

基于大模型的柴油机装配冷试数据分析

赵旭辉¹, 徐卓¹, 王相成², 闫伟^{1*}, 李国祥¹

1. 山东大学核科学与能源动力学院, 山东 济南 250061; 2. 山东浪潮傲林大数据科技有限公司, 山东 济南 250101

摘要:为提高柴油机装配质量和冷试性能,基于柴油机装配冷试基础数据集,选取加州大学欧文分校(University of California Irvine, UCI)机器学习资料库标准数据集中的 Seeds、Wine、Wdbc 三种数据集,对比支持向量机(support vector machines, SVM)模型、组合智能算法改进后 SVM 模型、Transformer 模型应用于冷试异常数据的分析效果。结果表明:SVM、改进后 SVM、Transformer 模型对正常数据和异常数据分类的准确率分别为 85.20%、92.54%、97.94%;相比 SVM、改进 SVM 模型,Transformer 模型的分类准确率有较大的提高,可用于分析参数异常;排气压力与转矩关系密切,排气压力较大造成转矩增大;排气门开启时间过长导致进气真空度异常,验证了 Transformer 模型对发动机装配异常识别方法的有效性。

关键词:柴油机装配;冷试;异常检测;SVM;Transformer 模型架构

中图分类号:TK422

文献标志码:A

文章编号:1673-6397(2026)01-0062-08

引用格式:赵旭辉,徐卓,王相成,等.基于大模型的柴油机装配冷试数据分析[J].内燃机与动力装置,2026,43(1):62-69.

ZHAO Xuhui, XU Zhuo, WANG Xiangcheng, et al. Data analysis of cold test for diesel engine assembly based on large models[J]. Internal Combustion Engine & Powerplant, 2026, 43(1): 62-69.

0 引言

为了保证柴油机装配质量,需在装配完成后进行性能测试。柴油机出厂前的性能检测分为热试和冷试,早期柴油机普遍采用热试技术,即通过通气、加油等操作,模拟柴油机在真实工作环境下的运行状况,测试不同工况的性能指标,并根据检测结果判断参数是否符合要求^[1]。虽然热试通用性较好,但存在测试周期长、效率低、油耗高等问题。相较于热试,冷试过程中,柴油机无需点火,由电机倒拖运行,该测试方法具有短耗时、零油耗、低噪声、零污染等特点,更能满足节能减排需求,有效解决了传统热试技术耗时长、排放高等问题,已成为当前主流的出厂检测方式。

分析冷试数据有助于提高发动机装配质量。冷试数据庞大且参数种类繁多,大数据技术可对海量、多维的测试数据进行深度耦合分析,已在测试数据处理中得到广泛应用。Peinado-asensi 等^[2]提出了 Criterion-360 评估体系,在分析冲压车间的海量数据时效率更高、成本更低;Li 等^[3]建立了基于数字孪生技术的装配质量评估方法,实现了航空发动机装配全过程的动态建模与误差分析,从而提高了装配质量评估的准确性与全面性;Wang 等^[4]提出了一种融合机械臂、异常检测与数字孪生的自动化装配缺陷检测系统,显著提高了柴油机装配检测的效率与准确性;Cui 等^[5]提出了一种基于 Transformer 模型自注意力

收稿日期:2025-09-30

基金项目:山东省工程机械智能装备创新创业共同体项目(GTT20240101);济南市科技型中小企业创新能力提升工程项目;山东省土方机械智慧施工技术重点实验室项目(PKL2024F13)

第一作者简介:赵旭辉(2002—),男,山西长治人,硕士研究生,主要研究方向为发动机冷试技术,E-mail:202434567@mail.sdu.edu.cn。

* **通信作者简介:**闫伟(1973—),男,济南人,工学博士,教授,博士生导师,主要研究方向为新能源汽车技术、发动机冷试技术等,E-mail:yanwei@sdu.edu.cn。

的对比学习轴承故障诊断方法,利用未标记数据进行自监督学习;Abdelrahman 等^[6]通过集成多种异常检测方法有效识别影响装配质量的关键变量,显著降低了废品率;Yun 等^[7]通过改进支持向量机(support vector machines, SVM)和构造故障特征集,提出了一种高识别精度的发动机故障分类器;向晓良等^[8]基于改进的 SVM,提出一种能在强噪声干扰下保持高准确率的轴承振动故障识别方法;王凯旋等^[9]基于改进 Transformer 模型的神经网络,提出了一种提升制动系统故障诊断的方法;Chen 等^[10]基于 Transformer 模型提出了一种多尺度注意力转换器模型,该模型显著提高了船用柴油机在复杂噪声环境下故障诊断的精度与鲁棒性;Wu 等^[11]提出了一种基于 Transformer 模型分类器,该分类器不仅能进行故障检测,还可有效识别故障类型和严重程度。

本文中基于加州大学欧文分校(University of California Irvine, UCI)机器学习资料库标准数据集,对比 SVM 模型、改进 SVM 模型、Transformer 模型分类效果,验证 Transformer 模型的有效性;针对柴油机冷试异常数据,采用 3 种模型分别对装配冷试数据集中的正常点和异常点进行分类分析,验证 Transformer 模型识别柴油机装配冷试数据异常的可行性。

1 冷试测试数据库构建

将大数据技术应用于柴油机装配冷试测试技术,需首先对采集到的冷试数据进行预处理,包括删除由于传感器异常等导致的无效数据、确定数据的合理控制阈值等,完成数据集的初步分析。

1.1 柴油机测试参数

冷试中离线数据分析系统主要测试起动转矩、排气压力、进气、高速振动及低速油压等多种项目,测试过程中柴油机转速为 60~1 500 r/min,具体测试参数为:最大曲轴转矩、最大排气压力、排气门开启时刻、最大进气真空度,其中排气门开启时刻以对应的曲轴转角表示。测试界面如图 1 所示。

1.2 数据统计

采用 3 σ 准则确定各参数正态分布阈值,利用 Box-Cox 转换对非正态分布数据进行处理,将其转化为正态分布后进行分析。数据统计结果如表 1 所示,表中 M_{min} 、 M_{max} 、 M_{ave} 、 M_{std} 分别为最大曲轴转矩采集样本的最小值、最大值、平均值、标准差; p_{min} 、 p_{max} 、 p_{ave} 、 p_{std} 分别为最大排气压力采集样本的最小值、最大值、平均值、标准差, θ_{min} 、 θ_{max} 、 θ_{ave} 、 θ_{std} 分别为排气门开启时刻采集样本的最小值、最大值、平均值、标准差, β_{min} 、 β_{max} 、 β_{ave} 、 β_{std} 分别为最大进气真空度采集样本的最小值、最大值、平均值、标准差。

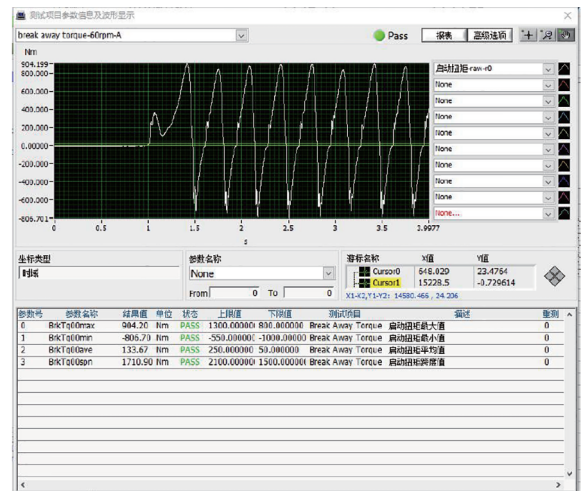


图 1 离线数据分析系统测试界面

表 1 数据统计结果

气缸 序号	M_{min} / (N·m)	M_{max} / (N·m)	M_{ave} / (N·m)	M_{std} / (N·m)	p_{min} / kPa	p_{max} / kPa	p_{ave} / kPa	p_{std} / kPa	θ_{min} / (°)	θ_{max} / (°)	θ_{ave} / (°)	θ_{std} / (°)	β_{min} / kPa	β_{max} / kPa	β_{ave} / kPa	β_{std} / kPa
1	716.7	1 060.7	891.3	55.1	701.2	838.5	777.6	20.7	259.8	281.3	272.2	2.9	-63.8	-36.4	-50.1	4.2
2	670.2	1 017.4	845.5	54.9	708.6	841.5	786.2	20.7	259.8	281.6	271.5	2.8	-63.5	-36.1	-50.2	4.4
3	690.1	1 045.9	872.3	58.3	707.2	845.4	783.8	21.1	258.4	281.3	271.3	2.8	-65.2	-35.5	-50.2	4.6
4	732.1	1 081.4	901.3	58.1	713.8	846.2	788.2	19.8	258.1	282.7	271.4	2.9	-64.4	-35.3	-50.2	4.6
5	694.1	1 039.1	869.1	55.3	703.3	841.7	783.5	21.9	257.3	282.3	271.2	2.9	-63.5	-35.9	-50.3	4.6
6	694.4	1 022.8	861.5	54.0	711.1	850.9	787.2	21.6	256.9	281.25	270.9	2.9	-67.0	-35.8	-52.1	4.9

采用智能算法改进基于密度的聚类算法(density-based spatial clustering of applications with noise, DBSCAN),对处理后的柴油机出厂数据进行聚类分析,将数据分为正常数据和异常数据两类,所有数据分析结果如图2所示。各项测试参数异常数据点共777个,异常数据分布如表2所示。

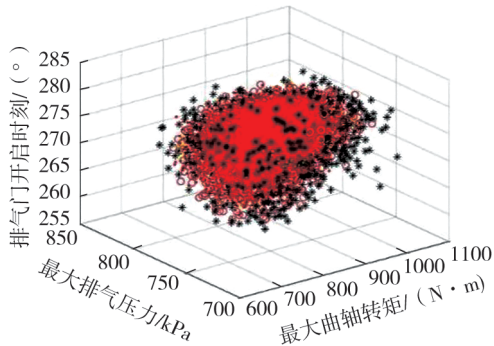


图2 数据分析结果

表2 异常数据点分布

气缸序号	异常数据数	占比/%
1	129	16.60
2	105	13.51
3	96	12.36
4	150	19.31
5	122	15.70
6	175	22.52

由图2、表2可知:柴油机6个缸测试参数异常点的数量分布不均匀,第6缸异常点占比最大,第3缸异常点占比最小,表明柴油机6个缸工作不均匀。为此,需在柴油机制造过程中建立有效的参数管控机制,实现对异常问题的及时预警与处理。

2 基于分类的大数据技术分析与应用

2.1 大数据技术分析

采用大数据技术分类方法对异常数据进行分类,常用的大数据技术分类方法有:SVM、改进SVM(KI-SVM)、Transformer模型。SVM是一种二分类模型,常用于分类和线性回归,本文中利用SVM对柴油机冷试测试数据进行分析,以实现异常数据识别^[12];KI-SVM是使用组合智能算法改进的SVM;Transformer模型凭借特有的自注意力机制可以精准捕捉数据之间的关联,且在处理海量、高维和非结构化数据方面展现出显著优势,因而应用比较广泛。

2.1.1 SVM基本原理

SVM的基本思想是基于结构风险最小化原则,寻找一个最优超平面作为决策边界,以使正负样本之间的分类间隔最大化^[13]。若训练集样本库 Y 中的数据可以通过线性函数 $\omega^T x_i + d = 0$ 划分为两类,则该线性函数应满足如下条件:

$$\begin{cases} \omega^T x_i + d \geq 1, y_i = 1 \\ \omega^T x_i + d \leq -1, y_i = -1 \end{cases}, \quad (1)$$

式中: x_i 为 Y 中的样本向量, $i=1,2,\dots,m$, m 为 Y 中的样本数; ω 为超平面的法向量, d 为超平面的截距, y_i 为样本的类别标签,通常为1或-1,代表两个不同的类别。

x_i 到超平面 S 的距离

$$l = |\omega^T x_i + d| / \|\omega\|. \quad (2)$$

若 $|\omega^T x_i + d| = 1$,则对应的数据点 x_j 为支持向量, x_j 到 S 的距离为 $\frac{1}{\|\omega\|}$,SVM目标可定义为:

$$\begin{aligned} \min \|\omega\| / 2, \\ \text{s. t. } y_i(\omega^T x_i + d) \geq 1. \end{aligned} \quad (3)$$

为方便计算,引入修正因子进行修正,最终的目标函数为:

$$\begin{aligned} \min & \frac{\|\boldsymbol{\omega}\|^2}{2} + C \sum_{i=1}^m \xi_i, \\ \text{s. t. } & y_i(\boldsymbol{\omega}^T \mathbf{x}_i + d) \geq 1 - \xi_i, \\ & \xi_i > 0, \end{aligned} \quad (4)$$

式中: C 为惩罚因子, ξ_i 为松弛变量。

将径向基核函数选为 SVM 核函数,核函数表达式如下:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2g^2}\right), \quad (5)$$

式中: g 为核函数带宽参数, $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ 为两个样本之间的相似度。

2.1.2 KI-SVM 基本原理及改进

KI-SVM 算法将 KOADE-IVYA 与 SVM 结合,KOADE-IVYA 在常春藤算法^[14](Ivy algorithm, IVYA)中引入开普勒算法^[15](Kepler optimization algorithm, KOA)的反向搜寻、差分进化(differential evolution, DE)算法^[16]的变异搜寻,并对 IVYA 中每个个体的生长或传播的重要参数 β 进行改进,最终构建而成的混合优化算法。

采用非随机方式确认参数 β ,计算式为:

$$\beta = \left[2 + \exp\left(2\left(\frac{t}{t_{\max}} - 1\right)\right)\right] / 2, \quad (6)$$

式中: t 为当前迭代次数, t_{\max} 为最大迭代次数。

在实际分类过程中,SVM 通过求解凸优化问题确定特征空间中的最优超平面 Q ^[17],此时优化目标函数为:

$$\begin{aligned} \max Q(\boldsymbol{\alpha}) &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{s. t. } & \sum_{i=1}^m \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \end{aligned} \quad (7)$$

式中 α_i 为对应的拉格朗日乘子。

将径向基核函数作为 SVM 核函数,结合式(5)(7),可以明显看出 SVM 与关键参数 C 、 g 密切相关,通过 KOADE-IVYA 对 SVM 核函数进行寻优,得到 KI-SVM,即改进后的 SVM 模型。

2.1.3 Transformer 模型

Transformer 模型是一种由 Vaswani 等^[18]提出的基于自注意力机制的神经网络架构,该模型架构摒弃了传统的循环神经网络(recurrent neural network, RNN)和卷积神经网络(convolutional neural network, CNN)结构,通过完全依赖自注意力机制,实现了模型并行训练的能力。

Transformer 模型结构由做分类任务的编码器和做预测任务的解码器组成,编码器由 N 个相同的层堆叠而成,每一层由一个多头注意力层和一个前馈神经网络层(feed-forward neural network, FFN)组成。

对于单头注意力层,注意力机制 f_A 的计算式为:

$$f_A(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = f_s(\mathbf{Q}\mathbf{K}^T / \sqrt{d_k}) \mathbf{V}, \quad (8)$$

式中: \mathbf{Q} 为查询向量,为输入向量与查询向量权重矩阵 \mathbf{W}_Q 的乘积; \mathbf{K} 为键向量,为输入向量与键向量权重矩阵 \mathbf{W}_K 的乘积; \mathbf{V} 为值向量,为输入向量与值向量权重矩阵 \mathbf{W}_V 的乘积; \mathbf{W}_Q 、 \mathbf{W}_K 、 $\mathbf{W}_V \in \mathbf{R}^{d_m \times d_k}$, f_s 为概率分布归一化函数, d_k 为键向量维度, d_m 为模型的特征向量维度。

Transformer 模型通过添加一种多头注意力机制,进一步完善自注意力层,多头注意力机制示意图3所示。具体做法为:若添加 h 个多头注意力机制,即采用 h 个线性变换对 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 进行函数 f_M 映射,每一组注意力将输入映射到不同的子表示空间,模型关注不同子表示空间中的不同位置,将不同的 f_M 拼接起来后,再进行线性变换。

多头注意力机制函数为:

$$f_M(Q, K, V) = f_c(q_1, \dots, q_h) W_o, \quad (9)$$

式中: f_c 为拼接函数; q_i 为第 i 个单头注意力机制函数, $q_i = f_A(QW_{i,Q}, KW_{i,K}, VW_{i,V})$, 其中, $W_{i,Q}$ 、 $W_{i,K}$ 、 $W_{i,V}$ 为参数矩阵; W_o 为附加权重矩阵; $W_{i,Q}$ 、 $W_{i,K}$ 、 $W_{i,V}$ 、 $W_o \in \mathbf{R}^{d_m \times d_k}$ 。

将式(9)结果输入到 FFN 层, 该层由两个线性变换和一个激活函数 f_j 组成, FFN 函数表达式:

$$F_F(x) = \max(0, f_j(xW_1 + b_1)) W_2 + b_2, \quad (10)$$

式中: b_1 、 b_2 为偏置向量; W_1 、 W_2 分别为输入层到隐藏层、隐藏层到下一层(或输出层)的权重矩阵。

由于每个编码器的多头注意力层和 FFN 层都有一个残差, 对其进行连接后对每一层执行归一化操作, 构造单个编码器模块, N 个相同的模块堆叠构成编码器。为了提高 Transformer 模型性能, 在训练过程中, 在残差连接和层归一化之前对编码器和解码器的每个子层输出使用 Dropout 操作, 即随机屏蔽一些数。Transformer 模型主要结构如图 4^[18] 所示。

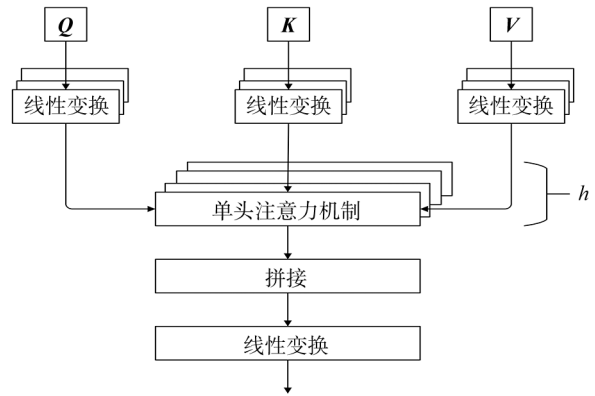


图3 多头注意力机制示意

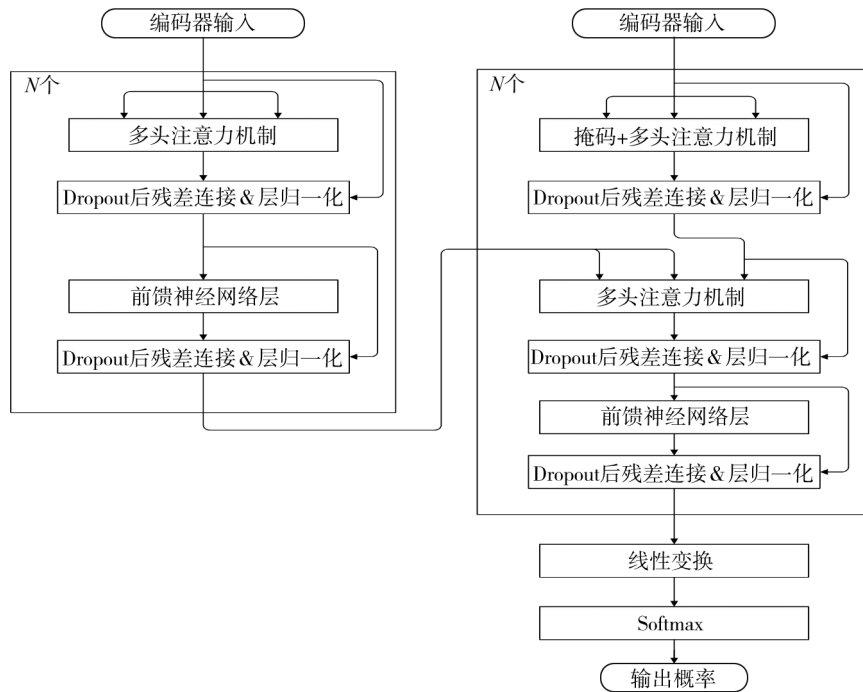


图4 Transformer 模型主要结构

2.2 性能测试

针对二分类、多分类等问题, 选取 UCI 机器学习资料库标准数据集^[19]中的 Seeds、Wine 与 Wdbc 数据集, 这些数据集可以避免模型对单一条件的依赖, 特征信息明确, 对模型的鲁棒性可以进行有效的测试, 适合机器学习和数据分类任务。3 种数据集信息如表 3 所示。

由于数据集数据个数、特征数目、类别数不同, 因此选择不同比例的训练集和测试集, 对于 Seeds 和 Wine 数据集, 随机选择 120 个样本作为训练集, 45 个样本作为测试集; 对于 Wdbc 数据集, 随机选择 120 个样本作为训练集, 60 个样本作为测试集。

表3 3种数据集信息

数据集	数据个数	特征数目	类别
Seeds	210	7	3
Wine	179	13	3
Wdbc	569	30	2

对于 SVM 模型,设置参数 C 与 g 均为 2;对于 KI-SVM 模型,设置参数 C 与 g 为 $[0.01, 100]$,种群数为 50,最大迭代次数为 200;对于 Transformer 模型,只使用了编码器,输入向量的维度为 64,多头注意力机制数为 4,单个编码器模块层数为 2,FFN 层的中间维度为 128,Dropout 参数为 0.1。

基于 3 种数据集,对比 3 种算法的试验结果,每种类型进行 30 次分类试验,每轮训练次数为 200,试验结果如表 4 所示。

表 4 3 种数据集分类试验结果

数据集	最优准确率/%			最差准确率/%			平均准确率/%			准确率标准差		
	SVM	KI-SVM	Transformer	SVM	KI-SVM	Transformer	SVM	KI-SVM	Transformer	SVM	KI-SVM	Transformer
Seeds	95.56	97.78	100	77.78	77.78	93.33	89.19	90.81	96.72	4.16	4.51	1.80
Wine	97.78	100.00	100	73.33	82.22	93.33	84.74	95.26	99.33	6.67	5.21	1.53
Wdbc	86.67	100.00	100	56.67	76.67	91.67	74.56	87.22	96.50	8.55	6.56	2.08

由表 4 可知:3 种数据集进行分类试验后,采用 Transformer 模型的最优准确率、最差准确率、平均准确率、准确率标准差均优于 SVM 模型和 KI-SVM 模型,验证了 Transformer 模型的有效性。

2.3 柴油机装配冷试分析及应用

随机选取 3 600 组正常数据,其中 3 200 组数据作为训练集,400 组数据作为测试集;随机选取 777 组异常数据中的 400 组作为训练集,剩下的 377 组数据作为测试集,正常数据标记为 0,异常数据标记为 1,训练次数均为 200。采用 SVM、KI-SVM、Transformer 模型分类测试结果如表 5 所示。

表 5 不同模型分类测试结果

模型	正常数据样本数	异常数据样本数	预测正确样本数	准确率/%
SVM	400	377	662	85.20
KI-SVM	400	377	719	92.54
Transformer	400	377	761	97.94

由表 5 可知:SVM、KI-SVM、Transformer 模型对正常数据和异常数据分类的准确率分别为 85.20%、92.54%、97.94%;Transformer 模型的准确率相较于 SVM、KI-SVM 均有较大提高,可以有效排查不合格的柴油机,验证了 Transformer 模型对柴油机装配冷试数据异常识别方法的可行性。

将采集到的异常点数据进一步进行统计分析,结果如表 6 所示。

表 6 各缸异常点数据对比

气缸 序号	$M_{min}/$ (N·m)	$M_{max}/$ (N·m)	$M_{ave}/$ (N·m)	$M_{std}/$ (N·m)	$P_{min}/$ kPa	$P_{max}/$ kPa	$P_{ave}/$ kPa	$P_{std}/$ kPa	$\theta_{min}/$ (°)	$\theta_{max}/$ (°)	$\theta_{ave}/$ (°)	$\theta_{std}/$ (°)	$\beta_{min}/$ kPa	$\beta_{max}/$ kPa	$\beta_{ave}/$ kPa	$\beta_{std}/$ kPa
1	719.3	1 060.7	901.2	86.8	701.2	833.9	757.4	32.2	259.8	281.3	273.5	5.8	-63.8	-37.6	-47.8	7.4
2	670.2	1 012.9	794.5	75.8	708.6	836.2	761.4	35.5	259.8	281.6	272.5	4.9	-63.5	-36.1	-47.1	8.0
3	690.1	1 045.9	865.3	89.9	707.2	845.4	770.6	41.9	258.4	281.3	269.6	5.9	-65.2	-35.5	-50.3	9.6
4	745.3	1 081.4	950.5	82.9	713.8	846.2	783.9	35.1	258.1	282.7	270.8	5.4	-64.4	-36.2	-49.9	7.8
5	700.1	1 039.1	844.9	84.1	703.3	840.86	762.6	40.7	257.3	282.3	271.7	5.7	-63.5	-35.9	-48.3	8.1
6	698.4	1 009.8	834.3	72.7	711.1	850.9	778.9	39.7	256.9	281.3	269.9	5.5	-67.0	-35.8	-54.9	9.0

通过统计 777 个异常点可知:排气压力与转矩呈显著正相关,排气压力增大导致转矩增大,这可能是因为装配间隙过小或润滑不足导致摩擦阻力增大,或气门升程不足导致排气受阻;此外,排气门开启时间过长与真空度异常同时出现,可能由于气门装配间隙精度偏低所致。

3 结论

1) 基于冷试装配数据集,采用智能算法改进的 DBSCAN 算法进行聚类分析,构建数据集。

2) 为了提高柴油机装配冷试测试性能,选取 UCI 机器学习资料库标准数据集中的 Seeds、Wine、Wdbc 三种数据集,对比 SVM、KI-SVM、Transformer 模型的分类效果,结果表明,Transformer 模型的最优准确率、最差准确率、平均准确率和准确率标准差优于 SVM 和 KI-SVM。

3) 基于 Transformer 模型对正常点和异常点分类的准确率达 97.94%,较 SVM、KI-SVM 分别提高了 14.61%、2.94%,验证了 Transformer 模型对柴油机装配冷试数据异常识别的可行性。

参考文献:

- [1] 林巨广,许华,谢峰,等.冷试技术在发动机装配质量在线检测中的应用[J].组合机床与自动化加工技术,2010(12):68-70.
- [2] PEINADO-ASENSI I, MONTÉS N, GARCÍA E. Industrial internet of things and big data techniques for the smart press shop 4.0 development in automotive industry [J]. IOP Conference Series: Materials Science and Engineering, 2023, 1284(1):012012.
- [3] LI J Y, ZHAO G, ZHANG P F, et al. A digital twin-based on-site quality assessment method for aero-engine assembly [J]. Journal of Manufacturing Systems, 2023, 71:565-580.
- [4] WANG K, WANG X, TAN C, et al. Digital twin-enhanced robotic system for remote diesel engine assembly defect inspection [J]. The Industrial Robot, 2025, 52(2):266-276.
- [5] CUI L, TIAN X C, WEI Q Z, et al. A self-attention based contrastive learning method for bearing fault diagnosis [J]. Expert Systems with Applications, 2024(238):13.
- [6] ABDELRAHMAN O, KEIKHOSROKIANI P. Assembly line anomaly detection and root cause analysis using machine learning [J]. IEEE Access, 2020, 8:189661-189672.
- [7] YUN K, CHONG Y, SONG E Z, et al. Intelligent fault diagnosis method of common rail injector based on composite hierarchical dispersion entropy and improved least squares support vector machine [J]. Digital Signal Processing, 2021, 114:103054.
- [8] 向晓良,苏春,张玉茹.基于改进 VMD 和 SVM 方法的滚动轴承故障诊断[J].东南大学学报(自然科学版),2026,56(2):322-332.
- [9] 王凯旋,张宏伟.基于改进 Transformer 的提升制动系统故障诊断研究[J].煤炭工程,2025,57(5):148-155.
- [10] CHEN M, GAN H B, WU H J. Research on fault diagnosis method for marine diesel engines based on multi-scale attention mechanism transformer [J]. Journal of Marine Science & Engineering, 2024, 12(12):2348.
- [11] WU H Y, TRIEBE M J, SUTHERLAND J W. A transformer-based approach for novel fault detection and fault classification/diagnosis in manufacturing: a rotary system application [J]. Journal of Manufacturing Systems, 2023, 67:439-452.
- [12] KHORAMPOUR S, GANDOMKAR M, SHAKIBA M. Enhancement of brain tumor classification from MRI images using multi-path convolutional neural network with SVM classifier [J]. Biomedical Signal Processing and Control, 2024, 93:106117.
- [13] JABARDI M. Support vector machines: theory, algorithms, and applications [J]. Infocommunications Journal, 2025, 17(1):66-73.
- [14] GHASEMI M, ZARE M, TROJOVSK P, et al. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm [J]. Knowledge-Based Systems, 2024, 295:111850.
- [15] MAGEED A A A E, ABOHANY A A A, HOSNY K M. Enhanced binary kepler optimization Algorithm for effective feature selection of supervised learning classification [J]. Journal of Big Data, 2025, 12(1):93.

- [16] 薛财文,刘通,邓立宝,等.基于主成分分析的DBSCAN分类差分进化算法改进[J].现代电子技术,2024,47(16):171-179.
- [17] WANG Y P, LIAO W J, SHEN H T, et al. Some notes on the basic concepts of support vector machines[J]. Journal of Computational Science, 2024, 82:102390.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of 31st Conference on Advances in Neural Information Processing Systems(NIPS 2017). Long Beach, USA: NIPS, 2017.
- [19] University of California, Irvine, School of Information and Computer Sciences. UCI machine learning repository[CP/OL]. (2020-10-29)[2024-11-30]. <https://archive.ics.uci.edu/>.

Data analysis of cold test for diesel engine assembly based on large models

ZHAO Xuhui¹, XU Zhuo¹, WANG Xiangcheng², YAN Wei^{1*}, LI Guoxiang¹

1. School of Nuclear Science, Energy and Power Engineering, Shandong University, Jinan 250061, China;

2. Shandong Inspur Aolin Big Data Technology Co., Ltd., Jinan 250101, China

Abstract: To improve the assembly quality and cold test performance of diesel engines, based on the basic dataset of diesel engine assembly and cold testing, three standard datasets, namely Seeds, Wine, and Wdbc, from the University of California Irvine(UCI) machine learning repository are selected. The analysis effects of the support vector machines (SVM) model, the SVM model improved by combined intelligent algorithm, and the Transformer model on abnormal cold test data are compared. The results show that the classification accuracies of the SVM, improved SVM, and Transformer models for normal and abnormal data are 85.20%, 92.54%, and 97.94%. Compared with the SVM and improved SVM models, the Transformer model has a significantly higher classification accuracy and can be used to analyze parameter anomalies. Exhaust pressure is closely related to torque, higher exhaust pressure leads to increased torque; an excessively long exhaust valve opening time results in abnormal intake vacuum. The effectiveness of the Transformer model in identifying engine assembly anomalies is verified.

Keywords: diesel engine assembly; cold test; anomaly detection; SVM; Transformer model architecture

(责任编辑:胡晓燕)