

基于粒子群算法优化支持向量机汽车故障诊断研究

余梓唐

(义乌工商职业技术学院 机电信息分院, 浙江 义乌 322000)

摘 要: 汽车故障检测和诊断技术一直是国内外研究热点问题。支持向量机用于汽车故障诊断时,其多分类组合决策对分类正确率及诊断时间有很大影响,为了有效提高汽车系统故障诊断的效率和精度,提出了一种基于粒子群算法优化层次支持向量机汽车故障诊断检测方法。针对分解支持向量机具有测试时间短、结构难以确定的特点,利用粒子群算法,依据最大间隔距离原则优化层次支持向量机模型,使每个节点的支持向量机具有最大分类间隔,减少了误差积累,从而优化了多级二叉树结构的 SVM,实现故障的分级诊断。仿真实验结果表明,提出的算法在所有参比模型中精度最高,能高效地对汽车系统的故障进行检测与定位,具有较强的泛化能力,同时缩短了故障诊断时间。

关键词: 粒子群算法; 支持向量机; 汽车故障诊断; 遗传聚类

中图分类号: TP301

文献标志码: A

文章编号: 1001-3695(2012)02-0572-03

doi:10.3969/j.issn.1001-3695.2012.02.045

Automotive fault diagnosis based on SVM and particle swarm algorithm

YU Zi-tang

(School of Electro-Mechanical & Information Technology, Yiwu Industrial & Commercial College, Yiwu Zhejiang 322000, China)

Abstract: Automobile fault detection and diagnosis technology has been a research hotspot. Support vector machine used in automobile fault diagnosis, the classification decision on the rate of correct classification and diagnosis time have great influence. In order to effectively improve the automobile fault diagnosis efficiency and accuracy, this paper proposed a method based on particle swarm optimization algorithm for hierarchical support vector machine fault diagnosis detection method. According to the decomposition support vector machine has short test time, is difficult to confirm the structure characteristics, this paper used the particle swarm algorithm, based on the maximum distance principle optimization of hierarchical support vector machine model, so that each node of the support vector machine had the maximal margin classification, reduced the error accumulation, thus optimized the multilevel binary tree structure of SVM, to realize fault hierarchical diagnosis. The simulation results show that the proposed algorithm, all the reference model of the highest accuracy, can be efficient for automobile system fault detection and location. This algorithm has strong generalization ability, and can shorten the time of fault diagnosis.

Key words: particle swarm algorithm; support vector machine; fault diagnosis; genetic clustering

0 引言

机械故障诊断是一种了解和掌握机器在运行过程的状态,确定其整体或局部正常或异常,早期发现故障及其原因,并能预报故障发展趋势的技术。油液监测、振动监测、噪声监测、性能趋势分析和无损探伤等为其主要的诊断技术方式。诊断技术发展几十年来,产生了巨大的经济效益,成为各国研究的热点。利用各种检查和测试方法,发现系统和设备是否存在故障的过程是故障检测;而进一步确定故障所在大致部位的过程是故障定位。故障检测和故障定位同属网络生存性范畴^[1,2]。要求把故障定位到实施修理时可更换的产品层次的过程称为故障隔离。故障诊断就是指故障检测和故障隔离的过程。故障诊断是系统控制领域中的一个关键而富有挑战性的难题^[3,4]。故障诊断和推理的主要任务是分析异常行为状态并分析出这些异常行为之间的因果关系,找出故障的原因。Petri网是一种严格的数学模型,其主要是对离散并行系统的数学表示,1960年由卡尔·A·佩特里发明的,适合于描述异步的、并发的计算机系统模型。它特别适合对具有异步、离散、并发、同

步等行为的系统进行表示和分析,已在故障诊断领域得到广泛应用。张怡哲等人^[5]提出采用蚁群算法和Petri网对柔性系统故障进行诊断,可以实现对柔性制造系统中操作过程进行动态调度,但没有给出诊断的具体推理步骤。王罗等人^[6]提出了神经网络汽车故障检测进行推理,并使用优先级表示优先诊断和处理严重程度高的故障,对每个故障的原因赋予一个权值,但权值的赋予需要专家系统的经验,具有很强的依赖性。

传统的支持向量机用于汽车故障诊断时,其多分类组合决策对分类正确率及诊断时间有很大影响,为了有效提高汽车系统故障诊断的效率和精度,提出了一种基于粒子群算法优化层次支持向量机汽车故障诊断检测方法。针对分解支持向量机具有测试时间短、结构难以确定的特点,本文利用粒子群算法,依据最大间隔距离原则优化层次支持向量机模型,使每个节点的支持向量机具有最大分类间隔,减少了误差积累,从而优化了多级二叉树结构的SVM,实现故障的分级诊断。仿真实验结果表明,本文提出的算法在所有参比模型中精度最高,能高效地对汽车系统的故障进行检测与定位,具有较强的泛化能力,同时缩短了故障诊断时间。

收稿日期: 2011-07-18; 修回日期: 2011-08-30

作者简介: 余梓唐(1966-),男,浙江义乌人,讲师,硕士,主要研究方向为智能计算、模式识别等(yztyw@ywu.cn)。

1 故障诊断基本原理

汽车故障是指汽车中的零部件部分散失或者全部散失汽车原来设计规定的功能的现象。汽车故障若按照影响汽车性能的情况,一般可以分为功能性故障和参数性故障。功能故障主要是指汽车不能继续完成本身的功能,比如行驶跑偏、发动机不能启动等;参数故障是指汽车的性能参数达不到规定的指标,由于这些故障会造成严重的后果,可以分为轻微故障、一般故障、严重故障和致命故障。

汽车故障是错综复杂的,每个故障都可能是由多种原因所引起的,目前传统的汽车故障诊断都是建立在手工检测的基础上,都是依赖于人工的推理和判断等,但这种方法对于现在复杂的汽车来说都显得力不从心,所以必须寻求一种智能方法对汽车进行故障检测与诊断。本文采用的基于模糊依赖度最大化汽车故障诊断算法,就是参数估计诊断方法,其主要思想是当故障由参数的变化描述时,可以通过采用已有的参数估计算法来监测故障的信息,并根据参数的估计值与正常值比较,从其偏差来作出判定。其要求找出模型参数和物理参数之间的一一对应关系,这样可以获得更好的故障检测。

2 改进的支持向量机汽车故障诊断

2.1 支持向量机

支持向量机是一种研究有限样本情况下统计及机器学习规律的通用学习方法,它建立在统计学习理论和结构风险最小化原理基础上,通过保持经验风险值固定并最小化置信范围来实现结构风险最小化,比传统的基于经验风险最小化原理的方法有明显的优点。支持向量机用于回归分析,其算法实现过程为:

对于给定的样本集 $\{x_i, y_i\}_{i=1}^k$, 其中, x_i 为输入值, y_i 为相应的预测值。其回归函数形式为 $f(x) = w \cdot \Phi(x) + b$, 其中, w 为权重矢量, b 为偏差。

通过最小化正则风险函数训练参数 w 和 b :

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \frac{1}{k} \sum_{i=1}^k \varepsilon(f(x_i) - y_i) \\ \text{s. t.} \quad & \varepsilon(f(x_i) - y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon & |w \cdot \Phi(x) + b - y_i| \geq \varepsilon \\ 0 & |w \cdot \Phi(x) + b - y_i| < \varepsilon \end{cases} \end{aligned} \tag{1}$$

其中: $\varepsilon(\cdot)$ 为 ε -intensive 损失函数, ε 为 ε -intensive 损失参数, C 为惩罚参数, 用来平衡经验风险和正则化部分。

寻求最优回归超平面的二次凸规划问题变成:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k (\xi_i + \xi_i^*) \\ \text{s. t.} \quad & \begin{cases} y_i - w \cdot \Phi(x) - b \leq \varepsilon + \xi_i & \xi_i \geq 0 \\ w \cdot \Phi(x) + b - y_i \leq \varepsilon + \xi_i^* & \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{2}$$

其中: ξ 和 ξ^* 为非负松弛变量。

引入拉格朗日乘子 a_i 和 a_i^* , 上式转换为对偶问题为

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i,j=1}^k \sum_{l=1}^k (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) - \\ & \varepsilon \sum_{i=1}^k (a_i + a_i^*) + \sum_{i=1}^k y_i (a_i - a_i^*) \\ \text{s. t.} \quad & \sum_{i=1}^k (a_i - a_i^*) = 0 \quad C \geq a_i, a_i^* \geq 0 \end{aligned} \tag{3}$$

其中: $K(x_i, x_j)$ 为核函数, 常用的核函数为径向基核函数 $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}$, $\sigma > 0$, 它构造的 SVM 具有较强的非线性预测

能力, 因而本文选择径向基核函数构造支持向量机。

得到 SVR 回归函数为

$$f(x) = \sum_{i=1}^k (a_i - a_i^*) K(x_i, x) + b \tag{4}$$

2.2 粒子群分层聚类优化支持向量机

本文将分层聚类算法与二叉树相结合, 在每个节点上通过分层聚类减少误差积累, 利用粒子群算法寻找最优类别划分, 采用类别进行编码, 把同属于一类故障的样本划到同一个聚类中心。粒子群优化算法^[10] (particle swarm ptimization, PSO), 初始化为一群随机粒子, 即随机的初始解, 然后粒子们根据自身的飞行经验和群体的飞行经验来调整自己的飞行轨迹, 参照最优粒子的位置在解空间中搜索, 并通过迭代找到最优解。遗传粒子群算法把遗传算法和粒子群算法有机结合起来, 既保证了遗传算法强大的全局搜索性能, 又同时融合了粒子群的位置转移思想^[11], 其寻优过程更有效率, 所得到的解精度更高。由于分层聚类问题不同于一般的组合优化问题, 在应用遗传粒子群算法求解分层聚类时, 根据聚类特点, 对遗传粒子群算法的模型进行如下改进:

为把同一类的样本划到同一个聚类中, 采用类别进行编码, 用二进制 0、1 辨识不同的聚类。对于待分类样本类别 n 的聚类问题, 令搜索空间的维数 $D = n, m$ 为粒子数量。定义第 i 个粒子位置 x_i 由长度为 n 的二进制编码构成, 每个粒子位置 $x_i = (x_{i1}, x_{i2}, \dots, x_{in}) (x_{ij} \in \{0, 1\}, i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ 对应于问题的一个解, 对第 i 个粒子 $x_{ij} = 0 (j = 1, 2, \dots, n)$ 表示该粒子把 j 类样本划为负类, $x_{ij} = 1$ 表示把 j 类样本划为正类。

在生成决策树的过程中, 应该让类间距最大的类最早分割出来, 即在决策树的上层节点处分割, 采用 SVM 的分类间隔作为两类间的可分离测度设计 GAPSO 适应度函数, 以使每个决策节点类间可分离度尽可能大。根据 SVM 理论, GAPSO 的适应度函数^[9] 设计为

$$\text{fitness} = 2 / \|w^*\| \tag{5}$$

其中:

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \tag{6}$$

式(6)中, α_i^* 为支持向量对应的 Lagrange 乘子。

2.3 遗传操作

a) 交叉。把适应度排在后面的 $m/2$ 个粒子两两配对, 随机在第 m 位 ($1 < m < n$) 进行两两交叉, 如果交叉后子代的适应度较高, 则代替原来的父代。

b) 变异。为算法在后期能跳出局部值, 变异率随迭代次数增加而增加^[12], P_m 的初值为 0.1, 一直增加到 0.5 为止。

$$p_m = p_1 + t(p_2 - p_1) / t_{\max} \tag{7}$$

其中: p_m 为当前变异概率; p_1 为初始变异概率, 取为 0.1; p_2 为最大变异概率, 取为 0.5; t 为当前迭代次数; $t_{\max}$ 为最大迭代次数。

建立粒子速度和位置更新方程。粒子速度更新和位置^[11]更新按下面公式进行:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 (P_{\text{best}} - P_{\text{present}}) + c_2 r_2 (G_{\text{best}} - G_{\text{present}}) \tag{8}$$

$$x_{\text{id}}^{t+1} = \begin{cases} 0 & \text{rand} \geq \text{sig}(v_{\text{id}}^{t+1}) \\ 1 & \text{others} \end{cases} \tag{9}$$
$$\text{sig}(v_{\text{id}}^{t+1}) = 1 / (1 + \exp(v_{\text{id}}^{t+1}))$$

2.4 基于分层聚类算法的 SVM 决策树生成

a) 对根节点根据所有样本用遗传粒子群算法分成两类。

b) 对子节点含样本类别数进行判断,所有样本的类别数都不大于 2,则转到步骤 d),否则转到步骤 c)。

c) 对各节点中类别数大于 2 的节点调用遗传粒子群算法聚类分析,转步骤 b)。

d) 结束循环,生成最优二叉树的朴树结构,根据朴树结构训练生成多分类 SVM。

3 实例应用结果与分析

本文所有的实验都是在 PC P4 T2310 1.86 GHz,2 GB RAM, Intel82865G 显卡的计算机上做的实验,利用人造数据进行仿真模拟来验证本文提出算法的有效实用性,通过样本集证明了算法的有效性。本文以桑塔纳轿车为例,采用本文算法对汽车故障进行仿真分析,图 1 给出了按照算法计算的分类精度编号情况,图 2 是本文方法故障特征识别图。从图 2 中可以明显发现,随着汽车故障特征的增加,故障的识别精度也呈递增趋势,这充分说明经本文提出的粒子群分层聚类优化支持向量机算法可以选择较少的故障特征而达到更高的分类精度,算法对汽车故障特征识别能力非常高。

需训练的二分类支持向量机个数为 45 个,测试时需要遍历所有的子分类器,本文方法只需训练 9 个二分类支持向量机,测试时遍历个数不大于 5 个,对训练和分类复杂度比较如表 1、2 所示。

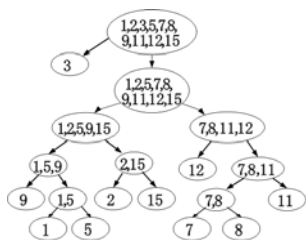


图1 最优决策树

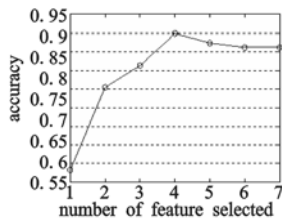


图2 故障特征识别

表 1 训练和分类复杂度比较

分类	传统算法	本文方法
需训练 SVM 的数目	45	9
分类遍历 SVM 平均数目	45	≤5

表 2 诊断结果

分类	传统支持向量机	本文方法
测试时间/s	0.9	0.4
训练正确率/%	96.5	97
测试正确率/%	96	96.5
样本总正确率/%	96.25	96.75

(上接第 571 页)

参考文献:

[1] 于森,王延章.一种基于角色网络模型的电子政务系统框架及其实现研究[J].计算机工程与应用,2003,39(12):31-35.
[2] 叶鑫,王延章.电子政务的层次角色网络模型研究[J].系统工程学报,2006,22(2):216-220.
[3] 叶鑫,王延章.基于角色模型的行政事务处理系统建模方法研究[J].计算机应用研究,2005,22(6):44-46.
[4] 王宁,于森,王延章.面向电子政务领域的角色网络模型[J].系统工程,2007,25(2):7-10.
[5] 王延章,高国伟,于森,等.信息系统中的角色网络模型和方法研究[J].计算机应用研究,2008,25(10):3116-3118.
[6] 许永涛,王延章,陈雪龙.基于角色网络理论的政府办公自动化系统开发平台的研究与实现[J].计算机应用研究,2007,24(2):

如表 1、2 所示,本文方法与传统的支持向量机相比,分类正确率相当,诊断时需要遍历的二值 SVM 个数却大大减少,减少了测试时间,提高了测试效率。

4 结束语

本文提出了一种基于粒子群算法优化层次支持向量机汽车故障诊断检测方法。针对分解支持向量机具有测试时间短、结构难以确定的特点,本文利用粒子群算法,依据最大间隔距离原则优化层次支持向量机模型,使每个节点的支持向量机具有最大分类间隔,减少了误差积累,从而优化了多级二叉树结构的 SVM,实现故障的分级诊断。仿真实验结果表明,本文提出的算法在所有参比模型中精度最高,能高效地对汽车系统的故障进行检测与定位,具有较强的泛化能力,同时缩短了故障诊断时间。

参考文献:

[1] 范贵生,虞慧群,陈丽琼,等.基于 Petri 网的服务组合故障诊断与处理[J].软件学报,2010,21(2):231-247.
[2] 李惠,崔成玲,李新宏.柔性制造系统中 CNC 机床故障诊断机理研究[J].计算机应用研究,2009,26(8):2982-2984.
[3] 徐滔,王祁.一种神经网络预测器在传感器故障诊断中的应用[J].传感技术学报,2005,18(2):235-236.
[4] 王源,胡寿松.基于支持向量机的非线性系统故障诊断[J].控制与决策,2001,16(5):617-620.
[5] 张怡哲,邓建华.舵面损伤在线故障模式预测及故障检测[J].西北工业大学学报,2011,21(3):299-300.
[6] 王罗,桂桂新,陈特效.回归型支持向量机在电机故障诊断中的研究[J].计算机测量与控制,2007,15(12):1691-1694.
[7] 桑海峰,何大阔,张大鹏.基于支持向量机与遗传算法的发酵过程软测量建模[J].东北大学学报,2010,28(6):781-784.
[8] LIN C F, WANG S D. Fuzzy support vector machines[J]. IEEE Trans on Neural Networks,2009,13(2):464-471.
[9] WANG T Y, CHIANG H M. Fuzzy support vector machine for multi-class text categorization[J]. Inform Process Manag,2010,43(4):914-929.
[10] 刘本德,胡昌华,蔡艳宁.基于聚类和 SVM 多分类的容差模拟电路故障诊断[J].系统仿真学报,2009,21(20):6479-6482.

209-211.
[7] JOSHI J B D, AREF W G, GHAFOR A, et al. Security models for Web-based applications[J]. Communications of the ACM,2001,44(2):38-72.
[8] JOSHI J B D, BERTINO E, GHAFOR A. Hybrid temporal role hierarchies in GTRBAC[J]. ACM Trans on Information and System Security,1999,2(1):3-33.
[9] CRAMPTON J, KHAMBHAMMETTU H. Delegation in role-based access control[C]//Proc of the 11th European Symposium on Research in Computer Security. Hamburg: Springer, 2006:174-191.
[10] WAINER J, KUMAR A. A fine-grained, controllable, user-to-user delegation method in RBAC[C]//Proc of the 10th ACM Symposium on Access Control Models and Technologies. New York: ACM Press, 2005:59-66.