

# 多分类 SVDD 混叠域识别的模拟电路故障诊断

全奇<sup>1</sup>, 胡双演<sup>1</sup>, 叶霞<sup>1</sup>, 张仲敏<sup>2</sup>, 李俊山<sup>1</sup>

(1. 第二炮兵工程大学 信息工程系, 西安 710025;

2. 西安通信学院, 西安 710106)

**摘要:** 针对多分类支持向量域数据描述 (SVDD) 方法中混叠样本诊断精度差的问题, 提出了一种带异类样本的多分类 SVDD 算法; 该方法在普通 SVDD 超球模型基础上, 对于存在混叠区域的类别, 以该类所有样本为目标类, 其他类与之混叠的样本为异类, 利用带异类样本的 SVDD 算法重新训练, 直至所有超球优化完毕; 仿真实验验证了文章算法消除混叠和提高精度的能力, 并将该算法应用于模拟电路故障诊断中; 相较于 SVDD 多分类算法、一对一和一对多 SVM 算法, 文章方法在模拟电路故障诊断中具有更高的诊断精度。

**关键词:** 支持向量域数据描述; 混叠; 异类样本; 故障诊断; 模拟电路

## An Approach to Discriminate Overlap Region of Multi-class Classification SVDD for Analog Circuits Fault Diagnosis

Tong Qi<sup>1</sup>, Hu Shuangyan<sup>1</sup>, Ye Xia<sup>1</sup>, Zhang Zhongmin<sup>2</sup>, Li Junshan<sup>1</sup>

(1. Department of Information Engineering, Second Artillery Engineering University, Xi'an 710025, China;

2. Xi'an Communications Institute, Xi'an 710106, China)

**Abstract:** To improve the discrimination accuracy of conventional multi-class classification support vector data description (SVDD) methods, a multiple classification Support vector data description algorithm with Negative Samples is proposed. Based on the general model of SVDD, the proposed algorithm treats the samples in the class as the target class, while the other classes of overlap and sample is heterogeneous for the overlap region. By using SVDD algorithm with Negative samples, the hypersphere model is trained again until all hypersphere models optimized. Simulated experimental results show that the proposed algorithm can eliminate overlap and improve the discrimination accuracy. The algorithm is applied in the implementation of analog circuits fault diagnosis, comparing with SVDD classification algorithm, one-to-one and one-to-many SVM algorithm, results show that the algorithm is more effective and higher accuracy in fault diagnosis.

**Keywords:** support vector data description; overlap region; heterogeneous samples; fault diagnosis; analog circuit

## 0 引言

随着电子技术的飞速发展, 导弹电子防护装备中的电路板集成度越来越高, 结构和功能也日趋复杂, 而模拟电路的故障诊断和维修能力相对较弱。模拟电路的测试由于受条件和人员水平的限制, 往往无法获得完备信息, 再加上当前电路中器件的紧密性, 单故障可能引发多故障, 导致故障诊断难度加大, 采用传统的故障诊断方法已不能满足现有装备的故障维修要求。因此, 研究合适的模拟电路混叠故障诊断方法对确保装备的有效性能及战争的胜利具有重要意义。

支持向量数据域描述 (support vector data description, SVDD)<sup>[1]</sup> 是一种常用的典型单分类器, 不少研究者<sup>[2-4]</sup> 将其扩展到多分类应用, 即多分类 SVDD (Multi-SVDD, MSVDD)。理想的 MSVDD 模型是各个超球之间彼此相互独立,

但在实际应用中很有可能出现两个或者多个超球相互混叠的情况, 如何对混叠区域样本进行正确分类, 并将改进算法应用到实际问题中, 是 MSVDD 算法研究的热点和难点之一。目前有许多研究学者提出各自的解决方案并应用到故障诊断问题中, 例如, 文献 [5] 从 SVDD 算法本身进行改进, 以消除混叠。文献 [6] 提出一种基于图谱空间映射 SVDD 的模拟电路故障诊断新方法, 改进了标准 SVDD 松弛的球体描述边界, 降低了因交叉区域而产生的误诊率。文献 [7-8] 采用了相对距离的判决方法, 提出一种基于隶属度函数的判决准则, 通过考虑两个混叠超球的大小, 来改善混叠问题。文献 [9] 利用绝对距离测度与相对距离测度实现了状态监测与故障分类的识别。文献 [10] 在 SVDD 算法基础上, 引入超球体核距离度量, 将多参数转化为单参数, 解决了参数过多相互矛盾的问题。

通过分析混叠产生的原因, 结合前期仿真实验结果, 证实如果在 SVDD 训练时加入少量异类样本, 则可以适当增强数据描述能力, 使得超球体更为紧致, 从而提高模型分类准确率。因此, 本文在 MSVDD 模型训练过程中加入异类样本, 提出一种带异类样本的 MSVDD 算法, 解决多分类混叠问题, 先利用仿真数据集验证算法消除混叠和提高精度的能力, 而后将该算法应用于模拟电路的故障诊断中, 获得了较理想的试验结果。

收稿日期: 2015-07-09; 修回日期: 2015-09-06。

**作者简介:** 全奇 (1988-), 男, 河南开封人, 硕士研究生, 主要从事装备故障诊断方向的研究。

叶霞 (1977-), 女, 江苏南京人, 副教授, 硕士生导师, 主要从数据库、指挥信息系统方向的研究。

李俊山 (1956-), 男, 陕西白水人, 教授, 博士生导师, 主要从事图像处理与目标识别、网络信息安全、电子对抗模拟与仿真方向的研究。

## 1 MSVDD 算法及其不足

相比于支持向量机“一对一<sup>[11]</sup>”、“一对多<sup>[12]</sup>”等多分类算法, 多分类 SVDD 算法具有模型简单、扩展性强、计算量少等优点。MSVDD 算法的基本思想是: 针对每类样本分别进行 SVDD 训练, 得到各自的超球模型, 然后利用各模型对测试样本进行分类。

假设有  $k$  类样本集  $X_k = \{x_1^k, x_2^k, \dots, x_{N_k}^k\}$ , 其中  $x_i^k \in \mathbf{R}^d$ ,  $N_k$  表示  $k$  类样本的数量, 算法的目标是寻找  $k$  个尽可能多的包含各自样本的最小超球体  $S_k$ ,  $S_k$  的半径  $R_k$  和球心  $a_k$  分别可由下式得到:

$$R_k^2 = 1 - 2 \sum_{i=1}^{N_k} \alpha_i^k K(x_i^k, x_{sv}^k) + \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \alpha_i^k \alpha_j^k K(x_i^k, x_j^k) \quad (1)$$

$$a_k = \sum_{i=1}^{N_k} \alpha_i^k \varphi(x_i^k) \quad (2)$$

其中:

$\alpha_i^k \geq 0$  为 Lagrange 乘子;

$x_{sv}^k$  为支持向量样本, 即少量  $\alpha_i^k > 0$  的样本;

$\varphi$  为映射函数;

$K(x_i^k, x_j^k)$  为核函数, 用以替代内积运算  $[\varphi(x_i^k), \varphi(x_j^k)]$ ,

这里选用高斯核函数, 即:

$$K(x_i^k, x_j^k) = \exp\left(-\frac{\|x_i^k - x_j^k\|^2}{2\sigma^2}\right) \quad (3)$$

其中:

$\sigma$  为高斯核参数。

任一测试样本  $x$  与超球  $S_k$  球心之间的距离可表示为:

$$R_{sk}^2 = \|\varphi(x) - a_k\|^2 = 1 - 2 \sum_{i=1}^{N_k} \alpha_i^k K(x_i^k, x) + \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \alpha_i^k \alpha_j^k K(x_i^k, x_j^k) \quad (4)$$

通过比较  $R_{sk}^2$  与  $R_k^2$  的大小就可以判别  $x$  落入哪类超球中, 从而对  $x$  进行归类。但传统 MSVDD 算法通常存在样本混叠问题, 这会给算法的分类精度带来很大影响。出现样本混叠区域的原因主要有两个方面:

1) 样本自身原因, 比如各类样本数量不平衡、样本之间的模糊性等。目前, 已有不少研究者针对样本自身原因提出相关的改进算法, 如文献 [13]。

2) 超球体紧致性不足。通过增大拒识率  $\text{fracrej}$  或减小高斯参数  $\sigma$  的方法可以提高超球紧致性, 但同时也使得模型的推广能力下降; 因此, 在模型训练前需要先对参数对 ( $\text{fracrej}$ ,  $\sigma$ ) 进行优化。

## 2 带异类样本的 MSVDD 算法

通过分析多分类 SVDD 混叠原因, 本文从另一角度出发, 在最佳参数对的前提下, 通过提高超球体紧致性, 从而改善多分类模型的样本混叠问题。为了提高超球的紧致性, 引入了带异类样本的 SVDD 算法 (SVDD with Negative Samples, NSVDD), NSVDD 算法是在 SVDD 训练时加入少量异类样本, 研究发现这样可以适当提高数据描述能力和模型分类准确率。当多分类存在混叠区域时, 通过 NSVDD 算法重新训练混叠超球, 可以消除混叠, 鉴于此, 构建了 NMSVDD 算法训练模

型, 提出了带异类样本的多分类 SVDD 算法 (NMSVDD)。

### 2.1 NSVDD 算法

带异类样本的 SVDD 算法的目标是建立一个包围所有正常样本, 而将故障样本排除在外的最小超球体。用  $i, j$  表示正常样本的下标,  $p, q$  表示故障样本的下标, 将正常样本标记为  $y_i = 1$ , 故障样本标记为  $y_p = -1$ , 两类样本的数目分别为  $m$  和  $n$ 。超球体半径为  $R$ , 球心为  $a$ , 则求解优化问题可以表示为:

$$\min F(R, a, \xi, \xi_p) = R^2 + \sum_i C_i \xi_i + \sum_p C_p \xi_p \quad (5)$$

$$\text{s. t. } \varphi(x_i) - a^2 \leq R^2 + \xi_i, \varphi(x_p) - a^2 \geq R^2 + \xi_p$$

$$\xi_i \geq 0, \xi_p \geq 0, \forall i, p \quad (6)$$

其中:

$\xi_i, \xi_p$  为松弛变量;

$C_i, C_p$  为惩罚因子。

式 (5) 对应的 Lagrange 函数如下:

$$L(R, a, \xi, \xi_p, \alpha, \beta) = R^2 + \sum_i C_i \xi_i + \sum_p C_p \xi_p - \sum_i \beta_i \xi_i - \sum_p \beta_p \xi_p - \sum_i \alpha_i \{ - [\varphi(x_i), \varphi(x_i)] - 2a\varphi(x_i) + a \cdot a \} + R^2 + \xi_i \} - \sum_p \alpha_p \{ [\varphi(x_p), \varphi(x_p)] - 2a\varphi(x_p) + a \cdot a \} - R^2 - \xi_p \} \quad (7)$$

其中:

$\alpha_i \geq 0, \alpha_p \geq 0, \beta_i \geq 0, \beta_p \geq 0$  为 Lagrange 乘子。

由极值条件并化简可得:

$$\sum_i \alpha_i - \sum_p \alpha_p = 1, \sum_i \alpha_i \varphi(x_i) - \sum_p \alpha_p \varphi(x_p) = a \quad (8)$$

$$\text{s. t. } 0 \leq \alpha_i \leq C_i, 0 \leq \alpha_p \leq C_p, \forall i, p$$

定义新变量  $\alpha_i^* = y_i \alpha_i$ , 这里  $i$  既表示正常样本的下标, 又表示异类样本的下标, 且只有少数支持向量样本的  $\alpha_i^*$  不为 0。引入核函数  $K(x_i, x_j)$  代替内积运算, 根据对偶规则, 则可以得到 NSVDD 算法与标准 SVDD 算法的求解形式是一致的, 只需需用  $\alpha_i^*$  替换  $\alpha_i$  即可。因此在训练 NSVDD 时, 只需赋予 Lagrange 乘子一个类别标签, 就可以得到与普通 SVDD 一致的求解形式和结果。

### 2.2 NMSVDD 算法流程

NMSVDD 的基本思路是: 如果某类超球与其他类超球存在混叠区域, 则以该类所有样本为目标类, 其他类与之混叠的样本为异类重新训练超球, 以增强超球体的紧致性, 削减甚至消除混叠区域, 同时提高模型精度。其算法的流程图如图 1 所示, 其详细步骤如下。

Step1: 对待测电路信号进行提取及预处理, 形成模拟电路所需的故障样本, 建立样本集并将其分为训练样本和测试样本;

Step2: 利用 SVDD 训练  $L$  类样本得到  $L$  个初始超球体  $S_1, S_2, \dots, S_L$ ;

Step3: 用式 (4) 依次判断各类超球与其余超球之间是否存在混叠样本, 若不存在, 则该类样本初始模型优化完毕, 否则转下一步;

Step4: 对于存在混叠区域的类别, 以该类所有样本为目标类, 其他类与之混叠的样本为异类按 NSVDD 算法重新训练超球;

Step5: 判断各类超球与其余超球之间是否存在混叠样本, 直至所有超球优化完毕, 否则返回执行 Step4。

通过前期仿真实验，一般经过 2~3 次循环，基本上可以完全消除混叠情况，因此设置算法循环次数上限为 5 次。

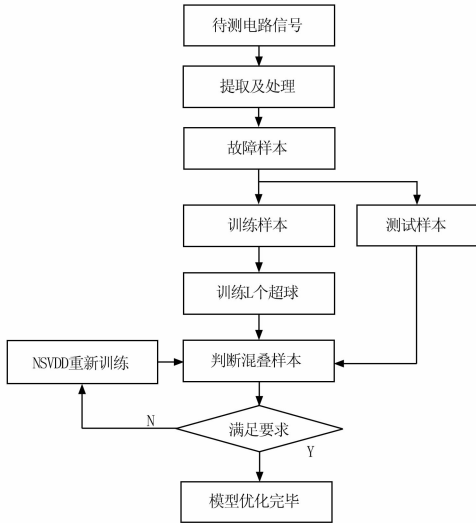


图 1 NMSVDD 算法流程

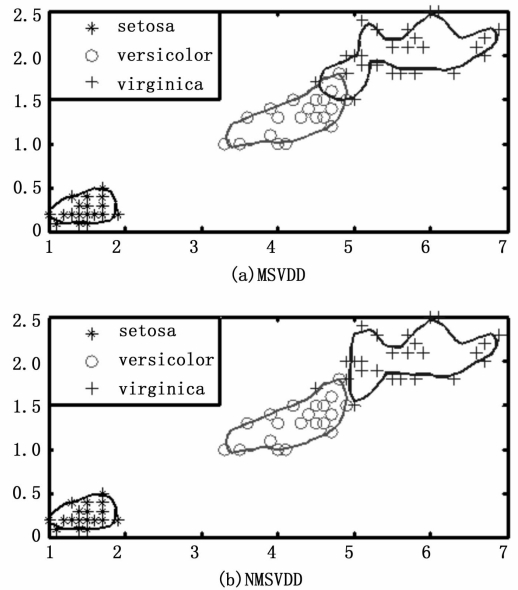


图 2 MSVDD 和 NMSVDD 算法分类

### 3 仿真与实例分析

#### 3.1 仿真实验与结果分析

选用鸢尾属植物数据集 (Iris dataset) 进行仿真分析, Iris 是标准多分类数据集, 含 setosa、versicolor、virginica 三类样本, 每类 50 个, 共 150 个样本点, 每个样本各有四维特征, 分别是: 萼片长度、萼片宽度、花瓣长度和花瓣宽度。为方便直观显示, 仅考虑花瓣长度与宽度两个属性, 各类前 25 个样本用于训练, 后 25 个样本用于测试。首先使用标准 MSVDD 算法训练各类超球, 应用交叉验证法确定参数对  $(fracrej, \sigma) = (0.05, 0.35)$ , 结果如图 2 (a) 所示。发现 versicolor 类与 virginica 类超球存在部分混叠, 然后利用 NMSVDD 算法对这两类样本重新训练, 参数对保持不变, 异类样本拒绝因子  $fracrej$  取 0.99, versicolor 类训练时, 将混叠域内的 virginica 类样本作为异类样本; 类似地, virginica 类训练时, 将混叠域内的 versicolor 类样本作为异类样本, 结果如图 2 (b) 所示。

由图 2 可以看出, 本文算法能够消除 versicolor 类与 virginica 类的混叠区域, 为进一步说明 NMSVDD 算法的效果, 分别对两种算法训练的 versicolor 和 virginica 超球的分类性能进行测试, 利用操作者工作曲线 (ROC) 进行对比, ROC 曲线越靠左上方, 曲线下的面积 (AUC 值) 越大, 表示分类器的分类性能越好, 图 3 上半部分是 versicolor 类的 ROC 曲线, 下半部分是 virginica 的 ROC 曲线。

图 3 直观显示了本文算法较传统 MSVDD 的优势, 特别是 virginica 类的 ROC 曲线, 两种算法的 AUC 值分别 0.960 和 0.996, 本文算法的分类性能有较大的提升。测试样本的分类结果如表 1 所示。

表 1 MSVDD 与 NMSVDD 分类结果

算法	setosa	versicolor	virginica
MSVDD	100%	92%	84%
NMSVDD	100%	96%	96%

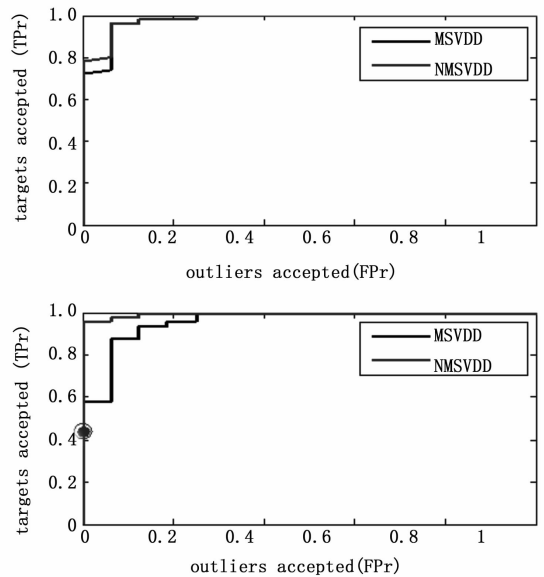


图 3 MSVDD 和 NMSVDD 算法的 ROC 曲线

通过仿真结果可知, NMSVDD 平均诊断结果为 96%, 较 MSVDD 具有更好的分类诊断效果, 验证了 NMSVDD 算法消除混叠和提高精度的能力。

#### 3.2 实例分析

为了验证带异类样本多分类 SVDD 算法应用于模拟电路故障诊断的效果, 以 Sallen-Key 带通滤波器电路为对象, 如图 4 所示。

##### 1) 激励信号和测试点的选择:

输入节点为 5 的激励信号是幅度为 5 V, 基频为 10 KHz 的电压信号, 节点 3 为输出测试点。

##### 2) 故障元件及故障模式选择:

采用 Multisim 软件以较难判断的软故障为主仿真待测电

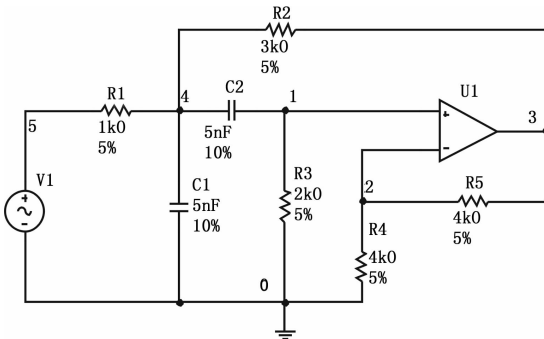


图 4 Sallen-Key 带通滤波电路

路。各元件的标称值已在图中标出, 其中电阻的容差范围为 5%, 电容的容差范围为 10%。通过对电路进行灵敏度分析, 发现 R1、R2、R3 和 C1、C2 的灵敏度较高, 对电路的输出影响较大, 因此电路故障模式设定如表 2 所示, 包括五类故障模式和电路正常模式。

表 2 电路故障模式设定表

电路状态	故障类型	故障模式	故障类别
正常	—	正常	F1
R1 故障	↑	超差 50%	
F2	R2 故障	↑	超差 50%
F3	R3 故障	↑	超差 50%
F4	C1 故障	↑	超差 50%
F5	C2 故障	↑	超差 50%

### 3) 故障样本的获取:

采集输出节点 3 的信号, 利用 Multisim 对每种故障模式进行 30 次的 Monte Carlo 分析, 共得到 180 个故障样本集, 90 个样本用来训练, 另外 90 个用来测试。

### 4) 故障诊断及结果分析:

首先使用 MSVDD 算法进行训练, 参数对  $(fracrej, \sigma) = (0.03, 5)$ , 分析后发现存在 4 种混叠故障, 对电路存在的混叠故障利用新算法重新训练, 异类样本拒绝因子取 0.99, 并在相同条件下将 NMSVDD 算法与基于 MSVDD 算法、一对一、一对多 SVM 诊断方法的测试结果进行对比, 诊断结果如表 3 所示。

表 3 算法分类器故障诊断结果比较

混叠故障	NMSVDD/(%)	MSVDD/(%)	一对一/(%)	一对多/(%)
F2F3	95.6	93.3	87.8	84.4
F2F6	91.1	87.8	77.8	78.9
F3F5	93.3	81.1	76.7	75.6
F3F6	96.7	90	80	74.4

由表 3 可以看出, 将 NMSVDD 算法应用于模拟电路的混叠故障诊断中, 其准确率比基于 MSVDD 算法、一对一和一对多 SVM 多分类算法的诊断准确率都要高许多, 说明当电路出现模糊故障即混叠故障, 模糊故障会陷入标准 SVDD 多个超

球体的交叉区域以及 SVM 分类面的交叉区域, 因此利用传统的故障诊断方法将很难正确判断。本文方法针对此问题, 将混叠的样本作为异类重新训练超球, 以消除混叠, 实验结果也证明了采用本文方法能够对模拟电路的混叠故障诊断实现快速故障检测和定位, 较传统的故障诊断算法有更好的诊断效果。

## 4 结论

本文研究了一种新的模拟电路故障诊断方法, 针对经典诊断方法在模拟电路混叠故障诊断中的不足, 提出了一种带异类样本的多分类 SVDD 算法。通过仿真实验验证了 NMSVDD 算法能够较好地消除混叠, 增强数据的描述能力, 使超球体更为紧致, 提高了模型分类准确率。另外, 在模拟电路的故障诊断实例中将本文算法与经典的故障诊断算法作对比, 从实验结果可以看出, NMSVDD 算法应用在模拟电路故障诊断中取得了较好的效果, 具有很大的应用潜力。

### 参考文献:

- [1] David M J T, Robert P W D. Support vector data description [J]. Machine Learning, 2004, 54: 45-66.
- [2] Kang W S, Im K H, Choi J Y. SVDD-Based Method for Fast Training of Multi-class Support Vector Classifier [A]. Third International Symposium on Neural Networks [C]. Chengdu, China; Springer, 2006: 991-996.
- [3] Lee D, Lee J. Domain Described Support Vector Classifier for Multi-classification Problems [J]. Pattern Recognition, 2007, 40 (1): 41-51.
- [4] Luo H, Wang Y R, Cui J. A SVDD Approach of Fuzzy Classification for Analog Circuit Fault Diagnosis with FWT as Preprocessor [J]. Expert Systems with Applications, 2011, 38 (8): 10554-10561.
- [5] Mu T T, Nandi A K. Multi class Classification Based on Extended Support Vector Data Description [J]. IEEE Transactions on Systems, Man, and Cybernetics—part B: Cybernetics, 2009, 39 (5): 1206-1216.
- [6] 罗 慧, 王友仁. 基于 GSM-SVDD 的模拟电路故障诊断方法 [J]. 电机与控制学报, 2013, 17 (1): 108-113.
- [7] Chiang J H, Hao P Y. A New Kernel-based Fuzzy Clustering Approach: Support Vector Clustering with Cell Growing [J]. IEEE Transaction on Fuzzy System 2003, 11 (4): 518-527.
- [8] Liu S, Liu Y K, Wang B, et al. An Improved Hyper-sphere Support Vector Machine [A] Third International Conference on Natural Computation [C]. Haikou, China; IEEE, 2007: 497-500.
- [9] 王 涛, 李艾华, 王旭平, 等. 基于 SVDD 与距离测度的齿轮泵故障诊断方法研究 [J]. 振动与冲击, 2013, 32 (11): 62-65.
- [10] 李 冬, 李本威, 王永华, 等. 基于聚类和多尺度优化的超球体核距离评估的航空发动机性能衰退 [J]. 推进技术, 2013, 34 (7): 977-983.
- [11] Debnath R, Takahide N, Takahashi H. A decision based on one-against-one method for multi-class support vector machine [J]. Pattern Anal Applic 2004, 27 (5): 164-175.
- [12] R. Rifkin, A Klautau. In defense of One-vs-all classification [J]. J Mach. Learning Res, 2004 (5): 101-141.
- [13] 李传亮, 王友仁, 罗 慧, 等. 全样本支持矢量数据描述模拟电路故障分类 [J]. 电工技术学报, 2012, 27 (8): 215-221.