

文章编号: 2095-4980(2018)05-0912-06

基于随机森林算法的模拟电路故障诊断

王 玲^{1,2}, 周东方¹, 生拥宏¹, 张 伟^{*2}

(1.信息工程大学 信息工程学院, 河南 郑州 450001; 2.河南农业大学 机电工程学院, 河南 郑州 450002)

摘 要: 为提高模拟电路参数故障的诊断率, 提出基于多特征向量提取和随机森林(RF)算法的模拟电路故障诊断新方法。采用时域和频域特征向量组合的多维特征向量以反映不同故障特征, 经 RF 算法进行决策, 并对决策树棵数及候选特征向量个数进行优化。故障诊断实验结果表明, 所提方法能较好地实现容差模拟电路故障诊断, 与支持向量机(SVM)方法相比, 表现出更好的分类性能; 与小波(包)特征提取方法相比, 简化了多维数据特征提取步骤, 易于实现在线故障诊断。

关键词: 随机森林; 多特征; 故障诊断; 决策树

中图分类号: TN129

文献标志码: A

doi: 10.11805/TKYDA201805.0912

Fault diagnosis of analog circuits based on Random Forest algorithm

WANG Ling^{1,2}, ZHOU Dongfang¹, SHENG Yonghong¹, ZHANG Wei^{*2}

(1.Communication Engineering College, Information Engineering University, Zhengzhou Henan 450001, China;

2.College of Mechanical & Electronic Engineering, Henan Agriculture University, Zhengzhou Henan 450002, China)

Abstract: In order to improve the diagnostic rate of analog circuit parametric fault, a new fault diagnosis method based on multi-feature extraction and Random Forest(RF) algorithm is presented. By combining the time domain and frequency domain feature vectors for multi-feature vector data to reflect different faults, the decision is made by RF algorithm, the number of decision tree and candidate feature vectors are optimized. The fault diagnosis experiment results show that the proposed method can better realize analog circuit fault diagnosis, and shows a better classification performance compared with Support Vector Machine(SVM) method, and compared with wavelet(packet) feature extraction method, the multi-dimensional data feature extraction is simplified for online fault diagnosis.

Keywords: Random Forest; multi-feature; fault diagnosis; decision tree

电路分析、电路综合和电路故障诊断是电路理论的三大研究领域^[1]。模拟电路故障诊断技术自 20 世纪 70 年代逐渐形成了一个较为完整的系统理论体系,是现代电路理论的难点,其任务是在已知电路网络拓扑结构的基础上,设计激励向量,通过对电路各可测节点的响应信号分析来定位故障元件及确定故障元件参数。尤其是近年,含有大量模拟电路的电子设备广泛用于各个领域,模拟电路故障诊断技术直接影响电子设备的可靠性。

随着人工智能及计算机的飞速发展,产生了基于知识的模拟电路故障诊断方法,此方法不需要精确的数学模型,且具有“智能”特性,是一种很有生命力的方法。基于知识的故障诊断方法主要有:专家系统方法、故障树方法、神经网络方法及信息融合故障诊断方法等^[2-4]。其中故障树诊断方法因诊断过程直观形象易懂、故障树图层次分明、可移植、诊断知识可共享的特点而成为模拟电路故障诊断工程应用的常用方法。但在模拟电路系统中,因器件容差及器件先后级联等存在模糊问题,传统单一故障树的诊断率较低。随机森林(RF)算法是基于统计学习理论的组合分类器,将 bootstrap 重抽样方法和决策树算法相结合,通过投票进行分类和预测,突破了单分类器性能无法提升的瓶颈^[5]。

为更好地对容差模拟电路故障进行故障分类,提高其分类性能,在故障树方法的基础上,对 RF 算法在模拟电路故障诊断中的应用进行讨论。

收稿日期: 2017-09-04; 修回日期: 2017-10-18

*通信作者: 张 伟 email:zwz04@163.com

1 RF 算法

RF 算法是机器学习、计算机视觉等领域应用较为广泛的一种算法^[6]，它不仅可以用来做分类(包括二分类和多分类)，也可用来做回归预测，还可以作为一种数据降维的手段^[7]。

在 RF 中，将生成 K 棵决策树 $\{h(x; \theta_k), k=1, 2, \dots, K\}$ ，通过 K 棵决策树对输入的数据变量 x 进行无加权投票，来确定 x 归属哪一类。其中 $\{\theta_k\}$ 是独立同分布的随机向量，是采用 bootstrap 重抽样方法从原始样本中抽取相同个数的样本所组成的特征子集构成。由于每棵树都是一个独立的判断分支，且特征子集是从原始特征空间中随机抽取的，故影响 RF 性能和效率的 3 个参数是：RF 中树的棵数 N_{tree} 、子集中候选特征数量 M_{try} 和叶节点的样本数 N_{size} 。若 N_{tree} 较小，则影响 RF 的分类精确度； N_{tree} 过大，则构建 RF 的时间较长，复杂度增大，且系统可解释性变差；子集中包含的候选特征在建树过程中是保持不变的， M_{try} 越小，决策树的随机性越强，则分类精确度就会降低； M_{try} 越大，则所选择的相关特性会更多，随机性降低，RF 的多样性也会降低，从而影响 RF 的决策精确度。因此， M_{try} 可用来协调 RF 的分类性能和多样性； N_{size} 表示叶节点的最小样本数，已有研究表明，RF 的性能对 N_{size} 不敏感^[8]。因 RF 是通过对样本集和样本子集的候选特征随机抽样产生的决策树，故未被抽中的数据称为袋外(Out Of Bag, OOB)数据^[9]，可用 OOB 来检测每棵树分类效果的好坏，计算 N_{tree} 棵决策树的平均 OOB 误分率，得到 OOB 误差估计来衡量 RF 的分类性能。

OOB 误差估计定义为：

$$OOB_{\text{error}} = 1 - \frac{\sum_{i=1}^{N_{\text{tree}}} I(f(x_i, y_i) = y_i)}{N_{\text{tree}}} \quad (1)$$

式中 I 是示性函数， $I(f(x_i, y_i) = y_i)$ 表示 OOB 决策树集合对数据 (x_i, y_i) 进行分类的正确结果。OOB 误差估计越小，则 RF 的分类性能越好。

2 基于 RF 算法的模拟电路故障诊断步骤

对有一定容差的模拟电路软故障进行诊断，首先需要提取各故障模式下的故障特征组成特征空间，然后输入分类器进行故障模式分类，由于 RF 是有监督的分类算法，需要大量的训练样本来构建模型，故 RF 算法用于模拟电路故障识别分为训练和故障分类测试 2 个阶段，如图 1 所示。分类器的性能取决于样本特征的选取，RF 算法的性能主要依赖于故障特征提取的效果，故需对特征提取方法进行研究。

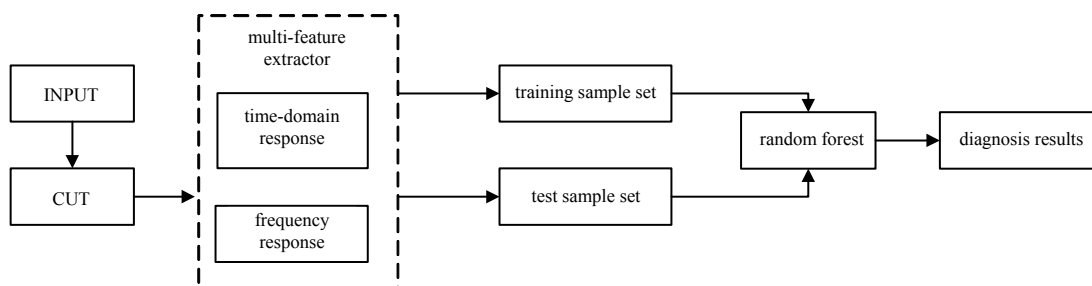


Fig.1 Fault diagnosis process of analog circuit

图 1 模拟电路故障诊断过程

2.1 模拟电路多特征提取

传统的故障诊断方法往往以一个特征信号(如电压值)作为测试指标，容易导致不同故障模式的响应区间重叠，诊断结果出现模糊。针对这一问题，文献[10]提出采用多个测点和多特征信息构建原始样本数据作为 BP 神经网络的输入来训练神经网络，获得较好的故障诊断率；文献[11-12]提出基于时域阶跃响应波形的多特征向量的可测性分析模型——多特征故障字典模型，提高了模拟电路的故障检测率和故障隔离率。

多特征向量提取方法是提取电路在不同激励下各测点的故障特征向量，来建立电路特征信息矩阵，以期全面反映电路的状态信息。对于电路特征信号的选用，科研人员进行了大量研究，并提出了众多故障特征指标，包含时域瞬态特征、稳态波形特征、频域相关特征、小波系数特征、统计特征等众多类别^[13]，这些特征提取方法大多需要复杂的求解计算，获取过程较为繁琐，不适合工程应用。且小波系数是基于时域瞬态响应进行变换的，统

计特征可以是时域或频域。综上,特征提取的基础数据均是时域和频域特征,为提高在线故障诊断效率,在不影响故障诊断效率的前提下,尽量缩减特征提取时间及步骤,故本文选用能在线直接读取的时域动态特征量和频域特征量作为电路故障特征。

2.1.1 时域动态特征提取

动态性能常用来衡量一个系统稳定性能,标准动态性能指标通常根据系统的阶跃响应曲线来计算。对于欠阻尼系统常用的动态性能指标有上升时间、峰值时间、过渡过程时间、最大超调量、振荡次数、延迟时间 6 个指标,也可添加时域统计量信息等。实验证明,特征量的增加提高了故障诊断的精确度,但特征量的增加使整个诊断过程的时间变长,最优的特征集合则可以在最短的时间内获得最优的测试结果^[14-15]。

为更好地反映时域瞬态响应曲线,且容易区分不同故障类型,优化特征量,本文采用上升时间(t_1)、峰值时间(t_2)、过渡过程时间(t_3)、最大过冲电压(U_1)、最小响应值(U_2) 5 个特征量,见图 2,分别定义如下。

$$t_1 = t_{0.9U_m} - t_{0.1U_m} \quad (2)$$

$$t_2 = t_{U_m} \quad (3)$$

$$t_3 = t_{U_0} \quad (4)$$

$$U_1 = U_m \quad (5)$$

$$U_2 = U_{\min} \quad (6)$$

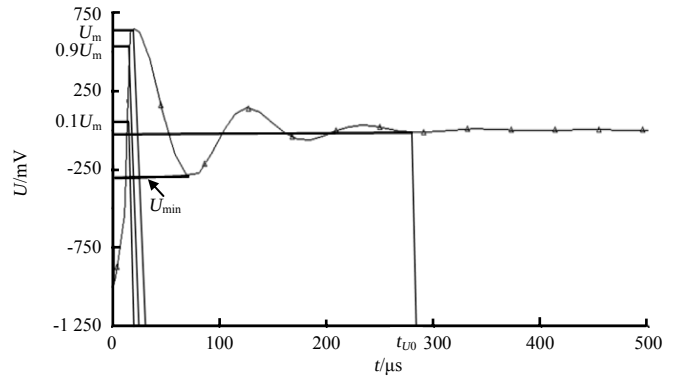


Fig.2 Time-domain response curve

图 2 时域响应曲线

2.1.2 频域特征提取

模拟电路的频域分析是电路系统常用的分析方法,电路的频域特征可通过示波器或频谱仪直接获取。以工程中常用的模拟滤波器为例,反应滤波器性能的重要指标是频响曲线,根据其性能分为低通滤波器、高通滤波器、带通滤波器、带阻滤波器,一般用中心频率、最大频响值、3 dB 截止频率、通带带宽、插入损耗、纹波等作为其性能指标。

2.2 RF 算法步骤

RF 算法需要通过大量训练样本建立故障诊断模型,在模型上进行故障诊断测试,具体流程如图 3 所示。

训练阶段的实现步骤如下:

- 1) 通过仿真获取原始训练样本数据集 $(x_i, y_i)_{NM}$, 其中 x_i 是 M 维样本向量, M 为样本属性个数, y_i 是 x_i 所对应的分类标号, N 为训练样本数目;
- 2) 在原始数据空间,采取有放回抽样的方式生成 K 个新训练样本子集;
- 3) 在新训练样本集中随机挑选 M_{try} 个属性,按照节点不纯度最小原则进行分支生长成 k 棵决策树,组成 RF 组合分类器。

故障分类的实现步骤如下:

- 1) 将未知故障样本输入到 RF 模型中,由各个决策树分别给出故障类别;
- 2) 集合所有决策树的判断结果,通过投票决定所属的故障类别;
- 3) 可通过式(1)的 OOB 误差估计来衡量 RF 算法的分类性能。

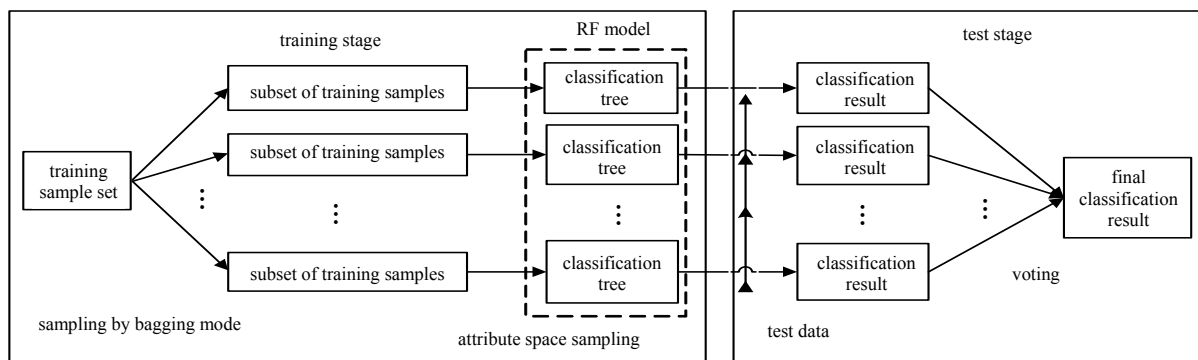


Fig.3 RF algorithm steps

图 3 RF 算法步骤

3 实验结果与分析

为比较、验证本文提出的故障诊断方法的先进性，选用参考文献[16]中的 ITC97 国际标准电路进行故障诊断，该电路的正常标称值如图 4 所示。设置电容和电阻的容差分别为正常值的 10%和 5%，超出容差 50%则为软故障。采用 multisim 软件进行电路仿真。经过灵敏度分析，故障敏感元件为 $R_1\sim R_4, C_1\sim C_2$ ，则故障模式分别为 $R_1 \pm 50\%$, $R_2 \pm 50\%$, $R_3 \pm 50\%$, $R_4 \pm 50\%$, $C_1 \pm 50\%$, $C_2 \pm 50\%$ ，再加上正常状态共有 13 种故障模式，如表 1 所示， $\uparrow$ 和 $\downarrow$ 表示其高于或低于正常容差值。

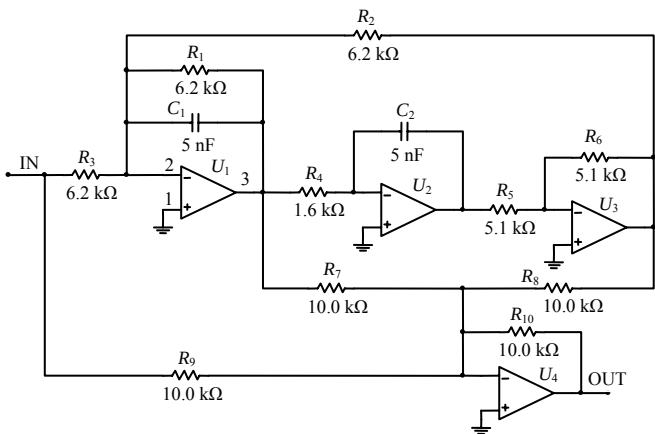


Fig.4 Four-op-amp high pass filter circuit
图 4 四运放高通滤波电路

表 1 四运放高通滤波电路故障的故障代码及类型
Table1 Four-op-amp high pass filter circuit fault codes and types

fault code	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
fault type	normal	$R_1\uparrow$	$R_1\downarrow$	$R_2\uparrow$	$R_2\downarrow$	$R_3\uparrow$	$R_3\downarrow$	$R_4\uparrow$	$R_4\downarrow$	$C_1\uparrow$	$C_1\downarrow$	$C_2\uparrow$	$C_2\downarrow$

3.1 特征参数提取

时域动态参数提取：输入激励信号电压为 1 V，周期为 500 μs ，脉冲宽度为 10 μs ，含有丰富频域信息的单脉冲电压信号。进行动态仿真采用上升时间、峰值时间、过渡过程时间、最大过冲电压、最小响应值 $[t_1, t_2, t_3, U_1, U_2]$ 5 个参数，每种故障模式进行 100 次 Monte Carlo 分析。

频域特征参数提取：对于高通滤波电路，选取中心频率、最大频响值、3 dB 截止频率 $[f_0, A_m, f_H]$ 3 个参数。

时域和频域特征参数组合得到每种故障模式 8 维特征向量 $[t_1, t_2, t_3, U_1, U_2, f_0, A_m, f_H]$ ，每种故障模式进行 100 次 Monte Carlo 分析，则总数据量为 $13 \times 8 \times 100$ 。

部分故障数据如表 2 所示：

表 2 部分故障特征数据
Table2 Partial fault characteristic data

	$t_1/\mu\text{s}$	$t_2/\mu\text{s}$	$t_3/\mu\text{s}$	U_1/V	U_2/V	f_0/MHz	A_m/V	f_H/MHz
F_0	2.455 9	20.361 2	316.225 4	0.646 140	-0.544 30	18.286 0	1.947 9	6.727 6
F_1	2.368 0	22.494 1	362.667 7	0.714 766	-0.421 82	18.736 4	3.149 7	6.419 1
F_2	1.973 1	17.758 5	130.229 8	0.501 883	-0.080 41	17.625 8	1.192 5	7.068 3
F_3	2.724 8	23.842 0	294.959 1	0.731 022	-0.270 88	22.961 5	2.490 0	1.989 6
F_4	2.043 6	16.348 8	335.831 1	0.529 521	-0.312 11	11.414 7	1.561 5	13.010 3
F_5	2.857 2	19.428 6	308.000 0	0.400 000	-0.181 48	16.607 9	1.414 2	8.502 5
F_6	4.535 1	20.975 1	349.773 2	1.334 700	-0.623 66	25.016 7	4.163 1	0
F_7	1.714 3	20.000 0	305.714 3	0.537 463	-0.201 59	19.778 5	1.654 9	5.579 2
F_8	3.401 3	18.707 5	356.009 1	0.901 967	-0.505 99	16.037 6	2.662 7	9.199 4
F_9	4.001 2	24.571 4	446.285 7	0.518 891	-0.289 26	19.994 1	2.465 9	5.237 7
F_{10}	3.401 3	16.439 9	134.920 6	0.924 916	-0.271 11	13.033 7	1.357 6	11.525 1
F_{11}	1.973 6	19.731 6	268.350 4	0.522 127	-0.189 90	19.463 5	1.5867	5.491 9
F_{12}	2.834 4	18.140 6	315.192 7	1.001 300	-0.620 20	14.571 1	3.020 7	10.264 2

3.2 RF 构建

获取多特征故障样本集后，按照 bagging 方式构建故障样本子集，随机选取 M_{try} 个特征，选取分类效果最好的一个特征作为该结点的分裂属性，按照二分递归分隔的方式生长决策树，决策树生长后，并不需要剪枝，重复此过程直到建立 k 颗决策树来构建 RF。

RF 中树的数量 N_{tree} 、子集中候选特征数量 M_{try} 对故障的识别率影响最大，需要对其参数进行优化。首先观察 N_{tree} 的取值，在 400~1 000 棵时对识别率的影响不大，故先固定 $N_{\text{tree}}=500$ ，调整 M_{try} 的值为 1~9，图 5 为 RF 算法性能和不同 M_{try} 之间的关系图，可知在 $M_{\text{try}}=3$ 和 $M_{\text{try}}=8$ 时，均能获得最优的分类正确率，为提高运算速度，降低算法复杂度，选择 $M_{\text{try}}=3$ 。

固定 $M_{\text{try}}=3$ ，调整 N_{tree} 的范围在 50~1 000 之间，间隔 50。得到分类正确率与决策树棵数之间的关系如图 6 所示，可知，RF 分类正确率随着决策树棵数的增加而增加，但达到一定棵数后分类正确率则基本保持不变，考虑到运行效率，选择 $N_{\text{tree}}=300$ 。诊断结果如表 3 所示，可见只有 F_7 和 F_{11} 出现误诊， F_7 的故障诊断率为 95.24%， F_{11} 的故障诊断率为 96.15%，其他故障诊断率均为 100%，平均故障诊断率达到 99.34%。

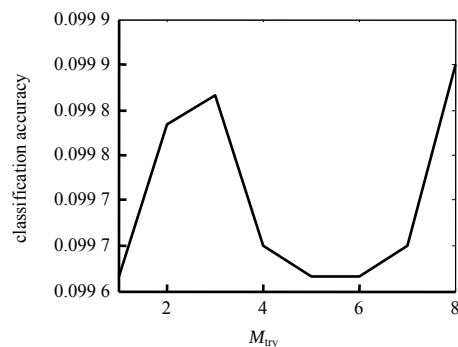


Fig.5 Relation between classification accuracy and number of eigenvectors
图 5 分类正确率与特征向量个数关系

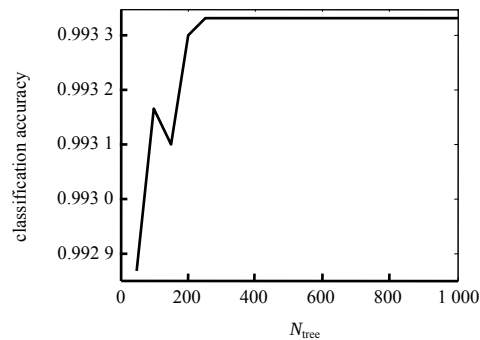


Fig.6 Relationship between classification accuracy and decision tree
图 6 分类正确率与决策树棵数关系

表 3 RF 的分类准确率(%)
Table3 Classification accuracy of RF(%)

	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
accuracy	100	100	100	100	100	100	100	95.24	100	100	100	96.15	100

目前比较常用的容差电路故障诊断方法中故障特征提取方法为基于时域动态响应信号的小波特征提取及小波包能量提取,分类采用 SVM,其中 SVM 的参数采用群智能算法,如混合入侵杂草算法、粒子群优化算法进行优化,其故障诊断率分别为 98.698%和 98.278%^[16]。

4 结论

基于 RF 算法的容差电路故障诊断方法,故障特征提取比较简单,工程应用中更易实现,与小波(包)特征提取方法相比,也更为可靠。经过 RF 算法优化可知,所需故障特征数据为 3 个,故障特征数据并不是越多越好,故障诊断率平均为 99.34%,高于现有文献中基于 SVM 分类方法的故障诊断结果。

参考文献:

[1] 马敏. 电路与电子系统故障诊断技术[M]. 北京:电子工业出版社, 2016. (MA Min. Fault diagnosis technology of circuit and electronic system[M]. Beijing:Electronic Industry Press, 2016.)

[2] SUN J,WANG C. Analog circuit fault diagnosis based on mRMR and optimized SVM[J]. Chinese Journal of Scientific Instrument, 2013,34(1):221–226.

[3] LI X,ZHANG Y,WANG S,et al. Analog circuits fault diagnosis by GA-RBF neural network and virtual instruments[C]// International Symposium on Instrumentation & Measurement, Sensor Network and Automation. Sanya,China:IEEE, 2011:1–5.

[4] TANG S,CAI H,LI Z. Fault diagnosis fusion method for analog circuits based on wavelet and neural network[J]. Journal of Central South University, 2015,46(1):127–134.

[5] 曹正凤. RF 算法优化研究[D]. 北京:首都经济贸易大学, 2014. (CAO Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing:Capital University of Economics and Business, 2014.)

[6] AURET L,ALDRICH C. Diagnostic monitoring of concentrator circuits with random forest models[C]// International Mineral Processing Congress,2010 IMPC. Brisbane,Australia:[s.n.], 2010:3111–3121.

[7] LIU X,YANG S,LI C,et al. SPAR:a random forest-based predictor for self-interacting proteins with fine-grained domain information[J]. Amino Acids, 2016,48(7):1655–1665.

[8] DIAZ-URIARTE R,DE ANDRES S A. Gene selection and classification of microarray data using random forest[J]. BMC Bioinformatics, 2006,7(1):3.

[9] MELLOR A,BOUKIR S,HAYWOOD A,et al. Exploring issues of training data imbalance and mislabeling on random forest performance for large area land cover classification using the ensemble margin[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015(105):155–168.

[10] PAN Q,YANG C. Fault diagnosis of analog circuit based on multi-test points and multi-feature information[J]. Applied Mechanics & Materials, 2013(313/314):277–280.

[11] LI M,WANG C L,LONG B,et al. Multi-feature fault dictionary and its application in testability analysis for analog circuits[J]. Journal of Electronic Measurement & Instrumentation, 2015(3):368–374.

- [12] LONG B,WANG C,GAO Y,et al. A testability index analysis method for analog circuits based on multi-feature model[J]. Chinese Journal of Scientific Instrument, 2013,34(4):914-919.
- [13] 魏子杰,陈圣俭,周校晨. 基于特征信息矩阵的容差模拟电路故障诊断[J]. 哈尔滨工程大学学报, 2016,37(9):1256-1260. (WEI Zijie,CHEN Shengjian,ZHOU Xiaochen. Fault diagnosis for analog circuits with tolerance based on circuit feature information matrix[J]. Journal of Harbin Engineering University, 2016,37(9):1256-1260.)
- [14] 王彩利. 基于多信号特征模拟电路可测性分析与故障诊断[D]. 成都:电子科技大学, 2014. (WANG Caili. Testability analysis and fault diagnosis of analog circuits based on multi-signal features[D]. Chengdu,China:University of Electronic Science and Technology of China, 2014.)
- [15] LONG B,LI M,WANG H,et al. Diagnostics of analog circuits based on LS-SVM using time-domain features[J]. Circuits Systems and Signal Processing, 2013,32(6):2683-2706.
- [16] HU Hongzhi,TIAN Shulin,GUO Qing,et al. The application of HIWO-SVM in analog circuit fault diagnosis[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2015,29(8):1-24.

作者简介:



王 玲(1980-),女,河南省周口市人,讲师,在读博士研究生,主要研究方向为电路故障诊断与预测,email:wangling0351@126.com.

周东方(1963-),男,浙江省诸暨市人,教授,博士生导师,主要研究方向为微波网络理论与技术.

生拥宏(1977-),男,江苏省海安县人,博士,副教授,主要研究方向为自动测试系统.

张 伟(1976-),男,河南省周口市人,硕士,实验师,主要研究方向为模拟电路设计.

第四届燃烧诊断技术学术研讨会

11月20-22日,绵阳 <http://www.csoe.org.cn/meeting/cd2018/>

主办单位:中国宇航学会光电专委会 承办单位:中国空气动力研究与发展中心

大会主席:乐嘉陵 院士(中国空气动力研究与发展中心)

程序委员会委员(按姓氏拼音排序):

蔡伟伟 上海交通大学,超星 清华大学,陈爽 中国空气动力研究与发展中心,冯大强 624所,洪流 西安航天动力研究所,胡志云 西北核技术研究所,阚瑞峰 中科院安徽光机所,李博 天津大学,李挺 北京航空航天大学,刘小勇 航天三院,尚守堂 606所,王广宇 航天工程大学,王智化 浙江大学,杨顺华 中国空气动力研究与发展中心,于欣 哈尔滨工业大学,张玉银 上海交通大学,朱家健 国防科技大学

会议议题/征文范围

- 先进/前沿燃烧诊断技术
- 复杂条件下燃烧场诊断技术
- 激光光谱诊断技术精度提升和可靠性研究
- 实际燃烧器(发动机)试验燃烧流场测量
- 湍流燃烧燃烧诊断技术与CFD结合

包括以上方面,但不局限于此,欢迎与燃烧诊断有关的所有来稿。

所有稿件务必做好保密审查工作。投稿请登录在线投稿系统<http://events.kjtxw.com/tougao/cd2018.html>

支持期刊:《推进技术》EI、《红外与激光工程》EI、《光学精密工程》EI、《实验流体力学》中文核心、《燃烧科学与技术》中文核心、《激光技术》中文核心、《太赫兹科学与电子信息学报》中文核心

全文截稿时间:11月10日,联系人:徐 涵,010-63728336, xuhan@csoe.org.cn