

【自动化技术】

基于分层聚类的支持向量机模拟电路故障诊断

罗沛清, 梁青阳, 江钦龙, 孙哲

(空军航空大学, 长春 130022)

摘要:提出分层聚类与支持向量机集成的算法,以多级二叉树结构的SVM实现故障的分级诊断。根据最大间隔距离原则,对各故障模式电路特征的逐次聚类二分获得二叉树,使每个节点的SVM具有最大分类间隔,减少了误差积累,从而优化了SVM的组合策略。聚类上利用遗传粒子群算法对样本进行聚类。经模拟电路仿真结果显示,该方法与一对一、一对多方法相比,在不影响分类精度下,减少测试时间,降低模型复杂度。

关键词:支持向量机;多分类;遗传粒子群算法

中图分类号:TP227

文献标识码:A

文章编号:1006-0707(2011)09-0092-04

模拟电路广泛应用于军工、自动控制、家用电器^[1]等各个方面。模拟电路故障占了电子设备故障的绝大部分。在模拟电路故障诊断经典方法上主要是故障字典法和参数辨识法,通过求解网络中所有元件的实际值来确定故障元件。这些方法要求测试节点多,计算量大,故难以付诸实际应用。随着模式识别技术的发展,小波分析结合神经网络以及支持向量机等技术应用到模拟电路故障诊断上。与传统方法相比,模式识别方法不需要电路的朴树结构,在对非线性电路故障诊断准确率以及实用性上都有一定的发展。但是神经网络在诊断过程中普遍存在网络收敛速度慢,容易收敛于“局部最优值”,且网络结构设计没有理论依据的缺点。基于支持向量机(support vector machine, SVM)的故障诊断方法建立了一套较好的有限样本下机器学习的理论框架和通用方法。该方法有严格的理论基础,较好地解决了小样本、非线性、高维数和局部极小点等问题,一定程度上能够克服神经网络故障诊断方法的缺陷。但是支持向量机用于模拟电路故障诊断时,其多分类组合决策对分类正确率及诊断时间有很大影响。针对二叉树支持向量机具有测试时间短,但是结构难以确定的特点,本文将遗传粒子群算法对故障特征进行聚类,优化支持向量机的结构。

1 支持向量机

支持向量机^[2]是一种建立在统计学习理论上的新型学习机器,通过核函数,将输入空间映射到一个高维特征空间,然后在特征空间中通过最大化分类间隔来构造最优分类超平面。

1.1 对线性可分问题

支持向量机通过以下算法决定最优分类超平面。设已知训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X, Y)^l, \\ x_i \in X = R^n, y_i \in Y = \{1, -1\}, i = 1, \dots, l$$

构造并求解最优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^l \alpha_j \quad (1)$$

$$\text{s. t.} \quad \sum_{i=1}^l y_i \alpha_i = 0 \quad (2)$$

$$\alpha_i \geq 0, i = 1, \dots, l$$

得到最优解

$$\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$$

计算

$$w^* = \sum_{i=1}^l y_i \alpha_i^* x_i \quad (3)$$

$$b^* = \sum_{i=1}^l y_i \alpha_i^* (x_i \cdot x_j) \quad (4)$$

构造分划超平面

$$(w^* \cdot x) + b^* = 0 \quad (5)$$

得到决策函数

$$f(x) = \text{sgn}((w^* \cdot x) + b^*) \quad (6)$$

1.2 对线性不可分问题

对线性不可分问题,将输入空间映射到一个高维特征空间,在高维空间中进行划分,但并不是直接在高位空间划分样本,而是通过内积运算在特征空间中通过最大化分类间隔来构造最优分类超平面。

支持向量机中采用不同的核函数将形成不同的算法,目前研究最多的核函数主要有以下3类:

$$\text{径向基核 } k(x, x') = \exp(-\|x - x'\| / \sigma^2) \quad (7)$$

其中 σ 为核宽度。

$$\text{多项式核 } k(x, x') = \exp(-(x \cdot x') + c)^d \quad (8)$$

其中 $d = 1, 2, \dots, n, c \geq 0$ 。

收稿日期:2011-07-06

作者简介:罗沛清(1984—),男,硕士研究生,主要从事故障诊断技术研究。

$$\text{sigmoid 核 } k(x, x') = \tan h(k \cdot x \cdot x') + v) \quad (9)$$

2 基于分层聚类多分类支持向量机

2.1 支持向量机多分类算法及其特点

标准支持向量机只能解决 2 分类问题。针对多分类算法,一种是将多个二分类支持向量机进行组合,实现多分类;另一种是直接构造出一个能进行多值分类的支持向量机分类器。但是,后者在训练速度和分类精度上也不太理想,所以当前支持向量机用于多分类大多是通过组合的方式构造。对 k 类样本问题,主要有 3 种组合方法:一对多方法(One-against-All SVM, OAASVM)^[3]、一对一方法(One-against-One SVM, OAOSVM)^[4]和决策导向无环图(Decision Directed Acyclic Graph SVM, DDAGSVM)^[5]。文献[6]对上述 3 类方法进行了比较,其中以 OAO 和 DDAG 方法的分类精度较高,但是 2 种方法需构建 SVM 数目多($k(k-1)/2$ 个),随着分类数的增加,所需的子分类器数急剧增加,模型复杂度急剧增长,会导致测试效率降低。OAA 方法需要的两分类支持向量机个数较少,只需 k 个 SVM,但每个子分类器需要处理所有的样本,训练时间较长(每个子分类器训练时间与样本数目有关),而且其在分类过程中会存在大量不可分区域,每个子支持向量机训练时需要处理所有的样本,训练时间比较长,同时,由于正负类样本不平衡,最后分类平面会向样本少的类别靠近,从而影响分类的正确率。文献[7-8]对二叉树的组合 SVM 方法进行多分类,精度上与一对一相近,测试时间得到很好改善,测试精度也得到一定提高,同时所需的分类器数目也更少,测试时不用遍历所有的子分类器,但是由于用的是以类别中心进行聚类方法,容易出现把同一类故障划分到不同的聚类中心。为了避免把同一类的样本划到不同的聚类中心,文献[9]利用类别进行编码聚类,但是编码设计比较复杂。对此,本文提出以遗传粒子群算法对样本进行分层聚类。

2.2 基于遗传粒子群分层聚类改进支持向量机分类

本文将分层聚类算法与二叉树相结合,通过分层聚类减少误差积累,在分层聚类上利用遗传算法寻找类别划分,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 optimization, 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。

遗传算法(genetic algorithm, GA)是借鉴生物界自然选择和群体进化机制形成的一种全局寻优算法^[11-12],按照一定的适配值函数及一系列遗传操作^[7]对各个体进行筛选,从而使适配值高的个体被保留下来,组成新的群体。

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,又同时融合粒

子群的位置转移思想^[13],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,因此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改进。

2.2.1 针对问题编码

为把同一类的样本划到同一个聚类中,采用类别进行编码编码,用二进制 0,1 表示不同的聚类。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。

2.2.2 适应度设计

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 GAPS0 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, GAPS0 的适应度函数^[9]设计为

$$\text{fitness} = 2 / \|w^*\| \quad (10)$$

其中

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad (11)$$

α_i^* 为支持向量对应的 Lagrange 乘子。

2.2.3 遗传操作

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。

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^[14], P_m 的初值为 0.1,一直增加到 0.5 为止。

$$p_m = p_1 + t \cdot (p_2 - p_1) / t_{\max} \quad (12)$$

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。

2.2.4 建立粒子速度和位置更新方程

粒子速度更新和位置^[13]更新按下面公式进行:

$$v_i(t+1) = \omega \cdot v_i(t) + c_1 \cdot r_1 \cdot (P_{\text{Best}} - P_{\text{Present}}) + c_2 \cdot r_2 \cdot (G_{\text{Best}} - G_{\text{Present}}) \quad (13)$$

$$\begin{cases} x_{id}^{t+1} = \begin{cases} 0, \text{rand} \geq \text{sig}(v_{id}^{t+1}) \\ 1, \text{others} \end{cases} \\ \text{sig}(v_{id}^{t+1}) = 1 / (1 + \exp(v_{id}^{t+1})) \end{cases} \quad (14)$$

2.3 基于遗传粒子群聚类流程

步骤 1 判断样本包含的类别数目,根据样本类别数进行种群初始化。

步骤 2 计算种群适应度,并按适应度对种群进行排序,计算全局最优及个体最优值。

步骤 3 对个体进行交叉变异操作,重新计算全局最优

点及个体最优值。

步骤4 更新粒子速度位置。

步骤5 重新判断种群适应度,满足要求则结束;否则转向步骤3。

步骤6 到达最大迭代次数或者连续4代最优解都不变,算法结束,输出最优解。

2.4 基于分层聚类算法的SVM决策树生成

步骤1 对根节点根据所有样本的总类别数用遗传粒子群算法分成2类。

步骤2 对子节点含样本类别数进行判断:如所有样本的类别数都不大于2个,则转到步骤4;否则转到步骤3。

步骤3 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2。

步骤4 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 应用实例

采用文献[8]的带通滤波器电路,用一对方法和本文方法进行比较,验证本文算法。电路见图1。

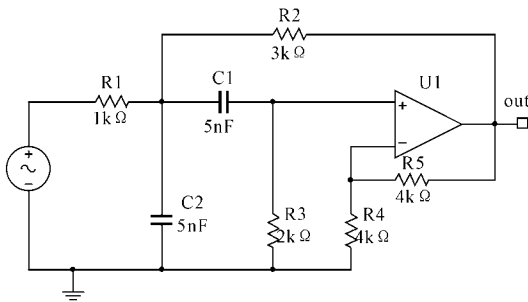


图1 带通滤波器电路

因为电路单硬故障占70%~80%,为使问题不至于太复杂,因此本文只对电路设置单硬故障,即设置R1、R2、R3、R4、R5、C1、C2开路故障及短路故障,加上电路正常模式共15种故障(表1)。本文利用Multisim提取幅频响应的有效点作为电路特征。

表1 单硬故障类

故障号	F1	F2	F3	F4	F5
故障类	正常	R1O	R1S	R2O	R2S
故障号	F6	F7	F8	F9	F10
故障类	R3O	R3S	R4O	R4S	R5O
故障号	F11	F12	F13	F14	F15
故障类	R5S	C1O	C1S	C2O	C2S

如图2所示,各种状态下电路的幅频响应曲线会有所不同所示。本文提取曲线在10、4K、8K、16K、25K、44K、64K、95KHz处的幅值作为诊断系统的特征输入。电路进行故障仿真时,用阻值为1Ω的电阻来表示短路故障,用阻值为100MΩ的电阻来表示开路故障。对元件取3%的容差,电路每种故障模式进行20次Monte Carlo仿真。随机提取其中10个作为训练样本,另外10个作为测试样本。

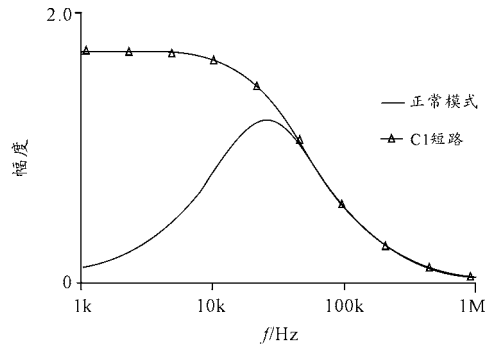


图2 电路各状态下幅频响应曲线

在各种故障模式下,F9、F10响应曲线相同,将F9、F10合为一类为F9。F3、F4、F6、F13响应曲线比较相近而且较弱。考虑到实际检测中无法检测到,将F3、F4、F6、F13、F14也合为一类为F3。故电路中共设置故障模式9个,加上正常模式共10种模式。本文采用径向基核函数,核宽度 $\delta=3.0$,依据上文对样本进行分层聚类得以下的最优决策树。

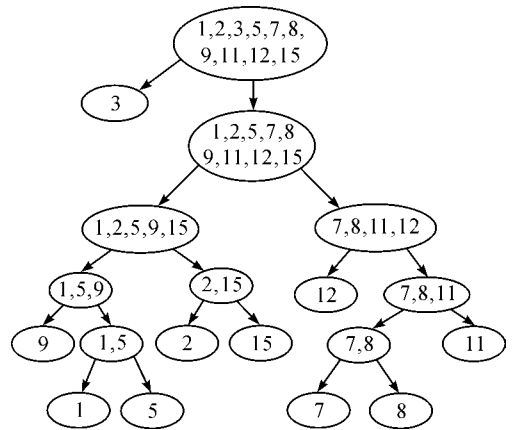


图3 最优决策树

一对方方法需训练的二分类支持向量机个数为45个,测试时需要历遍所有的子分类器,本文方法只需训练9个二分类支持向量机,测试时历遍个数不大于5个,对训练和分类复杂度比较如表2、3所示。

表2 训练和分类复杂度比较

	OVO	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3 诊断结果

	OVO	本文方法
测试时间/s	0.9	0.4
训练正确率/%	98	98
测试正确率/%	97	97
样本总正确率/%	97.5	97.5

4 结束语

本文将分层聚类分析与 SVM 相结合,设计了遗传粒子群算法实现对数据分层聚类,解决以聚类中心进行聚类时容易把同一类样本划分到两个聚类中心中,通过滤波器故障模式的诊断,获得了良好的诊断效果。解决二叉树支持向量机朴素树结构难确定的问题,并一对一多分类方法进行比较。结果表明,在不影响分类精度下,减少测试时间,降低模型复杂度。

参考文献:

- [1] 彭敏放. 容差模拟电路故障诊断屏蔽理论与信息融合方法研究[D]. 长沙:湖南大学,2006.
- [2] 许建华,张学工. 统计学习理论[M]. 北京:电子工业出版社,2004.
- [3] Bredensteiner E J, Bennett K P. Multicategory Classification by Support Vector Machines[J]. Computational Optimization and Applications, 1999, 12(1): 53 - 79.
- [4] Krebel U H. Cx Pairwise Classification and Support Vector Machines[M]. Advances in Kernel Methods; Support

Vector Learning, 1999:255 - 268.

- [5] Platt J C, Cristianini N, Shawe-Taylor J. Large Margin DAG's for Multi-class Classification[J]. Advances in Neural Information Processing Systems, 2000, 12:547 - 553.
- [6] H SU C W, L N C J. A comparison of methods for multi-class support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):415 - 425.
- [7] 连可,黄建国,王厚军,等. 一种基于遗传算法的 SVM 决策树多分类策略研究[J]. 电子学报, 2008, 36(8): 1502 - 1507.
- [8] 宋国明,王厚军,姜书艳,等. 一种聚类分层决策的 SVM 模拟电路故障诊断方法[J]. 仪器仪表学报 2010, 31(5):998 - 1000.
- [9] 陈世杰,连可,王厚军. 遗传算法优化的 SVM 模拟电路故障诊断方法[J]. 电子科技大学学报,2009, 38(4): 553 - 555.
- [10] 李丽,牛奔. 粒子群优化算法[M]. 北京:冶金工业出版社,2010:27 - 28.
- [11] 刘文涛. 基于改进遗传算法的网络入侵检测[J]. 重庆工商大学学报:自然科学版,2010,27(5):476 - 480.
- [12] 杨华芬. 一种改进的遗传算法在函数优化中的应用[J]. 重庆工商大学学报:自然科学版,2009,26(2): 148 - 151.
- [13] 张超杰,贺国,梁述海,等. 基于改进粒子群算法的模拟电路测试点选择[J]. 华中科技大学学报:自然科学版, 2009,11(37):32.
- [14] 沈春华,卢晶,徐柏龄. 浮点数编码的遗传算法在系统辨识中的应用[J]. 应用科学学报,2001,19(4):299 - 301.

(责任编辑 刘 舸)

(上接第 85 页)装备的不能电磁兼容,严重影响雷达各种功能的发挥。在作战中,除了侦察校射雷达之外,还有很多无线电通信、卫星通信、散射通信、移动通信、炮兵无人机和指挥自动化系统都要使用电磁频谱,如不能合理地频谱进行分配和管理,必然会出现互相干扰状况。因此,战场频谱管理机构要制定统一的频谱分配表,统一为己方各类电子信息分配频率,确保活动目标侦校雷达在使用不和其它用频装备在频率上相冲突。

3) 发展活动目标侦校雷达自身的电子对抗能力。活动目标侦校雷达除完成自身的功能和任务外,还能对敌方雷达、电台和光电设备实施干扰,利用有源或无源干扰器材设置电磁屏障,对敌方干扰设备进行有效的屏蔽。

参考文献:

- [1] 刘尚合. 武器装备的电磁环境效应及发展趋势[J]. 装

备指挥技术学院学报,2005(1).

- [2] 邹阳森,刘东,杨建文,等. 复杂电磁环境系统的效能评估[J]. 兵工自动化,2009,28(11):86 - 88.
- [3] 王旭豪,王文发,黄星. 复杂电磁环境下炮兵指挥信息系统面临的问题及对策[J]. 兵工自动化,2009,28(5):61 - 62.
- [4] 江振华. 无线电台抗复杂电磁干扰的几点措施[J]. 四川兵工学报,2010,31(6):46 - 47.
- [5] 王汝群. 战场电磁环境效应[M]. 北京:解放军出版社,2006.
- [6] 甄新究,张德勤,王进树. 复杂电磁环境下炮兵作战行动研究[M]. 北京:海潮出版社,2007.
- [7] 刘军. 信息化战场复杂电磁环境对作战的影响[J]. 西安政治学院学报,2007(4):66 - 68.

(责任编辑 刘 舸)