

基于关联规则数据挖掘技术的电网故障诊断

聂倩雯, 高 玮

(西南交通大学电气工程学院, 四川 成都 610031)

摘要: 在实际的电网故障诊断中, 面临如何从海量数据找到真正对于诊断结果有帮助的关键数据以及当故障信息存在不完整或不确定性, 甚至关键信息丢失时, 会导致故障诊断难以得出正确结论的问题。针对此问题, 将关联规则数据挖掘 DLG (Direct Large Itemsets Generation) 算法引入到电网故障诊断中。首先以保护、断路器作为条件属性, 故障区域作为决策属性, 考察各种故障情况并建立原始决策表, 然后利用关联规则挖掘进行属性约简, 通过修改阈值进行交互式挖掘, 直接提取最佳属性约简组合, 然后利用最佳属性约简组合形成的约简决策表和关联规则交互式挖掘, 针对各种情况的故障信息进行诊断推理。运用 C 编写了基于该方法的故障诊断软件, 采用四母线配电网系统作为仿真对象, 算例结果表明该算法在一定电网规模和保护动作信息不完备的情况下, 故障诊断正确性高、容错性好, 实用性强。

关键词: 电力系统; 故障诊断; 数据挖掘; 关联规则; 约简; DLG 算法

A power network fault diagnosis method based on data mining association rules

NIE Qian-wen, GAO Wei

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: When diagnosing the power network fault, it faces to the puzzle problem that is how to find the key data acquirement in the real complex power system fault diagnosis with mass data. And if the fault information is imperfect and indeterminate or even the key information is lost, it may result in the condition that correct conclusion could not be given by fault diagnosis. To settle these problems, the paper introduces the DLG (Direct Large Itemsets Generation) algorithm based on data mining association rules to diagnose the power network faults. At first, the protections and circuit breakers are taken as conditional attributes and faulty region as decision-making attribute, various faults are investigated and decision table is established. Then by use of attribute reducing method based on data mining association rules and modifying thresholds based on interactive data mining, the optimal attribute reduction combination is directly extracted. Finally, by means of the reduction decision table formed by optimal attribute reduction combination and interactive data mining based on association rules, the diagnosis reasonings to diversified conditions are obtained. The fault diagnosis software is programmed by C programming language. Results of calculation examples show that the proposed method is correct and effective, and can improve the effectiveness of the fault diagnosis system, so this method is available.

Key words: power system; fault diagnosis; data mining; association rules; reduction; DLG algorithm

中图分类号: TM73 文献标识码: A 文章编号: 1674-3415(2009)09-0008-07

0 引言

随着电网规模的不断扩大, 当电网发生故障时, 会有大量丰富的故障信息送入调度中心, 如何从海量故障数据中有效提取信息, 从信息中及时地发现知识; 如何提高在故障信息具有保护、断路器拒动误动或通信线路传输错误等不确定因素时的容错性是急需解决的问题。目前的电网故障诊断系统大多停留在人工诊断阶段, 少数投入运行的电网故障诊断系统^[1,2]在故障信息不确定时难以保证容错性。因此电网故障诊断一直是理论和应用研究的热点。

国内外学者已先后提出了多种电网故障诊断方法, 如专家系统^[3]、模糊理论^[4]、信息理论^[5]等。这些方法各有特点, 也得到了了一定的应用, 但对于处理海量数据信息的能力和不确定故障信息的容错性

仍不甚理想。因此近年来有一些更有效的方法, 如数据挖掘技术中基于事件序列数据挖掘、粗糙集理论^[6-12]、贝叶斯理论、决策树等被引入电网故障诊断中。文献[14]据电力系统发生故障时形成具有时空特征的离散事件序列信息, 提出基于事件序列 DM 原理的输电线路故障诊断模型。文献[15]是用基于决策树的产生式稳定判别规则提取方法, 从规则自身获得有关在线监测重点和电网稳定增强措施的参考信息。文献[6]是用改进的粗糙集约简算法来解决大电网大信息量情况下的故障诊断问题。文献[12]是用基于粗糙集理论和朴素贝叶斯网络的电网故障诊断方法, 提高了容错性。

而关联规则挖掘作为数据挖掘技术中一种重要方法, 因能发现隐藏在海量数据中的有关联关系有价值的模式, 对决策的生成具有重要的实用价

值而备受关注,主要用在商业智能解决方案体系中,在目前的电网故障诊断研究中应用不多。文献[16]就是分析了关联规则挖掘在商业智能中的应用,提供了一个模块化、易于集成、适合于最终用户使用的框架。如文献[17]只是在电力营销策略的制定方面,从理论上指出了关联规则数据挖掘技术在电力行业中的应用前景是巨大的。而文献[18]中作者虽然把关联规则数据挖掘应用到机车牵引变流器的故障诊断,但电网故障诊断中与之不同处有很多,例如要考虑保护动作的不确定信息处理等。

基于此,本文尝试将基于关联规则挖掘改进的Apriori算法即DLG算法应用于电网故障诊断中。首先以保护、断路器作为条件属性,故障区域作为决策属性,考察各种故障情况并建立决策表,然后利用关联规则挖掘进行属性约简,通过修改阈值进行交互式挖掘,能有效省去文献[7]中用粗糙集约简属性后再采用基于平均互信息的最佳属性约简组合选择法的计算量,可直接提取最佳属性约简组合,最后利用最佳属性约简组合形成的约简决策表和关联规则交互式挖掘来诊断推理,得出诊断结果。

1 基于关联规则的DM技术

1.1 关联规则基本原理

关联规则挖掘的对象一般是事物数据集。设^[13] $I=\{i_1, i_2, \dots, i_k\}$ 是二进制文字的集合,其中的元素称为项(item)。记 D 为交易 T (transaction)的集合,这里交易 T 是项的集合,并且 $T \subseteq I$ 。对应每一个交易有惟一的标识,如交易号,记作 TID 。设 X 是一个 I 中项的集合,如果 $X \subseteq T$,那么称交易 T 包含 X 。

一个关联规则是形如 $X \Rightarrow Y$ 的蕴涵式,这里 $X \subseteq I, Y \subseteq I$,并且 $X \cap Y = \emptyset$ 。度量关联规则有三个重要指标,它们分别是支持度、置信度和相关度。为方便计,我们用0%~100%之间的值来对应其支持度和置信度。

关联规则 $X \Rightarrow Y$ 的支持度(s)表示在交易集 D 中包含 $X \cup Y$ 的元组所占的比例,用概率论的表达方法即为:

$$s(X \cup Y) = \Pr(X \cup Y) \quad (1)$$

关联规则 $X \Rightarrow Y$ 的置信度或浓度(α)表示交易集 D 中包含 $X \cup Y$ 元组的数目和包含 X 的元组数目的比值,用概率论表达方法即为:

$$\alpha(X \cup Y) = \Pr(X \cup Y) / \Pr(X) \quad (2)$$

第三个指标是用以表征 X 和 Y 之间的相关度(σ),用概率论表达方法即为:

$$\sigma_{X,Y} = \{\Pr(X \cup Y)\} / \{\Pr(X)\Pr(Y)\} \quad (3)$$

若 $\Pr(X \cup Y) = \Pr(X)\Pr(Y)$ 成立,则说明模式 X 的出现是独立于模式 Y 的。因此,若相关度大于1则证明模式 X 和 Y 是正相关的。反之,相关度小于1则证明模式 X 和 Y 是负相关的。若规则负相关,即规则的前件和后件的出现概率是互逆的,此类规则不符合客观逻辑,应该删除。

项的集合称为项集(itemset)。包含 k 个项的项集称为 k -项集。项集的出现频率是包含项集的事务数,简称为项集的频率、支持度或计数。如果项集满足最小支持度(由用户或领域专家设定),则称它为频繁项集。给定一个交易集 D ,挖掘关联规则问题就是产生支持度和置信度分别大于最小支持度(minsupp)和最小置信度(minconf)的关联规则。

基于以上三个指标,有趣的关联规则可定义如下:

关联规则 $X \Rightarrow Y$ 若满足下面三个条件,则称它为有趣的关联规则:

$$s(X \Rightarrow Y) \geq s_{\min} \quad (4)$$

$$\alpha(X \Rightarrow Y) \geq \alpha_{\min} \quad (5)$$

$$\sigma(X \Rightarrow Y) \geq \sigma_{\min} \quad (6)$$

这里 $s_{\min}, \alpha_{\min} > 0, \sigma_{\min} > 1$ 皆可为自定义的数据挖掘阈值。

在电网故障诊断中,我们所要得到的是就是满足式(4)、(5)、(6)的有趣的有价值的强关联规则。

1.2 Apriori算法和DLG算法

最大频繁项集的生成是影响关联规则挖掘的关键问题。DLG算法是基于经典的Apriori算法上的一个改进算法。Apriori算法是一种最有影响的挖掘布尔关联规则频繁项集的算法,算法使用频繁项集性质的先验知识,用逐层搜索的迭代方法来获得频繁项集。 k 项集用于搜索 $k+1$ 项集。首先找到频繁1项集的集合,记为 L_1 。 L_1 用于找频繁2项集的集合 L_2 ,如此下去,直到不能找到频繁 K 项集。利用Apriori性质:频繁项集的所有非空子集都必须也是频繁的,通过连接和剪枝两步过程来实现。

连接步:通过对 L_{k-1} 中的每个元素执行连接,得到了 L_k 的候选集合 C_k 。

剪枝步: C_k 是 L_k 的超集,即它的成员中也有不是频繁的。首先,根据Apriori性质,缩小 C_k 的范围,然后扫描数据库,确定 C_k 中每个候选的计数,从而确定 L_k 频繁项集的挖掘。

而DLG是基于关联图结构,通过减少事务数据库的扫描次数,进而有效减少挖掘过程的I/O代价。

通过关联挖掘出的结果中一个合理的关联规则对应于含有一个属性的约简组合。具体实现如下所述。DLG算法共有三步。第一步是产生频繁1项目集阶段:它产生频繁1项目集并记录相关信息;第

二是关联图的构造,用以展示频繁项目之间的关联;最后一步是频繁 k -项目集($k>2$)产生阶段,它基于关联图的构造来产生 k 频繁项目集。

1.2.1 频繁 1 项集产生阶段

DLG 算法首先扫描一遍数据库以统计其支持度,同时为每个项目建立一个比特向量,每个比特向量的长度是数据库中事物的数目。与项目 i 相关联的比特向量表示为 BVi 。如果一个项目 i 出现在第 j 个事务中,则 BVi 的第 j 个比特置 1;否则, BVi 的第 j 个比特置 0。 BVi 中 1 的数目等于支持项目 i 的事物数,即项目 i 的支持度。

1.2.2 关联图的构造

项集 $\{i_1, i_2, \dots, i_k\}$ 的支持度是 $BVi_1 \wedge BVi_2 \wedge \dots \wedge BVi_k$ 中 1 的数目,这里符号“ \wedge ”是逻辑“与”运算。在经过第一步后,不必再扫描数据库。在关联图构造阶段,DLG 构造一个关联图以表示项目之间的关联。如果 $BVi \wedge BVj(i < j)$ 中 1 的数目不小于最小支持度,则构造一条从项目 i 到项目 j 的有向边,而且项集 $\{i, j\}$ 是一个频繁 2 项集。

1.2.3 频繁项集的产生

基于关联图可以产生频繁 k 项集($k>2$)。在关联图构造阶段,可以产生频繁 2 项集 $L2$ 。在频繁项集产生阶段,DLG 算法产生频繁 $k2$ 项集 $Lk(k>2)$ 对每一个 $Lk(k>2)$ 中的频繁 $k2$ 项集, $k2$ 项集的最后项用来扩展项集为 $(k+1)2$ 项集。设 $\{i_1, i_2, \dots, i_k\}$ 是一个频繁 $k2$ 项集,如果有一个从项目 ik 到项目 j 的有向边,那么项集 $\{i_1, i_2, \dots, i_k\}$ 被扩展为 $k+1$ 项集。如果 $BVi_1 \wedge BVi_2 \wedge \dots \wedge BVi_k \wedge BVj$ 中 1 的数目不小于最小支持度,那么 $\{i_1, i_2, \dots, i_k, j\}$ 是频繁 $k+1$ 项集。如果不能再继续产生频繁 k 项集,则 DLG 算法终止。

2 基于关联规则数据挖掘电网故障诊断方法

2.1 基本思路

利用上述概念,根据本文要解决的电网故障诊断问题,首先,可定义关联规则。

定义:给定一组故障信息 $C = \{C1, C2, \dots, Cm\}$ 、故障区域 $S = \{Sec1, Sec2, \dots, SecM\}$,故障诊断关联规则是形如 $X \Rightarrow Y$ 的蕴含式,其中 X, Y 分别是 C 和 S 的模式。

假若算法实现得到的关联规则的表达式为 $\{C1, C2, C3 \Rightarrow Sec2\}$,即说明由 $C1, C2, C3$ 这些断路器动作信息和保护动作信息来诊断得出的故障区域是 $Sec2$ 。

2.2 算法流程与分析

本文所述算法的流程如图 1 所示。

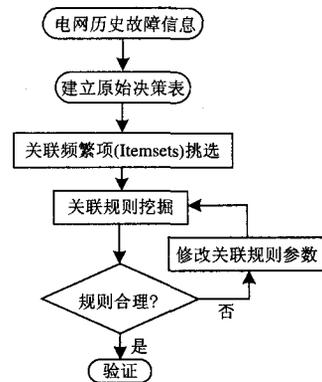


图 1 电网故障诊断关联规则挖掘流程

Fig. 1 Flow chart of power network fault diagnosis

在实际的系统处理中,事务数往往是成千上万的,为仔细说明操作流程,以下面这个简单的配电网为例,该系统有三个区域 $Sec1, Sec2$ 和 $Sec3$ 分别配有过流保护 $CO1, CO2$ 和 $CO3$ 。 $Sec1$ 配有距离保护 $RR1$, 为 $Sec2$ 和 $Sec3$ 提供后备保护。 $CB1, CB2$ 和 $CB3$ 是断路器。其故障信息分别为断路器 $CB1, CB2, CB3$, 过流保护 $CO1, CO2, CO3$, 距离保护 $RR1$, 条件属性取值为“0”或“1”,其中“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,分别为线路 $Sec1, Sec2, Sec3, Sec2/3$ (表示故障发生在 $Sec2$ 或 $Sec3$)、 No (表示无故障)。

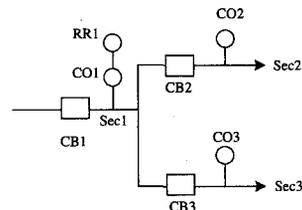


图 2 简单配电网示例结构图

Fig.2 The model of simple distribution network

以保护、断路器动作信息作为故障信息,以故障元件作为决策属性,形成需挖掘的事务,建立原始决策表。每个属性都对应于数据挖掘处理的决策表中的项。样本中的元素值为“1”的属性编号,如事务1为 $\{CB1, CO1, Sec1\}$,共有7个这样的事务,事务数为7。表1是由7个电网历史故障事务组成的原始决策表。

其中事务 1 表示故障发生在 $Sec1$, 过流保护动作 $CO1$ 跳开断路器 $CB1$;

事务 2 表示故障发生在 $Sec2$, 过流保护动作 $CO2$ 跳开断路器 $CB2$;

事务 3 表示故障发生在 Sec3,过流保护动作 CO3 跳开断路器 CB3;

事务 4 表示故障发生在 Sec2,过流保护 CO2 动作,但断路器 CB2 拒动,后备距离保护 RR1 动作跳开断路器 CB1;

事务 5 表示故障发生在 Sec3,过流保护 CO3 动作,但断路器 CB3 拒动,后备距离保护 RR1 动作跳开断路器 CB1;

事务 6 表示故障发生再 Sec2 或者 Sec3,记为 Sec2/3,但过流保护拒动,后备保护 RR1 动作跳开断路器 CB1;

事务 7 表示没有故障发生的情况(记为 NO)。

首先将电网故障事务表输入数据库,形成数据源,作为关联规则挖掘处理的事务表;然后用 C 语言在.NET 平台下实现 DLG 算法即得到其频繁项集。

将算例结果中得到的所有的频繁项目集,存放入文件名为“itemset.txt”文件中。

表 1 电网故障决策表 1

Tab. 1 Decision table of power network fault diagnosis

样本	条件属性(故障信息)							故障区域
	CB1	CB2	CB3	CO1	RR1	CO2	CO3	
1	1	0	0	1	0	0	0	Sec1
2	0	1	0	0	0	1	0	Sec2
3	0	0	1	0	0	0	1	Sec3
4	1	0	0	0	1	1	0	Sec2
5	1	0	0	0	1	0	1	Sec3
6	1	0	0	0	1	0	0	Sec2/3
7	0	0	0	0	0	0	0	No

表 2 关联规则提取表

Tab. 2 The table of the extracted association rules

序号	关联规则	置信度	序号	关联规则	置信度
1	{CO1,CB1=>Sec1}	100%	6	{RR1,CO2=>Sec2}	100%
2	{CO2=>Sec2}	100%	7	{RR1,CO3=>Sec3}	100%
3	{CO3=>Sec3}	100%	8	{RR1=>Sec2/3}	66.7%
4	{CB1,CO2=>Sec2}	100%	9	{CO1=>Sec1}	100%
5	{CB1,CO3=>Sec3}	100%	10	{CB1=>Sec2/3}	75%

需要说明的是若支持度和置信度的阈值过小,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。实用中应先较小的样本中通过修改阈值进行交互式挖掘,以确定合适的阈值,然后再对海量数据进行挖掘,并辅之以支持度或置信度排序技术。最后,在“itemset.txt”中提取出满足 $s(X \Rightarrow Y) \geq s \min, \alpha(X \Rightarrow Y) \geq \alpha \min, \sigma(X \Rightarrow Y) \geq \sigma \min$ 的关联规则。本算例中所用的域值为支持度 $s \min=14.2%$,置信度 α

$\min=65%$,相关度 $\sigma \min=1$ 。若出现关联规则 {CB1=>Sec2/3} 的 $s=14.3%, \alpha=66.7%, \sigma > 1$ 。即说明这条规则是可信正相关的强关联规则,能够被提取。即由属性 RR1,CO2 可推断出故障区域是 Sec2。因此,同时满足 $s(X \Rightarrow Y) \geq s \min, \sigma(X \Rightarrow Y) \geq \sigma \min$ 的条件下,提取的关联规则各置信度如表 2 所示。

因而可构成更为精简的电网故障决策表,如表 3、表 4 所示。

表 3 电网故障样本决策表

Tab. 3 Decision table of power network fault diagnosis

样本	故障信息(条件属性)				故障区域
	CO1	CB1	CO2	CO3	
1	1	1	0	0	Sec1
2	0	0	1	0	Sec2
3	0	0	0	1	Sec3
4	0	1	1	0	Sec2
5	0	1	0	1	Sec3
6	0	1	0	0	Sec2/3
7	0	0	0	0	No

表 4 电网故障样本决策表

Tab. 4 Decision table of power network fault diagnosis

样本	故障信息(条件属性)				故障区域
	CO1	RR1	CO2	CO3	
1	1	0	0	0	Sec1
2	0	0	1	0	Sec2
3	0	0	0	1	Sec3
4	0	1	1	0	Sec2
5	0	1	0	1	Sec3
6	0	1	0	0	Sec2/3
7	0	0	0	0	No

此外,通过表 5 检测样本本来验证该算法得出结果的有效性和容错性。

表 5 检测样本表

Tab. 5 The table of detected samples

序号	设置故障区域	故障信息				输出诊断结果	
		CO1	CB1	CO2	CO3	关联规则	置信度
1	Sec2	0	0	1	0	{CO2=>Sec2}1	100%
2	Sec2	0	1	1	0	{CB1,CO2=>Sec2}	87%
3	Sec2	0	0	1	*	{CO2=>Sec2}2	66.7%
4	Sec2	0	*	1	0		
5	Sec2	*	0	1	0		

表 5 中第 1 组为正确故障信息,第 2 组含有错误故障信息(CB1 本应该是“0”,而错误给出了“1”),第 3, 4, 5 组为不完备故障信息,其中“*”表示未知故障信息)。

诊断结果表明,当故障信息准确且完备时可得

到正确的诊断结果；当故障信息有部分错误时候，仍可得到正确的诊断结果；当故障信息不完备时，由于算法中能同时处理属性为0和不完备即(缺失状态项)，所以仍可得到正确的诊断结果或设定故障区域在输出的诊断结果中。

3 算例分析

针对本文所用的方法，笔者基于C语言编写了故障诊断软件，并用图3所示配电网系统进行了仿真分析。系统由4条母线B1-B4，一台变压器T和四条输电线路L1-L4组成。其中CB1-CB10表示所对应的断路器 MB1, MB2, MB3, MB4, MT, ML2, ML4, ML5, ML6, ML7, ML8, ML9, ML10, BL2, BL4, BL5, BL6, BL7, BL8, BL9, BL10, BT 表示 22 个保护，MB 表示母线保护，ML 表示输电线路主保护，BL 表示输电线路后备保护，ML 表示变压器主保护。根据保护动作原理，考虑下列故障情形，1) 9 个单

重故障：保护和断路器均正常动作；2) 20 个单重故障：考虑了 1 个保护或断路器拒动的情况；3) 11 个双重故障：保护和断路器均正常动作；可建立由 32 个条件属性、41 个样本组成的原始事务表(限于篇幅，此处不列出该事务表)。

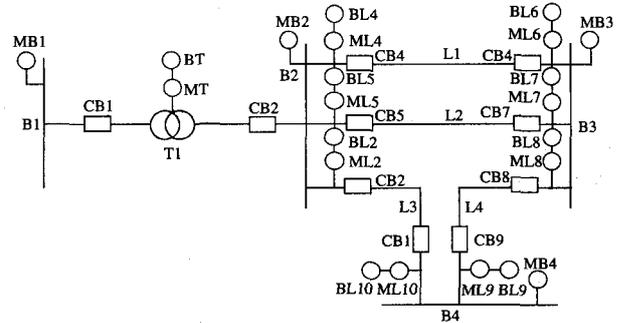


图3 四母线配电网系统

Fig. 3 The four-bus distribution network for simulation

表6 最佳约简决策表

Tab. 6 The best reduction decision table

样本	CB1	CB2	CB4	CB5	CB6	CB7	CB10	MB1	MT	ML2	ML7	ML8	ML9	BL4	BL7	BT	故障
1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	B1
2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	T
3	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	B2
4	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	B3
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	B4
6	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	L1
7	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	L2
8	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	L3
9	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	L4
10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	B1
11	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	B2
12	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	B3
13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	B4
14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	L1
15	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	L1
16	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	L2
17	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	L2
18	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	L3
19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	L3
20	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	L4
21	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	L4
22	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	L1
23	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	L1
24	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	L2
25	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	L3
26	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	L4
27	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	L4
28	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	B1,T
29	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	B2,T
30	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	B2,L1
31	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	B2,L2
32	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	B2,L3
33	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	L1,L2
34	0	1	0	1	0	1	1	0	0	1	1	0	0	0	0	0	L2,L3
35	0	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	L3,L4
36	0	0	0	1	0	1	0	0	0	0	1	1	1	0	0	0	L2,L4
37	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	B3,L4
38	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	B3,L1
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	No

首先将电网故障决策表输入数据库，形成数据

源,作为关联规则挖掘处理的事务表;然后用所编写

的软件实现 DLG 算法得到其频繁项集。通过修改阈值进行交互式挖掘, 以确定合适的阈值, 该仿真中, 确定的适合的域值为支持度 $s \min = 20\%$, 置信度 $\alpha \min = 60\%$, 相关度 $\sigma \min = 1$ 。最后, 提取出满足 $s(X \Rightarrow Y) \geq s \min$, $\alpha(X \Rightarrow Y) \geq \alpha \min$, $\sigma(X \Rightarrow Y) \geq \sigma \min$ 的若干关联规则。根据提取出的符合条件的关联规则即可得到最佳属性约简组合, 由它形成的最佳约简决策表如表 6 所示。约简后的决策表保持了与原决策表完全相同的分类能力, 但所需的条件属性从原有的 32 个减少到 16 个。

4 测试结果与比较

预先设定故障位置, 将引起此故障的故障信息输入本文所编写的电网故障诊断软件, 得到诊断结果, 将 2 者进行比较, 如表 7 所示。

由表 7 可见, 样本 1-4 为第一类故障(单重故障), 样本 5-9 为第二类故障(保护或断路器拒动), 样本 10-21 为第三类故障(双重故障)。当故障信息准确且

完整时(如样本 1、3、4、5、7、8、9、11、12、15、21), 可得到正确的诊断结果(此时, 关联规则的置信度为 100%); 当丢失或出错的故障信息不是核属性时(如样本 17、18、19、20), 可得到正确的诊断结果(此时对应的关联规则置信度分别为: 66.7%, 66.7%, 75%, 75%); 当核属性信息部分丢失或出错时(如样本 2、6、10、13、14、16), 其中样本 2 和样本 10 是属于互补的情况, 所以当核属性 MB1 受到干扰时无论用何种诊断算法都无法识别; 样本 13 和样本 16 仍可得到正确的诊断结果(此时对应的关联规则置信度分别为 64.3%, 75%); 而样本 6 和样本 14 不能得到正确的诊断结果, 设定故障区域的故障概率不是最高的(此时对应的关联规则置信度分别为 50%, 58%), 但也排在第二位(“/”前是诊断概率最高的故障区域, “/”后边是故障概率排在第二的故障区域), 对调度人员的决策有辅助作用。

表 7 电网故障诊断结果

Tab.7 The result of fault diagnosis for power network

样本	设定故障位置	CB1	CB2	CB4	CB5	CB6	CB7	CB10	MB1	MT	ML2	ML7	ML8	ML9	BL4	BL7	BT	诊断故障位置及其关联规则置信度
1	B1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	B1; 100%
2	T	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	B1; 60% / T; 50%
3	B2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	B2; 100%
4	B3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	B3; 100%
5	B4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	B4; 100%
6	L1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	L2; 61.5% / L1; 50%
7	L2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	L2; 100%
8	L3	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	L3; 100%
9	L4	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	L4; 100%
10	B1,T	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	T; 58% / B1.T; 56%
11	B2,T	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	B2.T; 100%
12	B2,L1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	B2.L1; 100%
13	B2,L2	0	1	1	1	0	1	0	—	0	0	1	0	0	0	0	0	B2.L2; 64.3%
14	B2,L3	0	1	1	1	0	0	1	—	0	1	0	0	0	0	0	0	B2.L2; 72.7% / B2.L3; 58%
15	L1,L2	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	0	L1,L2; 100%
16	L2,L3	0	1	0	1	0	1	1	1	0	1	1	0	0	0	0	0	L2,L3; 75%
17	L3,L4	0	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0	L3,L4; 66.7%
18	L2,L4	0	0	0	1	0	1	0	1	0	0	1	—	1	0	0	0	L2,L4; 66.7%
19	B3,L4	0	0	0	0	1	—	0	1	0	0	0	1	1	0	0	0	B3,L4; 75%
20	B3,L1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	B3,L1; 75%
21	No	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	No; 100%

5 结论

本文将关联规则的数据挖掘方法 DLG 算法应用于电网故障诊断中, 进行故障样本集的属性约简, 通过算例实现, 得到以下结论:

(1) 用关联规则挖掘进行属性约简时, 通过修改阈值进行交互式挖掘, 能直接得到最佳约简事务

表, 能省去参考文献[12]中用粗糙集约简属性后再采用基于平均互信息的最佳属性约简组合选择法的计算量。

(2) 在故障信息准确且完整, 丢失或出错的故障信息不是核属性的情况下, 算例结果表明仍能迅速得到正确的故障诊断结果, 正确率高。当核属性信息部分丢失或出错时, 正确诊断率不高, 但当出现

设定故障区域的故障概率不是最高时,也排在第二位,对调度人员的决策有辅助作用,表明该方法有一定的容错性。

(3) 利用 DLG 算法,其关联图结构减少了 Apriori 系列算法两种可能的开销:可能需要产生大量候选集和可能需要重复扫描数据库,通过模式匹配检查一个很大的候选集合,进而有效减少挖掘过程的 I/O 代价,提高挖掘模型的计算效率,对提高电网故障诊断的有效性和实时性是极其重要的。

参考文献

- [1] 史述红,杨宛辉,王克文.城市 110 kV 电网故障元件快速识别[J].电网技术,2001,25(6):37-40.
SHI Shu-hong, YANG Wan-hui, WANG Ke-wen. Identification of Faulty Components in 110 kV Urban Power Networks[J]. Power System Technology, 2001, 25(6): 37-40(in Chinese).
- [2] 文福拴,韩祯祥,田磊,等.基于遗传算法的电力系统故障诊断的解析模型与方法,第一部分:模型与方法[J].电力系统及其自动化学报,1998,10(3):1-7.
WEN Fu-shuan, HAN Zhen-xiang, TIAN Lei, et al. An Analytic Model and Genetic Algorithm Based Methods for Fault Diagnosis in Power Systems, Part I: the Model and Method[J]. Proceedings of the EPSA, 1998, 10(3): 1-7(in Chinese).
- [3] 杜一,张沛超,郁惟镛.基于事例和规则混合推理的变电站故障诊断系统[J].电网技术,2004,28(1):34-37.
DU Yi, ZHANG Pei-chao, YU Wei-yong. A Substation Fault Diagnosis System Based on Case-based Reasoning and Rule-based Reasoning [J]. Power System Technology, 2004, 28(1): 34-37(in Chinese).
- [4] 周明,任建文,李庚银,等.基于模糊推理的分布式电力系统故障诊断专家系统[J].电力系统自动化,2001,25(24):33-36.
ZHOU Ming, REN Jian-wen, LI Geng-yin, et al. Distributed Power System Fault Diagnosis Expert System Based on Fuzzy Inference[J]. Automation of Electric Power Systems, 2001, 25(24): 33-36(in Chinese).
- [5] 汤磊,孙宏斌,张伯明,等.基于信息理论的电力系统在线故障诊断[J].中国电机工程学报,2003,23(7):5-11.
TANG Lei, SUN Hong-bin, ZHANG Bo-ming, et al. Online Fault Diagnosis for Power System Based on Information Theory[J]. Proceedings of the CSEE, 2003, 23(7): 5-11(in Chinese).
- [6] 孙秋野,张化光,戴璟.基于改进粗糙集约简算法的在线故障诊断[J].中国电机工程学报,2007,27(11):58-61.
SUN Qiu-ye, ZHANG Hua-guang, DAI Jing. On-line Fault Diagnose of Distribution Dystem Based on Modified Rough Sets Reduction Algorithm[J]. Proceedings of the CSEE, 2007, 27(11): 58-61(in Chinese).
- [7] 束洪春,孙向飞,司大军,等.基于粗糙集理论的配电网故障诊断研究[J].中国电机工程学报,2001,21(10):73-82.
SHU Hong-chun, SUN Xiang-fei, SI Da-jun, et al. A Study of Fault Diagnosis in Distribution Line Based on Rough Set Theory [J]. Proceedings of the CSEE, 2001, 21(10): 73-82(in Chinese).
- [8] 束洪春,孙向飞,司大军.基于故障投诉电话信息的配电网故障定位粗糙集方法[J].电网技术,2004,28(1):64-66,70.
SHU Hong-chun, SUN Xiang-fei, SI Da-jun. A Study of Fault Diagnosis in Distribution Line Based on Rough Set Theory[J]. Power System Technology, 2004, 28(1): 64-66,70(in Chinese).
- [9] 赵冬梅,韩月,高曙.电网故障诊断的决策表约简新算法[J].电力系统自动化,2004,28(4):63-66.
ZHAO Dong-mei, HAN Yue, GAO Shu. New Reduct Algorithm of Decision Table for Electric Power Grid Fault Diagnosis[J]. Automation of Electric Power Systems, 2004, 28(4): 63-66(in Chinese).
- [10] 束洪春,孙向飞,于继来.粗糙集理论在电力系统中的应用[J].电力系统自动化,2004,28(3):90-95.
SHU Hong-chun, SUN Xiang-fei, YU Ji-lai. A Survey on the Application of Rough Set Theory in Power Systems[J]. Automation of Electric Power Systems, 2004, 28(3): 90-95.
- [11] Lambert-Tones G. Application of Rough Set in Power System Control Center Data Mining[A]. In: Proceeding of the First International Conference on Machine Learning and Cybernetics[C]. Piscataway (NJ): 2002.627-631.
- [12] 张耀天,何正友,赵静.基于粗糙集理论和朴素贝叶斯网络的电网故障诊断方法[J].电网技术,2007,31(1).
ZHANG Yao-tian, HE Zheng-you, ZHAO Jing. A Power Network Fault Diagnosis Method Based on Rough Set Theory and Naive Bayesian Networks[J]. Power System Technology, 2007, 31(1).
- [13] HAN Jia-wei, Kamber M. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann Publishers, 2001.
- [14] 廖志伟,孙雅明.基于数据挖掘模型在高压输电线故障诊断中的应用[J].电力系统自动化,2001,25(15):15-19.
LIAO Zhi-wei, SUN Ya-ming. A New Data Mining Approach for Fault Diagnosis of High Voltage Transmission Line Based on Rough Set Theory[J]. Automation of Electric Power Systems, 2001, 25(15): 15-19.
- [15] 王同文,管霖.基于模式发现的电力系统稳定评估和规则提取[J].中国电机工程学报,2007,27(19):25-31.

- 2005, 29 (9): 82-85.
- [6] Jeon Y J, Kim J C. An Efficient Simulated Annealing Algorithm for Network Reconfiguration in Large Scale Distribution Systems[J]. IEEE Trans on Power Delivery, 2002, 17 (4): 1070-1078.
- [7] 刘莉, 陈学允. 基于模糊遗传算法的配电网网络重构[J]. 中国电机工程学报, 2000, 20 (2): 65-67.
LIU Li, CHEN Xue-yun. Reconfiguration of Distribution Networks Based on Fuzzy Genetic Algorithms[J]. Proceedings of the CSEE, 2000, 20 (2): 65-67.
- [8] 屠强, 郭志忠. 辐射型配电网重构的二次电流矩法[J]. 中国电机工程学报, 2006, 26 (16): 57-61.
TU Qiang, GUO Zhi-zhong. The Quadratic Current Moment Method for Radial Distribution Network Reconfiguration[J]. Proceedings of the CSEE, 2006, 26 (16): 57-61.
- [9] Kashem M A, Ganapathy V, Jasmon G B. A Novel Approach for Network Reconfiguration Based load Balancing in Distribution Networks[J]. Electric Machines and Power Systems, 2000 (28): 415-431.
- [10] 刘蔚, 韩桢祥. 基于支持向量机的配电网重构[J]. 电力系统自动化, 2005, 29 (7): 48-52.
LIU Wei, HAN Zhen-xiang. Distribution System Reconfiguration Based on the Support Vector Machine[J]. Automation of Electric Power Systems, 2005, 29 (7): 48-52.
- [11] 孙健, 江道灼. 一种多目标配电网网络重构算法[J]. 电力系统自动化, 2003, 27(20): 57-61.
SUN Jian, JIANG Dao-zhuo. A New Multi-objective Algorithm for Distribution Network Reconfiguration[J]. Automation of Electric Power Systems, 2003, 27 (20): 57-61.
- [12] 刘健, 武晓滕, 余健明. 考虑负荷不确定性和相关性的配电网重构[J]. 电工技术学报, 2006, 21(12): 54-59.
LIU Jian, WU Xiao-meng, YU Jian-ming. Distribution Network Reconfiguration Considering Load Uncertainty and Dependence[J]. Transactions of China Electrotechnical Society, 2006, 21 (12): 54-59.
- [13] 刘健, 徐精求, 董海鹏. 配电网静态安全分析[J]. 电力系统自动化, 2003, 17 (27): 26-30.
LIU Jian, XU Jing-qiu, DONG Hai-peng. Security Analysis of Distribution Networks[J]. Automation of Electric Power Systems, 2003, 17 (27): 26-30.
-
- 收稿日期: 2008-06-19; 修回日期: 2008-10-20
- 作者简介:
伍利(1970-), 女, 博士, 研究方向为电力系统规划与运行; E-mail: wuli@xaut.edu.cn
姚李孝(1962-), 男, 博导, 研究方向为电力系统规划与运行;
彭金宁(1971-), 男, 硕士, 研究方向为电力系统发电与调试.
-
- (上接第 14 页 continued from page 14)
- WANG Tong-wen, GUAN Lin. Power System Stability Assessment and Rule Extraction Based on Pattern Discovery[J]. Proceedings of the CSEE, 2007, (19): 25-31. (in Chinese).
- [16] 郭迅华, 陈国青. 基于 XML 的通用关联规则挖掘应用模式[J]. 管理工程学报, 2005, (4): 53-59.
GUO Xun-hua, CHEN Guo-qing. An XML-based General-purpose Application Model for Association Rule Mining[J]. Journal of Industrial Engineering and Engineering Management, 2005, (4): 53-59.
- [17] 张斌, 孙哲. 关联规则技术在电力系统中的应用[J]. 电网技术, 2007(S1): 63-66.
ZHANG Bin, SUN Zhe. Application of Association Rules Techniques on Power Systems[J]. Power System Technology, 2007, (S1): 63-66.
- [18] 王轶. 基于数据挖掘的机车牵引变流器故障诊断[D]. 成都: 西南交通大学, 2005.
WANG Yi. The Fault Diagnosis of Locomotive Converter Based on Data Mining[D]. Chengdu: Southwest Jiaotong University, 2005.
-
- 收稿日期: 2007-06-09; 修回日期: 2008-07-31
- 作者简介:
聂倩雯(1987-), 女, 硕士研究生, 研究方向为牵引供电综合自动化、电网故障诊断技术; E-mail: Wonderful0305@163.com
高玮(1986-), 女, 硕士研究生, 研究方向为牵引供电综合自动化、电网故障诊断技术.