域
	QF ₂	CO ₂	CO ₃	CO ₄	CO ₅	
1	1	1	0	0	0	Sec ₂
2	0	0	1	0	0	Sec ₃
3	0	0	0	1	0	Sec ₄
4	0	0	0	0	1	Sec ₅
5	1	0	1	0	0	Sec ₃
6	1	0	0	1	0	Sec ₄
7	1	0	0	0	0	Sec _{3/4}

表 5 电网故障样本决策表

样本	故障信息(条件属性)					故障区域
	QF ₁	RR ₁	CO ₃	CO ₄	CO ₅	
1	1	0	0	0	0	Sec ₁
2	0	0	1	0	0	Sec ₃
3	0	0	0	1	0	Sec ₁
4	0	0	0	0	1	Sec ₅
5	0	1	1	0	0	Sec ₃
6	0	1	0	1	0	Sec ₄
7	0	1	0	1	0	Sec ₅
8	0	1	0	0	0	Sec _{2/5}

表 6 检测样本表

序号	设置故障位置	故障信息					输出诊断结果	
		CO ₂	CO ₃	CO ₄	QF ₂	RR ₁	关联规则	置信度/%
1	Sec ₃	0	1	0	0	0	{CO ₃ →Sec ₃ }	100
2	Sec ₃	0	1	0	1	0	{CO ₃ , QF ₂ →Sec ₃ }	86.3
3	Sec ₃	0	1	*	0	0		
4	Sec ₃	0	1	0	*	0	{CO ₃ →Sec ₃ }	64.7
5	Sec ₃	0	1	0	0	*		

通过诊断结果可以表明:当故障信息准确而且完备时可以获得准确的诊断结果;当故障信息出现部分错误时,由于通过关联规则挖掘算法筛选出了诊断核属性信息;当故障信息出现由于多种不确定因素而导致部分丢失时,依旧可以得到正确的诊断结果。

3 算例仿真

为验证本文所提方法在大电网拓扑结构中实现双重故障和稀有故障的准确挖掘,同时算法的时间复杂度和空间复杂度达到最优,本算例采用四母线系统(如图 6 所示)作为仿真对象。系统由 4 条母线 B_i(i=1,2,3,4),1 台变压器 T,4 条输电线路 L_i(i=1,2,3,4)组成。图 6 中保护和断路器编号及含义如表 7 所示。根据保护和断路器的动作原理,所考虑的故障类型汇总如表 8 所示,共建立含有 32 个条件属性,21 个故障样本的故障初始决策表。

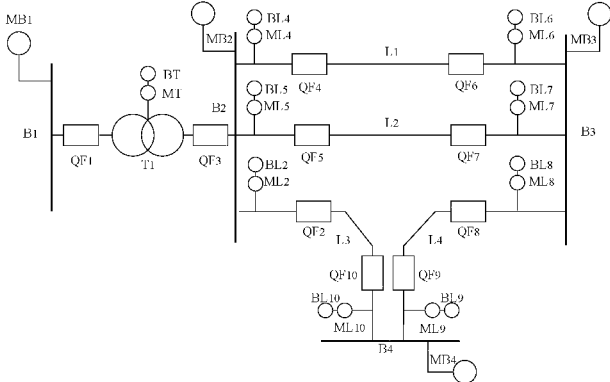


图 6 四母线配电系统

表 7 故障类型汇总表

序号	故障类型	数量	故障性质
①	单重故障	4	保护和断路器均正常工作
②	单重故障	2	保护和断路器拒动
③	单重故障	2	非核属性信息部分丢失
④	单重故障	1	核属性信息部分丢失
⑤	双重故障	1	核属性信息部分丢失
⑥	双重故障	2	非核属性信息部分丢失
⑦	双重故障	4	保护和断路器均正常工作
⑧	双重故障	3	保护和断路器拒动
⑨	稀有故障	1	保护和断路器均正常工作

表 8 断路器及保护编号表

属性	标号
线路断路器	QF ₁ , QF ₂ , QF ₃ , ..., QF ₁₀
母线主保护	MB ₁ , MB ₂ , MB ₃ , MB ₄
线路主保护	ML ₂ , ML ₃ , ML ₄ , ..., ML ₁₀
线路后备保护	BL ₂ , BL ₃ , BL ₄ , ..., BL ₁₀
变压器主保护	MT
变压器后备保护	BT

首先截取小范围数据形成简化故障初始决策表进行动态阈值挖掘确定相关度量标准,再于全局数据库中进行实验测试。测试设备硬件条件为 AMD Ryzen7 5800H with Radeon Graphics 3.20 GHz CPU。挖掘结果如图 7 所示。

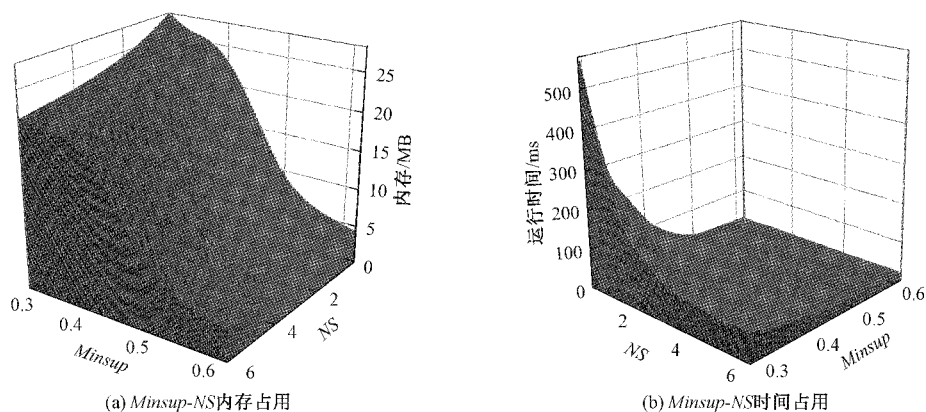


图7 阈值动态挖掘结果

在数据集中,*NS*参数对不同最小支持度阈值下的算法运行时间和内存占用率均有较大的影响,当*NS*参数越大时,算法所需要的内存占用量和运行时间就越少。因此,通过适当提高*NS*参数,可以有效实现时间复杂度与空间复杂度的双重优化。通过实验证明,当*NS*=6、最小支持度阈值 $Minsup = 41.7\%$ 时,ACT-Apriori算法的内存占用量与运行时间达到最优。

根据前文动态阈值挖掘所确定的相关阈值,并根据 Sharmila 等^[17]提出最小支持度阈值为依据,确定本算例相关阈值为 $NS = 6$ 、 $Minsup = 41.7\%$ 、 $Minconf = 55.1\%$ 的并预先设定前文所提出的所有故障类型及位置,将引起此故障的故障信息输入故障诊断软件,得到诊断结果如图8及表9所示,其中图8横坐标代表故障预先设置的位置,纵坐标代表经算法迭代分析后的故障诊断结果,灰度越大表示置信度越高进而表示诊断结果越精确,灰度越小代表置信度越低进而表示故障不发生在该位置;以坐标(B_1, B_1)

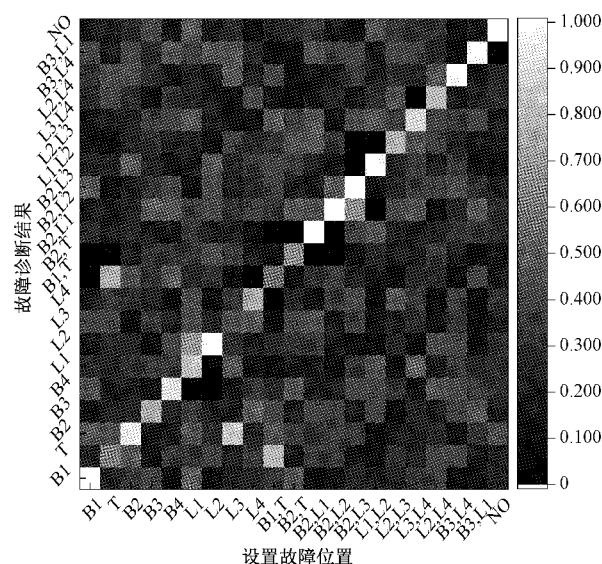


图8 故障诊断结果

表9 故障诊断结果

样本	故障类型	设定故障位置	故障诊断结果	关联规则置信度/%	样本	故障类型	设定故障位置	故障诊断结果	关联规则置信度/%
1	①	B_1	B_1	99.6	12	⑥	B_2, L_1	B_2, L_2	99.3
2	②	T	T	72.2	13	⑦	B_2, L_2	B_2, L_2	99.8
3	①	B_2	B_2	95.3	14	⑦	B_2, L_3	B_2, L_3	99.5
4	①	B_3	B_3	75.4	15	⑦	L_1, L_2	L_1, L_2	98.5
5	①	B_4	B_4	91.5	16	⑦	L_2, L_3	L_2, L_3	78.0
6	②	L_1	L_1	82.4	17	⑧	L_3, L_4	L_3, L_4	88.6
7	③	L_2	L_2	98.8	18	⑧	L_2, L_4	L_2, L_4	80.2
8	④	L_3	B_2	85.1	19	⑧	B_3, L_1	B_3, L_1	98.8
9	③	L_4	L_4	75.2	20	⑨	B_3, L_1	B_3, L_1	97.9
10	⑤	B_1, T	B_1, T	79.6	21		NO	NO	99.9
11	⑥	B_2, T	B_1	56.2					
			B_2, T	66.2					

为例,热图中灰度值为 255(白色),置信度为 99.6%,表明经算法迭代分析后得出故障位置为 B_1 ,与预先设定的故障位置一致;再以坐标 (B_1, B_1) 为例,热图中灰度值为 0(黑色),置信度为 13.8%,低于最小置信度阈值,表明经算法迭代分析后得出故障位置 B_4 不为预先设定的故障位置,因而得到较低的置信度。根据表 9 可知,当故障诊断信息完备时(样本 1、3、4、5),可以得到正确的诊断结果;当保护和断路器拒动时(样本 2、6、17、18、19),通过置信度排序技术可以实现正确的诊断结果,为检修人员提供辅助决策意见;当非核属性信息丢失时(样本 7、9、11、12),对故障诊断结果无较大影响且置信度可以达到较高水平;当核属性信息丢失时,由于样本 10 与样本 2 条件属性互为互补,可以得到较为正确的诊断情况,但样本 8 因条件属性无互补的情况从而导致诊断结果错误,所以当核属性信息丢失时,在条件属性无互补的情况下,无论何种算法都无法得到正确的诊断结果。当发生稀有故障(样本 20),由于其核属性数据未丢失,因此仍然可以产生正确的诊断结果。

为更好的验证所提算法在时间复杂度与空间复杂度的优越性,现将 ACT-Apriori 算法与 Apriori 算法、FP-growth 算法和 FP-Network 算法在不同最小支持度阈值情况下进行对比分析,结果如图 9 所示。由图 9 可知,当 $Minsup \in [0.285, 0.426]$ 时,ACT-Apriori 算法的运行时间明显优于 Apriori 算法和 FP-growth 算法,最优处 ($Minsup = 0.285$) ACT-Apriori 算法的运行时间比 Apriori 算法和 FP-growth 算法分别缩减了 90.69% 和 83.55%,相比于 FP-Network 算法,最优缩减率仅为 17.74%,这是因为 FP-Network 算法与 ACT-Apriori 算法均采用了相同的 Tree 结构对高维数据进行降维,但 ACT-Apriori 算法引入了 $TFISup$ 数组,可进一步压缩数据量;在内存占用方面,当 $Minsup \in [0.285, 0.377]$ 时,ACT-Apriori 算法内存占用情况均优于其余 3 种算法,最优处 ($Minsup = 0.285$) ACT-Apriori 算法的内存占用比其余三种算法分别缩减了 21.43%、15.38%、12.01%。当 $Minsup \in (0.426, 0.621]$ 时,虽然 ACT-Apriori 算法在时间和内存占用上比其余

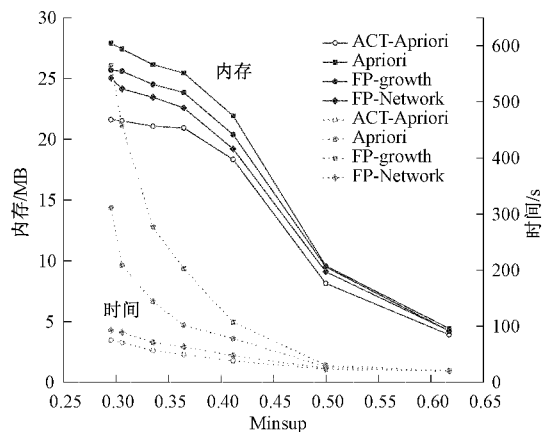


图 9 内存占用和运行时间横向对比分析

3 种算法的优化程度并不明显,但是在数据挖掘过程中,一旦 $Minsup$ 设置过高,大量的关联规则将被遗漏,因此,应尽量将 $Minsup$ 设置在较低的阈值区间,这同时也顺应了本文所提算法的最佳优化区间,保证了本算算法应用的优越性。

4 结 论

本文将自编码关联规则算法用于电网故障诊断之中,通过仿真算例可得到以下结论:

通过关联规则挖掘技术,可以有效对海量的故障信息进行降维处理,快速提取核属性信息;通过动态阈值交互式挖掘技术可以准确确定未知样本集的最佳阈值,保障强耦合关联规则的快速挖掘。在故障信息完备的情况下,无论是单重、双重还是稀有故障,均可以得到正确的诊断结果;当出现保护和断路器拒动导致故障信息不完备或非核属性信息缺失时,通过置信度排序技术可以对所有可能故障位置降序排列,为检修人员提供了辅助决策意见,提高了故障诊断的冗余性和快速性。将自编码技术引入 Apriori 算法,仅需扫描 $TFISup$ 数组即可实现频繁项集的筛选,摒弃了 Apriori、FP-growth 等传统算法需要多次扫描全局数据库的问题,降低了算法的时间复杂度;通过适当设置 NS 参数,可有效降低算法的内存占用率,进而有效减少挖掘过程中 I/O 占用率,降低了空间复杂度,提高了挖掘模型的计算效率,对电网故障快速诊断具有极其重要的应用价值。

然而,本文所提算法在空间复杂度的优化程度上仍有不足, NS 参数选取的恰当与否直接影响算法整体的运行性能,但 NS 参数的选取过程仍较为繁琐;另外,当应用于核属性信息丢失情况下的双重连锁故障诊断中时,无法得到较高的真实故障位置置信度。因此,如何建立更为准确的双重连锁故障诊断模型,保障核属性信息丢失时的精确诊断,是下一步的研究方向。

参 考 文 献

- [1] 侯思祖,郭威. 基于改进 EEMD 与混沌振子的配电网故障选线[J]. 仪器仪表学报,2019,40(4):77-87.
- [2] 郭榕,杨群,刘绍翰,等. 电网故障处置知识图谱构建研究与应用[J]. 电网技术,2021,45(6):2092-2100.
- [3] 李桐,王刚,刘扬,等. 配电网智能终端信息安全风险分析[J]. 智慧电力,2020,48(9):118-122.
- [4] 毛钧毅,韩松,李洪乾,等. 采用 Spiked 协方差模型与“相变”现象的电网不平衡扰动评估[J]. 仪器仪表学报,2020,41(12):208-216.
- [5] 洪翠,付宇泽,郭谋发,等. 改进多分类支持向量机的配电网故障识别方法[J]. 电子测量与仪器学报,2019,33(1):7-15.
- [6] 袁川来,廖庸邑,孔玲爽. 时间约束的改进分层模糊 Petri 网的配电网故障诊断方法[J]. 电子测量与仪器学报,2020,34(3):126-134.

- [7] LIU W, WANG T, ZANG T, et al. A fault diagnosis method for power transmission networks based on spiking neural P systems with self-updating rules considering biological apoptosis mechanism [J]. Complexity, 2020, DOI: 10.1155/2020/2462647.
- [8] AKER E, OTHMAN M L, VEERASAMY V, et al. Fault detection and classification of shunt compensated transmission line using discrete wavelet transform and naive bayes classifier [J]. Energies, 2020, DOI: 10.3390/en13010243.
- [9] BENKERCHA R, MOULAHOU M S. Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system[J]. Solar Energy, 2018, 173(OCT.):610-634.
- [10] 孙铁军, 曲丽萍, 关海爽, 等. 基于智能优化的模糊 Petri 网的电网故障诊断[J]. 控制工程, 2021, 28(1): 127-134.
- [11] 夏昌浩, 胡爽, 李伶俐, 等. 基于贝叶斯-粗糙集理论的诊断模型及诊断方法的应用[J]. 三峡大学学报(自然科学版), 2020, 42(5):82-87.
- [12] 王磊, 陈青, 高洪雨, 等. 基于大数据挖掘技术的智能变电站故障追踪架构[J]. 电力系统自动化, 2018, 42(3): 84-91.
- [13] 邵晓东, 魏洪昌, 陈宏达, 等. 依赖关联规则映射的多维信息 SDN 数据中心优化[J]. 电子测量技术, 2019, 42(19):74-78.
- [14] 詹柳春, 黄长江. 云计算下物联网密集场景大数据挖掘技术[J]. 电子测量技术, 2019, 42(23):164-168.
- [15] HE K, LIU B, XIE M, et al. Reliability analysis of systems with discrete event data using association rules [J]. Quality and Reliability Engineering International, 2021, DOI:10.1002/QRE.2942.
- [16] 聂倩雯. 基于关联规则数据挖掘和扩展贝叶斯网络的电网故障诊断方法研究[D]. 成都: 西南交通大学, 2010.
- [17] SHARMILA S, VIJAYARANI S. Association rule mining using fuzzy logic and whale optimization algorithm [J]. SoftComputing, 2020, DOI: 10.1007/s00500-020-05229-4.

作者简介

程江洲, 副教授, 主要研究方向为电力设备状态在线检测、智能电网等。E-mail:chengjiangzhou@ctgu.edu.cn

闫冉阳, 硕士研究生, 主要研究方向为智能配电网。E-mail:2557478431@qq.com

冯梦婷, 硕士研究生, 主要研究方向为综合能源系统。E-mail:842764286@qq.com

冯馨以, 硕士研究生, 主要研究方向为智能电网。E-mail:1121774404@qq.com