

# 基于粗糙集的改进约简算法和决策树的电网故障诊断模型

栗然<sup>1</sup>, 黎静华<sup>1</sup>, 苏立军<sup>2</sup>

(1. 华北电力大学电气工程学院, 河北 保定 071003; 2. 河北晋州社会保险所, 河北 晋州 050000)

**摘要:** 采用基于粗糙集的决策树方法建立电网故障诊断模型, 可以方便地处理含有非数值特征的、不精确的故障样本, 且当丢失或出错的故障信息不是关键信号时, 不会影响诊断结果, 具有较强的容错能力和适应性。该文基于粗糙集理论, 首先利用可辨识矩阵的改进算法对由断路器和保护为条件属性、考虑各种故障情况所组成的诊断决策表进行简化; 然后采用加权平均粗糙度的概念, 作为选择分离属性的标准, 构造电网故障决策树, 从而实现对电网的故障诊断。通过实例表明, 该方法能有效地约简知识, 具有很强的容错能力, 能准确地判断电网故障以及对其进行定位。与规则表示相比, 决策树直观、易于理解、维护和修改。

**关键词:** 故障诊断; 粗糙集; 加权平均粗糙度; 辨识矩阵; 决策树; 电网

**中图分类号:** TM711 **文献标识码:** A **文章编号:** 1003-4897(2005)18-0001-05

## 0 引言

电网故障诊断主要是对各级各类保护装置产生的报警信息, 断路器的状态变化信息以及电压电流等电气量测量的特征进行分析, 根据保护动作的逻辑和运行人员的经验来推断可能的故障位置和故障类型。由于这一过程很难用传统的数学方法描述, 因此, 基于专家系统 (expert power)<sup>[1]</sup>、人工神经网络 (artificial neural network)<sup>[2]</sup>、模糊理论 (fuzzy theory)<sup>[3]</sup>、遗传算法 (genetic algorithms)<sup>[4]</sup>和 Petric 网络 (Petric net)<sup>[5]</sup>等人工智能技术的电网故障诊断系统已经得到广泛的研究。但是以上的这些方法不仅需要进行复杂的计算, 还需要大量的统计信息与专家经验, 由于电网故障的复杂性, 造成故障知识不完备或检测信息缺失, 使得它们在诊断能力、适用性和知识获取方面还不尽人意。

粗糙集处理不精确问题时, 不需要提供待求解问题所需处理的数据集合之外的任何先验信息, 且通过属性约简具有较强的容错能力, 适合于因保护装置和断路器误动或拒动、通信装置故障等原因造成的不完备信号模式下的电网故障诊断。怎样更有效地利用粗糙集进行属性特征的约简以及构造准确的电网故障诊断模型成为研究的热点。一般的约简算法只能适用于故障样本相容的情况, 对于不相容的样本显得无能为力。在属性约简的基础上建立决策树模型, 能直观得出各个特征信息对诊断决策的贡献的大小, 从而能更有效地处理不完备信息, 比直接建立规则具有更强的容错能力。但是在传统的 D3 算法<sup>[5]</sup>中, 基于信息熵的计算非常复杂, 且对属

性进行了重复的计算和遍历, 对于故障样本较多的情况计算量很大。针对这些问题, 本文首先利用可辨识矩阵的改进算法对由断路器和保护为条件属性、考虑各种故障情况所组成的诊断决策表进行简化, 求出核属性后进一步简化泛化的过程, 比以往的约简算法具有更高的计算效率, 可以有效解决具有大量复杂故障诊断样本的情况; 然后应用粗糙集理论中的加权平均粗糙度的概念, 作为选择分离属性的标准, 构造电网故障决策树, 基于加权平均粗糙度的方法比基于传统信息熵的方法所占用的空间少, 同时避免了在决策树的一条路径上多次检验某一属性的问题。通过实例表明, 该方法能有效地约简知识, 具有很强的容错能力, 能准确地判断电网故障以及对其进行定位。与规则表示相比, 决策树直观、易于理解、维护和修改。

## 1 基于粗糙集的电网故障决策树模型

电网发生故障时, 往往会产生大量的警报信号, 这些信号特征中, 有些是相关的, 有些是独立的。独立的特征能提供互补信息, 因而应加以保留; 相关的特征产生冗余信息, 同时会增加计算工作量, 因而需要加以消除。基于粗糙集理论的属性约简正好为去除这种冗余性特征提供了方便。在诊断过程中, 特征信号往往不完备, 例如保护和断路器发生了误动或拒动, 通信装置故障造成的信号改变或丢失等。这就要求诊断系统具有一定的容错能力, 而粗糙集理论恰恰具有较强的容错能力, 它利用信号特征集合的冗余性, 通过避开传输出错或已丢失的信号, 实现正确的诊断。

基于粗糙集形成电网故障决策树模型的基本步骤如下:

首先将保护和断路器信号作为对故障分类的条件属性集,建立考虑各种可能发生的故障情况的决策表<sup>[6,7]</sup>;然后用基于可辨识矩阵的改进算法对决策表进行约简,得到最优属性约简集,形成新的决策表。这表明,即使信号信息缺少或出现错误时,若该信号信息不属于约简属性集,则得到的诊断结果仍然是正确的,故基于粗糙集的诊断方法具有较强的容错能力。用加权平均粗糙度方法对新的决策表进行分类,最终形成电网故障诊断决策树。决策树为确定的某一故障类型的决策方法提供了清晰的陈述。根据保护和断路器的信号沿着树的分支序进行检验,树的叶结点即为电网故障类型。

图 1 为一个简单电力系统接线图,有 3 个区域 Sec1, Sec2 和 Sec3,分别配有过流保护  $CO_1$ 、 $CO_2$  和  $CO_3$ , Sec1 配距离保护  $RR_1$ ,并为 Sec2 和 Sec3 提供后备保护。 $CB_1$ 、 $CB_2$  和  $CB_3$  是断路器。根据保护动作原理,考虑单一故障列出决策表 1,共 8 个样本。决策表有 7 个条件属性,即  $\{CB_1\}$ ,  $\{CB_2\}$ ,  $\{CB_3\}$ ,  $\{RR_1\}$ ,  $\{CO_1\}$ ,  $\{CO_2\}$ ,  $\{CO_3\}$ 。表 1 中“1”表示断路器打开或保护动作;“0”表示断路器闭合或保护未动作;样本 4 和 5 为断路器拒动情况;样本 6 为保护拒动情况;样本 7 为保护误动情况;No 表示无故障。

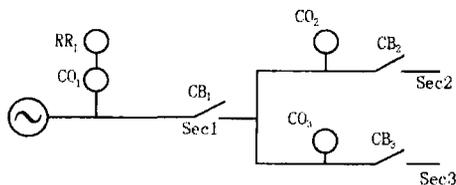


图 1 简单配电系统

Fig 1 A simple distribution network

## 2 属性约简的改进算法

电力系统中大量的故障诊断信号给故障诊断系统快速准确地识别故障造成了一定的困难。

因此,建立故障决策树前进行特征信号空间的压缩具有重要的意义。国内外很多学者对这一问题进行了大量的研究,给出了一系列的约简算法。本文采用基于可辨识矩阵和逻辑运算的改进算法,该算法可以处理具有不相容的故障诊断样本,而一般的约简算法仅仅局限于相容的样本<sup>[8]</sup>。且本文算法在求出核属性后进一步简化了泛化的过程,比以往的约简算法具有更高的计算效率,这对于具有大

量复杂数据的电网故障诊断样本是非常实用的。

表 1 电网故障诊断决策表

Tab 1 Decision table of electric network fault diagnosis

样本	CB <sub>1</sub>	CB <sub>2</sub>	CB <sub>3</sub>	CO <sub>1</sub>	RR <sub>1</sub>	CO <sub>2</sub>	CO <sub>3</sub>	Fault
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
7	1	1	0	1	0	1	0	Sec2
8	0	0	0	0	0	0	0	NO

定义 1: 设  $S = (U, A, V, f)$  为一个决策表,  $|U| = n_0$ 。决策表  $S$  的可辨识矩阵  $M(S)$  是一个  $n \times n$  矩阵,其第  $i$  行第  $j$  列的元素为:

$$C_{ij} = \begin{cases} |a| & a \in c, f(u_i, a) \neq f(u_j, a), \\ & \text{且 } u_i, u_j \text{ 中至少有一个在 } pos_x(D) \text{ 中} \\ 0 & \text{其它} \end{cases} \quad (1)$$

该辨识矩阵定义适用于相容的和不相容的决策表约简。下面给出改进约简算法的步骤:

步骤 1: 按定义 1 计算决策表的可辨识矩阵  $M(S)$ ; 根据决策表 1, 按照定义 1 得到其相应的辨识矩阵如表 2 所示; 因为可辨识矩阵是关于对角线对称的, 所以只需计算下三角的元素即可。

步骤 2: 令  $P = \{a \mid C_{ij} = \{a\}, a \subseteq c\}$ , 即  $P$  为  $M(S)$  中所有单属性元素所组成的属性集合, 若  $M(S)$  中元素  $C_{kl}$  包含  $P$  中元素, 则令  $C_{kl} = 0$ , 得一新矩阵。对新矩阵中的所有非零元素  $C_{ij}$ , 建立相应的析取逻辑表达式  $L_{ij}$ ,  $L_{ij} = a_i \wedge c_{ij}$ ,  $a_i$ ; 在表 2 中仅含有  $CO_2$  和  $CO_3$  两个单属性元素。因此, 这个故障诊断的决策信息系统存在两个核属性, 即  $CO_2, CO_3$ 。把辨识矩阵中含有核属性的元素置 0, 从而得到新的辨识矩阵如表 3 所示。

步骤 3: 将所有的析取逻辑表达式  $L_{ij}$  进行合取运算, 得到一合取范式  $L$ ; 即

$$L = \bigwedge_{i,j} c_{ij} \wedge L_{ij}$$

步骤 4: 将合取范式  $L$  转换为析取范式, 得

$$L = \bigvee_i L_i$$

步骤 5: 输出属性约简结果。将  $P$  中所有属性加入到析取范式中的每一个合取项就对应于一个属性约简的结果。

在该例中, 决策表 1 的属性约简为  $P = \{CO_2, CO_3\}$

$RR_1\} \{CB_1, CO_1\} \{CB_1, RR_1\}$ 。得到三个属性约简为  $\{CO_1, RR_1, CO_2, CO_3\}, \{CB_1, CO_1, CO_2, CO_3\}, \{CB_1, RR_1, CO_2, CO_3\}$ 。下面以约简  $\{CB_1, CO_1, CO_2, CO_3\}$ 为例构建决策树。

该改进约简算法与一般文献的约简算法相比,因为省略了一些不必要计算元素的计算过程,在能够得到相同约简结果的同时,具有更高的效率<sup>[8]</sup>。

表 2 对应于表 1 的辨识矩阵

Tab 2 Discernibility matrix according to Tab 1

	1	2	3	4	5	6	7	8
1	0							
2	$CB_1, CB_2, CO_1, CO_2$	0						
3	$CB_1, CB_3, CO_1, CO_3$	$CB_2, CB_3, CO_2, CO_3$	0					
4	$CO_1, RR_1, CO_2$	0	$CB_1, CB_3, RR_1, CO_2, CO_3$	0				
5	$CO_1, RR_1, CO_3$	$CB_1, CB_2, RR_1, CO_2, CO_3$	0	$CO_2, CO_3$	0			
6	$CO_1, RR_1$	0	$CB_1, CB_3, RR_1, CO_3$	$CO_2$	$CO_3$	0		
7	$CB_2, CO_2$	0	$CB_1, CB_2, CB_3, CO_1, CO_2, CO_3$	0	$CB_2, CO_1, RR_1, CO_2, CO_3$	0	0	
8	$CB_1, CO_1$	$CB_2, CO_2$	$CB_3, CO_3$	$CB_1, RR_1, CO_2$	$CB_1, RR_1, CO_3$	$CB_1, RR_1$	$CB_1, CB_2, CO_1, CO_2$	0

表 3 对应于表 2 的新辨识矩阵

Tab 3 New discernibility matrix according to Tab 2

	1	2	3	4	5	6	7	8
1	0							
2	0	0						
3	0	0	0					
4	0	0	0	0				
5	0	0	0	0	0			
6	$CO_1, RR_1$	0	0	0	0	0		
7	0	0	0	0	0	0	0	
8	$CB_1, CO_1$	0	0	0	0	$CB_1, RR_1$	0	0

### 3 基于加权平均粗糙度的决策树

在构造决策树的过程中,分离属性选择的标准直接影响分类的结果。本文采用加权平均粗糙度 (Weighted and Roughness)作为分离属性的标准。用该方法构造的决策树与传统的基于信息熵方法构造的决策树相比,复杂性低,且能有效提高分类结果<sup>[10]</sup>。

定义 2:对决策系统  $S = (U, A, V, f)$ ,  $X \subseteq U, B \subseteq A, u_B = \text{Card}(B(X)) / \text{Card}(B(X))$ 是定义在空间  $B$  的精确度,  $0 \leq u_B \leq 1$ 。  $X$ 关于  $B$  的加权平均粗糙度定义为:

$$r_B(i) = 1 - \left( \sum_{j=1}^m w_j u_B(X_j) \right) \quad (2)$$

其中:  $i$ 表示第  $i$ 个条件属性;  $j$ 表示决策属性的第  $j$ 个等价类,  $j = 1, \dots, m$ , 其中  $m$ 是决策属性等价类的数目;  $X_j$ 表示决策属性第  $j$ 个等价类集合,  $X_j$

$\subseteq U$ ;  $w_j$ 表示  $X_j$ 在整个论域  $U$ 上所占的比例  $w_j = \text{Card}(X_j) / \text{Card}(U)$ 。

$r_B(i)$ 的取值范围是  $[0, 1]$ ,若  $r_B(i) = 0$ ,在第  $i$ 个属性上将是一个确定的划分,其中不包含不确定因素;若  $r_B(i) = 1$ ,在第  $i$ 个属性上将是一个最大的不确定划分。因此  $r_B(i)$ 的值越小,在第  $i$ 个属性上的划分就越确定,在这里加权值就是为了更全面地考虑决策属性的每个划分对分类的贡献。根据该原理,得到构造基于加权平均粗糙度的决策树的步骤:

步骤 1:根据划分属性的度量标准 (加权平均粗糙度最小)选择最佳条件属性。

步骤 2:用选择的属性去划分训练集,相应于该属性的每一个取值产生一个分支 (子表),再在每一个分支中用同样的方法选择划分属性。

步骤 3:递归地应用到每一个分支上。如果某一分类中的所有事例都来自于一个类别,那么就产生一个标有该类别名的叶结点,直到树中所有的分支都到达叶结点为止。

根据介绍的加权平均粗糙度来构造决策树。首先,由粗糙集理论计算条件属性相对于不同决策属性集合的上、下近似及加权平均粗糙度,通过计算得出 4 个条件属性的加权平均粗糙度为:

$$(CB_1) = 1; \quad (CO_1) = 1$$

$$(CO_2) = 23/32; \quad (CO_3) = 24/32$$

显然,条件属性  $CO_2$  的加权平均粗糙度最小,所以将  $CO_2$  作为根。并由它在表 1 中的 2 个属性值

构造出 2 个分支。每个分支对应于原信息表的一个子表,当  $CO_2 = 1$  时,分支中子表里的事例来自于同一类别,因此将 Sec2 作为该分支的叶结点。当  $CO_2 = 0$  时,计算其余 3 个属性的加权平均粗糙度为:

$(CB_1) = 1, (CO_1) = 1, (CO_3) = 3/5$ , 因此选择  $CO_3$  作为该分支的划分属性。依次得到如图 2 所示的决策树。

基于加权平均粗糙度的方法比基于传统信息熵的方法所占用的空间少;又由于前者比后者得到的叶结点更少,更易于决策;同时避免了在决策树的一条路径上多次检验某一属性的问题<sup>[10]</sup>。

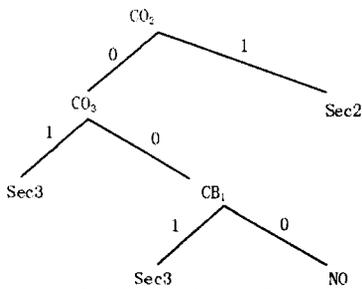


图 2 对应表 1 的决策树

Fig 2 Decision tree according to Tab 1

#### 4 故障决策分析

由图 2 的决策树可知,当所给的信息中有  $CO_2 = 1$  时,可判定故障区域为 Sec2;当所给的信息中  $CO_2 = 0, CO_3 = 1$  时,可判定故障区域在 Sec3;所给信息中  $CO_3 = 0, CO_2 = 0, CB_1 = 0$  时,可以判定该电网无故障;当所给信息中  $CO_3 = 0, CO_2 = 0, CB_1 = 1, CO_1 = 0$  时,可以判定该电网的故障区域在 Sec1。在实际应用中,当所给信息有误时,如:  $CB_1 = 0, CO_1 = 1, CB_2 = 1, CO_2 = 1$  时,由于保护的  $CO_1$  误动,导致信息有误,此时,从上述决策树中推断:  $CO_2 = 1$  电网故障区域在 Sec2。可见,仍能得出正确结果。用决策树进行电网的故障诊断较易于维护和修改,当需要增加记录时,只需在原来的基础上沿途增加即可,无需对新的决策表进行重新计算,比用规则表示易于维护。由于决策树是基于粗糙集理论来建造的,当有些信息丢失或传输发生错误时,可通过冗余约简满足诊断要求,因此该方法具有较强的容错能力。

#### 5 结论

本文给出了利用可辨识矩阵的属性约简改进算法对属性进行约简的方法,实现了在不完备信号信息的情况下电网的故障诊断,体现了良好的容错性

能;采用加权平均粗糙度构造电网故障诊断决策树,该算法比传统的 D3 算法简单、准确易于实现,可方便地处理具有非数值特征的模式,能从故障样本中学习判定规则,适用于大量样本的自动处理,故用于电网故障诊断是适宜的;与规则表示相比,决策树更易于维护和修改,当有记录加入或修改时,只需在原来的基础上进行修改即可,而不用像规则表示那样全部需要重新计算且用树型结构表示知识非常直观,使人容易理解。

#### 参考文献:

- [1] XU Ka-li, ZHOU Ming, REN Jian-wen An Object-oriented Power System Fault Diagnosis Expert System [A]. International Conference on Electrical Engineering Hong Kong 1999.
- [2] WANG Xiao-ru, WU Si-tao, QIAN Qing-quan Neural Network Approach to Power Transmission Line Fault Classification [A]. International Conference on Information, Communication and Signal Processing USA: 1997.
- [3] Monsef H, Ran Jbar A M, Jadid S Fuzzy Rule-based Expert System for Power System Fault Diagnosis [J]. IEEE Proceedings—Gener, Transm and Distrib, 1997, 144 (2): 186-192.
- [4] WEN Fu-shuan, HAN Zhen-xiang A Refined Genetic Algorithm for Fault Section Estimation in Power Systems Using the Time Sequence Information of Circuit Breakers [J]. Electric Machines and Power Systems, 1996, 24: 801-815.
- [5] 马瑞,颜宏文.基于判决树的电力市场中发电商竞标能力数据挖掘的新思路 [J]. 电力系统自动化, 2002, 26 (15): 22-25.  
MA Rui, YAN Hong-wen A New Data-mining Framework Based on Decision-tree for Competitive Bidding Assessment in Power Market [J]. Automation of Electric Power Systems, 2002, 26 (15): 22-25.
- [6] Lo KL, Ng H S, Trecat J. Power Systems Fault Diagnosis Using Petri Nets [J]. IEE Proceedings—Gener, Transm and Distrib, 1997, 144 (3): 231-236.
- [7] 束洪春,孙向飞,司大军.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): 73-77.  
SHU Hong-chun, SUN Xiang-fei, SIDa-jun The Study of Distribution System Fault Diagnosis Based on Rough Set Theory [J]. Proceedings of the CSEE, 2001, 21 (10): 73-77.
- [8] 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 Dongmei, 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
- [9] 刘文军,谷云东,冯艳宾,等. 基于可辨识矩阵和逻辑运算的属性约简算法的改进 [J]. 模式识别与人工智能, 2004, 17(1): 119-123.
- LI Wen-jun, GU Yun-dong, FENG Yan-bin, et al An Improved Attribute Reduction Algorithm of Decision Table [J]. Pattern Recognition and Artificial Intelligence, 2004, 17(1): 119-123.
- [10] 蒋芸,李战怀,张强,等. 一种基于粗糙集构造决策树的新方法 [J]. 计算机应用, 2004, 24(8): 21-23.
- JIANg Yun, LI Zhan-huai, ZHANG Qiang, et al New Method for Constructing Decision Tree Based on Rough Sets Theory [J]. Computer Application, 2004, 24(8): 21-23.
- 收稿日期: 2004-12-23; 修回日期: 2005-06-08
- 作者简介:  
栗 然 (1965 - ),女,副教授,硕士生导师,从事人工智能在电力系统中的应用、电网调度自动化、电网调度运行管理;数据仓库和数据挖掘技术在电力系统中的应用等; E-mail: liranhb@sohu.com  
黎静华 (1982 - ),女,硕士研究生,从事人工智能在电力系统中的应用;  
苏立军 (1964 - ),男,工程师,从事人工智能应用研究。

### Application of rough set to build electric power grid fault diagnosis model based on decision tree

L I Ran<sup>1</sup>, L I Jing-hua<sup>1</sup>, S U Li-jun<sup>2</sup>

(1. School of Electrical Engineering, North China Electric Power University, Baoding 071003, China;

2. Jinzhou Society Insurance Institute, Jinzhou 050000, China)

**Abstract:** By employing decision tree to build the model of electric power grid fault diagnosis, fault sample with non numerical and inaccuracy values can be processed, diagnosis result can be learned when information is corrupted, erroneous and even missing because of its strong fault tolerant ability and better adaptive capability. Based on rough set theory, the paper proposes an improved discernibility matrix to reduce the decision table which includes all kinds of fault cases with the signals of protection relays and takes circuit breakers as condition attributes and fault sections as value attributes. Then the decision tree of electric power grid diagnosis is built by using weighted mean roughness as separating attributes standard to realize electric power grid fault diagnosis. An example shows that the presented method can reduce attributes effectively with strong fault tolerant ability and the fault of electric power grid can be identified accurately. Compared with rule sets, decision tree is easier to maintain and modify. Furthermore, the decision tree is simple and easy to comprehend.

**Key words:** fault diagnosis; rough sets; weighted mean roughness; discernibility matrix; decision tree; electric power grid

### 2006年《继电器》征订启事

电力系统继电保护类专业期刊,集三十年功力,精心打造“继电保护”专栏。专业范围:电力系统继电保护;电力系统分析与控制;电力系统自动化;电力市场;电能质量;人工智能技术在电力系统的应用等电力相关专业。

设置栏目:继电保护理论研究、继电保护运行技术、电力系统分析与控制、电力系统规划设计与可靠性、自动化及运动技术等。

郑重承诺:推行试订,若读者对杂志不满意,可随时退订,余款退回。

1. 全国邮局均可订阅:邮发代号 36 - 135。国内定价:12元/期,全年 24期共 288元。

2. 登陆中国刊物订阅网(www. bk114. com)、看看网(www. kankan. com. cn)可网上订阅。

3. 杂志社订阅:

邮局汇款 地址:河南省许昌市许继大道 32号 461000 收款人:继电器杂志社

银行转帐 户名:许昌继电器研究所 帐号:1708023029021010058 - 051 开户行:许昌市工行五一路支行

针对在本杂志社订阅者,全年订阅享受 8.5折优惠,连续两年或订阅 2套杂志以上者享受 8折优惠,并有机会赠送 04、05年合订本光盘。

订阅电话:0374-3212444 传真:0374-3360297 E-mail: faxing@xjgc.com

注:发邮件告知您的姓名、联系电话、地址和工作单位及部门,经核实可获得免费赠刊。