

粗糙集约简算法在变压器故障诊断中的应用研究

吕宏丽

(唐山学院 信息工程系, 河北 唐山 063000)

摘要:将粗糙集理论和神经网络技术应用于变压器故障诊断中,粗糙集约简作为神经网络的前置单元,采用基于属性重要性的约简算法。详细阐述了基于属性重要性的约简算法和实现方法,经实际数据训练和测试结果表明,该算法减少了输入样本数,提高了训练速度效率和故障诊断准确率,验证了该算法应用于变压器故障诊断系统的可行性和有效性。

关键词:故障诊断;神经网络;粗糙集;属性约简;属性重要性

中图分类号:TP299 **文献标志码:**A **文章编号:**1672-349X(2015)03-0029-03

DOI:10.16160/j.cnki.tsxyxb.2015.03.010

Application of Rough Set Reduction Algorithm in Fault Diagnosis of Power Transformer

LV Hong-li

(Department of Information Engineering, Tangshan College, Tangshan 063000, China)

Abstract: In this paper, the rough set theory and neural network technology are applied to transformer fault diagnosis, through reduction algorithm with rough set reduction as the pre-unit of neural network. Reduction algorithm and its implementation are discussed in detail. The actual training data and test results show that rough set reduction algorithm reduces the number of input samples, improves the training speed and efficiency and accuracy of fault diagnosis. In conclusion, the application of rough set reduction algorithm to fault diagnosis system of power transformers is feasible and effective.

Key Words: fault diagnosis; neural network; rough set; attribute reduction; attribute significance

0 引言

电力变压器是电力系统中输电、变电、配电设备中最重要和最昂贵的设备之一,其运行状态的安全与否直接关系到整个电力系统的安全性和经济性。变压器故障诊断技术一直是学术界研究的热点,变压器故障的因果关系复杂,其中蕴涵着较大的冗余性。采用传统的神经网络诊断方法,由于不能有效地简化知识,压缩样本空间,影响了诊断的实时性与效率。由于目前智能故障诊断技术发展趋势是将多种诊断技术融合起来,从而实现优势互补,因此本设计将粗糙集理论引入变压器故障诊断中,用于简化数据,提高诊断速度。

粗糙集理论是由波兰华沙理工大学 Z. Pawlak 教授于 1982 年提出来的,是一种处理不确定信息的数学工具,具有对

数据进行压缩和约简的能力,同时不需要任何先验信息,因此更具有客观性^[1],被广泛应用于人工智能、模式识别和故障诊断等领域。粗糙集理论的主要思想是通过属性约简和值约简,导出问题的决策或分类规则,其核心内容是属性约简。

属性约简是在保持信息系统的分类能力不变的前提下,去除其中的冗余属性。经过属性约简能形成精简的规则库以便作出快速、准确的决策。目前属性约简算法主要有:启发式约简算法^[2-4]、基于差别矩阵的属性约简算法^[5]和基于属性重要性的属性约简算法^[6]。这些属性约简算法各有优缺点,而本设计选用基于属性重要性的约简算法。

1 基于属性重要性的属性约简算法

属性的重要性是建立在属性的分类能力上的,属性的重

收稿日期:2014-06-25

基金项目:2014 年度河北省高等学校科学技术研究项目(Z2014037)

作者简介:吕宏丽(1974-),女,甘肃宁县人,副教授,硕士,主要从事电力电子技术与现代传动研究。

要程度是当删除该属性时,考察对信息系统的分类产生的影响。如果删除前后分类没有变化,说明属性重要程度低;如果产生相应的变化,说明该属性重要。

1.1 信息系统的定义

粗糙集理论研究的信息系统通常用一个数据表来表示^[7]。 $S = (U, A, V, f)$ 是一个信息系统,其中 U 是非空的对象集,即 $U = \{X_1, X_2, \dots, X_n\}$, U 中的每个 x_i 称为一个对象; $A = C \cup D$ 是表示属性的非空有限集合, $C = \{c_1, c_2, \dots, c_m\}$ 称为条件属性集合, $D = \{d_1, d_2, \dots, d_k\}$ 表示决策属性集合; $V = \bigcup v_a, \forall a \in A, v_a$ 表示属性的值域; $f = U \times A \rightarrow V$ 是一个信息函数,它对一个对象的每一个属性赋予一个信息值,即 $x \in U, a \in A$ 有 $f(x, a) \in v_a$ 。

1.2 属性重要性的定义

设 $S = (U, A, V, f)$ 是一个信息系统, $X \subseteq A, \forall x \in (A - X)$ 的重要性定义为:

$$Sig_x(x) = \frac{|X| - |X \cup \{x\}|}{|X|} = 1 - \frac{|X \cup \{x\}|}{|X|} \quad (1)$$

其中 $|X| = |IND(X)|$, $IND(X)$ 是由 X 构造的不可分辨关系。具有不可分辨关系的对象是属性值完全相同的对象,所以只要通过判断离散决策表中各样本的不可分辨关系,就能够找到重复的实例。

$$\frac{|X| - |X \cup \{x\}|}{|X|} \text{ 代表不可分辨率。}$$

信息系统 $I = (U, A)$, X 是 A 中的属性子集,属性 $x \in A$, 当 x 加入到集合 X 使得其分辨率越大,就说明属性 x 对集合 X 的重要性越大。

定理^[8]: 设 $S = (U, A, V, f)$ 是一个信息系统, $P \subseteq A$, 若 $|U/P| = |U/A|$, 且 $\forall x \in P$, 有 $Sig_{P-x}(x) > 0$, 则 P 为 A 的一个约简。

令 $X \subseteq A$, 如果 $X - CORE(X) \neq \phi$, $CORE(X) = RED(X)$ 的充分必要条件是 $Sig_{CORE(X)}(x) = 0, x \in X - CORE(X)$ 。

其中 $CORE(X)$ 为信息系统的“核”。“核”可以作为所有约简的计算基础,因为核包含在所有的约简之中。 $RED(X)$ 表示信息系统的“约简”。

基于属性重要性的属性约简算法分成两个阶段:第一阶段先计算出决策表中单个属性的重要性,当某一属性的重要性大于零时此属性为核属性,并入到核属性集 $CORE$ 中;第二阶段以核属性为基础,计算条件属性集中除核以外其他属性的重要性,取重要性最大的属性加入到核集中形成新的集合 RED ,再以 RED 为基础依次循环下去直至剩下所有属性的重要性都为零,最后就得到了属性约简结果。图 1 为算法流程图。

2 变压器故障诊断系统

油中溶解气体分析方法(Disoslvde Gases Analyssi, 简称 DGA)作为一种有效的油浸式电力变压器异常监测手段,在

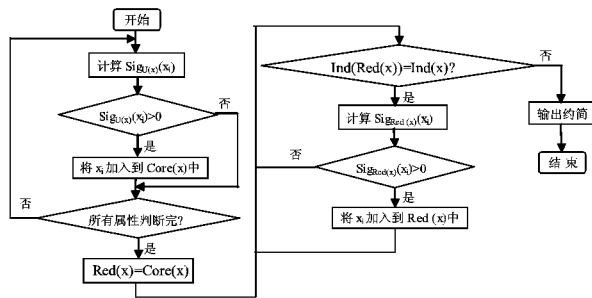


图 1 属性约简算法流程图

电力系统得到广泛的应用。其诊断原理是运行中的变压器会产生一定数量的氢气(H_2)、甲烷(CH_4)、乙烷(C_2H_6)、乙烯(C_2H_4)、乙炔(C_2H_2) 等气体,正常运行和故障运行时气体的组分、含量及产气速率不同从而判断是否发生故障。典型故障分为低能放电、高能放电、中低温过热、高温过热^[9]。

以 DGA 为基础,建立粗糙集神经网络(Rough Set Neural Network, 简称 RSNN)故障诊断系统。神经网络可以确定变压器油中溶解气体组分含量与变压器故障之间的非线性映射关系;粗糙集理论在变压器故障知识不完备的情况下,采用基于属性重要性的属性约简算法对样本数据进行约简,去除冗余属性和冗余样本,作为神经网络系统的前置单元。图 2 为 RSNN 系统框图。

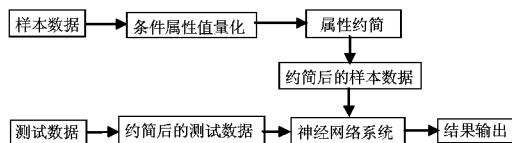


图 2 RSNN 系统框图

3 属性约简算法的应用

结合 DGA 方法,选取氢气(H_2)、甲烷(CH_4)、乙烷(C_2H_6)、乙烯(C_2H_4)、乙炔(C_2H_2) 为故障诊断的五种特征气体,以其组分含量占总含量的百分比作为神经网络的输入向量,输出层表示变压器工作状态:正常、低能放电、高能放电、中低温过热、高温过热,分别表示为 1,2,3,4,5。选取 15 组数据作为训练样本,离散化处理得到原始决策表(见表 1),8 个样本输入数据作为条件属性分别用 $\{C_1, \dots, C_8\}$ 表示,样本输出数据作为决策属性用 D 表示。

条件属性集 C 对决策表的划分是:

$$U/C = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}\}.$$

在条件属性集 C 中依次去除属性 $C_1, \dots, C_8$, 新的划分为:

$$U/C - \{C_1\} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}, x_{15}\}, \{x_{14}\}\},$$

$$U/C - \{C_5\} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}\},$$

$$U/C - \{C_2\} = U/C - \{C_3\} = U/C - \{C_4\} = U/C - \{C_6\} \\ = U/C - \{C_7\} = U/C - \{C_8\} = U/C.$$

表1 训练样本原始决策表

U	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	D
	$\frac{\text{CH}_4}{\text{H}_2}$	$\frac{\text{C}_2\text{H}_2}{\text{C}_2\text{H}_4}$	$\frac{\text{C}_2\text{H}_4}{\text{C}_2\text{H}_6}$	$\frac{\text{C}_2\text{H}_6}{\text{CH}_4}$	总烃	$\frac{\text{C}_2\text{H}_4}{\text{总烃}}$	$\frac{\text{C}_2\text{H}_6}{\text{总烃}}$	$\frac{\text{C}_2\text{H}_2}{\text{总烃}}$	
x ₁	1	2	0	2	1	0	2	2	1
x ₂	2	1	1	2	0	2	1	2	1
x ₃	0	1	0	2	0	0	2	1	1
x ₄	0	0	0	2	1	0	2	0	2
x ₅	0	0	0	1	2	0	2	0	2
x ₆	0	0	0	2	1	0	2	0	2
x ₇	1	2	2	0	0	1	0	2	3
x ₈	1	2	2	0	1	1	0	2	3
x ₉	0	2	2	1	0	1	0	2	3
x ₁₀	2	0	0	2	1	1	2	0	4
x ₁₁	1	1	1	1	2	1	1	1	4
x ₁₂	1	0	1	1	2	1	1	0	4
x ₁₃	1	1	2	0	1	2	0	1	5
x ₁₄	2	0	1	2	0	2	1	0	5
x ₁₅	2	1	2	0	1	2	0	1	5

根据公式(1) 推演出去除各属性后的重要性公式为

$$\text{Sig}_{C-(x_i)}(x_i) = 1 - \frac{|U/C|}{|U/C - \{x_i\}|}. \quad (2)$$

计算出各属性的重要性如表2所示。

表2 各属性的重要性

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
重要性	0.11	0	0	0	0.11	0	0	0

根据属性重要性的计算结果,得到 $\text{CORE}(C) = \{C_1, C_5\}$ 。

令 $\text{RED}(C) = \text{CORE}(C)$, 约简对U的划分为

$$U/\text{RED} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}, \{x_7\}, \\ \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}, x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}\},$$

$$U/\text{RED} + \{C_2\} = U/\text{RED} + \{C_8\} = \{\{x_1\}, \{x_2\}, \{x_3\}, \\ \{x_4, x_6\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \\ \{x_{14}\}, \{x_{15}\}\},$$

$$U/\text{RED} + \{C_3\} = U/\text{RED} + \{C_4\} = U/\text{RED} + \{C_6\} = \\ U/\text{RED} + \{C_7\} = U/\text{RED}.$$

根据公式(1) 计算基于核属性的属性重要性,见表3。

表3 基于核属性的各属性重要性

	C ₂	C ₃	C ₄	C ₆	C ₇	C ₈
重要性	0.11	0	0	0	0	0.11

根据以上计算,将属性 $\{C_2, C_8\}$ 加到核属性集中,得到属性约简集合 $\text{RED}(C) = \{C_1, C_2, C_5, C_8\}$,即RSNN系统的输入样本减少为4个,分别为 $\frac{\text{CH}_4}{\text{H}_2}, \frac{\text{C}_2\text{H}_2}{\text{C}_2\text{H}_4}, \frac{\text{CH}_4}{\text{总烃}}, \frac{\text{C}_2\text{H}_4}{\text{总烃}}$ 。

将15组原始样本数据经过粗糙集约简后,作为三层BP神经网络系统的输入层数据,选取LM算法,训练目标误差为0.001,依次采用不同的隐含层节点数,当隐含层为10时

训练次数和测试正确率最佳,RSNN诊断系统训练曲线如图3所示。与之对比,将15组原始样本数据直接作为三层BP神经网络系统的输入层数据,其它参数与前设置相同,当隐含层为7时训练效果最佳,图4为NN诊断系统训练曲线图。

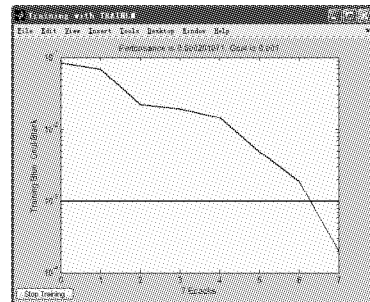


图3 RSNN诊断系统训练曲线图

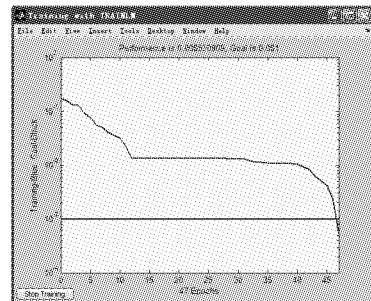


图4 NN诊断系统训练曲线图

选取10组数据作为RSNN和NN诊断系统的测试样本,测试结果如表4所示。

表4 两种方法的测试结果比较表

测试方法	训练步数 / 步	诊断正确率(%)	网络结构
NN系统	47	80	8-7-5
RSNN系统	7	90	4-10-5

4 结论

通过将粗糙集理论与BP神经网络相结合形成RSNN智能方法进行变压器故障诊断,把粗糙集约简作为神经网络的前置单元。测试结果表明,基于属性重要性的约简算法在变压器故障诊断系统中实现了样本化简,较之常规的神经网络诊断系统的变压器故障诊断训练效率提高,故障诊断准确率提高。

参考文献:

- [1] Pawlak Z. Rough set[J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356.
- [2] 蔡莉,胡学刚. 一种基于依赖度的决策表属性约简算法[J]. 安庆师范学院学报:自然科学版, 2008, 14(1): 37-39.
- [3] 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766. (下转第86页)

25 g, 粒径 100~160 目, 吸附时间 60 min, 石灰量 0.25 g。此条件下, 高温改性粉煤灰对废水 COD 和氨氮的去除率分别达到了 85.2% 和 89.6%; 酸改性粉煤灰对废水 COD 和氨氮的去除率分别达到了 78.3% 和 82.7%。

2.3 改性粉煤灰深度处理焦化废水的原理

原状粉煤灰主要由晶体矿物石英和莫来石等组成, 粉煤灰在高温和助熔剂的作用下, SiO_2 和 Al_2O_3 的结合键大大减弱, 导致晶相结构的破坏, 玻璃体熔融生成新的活性物质, 称之为类沸石^[15], 沸石非常大的比表面积大大增强了粉煤灰的吸附能力, 尤其在弱酸性环境下对氨氮的化学吸附效果好, 因此高温改性粉煤灰对焦化废水的氨氮去除率较高。而酸改性粉煤灰是利用酸的腐蚀作用打通粉煤灰的通道, 增大了粉煤灰的孔洞, 增加了粉煤灰的吸附作用。

3 结论

(1) 通过高温改性和酸改性粉煤灰深度处理焦化废水中的 COD 和氨氮的实验对比, 表明高温改性粉煤灰去除效果要好于酸改性粉煤灰。

(2) 改性粉煤灰深度处理焦化废水的最佳工艺条件: 焦化废水 150 mL, pH 值 5 左右, 改性粉煤灰投加量 25 g, 粒径 100~160 目, 吸附时间 60 min, 石灰量 0.25 g, 高温改性粉煤灰对焦化废水 COD 和氨氮的去除率分别达到了 85.2% 和 89.6%; 酸改性粉煤灰对焦化废水 COD 和氨氮的去除率分别达到了 78.3% 和 82.7%。

参考文献:

- [1] 江吉吉, 宁平, 普红平, 等. 改性沸石去除水中低浓度氨氮的研究[J]. 安全与环境学报, 2004, 4(2): 40-43.
- [2] 单明军, 吕艳丽, 丛蕾. 焦化废水处理技术[M]. 北京: 化学工业出版社, 2007.
- [3] 皮科武, 罗永强, 龚文琪. 混凝沉淀工艺处理酚、氰废水工程实例[J]. 湖北工业大学学报, 2009, 24(1): 17-19.

- [4] 杨国义, 王婷. A/O 工艺处理焦化废水的实践研究[J]. 2010, 10(2): 185-188.
- [5] 郭建英, 杨斌, 鲁红志, 等. 内循环 A/O 工艺处理焦化废水运行分析[J]. 节能与环保, 2012, 38(11): 114-117.
- [6] Zhao Yuxiang, Qi Rong, Yin Yanjun. Treatment of coke plant wastewater by A/O fixed biofilm system [J]. Science in China Ser. B Chemistry, 2005, 48(5): 489-496.
- [7] 张能一, 唐秀华, 邹平, 等. 我国焦化废水的水质特点及其处理方法[J]. 净水技术, 2005, 24(2): 42-47.
- [8] 王纪军. A²/O 工艺处理焦化废水工程实例[J]. 技术应用, 2006, 38(3): 14-16.
- [9] 李亚新, 李林永. A²/O 工艺处理焦化废水[J]. 科技情报开发与经济, 2004, 14(1): 128-131.
- [10] Li Bing, Sun Yinglan, Li Yuying. Pretreatment of coking wastewater using anaerobic sequencing batch reactor (ASBR) [J]. Journal of Zhejiang University Science, 2005, 6B(11): 1115-1123.
- [11] 剑平, 赵娜. 混凝气浮在焦化废水生化尾水深度处理中的应用[J]. 工业水处理, 2010(7): 84-86.
- [12] 周静, 李素, 苍大强, 等. 粉煤灰深度处理焦化废水中氨氮的研究[J]. 能源环境保护, 2007, 21(6): 30-32.
- [13] 金文杰, 左宇, 于群. 混凝吸附法深度处理焦化废水[J]. 环境工程, 2009, 27(2): 3-5.
- [14] 国家环境保护总局水和废水监测分析方法编委会. 水和废水监测分析方法[M]. 4 版. 北京: 中国环境科学出版社, 2002.
- [15] 滕宗焕, 陈建中. 改性粉煤灰的吸附机理及其在废水处理中的应用[J]. 西南给排水, 2007, 29(4): 24-26.

(责任编辑: 李秀荣)

(上接第 31 页)

- [4] 卢佳华. 基于属性频率函数的粗糙集属性约简算法[J]. 武汉大学学报: 理学版, 2006, 52(3): 331-334.
- [5] 蒋瑜, 王燮, 叶振. 基于差别矩阵的 Rough 集属性约简算法[J]. 系统仿真学报, 2008, 20(14): 3717-3720.
- [6] 廖启明. 基于粗糙集理论的属性约简与求核算法研究[D]. 长沙: 长沙理工大学, 2012.
- [7] 周献中, 黄兵, 李华雄, 等. 不完备信息系统知识获取的

粗糙集理论与方法[M]. 南京: 南京大学出版社, 2010: 120-121.

- [8] 丁军, 高学东. 一种信息系统的快速属性约简算法[J]. 计算机工程与应用, 2007, 3(14): 173-176.
- [9] 李俭. 大型电力变压器以油中溶解气体为特征量的内部故障诊断模型研究[D]. 重庆: 重庆大学, 2001.

(责任编辑: 李高峰)