

# 基于 Lasso 回归的高速动车组齿轮箱性能检测方法

王远霏 孙海荣 王东星

(中车唐山机车车辆有限公司, 063035, 唐山//第一作者, 正高级工程师)

**摘要** 传动齿轮箱作为高速动车组走行系统的核心部件, 直接决定列车的运行速度和运行状况, 对列车的行车安全起着至关重要的作用。提出了一种基于 Lasso 回归的动车组齿轮箱性能检测方法, 通过分析动车组齿轮箱实际运行时的性能指标与其期望值的偏离情况, 检测动车组齿轮箱的异常状态, 进而识别动车组齿轮箱的早期故障。实例验证了基于 Lasso 回归的动车组齿轮箱性能检测方法的有效性和可行性。

**关键词** 高速动车组; 齿轮箱; Lasso 回归; 性能检测

**中图分类号** U270.1<sup>+</sup>4

**DOI**: 10.16037/j.1007-869x.2020.03.029

## Performance Examination Method for High-speed EMU Gearbox Based on Lasso Regression

WANG Yuanfei, SUN Hairong, WANG Dongxing

**Abstract** As the core component of high-speed EMU running system, driving gearbox directly determines the running speed and operation status of the train, which plays an important role in the operation safety. A performance examination method for high-speed EMU gearbox based on the Lasso (Least Absolute Shrinkage and Selection Operator) regression was proposed. By analyzing the deviation between actual values and expected values of the performance index for high-speed EMU gearbox, the error state was detected and the initial failure was identified. Furthermore, simulation example results show the effectiveness and feasibility for Lasso regression to be applied to detect the performance of gearbox.

**Key words** high-speed EMU; gearbox; Lasso regression; performance examination

**Author's address** CRRC Tangshan Co., Ltd., 063035, Tangshan, China

传动齿轮箱作为高速动车组走行系统的核心零部件, 负责将牵引电机的高速转速转化为带动列车前进的巨大扭矩, 相当于汽车的变速箱, 直接决定列车的运行速度和运行状况, 对列车的行车安全

起着至关重要的作用。随着列车运行速度的提升以及运行环境的不断变化, 传动齿轮箱的正常运行也相应地面临多重问题, 对动车组齿轮箱进行早期故障预警已迫在眉睫。

为了确保动车组走行系统的安全可靠运行, 在动车组的走行系统上加装了大量的传感器, 能够对齿轮箱的工作参数进行监测, 可为齿轮箱的故障诊断提供可靠的数据基础。目前最常用的齿轮箱故障诊断方法主要有振动诊断和热分析方法。振动诊断方法是齿轮传动系统最常见的检测技术, 如频域分析、时域分析、时频分析等<sup>[1-2]</sup>; 热分析方法主要有热网络方法<sup>[3]</sup>和计算热传学方法等<sup>[4-5]</sup>。但是, 目前的动车组齿轮箱状态检测和日常检修主要侧重于已经发生的故障, 对于故障的早期检测尚缺乏适用的方法和有效的技术; 动车组齿轮箱的维修策略主要是定时维修和事后维修, 不能满足提高动车组运行效率及降低运营成本的要求。

本文提出了一种基于 Lasso 回归的动车组齿轮箱性能检测方法, 通过分析动车组齿轮箱实际运行时的性能指标与其期望值的偏离情况, 检测动车组齿轮箱的异常状态, 进而识别动车组齿轮箱的早期故障。同时, 可以在不改变动车组现有设备和检测条件的情况下, 充分利用动车组现有数据对动车组齿轮箱进行性能预测, 提前发现齿轮箱故障, 从而进一步减少齿轮箱故障对动车组正常运营的影响。

## 1 Lasso 回归

Lasso (Least Absolute Shrinkage and Selection Operator, 简称为 Lasso) 回归, 又称套索回归, 是 Robert Tibshirani 于 1996 年提出的一种新的变量选择技术<sup>[7]</sup>。Lasso 回归是一种收缩估计方法, 其基本思想是在回归系数的绝对值之和小于一个常数的约束条件下, 使残差平方和最小化, 从而能够产生某些严格等于 0 的回归系数, 进一步得到可以解释的模型<sup>[8]</sup>。

在进行 Lasso 回归之前,需要对原始数据进行去中心标准化处理,从而消除不同指标量纲的影响,即使观测数据满足以下条件:

$$\begin{aligned} \sum_{i=1}^n y_i &= 0, \sum_{i=1}^n x_{ij} = 0, \\ \sum_{i=1}^n x_{ij}^2 &= 1, j = 1, 2, \cdots, p \end{aligned} \tag{1}$$

式中:

$y_i$ ——目标变量或响应变量第  $i$  个值;  
 $x_{ij}$ ——解释变量对应的数据矩阵中第  $i$  行、第  $j$  列的值。

Lasso 回归是在残差平方和( $R_{ss}$ )最小化的计算中加入一个  $L_1$  范数作为罚约束:

$$\begin{aligned} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \\ = R_{ss} + \lambda \sum_{j=1}^p |\beta_j| \end{aligned} \tag{2}$$

式中:

$\beta$ ——回归系数向量;  
 $\lambda$ ——系数。

Lasso 回归在线性模型中所对应的目标函数为:

$$\hat{\beta}_{\text{Lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ R_{ss} + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

在运用 Lasso 回归进行建模分析时,需要特别注意公式中系数  $\lambda$  的设置, $\lambda$  控制了惩罚的严厉程度。若设置过大,最后的模型参数均将趋于 0,形成拟合不足;若设置过小,又会形成拟合过度。所以  $\lambda$  的取值一般需要通过交叉验证来确定。

Lasso 回归之所以优于其他回归模型,在于其采用  $L_1$  范数的作为罚约束。采用  $L_1$  范数的好处是当  $\lambda$  充分大的时候,可以把某些待估系数精确地收缩到 0。缩减系数的目的是为了消除噪声特征(去除对模型影响不大的特征)和关联的特征(消除多重共线的影响)。这也是本文采用 Lasso 回归进行建模分析的一个很重要的原因。

## 2 齿轮箱正常行为模型构建

正常行为模型是部件处于健康状态(即部件无故障、无性能退化时的运行状态)下构建的模型,是用来定义所研究部件的性能指标及其影响因素之间的函数关系的模型。通过监测性能指标的变化过程,来对所研究部件的性能退化及早期故障进行

检测。

### 2.1 建模数据选择

目前,我国在线运营的动车组均属于动力分散式电力动车组<sup>[6]</sup>。不同速度级别下的动车组车型也有所差异,但在动车组齿轮箱的数据采集点位方面却有着相似之处。本文以 CRH380B 车型为例进行研究,该车型与齿轮箱数据采集相关的部分参数字典如表 1 所示。

表 1 CRH380B 齿轮箱部分参数字典

参数代码	参数名称
CRH3-X1	电机 1 大齿轮温度-电机侧
CRH3-X2	电机 2 大齿轮温度-电机侧
CRH3-X3	电机 3 大齿轮温度-电机侧
CRH3-X4	电机 4 大齿轮温度-电机侧
CRH3-X5	电机 1 小齿轮温度-轮侧
CRH3-X20	电机 2 小齿轮温度-轮侧
CRH3-X21	电机 3 小齿轮温度-轮侧
CRH3-X22	电机 4 小齿轮温度-轮侧
CRH3-X23	电机 1 大齿轮温度-轮侧
CRH3-X24	电机 2 大齿轮温度-轮侧
CRH3-X25	电机 3 大齿轮温度-轮侧
CRH3-X26	电机 4 大齿轮温度-轮侧
CRH3-X6	电机 1 小齿轮温度-电机侧
CRH3-X7	电机 2 小齿轮温度-电机侧
CRH3-X8	电机 3 小齿轮温度-电机侧
CRH3-X9	电机 4 小齿轮温度-电机侧
CRH3-X10	列车运行速度
CRH3-X11	环境温度

鉴于齿轮箱对动车组走行安全的重要性,需要随时对其温度进行监控。在高速动车组处于正常运行状态时,齿轮箱温度传感器每一天针对每一个点位都会采集到成百上千条数据。为了更好地构建齿轮箱正常行为模型,对齿轮箱温度传感器采集到的数据进行特征提取,即将齿轮箱温度传感器采集到的数据按天进行分组;选取每个分组下温度较高的点(列车正常运行状态下),并计算这些点的平均值,使得每一分组仅有一个样本点,多个这样的样本点构成一个建模学习的样本。

本次建模的目的是通过选取某一个电机齿轮温度(如 CRH3-X6 的温度)作为动车组齿轮箱性能检测指标,将其相同部件类型同侧不同电机齿轮温度(如 CRH3-X7、CRH3-X8、CRH3-X9)、列车运行速度、环境温度作为性能指标的影响因素,构建动车组齿轮箱性能检测模型。

### 2.2 正常行为模型构建

以 CRH380B 为例,齿轮箱正常行为模型构建主要分为以下几个步骤:

1) 选取某一电机齿轮温度(如 CRH3-X6 的温度)作为动车组齿轮箱性能检测的指标,即性能指标。

2) 将步骤 1) 中选取的性能指标相同部件类型同侧不同电机齿轮温度(如 CRH3-X7、CRH3-X8、CRH3-X9)、列车运行速度(CRH3-X10)、环境温度(CRH3-X11)作为性能指标的影响因素。

3) 定义动车组齿轮箱性能检测的正常行为模型,即定义步骤 1) 中确定的性能指标与步骤 2) 中确定的影响因素之间的函数关系。

模型采用以下方程表示:

$$\tilde{Y}_1=f(Y_2,Y_3,Y_4,v,T_{out})+\varepsilon \tag{3}$$

式中:

- $\tilde{Y}_1$ ——步骤 1) 中确定的性能指标 $Y_1$ 的期望值;
- $Y_2,Y_3,Y_4$ ——与性能指标 $Y_1$ 相同部件类型同侧其他 3 个电机齿轮温度;
- $v$ ——列车运行速度;
- $T_{out}$ ——环境温度;
- $\varepsilon$ ——模型拟合误差;
- $f(\cdot)$ ——齿轮箱性能检测的正常行为模型,用于描述步骤 1) 中定义的性能指标与步骤 2) 中定义的影响因素之间的函数关系。

4) 为了更好地确定齿轮箱性能检测的影响因素,本文采用 Lasso 回归对步骤 3) 中建立的模型进行拟合。

5) 选择动车组齿轮箱无故障、无性能退化并且处于正常运行状态下的数据构建训练样本集和测试样本集,将其应用于步骤 4) 中构建的模型训练和测试,得到描述动车组齿轮箱正常工作时的电机齿轮温度变化的性能检测模型。具体的数据挑选方式详见本文 2.1 节中的描述。

6) 采用步骤 5) 中构建的样本对步骤 4) 中的 Lasso 回归模型进行训练和测试。

7) 基于动车组齿轮箱性能检测模型进行动车组齿轮箱性能检测,具体方式为:按照步骤 5) 的方法构建齿轮箱性能检测模型的输入样本,得到步骤 1) 中定义的性能指标的期望值  $\tilde{Y}_1$ ,同时从性能指标对应的传感器得到其实际测量值  $Y_1$ ,计算偏差值  $\Delta Y=|\tilde{Y}_1-Y_1|$ 。

对  $\Delta Y$  及其构成的时间序列的分析如下:

(1) 如果  $\Delta Y$  超过给定的阈值  $\delta$ ,则表明动车组

齿轮箱性能超限,需要进行检修,排除故障或其他因素;

(2) 如果  $\Delta Y$  时间序列发生了突变,则表明动车组齿轮箱有可能发生了故障,需要进行检修和故障排除;

(3) 如果  $\Delta Y$  时间序列产生了趋势,则表明动车组齿轮箱发生了故障或者其性能在比较稳定地下降,需要密切关注并进行故障诊断。

关于齿轮箱性能检测模型的大体构建流程如图 1 所示。

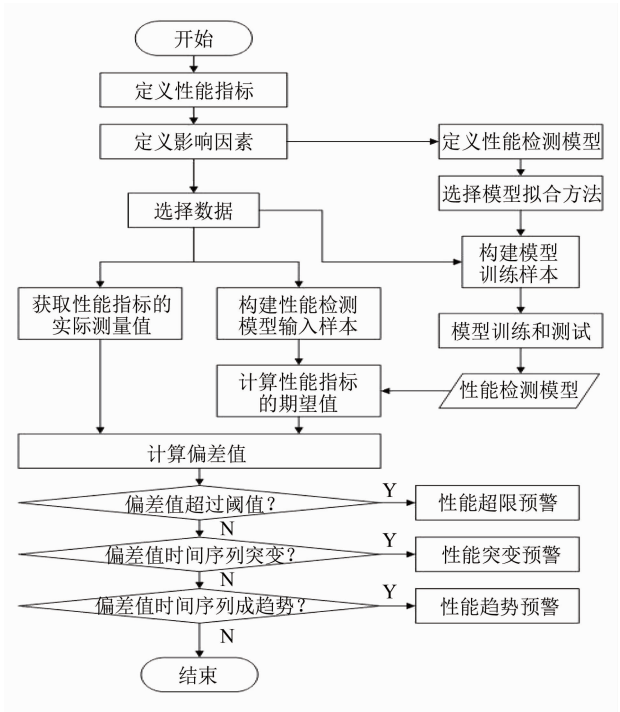


图 1 基于 Lasso 回归的齿轮箱性能检测流程图

### 3 基于 Lasso 回归的动车组齿轮箱性能检测方法实例分析

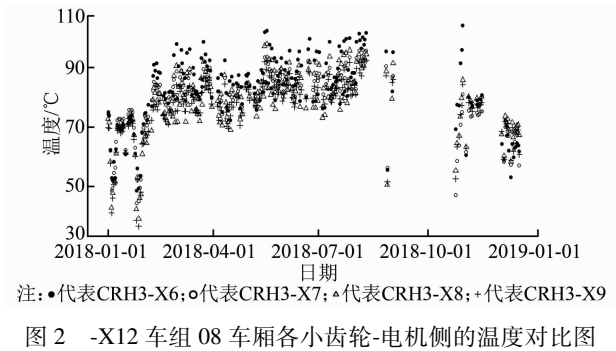
#### 3.1 动车组齿轮箱故障描述

通过对近年来动车组齿轮箱故障信息的统计分析发现,报出故障最多的是小齿轮电机侧温度。本文选择 2018 年 10 月 29 日报出故障信息的-X12 车组(CRH380B 车型)进行模型验证。具体的故障信息为:08 车 1 轴齿轮箱,CRH3-X6 温度最高达 114.67℃,最高温差为 30.71℃。由于齿轮箱及齿轮箱侧半联轴器(VOITH)(一位)温度过高,导致轮对故障,更换轮对。表 2 为-X12 车组 08 车厢 2018 年 1 月 1 日部分参数的数据信息。

表 2 -X12 车组 08 车厢 2018 年 1 月 1 日部分参数数据信息

时刻	温度/℃					
	CRH3-X6	CRH3-X7	CRH3-X8	CRH3-X9	CRH3-X10	CRH3-X11
17:10:17	64.63	71.45	63.49	65.76	239.48	8.70
17:20:18	70.31	72.59	66.90	66.90	114.23	8.30
17:21:18	68.04	70.31	65.76	64.63	70.98	7.80
17:26:19	58.94	61.22	57.80	55.53	65.66	7.40
17:27:20	58.94	62.35	57.80	56.67	130.31	7.40
17:28:21	60.08	65.76	57.80	58.94	196.75	7.50

为了更好地进行建模分析,采用本文 2.1 节中的数据选择方式对建模数据进行了处理,图 2 为 -X12 车组 08 车厢经过特征提取之后的数据。



由图 2 可见,-X12 车组 08 车厢电机 1 小齿轮电机侧温度(CRH3-X6)从 2018 年 2 月份月上旬开始就明显高于同车厢同侧其他点位的温度值,但是 -X12 车组直至 2018 年 10 月 29 日才报出故障。由此可见,目前动车组采取的齿轮箱温度报警方式仅侧重于高温报警,不能进行早期检测。

### 3.2 齿轮箱性能检测模型构建

鉴于-X12 车组 08 车厢 CRH3-X6 从 2 月份开始就明显高于同车厢同部件类型同侧其他点位的温度值,在此选取 2018 年 1 月份 CRH3-X6 及其影响因素处于正常状态下的数据作为模型的训练数据,构建基于 Lasso 回归-X12 车组 08 车厢小齿轮的性能检测模型;并通过 9 折交叉验证,绘制均方误差 ( $M_{SE}$ ) 随回归参数  $\log \lambda$  的变化图,如图 3 所示。

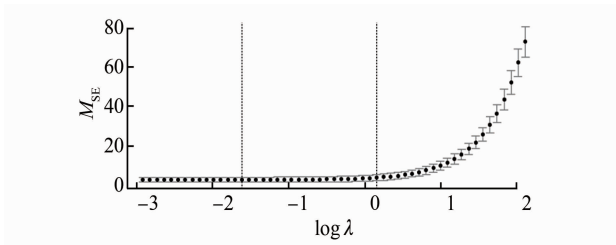


图 3  $M_{SE}$  随  $\log \lambda$  的变化关系图

由输出的结果可知:当  $\lambda_{\min} = 0.054\ 254\ 45$ ,即

$\log \lambda_{\min} = -2.914\ 07$  (图 3 中左侧虚线) 时,  $M_{SE}$  达到最小值;当  $\lambda_{1se} = 1.065\ 037$ , 即  $\log \lambda_{1se} = 0.063\ 009\ 4$  (图 3 中右侧虚线) 时,  $\lambda_{1se}$  即表示一个标准误差内的最佳  $\lambda$  值。一般情况下,最终输出的最佳  $\lambda$  值就是一个标准误差内最简洁的模型的系数。结合实际情况及其建模原理,本模型采用  $M_{SE}$  处于最小值时所对应的模型系数。系数不为 0 的自变量被选为齿轮箱正常行为模型的影响因素。

### 3.3 故障预警实例验证

为探究上述模型的性质以及对齿轮箱性能检测的效果,本文将-X12 车组 08 车厢 2018 年 2 月份至 12 月份的数据作为模型验证数据,在 Windows 7 操作系统、Intel 处理器 3.30 GHz、8.00 GB 内存、R-3.2.5 试验环境下,对本文 3.2 节构建的齿轮箱性能检测模型进行检验,模型预测结果如图 4 所示。

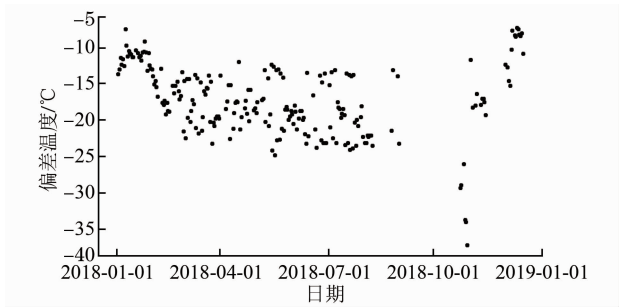


图 4 -X12 车组 08 车厢电机 1 小齿轮温度-电机侧温度偏差时序图

结合本文 2.2 节中的偏差值预警方式可知:偏差值  $\Delta Y$  (为负值,是因为本文采用  $\Delta Y = \hat{Y}_1 - Y_1$ ) 从 2 月份开始有明显的下降趋势,并且在 10 月末发生突变;初步认为-X12 车组 08 车厢电机 1 侧小齿轮的性能在 2018 年 2 月份的时候开始衰退。从图 2 可以看出,2018 年 9 月份并未采集到数据,有可能是动车组对该部位齿轮箱进行了大修,大修之后部件之间的磨合导致了 10 月末的突变。

为进一步降低预警之后的检修成本,同时结合上述故障验证结果,建议将该部件性能检测预警方式按照短、中、长期预警的形式进行分类。短期预警主要针对突变、超限预警,该类预警在短时间内的数据会呈现出明显异常,可以结合检修人员以往的检修经验,给予及时有效的处理;中期预警(可按月/季度预警)和长期预警(可按年预警)主要是对部件的性能趋势进行监测,部件的性能退化程度实时地呈现给检修人员,使检修人员对该部件的健

(下转第 135 页)

的优化,比如选择在地板局部进行拓扑优化。

## 5 结论

1) 本文通过模态贡献量分析找出对车内噪声贡献最大的模态,并由此发现场点附近的地板振动最为剧烈。在仅考虑结构噪声时,场点侧端部地板对于转向架上方场点的声压级贡献量最大。

2) 在 80 Hz 频率附近,仿真和实测的声压级均出现峰值。仿真分析数据与实测数据的对比表明,车体板件辐射噪声对转向架上方场点处的噪声具有一定影响。

3) 将车体场点侧端部地板厚度降低 2 mm 后,场点 43 Hz 和 82 Hz 频率处的线性声压级均降低了 6 dB 以上,并且在 0~200 Hz 大部分频段内噪声声压级均有所下降,该优化方案使场点处结构噪声水平得到了改善。

## 参考文献

[1] 何理,张立民.基于静态台架试验的车体模态贡献量分析[J].

噪声与振动控制,2018(4): 96.

- [2] 李增刚,詹福亮.声学仿真计算高级应用实例[M].北京:国防工业出版社,2010: 214-215.
- [3] LIM T C. Automotive panel noise contribution modeling based on finite element and measured structural-acoustic spectra[J]. Applied Acoustics, 2000(4): 505.
- [4] 梁静.车体模态贡献分析及其对振动影响研究[D].重庆:西南交通大学,2011.
- [5] 潘震,方宇,胡定玉,等.A 型地铁车辆车内噪声预测与控制[J].城市轨道交通研究,2018(8): 67.
- [6] 王晋鹏,常山,刘更,等.结合模态声学贡献量与板面声学贡献量的减速箱降噪技术研究[J].振动与冲击,2016(4): 210.
- [7] 宋士轲.中空挤压铝型材的等效模型研究[D].北京:北京交通大学,2014.
- [8] 张克姝.地铁 A 型车内装结构对车内噪声的影响分析[J].城市轨道交通研究,2012(12): 50.
- [9] 张磊.高速列车室内低频噪声的预测与控制[D].大连:大连交通大学,2008.
- [10] 庞剑.汽车车身噪声与振动控制[M].北京:机械工业出版社,2015.

(收稿日期:2019-06-12)

(上接第 130 页)

康状况能够有更加清晰、全面的了解,能够对部件的早期故障进行及时处理,降低部件出现重大损坏的可能性,从而进一步减少齿轮箱故障对动车组正常运营的影响,保障高速动车组能够安全可靠地运行。

经过故障统计分析,发现-X12 车组 2018 年关于齿轮箱的故障仅有 10 月 29 日一条故障记录,由此可以得知,本文提出的基于 Lasso 回归的高速动车组齿轮箱性能检测模型能够很好地对动车组齿轮箱进行早期故障预警,可从一定程度上解决动车组齿轮箱早期故障检测及其性能量化评估问题。

## 4 结语

本文将 Lasso 回归用于检测高速动车组齿轮箱性能的相关问题,根据齿轮箱的具体特性及其工作原理,设计了解决齿轮箱性能检测的方法。Lasso 回归采用罚函数的方式对目标函数进行简化求解,能够更加快速、有效地挑选出性能指标合理的影响因素,去除其他噪声影响因素。试验结果显示出了

基于 Lasso 回归的动车组齿轮箱性能检测模型的可能性、有效性和优越性,研究结论也为动车组齿轮箱性能检测的方式提供了试验基础和探究思路。

## 参考文献

- [1] 万国强,林建辉,易彩.高速列车齿轮箱振动特性分析与故障识别方法[J].机械科学与技术,2018(1): 115.
- [2] 吴欣远,林建辉.高速列车齿轮箱振动信号诊断研究[J].硅谷,2015(3): 55.
- [3] 刘杰,刘世军.基于热网络法的高速列车齿轮箱热平衡温度计算分析[J].机车电传动,2018(3): 70.
- [4] 陈晓玲,张武高,黄智勇,等.高速列车运行速度对传动齿轮箱平衡温度的影响[J].上海交通大学学报,2007(9): 1510.
- [5] 魏领会,王海龙,姚灿江.环境温度对高速列车齿轮箱温度场的影响[J].北方工业大学学报,2017(5): 63.
- [6] 宇善良.动车组常见故障分析及处理[M].北京:北京交通大学出版社,2014.
- [7] TIBSHIRANI R. Regression shrinkage and selection via the Lasso[J]. Journal of the Royal Statistical Society, 1996, 58: 267.
- [8] 游皓麟.R 语言预测实战[M].北京:电子工业出版社,2016.

(收稿日期:2019-10-25)