

基于 PCA 和 SVM 的轨道不平顺状态识别

林怀青¹, 柴晓冬², 苏圣超³, 郑树彬²

(1. 上海工程技术大学 机械工程学院, 上海 201620; 2. 上海工程技术大学 城市轨道交通学院, 上海 201620;

3. 上海工程技术大学 工程实训中心, 上海 201620)

摘要:为快速识别轨道不平顺中存在的短波不平顺类型,提出基于主成分分析(PCA)和支持向量机(SVM)进行轨道不平顺状态识别的方法。首先提取轴箱加速度的特征参数,并采用主成分分析法对特征参数进行降维处理,提取出轨道不平顺的主元特征;然后构建支持向量机多分类器,以不同不平顺类型下轴箱加速度数据来验证模型的准确性;最后对实测数据进行轨道不平顺识别。通过对不同轨道不平顺下轴箱加速度的分析,结果表明该方法能够有效地实现一定区段内轨道不平顺类型的识别。

关键词:轨道不平顺;轴箱加速度;主成分分析(PCA);支持向量机(SVM)

中图分类号:TP391.4 **文献标识码:**A **文章编号:**1000-8829(2016)05-0025-04

Track Irregularity State Recognition Based on PCA and SVM

LIN Huai-qing¹, CHAI Xiao-dong², SU Sheng-chao³, ZHENG Shu-bin²

(1. College of Mechanical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. College of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai 201620, China;

3. Engineering Training Center, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: In order to identify the type of short wave irregularity in track irregularity, a method based on principal component analysis and support vector machine is proposed for the identification of track irregularity. Firstly, extracting the characteristic parameters of the axle box vibration acceleration, and then the main elements of the track irregularity are extracted by using principal component analysis to analyze the feature parameters. Secondly, the support vector machine multi-classifier is constructed to verify the accuracy of the model by the use of different types of axle box acceleration data. At last, the measured data are analyzed to identify the track irregularity. Results indicate that the proposed method can effectively realize the identification of the track irregularity in a certain range according to the analysis of the acceleration of the axle box with different track irregularity.

Key words: track irregularity; axle box acceleration; principal component analysis(PCA); support vector machine (SVM)

轨道不平顺是引起列车产生振动的主要原因^[1-2],即使幅值很小的轨面不平顺也可能引起轮轨间产生强烈的振动。列车的轴箱振动信号中蕴藏着大量的轨道状态以及病害信息,因此基于列车的轴箱加速度来研究轨道不平顺具有现实意义。

收稿日期:2015-09-11

基金项目:国家自然科学基金项目(51478258);上海市科委重点支撑项目(13510501300);上海市研究生教育创新计划学位点引导布局与建设培育项目(13sc002);上海工程技术大学研究生科研创新项目(14KY1007)

作者简介:林怀青(1990—),男,硕士研究生,主要研究方向为基于数据驱动的轨道不平顺估计。

目前,基于列车轴箱振动进行轨道状态估计的研究较少,已有的相关工作主要就轨道不平顺与轴箱加速度的相关关系进行研究。文献[3]利用希尔伯特-黄变换方法分析轴箱加速度信号,检测轨道线路中的短波不平顺;文献[4]利用 EMD 和 Cohen 核分布相结合的方法,抑制二次分布中的交叉项,实现基于轴箱加速度信号的轨道不平顺分析;文献[5]利用小波能量谱的方法实现轨道不平顺状态的评估。这些算法都对轨道不平顺识别进行了研究,但只能判断是否存在短波不平顺,对于短波不平顺的类型尚无法判断;同时计算量比较大,对轨道短波不平顺识别的效率较低。

为有效地识别轨道不平顺中存在的短波不平顺类

型,本文基于 PCA 和 SVM 实现轨道状态的快速识别。采用谐波型模型描述轨头压陷、焊接凹陷和波形磨耗三类短波不平顺,然后与美国 4 级轨道谱(美国联邦铁路局对大量实测数据进行研究,得出符合美国的六个轨道级别的轨道谱)相叠加,经过 SIMPACK(德国 INTEC GmbH 公司的多体动力学分析软件包)车辆模型的仿真,获取对应不平顺下轴箱加速度。然后采用 13 个时域特征参数和 8 个频域特征参数来表征轴箱振动特征,同时,针对特征参数的信息冗余以及维数过高,采用 PCA 方法来实现数据降维,结合 SVM 法识别轨道不平顺状态。最后采用 PCA 和 SVM 方法^[6-7]对实测数据进行轨道不平顺状态识别。

1 算法原理

1.1 主成分分析原理

主成分分析法(PCA)是把多指标之间互相关联的复杂关系进行简化分析的方法^[8]。它把几个相关指标通过线性变换转变为其他几个不相关的指标,新指标按照方差递减顺序依次排列。其中第一主成分具有最大方差,重要性最大,第二主成分次之,依次类推。从机车轴箱振动状态的角度推断轨道不平顺状态的过程中,必须考虑许多时频域指标,而这些指标间存在一定的相关性,在某种程度上存在信息的重叠。经过主成分分析,在保证数据信息丢失最小的原则下,用较少的互不相关的综合指标替代原来较多的指标,实现降维的目的。其主要步骤如下:

有原始数据矩阵

$$X_{m \times n} = (x_1, x_2, \dots, x_n) \quad (1)$$

式中, m 为某种类型的轴箱振动样本数; n 为分析变量数。为消除变量的量纲不同、数值差异过大带来的影响,需进行标准化处理

$$x_i^* = \frac{x_i - \text{mean}(x_i)}{\sqrt{\text{std}(x_i)}} \quad (2)$$

① 其中 $i = 1, 2, \dots, n$; 并计算 $X_{m \times n}^*$ 的协方差矩阵 $Y_{n \times n}$ 。

② 根据特征方程 $(\lambda I - Y_{n \times n}) \times \mu = 0$, 求出 $Y_{n \times n}$ 的各特征值 λ 和对应的特征向量 μ 。

③ 计算主成分 $\{Z_1, Z_2, \dots, Z_n\}$; 主成分 Z_i 的贡献率 $\lambda_i / \sum_{i=1}^n \lambda_i$; 设 $k < n$, 计算主成分 Z_1, \dots, Z_k 的累计贡献率 $\sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i$ 。选取使主成分的累计贡献率大于 85% 的最小整数 k , 使分析对象从 n 维降为 k 维。

1.2 支持向量机原理

支持向量机(SVM, support vector machine)是基于统计学习理论中结构风险最小化原理和 VC 维(Vap-

nik-Chervonenkis dimension)理论的一种小样本统计学习方法,通过应用核函数(也称内积函数),实现从样本空间到高维特征空间的非线性映射^[9]。在 R^n 空间中,分类预测的本质是寻找通过学习样本诱导的函数,使函数有良好的泛化性。其原理如下:

设有样本集 (x_i, y_i) , $i = 1, 2, \dots, n$, 其中 $x \in R^d$ 表示输入, $y \in \{\pm 1\}$ 表示输出。支持向量机的目标是找到使训练数据平均错误分类误差最小的超平面,从而可以推导出优化问题。设最优平面为

$$\Phi(\omega, \xi) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \xi_i \quad (3)$$

则偏置 b 和权重向量 ω 需要满足以下约束条件

$$y_i(\theta) \geq \lambda_i \quad (4)$$

式中, $\theta = \omega^T x_i + b$; $\lambda_i = 1 - \xi_i$, ξ_i 为松弛变量,表示模式与理想情况的偏差程度; C 为惩罚系数,表示对错分样本的惩罚程度。

依据拉格朗日乘子法,可将最优平面的求解转化为以下的约束优化问题

$$Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j K(x_i, x_j) \quad (5)$$

$$\sum_{i=1}^n a_i y_i = 0 \quad a_i \geq 0 \quad (6)$$

式中, $\{a_i\}_{i=1}^n$ 为拉格朗日乘子。拉格朗日乘子中大部分为 0, 而不等于 0 的 a_i 所对应的样本称为支持向量机。同时,采用满足 Mercer 条件的核函数 $K(x_i, x_j)$ 代替高维特征空间的内积运算,避免了因维数过高而引起的“维数灾难”问题。

2 仿真与分析

2.1 仿真数据获取

本文采用谐波型模型来描述轨头压陷、焊接凹陷和波浪形磨耗三类短波不平顺,将短波不平顺叠加在美国 4 级轨道谱上模拟存在短波不平顺的轨道线路。将短波不平顺作为 SIMPACK 轨道车辆的外部激励输入,车速设置为 36 km/h,采样频率为 400 Hz,轨道长度为 1000 m,得到短波不平顺下的轴箱加速度。以 10 m 为研究单元,验证 PCA-SVM 模型的有效性。

2.2 轴箱振动加速度特征分析

轨道不平顺是引起列车产生振动的主要原因,即使幅值很小的轨面不平顺也可能引起轮轨间产生强烈的振动。可以说列车的轴箱加速度信号中蕴藏着大量的轨道状态以及病害信息。根据轴箱加速度时域波形无法确定轨道状况,为进一步对轴箱加速度进行分析,可以对其时域特征以及频域特征进行分析。提取轴箱振动加速度的均值、峰值、均方根值、方根幅值、绝对平

均值、方差、歪度、峭度、波形指标、峰值指标、脉冲指标、裕度指标和峭度指标 13 个时域特征参数。对轴箱振动加速度进行三层小波包分析,将频域平分为 8 个频带,分别计算各个频带的相对功率谱能量,提取 8 个频域特征参数。共计提取 21 种特征参数表征轴箱振动特性,各参数定义及计算方法见文献[10]和文献[11]。

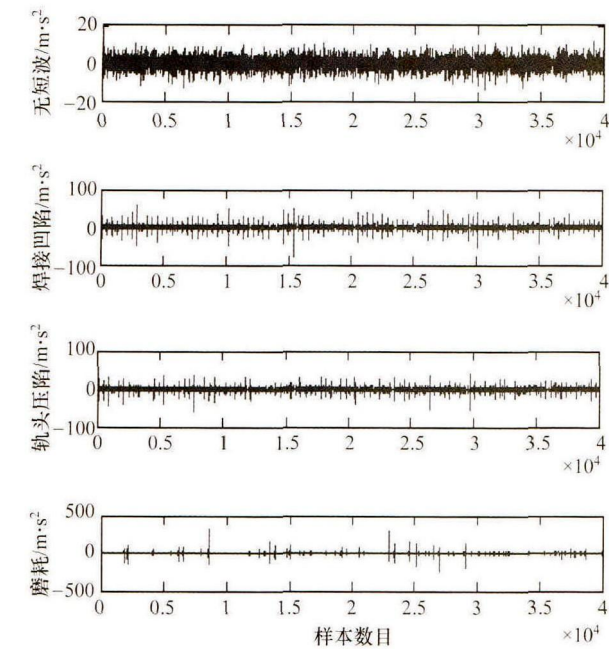


图 1 轴箱加速度数据

2.3 PCA 数据降维

轴箱加速度的 21 个时频域特征参数所反映的特征状态各不相同,难以判断哪些特征参数规律性好以及敏感性强。同时,采用 21 个特征向量对轨道状况进行分类预测存在信息冗余以及维数过高的问题,会因计算量过大而导致耗时较长,增加构建分类器的难度,所以采用 PCA 方法对特征向量进行降维和优化处理。

分别抽取不含短波不平顺、轨头压陷、焊接凹陷和波浪形磨耗 4 种状况的轴箱加速度各 50 组,运用主成分分析法对轴箱振动加速度样本数据进行数据降维分析,得到的主成分贡献率以及特征权系数如图 2 及表 1 所示。

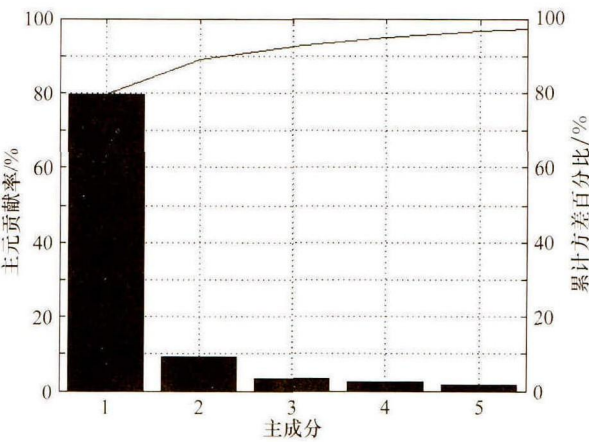


图 2 主成分贡献率

表 1 主成分及特征参数权系数

主成分编号	特征值	累计方差百分比	特征参数	主成分权系数		
				主成分 1	主成分 2	主成分 3
1	5861.852	79.667	均值	0.000	-0.001	-0.001
2	681.485	88.930	峰值	0.410	0.320	-0.005
3	255.732	92.406	均方根值	0.031	-0.006	0.017
4	190.774	94.999	方根幅值	0.004	-0.003	0.010
5	127.340	96.729	绝对平均值	0.010	-0.007	0.013
6	75.654	97.758	方差	0.602	-0.358	-0.414
7	50.218	98.440	歪度	0.000	-0.018	-0.017
8	42.991	99.024	峭度	0.200	0.590	0.054
9	35.056	99.501	波形指标	0.003	0.004	0.003
10	20.464	99.779	峰值指标	0.021	0.055	0.021
11	10.953	99.928	脉冲指标	0.066	0.146	0.034
12	3.575	99.977	裕度指标	0.118	0.188	0.022
13	1.075	99.991	峭度指标	0.000	0.001	-0.002
14	0.548	99.999	子频带能量 1	0.055	0.022	0.077
15	0.062	99.999	子频带能量 2	0.267	-0.257	0.870
16	0.023	100.000	子频带能量 3	0.197	-0.015	-0.090
17	0.009	100.000	子频带能量 4	0.356	-0.388	0.023
18	0.005	100.000	子频带能量 5	0.196	0.297	0.189
19	0.004	100.000	子频带能量 6	0.240	0.167	0.006
20	0.001	100.000	子频带能量 7	0.195	0.098	-0.122
21	0.000	100.000	子频带能量 8	0.189	0.131	-0.034

从图2以及表1可以看出,前5个主成分的累计方差百分比达到96%以上,完全可以解释原有的21个变量所反映的轴箱加速度时频域信息,大大地降低了分析的维数。在轴箱加速度时频域特征参数中,峰值、方差和子频带能量4对主成分1的权系数在0.3以上,这些特征参数对主成分1起主导作用。峰值、方差、峭度和子频带能量4对主成分2起主导作用,方差和子频带能量2对主成分3起主导作用。

2.4 SVM 分类识别

SVM 是基于二分类问题而提出来的,其原始模型只能解决二分类问题。多分类问题可以通过对二分类器的一系列组合向多个二分类问题进行转化。常用的组合方法有一对一、一对多、有向无环图和二叉树等方法^[12]。对于从轴箱振动来实现轨道不平顺估计的分类问题,由于不平顺的类别不多,同时要求具有较高的分类精度,采用一对一方法构建多分类模型具有明显的优越性。

轴箱加速度的21个特征参数经PCA降维后,可以选用前5个主成分来表述原来的轴箱加速度数据。再分别抽取不含短波不平顺、轨头压陷、焊接凹陷和波浪形磨耗4种状况的轴箱加速度各50组,先进行归一化,然后乘以前200组数据得到的主成分系数。5个主成分特征构成一个 400×5 的特征矩阵,轨道不平顺的每种状态对应一个 100×5 的特征矩阵,每种状态的前50组数据用于训练,后50组数据用于测试。选取径向基核(RBF)作为SVM的核函数,采用网格搜索法来确定SVM的惩罚系数C和核函数参数。训练结果如图3所示。

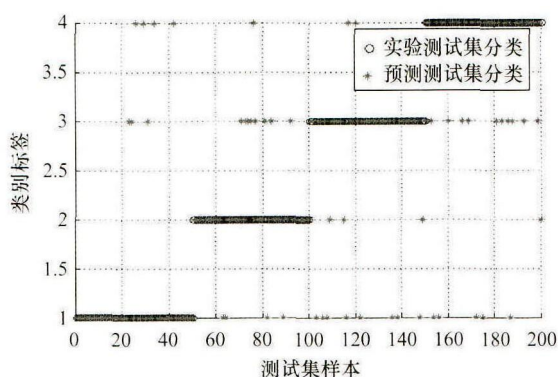


图3 轨道不平顺状况识别分类图

其中标签1表示不含短波不平顺,标签2表示存在轨头压陷,标签3表示存在焊接凹陷,标签4表示存在波形磨耗。200个测试样本中有151个分类正确,准确率为75.5%,可以较为准确地识别轨道不平顺的类型。

3 实例分析

机车轴箱振动加速度数据选用上海地铁2号线龙

阳路至张江高科站之间一段轨道采集的46 s数据,机车行驶速度为37.1 km/h,采样频率为5000 Hz。



图4 轴箱加速度传感器安装位置

为尽量保持仿真数据的采样频率一致,通过对每10个数据中抽取1个将采样频率调整为500 Hz,图5是经过低通滤波后画出的轴箱轴箱加速度图。

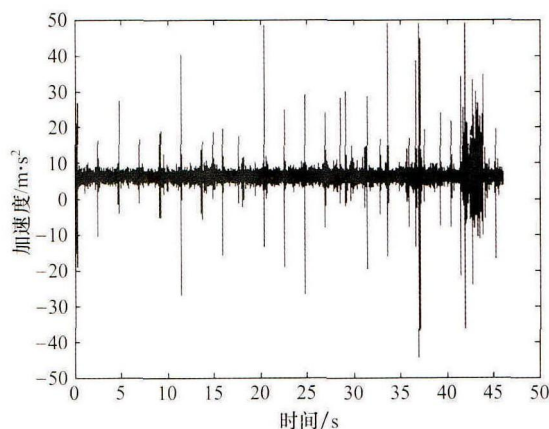


图5 轴箱加速度数据

将每秒采集的数据作为研究单元,提取46组实测数据的21个特征参数,然后进行归一化,再根据上述训练数据得到的主成分系数,实现数据的降维处理,构成一个 46×5 的测试样本矩阵,设置类型标签为1。根据上述已训练好的模型进行分类预测,训练结果如图6所示。

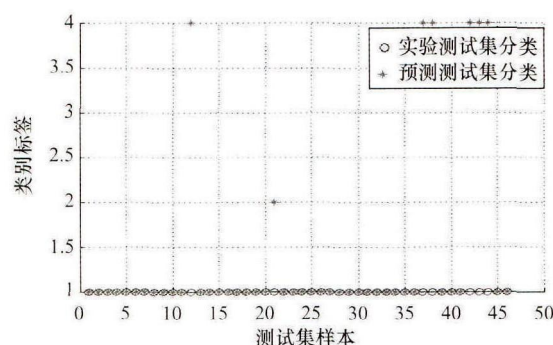


图6 实测数据轨道不平顺状况识别分类图

(下转第36页)


```

agram -> size(), &sender, &senderPort);
TAGORG tag;
TagTempList. clear();
QDataStream in( datagram, QIODevice::ReadOnly);
while( false == in. atEnd() ) {
in >> tag. SID >> tag. type >> tag. IFF;
in >> tag. x >> tag. y >> tag. z;
in >> tag. da >> tag. db >> tag. v;
if( QDataStream::Ok != in. status() ) {
qDebug() << "get Org Target UDP datagram error !";
break;
}
.....// 坐标中心转换
if( false == filtTarget(&tag) ) continue;
TagTempList. append( tag );//写入目标临时表
}
delete datagram;
}
// 更新目标集
QWriteLocker locker( m_pOrgTagListLock);
m_pOrgTagList -> clear();
m_pOrgTagList -> append( TagTempList);
m_OrgTagTimeOver. restart(); //目标超时计时器清零
}

```

用户操作训练时,数据采集与驱动板读取面板元件和组合元件的动作,将信息以 UDP 报文格式发送给通信控制板,再经处理、分发后通过网络交换机发送给主控计算机。计算机主控软件根据这些信息进行综合处理,将处理结果以广播的形式送往 CAN 总线。以太

网内各设备根据需求组合的单片机板号,将 UDP 报文通知该电路板并驱动对应状态指示灯点亮,或在显示屏上显示相应图像。

3 结束语

系统运行过程中解决了不同型号模拟器之间数据采集与驱动电路、通信控制电路的模块化应用和通信模块的可移植应用,效果较好。在系统训练过程中也出现过如下不足,模拟训练机柜与三维战场视景仿真之间庞大的数据交互量在保证高速传输的条件下,当空情目标批次超过 36 批时,CAN 总线和 UDP 数据通信之间偶尔出现数据掉包现象。下一步将修改完善 UDP 传输协议,使用 CAN 总线专用调试设备进行综合通信调试,确保解决此问题。

参考文献:

- [1] Microchip Technology Inc. PIC18F25K80 Datasheet[Z]. 2007-10.
- [2] 李春光,孙朝霞. CAN 总线式多功能传感器节点的设计[J]. 长春师范学院学报(自然科学版),2011,30(4).
- [3] 陈继海,周凤星,严保康. 基于 Cortex-M3 的多通道数据采集系统设计[J]. 仪表技术与传感器,2013(12).
- [4] 程小艳. 嵌入式 STM32F107VCT6 微处理器接口模块开发及应用研究[D]. 合肥:合肥工业大学,2012.
- [5] 李俊舟,赵继敏,邓伟,等. 基于 DSP 和 UDP 协议的数据采集系统[J]. 仪表技术与传感器,2013(4):41-42.

□

(上接第 28 页)

从图 6 可以看出,46 组实测数据中有 1 组存在轨头压陷,有 6 组存在波形磨耗,不存在焊接凹陷。所以应将这些路段作为轨道工务维修的重点,以保证机车的行车安全。

4 结束语

综上所述,本文提出的基于主成分分析和支持向量机的轨道不平顺状态识别方法,具有较高的算法效率以及较好的识别精度,能够有效地实现一定区段内轨道不平顺类型的识别。有助于铁路公务部门针对性地对轨道不平顺进行养护维修,提高工作效率,为城市轨道交通的运营安全提供保障。

参考文献:

- [1] 罗林,张格明,吴旺青,等. 轮轨系统轨道平顺状态的控制[M]. 北京:中国铁道出版社,2006.
- [2] 翟婉明. 车辆-轨道耦合动力学[M]. 4 版. 北京:科学出版社,2015.
- [3] 曹西宁,柴晓冬,郑树彬. 基于 Hibert-Huang 变换的轨道

车辆轴箱加速度信号分析[J]. 仪表技术与传感器,2015(3):92-95.

- [4] 宁静,诸昌铃,张兵. 基于 EMD 和 Cohen 核的轨道不平顺信号分析方法[J]. 振动与冲击,2013,32(4):31-38.
- [5] 徐磊,陈宪麦,徐伟昌,等. 小波能量谱在铁路轨道检测中的应用[J]. 振动工程学报,2014,27(4):605-612.
- [6] 吴峰,陈后金,姚畅,等. 基于网格搜索的 PCA-SVM 道路交通标志识别[J]. 铁道学报,2014,36(11):60-64.
- [7] 刘永斌,何清波,孔凡让,等. 基于 PCA 和 SVM 的内燃机故障诊断[J]. 振动、测试与诊断,2012,32(2):250-255.
- [8] 张鹏. 基于主成分分析的综合评价研究[D]. 南京:南京理工大学,2005:24-26.
- [9] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995,20(3):273-297.
- [10] 陈可兴,李川奇. 设备状态监测与故障诊断技术[M]. 北京:科学技术文献出版社,1991:40-87.
- [11] 何正嘉,陈进,王太勇,等. 机械故障诊断理论及应用[M]. 北京:高等教育出版社,2010:31-67.
- [12] 朗宇宁,蔺娟如. 基于支持向量机的多分类方法研究[J]. 中国西部科技,2010,9(17):28-29.

□