

# 基于粒子群优化决策树的齿轮箱故障诊断\*

程 珩, 黄超勇, 张永刚

(太原理工大学机械电子工程研究所 太原, 030024)

**摘要** 针对现有支持向量机(support vector machines,简称SVM)在构造多类分类器的过程中存在计算费时、搜索率不高的问题,提出了一种新的SVM决策树设计算法。引入具有优良的全局搜索性能的粒子群算法,将其应用于优化决策树,构造出一种自适应性强、识别率高的多元分类器,实现SVM的有效多值分类。将其结果应用于齿轮箱故障诊断中,试验结果证明改进后的SVM构造方法的有效性和准确性。

**关键词** 粒子群; 决策树; 支持向量机; 故障诊断

**中图分类号** TH165.3

## 引 言

齿轮箱作为机械设备中一种必不可少的连接和传递动力的通用零部件,在现代工业发展中具有广泛的应用。因此,对其进行故障诊断,保证设备正常运行具有重要的实际意义<sup>[1]</sup>。故障辨识技术突破的关键在于特征的提取和分类器的设计。对于分类器设计部分,由于支持向量机在解决小样本、非线性、高维模式问题时表现出独特的优越性而得到广泛应用<sup>[2-3]</sup>,但支持向量机最初的两类分类模型已无法满足实际中遇到的多类故障。目前,提出的多分类处理方法主要有:经典的一对一分类(one-versus-one,简称OVO)、一对多分类(one-versus-rest,简称OVR)、有向非循环图(directed acyclic graph,简称DAG)支持向量机(DAG-SVM)和决策树(decision tree based,简称DT)支持向量机(DT-SVM)。后一种方法无论从训练速度还是分类精度都优于前者,但也存在计算费时、搜索率不高的问题。

笔者提出了一种新的SVM决策树设计算法。将粒子群算法引入决策树的优化,根据不同的故障信号构造最优决策树,使决策树的构造具有自适应性,提高了向量机对故障分类的识别率。试验结果表明,本研究方法比传统方法有较高的分类精度。

## 1 粒子群优化算法模型的建立

假设优化问题所有的潜在解都是搜索空间的粒子,所有粒子都有一个被优化的函数决定的适应值,每个粒子都有一个速度向量决定他们飞行的方向和距离<sup>[4-5]</sup>。这样粒子就追随当前的最优粒子在解空间中进行搜索,跟踪自身当前找到的最优解  $p_{best}$  和种群当前找到的最优解  $p_{gbest}$ ,逐代搜索,直到最后得到最优解。每个粒子根据下式更新自己的速度和位置。

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id} - x_{id}^k) + c_2 r_2 (p_{gd} - x_{id}^k) \quad (1)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (2)$$

其中:下标“ $d$ ”表示粒子的第 $d$ 维;“ $i$ ”表示粒子 $i$ ;“ $k$ ”表示第 $k$ 代; $c_1, c_2$ 为学习因子,也叫加速度常数; $r_1, r_2$ 为介于 $[0, 1]$ 之间的随机数<sup>[2]</sup>;  $\omega$ 为惯性权重函数,  $\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} k$ ;  $\omega_{max}$ 为初始权重;  $\omega_{min}$ 为最终权重;  $iter_{max}$ 为最大迭代次数。

式(1)右边由3部分组成,第1部分为“惯性”,反映了粒子的运动“习惯”,代表粒子有维持自己先前速度的趋势;第2部分为“认知”部分,反映了粒子对自身历史经验的记忆,代表粒子有向自身历史最佳位置逼近的趋势;第3部分为“社会”部分,反映了粒子间协同合作与知识共享的群体历史经验,代表粒子有向群体或邻域历史最佳位置逼近的趋势<sup>[6]</sup>。

\* 山西省自然科学基金资助项目(2011011026-3)  
收稿日期:2011-10-06;修改稿收到日期:2012-01-10

本研究的特点是把惯性权重设为一个线性减少的函数,使得粒子群算法在开始的时候倾向于“开掘”,然后逐渐转向于“开拓”,从而在局部区域内调制解,使粒子群算法性能得到很大提高<sup>[7]</sup>。粒子群优化(particle swarm optimization,简称PSO)算法是从随机解出发,通过迭代寻找最优解,这种算法具有实现容易、精度高和收敛快等优点。

## 2 PSO优化SVM决策树模型的建立

SVM决策树是一种解决多类问题的有效方法,它采用自上而下的划分算法或自下而上的凝聚算法生成决策树的层次。树中每个节点都是完成一个预定义的分类子任务的二值分类器,在分类阶段只需遍历部分分类器,因此分类效率和精度比较高。SVM决策树的分类性能与树的层次结构紧密相关,分类错误发生的地方离根结点越近,树的错分积累越大。为了获得更好的泛化性能,由可分性强的类为决策树的上层结点定义分类子任务。笔者引入PSO算法,对SVM决策树算法进行改进。

### 2.1 确定目标的适应度评价函数

评价函数是PSO算法与具体应用问题的惟一接口,是种群个体优劣的一种量化反映,它的构造直接影响到问题求解的质量与效率。因此,决策树的错误分类率是评价决策树的重要标准。设 $x_t = \{x_1, x_2, \dots, x_{N_t}\}$ 为到达决策树节点 $t$ 的训练样本集, $x_t = x_t^1 \cup x_t^2 \dots x_t^c$ , $c$ 为代表类的数目, $x_t^i$ 代表 $x_t$ 中属于 $w_i$ 的所有样本。定义 $x_t$ 的错误率为

$$i(x_t) = \sum_{i=1}^c \sum_{j \neq i} p(w_i | x_t) p(w_j | x_t) \quad (3)$$

$$(i = 1, 2, \dots, c; j = 1, 2, \dots, c)$$

其中: $p(w_i | x_t) = \frac{\# w_i}{\# x_t}$ ;"#"表示能够代表样本集的样本数目的大小;当 $x_t$ 仅包含来自同一类的样本时, $i(x_t) = 0$ 。

设节点 $t$ 的权值矢量为 $w$ ,将 $x_t$ 分为左子树 $x_{tL}(w)$ 和右子树 $x_{tR}(w)$ 两个部分

$$x_{tL}(w) = \{x | x \in x_t, w^T < 0\} \quad (4)$$

$$x_{tR}(w) = \{x | x \in x_t, w^T > 0\} \quad (5)$$

取分割后的错误率 $i(x_t, w)$ 为节点 $x_{tL}(w)$ 和 $x_{tR}(w)$ 错误率的加权平均值。

样本分割后的错误率为

$$i'(x_t, w) = p(x_{tL}(w) | x_t) i(x_{tL}(w)) + p(x_{tR}(w) | x_t) i(x_{tR}(w))$$

$$\text{其中: } p(w_{iL} | x_t) = \frac{\# x_{iL}}{\# x_t}; p(w_{iR} | x_t) = \frac{\# x_{iR}}{\# x_t}。$$

错误率的减少量为 $\Delta i(x_t, w) = i(x_t, w) - i'(x_t, w)$ ,它表示节点 $t$ 的样本分割前、后的错误率减少量为 $\Delta i(x_t, w)$ 。选择 $f(y) = \Delta i(x_t, w)$ 作为评价函数可以使错误率的减少量最大,决策树的分类精度更高。

### 2.2 最优决策树的生成

最优决策树指适应函数值最小的决策树。在粒子运动演化过程中,最后一代群体中并不一定包含最优的决策树,因此要将在整个过程中最好的决策树保存起来,每当产生下一代,则将之与保存的做比较,淘汰相对较差的,这样直到进化结束,保存的决策树基本上就是最好的。图1为基于PSO优化决策树建模流程图。

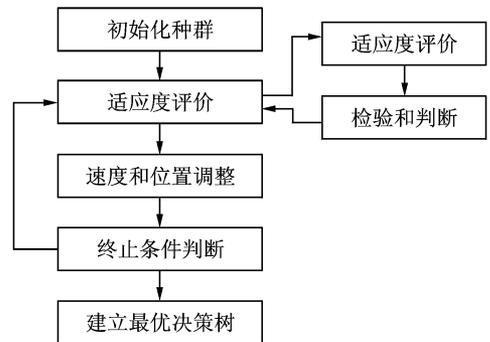


图1 PSO优化决策树建模流程图

### 2.3 几种类别策略的对比

支持向量机在对于故障数据的分类时,传统支持向量机的训练OVO与DAG采用“一对一”的方法,需要训练的SVM最多,且随类别数的增加成平方指数增加;OVR采用“一对余”的方法,训练的复杂程度大大减少;DT-SVM和PDT-SVM由于采用了树结构,训练的复杂程度最小,可以有效减少训练的复杂度,提高分类效率。

## 3 试验验证

试验故障信号采集系统的组成如图2所示,系统由变频器、电机、齿轮箱、4个加速度传感器、耦合放大器、16通道动态信号分析仪和导线组成。本试验分别采集轴承内、外圈故障、齿轮齿面磨损和轮齿

表1 齿轮箱振动信号特征参量表

| 故障模式 | 采样频率/kHz | 峰值 $X_1$  | 峭度指标 $X_2$ | 裕度指标 $X_3$ | 偏态指标 $X_4$ | 波形指标 $X_5$ | 频谱质心 $X_6$ | 相关因子 $X_7$ |
|------|----------|-----------|------------|------------|------------|------------|------------|------------|
| 正常   | 2        | 0.078 125 | 0.139 726  | 0.216 339  | 0.245 816  | 0.176 027  | 0.028 755  | 0.879 836  |
|      |          | 0.088 542 | 0.123 769  | 0.094 873  | 0.529 369  | 0.153 268  | 0.067 664  | 0.752 438  |
|      |          | 0.072 315 | 0.061 758  | 0.074 636  | 0.000 000  | 0.133 723  | 0.155 013  | 0.566 593  |
|      |          | 0.082 138 | 0.116 713  | 0.094 873  | 0.529 369  | 0.153 268  | 0.067 664  | 0.752 438  |
| 内圈故障 | 2        | 0.380 208 | 0.188 926  | 0.363 164  | 0.319 247  | 0.293 462  | 0.157 204  | 0.456 792  |
|      |          | 0.270 833 | 0.141 669  | 0.104 706  | 0.513 807  | 0.218 264  | 0.157 865  | 0.512 942  |
|      |          | 0.239 583 | 0.151 229  | 0.141 953  | 0.538 466  | 0.228 753  | 0.246 477  | 0.406 768  |
|      |          | 0.406 252 | 0.092 834  | 0.108 911  | 0.653 317  | 0.117 216  | 0.256 157  | 0.369 456  |
| 外圈故障 | 2        | 0.192 708 | 0.107 727  | 0.288 272  | 0.078 846  | 0.194 676  | 0.104 676  | 0.423 811  |
|      |          | 0.088 542 | 0.000 000  | 0.021 763  | 0.259 672  | 0.000 000  | 0.064 462  | 0.558 399  |
|      |          | 0.000 000 | 0.052 845  | 0.060 008  | 0.192 789  | 0.100 037  | 0.000 000  | 0.578 923  |
|      |          | 0.072917  | 0.020947   | 0.000000   | 0.158931   | 0.060776   | 0.088641   | 0.439346   |
| 齿轮磨损 | 2        | 0.354 167 | 0.179 249  | 0.320 684  | 0.270 484  | 0.265 633  | 0.186 755  | 0.836 583  |
|      |          | 0.192 708 | 0.071 498  | 0.135 377  | 0.464 192  | 0.091 137  | 0.116 541  | 1.000 000  |
|      |          | 0.432 292 | 0.132 016  | 0.247 055  | 0.252 905  | 0.217 852  | 0.359 358  | 0.914 079  |
|      |          | 0.282 125 | 0.035 932  | 0.074 349  | 0.482 486  | 0.015 642  | 0.178 913  | 0.723 688  |
| 轮齿折断 | 2        | 0.406 225 | 0.098 234  | 0.108 911  | 0.653 317  | 0.117 216  | 0.256 157  | 0.323 811  |
|      |          | 0.984 375 | 1.000 000  | 0.924 444  | 1.000 000  | 1.000 000  | 0.285 597  | 0.458 399  |
|      |          | 1.000 000 | 0.622 266  | 1.000 000  | 0.911 888  | 0.518 161  | 0.190 852  | 0.243 734  |
|      |          | 0.568 125 | 0.223 941  | 0.298 016  | 0.120 366  | 0.337 017  | 1.000 000  | 0.539 346  |

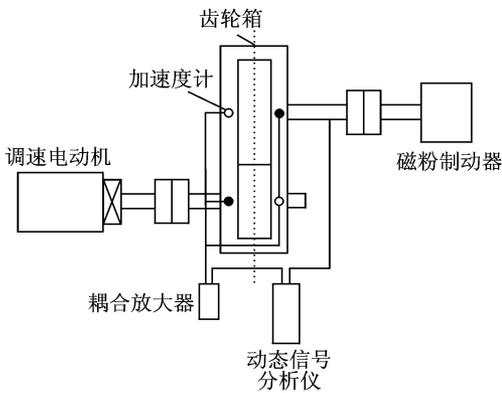


图2 齿轮箱故障信号采集系统示意图

折断等5种典型故障信号和正常信号。采样频率为2 kHz,采样点数为1 024点。齿轮箱振动信号的各种时、频域特征参量对故障信息的表达能力各有侧重,只有进行综合分析才能得到比较完整的故障信息<sup>[8-9]</sup>。所建诊断模型选用5个时特征参量:峰值、峭度指标、裕度指标、偏态指标和波形指标。2个频域特征参量为频谱质心和相关因子。由于这些特征参量存在量纲差别,因此在建模之前必须进行归一化处理。表1给出了20组特征参量(已归一化),其中10组用于训练,另外10组用于测试结果。

在优化过程中,粒子群优化结构参数的设置为:

粒子群初始化种群个数为20;搜索空间维数为4;最大迭代次数 $k=200$ ;  $c_1, c_2$  为2.0;  $w = w_{max} - (w_{max} - w_{min}) \times k / iter_{max}$ ;  $w_{max}$  取为1;  $w_{min}$  取为0.3。利用粒子群优化决策树,训练过程如图3所示。

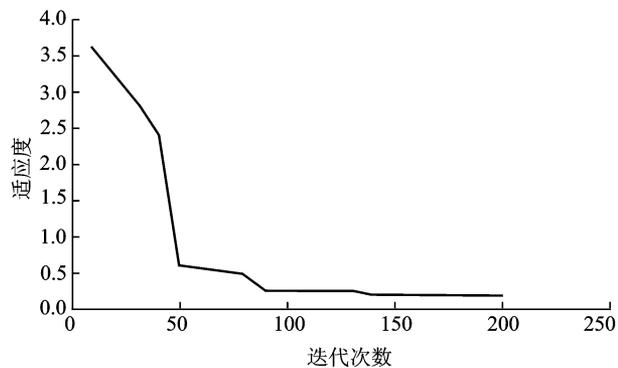


图3 粒子群优化决策树过程图

生成决策树之后,选用的支持向量机核函数为径向基核函数,惩罚因子 $c$ 取1,径向基参数 $\sigma$ 取1。对比传统的OVO,OVR,DAG和笔者提出的基于PSO优化决策树的分类策略的支持向量机(PSO-decision tree,简称PDT),分类结果如表2所示。可以看出,OVO方法和OVR虽然较为简单,但是识别效率不高,主要原因为投票策略采用的是概率统计的方法,即通过式(6)得到后验概率(最终得票概率)

来获得每类投票的最终得分值<sup>[10]</sup>。

$$P(i|x) = \sum_{j=1, j \neq i}^N \{P_{i,j}(i|x)P(j)\} \quad (i = 1, 2, \dots, N) \quad (6)$$

表2 支持向量机分类构造分类精度比较

| 分类策略  | OVR | OVO | DT | DAG | PDT |
|-------|-----|-----|----|-----|-----|
| 正确率/% | 87  | 85  | 89 | 88  | 99  |

当得票相同时存在不可分区域。当分类较多、分类数据量较大时，这一问题显得较为突出。采用PDT的分类方法使支持向量机能够根据实际故障主动建立相应的决策树，优化后的支持向量机对于故障的识别效果大大增强，有效地对故障进行诊断。

## 4 结束语

通过试验分析，得出传统的分类策略对同一样本的分类准确率有一定的缺陷，而笔者提出的利用粒子群优化算法优化支持向量机的决策树方法可有效地对故障进行精确诊断，为机械故障诊断提供了一种准确、有效的方法。

### 参 考 文 献

- [1] 熊诗波, 黄长艺. 机械工程测试基础[M]. 北京: 机械工业出版社, 2006: 188-211.
- [2] Olivier C, Vladimir V, Olivier B, et al. Choosing multiple parameter for support vector machines[J]. Machine Learning, 2002, 46: 131-159.
- [3] 杜京义, 侯媛彬. 基于核方法的故障诊断理论及其方法的研究[M]. 北京: 北京大学出版社, 2010: 34-38.
- [4] 李旭渊, 许化龙, 许哲. 基于改进粒子群优化算法的发动机故障仿真[J]. 火力指挥与控制, 2010, 35(4): 85-87.

Li Xuyuan, Xu Hualong, Xu Zhe. Fault simulation of engine based on improved particle swarm optimization[J]. Fire Control & Command Control, 2010, 35(4): 85-87. (in Chinese)

- [5] Cheong S M, Oh S H, Lee S Y. Support vector machines with binary tree architecture for multi-class classification[J]. Neural Information Processing-Letters and Reviews, 2004, 2(3): 47-51.
- [6] 纪震, 廖惠连, 吴青华. 粒子群算法及其应用[M]. 北京: 科学出版社, 2009: 16-70.
- [7] 丁华, 王秀坤, 孙焱. 基于PSO改进决策树算法的研究[J]. 小型微型计算机系统, 2005, 26(7): 1206-1210. Ding Hua, Wang Xiukun, Sun Tao. Research on decision tree method based on improved PSO[J]. Mini-micro Systems, 2005, 26(7): 1206-1210. (in Chinese)
- [8] 钟秉林, 黄仁. 机械故障诊断学[M]. 北京: 机械工业出版社, 2002: 320-400.
- [9] 丁康, 李巍华, 朱小勇. 齿轮及齿轮箱故障诊断实用技术[M]. 北京: 机械工业出版社, 2005: 10-22.
- [10] 王晓红. 基于概率投票策略的多类支持向量机及应用[J]. 计算机工程, 2009, 35(2): 180-183. Wang Xiaohong. Multi-class SVMs based on probability voting strategy and its application[J]. Computer Engineering, 2009, 35(2): 180-183. (in Chinese)



第一作者简介: 程珩, 女, 1956年11月生, 教授。主要研究方向为机械测试、故障诊断和信号处理。曾发表《工作模式分析在减速器监测和诊断中的应用》《《振动、测试与诊断》2010年第2期》等论文。E-mail: chenghang@tyut.edu.cn