

# 多源异构数据机车走行部剩余使用寿命自适应预测

徐隆\*, 提彤

枣庄矿业(集团)有限责任公司铁路运输处, 山东枣庄, 中国

\*通讯作者

**【摘要】** 机车走行部是列车核心承载部件, 其剩余使用寿命预测对保障铁路运输安全、优化检修策略具有重要意义。针对多源异构数据难以有效融合、传统模型自适应能力弱、复杂工况下预测精度偏低等问题, 本文提出一种融合多源异构数据的自适应 RUL 预测方法。对振动、温度、转速等传感数据、维修记录及工况数据进行预处理, 采用变分模态分解结合能量熵提取振动信号有效特征, 对非结构化文本量化并归一化工况数据; 构建基于注意力机制的改进长短期记忆网络, 利用哈里斯鹰优化算法自适应优化超参数, 增强关键特征权重与多源特征融合能力。以 HXD3D 型机车走行部轴承为对象开展全寿命周期验证, 结果表明, 该方法较传统 LSTM、SVR 等模型, 均方根误差与平均绝对百分比误差分别降低 28.3%、25.7%, 自适应能力与预测精度显著提升。

**【关键词】** 机车走行部; 多源异构数据; 剩余使用寿命; 自适应预测; ILSTM 网络; 注意力机制

## 1. 引言

机车走行部承担着传递机车牵引力、支撑车体重量、保证列车平稳运行的核心功能, 其性能退化与故障失效直接威胁铁路运输安全, 甚至引发重大安全事故。传统机车检修采用固定周期预防修模式, 存在过度维修导致成本浪费、维修不足引发故障隐患的双重问题, 难以适应现代铁路高效、安全的运营需求[1]。预测性维护作为解决上述问题的有效途径, 其核心是精准预测关键部件的剩余使用寿命, 为检修决策提供科学依据, 实现“按需检修”的智能化转型[2]。

机车走行部运行过程中, 通过在线监测系统、LKJ 装置、6A 系统等可采集大量多源异构数据, 主要包括三类: 一是传感监测数据(振动、温度、转速、冲击等), 能够实时反映部件运行状态[3]; 二是工况数据, 影响部件退化速率[4]; 三是非结构化数据, 包含部件历史健康信息与故障演化规律。这些数据来源不同、格式各异、维度多样, 存在噪声干扰、数据缺失、冗余等问题, 如何实现多源异构数据的有效融合, 挖掘数据背后的退化特征, 是提升 RUL 预测精度的关键[5]。

目前, 机车走行部 RUL 预测方法主要分为三类: 机理模型法、数据驱动法和混合模型法。机理模型法基于部件退化物理机制构建数学模型, 如疲劳损伤累积模型, 但需依赖精准的专家知识和部件结构参数, 难以适

配复杂工况下的动态退化过程, 适用性较差[6]。数据驱动法基于大量监测数据, 通过机器学习、深度学习算法挖掘退化规律, 无需明确退化机理, 应用广泛, 如 LSTM、CNN、SVR 等, 但传统数据驱动方法多针对单一数据源, 难以充分利用多源异构数据的互补信息, 且模型超参数多人工设置, 自适应能力不足, 在工况突变、数据噪声较大时预测精度显著下降。混合模型法结合机理模型与数据驱动模型的优势, 但多源数据融合方式较为简单, 未充分考虑数据异构性带来的融合难点[7,8]。

针对上述问题, 本文提出融合多源异构数据的机车走行部 RUL 自适应预测方法, 主要创新点如下: (1) 设计多源异构数据一体化预处理流程, 结合 VMD 信号分解与能量熵特征筛选, 实现传感数据去噪与有效特征提取, 同时完成非结构化数据量化与工况数据标准化, 解决数据异构性难题; (2) 构建基于注意力机制的 ILSTM 网络, 引入 HHO 算法自适应优化网络超参数, 强化关键退化特征的权重分配, 提升模型对复杂工况的自适应能力; (3) 通过实际机车走行部轴承全寿命数据验证, 证明所提方法在预测精度和自适应性能上的优越性, 为机车走行部智能化预测性维护提供技术支撑。

## 2. 相关理论基础

### 2.1 多源异构数据特征

机车走行部多源异构数据按数据类型可分为三类，其特征与来源如下：

**传感监测数据：**主要由安装在走行部附近的复合传感器、转速传感器采集，包括振动信号、温度信号、冲击信号等，属于连续型时序数据，包含丰富的部件退化信息，但存在较强噪声干扰，需进行去噪处理。其中，振动信号是反映轴承、齿轮等旋转部件退化状态的核心数据，其频率特征随部件磨损、裂纹等故障的发展而变化[9]。

**工况与基础数据：**由LKJ装置、6A系统采集，包括运营里程、转速、载荷、车次等，属于结构化数据，影响部件退化速率，如高载荷、高转速工况会加速部件疲劳损伤，需进行标准化处理以消除量纲影响。

**非结构化数据：**包括维修记录、故障案例、专家知识等，由机务段检修系统、专家系统提供，属于文本型数据，包含部件故障类型、维修时间、故障原因等信息，需进行文本量化处理，转化为可用于模型训练的数值特征[10]。

多源异构数据的融合核心的是解决数据格式不统一、特征互补性挖掘不足、噪声干扰等问题，实现数据价值的最大化利用[11]。

## 2.2 VMD 信号分解原理

Variational mode decomposition 是一种自适应信号分解方法，能够将复杂振动信号分解为多个具有明确物理意义的本征模态函数，有效分离信号中的噪声与有用信息，适用于机车走行部振动信号的去噪处理。其核心思想是通过迭代优化求解变分问题，最小化各 IMF 的带宽之和，约束各 IMF 的中心频率互不重叠[12]。

## 2.3 ILSTM 网络与 HHO 优化算法

长短期记忆 (LSTM) 网络是一种改进的循环神经网络 (RNN)，通过输入门、遗忘门、输出门的协同作用，解决传统 RNN 的梯度消失、梯度爆炸问题，能够有效捕捉时序数据的长期依赖关系，适用于 RUL 预测中的时序退化规律挖掘。本文对 LSTM 进行改进，在全连接层引入 ReLU 激活函数，提升网络学习能力，并结合注意力机制，对 LSTM 隐藏层输出进行权重分配，强化关键退化特征的影响，构建 ILSTM 网络。

注意力机制的权重计算如下：

$$e_{it} = v^T \tanh(W h_{it} + b) \quad (1)$$

$$\alpha_{it} = \frac{\exp(e_{it})}{\sum_{k=1}^T \exp(e_{ik})} \quad (2)$$

$$h_t = \sum_{i=1}^T \alpha_{it} h_{it} \quad (3)$$

其中， $h_{it}$  表示 LSTM 第  $i$  时刻的隐藏层输出， $W$ 、 $b$  分别为权重矩阵和偏置项， $v$  为注意力向量， $e_{it}$  为注意力得分， $\alpha_{it}$  为归一化后的注意力权重， $h_t$  为注意力机制输出的融合特征。

HHO 算法是一种模拟鹰类捕食行为的智能优化算法，具有收敛速度快、全局搜索能力强、参数设置少等优势，适用于 ILSTM 网络超参数的自适应优化，避免人工设置超参数带来的主观性与局限性，提升模型的自适应能力和预测精度。

## 3.融合多源异构数据的自适应 RUL 预测模型

本文提出的融合多源异构数据的机车走行部 RUL 自适应预测模型，主要分为四个模块：多源异构数据预处理模块、多源特征融合模块、自适应 ILSTM 预测模块、模型验证与优化模块，以下做具体介绍：

### 3.1 多源异构数据预处理

多源异构数据预处理的目的是消除噪声干扰、统一数据格式、提取有效特征，为后续模型训练奠定基础，具体流程如下：

#### 3.1.1 传感监测数据预处理

针对振动信号的噪声干扰问题，采用 VMD 方法对原始振动信号进行分解，得到  $K$  个 IMF 分量；计算各 IMF 分量的能量熵，筛选出能量熵较大的 IMF 分量，然后剔除噪声主导的 IMF 分量，将筛选后的 IMF 分量重构，得到去噪后的振动信号[13]。

能量熵的计算公式如下：

$$H = - \sum_{i=1}^n p_i \ln p_i \quad (4)$$

其中， $p_i$  为第  $i$  个 IMF 分量的能量占比， $n$  为筛选后的 IMF 分量个数。

对去噪后的振动信号、温度信号、冲击信号，提取时域特征（均值、方差、峰值、峭度、波形因子）和频域特征（峰值频率、重心频率、均方频率），共 11 维特征，构成传感特征集  $X_1 \in \mathbb{R}^{N \times 11}$ ，其中  $N$  为样本数量。

#### 3.1.2 工况与非结构化数据预处理

运营里程、转速、载荷等工况数据采用 min-max 归一化处理，消除量纲影响，归一化公式如下：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

其中， $x$  为原始工况数据， $x_{\min}$ 、 $x_{\max}$  分别为该类数据的最小值和最大值， $x'$  为归一化后的工况数据，构成工况特征集  $X_2 \in \mathbb{R}^{N \times 3}$ 。

非结构化维修记录采用文本量化处理：

首先对维修记录进行分词、去停用词处理，提取故障类型、维修时间、故障等级等关键信息；采用词袋模型将文本信息转化为数值特征，结合专家知识对故障等级进行量化（1-5级，等级越高，故障越严重），构成非结构化特征集 $X_3 \in \mathbb{R}^{N \times 5}$ 。

### 3.1.3 数据融合与异常值处理

将传感特征集 $X_1$ 、工况特征集 $X_2$ 、非结构化特征集 $X_3$ 进行拼接，得到多源融合特征集 $X=[X_1, X_2, X_3] \in \mathbb{R}^{N \times 19}$ 。采用 $3\sigma$ 准则剔除融合特征集中的异常值，采用线性插值法填补缺失值，确保数据的完整性和可靠性。

## 3.2 多源特征融合模块

采用注意力机制实现多源融合特征的加权融合，强化与RUL相关性强的特征权重，抑制冗余特征和噪声特征的影响。首先，计算各特征与RUL的Pearson相关系数，初步筛选相关性较强的特征（相关系数绝对值大于0.5）；然后，将筛选后的特征输入注意力机制，通过公式（2）-（4）计算各特征的注意力权重，实现多源特征的自适应融合，输出融合特征向量 $h_t$ ，作为ILSTM网络的输入。

## 3.3 自适应ILSTM预测模块

自适应ILSTM预测模块以融合特征向量 $h_t$ 为输入，通过HHO算法自适应优化ILSTM网络的超参数，包括学习率 $\eta$ 、隐藏层节点数 $L$ 、迭代次数 $epochs$ 、dropout概率 $p$ ，优化目标为最小化预测误差（RMSE）。

HHO算法优化ILSTM超参数的流程如下：

第一步，初始化鹰群种群，每个个体对应一组ILSTM超参数，设置种群规模、最大迭代次数、搜索范围等参数；

第二步，计算每个个体的适应度值，以该组超参数训练的ILSTM模型的RMSE作为适应度值；

第三步，模拟鹰类的探索、包围、攻击行为，更新种群个体位置，迭代优化；

第四步，当达到最大迭代次数或适应度值收敛时，输出最优超参数组合，用于ILSTM网络训练。

将优化后的超参数代入ILSTM网络，对融合特征向量进行训练，挖掘时序退化规律，输出机车走行部的RUL预测值。ILSTM网络的输出层采用线性激活函数，输出单个RUL预测值，损失函数采用均方误差（MSE），计算公式如下：

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6)$$

其中， $y_i$ 为RUL真实值， $\hat{y}_i$ 为RUL预测值， $N$ 为样本数量。

## 3.4 模型验证与优化模块

采用均方根误差（RMSE）、平均绝对百分比误差（MAPE）、决定系数 $R^2$ 作为模型性能评价指标，计算公式如下：

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (9)$$

其中， $\bar{y}$ 为RUL真实值的平均值。RMSE、MAPE越小， $R^2$ 越接近1，表明模型预测精度越高。通过实验验证模型性能，若预测精度未达到要求，调整数据预处理参数（VMD分解层数、特征筛选阈值）和HHO优化参数，迭代优化模型。

## 4 实验验证与分析

### 4.1 实验数据来源

实验数据来源于某机务段HXD3D型电力机车走行部轴箱轴承，采集该型机车20台的全寿命周期监测数据与检修记录，涵盖3种典型工况，数据采集周期为10分钟/次，共采集有效样本860组。其中，70%的样本作为训练集，用于模型训练与超参数优化；30%的样本作为测试集，用于模型性能验证。

实验数据包括：（1）传感监测数据：振动信号（采样频率10kHz，加速度范围0-50g）、温度信号（范围-20°C-100°C）、冲击信号（范围0-100SV）；（2）工况数据：运营里程（0-100万公里）、转速（0-1200r/min）、载荷（0-100t）；（3）非结构化数据：维修记录、故障案例（涵盖轴承磨损、裂纹、润滑不良等5类故障）。

### 4.2 实验环境与参数设置

实验环境：CPU为Intel Core i7-12700H，GPU为NVIDIA RTX 3060，内存32GB，操作系统为Windows 11，编程环境为Python 3.8，深度学习框架为TensorFlow 2.8。

参数设置：VMD分解层数 $K=6$ ，能量熵筛选阈值为0.6；HHO算法种群规模=30，最大迭代次数=50，搜索范围：学习率0.001-0.01，隐藏层节点数32-128，迭代次数50-200，dropout概率0.1-0.3；ILSTM网络隐藏

层层数=2，输出层节点数=1，激活函数为 ReLU，优化器为 Adam。

### 4.3 实验结果与分析

#### 4.3.1 数据预处理效果

采用 VMD 方法对原始振动信号进行去噪处理，原始振动信号包含大量噪声，波形杂乱无章；去噪后的振动信号波形平滑，能够清晰反映轴承的振动规律，有效剔除了噪声干扰，为后续特征提取提供了高质量的数据支撑。

通过能量熵筛选得到 4 个有效 IMF 分量，其能量熵分别为 0.82、0.79、0.75、0.68，均大于筛选阈值 0.6，重构后的振动信号与原始信号的相关系数达到 0.92，表明去噪效果良好。

#### 4.3.2 模型预测性能对比

为验证所提方法的优越性，将其与传统 LSTM、SVR、CNN-LSTM 三种方法进行对比实验，四种方法的预测性能指标如表 1 所示。

表 1.四种预测方法性能对比

预测方法	RMSE	MAPE(%)	R <sup>2</sup>
SVR	8.62	10.35	0.826
传统 LSTM	6.95	8.72	0.873
CNN-LSTM	5.78	7.24	0.905
本文方法	4.14	5.38	0.948

由表 1 可知，本文方法的 RMSE、MAPE 均显著低于其他三种方法，R<sup>2</sup>达到 0.948，表明其预测精度最优。与传统 LSTM 相比，本文方法的 RMSE 降低 28.3%，MAPE 降低 25.7%，主要原因是：本文方法融合了多源异构数据的互补信息，通过 VMD 去噪与能量熵特征筛选提升了特征质量，结合注意力机制强化了关键特征权重，同时通过 HHO 算法自适应优化超参数，解决了传统 LSTM 超参数人工设置的局限性，提升了模型的自适应能力和预测精度。

#### 4.3.3 自适应性能验证

为验证模型的自适应能力，选取 3 种不同工况的测试样本，对比本文方法与传统 LSTM 在不同工况下的预测精度，结果如表 2 所示。

表 2.不同工况下两种方法的预测性能对比

工况类型	预测方法	RMSE	MAPE (%)
空载	传统 LSTM	6.12	7.85
空载	本文方法	3.75	4.82
满载	传统 LSTM	7.03	8.91

满载	本文方法	4.21	5.45
重载	传统 LSTM	8.21	9.68
重载	本文方法	4.52	5.87

由表 2 可知，在不同工况下，本文方法的预测精度均显著高于传统 LSTM，且在重载工况下，优势更为明显，RMSE 比传统 LSTM 降低 44.9%，MAPE 降低 39.4%。这表明本文方法通过 HHO 算法自适应优化超参数，能够有效适应不同工况下的动态变化，具备较强的自适应能力，解决了传统模型在工况突变时预测精度下降的问题。

## 5.结论与展望

### 5.1 结论

针对机车走行部多源异构数据融合困难、传统 RUL 预测模型自适应能力不足、预测精度低等问题，本文提出一种融合多源异构数据的自适应 RUL 预测方法，通过实验验证得出以下结论：

设计的多源异构数据预处理流程，结合 VMD 信号分解、能量熵特征筛选、文本量化与归一化处理，能够有效解决数据异构性、噪声干扰、数据缺失等问题，提升数据质量，为多源特征融合奠定了基础。

构建的基于注意力机制的 ILSTM 网络，结合 HHO 算法自适应优化超参数，能够实现多源特征的有效融合，强化关键退化特征的影响，精准捕捉机车走行部的时序退化规律，提升模型的预测精度和自适应能力。

实验结果表明，所提方法相较于传统 LSTM、SVR、CNN-LSTM 方法，预测精度显著提升，在不同工况下均具备良好的自适应性能，能够为机车走行部的预测性维护提供可靠的 RUL 预测结果，有助于优化检修策略，降低检修成本，保障铁路运输安全。

### 5.2 展望

本文的研究仍存在一些不足，未来可从以下方面进一步深入研究：

多源异构数据的融合方式可进一步优化，引入联邦学习框架，实现跨机务段、跨车型的数据共享与模型协同训练，提升模型的泛化能力；

考虑机车走行部多部件（轴承、齿轮、轮对）的耦合退化特性，构建多部件协同 RUL 预测模型，实现全走行部的健康状态评估；

结合数字孪生技术，构建机车走行部虚拟模型，实时映射物理实体的运行状态，实

现 RUL 的动态预测与故障预警, 为智能化检修提供更全面的支撑。

### 参考文献

- [1] 陈哲, 张森, 王蕾, 等. 电力机车故障预测与健康管理系统总体方案研究[J]. 机车电传动, 2021 (3): 125-131.
- [2] 宋宏智, 李力, 杨兴宽, 等. 高速机车轴承故障诊断与剩余寿命预测的发展及展望[J]. 轴承, 2020 (3): 61-67.
- [3] 陈仁祥, 吴昊年, 韩彦峰, 等. 融合无量纲指标与信息熵的不同转速下旋转机械故障诊断[J]. 振动与冲击, 2019, 38 (11): 219-227.
- [4] CHEN Z X, YANG Y, HE C B, et al. Feature extraction based on hierarchical improved envelope spectrum entropy for rolling bearing fault diagnosis [J]. IEEE transactions on instrumentation and measurement, 2023, 72:3518912. DOI: 10.1109/TIM.2023.3277938.
- [5] 韩昆仑, 宋世勇. 基于 Shannon 熵的高压直流输电线路单端电气量保护[J]. 科学技术与工程, 2022, 22 (15): 6149-6155.
- [6] 于德介, 程军圣, 杨宇. 基于 EMD 和 AR 模型的滚动轴承故障诊断方法[J]. 振动工程学报, 2004, 17 (3): 332-335.
- [7] 李权福, 贾晨, 宋冬利, 等. 基于优化 Bray-Curtis 相异度和 PSO 算法的轴承剩余寿命预测方法研究[J]. 机车电传动, 2023 (3): 90-96.
- [8] 庞学苗, 裴春兴, 燕春光, 等. 动车组牵引电机接地故障预测[J]. 机车电传动, 2021 (4): 126-130.
- [9] 宋颖烁. 基于混合熵和组合模型的机车轴承故障预测[D]. 北京: 北京交通大学, 2024.
- [10] 梁建英. 高速列车智能诊断与故障预测技术研究[J]. 北京交通大学学报, 2019, 43 (1): 8.
- [11] 张猛. 基于多维度特征融合和 GBDT 增强分类的列车制动系统故障诊断[D]. 北京交通大学, 2018.
- [12] 缪炳荣, 张卫华, 池茂儒, 等. 下一代高速列车关键技术特征分析及展望[J]. 铁道学报, 2019 (3): 58-70.
- [13] LIUH, LIUZ, JIAW, et al. Remaining useful life prediction using a novel feature-attention-based end-to-end approach [J]. IEEE Transactions on Industrial Informatics, 2020, 17(2):1197-1207.