

# 基于支持向量机的健康状态评估方法<sup>①</sup>

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<sup>[1]</sup>, 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<sup>[2]</sup>. 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

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<sup>[3]</sup>,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<sup>[4]</sup>,自回归滑动平均 (ARMA) 特征<sup>[5]</sup>,直方图特征<sup>[5]</sup>或小波特征<sup>[6]</sup>,常用的特征选取技术包括主成分分析 (PCA)<sup>[7]</sup>,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<sup>[8]</sup>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,目前已经成功应用在各大领域和系统中,包括智能电网<sup>[8]</sup>、汽车液压制动系统<sup>[9]</sup>、汽油发动机气门<sup>[10]</sup>、轴承<sup>[12,13]</sup>等.支持向量机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<sup>[14]</sup>.

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, SVM-DT-SVM)<sup>[15]</sup>.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;DDAG-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的提高.相对 DDAG-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.

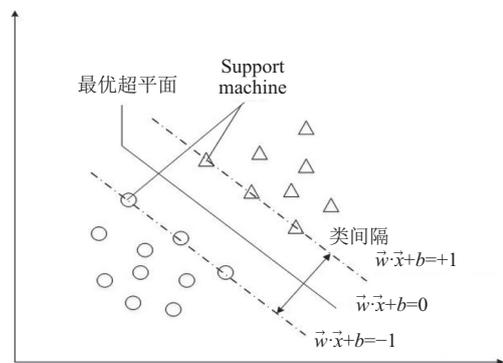


图 1 线性可分情况

计算模型:给定训练样本集  $\{x[i], y[i]\}, i=1, 2, 3, \dots, n, y$ , 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  $w^T x + b = 0$ , 则任意一点  $(x, y)$  到这个超平面的距离为  $\frac{|w^T x + b|}{\|w\|_2}$ , 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:

$$\max \frac{|w^T x + b|}{\|w\|_2} \quad (1)$$

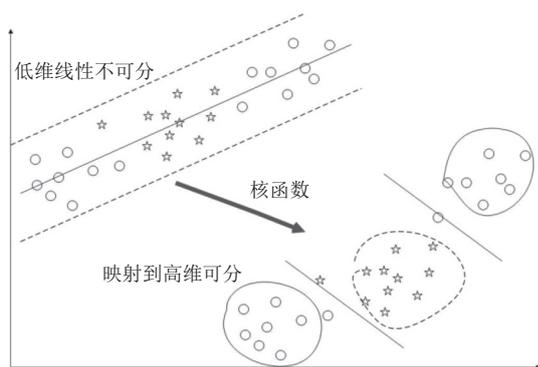


图2 非线性可分情况

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  $w^T x + b$  值为 1 或 -1. 所以这时目标公式 (1) 就简化为:

$$\max \frac{1}{\|w\|_2} \quad (2)$$

而此时的约束条件可以表达为:

$$y_i (w^T x_i + b) \geq 1, i = 1, 2, \dots, m \quad (3)$$

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).

$$\min \frac{1}{2} \|w\|^2 \quad (4)$$

$$y_i (w^T x_i + b) \geq 1, i = 1, 2, \dots, m \quad (5)$$

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

$$\min f(w, b) = \frac{1}{2} \|w\|^2 \quad (6)$$

$$g_i(w, b) = 1 - y_i (w^T x_i + b) \leq 0, i = 1, 2, \dots, m \quad (7)$$

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):

$$\max \theta_D(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_i a_j y_i y_j \langle x_i, x_j \rangle \quad (8)$$

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  $a$ , 求得  $w, b$ , 从而得到最优分类面.

$$w = \sum_{i=1}^m a_i y_i x_i$$

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

### 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.

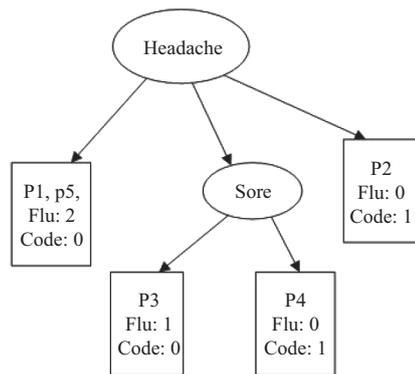


图3 决策树实例

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=severe) 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定量相结合的基于运筹学理论的层次权重决策分析方法<sup>[16]</sup>, 其主要特点是用











