

基于粗糙集约简技术的输电线路运行风险评估

徐守坤, 韩波, 马正华, 李宁

(江苏工业学院信息科学与工程学院, 江苏 常州 213164)

摘要: 知识约简是粗糙集理论中获取决策规则的重要方式。结合知识约简的概念和定义, 介绍了 Semi-Naive-Scaler 属性离散化算法, 给出了基于可辨识矩阵的属性约简算法, 并采用了基于可辨识矩阵的二值化数据过滤和贪心算法相结合的规则约简算法。通过属性约简和规则约简, 得出决策规则。最后在输电线路运行风险评估方面给出了应用算例。

关键词: 粗糙集; 离散化; 约简; 决策规则; 风险评估

Risk assessment for transmission lines running based on reducing technology of rough sets

XU Shou-kun, HAN Bo, MA Zheng-hua, LI Ning

(School of Information Science & Engineering, Jiangsu Polytechnic University, Changzhou 213164, China)

Abstract: Knowledge reducing is an important way to obtain decision rules in rough sets theory. Combining the concepts and definitions of knowledge reducing, it introduces the attribute discretization algorithm of the Semi-Naive-Scaler, gives attributes reducing algorithm based on discernibility matrix, and also adopts the rules reducing algorithm combining with binarization data filtration based on discernibility matrix and greedy algorithm. Through attribute reduction and rule reduction, it obtains decision rules. Finally, an application example in risk assessment of transmission lines running is given.

Key words: rough sets; discretization; reducing; decision rules; risk assessment

中图分类号: TM75; TM712 文献标识码: A 文章编号: 1674-3415(2009)24-0190-04

0 引言

粗糙集理论是一种新的处理模糊和不确定性问题的数学工具^[1]。其主要思想就是在保持分类能力不变的前提下, 通过知识约简, 导出问题的决策或分类规则。它最显著的特点是除了数据集本身外不需要任何预备的或额外的有关数据信息。粗糙集理论目前已开始应用于我国的电力系统领域中, 给输电线路运行状态下的管理带来了科学性和高效性。

本文中输电线路运行状态下的风险评估是根据线路各相关参数, 应用粗糙集约简方法来分析、计算、衡量输电线路存在的风险性水平的各指标的过程^[2]。之所以采用粗糙集约简方法是因为对于输电线路, 各项状态评估信息表是关系数据库表, 各字段可以看作条件属性, 将总字段离散化后作为决策属性, 与粗糙集理论中的决策表形式吻合。具体来说就是试图发现输电线路各项状态信息与线路运行状态之间的一些内在规律, 提供一些潜在的、有价值的决策信息以供参考。

1 粗糙集中知识约简

知识约简是粗糙集理论的核心内容之一。众所周知, 知识库中的知识(属性)并不是同等重要的, 甚至其中某些知识是冗余的。所谓知识约简, 就是在保持知识库分类能力不变的条件下, 删除其中不相关或不重要的知识。知识约简一般包括属性约简和规则约简两个步骤, 前者是在保持分类能力不变的前提下, 删除不相关或不重要的属性, 即用约简后的属性集合代替原来的属性集合而不降低分类效果。目前人们在属性约简方面已经做了大量的工作, 提出了各种各样的属性约简算法^[3]。后者则在属性约简基础上对决策表中的规则进行约简, 即在不改变规则分类能力的前提下, 删除规则中含有的冗余信息, 用最少的属性值来表达规则所含内容。规则有确定性和不确定性之分, 前者是确定成立的, 后者是以一定概率成立的。

2 约简算法

2.1 连续属性离散化

使用粗糙集约简方法对信息库进行处理, 首先

必须将连续属性进行离散化。离散化常用的方法包括: 经验分割法、等频分割法、等距分割法、Naive Scaler算法和Semi-Naive-Scaler算法、自组织特征映射(SOM)网络方法等。对于不同的数据类型, 上述各种方法的效果不同, 没有一个是绝对最好的。但是, 离散化的结果对于分析结果有很大的影响。本文采用可行性较高的Semi-Naive-Scaler算法对案例表进行离散化处理^[4]。该算法按照属性值的大小关系, 对案例表中的实例进行从大到小排序:

$$X_i = \{x \in U \mid a(x) = c_i\}, \quad c_0 < c_1 < \dots < c_i < \dots < c_n \quad (1)$$

然后按照如下公式处理断点:

$$D_i = \{v \in V_d \mid v = \arg \max \{x \in X_i \mid d(x) = d_i\}\} \quad (2)$$

$$B = \left\{ \frac{c_i + c_{i+1}}{2} \mid D_i \not\subset D_{i+1}, D_{i+1} \not\subset D_i \right\} \quad (3)$$

式中: c_i 和 d_i 分别是决策表中的各实例的条件属性和决策属性的取值, B 为属性断点集合, 集合 D_i 是 X_i 中主要的决策属性值, 即等价类 X_i 中出现频率最高的决策属性值。

Semi-Naive-Scaler 算法的思想为:

1) 根据式(1)和式(2)计算 X_i 和 D_i 的值;

2) 如果 $D_i \not\subset D_{i+1}$ 且 $D_{i+1} \not\subset D_i$, 则根据式(3)计算出此断点值; 否则, 不选取此断点。

2.2 属性约简

在离散化以后, 即可以进行属性约简。属性约简可以采取许多方法, 如基于属性重要性的启发式约简算法、基于可辨识矩阵的约简算法等。但不论采用哪一种方法, 目标都是求取最佳的属性约简, 它须满足一个标准: 在不损失原有信息的前提下, 或者约简后得到最少的属性, 或者最终得到最优的决策规则, 或者总的属性约简最大。通常的属性约简算法可定义为:

1) 删除多余的条件属性, 即消除表中的某一行;

2) 消去重复的行;

3) 消去每一块决策规则的多余属性。

本文采用基于可辨识矩阵的属性约简算法, 此算法能够得到全部的可能约简组合, 也就是说能够获得理论上的最优约简^[5]。令 $Reduct$ 是决策表 T 经过属性约简之后得到的条件属性集合, 可对本属性约简算法描述如下:

第一步: 将决策表转换为可辨识矩阵的形式, 并将矩阵中属性组合数为 1 的属性列入最终的属性约简集合, 即 $Reduct = C_0$ 。

第二步: 从可辨识矩阵中找出不包含核属性组合 S , 即

$$Q = \{B_i : B_i \cap Reduct \neq \emptyset, i = 1, 2, \dots, S\}, \\ S = S - Q.$$

第三步: 将所有不包含核属性的条件属性组合表示为合取范式的形式, 即

$$p = \wedge \{ \vee b_{i,k} : (i = 1, 2, \dots, s; k = 1, 2, \dots, m) \}.$$

第四步: 将 p 转换为析取范式的形式, 并进行化简。

第五步: 所有约简得到的合取式与核属性组成属性约简组合。

利用以上属性约简算法, 可以得到在保证原决策表不损失潜在知识的前提下进行最佳属性约简的充分条件。

2.3 规则约简

属性约简可以简化信息系统的复杂程度, 但对于属性约简后的信息表, 并不是表达每条记录的所有条件属性都是必须的。因此在获得最优属性约简组合的基础上, 还需进行规则约简(或值约简), 信息系统的规则约简过程存在值核的概念, 它是指信息系统的每条记录中不可删除的属性值, 否则将破坏原系统的不可区分关系。规则约简的过程正是逐条对表中记录进行考察, 删除所有不影响规则表达的冗余条件属性值的过程。目前, 在运用粗糙集理论规则约简方面有不少方法, 如通过分析信息系统中决策间的不可区分关系方法、采用属性值重要性度量作为附加信息的启发式算法等。本文采用基于可辨识矩阵的二值化数据过滤约简算法^[6]及贪心算法来进行规则约简。

假设信息系统 $I = \langle U, \Omega, V, f \rangle$, 对应的二值化信息系统为 $I = \langle U, \Omega^B, \{0, 1\}, f^B \rangle$, 式中, Ω^B 为二值化后的属性集合; f^B 为信息函数, 其值域为 $\{0, 1\}$ 。

设属性 $a \in \Omega$ 有 n 个属性值, 那么用一个 n 维的二值向量来表示这个属性值。对于这个向量的第 i 位定义如下:

$$V_{ai}(x) = \begin{cases} 1, & V_a(x) = i \\ 0, & \text{其他} \end{cases} \quad (4)$$

式中: $V_{ai}(x)$ 为实例 x 的二值化属性值的第 i 位; $V_a(x)$ 为非二值化属性值。

通过这种二值化方法属性值对应到某个扩展的二值属性上, 从而将属性值化简问题变成一个属性化简问题, 可以用统一的算法进行处理。以表 1

为例来描述二值化过滤方法的步骤。表 1 中 P 和 L 为条件属性； W 为决策属性。经过二值化后的决策表见表 2，结合贪心算法来进一步化简决策表 2，贪心算法的具体步骤如下：

第一步：根据原来的信息表 2 构造新的信息表。

第二步：初始化最佳属性值集。

第三步：选取信息表所有列中 1 的个数最多的条件属性值加入到最佳属性值集中，去掉此条件属性值所在的列和在此条件属性为 1 的所有行；当有一个以上的条件属性值的列 1 的个数相同时，把列对应的条件属性值所在的列值为 1 的行的 1 的数目相加。

第四步：如果新的信息表中的元素不为空，则转至第三步，否则停止。

按照上述方法化简决策表 2，得到属性值化简后的决策表，见表 3。表中的符号*表示可约简的属性值。

表 1 初始决策表

Tab.1 Original decision table

	P	L	W
1	1	2	2
2, 8	3	2	2
3, 6	1	1	1
4	2	1	1
5, 7	2	2	2

表 2 二值化处理后的决策表

Tab.2 Decision table after binarization

	P			L		W
	1	2	3	1	2	
1	1	0	0	0	1	2
2, 8	0	0	1	0	1	3
3, 6	1	0	0	1	0	1
4	0	1	0	1	0	1
5, 7	0	1	0	0	1	2

表 3 化简后的决策表

Tab.3 Decision table after simplification

	P	L	W
1, 5, 7	*	2	2
2, 8	3	2	3
3, 4, 6	*	1	1

事实上，经过约简之后的信息系统就可以作为所求的规则知识。由表 4 可以得到下面三条规则：

- 1) If $P=3, L=2$ then $W=3$;
- 2) If $L=2$ then $W=2$;
- 3) If $L=1$ then $W=1$.

3 应用算例

输电线路风险评估主要是对各级线路运行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预测运行状态下线路存在的风险指标。本实例采用上述属性约简和规则约简的方法，来说明如何得到风险策略规则^[6]。

表 4 是一个输出输电线路风险评估决策表， D 为风险类型，共存在三种类型风险 D_1 （电压过高）、 D_2 （温度过高）、 D_3 （电流过高）；选取参数 C_1 、 C_2 、 C_3 （ C_1 、 C_2 、 C_3 为输电线路风险因素离散化后的具体含义）； P 为条件属性， U 为影响风险的类别。

表 4 输电线路风险评估决策表

Tab.4 Decision table of risk assessment for transmission lines

U	C_1	C_2	C_3	P	D
1	1	0	1	0	D_1
2	1	1	1	0	D_1
3	1	2	0	0	D_2
4	2	2	2	0	D_2
5	2	0	2	0	D_3
6	2	1	2	2	D_3

根据前面提出的基于可辨识矩阵的属性约简算法，对上面的决策表进行约简，仅含有一个核属性 C_2 。除去所有包含 C_2 的属性组合，得到一组不包含核属性的属性组合 C_1 、 C_3 ，其合取范式为 $C_1 \wedge C_3$ 。将合取范式与核属性组成属性组合，得到一个属性约简结果见表 5。

表 5 属性约简结果

Tab.5 Results of attributes reducing

	C_1	C_2	C_3	D
1	1	0	1	D_1
2	1	1	1	D_1
3	1	2	0	D_2
4	2	2	2	D_2
5	2	0	2	D_3
6	2	1	2	D_3

利用可辨识矩阵的二值化数据过滤约简算法对上面得到的属性约简进行规则约简，得到转换表结果如表 6。

表 6 二值化数据过滤转换表

Tab.6 Conversion table of binarization data filtration

U	C_1			C_2			C_3			D
	0	1	2	0	1	2	0	1	2	
1	0	1	0	1	0	0	0	1	0	D_1
2	0	1	0	0	1	0	0	1	0	D_1
3	0	1	0	0	0	1	1	0	0	D_2
4	0	0	1	0	0	1	0	0	1	D_2
5	0	0	1	1	0	0	0	0	1	D_3
6	0	0	1	0	1	0	0	0	1	D_3

将转换表用可辨识矩阵方法并结合贪心算法,进行约简后得到最终约简表,见表7。

表7 最终约简表

Tab.7 Final reducing table

U	C ₁	C ₂	C ₃	D
1, 2	1	*	1	D ₁
3	1	2	0	D ₂
3	2	2	2	D ₂
5, 6	2	*	2	D ₃

经过约简的决策表去掉了很多冗余信息,但仍能反映各潜在因素对风险的影响关系,下面列出本算例中输电线路风险策略规则:

1) $(C_1=1 \wedge C_3=1) \rightarrow$ 风险类型=D₁ (电压过高);

2) $(C_1=1 \wedge C_2=2 \wedge C_3=0) \rightarrow$ 风险类型=D₂ (温度过高);

3) $(C_1=2 \wedge C_2=2 \wedge C_3=2) \rightarrow$ 风险类型=D₂ (温度过高);

4) $(C_1=2 \wedge C_3=2) \rightarrow$ 风险类型=D₁ (电流过高)。

4 结语

粗糙集理论能够分析隐藏在数据中的真实特性,而不需要任何附加信息。本文结合粗糙集理论中知识约简方法,通过属性约简和规则约简,来获取决策规则。并通过对输电线路运行状态下存在的风险特征进行规则提取。结果表明,这种方法能够有效地获取决策规则,为进行风险评估打下了基础。这种方法同样适应在其他涉及决策规则获取或重要属性选择的领域。

参考文献

- [1] Pawalk Z. Rough Sets [M].Theoretical Aspects of Reasoning about Data[J]. Boston,America: Kluwer

Academic Publishers,1991.60-67.

- [2] 陶鹏,张焰,邱生.基于GIS的配电网供电可靠性定量评估系统研究[J].继电器,2006,34(13):67-71.

TAO Peng, ZHANG Yan, QIU Sheng. Research of Quantitative Assessment System for Distribution Network Reliability Based on GIS[J]. Relay,2006,34(13): 67-71.

- [3] 姜涛,韩富春,范卫星.基于粗糙集理论的架空输电线路运行状态评估[J].电气技术,2007,4:58-60.

JIANG Tao, HAN Fu-chun, FAN Wei-xing. Evaluating the Condition of Overhead Transmission Lines Based on Rough Sets Theory[J]. Electrical Engineering,2007,4: 58-60.

- [4] Knowledge System Group of Dept. of Computer and Technology Rosetta Technical Reference Manual[M]. Norwegian University of Science and Technology,1999.

- [5] 张文修,吴伟志,梁吉业,等.粗糙集理论与方法[M].北京:科学出版社,2001.

ZHANG Wen-xiu, WU Wei-zhi, LIANG Ji-ye, et al. Rough Sets Theory and Method[M].Beijing:Science Press,2001.

- [6] 孙秋野,黎明.粗糙集理论及其电力行业应用[M].北京:机械工业出版社,2009.

SUN Qiu-ye, LI Ming. Rough Sets Theory and Application of Electric Power Industry[M]. Beijing: Machinery Industry Press,2009.

收稿日期:2009-06-20

作者简介:

徐守坤(1972-),男,博士,副教授,主要研究方向为智能信息处理,数据库;E-mail: shoukxun@163.com

韩波(1983-),男,硕士研究生,主要研究方向为数据库与智能信息处理。

(上接第189页 continued from page 189)

高频信号衰耗小,两侧均收到28 dB左右信号,在两侧均发信时会产生功率倒灌现象,造成功放损伤。现在两侧收发信机的发信滤波板中投入4 dB衰耗,使得收信电平为19 dB。

4 结论

经反复试验发现当南昌变手动交换信号时,在正常时昌东变发30 dB(南昌变收28 dB),此时前放为2.93 dB,功放为0.98 dB,而在不正常时昌东变发18 dB(南昌变收15 dB),此时前放为2.93 dB,功放为-11 dB。收信电平过高在两侧均发信时会产

生功率倒灌现象,造成功放损伤,引起功放的稳定性不好是造成此次事故的根本原因。

将收信电平降低后更换收发信机功放插件,重复上述试验,经过近一年的观察,昌东变每次发信均正常。图4、5是收信电平降低前后昌东变手动交换信号时前200 ms发信情况录波图。

收稿日期:2009-02-12; 修回日期:2009-03-05

作者简介:

曾庆汇(1976-),男,本科,工程师,从事继电保护运行工作。E-mail:zqhjbyb@sina.com