计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(3):18−26 [doi: 10.15888/j.cnki.csa.006285]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于支持向量机的健康状态评估方法①

张　春, 舒　敏

(北京交通大学 高速铁路网络管理教育部工程研究中心, 北京 100044)
(北京交通大学 计算机与信息技术学院, 北京 100044)
通讯作者: 舒　敏, E-mail: 15120435@bjtu.edu.cn

摘　要: 随着高速铁路动车组的快速发展和应用, 其安全性和可靠性引起了广泛关注. 为了准确判别高速铁路动车组轴箱轴承 (以下简称轴箱轴承) 的健康状态情况, 提出通过采集轴箱轴承温度及在相同和不同转向架驱动侧、非驱动侧各个部件的温度数据, 利用主成分分析法 (PCA) 进行特征降维, 将基于决策树的支持向量机 (DT-SVM) 多分类算法作为判别算法, 同时结合层次分析法 (AHP) 确定向量值权重, 从而进一步提高分类精度. 大量实验表明该方法可使分类准确率提升 5% 左右, 此外通过建立健康状态评估模型,将轴箱轴承健康状态分为健康、温升、强温和激温四类, 有助于提高轴箱轴承健康状态的判别力和运维决策的准确性.

关键词: 轴箱轴承; 支持向量机; 主成分分析法; 层次分析法; 健康状态评估; 多分类方法

引用格式: 张春,舒敏.基于支持向量机的健康状态评估方法.计算机系统应用,2018,27(3):18–26. http://www.c-s-a.org.cn/1003-3254/6285.html

Health Assessment Method Based on Support Vector Machine

ZHANG Chun, SHU Min

(Engineering Research Center of Network Management Technology for High Speed Railway Ministry of Education, Beijing Jiaotong University, Beijing 100044, China)
(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the rapid development and application of high-speed railway EMU, its safety and reliability have attracted wide attention. In order to estimate the health status of axle box bearing of high-speed railway EMU accurately (hereinafter referred to as axle box bearing), this study proposes a classification algorithm based on Decision Tree and Support Vector Machine, and utilizes Principal Component Analysis (PCA) to reduce the feature dimension simultaneously. In addition, the performance of the classification can be further improved by collecting the temperature data of axle box bearing and various components on the drive side and non-drive side of which either in the same bogie or in different bogie. And the Analytic Hierarchy Process (AHP) has been used to distribute the weight of vectors. Extensive experiments demonstrate the effectiveness of the classification model, and the accuracy has been increased by about 5%. Furthermore, the judgment ability of health status of axle box bearing and the precision of the operation and maintenance policy can be enhanced if we establish a health assessment model through dividing the health status of axle box bearing into four parts including health, temperature rising, strong temperature, and irritative temperature.

Key words: axle box bearing; SVM; PCA; AHP; health status assessment; multi-class methods

近年来, 高速列车 (HSTS) 的速度已经提升到 300 多公里/小时[1], 根据国际铁路联盟的报告, 截至 2015 年 4 月, 整个世界运行的高速铁路数量已经达到了 3603 辆, 使得列车的安全性和可靠性备受关注[2]. 轴

① 基金项目: 国家“八六三”高技术研究发展计划基金 (2015AA043701); 中国铁路总公司科技研究开发计划课题 (2015J006-C)
收稿时间: 2017-07-04; 修改时间: 2017-07-20; 采用时间: 2017-08-07; csa 在线出版时间: 2018-01-25

© 中国科学院软件研究所 http://www.c-s-a.org.cn

箱轴承作为高速铁路动车组的关键部件之一, 有着举足轻重的位置[3], 其主要作用是限制相对运动和减少机器旋转部件之间的摩擦力. 在高速铁路动车组中, 每对轮轴都需要配备一个轴箱轴承, 由于高应力和载荷的影响, 轴箱轴承会磨损、破坏乃至失效, 这将会不同程度地影响列车的安全性与经济适用性, 因此, 许多科学家和企业致力于研究轴箱轴承的故障监测与剩余寿命预测. 通过比较在正常情况和故障情况下运行轴箱轴承的温度范围, 发现轴箱轴承在不同的健康状态下, 由于载荷以及径向力的不同, 导致摩擦力不同, 从而伴随产生的热量不一致的现象. 因此, 使用基于机器学习的判别轴箱轴承的不同健康状态的分类方法应运而生, 其根据温度范围做出健康状态评估, 指导运维决策. 基于机器学习的分类方法主要包括三个步骤: 特征提取、特征选择和特征分类. 常被提取的特征可以是统计特征[4], 自回归滑动平均 (ARMA) 特征[5], 直方图特征[5]或小波特征[6], 常用的特征选取技术包括主成分分析 (PCA)[7], 遗传算法 (GA) 和决策树 (DT). 在本文的研究中使用了统计特征和基于决策树方法的特征选取技术.

本文提出了一种以支持向量机 (SVM) 为算法基础, 基于决策树与层次分析法进行改进的健康状态评估的方法, . 支持向量机[8]是一种以数据为驱动, 有效进行故障预测与诊断的机器学习方法, 目前已经成功地应用在各大领域和系统中, 包括智能电网[8]、汽车液压制动系统[9]、汽油发动机气门[10]、轴承[12,13]等. 支持向量既可以克服神经网络等方法所固有的过拟合和欠拟合问题, 又可以在样本稀疏的条件下实现较高的分类效果. 此外, 其通过寻求最小化结构化风险来提高模型的泛化性能, 实现经验风险和置信范围的最小化[14].

目前基于支持向量机的多分类方法主要有一对一法 (One-Against-One SVM, OAO-SVM)、一对多法 (One-Against-All SVM, OAA-SVM)、基于有向无环图法 (Decision Directed Acyclic Graph SVM, DDAG-SVM) 和决策树方法 (Decision Tree, SVMDT-SVM)[15]. OAA-SVM 与 OAO-SVM 的分类准确率比较高, 但是前者需要训练 N 个决策树来遍历整个数据集, 后者需要训练 $N(N-1)/2$ 个决策面, 所需时间均较长; DDAV-SVM 需要训练 $N(N-1)/2$ 个决策面, DT-SVM 需要训练 $N-1$ 个决策面, 由于这两种方法的训练数据结构为树形, 其遍历过程自顶向下逐层递减, 所以训练速度比前两者有大大的提高. 相对 DDAV-SVM 需要遍历 $N-1$ 个决策面, DT-SVM 方法在识别阶段只需要遍历 $\log_2 N$ 个决策面, 时间优势更加明显. 另外决策树很容易理解和解释, 还可以通过更紧凑的结合形成影响图, 增强对事件与关系的关注度, 同时在部分数据缺失的情况下, 也可以达到较好的分类效果. 综合考虑轴箱轴承故障样本数据的稀缺性、各类样本分布的不规律性以及各特征权重不一致的特点, 本文优先选取决策树算法.

1 相关工作

1.1 支持向量机理论基础

指导思想: 寻找一个分类超平面, 将两类样本分别划分到超平面两侧, 并且使得每类的样本与分类超平面的距离达到最大. 其中, 每个类别中与分类超平面最近的点被称为支持向量. 并且, 每个类别的支持向量到分类超平面的距离相等的时候才能达到最优分类超平面. 针对二分类问题, 若问题线性可分, 则找到其最优分类面将训练样本完全分开 (即使得每一类数据与超平面距离最近的向量与超平面之间的距离最大), 如图 1 所示; 若问题线性不可分, 则通过使用核函数将特征向量从低维空间映射到高维空间使其线性可分, 如图 2 所示.

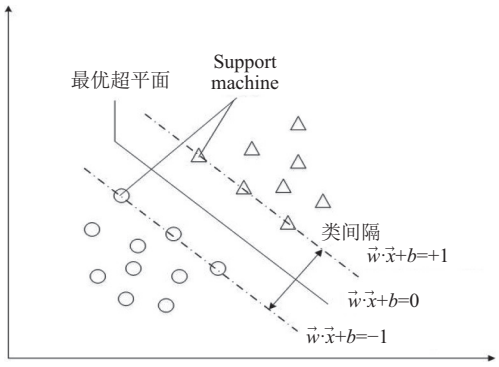


图 1 线性可分情况

计算模型: 给定训练样本集 $\{x[i], y[i]\}$, $i=1, 2, 3, \cdots, n$, y, 其中, $y=+1$ 为类别 1, $y=-1$ 为类别 2.

假设 n 维空间的分类超平面是 $w^{\mathrm{T}}x+b=0$, 则任意一点 (x, y) 到这个超平面的距离为 $\frac{|w^{\mathrm{T}}x+b|}{\|w\|_2}$, 而 SVM 的目标就是寻找可以正确区分所有样本的 w, b, 并使得对于任意一个支持向量 (x, y) 有:

$$\frac{\max|w^{\mathrm{T}}x+b|}{\|w\|_2} \tag{1}$$

© 中国科学院软件研究所 http://www.c-s-a.org.cn

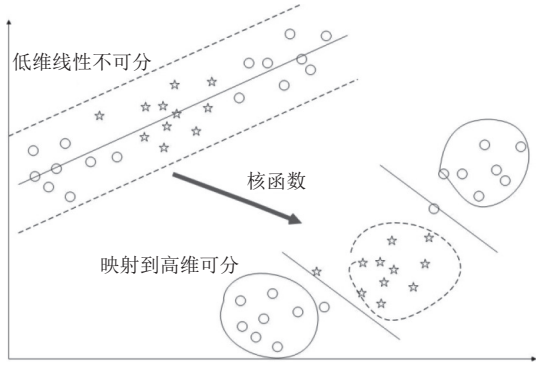


图 2 非线性可分情况

由于同一个分类超平面 w, b 可以成比例的放缩. 因此总可以经过适当的放缩找到合适的 w, b 使得支持向量处的$w^{\mathrm{T}}x+b$值为 1 或–1. 所以这时目标公式 (1) 就简化为:

$$max\frac{1}{\|w\|_2} \tag{2}$$

而此时的约束条件可以表达为:

$$y_i\left(w^{\mathrm{T}}x_i+b\right)\geq 1, i=1,2,\cdots,m \tag{3}$$

在实际应用过程中一般采用的公式 (2) 的另一种等价凸函数形式, 见下述公式 (4), 因此问题转化为下述公式 (4) 和公式 (5).

$$min\frac{1}{2}\|w\|^2 \tag{4}$$

$$y_i\left(w^{\mathrm{T}}x_i+b\right)\geq 1, i=1,2,\cdots,m \tag{5}$$

将公式 (4) 和公式 (5) 转换为标准形式公式 (6) 和公式 (7),

$$minf(w,b)=\frac{1}{2}\|w\|^2 \tag{6}$$

$$g_i(w,b)=1-y_i\left(w^{\mathrm{T}}x_i+b\right)\leq 0, i=1,2,\cdots,m \tag{7}$$

然后根据对偶理论, 可以通过求解该问题的对偶问题得到最优解, 其对应的对偶问题为公式 (8):

$$\max\theta_D(a)=\sum_{i=1}^{m}a_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}a_ia_jy_iy_j\left\langle x_i,x_j\right\rangle \tag{8}$$

其中, $\theta_D(a)=min_{w,b}L(w,b,a)=max_{a_i\geq 0}min_{w,b}$, 根据对偶问题的解 α, 求得 w, b, 从而得到最优分类面.

$$w=\sum_{i=1}^{m}a_iy_ix_i$$

$$b=-\frac{1}{2}(max_{y_i=-1}w^{\mathrm{T}}x_i+min_{y_i=1}w^{\mathrm{T}}x_i)$$

1.2 决策树理论基础

决策树是一种用于对实例进行分类的树形结构. 决策树由节点 (node) 和有向边 (directed edge) 组成. 节点的类型有两种: 内部节点和叶子节点. 其中, 内部节点表示一个特征或属性的测试条件 (用于分开具有不同特性的记录), 叶子节点表示一个分类. 一旦我们构造了一个决策树模型, 以它为基础来进行分类将是非常容易的. 具体做法是, 从根节点开始, 对实例的某一特征进行测试, 根据测试结构将实例分配到其子节点 (也就是选择适当的分支); 沿着该分支可能达到叶子节点或者到达另一个内部节点时, 那么就使用新的测试条件递归执行下去, 直到抵达一个叶子节点. 当到达叶子节点时, 我们便得到了最终的分类结果.

如图 3 所示为决策树的一个实例, 仅仅用两个 feature 就可以对数据集 (表 1) 中的 5 个记录实现了准确的分类.

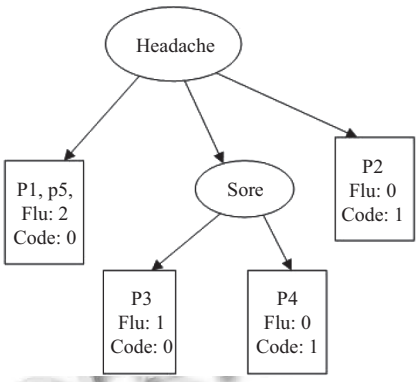


图 3 决策树实例

表 1 数据集

Patient	Headache	Cough	Temperature	Sore	Diagnosis
P1	Severe	Mild	High	Yes	Flu
P2	Mild	severe	Normal	Yes	Cold
P3	Mild	Mild	Normal	Yes	Flu
P4	Mild	No	Normal	No	Cold
P5	Severe	Sever	Normal	Yes	Flu

得到规则如下:

Rule 1: if(headache=servere) then it is Flu

Rule 2: if(headache=mild) and (Sore=yes) then it is Flu

Rule 3: if(headache=mild) and (Sore=no) then it is Cold

Rule 4: if(headache=no) then it is Cold

1.3 AHP 算法

层次分析法是一种定性与定量相结合的基于运筹学理论的层次权重决策分析方法[16], 其主要特点是用

© 中国科学院软件研究所 http://www.c-s-a.org.cn

两两重要性程度表示出两个方案的相应 1~9 级重要性程度等级 (如表 2 所示), 以此判断各个因素的相对重要性.

表 2 重要性等级标度表

标度值	代表含义(因素x与因素y比较)
1	因素x与因素y同等重要
3	因素x比因素y稍微重要
5	因素x比因素y重要
7	因素x比因素y强烈重要
9	因素x比因素y极端重要
2,4,6,8	因素x与因素y重要程度居于以上两者之间
1/3	因素x与因素稍微次要
1/5	因素x比因素y次要
1/7	因素x比因素y强烈次要
1/9	因素x比因素y极端次要
1/2,1/4,1/6,1/8	因素x与因素次要程度居于以上两者之间

层次分析法按照上述标准将复杂系统简单化, 划分为目标层、准则层和指标层, 对每层构建判断矩阵, 通过比较和计算获得不同的权重, 来对每一层进行决策, 同时有效利用专家经验的干预, 得到更加合理的方案.

1.4 基于决策树的 SVM 算法

基于决策树的 SVM 分类方法的基本思想是将一个 N 类可分的问题分解为若干个二分类问题. 首先将数据集划分为两个子集, 然后对两个子集继续划分, 依次类推, 直至最后的子集不能被划分为止, 在这个过程中会动态生成一棵二叉树.

DT-SVM 生成的决策树有两种不同的方式[17]:

1) 偏态树. 根节点表示所有类元素的集合, 每一次 SVM 都通过二分类法单独剥离出一个类别作为叶子节点, 其他类别继续作为子树根节点, 循环直到所有类别都为叶子节点结束. 如图 4 所示.

2) 正态树. 根节点是所有类元素集合, 首先 SVM 将所有类元素分成两部分, 然后再对这两部分进行二分类, 直到所有类别均变成子节点结束, 如图 5 所示.

虽然 DT-SVM 对于 N 分类问题只需要构造 N–1 个分类器, 具有较高的训练速度和分类速度, 但是 DT-SVM 存在错分类现象, 即分类错误越靠近树根的地方, 错误延续的节点数据越多, 其分类性能越差. 所以, 在构造决策树的过程中必须秉持着先易类后难类的原则, 先将比较容易分开的类放在根节点, 逐层类推直至叶子结点. 因此如何确定决策树的结构是个关键问题.

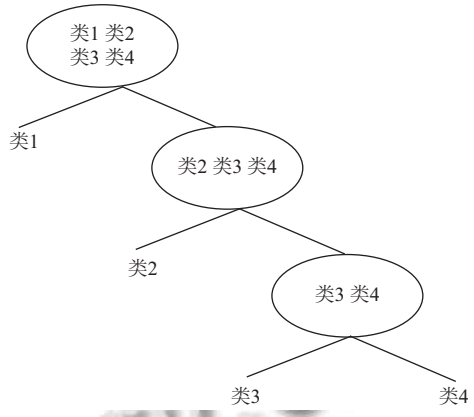


图 4 DT-SVM 偏态树

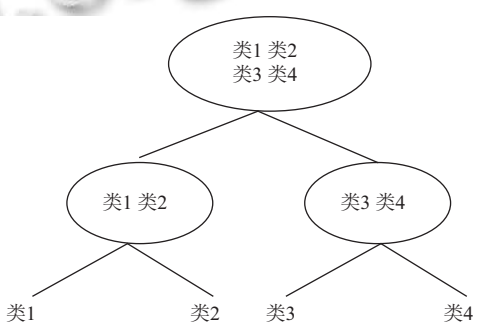


图 5 DT-SVM 正态树

另一方面, 由于高速铁路动车组的安全可靠性, 基于故障状态的现场数据特别少, 导致数据不平衡的问题成为一大挑战, 大类的数据远远超过了小类的数据, 故障样本数据的稀缺导致分类精度存在偏差、不同样本数据映射到高维空间的分布情况不同、分类过程中特征值所占权重不同. 另外二叉树的遍历都是从根节点开始, 带有一定的盲目性, 会导致资源的浪费. 因此本文提出对基于决策树的支持向量机方法进行改进, 平衡掉不同各类别数据量差异过大问题, 弥补上述不足.

2 改进的 AHP-DT-SVM 算法

DT-SVM 构造的分类二叉树结构不统一, 并且不同的二叉树结构对分类的结果和精度影响较大, 本文对 DT-SVM 算法进行优化, 结合层次分析法确定特征值的权重, 让权重高的特征值位于二叉树上层, 保证尽可能高效的首先识别. 基于此的 DT-SVM 每次都会得到较好的分类二叉树, 避免随机性造成的结果不准确与精度不够高的问题.

轴箱轴承在使用过程中, 会受到载荷、润滑、使用时间、维护措施以及疲劳扩展的随机性等多种因素

© 中国科学院软件研究所 http://www.c-s-a.org.cn

的影响, 而轴箱轴承的损坏主要是由摩擦产生, 其次是轴箱径向受力和轴向受力. 根据速度、表面破坏程度等影响因素的不同, 轴箱轴承间产生的摩擦力度会有所差异, 进而导致产生的热量不同, 轴箱轴承产生的温度高低也就会有所差别. 由于金属的导热性, 跟轴箱有关的零部件也会接受热量产生温升. 此外, 虽然同一转向架不同轴箱运行环境一致, 但由于故障的产生, 导致各自当前温度可能有差异. 因此可通过对比温度范围来进一步确定轴箱健康状态. 另外, 考虑到不同的特征值具有不同的影响程度, 基于层次分析法的 AHP-DT-SVM 算法对轴箱轴承进行健康状态分类的过程如图 6 所示.

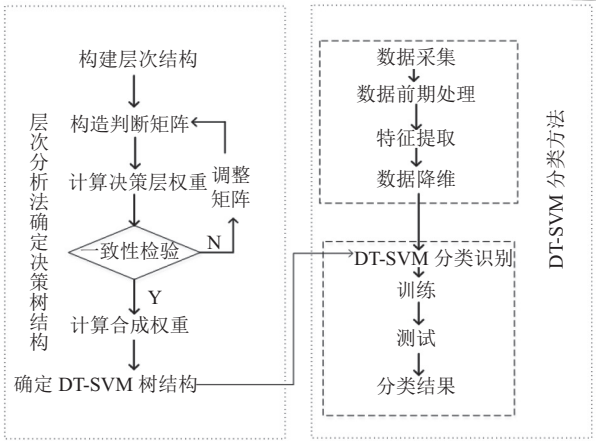


图 6 基于 AHP-SVM 的健康状态评估模型

1) 建立层次分析结构, 在深入分析待研究问题的基础上, 将分析指标划分为不同的层次, 建立多层次评价模型.

2) 构造判断矩阵, 设置评价模型的重要性比较评价指标集 $V=\{v_1, v_2,\cdots, v_n\}$, v_i 为第 i 个需要进行比较的指标, 对同一层次的各因子关于上一层次某一准则的重要性两两比较, 然后构建比较判断矩阵.

3) 计算权重, 采用算数平均法对判断矩阵进行计算, 求得权重值, 同时计算一致性比率 CR, 反复调整矩阵, 直到 CR<0.1 停止.

4) 根据权重中确定 DT-SVM 决策树的结构, 同时计算类间分离性测度来对类进行不断地分离与合并, 每次取最后一个合并的类与其他类做正负样本训练分类器, 并将其作为根节点, 依次类推, 保证容易分的类先分离.

训练 SVM 分类器的算法步骤如下所述.

假设类别数为 N, 训练样本集 X 由类 X_i, i=1, 2,⋯, N, j=2, 3,⋯, N, 其中, $i<j$.

DT-SVM 过程具体如下.

1) 利用基于类分布的类间分离性测度计算公式, 计算各类间的分离性测度$\delta_{ij}, i=1,2,\cdots,N, j=2,\cdots,N$, 其中, $i<j$.

$$\delta_{ij}=\frac{d_{ij}}{\delta_i+\delta_j}$$

其中, $d_{ij}=\|c_i-c_j\|$,表示类 i 与类 j 之间的中心距离, c_i 于 c_j 是根据训练样本计算出来的类中心.

$$c_i=\frac{1}{n_i}\sum_{x\in X}x, i=1,2,\cdots,k$$

n_i 为类 X 的样本个数, K 为类别数, 式中, δ_{ij} 为类方差, 表示样本分布:

$$\delta_i=\frac{1}{n_i-1}\sum_{x\in X}\|x_i-c_i\|, i=1,2,\cdots,k$$

若 δ_{ij}, 则类 i 与类 j 间无交集, 反之, 有交集. δ_{ij} 越大, 则表明两个类之间的分离性更加明显.

2) 筛选出分离测度最小的两类合并为一个类别, 计算该类的类中心与方差, 此时类别数据由 N 变为 N–1.

3) 计算 2) 中得到的类与其余类的分离性测度, 选择; 测度最小的类与之合并, 计算类中心与方差, 类别数目继续减 1.

4) 当类别数据>2, 继续执行步骤 3), 否则, 算法结束.

3 基于 AHP-DT-SVM 的健康状态评估过程

为了更清楚地说明此算法的关键步骤和和原理, 下文结合轴箱轴承实际的健康状态评估过程来对此优化算法以及建模过程进行解释说明.

3.1 数据处理

1) 数据预处理-去噪与归一化

步骤1. 利用聚类方法进行去噪处理

步骤2. 利用Z-score标准化方法进行归一化处理. 该方法给予原始数据的均值和标准差进行数据的标准化, 经过处理的数据符合标准正态分布, 即均值为0, 标准差为1, 转化函数为: $x^*=\frac{x-\mu}{\sigma}$其中μ为所有样本数据的均值, σ为所有样本数据的标准差.

步骤3. 对轴箱轴承温度进行回归拟合, 在确定拟合线的情况下, 框定上下温度界限, 数据包含率达到95%即可, 超出范围的数据可判定为噪声.

2) 主成分分析法降维

可提取的有价值的轴箱轴承的特征值如表 3 所示.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

表 3 可提取和降维后的特征值表

可提取的特征值	降维后的特征值
驱动侧温度	轴箱轴承温度
同车驱动侧温度	轴箱轴承温度均值
整列驱动侧均值	驱动侧温度均值
同车非驱动测温度	非驱动测温度均值
同车非驱动测均值	同轴轴端温度
整列非驱动测均值	大齿轮温度
同组拖车温度均值	小齿轮箱温度
整列拖车温度均值	
同车大齿轮温度	
同车小齿轮温度	
同车定子温度	
同车转子温度	
转向架同侧轴端温度	

考虑到以上特征有重复性并且有些特征对排序作用效果不明显, 本文采用主成分分析法对特征向量做进一步的特征降维.

主成分分析法 (Principal Component Analysis, PCA) 是目前应用最为广泛的一种数据驱动方法[17], 其可以从多元事物中解析出主要影响因素, 揭示事物的本质, 简化复杂问题, 从而实现对高维数据的线性降维. 本质是求得这个投影矩阵, 用高维的特征乘以这个投影矩阵, 便可以将高维特征的维数下降到指定的维数. PCA 的目标是寻找 $r(r<n)$ 个新变量, 这 r 个新变量称为“主成分”, 它们可以在很大程度上反映原来 n 个变量的影响, 并且这些新变量是互不相关的, 也是正交的, 并且能反映事物的主要特征, 压缩原有数据矩阵的规模. 那么, 如何确定新变量的个数 R 是主成分分析的关键, 我们需要进一步分析每个主元素对信息的贡献, 通过计算每一个特征值对于降低 entropy(熵) 的贡献来进行排序, 选择排序较高的留下来, 去掉排序较低的特征值, 从而达到降维的目的.

PCA 算法步骤如下所述, 这里假设有 w 条 v 维数据需要进行降维处理. 降维后的特征值如表 3 所示.

1. 将原始数据按列组成v行w列矩阵R
2. 将R的每一行(代表一个属性字段)进行零均值化, 即减去这一行的均值
3. 求出协方差矩阵$C=\frac{1}{w}RR^{\mathrm{T}}$,
4. 求出协方差矩阵的特征值及对应的特征向量
5. 将特征向量按对应特征值大小从上到下按行排列成矩阵, 取前i行组成矩阵P
6. $Y=PR$,即为降维到i维后的数据

3.2 构造轴承轴承健康状态类型状态集

本文首先根据赵佳颖和马千里老师提供的温度分类标准[18,19], 确定轴箱轴承的 4 个健康状态, 包括{A 正常, B 温升, C 强温, D 激温}4 类, 相应的状态标识函数值为 $f(s)$={health, subhealth, deterioration, failure}; 然后利用支持向量机算法对数据进行分类测试, 根据准确率对分类界限进行调整, 确定最终健康状态分类判别标准, 构建健康状态模型.

3.3 层次结构评估指标体系构建

整个评估系统可以分为目标层、准则层和方案层, 具体架构如图 7 所示.

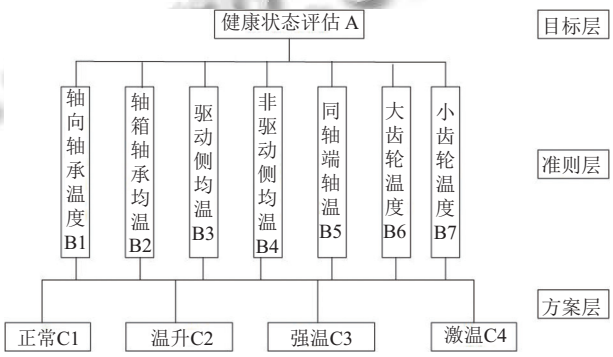


图 7 轴箱轴承健康状态评估的层次结构

3.4 评估指标权重计算与一致性检验

1) 评估指标权重计算

AHP 确定评估指标的相对重要程度计算公式为:

$$R_{ij}=\begin{cases}1,\tilde{R}_{ij}<1\\ \left[\dfrac{v_i}{v_j}\right],1\le\ \tilde{R}_{ij}\le 9,\\ 9,\tilde{R}_{ij}>9\end{cases}\tag{9}$$

式中, V_i 和 V_j 都是影响程度值[20].

层次分析法有四种方法计算权重: 几何平均法、算术平均法、特征向量法和最小二乘法. 本文主要利用算数平均法计算权重. 计算公式如下:

$$W_i=\frac{1}{n}\sum\nolimits_{j=1}^{n}\frac{R_{ij}}{\sum_{k=1}^{n}R_{kj}},i=1,2,\cdots,n\tag{10}$$

根据重要性等级标度表以及公式 (9) 和 (10), 可得轴箱轴承健康状态评估准则层的判断矩阵和权重如表 4 所示.

2) 一致性检验

为了避免主观影响, 尽量实现客观化描述, 对矩阵的一致性检验必不可缺. 只有当一致性比例 $CR<0.1$ 时, 权重矩阵才可以接受, 否则, 需要进行适当修改, 降低主观因素影响.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

表 4 准则层判断矩阵

R_{ij}	V_1	V_2	V_3	V_4	V_5	V_6	V_7
V_1	1	3	5	7	9	9	9
V_2	1/3	1	3	5	5	5	5
V_3	1/5	1/3	1	2	3	3	3
V_4	1/7	1/5	1/2	1	2	2	2
V_5	1/9	1/5	1/3	1/2	1	2	2
V_6	1/9	1/5	1/3	1/2	1/2	1	1
V_7	1/9	1/5	1/3	1/2	1/2	1	1
权重值	0.4562	1/4	0.1120	0.0696	0.0539	0.0381	0.0381

一致性比例 CR(Consistency Ratio): $CR=\dfrac{CI}{RI}$, 其中, 一致性指标$CI=\dfrac{\lambda_{\max}-n}{n-1}$, $\lambda_{\max}$ 为最大特征向量值平均随机一致性指标查表 (表 5) 可获得经矩阵的一致性检验的 CR=0.02559<0.1, 这表明了合成后的权重一致性满足要求.

表 5 平均随机一致性指标

阶数	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

3) 计算准则层对于目标层的合成权重

3.5 确定诊断树结构

上文已经提到构建的健康状态分类为 (健康, 温升, 强温, 激温), 根据合成权重值可以得到 DT-SVM 的诊断结构如图 8 所示[10].

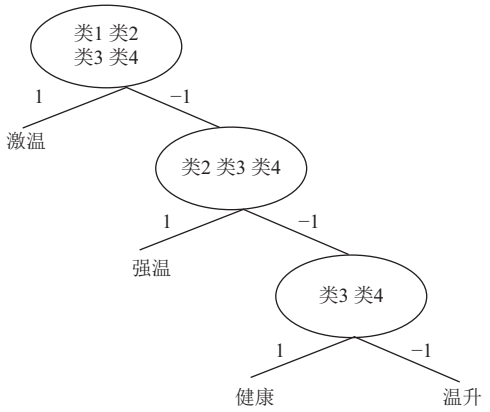


图 8 DT-SVM 诊断结构图

确定诊断结构后, 根据 1.2 节提到的改进算法进行求解, 步骤如下.

1. 确定SVM分类性能参数;
2. 输入训练样本, 完成训练过程;
3. 将测试样本输入到训练完成的DT-SVM中, 进行检验;
4. 在保证样本数据一致的情况下, 基于AHP-DT-SVM方法、SVM方法和决策树算法对轴箱轴承进行健康状态分类评估, 得到对比实验结果.

4 实验分析和模型评价

4.1 实验平台

本实验使用 Weka(怀卡托智能分析环境) 软件作为分类平台. Weka 是一款数据挖掘工具, 基于 Java 环境且开源、不收费, 受到广大数据挖掘工程师的欢迎, 其只需要对数据进行特征提取、选择合适的的算法算法、对各类参数进行调优即可. 本实验采用的是 Weka3.6 稳定版本, 植入了 SVM 的扩展包.

4.2 实验数据

本实验数据来源于国内某动车组生产制造基地于 2016 年的所有动车组转向架及其零部件服役数据以及所属实验室的动车组试验台上的模拟数据. 经过数据预处理操作之后, 选取大约 1 万条数据.

试验台模拟数据比较容易获得, 并且完整性较好, 包含轴承全生命周期的所有数据, 缺点是缺乏列车实际运行过程中环境因素等的影响, 会有偏差; 列车运行数据噪声比较多, 完整性较差, 尤其是故障数据极度缺乏; 综上所述, 采用以上两种数据结合的方式进行分类分析. 部分服役数据样例展示如表 6 所示, 其中健康数据共 5000 条，温升数据 3000 条，强温数据 4500 条，激温数据 1500 条，并且每个类别的数据都展示了两个温度值，代表这两个温度只都满足于这个类别.

表 6 目标层合成权重

分类	健康	温升	强温	激温
合成权重	0.2142	0.1577	0.1648	0.2663

说明. 轴箱轴承温度均值为同轴箱轴承其余传感器所测温度均值; 驱动侧温度均值为同一转向架驱动侧轴箱轴承传感器所测温度均值; 非驱动侧温度均值为同一转向架非驱动侧轴箱轴承传感器所测温度均值; 同侧轴承轴端温度为所测轴箱轴承同侧转向架轴的轴端温度.

4.3 实验分析

在 Weka 平台上, 分别利用 AHP-DT-SVM 算法、SVM 算法和 DT-SVM 算法对同一样本数据进行分类, 结果展示如表 7 所示. 同时图 9 对分类后的轴箱轴承温度做了具体说明.

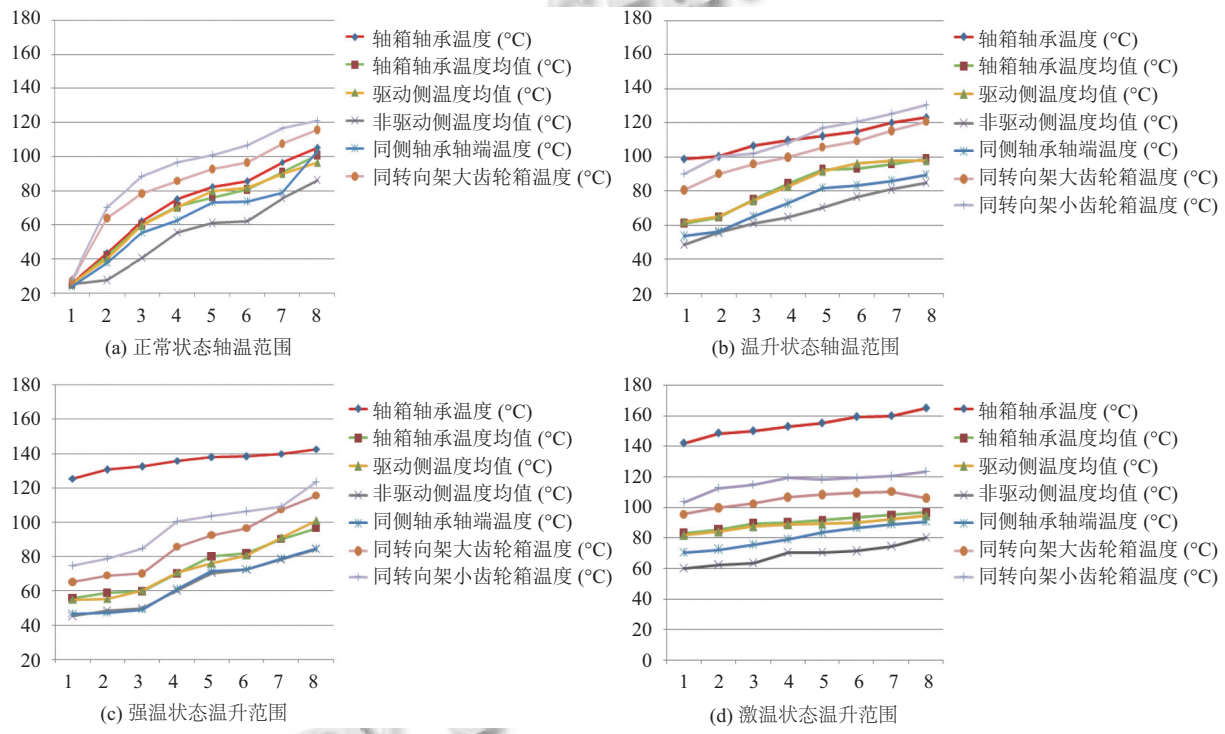
通过表 8 可知: 基于 DT-SVM 算法时间性能最好, 但是分类精度偏低并且由于根的不确定性导致的盲目性比较大; 基于 SVM 的分类算法虽然分类性能有所提高, 但是所需分类器多且用时最长; 基于 AHP-DT-

© 中国科学院软件研究所 http://www.c-s-a.org.cn

SVM 算法的分类精度最高, 同时分类器个数也少, 时间性能介于其他两者算法之间. 因此, 综合分类器数量、所需时间以及分类精度三个因素, 优化后的 AHP-SVM 算法相对于其他算法优越性更高.

表 7 部分数据样例展示表

	轴箱轴承温度(℃)	轴箱轴承温度均值(℃)	驱动侧温度均值(℃)	非驱动测温度均值均值(℃)	同侧轴承轴端温度(℃)	同转向架大齿轮箱温度(℃)	同转向架小齿轮箱温度(℃)
健康	25.96 85.67	25.96 81.89	25.36 80.65	25.42 62.35	23.96 73.65	27.1 96.52	28.24 106.84
温升	100.3 120.4	64.85 95.87	65.38 97.98	55.74 81.2	56.38 85.96	90.32 115.43	100.56 125.34
强温	125.6 138.4	55.96 81.89	54.86 80.65	45.23 72.35	46.85 72.56	65.14 96.52	74.65 106.24
激温	148.9 153.2	85.6 90.24	84.35 88.96	62.48 70.62	72.65 79.36	100.03 106.87	112.9 119.8



(a) 正常状态轴温范围 (b) 温升状态轴温范围 (c) 强温状态温升范围 (d) 激温状态温升范围

图 9 四种状态下特征向量值温度范围

表 8 三类算法分类对比图

分类算法	分类器(个)	时间性能(min)	分类精度(%)
AHP-DT-SVM	3	10.89	96.5
一对一SVM	4	12.6	92.4
DT-SVM	3	8.3	91.9

此外, 图 9 也隐含了 4 种健康状态下的数值分类规律: 图 9(a) 表示的正常状态下, 轴温低于 105℃, 与均温差值不超过 10℃; 图 9(b) 表示的温升状态, 轴温范围大约在 100℃~125℃ 之间, 在临界值波动时候, 与轴温均值超过了 35℃, 与驱动侧温度均值差超过了 40℃; 图 9(c) 表示的强温状态, 轴温范围大约在 120℃~140℃ 之间, 在临界值波动时候, 与轴温均值差超过了 55℃, 与驱动侧温度均值差超过了 60℃; 图 9(d) 表示的激温状态, 轴温超过了 140℃, 并且与轴温均值差超过了 65℃.

5 健康状态评估模型

本文采用多特征向量分类方式来提高判别度, 同时采用基于向量机的分类方式, 提高判断准确度, 从而对健康状态进行预测.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

基于上述健康状态评估模型以及 AHP-DT-SVM 分类结果, 我们可重新得到轴箱轴承健康状态评估判断标准: 1) 健康状态: 轴温低于 105℃, 与均温差值不超过 10℃; 2) 温升状态: 轴温范围大约在 100℃~125℃ 之间, 与轴温均值大于 35℃, 与驱动侧温度均值差大于 40℃; 3) 强温状态: 轴温范围大约在 120℃~140℃ 之间, 与轴温均值差大于 55℃, 与驱动侧温度均值差大于 60℃; 4) 激温状态: 轴温大于 140℃, 同时与轴温均值差大于 65℃. 当轴温处于两状态临界值时候, 用与轴温均值差和驱动侧温度均值差来明确划分状态.

6 结束语

本文提出了一种基于支持向量机和决策树方法的 AHP-DT-SVM 算法, 使用层次分析法计算特征值权重比例, 然后结合决策树和支持向量机方法对数据进行分类, 从而可以准确快速的实现轴箱轴承温度数据健康状态分类, 且基于此构建健康状态评估模型, 以规范分类过程和评估结果, 提高分类精度, 指导修程修制优化, 提高动车组运维效率.

参考文献

1 He QB, Wang J, Hu F, *et al*. Wayside acoustic diagnosis of defective train bearings based on signal resampling and information enhancement. Journal of Sound and Vibration, 2013, 332(21): 5635–5649. [doi: 10.1016/j.jsv.2013.05.026]

2 Fumeo E, Oneto L, Anguita D. Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Computer Science, 2015, 53: 437–446. [doi: 10.1016/j.procs.2015.07.321]

3 王勇, 韩伟, 王凤才. 高速铁路轴箱轴承接触润滑机理. 机械设计与制造, 2017, (5): 131–134.

4 Sakthivel NR, Indira V, Nair BB, *et al*. Use of histogram features for decision tree-based fault diagnosis of monoblock centrifugal pump. International Journal of Granular Computing, Rough Sets and Intelligent Systems, 2011, 2(1): 23–36. [doi: 10.1504/IJGCRSIS.2011.041458]

5 Soman KP, Ramachandran KI. Insight into Wavelets: From Theory to Practice. 2nd rev. ed. Prentice-Hall of India Private Limited, 2005.

6 Peng YH, Flach PA, Brazdil P, *et al*. Decision tree-based data characterization for meta-learning. ECML/PKDD-2002 Workshop IDDM-2002. Helsinki, Finland. 2002.

7 Zhao W, Tao T, Zio E. System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. Applied Soft Computing, 2015, 30: 792–802. [doi: 10.1016/j.asoc.2015.02.026]

8 Jindal A, Dua A, Kaur K, *et al*. Decision tree and SVM-based data analytics for theft detection in smart grid. IEEE Transactions on Industrial Informatics, 2016, 12(3): 1005–1016. [doi: 10.1109/TII.2016.2543145]

9 Jegadeeshwaran R, Sugumaran V. Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. Mechanical Systems and Signal Processing, 2015, 52–53: 436–446.

10 Lei YG, Jia F, Lin J, *et al*. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3137–3147. [doi: 10.1109/TIE.2016.2519325]

11 Li N, Zhou R, Hu QH, *et al*. Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. Mechanical Systems and Signal Processing, 2012, 28: 608–621. [doi: 10.1016/j.ymssp.2011.10.016]

12 Liu RN, Yang BY, Zhang XL, *et al*. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. Mechanical Systems and Signal Processing, 2016, 75: 345–370. [doi: 10.1016/j.ymssp.2015.12.020]

13 郭磊, 李兴林, 吴参, 等. 基于支持向量机的滚动轴承性能退化评估方法. 轴承, 2012, (8): 46–50.

14 董宝玉. 支持向量技术及其应用研究[博士学位论文]. 大连: 大连海事大学, 2016.

15 Saaty TL. A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 1977, 15(3): 234–281. [doi: 10.1016/0022-2496(77)90033-5]

16 胡俊, 滕少华, 张巍, 等. 支持向量机与哈夫曼树实现多分类的研究. 广东工业大学学报, 2014, 31(2): 36–42.

17 He QB, Yan RQ, Kong FR, *et al*. Machine condition monitoring using principal component representations. Mechanical Systems and Signal Processing, 2009, 23(2): 446–466. [doi: 10.1016/j.ymssp.2008.03.010]

18 赵佳颖, 曹成鹏, 孙景辉, 等. 高速动车组轮对轴箱轴承温度监控技术优化改进. 铁道机车与动车, 2013, (5): 37–38, 50.

19 马千里. 车辆轴温智能探测系统 (THDS) 探测客车热轴预报规律的研究. 铁道车辆, 2016, 54(2): 8–11.

20 周晓兰, 谢红. 高校图书馆网站评价指标体系建立与指标权重的计算研究. 现代情报, 2009, 29(8): 99–102.

© 中国科学院软件研究所 http://www.c-s-a.org.cn