

DOI :10.3880/j. issn. 1006-7647. 2009. 03. 010

基于变精度粗糙集的水电机组故障诊断

王荣荣 ,王利平 ,侯新刚

(西京学院电子教研室 ,陕西 西安 710123)

摘要 :将变精度粗糙集理论引入水电机组故障诊断中 ,利用变精度粗糙集属性约简方法对水电机组故障的检测信息进行约简 ,提取对故障分类起主要作用的信息 ,并用 RBF 神经网络对粗糙集处理后的故障信息进行诊断。该方法不仅克服了神经网络对冗余信息和有用信息识别的局限性 ,有效地降低了神经网络的输入信息空间维数 ,减小了神经网络规模 ,还可以弥补经典粗糙集方法对输入信息中的噪声较敏感、抗干扰能力差的不足 ,进而达到提高诊断准确性的目的。水电机组振动故障实例的诊断分析结果证明了该诊断方法的有效性和优越性。

关键词 :变精度粗糙集 ;RBF 神经网络 ;水电机组 ;故障诊断

中图分类号 :TM312 文献标识码 :A 文章编号 :1006-7647(2009)03-0037-04

Research on the fault diagnosis of hydroelectric units based on variable precision rough set //WANG Rong-rong , WANG Li-ping , HOU Xin-gang Teaching and Research Section of Electron Xijing University , Xi 'an 710123 , China)

Abstract :The variable precision rough set theory is introduced into the fault diagnosis of hydroelectric units. The inspection information of hydroelectric units ' faults is reduced by the reduction method of the variable precision rough set , and the information that plays a major role in fault classification can be obtained. The information treated by the rough set is diagnosed with a radial basis function (RBF) neural network. The method not only bypasses the limitations of neural networks , which may fail to determine the redundancy or usefulness of information , it also decreases the spatial dimension of the network input information effectively , reducing the scale of the neural network. The method makes up for the defects of the classical rough set , such as the sensitivity to noise of input information and the weak anti-jamming capability , and accordingly improves the accuracy of fault diagnosis. The result of a fault example proves that the proposed method is valid and has significant advantages.

Key words : variable precision rough set ; radial basis function (RBF) neural network ; hydroelectric units ; fault diagnosis

随着故障诊断技术的不断发展 ,应用于水电机组故障诊断的技术和方法也在不断改进与完善。目前存在的各种诊断方法分别从不同的途径解决水电机组故障诊断问题 ,神经网络诊断技术是其中最常用的方法。如文献 [1] 用一种改进的 BP 神经网络训练方法对水电机组的故障进行分析诊断 ,并使用 BP 子网络方法对水轮发电机组故障分类构造诊断网络 ;文献 [2] 将神经网络与专家系统相结合 ,进行水电机组故障诊断研究 ;文献 [3] 将神经网络和信息融合技术相结合 ,通过故障特征信息的有效组合 ,用各种种子神经网络从不同侧面对水电机组故障进行初步诊断 ,然后对诊断结果应用 Dempster-Shafer 证据理论进行决策融合。由于神经网络不能确定哪些知识是冗余的哪些知识是有用的 ,而且当样本数据量大时网络规模非常庞大 ,网络的归纳、联想能力将大

幅度下降 ,限制了神经网络在水电机组故障诊断中的应用。

笔者将变精度粗糙集 (variable precision rough set ,简称 VPRS) 理论引入水电机组故障诊断中 ,利用它在处理不确定信息方面的优势对水电机组故障信息进行去余处理 ,并将其与径向基函数 (radial basis function ,简称 RBF) 神经网络^[4-6]结合 ,充分发挥了两者各自的优点 ,克服了神经网络对冗余信息和有用信息识别的局限性和经典粗糙集方法对输入信息中的噪声敏感、抗干扰能力差的不足 ,为神经网络在水电机组故障诊断中得到更好的应用提供了一个新思路。

1 粗糙集理论基础

粗糙集理论^[7]是一种处理不确定和不完全信息的强有力的工具。1982 年波兰学者 Pawlak 首先提

作者简介 :王荣荣 (1971 —) ,女 ,陕西渭南人 ,讲师 ,硕士 ,从事水电机组状态监测与故障诊断、变压器设计与故障诊断研究。E-mail : wangrongrong2003@163. com

出的经典粗糙集模型所处理的分类必须是完全正确或肯定的,它是严格按照等价类来分类的,这种分类是精确的,没有某种程度上的“包含”或“属于”,当信息中存在噪声时,会影响诊断结果的准确性,常表现为结果不稳定。针对 Pawlak 经典粗糙集模型的缺点,1993 年 Ziarko 提出了变精度粗糙集模型(VPRS 模型),它在经典粗糙集模型的基础上引进了阈值参数 β ,也就是允许一定程度的分类误差率存在,放松了对经典粗糙集近似边界的严格定义,将决策属性对条件属性的完全依赖转变为一定程度上的依赖,增强了经典粗糙集模型的抗干扰能力和对新数据的预测能力,为研究处理由于噪声引起的数据不一致性问题提供了很好的方法。

1.1 VPRS 模型

在 VPRS 模型中,阈值参数 β 有 2 种不同的定义和取值范围,一种是 Ziarko 所定义的分类误差率,其取值范围为 $0 < \beta \leq 0.5$;另一种是 An 等^[8]所定义的分类正确率,其取值范围为 $0.5 < \beta \leq 1$ 。笔者主要研究 An 等所定义的 VPRS 模型。

定义 1 设 X 和 Y 表示有限论域 U 的 2 个非空子集,令

$$P(Y/X) = \begin{cases} \frac{|X \cap Y|}{|X|} & |X| > 0 \\ 1 & |X| = 0 \end{cases} \quad (1)$$

式中: $|X|$ 为集合 X 的基数; $|X \cap Y|$ 为集合 $X \cap Y$ 的基数; $P(Y/X)$ 为集合 X 关于集合 Y 的相对分类正确率或包含度,即集合 Y 基于集合 X 的条件概率。

定义 2 在信息系统 $S = (U, A, V, f)$ 中, U 为对象集; $A = C \cup D$, 其中 C 为条件属性集, D 为决策属性集, $C \cap D = \emptyset$; $V = \bigcup_{a \in A} V_a$, V_a 是属性 a 的值域; $f: U \times A \rightarrow V$ 是一个信息函数,它为每个对象 x 的每个属性 a 赋予一个信息值,即 $x \in U, \forall a \in A, f(x, a) \in V_a$ 。

给定阈值参数 $0.5 < \beta \leq 1, X \subseteq U, R \subseteq A[x]_R$ 为 X 的 R 等价类, $x \in X$ 则 X 的 β 下近似集、上近似集分别为

$$R_{*\beta}(X) = \bigcup \{ [x]_R : P(X|[x]_R) \geq \beta \} \quad (2)$$

$$R_{\beta}^*(X) = \bigcup \{ [x]_R : P(X|[x]_R) \geq 1 - \beta \} \quad (3)$$

式中: $R_{*\beta}(X)$ 也称为 β 的正域,记作 $\text{pos}_{\beta}^R(X)$,可理解为将 U 中的对象以不小于 β 的分类正确率分于 X 的最大集合; $R_{\beta}^*(X)$ 可理解为将 U 中的对象以不小于 $1 - \beta$ 的分类正确率分于 X 的最大集合。随着 β 的增大, β 粗糙集的近似边界区域变窄,即 VPRS 意义下的不确定区域变小,当 $\beta = 1$ 时, VPRS 即为经典粗糙集。

1.2 VPRS 的属性约简

VPRS 属性约简是从条件属性中删去重要性小的属性,保留重要的属性。VPRS 中决策属性集 D 与条件属性集 C 的 β 依赖度定义如下:

$$K = \chi(C, D, \beta) = \frac{|\text{pos}(C, D, \beta)|}{|U|} \quad (4)$$

式中: $\text{pos}(C, D, \beta)$ 是决策属性集 D 相对条件属性集 C 在 β 下的正域,则 C 关于 D 的 β 相对约简是 C 的一个子集 $\text{red}(C, D, \beta)$,其应满足下列条件:

$$\chi(C, D, \beta) = \chi(\text{red}(C, D, \beta), D, \beta) \quad (5)$$

且从 $\text{red}(C, D, \beta)$ 中除去任何一个属性都将使式(4)不成立。

2 VPRS 在故障诊断中的应用

采用将 VPRS 与 RBF 神经网络相结合的方法^[9],即将 VPRS 作为 RBF 神经网络的前端处理系统,用 VPRS 理论先对信息进行约简预处理,去除冗余信息,保留起主要作用的信息,然后把预处理后的信息输入 RBF 神经网络进行诊断。VPRS-RBF 神经网络故障诊断原理见图 1。

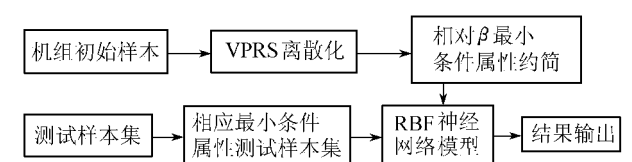


图 1 VPRS-RBF 神经网络故障诊断原理

2.1 VPRS 的数据处理

VPRS 应用于故障诊断数据处理的关键技术包括:①初始信息表的建立:水电机组结构庞大,故障原因复杂,实际监测数据量很大。可以把某次故障监测作为对象,对监测所得的机组信息数据经过相关分析、变换、处理后,把获得的机组故障状态的特征量及一些定性征兆作为信息表的条件属性,对应的故障类别作为信息表的决策属性形成故障初始样本决策表。②数据离散化^[10]粗糙集理论只对量化的属性值进行分析处理,而水电机组振动故障诊断中的各种特征参数多为连续量。因此,在进行约简处理之前,需要对诊断数据中的连续属性值进行离散处理。③属性约简:设定阈值参数 β ($0.5 < \beta \leq 1$) 利用式(4)和式(5),在保持 β 的前提下,删除与决策属性无关或相关性小的条件属性,得到约简条件属性。一般情况下,经典粗糙集的属性约简算法不适于 VPRS。笔者采用基于遗传算法寻找 VPRS 属性的属性约简方法^[11-12]。

2.2 RBF 神经网络故障诊断

训练 RBF 神经网络模型:以粗糙集约简后所得

的相对 β 最小条件属性作为 RBF 神经网络的输入神经元 ;以故障分类作为 RBF 神经网络的输出神经元 构造相应的 RBF 神经网络 ,并对其进行学习和训练 根据试验确定出 RBF 神经网络的结构参数。

检验诊断结果 :将测试样本输入神经网络进行检验 ,反复调整参数 ,直到输出诊断结果满意为止。

3 应用实例

3.1 初始样本表的形成

水电机组故障大多以振动的形式表现出来 ,因此对某水电厂机组监测系统的实测振动数据进行了分析^[13] ,得到各特征频段的频谱特征向量 ,经过归一化处理后作为粗糙集的条件属性 ,对应的故障类别作为决策属性形成诊断决策表。

选取振动信号频谱中的 6 个频率特征量 :0 ~ 0.2(f 为转子基频) , $1/6f \sim 1/2f$, $1f$, $2f$,大于或等于 $3f$,50 Hz 或 100 Hz 频率及 4 个定性征兆属性 :振动与转速的关系、振动与负荷的关系、振动与励磁电流的关系、振动与流量的关系作为条件属性 ,分别用 $c_1 \sim c_{10}$ 表示。选择转子不平衡、转子弓形弯曲、轴承间隙过大、尾水管偏心涡带、卡门涡列、转子不对中、转子匝间短路、磁极不均匀 8 类故障作为决策属性 ,在决策表中分别用 1 ,2 ,... ,8 表示。选择这 8 个故障样本 分别用 $u_1 \sim u_8$ 表示 构成由 10 种条件属性和 8 种决策属性组成的水电机组振动故障决策表(表 1)。

3.2 VPRS 属性约简

利用前面所述数据处理方式对样本数据进行离

散处理和属性约简 ,取 $\beta = 0.85$ 。获得相对 β 最小约简为 $\{c_2 ,c_4 ,c_7 ,c_8 ,c_9 \}$,见表 2。可见经过 VPRS 分析处理后 ,在保证 β 的前提下 ,除去了冗余的条件属性 ,保留了重要条件属性 ,减少了原来决策表中条件属性的个数。

3.3 训练 RBF 神经网络

对原始数据和 VPRS 处理后的数据分别构造 RBF 神经网络 ,用学习样本对它们进行训练 ,然后用测试样本进行检验 ,比较诊断结果。

对原始数据构造神经网络时 ,学习样本见表 1 ,其输入层节点为 10 个 ,隐含层节点取 8 个 ,输出目标为 8 个 ,构成 10-8-8 的网络结构。对约简后的属性构造神经网络时 ,学习样本见表 2 ,其输入层节点为 5 个 ,隐含层节点取 6 个 ,输出目标为 8 个 ,构成 5-6-8 的网络结构。故障诊断测试样本见表 3 ,神经网络诊断阈值取 0.5 ,小于 0.5 该故障不存在 ,大于 0.5 该故障存在 测试样本诊断结果见表 4。

从表 4 可以看出 ,2 种方法都诊断出了 3 种故障 (转子不平衡、尾水管偏心涡带、磁极不均匀) ,输出值都大于阈值 0.5。RBF 神经网络方法对样本 2、样本 3 的诊断结果比较理想 ,而对样本 1 的故障诊断结果不理想 ,输出值仅高出阈值 0.086 2。VPRS-RBF 神经网络方法对 3 个测试样本的诊断结果都比较理想 ,样本 1 的故障诊断结果高出阈值 0.299 1 ,比 RBF 神经网络诊断结果好。2 种方法相比 ,VPRS-RBF 神经网络结构更为简单 ,诊断准确率更高。

表 1 故障决策表

故障 样本	条件属性										决策 属性
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	
u_1	0.01	0.12	0.93	0.02	0.20	0.07	0.97	0.17	0.01	0.06	1
u_2	0.03	0.01	1.00	0.80	0.10	0.01	1.00	0.02	0.01	0.02	2
u_3	0.01	0.02	0.04	0.03	0.01	0.03	0.01	0.96	0.06	0.01	3
u_4	0.05	0.95	0.07	0.02	0.01	0.02	0.10	0.95	0.08	0.10	4
u_5	0.93	0.04	0.10	0.11	0.06	0.02	0.02	0.07	0.09	0.95	5
u_6	0.01	0.03	0.68	0.96	0.82	0.06	0.97	0.95	0.01	0.45	6
u_7	0.03	0.06	0.97	0.13	0.04	0.02	0.05	0.11	0.80	0.03	7
u_8	0.04	0.02	0.98	0.06	0.02	0.01	0.98	0.03	0.98	0.07	8

表 2 约简决策表

故障 样本	输入样本条件属性					故障 样本	输入样本条件属性				
	c_2	c_4	c_7	c_8	c_9		c_2	c_4	c_7	c_8	c_9
u_1	0.12	0.02	0.97	0.17	0.01	u_5	0.04	0.11	0.02	0.07	0.09
u_2	0.01	0.80	1.00	0.02	0.01	u_6	0.03	0.96	0.97	0.95	0.01
u_3	0.02	0.03	0.01	0.96	0.06	u_7	0.06	0.13	0.05	0.11	0.80
u_4	0.95	0.02	0.10	0.95	0.08	u_8	0.02	0.06	0.98	0.03	0.98

表 3 故障诊断测试样本

故障测 试样本	条件属性									
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
样本 1 (转子不平衡)	0.01	0.02	0.98	0.05	0.05	0.07	0.91	0.10	0.24	0.01
样本 2 (尾水管偏心涡带)	0.03	0.97	0.07	0.02	0.01	0.02	0.10	0.95	0.08	0.16
样本 3 (磁极不均匀)	0.05	0.02	1.00	0.06	0.02	0.02	0.97	0.03	0.98	0.13

表 4 测试样本诊断结果

故障类型	RBF 神经网络方法诊断结果			VPRS-RBF 神经网络方法诊断结果		
	样本 1	样本 2	样本 3	样本 1	样本 2	样本 3
转子不平衡	0.5862	0.1304	0.1072	0.7991	0.1001	0.0988
转子弓形弯曲	0.1889	0.0917	0.0975	0.0257	0.0989	0.1189
轴承间隙过大	0.1011	0.0870	0.0998	0.1123	0.0997	0.1052
尾水管偏心涡带	0.0995	0.8955	0.1000	0.0961	0.9000	0.1001
卡门涡列	0.1002	0.1002	0.1001	0.0339	0.1017	0.0719
磁极不均匀	0.2871	0.0963	0.8819	0.2618	0.1001	0.8984
转子匝间短路	0.1427	0.0991	0.1035	0.1616	0.0994	0.1101
转子不对中	0.0944	0.0998	0.1000	0.1096	0.1002	0.0967

4 结 语

笔者研究了 VPRS 在处理水电机组故障诊断数据中的应用,并将其与神经网络方法相结合进行故障诊断。在 VPRS 预处理数据时通过设定阈值参数 β ,可以克服经典粗糙集模型对数据噪声过于敏感的缺点。通过对水电机组振动故障实例进行诊断分析,表明了该诊断方法可以有效地简化神经网络的规模和结构,克服了传统神经网络方法处理高维数据时网络结构复杂的不足,并增强了对含噪声数据分析和处理的鲁棒性,具有更高的容错性和诊断能力。

参考文献:

[1] 符向前,刘光临,蒋劲. BP 神经网络在水轮发电机组状态监测与诊断系统中的应用[J]. 武汉大学学报 :工学版,2002,35(1) :24-28.

[2] 贾嵘,白亮,罗兴,等. 基于神经网络的水轮发电机组振动故障诊断专家系统[J]. 水力发电学报,2004,23(6) :120-123.

[3] 杨晓萍,南海鹏,张江滨. 信息融合技术在水轮发电机组故障诊断中的应用[J]. 水力发电学报,2004,23(6) :111-115.

(上接第 8 页)

b. 入汇口上游主流断面在 $z^* = 0.2$ 处水流脉动强度沿 y 方向分布很均匀,峰值很小;而入汇口及其下游主流断面在 $z^* = 0.2$ 处各方向的水流脉动强度沿 y 方向变化很大,靠近入汇口一侧(垂线 Line0.45 ~ Line0.55 附近) x 和 y 方向的水流脉动强度值很大,远离入汇口处垂线的水流脉动强度又变小并趋于稳定。

参考文献:

[1] 兰波,汪勇.干支流交汇水面形态特征分析[J]. 重庆交通学院学报,1997,16(4) :109-114.

[2] 徐孝平,彭文启.直角交汇河段流场特性分析[J]. 水利学报,1993,24(2) :22-31.

[3] 吴迪,郭维东,刘卓也.复式断面河道“Y”型交汇河口水流水力特性[J]. 水利水电科技进展,2007,27(3) :21-23.

[4] 张光碧,邓军,张法星,等. VOF 模型在有支流汇入的河

[4] 朱明星,张德龙. RBF 网络基函数中心选取算法的研究[J]. 安徽大学学报,2000,24(1) :72-78.

[5] 葛强,徐俩俩,仇宝云,等. 基于 RBF 人工神经网络的电动机振动故障诊断[J]. 河海大学学报 :自然科学版,2008,36(6) :842-845.

[6] 丁晓群,蔡志慧. 人工神经网络在电力系统中的应用与研究现状[J]. 水利水电科技进展,1996,16(4) :12-16.

[7] PAWLAK Z. Rough sets theory and its applications to data analysis[J]. Cybernetics and Systems,1998,29,661-668.

[8] AN A,SHAN N,CHAN C,et al. Discovering rules for water demand prediction :an enhanced rough-set approach [J]. Engineering Application and Artificial Intelligence,1996,9(6) :645-653.

[9] 王荣荣. 基于变精度粗糙集与神经网络的水电机组振动故障诊断研究[D]. 西安 :西安理工大学,2007.

[10] 代建华,李元香. 粗糙集理论中基于遗传算法的离散化方法[J]. 计算机工程与应用,2003(8) :13-14.

[11] 陈纯毅,崔广才. 基于变精度粗糙集模型的属性约简方法研究[J]. 长春理工大学学报,2005,28(3) :52-54.

[12] 陶志,许宝栋,汪定伟,等. 基于遗传算法的粗糙集知识约简方法[J]. 系统工程,2003,21(4) :116-122.

[13] 梁武科,罗兴,张彦宁,等. 水力发电机组振动故障诊断系统中的信号预处理[J]. 水力发电学报,2003,22(3) :114-120.

(收稿日期 2008-10-13 编辑 骆超)

道复杂流场中的应用[J]. 河海大学学报 :自然科学版,2007,35(5) :592-595.

[5] WEBER L J,SCHMIDT E D,MAWER N. Experiments on flow at a 90° open-channel junction[J]. Journal of Hydraulic Engineering,ASCE,2001,127(5) :340-350.

[6] 茅泽育,罗昂,罗日升,等. 矩形断面压力管道汇流口局部能量损失[J]. 水利水电科技进展,2006,26(3) :62-66.

[7] 肖洋,唐洪武,毛野,等. 新型声学多普勒流速仪及其应用[J]. 河海大学学报 :自然科学版,2002,30(3) :15-18.

[8] 王协康,杨青远,王宪业,等. 卵砾石床面时均流速分布的试验研究[J]. 四川大学学报 :工程科学版,2007,39(2) :14-19.

[9] 刘同官,郭炜,王协康,等. 入汇角为 30°时交汇区水流结构水槽试验研究[J]. 长江科学院院报,2007,24(4) :75-78.

[10] 刘艾明,徐海涛,卢金友. 矩形水槽水流紊动特性分析[J]. 长江科学院院报,2006,23(1) :12-15.

[11] 吕升奇,唐洪武,闫静. 有无植物条件下明渠水流紊动特性对比[J]. 水利水电科技进展,2007,27(6) :64-68.

(收稿日期 2008-07-15 编辑 骆超)