



基于邻域粗糙集与 RVM 的制粉系统故障诊断

张文涛^{1,2}, 钟文晶², 胡伯勇², 陆豪强², 马永光¹, 董子健¹

(1. 华北电力大学控制与计算机工程学院, 河北 保定 071003;

2. 浙江浙能技术研究院有限公司, 浙江 杭州 310000)

摘要: 针对火电厂制粉系统的故障征兆参数复杂、不易诊断的特点, 提出一种基于邻域粗糙集(NRS)与相关向量机(RVM)的故障诊断方法。该方法首先利用邻域粗糙集约简输入的特征向量, 并将约简得到的最优决策表作为 RVM 的输入, 采用组合核函数代替传统的单一核函数, 利用网格搜索和交叉验证的方法确定最佳的核函数参数和组合核系数, 建立二叉树 RVM 多分类模型, 从而进行制粉系统故障识别和诊断。实验结果表明, 该方法故障诊断准确率可达 95%, 且泛化能力强。

关键词: 制粉系统; 邻域粗糙集; 相关向量机; 故障诊断

中图分类号: TP27

文献标志码: A

文章编号: 1674-5124(2019)08-0151-05

Fault diagnosis based on neighborhood rough set and RVM for pulverizing system

ZHANG Wentao^{1,2}, ZHONG Wenjing², HU Boyong², LU Haoqiang², MA Yongguang¹, DONG Zijian¹

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China;

2. Zhejiang Energy Group Research Institute, Hangzhou 310000, China)

Abstract: On account of the complex and difficult diagnosis characteristic of the fault sign parameters of the power plant, a fault diagnosis method based on the neighborhood rough set (NRS) and the relevance vector machine (RVM) is proposed. Firstly, the feature vectors of input are reduced by neighborhood rough set, and the optimal decision table is used as the input of RVM. Then the combination kernel function is used instead of the traditional single kernel function. The grid search and cross validation are used to determine the optimal parameters of the kernel function and the combination kernel coefficients. Finally, the two fork tree RVM multi-classification model is established, therefore it can be applied for the fault recognition and diagnosis of pulverizing system. The experimental results show that the accuracy of fault diagnosis can reach 95%, and the generalization ability is strong.

Keywords: pulverizing system; neighborhood rough set (NRS); relevance vector machine (RVM); fault diagnosis

收稿日期: 2018-05-08; 收到修改稿日期: 2018-06-20

基金项目: 中央高校基本科研业务费专项资金(9160316004)

作者简介: 张文涛(1992-), 男, 山东泰安市人, 硕士研究生, 专业方向为发电设备的故障诊断。

0 引言

制粉系统是电厂锅炉的一个非常重要的子系统,其工作状况对于整个火电机组的安全和经济运行具有非常重要的意义,一旦制粉系统发生故障,其后果不可估计^[1]。20世纪80年代之前,一般都是依靠人工分析和信号处理来对故障进行诊断和分析,这种方法效率低且准确率不高。随着技术的不断发展,许多新的方法开始运用到故障诊断之中。专家系统(ES)依据专家知识,通过模拟人的思维对故障进行推理和判断,但是这种方法对知识的获取比较困难,且自动学习能力差^[2]。支持向量机(SVM)是一种得到广泛应用的机器学习算法,文献[3]利用核主元分析(KPCA)提取特征向量,将提取后的主元作为SVM的输入,对故障进行诊断和分类。随着神经网络的不断发展,许多专家学者利用神经网络进行制粉系统的故障诊断和故障识别。文献[4]使用RBF神经网络,但训练样本过大时,训练速度会很慢。文献[5]提出基于粗糙集(RS)和概率神经网络的故障诊断方法,解决了样本维数和训练速度的问题,但是粗糙集理论^[6-7]会对数值型数据进行离散化处理,造成信息损失。

相关向量机(relevance vector machine, RVM)是在SVM的基础上提出的一种基于贝叶斯理论的稀疏学习算法^[8],与SVM相比,具有以下优点:核函数的选取不受Mercer定理的限制;相关向量的数目远小于支持向量的数目;RVM无需手动设置惩罚因子。RVM已在模式识别中得到了广泛应用^[9]。邻域粗糙集理论^[10]是对传统粗糙集理论的扩展,该方法可以直接处理数值型数据,避免在离散化中数据损失的问题。因此本文提出一种基于邻域粗糙集和RVM的制粉系统故障诊断方法,实现对多种故障的识别和诊断。

1 邻域粗糙集理论

经典粗糙集理论在处理连续型数据时需要将数据进行离散化处理,但是离散化处理后会改变数据原始的属性性质。邻域粗糙集是为了解决经典粗糙集不便于处理数值属性的数据集而提出来的。

对于一个决策系统 $DS=(U, C \cup D, V, f)$,其中, U 表示论域, C 为条件属性, D 为决策属性, V 为各属性值的集合, f 代表信息函数,表示样本、属性和属性值之间的映射关系。

属性约简就是将邻域决策系统中不必要、冗余

的属性删除,但又不影响系统本身的分类能力。通常情况下,属性的约简并不是唯一的,可以有多个约简集合,所有约简集合的交集被称为属性的核,这是不能被删除的^[11]。在本文中,邻域粗糙集约简使用前向贪心算法,其基本思想是以空集为起点,计算所有属性的重要度,保留属性重要度最大的为约简后的集合,基于邻域粗糙集的属性约简算法如下:

- 1) 输入邻域决策系统 $NDS=<U, A, D>$,设定重要度下限为 $Sig_{min}(Sig_{min} \geq 0)$,其中 U 为给定样本集合, D 为决策属性, A 表示生成邻域的邻域关系。
- 2) 初始化约简集合 red 为空集。
- 3) 对每一个属性 a_i 计算重要度 $Sig(a, B, D) = \gamma_B(D) - \gamma_{B-\{a\}}(D)$,式中 B 为条件属性。
- 4) 选择最大的重要度,此时 $Sig(a_k, B, D) = \max(Sig(a_i, B, D))$ 。
- 5) 如果 $Sig(a_k, B, D) > Sig_{min}$, $red \leftarrow red + a_k$,返回到步骤3),否则就输出 red 。

2 相关向量机

2.1 RVM分类模型

RVM是一种可用于回归或分类的基于贝叶斯的稀疏学习算法,下面针对二分类的RVM方法进行介绍。给定训练样本集 $\{x_n, t_n\}_{n=1}^N$,其中 x 是输入矩阵, $t_n \in \{0, 1\}$ 为分类的类别号, N 是训练样本数,RVM的分类函数可以定义:

$$y(x, w) = \sum_{i=1}^N w_i K(x, x_i) + w_0 \quad (1)$$

式中: w ——权值向量;

$K(x, x_i)$ ——核函数。

设 $P(t|x)$ 为伯努利分布,则样本数据的似然估计概率分布为

$$P(t|w) = \prod_{n=1}^N \sigma\{y(x_n; w)\}^{t_n} [1 - \sigma\{y(x_n; w)\}]^{1-t_n} \quad (2)$$

式中, $\sigma(\cdot)$ 为sigmoid函数。

为了保持RVM模型的稀疏性,假设 w 服从均值为0的高斯先验概率分布:

$$P(w|\alpha) = \prod_{i=0}^N N(w_i|0, \alpha_i^{-1}) \quad (3)$$

其中, α 为 $N+1$ 维的超参数向量, $N(w_i|0, \alpha_i^{-1})$ 为高斯分布函数。

对于新的输入向量 x^* ,根据概率预测公式,目标

向量 t^* 的条件概率为

$$p(t^*|t) = \int p(t^*|w, \alpha) p(w|t, \alpha) p(t|\alpha) dw d\alpha \quad (4)$$

由于不能通过解析的方法求解权重系数的后验概率 $p(w|t, \alpha)$ 和边缘似然估计函数 $p(t|\alpha)$, 因此可以采用一种基于拉普拉斯估计的积分变换法来估算后验权重概率分布, 具体可以参考文献 [12], 本文在此不再赘述。

经过多次迭代, 大部分的权值都为 0, 剩下的少部分非零权值向量被称为相关向量, 只有相关向量对样本数据的分类起作用, 则 RVM 的分类函数可以表示为:

$$y(x^*; w) = \sigma(\sum_{x_i \in RVs} w_i K(x^*; x_i) + w_0) \quad (5)$$

2.2 组合核函数的构建

RVM 的核函数可以有很多种, 比如 Gauss 核函数、Poly 核函数、Laplace 核函数等, 不同的核函数有不同的特点, 在训练时产生的相关向量的个数也不同, 预测精度也不一样^[13]。本文采用一种局部 Gauss 核函数和全局 Poly 核函数组合核函数的形式, 表达式为:

$$K(x_i, x_j) = sK_{Gauss}(x_i, x_j) + (1 - s)K_{Poly}(x_i, x_j) \quad (6)$$

其中, s 表示组合核函数权重系数, $0 \leq s \leq 1$ 。Gauss 核函数的表达式为 $K_{Gauss}(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \gamma^2)$, γ 为核函数的宽度; Poly 核函数的表达式为 $K_{Poly}(x_i, x_j) = (x_i^T x_j + 1)^2$, 本文中多项式级数为 2。

2.3 多分类 RVM

单一的 RVM 只能解决二分类问题, 对于多分类问题, 通常通过构建多个二分类 RVM 实现, 常用的方式有一对一、一对多、有向无环图和二叉树等^[14]。由于二叉树 RVM 具有较高的训练和测试速度, 且需要较少的分类器, 故本文采用一种基于二叉树 RVM 的多分类策略。

二叉树 RVM 多分类算法的思想是: 先将所有样本数据的类别分成两大子类, 然后将子类进一步划分成两个次子类, 如此划分下去, 直到将所有的类划分结束。对于 N 分类问题, 只需要构造 $N-1$ 个分类器, 因此计算效率高。基于二叉树 RVM 多分类的制粉系统故障诊断模型如图 1 所示。

3 基于邻域粗糙集和 RVM 的制粉系统故障诊断方法

火电厂运行环境复杂, 而制粉系统故障类别

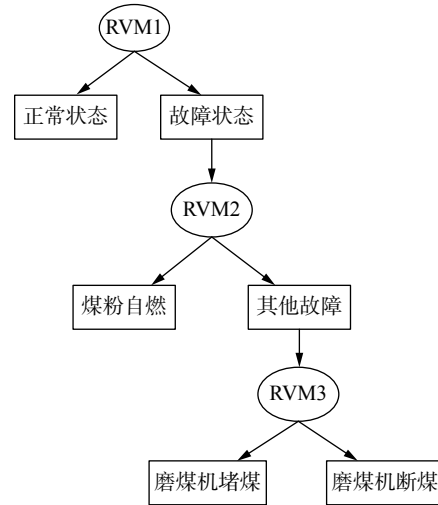


图 1 故障多分类模型

多, 且难以通过故障征兆判断故障原因^[15]。邻域粗糙集在处理不完整数据和不确定知识时, 可以有效挖掘出数据内部的信息, 并将数据通过一定的原则进行约简和提炼。基于此, 本文提出基于邻域粗糙集和 RVM 的制粉系统故障诊断方法, 故障诊断的流程见图 2, 具体步骤如下:

1) 选取训练样本, 构造邻域决策系统。

2) 对样本数据进行归一化处理, 消除量纲的影响, 归一化公式如下:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

3) 基于邻域粗糙集理论, 对训练样本进行属性约简, 得到与决策属性最相关的条件属性。

4) 对于属性约简后的样本数据, 构造二叉树 RVM 多分类器, 并采用网格搜索和 K 折交叉验证的方法, 确定最佳的核函数参数和组合核系数, 从而建立 RVM 故障诊断模型。

5) 对测试样本进行故障诊断和识别。

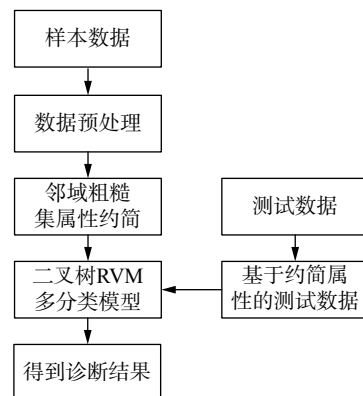


图 2 基于 NRS-RVM 的故障诊断流程图

4 仿真研究

4.1 性能分析

为了验证基于 NRS-RVM 故障诊断模型的有效性,选取 UCI 标准数据集 (<http://archive.ics.uci.edu/ml>) 进行仿真实验。主要包括两个实验,实验一验证经过邻域粗糙集属性约简之后的数据,可以提高分类精度,实验二主要比较 RVM 与 SVM 模型的性能,实验中选用的数据集如表 1 所示。

表 1 实验数据集信息

数据集	属性个数	类别数	训练样本	测试样本	样本总数
Breastw	9	2	444	239	683
Vehicle	18	4	400	446	846
Diabetes	8	2	506	253	759

将经过约简后的数据和未约简的测试数据的进行分类对比,分类算法采用 RVM 算法,对比结果如表 2 所示。

表 2 数据集约简前后分类准确率比较

数据集	原始属性个数	分类准确率/%	约简后属性个数	约简后分类准确率/%
Breastw	9	96.23	4	97.49
Vehicle	18	94.17	8	96.86
Diabetes	8	77.07	3	77.86

从表 2 中可以看出经过邻域粗糙集约简之后的数据,分类准确率都得到了提高。这说明经过邻域粗糙集的属性约简,不仅可以减少计算量,去除冗余信息,而且可以提高分类准确率。

利用前面所提的约简后的数据集,分别采用 SVM 和 RVM 对测试数据进行分类,从训练时间、测试时间、向量个数和分类准确率 4 个方面比较它们的性能,对比结果如表 3 所示。

表 3 SVM 与 RVM 性能比较 (仿真)

数据集	模型	训练时间/s	测试时间/s	向量个数	分类准确率/%
Breastw	SVM	0.875 6	0.203 5	89	97.07
	RVM	6.234 4	0.016 3	17	97.49
Vehicle	SVM	0.706 3	0.199 8	106	96.41
	RVM	4.965 4	0.008 7	18	96.86
Diabetes	SVM	0.900 1	0.206 4	198	77.86
	RVM	6.998 0	0.020 4	22	77.86

从表 3 中可以很直观地看出,在分类准确率方面,RVM 和 SVM 都能取得不错的效果,RVM 略高于 SVM。RVM 的训练时间虽然要比 SVM 长,但是 RVM 有着比 SVM 更短的测试时间,并且相关向量的个数远远少于 SVM 的支持向量的个数,说明了 RVM 模型的优越性,可以应用在故障诊断方面。

4.2 实例分析

制粉系统的主要故障有磨煤机堵煤、磨煤机断煤、煤粉自燃、一次风管堵塞等。本文选取决策属性 $D = \{1, 2, 3, 4\}$, 分别对应正常状态、磨煤机堵煤、磨煤机断煤和煤粉自燃等 4 种状态。选择跟制粉系统相关的 11 个条件属性 $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$, 分别为磨煤机电流、磨煤机齿轮箱推力瓦温度、磨煤机润滑油温度、磨煤机入口温度、磨煤机一次风流量、磨煤机入口压力、磨煤机进出口压差、磨煤机出口温度、磨密封风与一次风差压、分离器出口压力、给煤量。一共选取 160 组数据,每类数据 40 组,其中取每类数据的前 30 组作为训练集,后面 10 组数据为测试集样本。

首先将数据进行归一化处理,然后采用前面所述的邻域粗糙集属性约简算法,对数据进行约简,得到的最优条件属性表为 $\{c_1, c_2, c_4, c_5, c_8, c_{11}\}$, 由原来 11 个条件属性,减少为 6 个,大大减少了样本集的数据量,这样既能保证原始数据信息不会损失,又能为后面的 RVM 建模提供方便。

本文采用网格搜索和 5 折交叉验证的方法寻找最优的核函数参数和组合核系数,根据最优参数,建立二叉树 RVM 多分类故障诊断模型。测试集样本的故障诊断结果如图 3 所示。

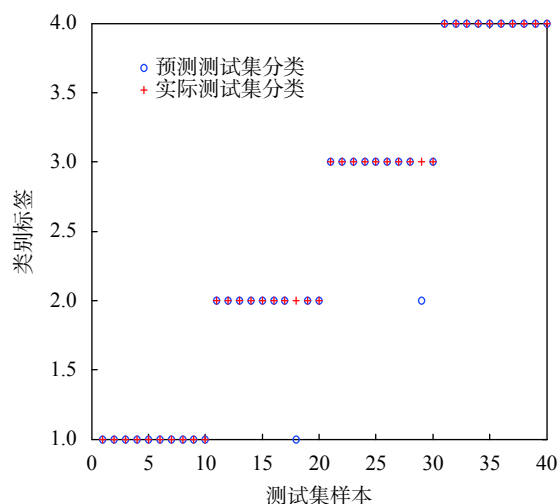


图 3 测试集样本的故障诊断结果

从图中可以看出, 一共 40 个测试样本, 每类有 10 个, 有 2 个测试样本被诊断错误, 故障诊断准确率可以达到 95%。

为了验证本方法的有效性和优越性, 再使用 SVM 算法进行制粉系统的故障诊断, 文中借助 LIBSVM 软件包, 选择 RBF 核函数, 并采用网格搜索和交叉验证的方法确定 SVM 的惩罚因子 C 和核函数参数 σ 。两种方法在测试样本数据上的训练时间、测试时间、向量个数和诊断准确率的比较如表 4 所示。

表 4 SVM 与 RVM 性能比较 (制粉系统)

模型	训练时间/s	测试时间/s	向量个数	准确率/%
SVM	0.659	0.235 4	67	92.5
RVM	3.254	0.017 8	11	95

从表 4 中可以很直观地看出, 在制粉系统故障诊断准确率方面, RVM 和 SVM 都能取得不错的效果, RVM 略高于 SVM。RVM 的训练时间要比 SVM 长, 这是由于 RVM 在计算过程中要进行大量的迭代, 因此 RVM 模型的训练时间比较长, 但在实际的工程应用中, 故障诊断模型都是提前训练好的, 并不会影响故障诊断的效率。RVM 有着比 SVM 更短的测试时间, 说明 RVM 的分类速度快, 泛化能力强。更重要的是, RVM 的相关向量的个数只有 11 个, 远远少于 SVM 的支持向量的个数, 这说明 RVM 故障诊断模型的稀疏性好, 在大样本数据时也能取得良好的效果。

5 结束语

针对火电厂制粉系统故障类别多, 难以通过故障征兆判断故障原因的问题, 本文提出一种基于邻域粗糙集和 RVM 的故障诊断方法, 不仅解决了经典粗糙集在离散化时的数据损失问题, 而且减少了测试时间, 大大提高了故障诊断准确率, 对于实际

工业过程中制粉系统的故障诊断具有一定的启示作用。

参考文献

- [1] 钱虹, 茅大钧, 叶建华, 等. 火电厂磨煤机出口温度异常分析和故障诊断系统 [J]. *自动化仪表*, 2009(7): 22-24.
- [2] 王禹新, 牛玉广. 基于专家系统的制粉系统故障诊断的研究 [J]. *仪器仪表与分析测*, 2010(3): 1-3.
- [3] 刘定平, 叶向荣, 陈斌源, 等. 基于核主元分析和最小二乘支持向量机的中速磨煤机故障诊断 [J]. *动力工程*, 2009(2): 155-158.
- [4] 陈斌源, 朱军. 基于径向基函数神经网络的中速磨煤机故障诊断 [J]. *发电设备*, 2011, 25(5): 323-326.
- [5] 费树岷, 李延红, 柴琳. 基于 RSPNN 的制粉系统故障诊断 [J]. *控制工程*, 2012, 19(3): 412-415.
- [6] SHEN L, TAY F, QU L, et al. Fault diagnosis using rough sets theory [J]. *Compute Ind*, 2000, 43(1): 61-72.
- [7] 王国胤, 姚一豫, 于洪. 粗糙集理论与应用研究综述 [J]. *计算机报*, 2009, 32(7): 1229-1245.
- [8] TIPPING M E. Sparse bayesian learning and the relevance vector machine [J]. *Journal of Machine Learning Research*, 2001, 1(3): 211-244.
- [9] 张亚男, 杨慧中. 基于 RVM 组合核优化的软测量模型研究 [J]. *系统仿真学报*, 2018(1): 272-277.
- [10] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简 [J]. *软件学报*, 2008, 19(3): 640-649.
- [11] 江峰, 王莎莎, 杜军威, 等. 基于近似决策熵的属性约简 [J]. *控制与决策*, 2015, 30(1): 65-70.
- [12] 卢锦玲, 绳菲菲, 赵洪山. 基于相关向量机的风机齿轮箱故障诊断方法 [J]. *华北电力大学学报 (自然科学版)*, 2017, 44(2): 91-96.
- [13] PSORAKIS I, DAMOULAS T, GIROLAMI M A. Multiclass relevance vector machines: an evaluation of sparsity and accuracy [J]. *IEEE Transactions on Neural Networks*, 2010, 21(10): 1588-1598.
- [14] 薛宁静. 多类支持向量机分类器对比研究 [J]. *计算机工程与设计*, 2011, 32(5): 1792-1795.
- [15] 樊帅, 肖军. 锅炉制粉系统故障诊断方法 [J]. *热力发电*, 2015, 44(2): 13-17, 23.

(编辑: 商丹丹)