

基于粗糙集理论和信息融合的变电站故障诊断方法

刘菁, 解大

(上海交通大学电气工程系, 上海 200030)

摘要: 由于变电站故障诊断信息中含有大量不确定、噪声数据, 提出了利用粗糙集理论对大量的诊断征兆信息进行分类, 从而获得简约的规则。然后从单个决策表中获得多个简约的决策规则集, 再利用信息融合来综合这些简约的规则集进行故障的诊断, 从而提高诊断的效率和准确性。从最后仿真比较的结果中可以看出该方法对含有噪声的数据也能获得较好的诊断准确率。

关键词: 粗糙集理论; 信息融合; 故障诊断; 电力系统

Substation fault diagnosis method based on rough set theory and information fusion

LIU Jing, XIE Da

(Department of Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Application of rough sets is made to classify a large number of fault diagnosis information which contains much uncertain noise data when they are transferred from substation. Rough set theory is used to reduce the decision table formed by fault diagnosis information and several decision rules will be gained from single decision table and they are reduced all together for sequent information fusion. All reduced rules from each decision table will be used to diagnose fault after data fusion. This approach can improve diagnosis's efficiency and precision. The simulation result at last shows that excellent precision can be obtained even if data contains noise.

Key words: rough set theory; information fusion; fault diagnosis; power system

中图分类号: TM76

文献标识码: A

文章编号: 1003-4897(2007)06-0005-05

0 引言

变电站故障诊断就是从变电站的某些检测量中得到故障征兆信息, 再通过对这些故障征兆信息的分析与处理, 判断出故障源的位置。其中检测量由厂站监控系统和故障录波器等提供, 故障征兆包括保护开关动作, 断路器跳闸等, 而需判断的故障源通常有母线、变压器、进线出线, 无功补偿设备等。

目前变电站故障诊断的常用方法有: 逻辑处理、专家系统^[1]、神经网络^[2]、Petri网^[3]、模糊理论^[4,5]及遗传算法^[6,7]等方法。这些方法对于准确、完整的故障征兆信息都具有较好判断分类能力, 但是由于电网本身的复杂性, 影响故障征兆信息不确定、不精确、带有噪声的原因很多, 如采样装置故障、信息传输线路受到干扰、误操作等, 这些因素给调度人员正确、快速的操作带来了困难。所以为了提高报警处理的快速性与正确率, 需要一个有力

的工具来帮助调度人员进行有效地处理。

粗糙集理论由波兰学者 Pawlak Z 于 1982 年首先提出, 是一种研究不确定和不精确性知识的数学工具。它的特点是: 即使对于缺少先验知识, 并带有不确定、不精确、含有噪声的数据, 也能够保持分类能力不变的情况下, 通过属性的简约, 得出概念的分类规则。本文针对变电站的保护开关、断路器动作, 利用粗糙集理论对变电站故障征兆数据进行有效地分类和提炼, 从而获得准确的规则。并利用不同的用户定义属性从一组决策数据中导出多组简约的决策规则集, 再利用基于证据理论的数据融合技术来综合这些规则集进行诊断, 更加提高了诊断的准确性。

1 粗糙集理论

1.1 粗糙集的基本模型

粗糙集的基本模型可以由一个四元组的信息

系统 $S = (U, Q, V, f)$ 表示。其中， $U = \{x_1, x_2, \dots, x_n\}$ 是一个有限的非空对象集合，称为论域； Q 是属性的有限集，通常将该集合分成两部分，即：条件属性集合 C 与决策属性集合 D ，且 $Q = C \cup D$ ； $V = \bigcup_{q \in Q} V_q$ 是所有属性值域的并集，其中： V_q 是属性 q 的值域； f 是一个映射，表示每个对象的每个属性对应的值，即： $f: U \times Q \rightarrow V$ 。

在一个信息系统中，对于任何一个属性集合 $P \subseteq Q$ ，不可分辨关系定义为^[9]：

$$IND(P) = \{(x, y) \in U \times U : f(x, a) = f(y, a)\} \quad (1)$$

其中： $\forall a \in P$ 。对于 P 上的不可分辨关系， U 可以分成几个等价类，用 $U / IND(P)$ 表示。

对于任何一个对象集合 $X \subseteq U$ 和属性集合 $P \subseteq Q$ ， P 的下近似集和上近似集定义为

$$\underline{apr}_P(X) = \bigcup \{Y \in U / IND(P) : Y \subseteq X\} \quad (2)$$

$$\overline{apr}_P(X) = \bigcup \{Y \in U / IND(P) : Y \cap X \neq \Phi\} \quad (3)$$

1.2 属性的简约

信息系统 $S = (U, Q, V, f)$ ， $Q = C \cup D$ 中：属性集 C 的分辨矩阵 $M(C) = (m_{i,j})_{n \times n}$ 定义为

$$(m_{i,j}) = \begin{cases} \Phi \\ \{c \in C : f(x_i, c) \neq f(x_j, c)\} \end{cases} \quad (4)$$

第一种情况为：当 $x_i, x_j \in D$ 的不同的等价类；

第二种情况为：当 $x_i, x_j \in D$ 的同一个等价类。

核心定义为：

$$CORE(C, D) = \{c \in C : m_{ij} = \{c\}, 1 \leq j < i \leq n\} \quad (5)$$

其中： m_{ij} 为分辨矩阵中的所有一维集合的元素。

下面将给出一个基于属性重要性的属性简约算法。

属性 P 对 R 的依赖度定义为

$$\gamma_R(P) = \frac{\text{card}(POS_R(P))}{\text{card}(U)} \quad (6)$$

其中

$$POS_R(P) = \bigcup_{X \in U / IND(P)} \underline{apr}_R(X) \quad (7)$$

式 (7) 表示属性集 R 在 $U / IND(P)$ 中的正区域，式 (6) 中的 $\text{card}(\bullet)$ 表示集合的基数。

属性的重要性定义为

$$SGF(a, R, P) = \gamma_R(P) - \gamma_{R-\{a\}}(P) \quad (8)$$

即在属性集 R 中去除属性 a 后，属性集 P 对 R

的依赖度的改变。如果 $SGF(a, R, P)$ 为 0 则说明 a 是冗余属性。令 P 为决策属性， R 为条件属性集，则当 a 为冗余属性时，即 $\gamma_R(P) = \gamma_{R-\{a\}}(P)$ 时就可以将 a 从 R 中剔除，此时并不会影响决策属性的分类 $U / IND(P)$ 。

基于属性重要性的属性简约算法如下：

输入：信息系统 $S = (U, Q, V, f)$ ， $Q = C \cup D$ ，

用户定义属性为 UA 。

输出：简约后的属性集 C 。

算法的步骤：

- (1) 计算核心集 $CORE$ ；
- (2) $REDU = REDU \cup UA$ ；
- (3) $AR' = AR - REDU$ ；
- (4) WHILE $\gamma(REDU, D) \neq \gamma(AR, D)$, DO
选择 $a \in AR'$ 满足：
 $SGF(a, AR', D) = \text{Max}\{SGF(a_i, AR', D)\}$
 $REDU = REDU \cup \{a\}$, $AR' = AR' - \{a\}$
END WHILE
- (5) FOR $I = 0$ to $|REDU| - 1$ DO
IF $a_i \notin CORE \cup UA$ THEN
 $REDU = REDU - \{a_i\}$
IF $\gamma(REDU, D) \neq \gamma(AR, D)$ THEN
 $REDU = REDU \cup \{a_i\}$
END IF
END IF
END FOR

以上算法步骤中 $a_i \in AR'$, $i = 1, 2, \dots, m$ 。

2 证据理论的基本原理

2.1 概率分配函数

假设 $D = \{x_1, x_2, \dots, x_n\}$ 是一个元素的有限集合，概率分配函数就是把 D 上的任意一个子集 A 都映射到 $[0, 1]$ 上的一个函数 M ，而 2^D 表示所有子集组成的集合。当 A 对应一个命题时， $M(A)$ 就对应命题不确定性的量度。且函数 M 满足：

$$M(\Phi) = 0, \sum_{A \subseteq D} M(A) = 1 \quad (9)$$

2.2 信任函数

如果函数 $Bel(x)$ 将集合 2^D 映射到区间 $[0, 1]$ 上，并且满足：

$$Bel(A) = \sum_{B \subseteq A} M(B), \forall A \in 2^D \quad (10)$$

则称 $Bel(x)$ 为信任函数。

2.3 似然函数

如果函数 $Pl(x)$ 将集合 2^D 映射到区间 $[0, 1]$ 上, 并且满足:

$$Pl(A) = \sum_{B \cap A \neq \Phi} M(B), \forall A \in 2^D \quad (11)$$

则称 $Pl(x)$ 为似然函数。

2.4 概率分配函数的正交和

有时对于相同的证据, 由于数据的来源不同, 可能会得到多个不同的概率分配函数 M_1, M_2, \dots, M_n , 此时它们的正交和 $M = M_1 \oplus M_2 \oplus \dots \oplus M_n$ 定义为:

$$M(\Phi) = 0$$

$$M(A) = K^{-1} \times \sum_{\bigcap_{A_i=A} 1 \leq i \leq n} \prod M_i(A_i) \quad (12)$$

其中 K 由下式得到:

$$K = \sum_{\bigcap_{A_i \neq \Phi} 1 \leq i \leq n} \prod M_i(A_i) \quad (13)$$

3 基于粗糙集理论和信息融合的变电站故障诊断模型

基于粗糙集理论的变电站故障诊断就是通过粗糙集理论对变电站故障数据处理后进行分类得到规则, 从而判断出故障所在的区域和故障所在的元件, 为调度人员提供参考的数据。用粗糙集理论对决策表进行的关键操作就是决策表的简约, 所以首先要确定决策表中的条件属性和决策属性在模型中所对应的实际量。决策属性显然就是故障区域或故障元件, 而条件属性就是引起这些故障的原因, 如保护开关、断路器的动作。运用粗糙集理论进行条件属性的简约后, 仍然能够保持信息的分类能力。这说明在故障征兆信息中存在许多冗余数据, 这些数据给故障诊断带来了大量的多余的工作量, 去除这些数据后故障诊断的规则变得简洁。故障诊断过程的流程图如图 1 所示。

诊断过程的两点说明:

(1) 某些时候调度人员在决策过程中要特别对一些保护开关、断路器的动作所反映的故障区域、故障元件进行观察, 此时可在用户定义属性集中设置相应的条件属性。如果这些属性对于决策属性是冗余的, 则可能得到一个不同的归约集, 但是与原始的决策表还是有相同的分类能力。

(2) 本文中采用在属性简约算法中设定不同的用户定义属性集来产生不同的简约规则集, 这些不同的

简约规则集既是用于信息融合的不同证据体。

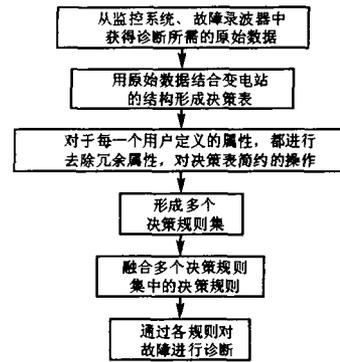


图 1 故障诊断流程图

Fig.1 Flow chart of fault diagnosis

(3) 当采用证据理论作为信息融合的方法进行故障诊断时, 对于每一个被诊断的样本, 在各个规则集中按照条件属性选出相应的诊断规则, 再通过证据理论得出诊断结论的信任度, 如果信任度大于 0.5 则认为该融合后的诊断规则是有效的, 然后就可以用之对样本进行诊断了。

对于不确定知识产生的规则^[9]:

$$\text{If } E \text{ Then } H = \{h_1, h_2, \dots, h_n\},$$

$$CF = \{c_1, c_2, \dots, c_n\}$$

结论 H 的信任度为:

$$f(H) = Bel(H) + \frac{|H|}{|D|} \times [Pl(H) - Bel(H)] \quad (14)$$

其中: $Bel(H) = \sum_{i=1}^n M(\{h_i\});$

$$Pl(H) = 1 - Bel(\sim H);$$

CF 为该条知识的可信度因子。

而 $|H|, |D|$ 分别为结论空间和样本空间的维数。

对于证据 E 的信任度则由经验给出, 如果证据 E 由多个证据合取而成, 即:

$$E = E_1 \cap E_2 \cap \dots \cap E_n \quad (15)$$

则 E 的信任度为合取式中所有证据信任度的最小值。

反之, 如果证据 E 由多个证据析取而成, 即:

$$E = E_1 \cup E_2 \cup \dots \cup E_n \quad (16)$$

则 E 的信任度为合取式中所有证据信任度的最大值。

此时就可以得到 H 的概率分配函数为:

$$M(\{h_1, \{h_2, \dots, \{h_n\}\}) = \{f(E) \times c_1, f(E) \times c_2, \dots, f(E) \times c_n\} \quad (17)$$

4 实例分析

图 2 为上海某变电站线路图。该站有 3 个电压等级, 110 kV、35 kV、10 kV 母线保护分别为: 110B、35B、10B; 变电站进线主保护与备用保护分别为: LLM、LLB; 主变压器和站用变的主保护与备用保护分别为: MTM、MTB、STM、STB; 35kV 母线与 10 kV 母线所带负载出线的主保护与备用保护分别为: 35L1M、35L1B、35L2M、35L2B、10L1M、10L1B、10L2M、10L2B; 无功补偿设备的主保护与备用保护分别为: SVC M、SVC B。另外, 进线上的断路器为: LLBR; 主变 3 个电压等级的断路器分别为: MT110BR、MT35BR、MT10BR; 35 kV、10 kV 母线所带负载出线上的断路器分别为: 35L1BR、35L2BR、10L1BR、10L2BR; 站用变与无功设备与 10 kV 母线间的断路器分别为: STBR、SVCBR。图中的小圈表示相应断路器的流变。

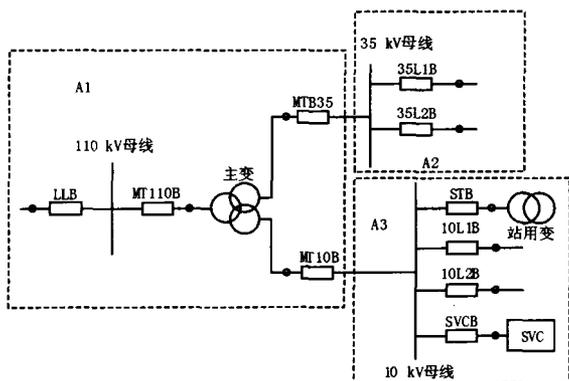


图 2 变电站线路图

Fig.2 Circuit diagram for the substation

图中故障设备有变电站进线 (LL)、110 kV 母线 (110B)、主变 (MT)、35 kV 母线 (35B)、35 kV 侧 2 条出线 (35L1、35L2)、10 kV 母线 (10B)、10 kV 侧 2 条出线 (10L1、10L2)、站用变 (ST)、无功设备 (SVC)。而诊断区域分为 110 kV、35 kV、10 kV 这三块, 如图 3 中的 A1、A2、A3。不同的故障设备所在的区域见表 1。

表 1 故障设备对应区域

Tab.1 Area corresponding to fault equipment

LL	110B	MT	35B	35L1	35L2	10B	10L1	10L2	ST	SVC
A1	A1	A1	A2	A2	A2	A3	A3	A3	A3	A3

首先对于不同的故障点与故障设备得到 50 个样本, 其中设备保护与断路器的动作在决策表中用 1 表示, 不动作用 0 表示。然后用其中的 19 个样本形成决定表, 用其形成诊断规则, 最后用这些规则

对剩余样本进行诊断来验证诊断的效果。

其中各保护与断路器的动作与否为条件属性, 而决策属性为故障区域。然后对该决策表进行简约, 用户定义的属性分别取 110 kV 母线 (110B)、35 kV 母线 (35B)、10 kV 母线 (10B)。三中不同定义属性对应的简约集分别为: {110B, 35L1BR, 35L2BR, 10L1BR, 10L2BR, STBR, SVCBR}、{35B, LLBR, MT110BR, 10L1BR, 10L2BR, STBR, SVCBR}、{10B, LLBR, MT110BR, 35L1BR, 35L2BR}。简约后决策规则集中的属性必然比原来的属性个数少, 这就导致相同的样本出现, 此时就需要进行值简约。即删去所有属性值相同的样本, 只保留一个; 对于只有一个条件属性不同的两个样本, 将其不同的条件属性用 * 表示, 然后合并成一个样本。这三个属性简约集对应的决策规则表分别如表 2、3、4 所示。

表 2 含用户定义属性为 110B 的决策规则

Tab.2 Decision rule containing customized attribute 110B

110B	35L1BR	35L2BR	10L1BR	10L2BR	STBR	SVCBR	AREA
*	0	0	0	0	0	0	1
0	1	*	0	0	0	0	2
0	*	1	0	0	0	0	2
0	0	0	1	1	1	1	3
0	0	0	1	0	0	0	3
0	0	0	0	1	0	0	3
0	0	0	0	0	1	0	3
0	0	0	0	0	0	1	3

表 3 含用户定义属性为 35B 的决策规则

Tab.3 Decision rule containing customized attribute 35B

35B	LLBR	MT110BR	10L1BR	10L2BR	STBR	SVCBR	AREA
0	1	*	0	0	0	0	1
0	*	1	0	0	0	0	1
*	0	0	0	0	0	0	2
0	0	0	1	1	1	1	3
0	0	0	1	0	0	0	3
0	0	0	0	1	0	0	3
0	0	0	0	0	1	0	3
0	0	0	0	0	0	1	3

表 4 含用户定义属性为 10B 的决策规则

Tab.4 Decision rule containing customized attribute 10B

10B	LLBR	MT110BR	35L1BR	35L2BR	AREA
0	1	*	0	0	1
0	*	1	0	0	1
0	0	0	1	*	2
0	0	0	*	1	2
*	0	0	0	0	3

然后模拟采样装置故障、信息传输线路受到干扰、误操作等原因造成的被诊断样本中的噪声, 即

如果原属性值为 1, 则置为 0; 如果原属性值为 0, 则置为 1, 并检测所含噪声比例不同时决策规则的诊断正确率。用所得的决策规则去诊断含有噪声的数据, 如果被诊断样本集中的某一个样本与规则集中的某一条规则相应的条件属性值都相同, 此时如果两者的决策属性值又相同, 说明决策正确, 而如果不相同则说明判断错误。但如果规则集中没有一条规则可以与这个样本相匹配, 则说明无法用该规则集去诊断这条样本。

本文用 19 个样本形成的三个决策规则集及其融合成的规则集分别对含有噪声的样本数据进行诊断。含噪声样本数据由程序在被诊断样本中随机产生, 分别对 6 种噪声占比进行比较。每种噪声占比时的含噪声样本集随机产生 10 次, 然后对每次所得的正确率取平均值作为该噪声占比下决策的正确率。随即产生 10 组样本集是因为噪声分布的不同对不同的决策规则的诊断效果是不同的。最后比较他们的诊断正确率的不同, 如表 5。可以看到融合后决策规则集的决策正确率比单个决策规则集高平均 10 个百分点。

表 5 不同噪声占比下采用不同决策规则与融合决策规则正确率的比较

Tab.5 Precision rate's comparison between different decision rule and fused decision rule as to different noise ratio

噪声占比 / (%)	决策规则 1	决策规则 2	决策规则 3	融合后的决策规则
1	0.9053	0.8912	0.7193	0.9650
2	0.8596	0.8474	0.6807	0.9193
4	0.7842	0.8228	0.6842	0.8877
5	0.7702	0.7649	0.6842	0.8702
10	0.5912	0.5965	0.5754	0.7070
20	0.4439	0.4316	0.4614	0.5246

5 结论

在进行变电站故障诊断时所得到的故障诊断信息中含有大量不精确的, 含有噪声的数据, 本文利用粗糙集理论在即使缺少部分正确数据的情况下, 能以对数据的分类能力处理不确定数据的特点对故障征兆数据进行有效地分类和提炼, 从而获得准确的规则。并利用不同的用户定义属性从单一决策表生成多个简约决策规则集, 再结合基于证据理论的信息融合原理进行故障诊断, 从而在原来的基础上更增加了诊断的准确性。最后用不同的决策规则集对随机产生噪声的样本集进行诊断, 从仿真比

较中可以看出对于不同的噪声占比, 信息融合获得的决策规则集的诊断效果是明显的。

参考文献

- [1] Lee Heung-Jae, Ahn Bok-Skin, Park Young-Moon. A Fault Diagnosis Expert System for Distribution Substations[J]. IEEE Trans on Power Delivery, 2000, 15(1): 92-97.
- [2] 张东英, 钟华兵, 杨以涵, 等. 基于 BP 神经网络和专家系统的变电站报警信息处理系统[J]. 电力系统自动化, 2001, 25(9): 45-47.
ZHANG Dong-yin, ZHONG Hua-bing, YANG Yi-han, et al. Substation Alarm Message Dealing System Based on ANN and ES[J]. Automation of Electric Power Systems, 2001, 25(9): 45-47.
- [3] 任惠, 米增强, 赵洪山. 基于编码 petri 网的电力系统故障诊断模型研究[J]. 中国电机工程学报, 2005, 25(20): 44-49.
REN Hui, MI Zeng-qiang, ZHAO Hong-shan. Power System Fault Diagnosis Modeling Techniques Based on Encoded Petri Nets[J]. Proceedings of the CSEE, 2005, 25(20): 44-49.
- [4] 杨莉, 尚勇, 周跃峰, 等. 基于概率推理和模糊数学的变压器综合故障诊断模型[J]. 中国电机工程学报, 2000, 20(7): 19-23.
YANG Li, SHANG Yong, ZHOU Yue-feng, et al. Probability Reasoning and Fuzzy Technique Applied for Identifying Power Transformer Malfunction[J]. Proceedings of the CSEE, 2000, 20(7): 19-23.
- [5] 钱政, 黄兰, 严璋, 等. 集成模糊数学与范例推理的变压器故障诊断方法[J]. 电网技术, 2001, 25(9): 24-27.
QIAN Zheng, HUANG Lan, YAN Zhang, et al. A Fault Diagnosis Method for Power Transformer Integrating Fuzzy Theory with Case-based Reasoning[J]. Power System Technology, 2001, 25(9): 24-27.
- [6] 邓宏贵, 罗安, 曹建, 等. 基因多点交叉遗传算法在变压器故障诊断中的应用[J]. 电网技术, 2004, 28(24): 1-4.
DENG Hong-gui, LUO An, CAO Jian, et al. Application of Multi-point Criss-cross Genetic Algorithm in Transformer Fault Diagnosis[J]. Power System Technology, 2004, 28(24): 1-4.
- [7] 张建华, 侯国莲, 张巍, 等. 一种基于模糊规则和遗传算法的凝汽器故障诊断方法的研究[J]. 中国电机工程学报, 2004, 24(4): 205-209.
ZHANG Jian-hua, HOU Guo-lian, ZHANG Wei, et al. A Research on Fuzzy Rules and Genetic Algorithm Based Method for a Condenser Fault Diagnosis[J]. Proceedings of the CSEE, 2004, 24(4): 205-209.

(下转第 73 页 continued on page 73)

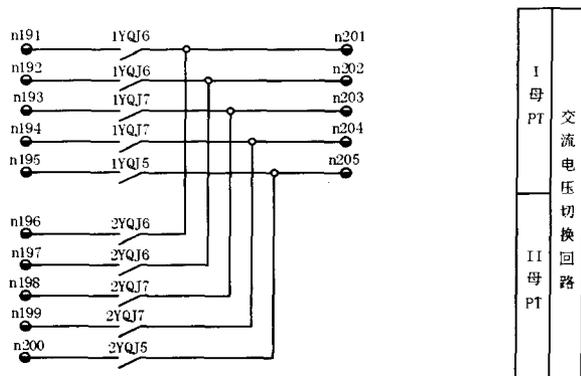


图 6 交流电压切换回路

Fig.6 Alternative voltage switch-over circuit

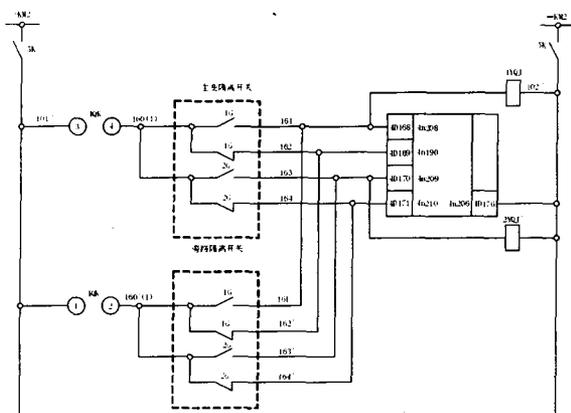


图 7 电压切换继电器控制回路

Fig.7 Controlling circuit of the voltage switch-over relay

经过分析发现问题出在电压切换箱的启动回路上，如上图所示：当主变由本开关 201 挂 I 母运行时，4QK3、4QK4 接通，4QK1、4QK2 打开，主变隔离开关 1G 常开触点闭合，常闭触点打开，161 带正电，1XD 亮；2G 常开触点打开，常闭触点闭合，

164 带正电；旁路隔离开关 1G、2G 常开触点打开，常闭触点闭合，164 的正电通过旁路隔离开关 2G 常闭触点，再通过旁路隔离开关 1G 常闭触点反送至 162，使得电压切换继电器 1YQJ4、5、6、7 返回，导致保护失压。

通过修改控制回路解决了这个问题，修改后的控制回路图如图 8 所示。

当本开关运行时，4QK3~4、4QK7~8 接通，4QK1~2、4QK5~6 断开；当用旁路开关代运行时，4QK1~2、4QK5~6 接通，4QK3~4、4QK7~8 断开。消除了寄生回路。

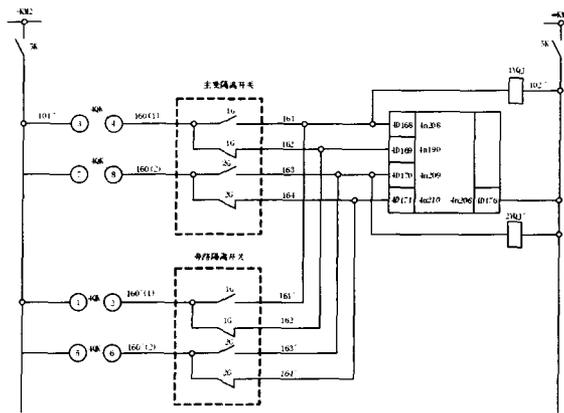


图 8 改进后的电压切换继电器控制回路

Fig.8 Controlling circuit of the voltage switch-over relay after improvement

收稿日期：2006-06-28

作者简介：

刘伟平 (1976-)，男，工程师，从事电力系统继电保护工作。Email: mxlxygz@163.com

(上接第 9 页 continued from page 9)

Application[M].Beijing:Higher Education Press,2004.

[8] 刘同明,等. 数据挖掘技术及其应用[M]. 北京:国防工业出版社,2001.

收稿日期：2006-10-16

作者简介：

刘 菁 (1982-)，男，硕士，主要研究方向为电力系统智能控制和电力系统仿真等；

解 大 (1969-)，男，博士，副教授，主要研究方向为电力系统 FACTS 和电力系统仿真等。E-mail:xieda@sjtu.edu.cn

LIU Tong-ming,et al.Data Mining Technology and Its Application[M].Beijing:National Defense Industry Press,2001.

[9] 张仰森. 人工智能原理与应用[M].北京:高等教育出版社. 2004.

ZHANG Yang-sen.Artificial Intelligence Theory and Its