

基于距离密度的道岔转辙机异常检测新方法

刘晋瑛¹, 杨喜旺², 黄晋英³, 常金方¹, 高乐乐³

(1. 中北大学 计算机科学与技术学院, 山西 太原 030051; 2. 山西电子科技学院 信息与通信工程学院, 山西 临汾 041000;
3. 中北大学 机械工程学院, 山西 太原 030051)

摘要: 转辙机在工业应用中存在的大量异常数据, 对其健康状态的预测造成了严重影响。传统的基于局部密度的局部异常因子算法(Local Outlierness Factor, LOF)适用于单簇数据环境, 在多簇数据环境中容易将处于簇边缘的正常点判为异常点。针对这一缺陷, 以铁路道岔转辙机为研究对象, 选择5000条转辙机从启动到接通的油压数据进行分析, 通过引入反向 k -近邻改进LOF算法, 提出了一种新的基于密度的影响域异常检测算法(Influenced Outlierness, INFLO)。根据影响域离群因子的分布情况设定阈值, 依据不同的 k 值在阈值下的异常点个数来量化并识别异常点, 与传统的四分位数、KNN和LOF算法相比, 识别的异常点个数由8.92%提高到了10.6%。最后, 设计了长短期记忆(Long Short-Term Memory, LSTM)网络模型, 将剔除异常点后的数据点输入到网络中, 与上述方法和原始数据进行对比实验来验证所提出的数据清洗方法的有效性。结果显示改进后新方法的RMSE曲线与原始数据的RMSE曲线更接近, 有更好的预测效果。所提出的新方法能较好地清除转辙机油压数据集的异常值, 进而提高故障检测和数据预测的性能, 同时该方法对异常数据的清除效果优于其他方法。

关键词: INFLO; 转辙机; 数据清洗; 异常检测; LSTM

中图分类号: TP274 **文献标识码:** A **doi:** 10.62756/jnuc.issn.1673-3193.2024.09.0043

引用格式: 刘晋瑛, 杨喜旺, 黄晋英, 等. 基于距离密度的道岔转辙机异常检测新方法[J]. 中北大学学报(自然科学版), 2025, 46(4): 517-524.

LIU Jinying, YANG Xiwang, HUANG Jinyin, et al. A new method for detecting anomalies in switch machines based on distance density[J]. Journal of North University of China (Natural Science Edition), 2025, 46(4): 517-524.

A New Method for Detecting Anomalies in Switch Machines Based on Distance Density

LIU Jinying¹, YANG Xiwang², HUANG Jinyin³, CHANG Jinfang¹, GAO Lele³

(1. School of Computer Science and Technology, North University of China, Taiyuan 030051, China;
2. College of Information and Communication Engineering, Shanxi University of Electronic Science and Technology,
Linfen 041000, China;
3. School of Mechanical Engineering, North University of China, Taiyuan 030051, China)

Abstract: There are a large number of abnormal data in switch machines in industrial practice, which seriously affects the health status prediction of switch machines. The traditional Local Outlier Factor (LOF)

收稿日期: 2024-09-13

基金项目: 山西省回国留学人员科研资助项目(2022-141); 山西省基础研究计划资助项目(202203021211096); 山西省自然科学基金资助项目(201901D111157)

作者简介: 刘晋瑛(1997-), 女, 硕士生, 主要从事大数据与视觉计算的研究。

通信作者: 杨喜旺(1971-), 男, 副教授, 博士, 主要从事计算机应用、大数据图像处理及故障诊断的研究。E-mail: yangxw@nuc.edu.cn.

algorithm based on local density is suitable for single cluster data environments. However, in multi-cluster environments, it tends to misclassify normal points at the cluster boundaries as outliers. In response to this deficiency, a study was conducted on railway switch machines, selecting 5 000 oil pressure data from start-up to connection for analysis, and a new density based influenced outlierness algorithm (INFLO) was proposed by introducing reverse k -nearest neighbor to improve the LOF algorithm. A threshold was set based on the distribution of INFLO, and the number of outliers at the threshold was quantified and identified according to different k values. The number of outliers identified was increased from 8.92% to 10.6% compared with traditional Quartile, KNN, and LOF algorithms. Finally, a Long Short-Term Memory (LSTM) network model was designed to input the data after removing outliers into the network. Comparative experiments were conducted with the above methods and the original data to verify the effectiveness of the proposed data cleaning method. The results show that the RMSE curve of the improved method is closer to that of the original data, indicating better prediction performance. The proposed new method can better eliminate outliers in the oil pressure dataset of the switch machine, and improve the performance of fault detection and data prediction, indicating that this method has a better effect on clearing outlier data than other methods.

Key words: INFLO; switch machine; data cleaning; anomaly detection; LSTM

0 引言

随着社会的持续发展,工业数据质量的研究日益成为关注焦点,数据清洗算法的研究愈加重要。转辙机是铁路道岔系统中的核心设备,负责在列车行驶过程中将铁路轨道按需转换,以确保列车安全高效地完成行驶任务。转辙机需要在各种气候和自然条件下运行,面对复杂的运输环境和高频率的使用压力,保障转辙机设备的健康状态和高效稳定的运行对于铁路运输系统的安全和可靠性具有不可替代的重要意义。

转辙机运行数据的采集、分析和对于确保列车安全和运行效率至关重要。这些数据不仅包括转辙机的位置、状态和操作记录,还涉及到各种环境条件下的运行参数和异常事件记录。多数学者通常将转辙机的研究的重点放在故障诊断和寿命预测上,忽略了高质量数据在数据挖掘工作中的关键性作用。高质量的数据清洗不仅有助于减少异常数据的影响,还能提升模型的准确性和稳定性,从而在实际应用中取得更好的效果。因此,重视数据清洗工作并采用先进的数据质量管理技术是至关重要的。随着大数据和人工智能技术的迅速发展,数据清洗的研究与实践也在不断深化和演进。目前的研究重点包括但不限于:开发高效的自动化清洗算法,识别和处理脏数据、缺失数据以及异常数据;优化数据清洗流程,

提高数据清洗的准确性和效率。这些改进研究不仅提升了数据质量,还为数据驱动决策、机器学习和预测建模等应用的可靠性和实用性奠定了坚实的基础。近年来出现了多种数据异常检测的方法和算法,其中常见的数据清洗方法有基于统计的方法、基于机器学习的聚类法、基于距离的方法、近邻排序算法、四分位数算法以及孤立森林(IF)和深度学习法。

杨茂等^[1]建立了类似于正态分布的 3σ 准则模型来对光伏功率的数据进行异常检测。孙建树等^[2]利用ARIMA-SVR时间序列算法对水文数据进行了异常检测,检测性能良好但适用性差。徐凯等^[3]利用粗泛化箱型图法对桥梁监测数据进行了异常值识别及剔除,提高了计算效率且适用于大多数的数据分布。许璟琳等^[4]利用 k 近邻算法对医院的用电能耗进行了检测,杨霞等^[5]利用高斯混合模型提取了电力大数据异常值的特征,Wang等^[6]选用局部离群因子算法对异常值进行了识别,贺潇瑞^[7]利用孤立森林算法对随机子空间技术提取的小干扰稳定特征参数结果进行了异常值检测,这些算法均具有高准确率和高效率性,但是检测精度低。邹春龙等^[8]利用改进的K-means聚类定心算法来剔除曲轴轴颈数据的噪声点,王婧怡等^[9]基于KNN模型提出了一种高效的计算不完备数据夏普利值数据清洗的算法,Zhang等^[10]提出了一种基于权重的离群点检测算法,并得出在低维条件下检测效果优于局部离群因子的

结论。张忠平等^[11]提出了一种基于映射距离比离群因子的离群点检测算法来对人工数据集进行异常检测,检测能力和稳定性高于其他算法,并且与传统的检测模型相比具有更高的检测精度。还有研究者将多种方法结合起来进行数据清洗,如: Morrison 等^[12]提出了变点群法与四分位数算法相结合的离群点清洗模型, Lin 等^[13]使用了孤立森林和网络结合的方法来识别并删除 SCADA 数据中的异常值,黄灶荣等^[14]使用基于 BlazePose 和随机森林的算法对人体步态进行异常检测,陈泓铭等^[15]基于优化生成对抗模型对高温环境下的涡轮发动机转子系统不平衡进行异常检测。

基于机器学习的聚类法、基于距离的方法、四分位数算法以及孤立森林都是把数据的异常值简单定义为离群点,对与正常值差异不大的异常数据检测时性能会下降。吴磊等^[16]的研究表明,长短记忆网络具有良好的预测性能。周棟淞等^[17]基于长短时记忆神经网络,在时间维度的基础上利用 N-gram 算法增加数据的信息维度,以此减小不同进程类型的特征差异,表现出较好的检测效果。刘云朋等^[18]提出了一种基于深度学习与遗传算法相融合的监测算法来对光纤网络数据进行异常检测,在精度、稳定性及时效性方面都有很大提高。随着大数据时代深度学习的广泛应用,越来越多的学者将数据清洗的研究重点转到深度学习上来。

LOF 算法仅适用于单簇数据分布环境,对多簇数据进行判定时容易出现错误。针对这个缺陷,本文提出了一种基于局部密度的异常点检测算法 INFLO,引入反向 k -近邻对象,在计算 k -近邻的同时也计算了反向 k -近邻,由此计算出来的影响域离群因子可以高效地检测出局部异常点。

1 转辙机油压数据特性分析

在一定的环境下,油压的大小能反映出转辙机输出作用力的变化。根据转辙机正常工作状态下的油压数据,使用 Pycharm 软件绘制出转辙机正常理想工作状态的油压曲线图如图 1 所示。

曲线中的不同峰值和转换时间反映了转辙机工作过程中的 5 个阶段(启动和解锁阶段(0~2.90 s)、转换阶段(2.90~8.25 s)、锁闭阶段(8.25~10.65 s)、电路接通阶段)。

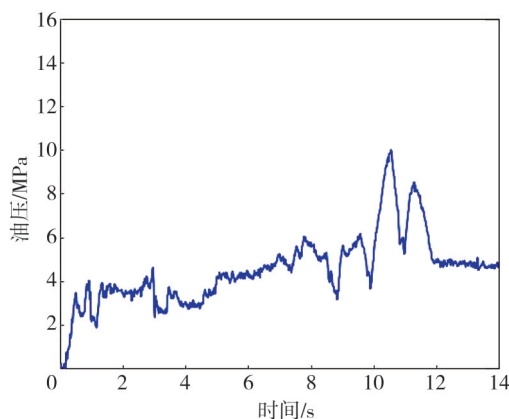


图 1 正常理想工作状态的油压曲线

Fig. 1 Oil pressure curve in normal working condition

2 基于密度的数据清洗算法

2.1 LOF 算法

LOF 算法是由 Breunig 等^[19]提出的一种用于异常检测的非监督学习方法,它可以通过计算每个数据点的局部离群因子来确定数据集中的异常值。具体如下:

定义 1 k -近邻 (k -Nearest Neighbors, k NN): 对于数据集中的每个数据点 p , 根据选定的距离度量方式, 找到与 p 距离最近的 k 个数据点, 这 k 个数据点即为 p 的 k 近邻。

定义 2 可达距离 (Reachability Distance): 用于评估一个数据点相对于另一个数据点的邻域密度。点 p 到 o 的可达距离等于 p 到其 k 个最近邻距离与 p 到 o 直接距离之间的最大值。

$$reach - distance_k(p, o) = \max(k - distance(o), d(p, o)). \quad (1)$$

定义 3 局部可达密度 (Local Reachability Density, LRD): 点 p 的局部可达密度为其 k -距离邻域中所有点相对于点 p 的可达距离的平均值的倒数。

$$lrd_k(p) = \left(\frac{\sum_{o \in N(p, k)} reach - distance_k(p, o)}{|N(p, k)|} \right)^{-1}, \quad (2)$$

式中: $\sum_{o \in N(p, k)} reach - distance_k(p, o)$ 为点 p 的 k -距离邻域中所有点相对于点 p 的可达距离。

定义 4 局部异常因子 (Local Outlier Factor, LOF): 点 p 的局部离群因子为点 p 邻域内其他点的局部可达密度的均值与点 p 的局部可达密度的比值。

$$lof_k(p) = \frac{\sum_{o \in N(p, k)} \frac{lrd_k(o)}{lrd_k(p)}}{|(p, k)|}, \quad (3)$$

式中: $\sum_{o \in N(p, k)} \frac{lrd_k(o)}{lrd_k(p)}$ 为点 p 的邻域内其他点的局部可达密度与点 p 的局部可达密度的比值的和。

通过计算可以得到数据集中每个点的 lof 值。lof 值越大, 表示该点相对于其邻域的离群程度越高, 从而更有可能是离群点。

然而使用 LOF 算法进行异常检测时, 一个常见的问题是在存在密度聚集的数据集中, 边界上的正常数据点可能被错误地标记为异常^[20]。这是因为 LOF 算法仅考虑了数据点的 k 近邻, 而没有考虑反向 k 近邻。这就导致了一个局限性, 即当数据点分布存在多个簇时, 边界上的正常数据点可能与异常数据点具有相似的 lof 值。LOF 算法对输入参数 k 的选择较为敏感, 不同的 k 可能导致不同的结果。

2.2 改进的算法

为了克服 LOF 算法的这一缺陷, 改进的 INFLO 算法中引入了反向 k 近邻^[21] (Reverse k -Nearest Neighbors, RkNN) 的概念, 它是指一个数据点的 k 个最近邻居中, 有多少个数据点将该数据点作为其最近邻居。RkNN 值可以用来衡量一个数据点在密度聚集时的影响力。引入 RkNN 可以使算法更准确地评估数据点的局部密度, 从而提高异常检测的性能。此外, RkNN 还可以用于识别数据点的聚类结构, 从而帮助理解数据集的特征。

具体而言, INFLO 算法不仅计算数据点的 k 近邻的局部可达密度 (lrd), 还计算 RkNN 的 lrd。与 LOF 算法相比, 在数据点分布超过两个甚至更多簇时, INFLO 算法可以更准确地判断数据点是否为异常。INFLO 算法基本定义如下:

定义 5 局部密度 (Local Density): 点 p 的局部密度是其 k 邻近距离的倒数。

$$den(p) = \frac{1}{k - distance(p)}. \quad (4)$$

定义 6 反向 k -近邻 (Reverse k -Nearest Neighbors, RkNN): 如果一个数据点 q 的 k 个最近邻中包含点 p , 那么数据点 p 就属于点 q 的反向 k 近邻。

$$rmn_k(p) = \{q | q \in NN_k(q)\}. \quad (5)$$

定义 7 影响域离群因子 (Influenced Outliers, INFLO): 影响域由 k -近邻和 RkNN 组成, 点 p 的影响域离群因子是其影响域内其他点的密度的均值与点 p 密度的比值。

$$inflo_k(p) = \frac{den_{avg}(IS_k(p))}{den(p)}, \quad (6)$$

式中: $den_{avg}(IS_k(p)) = \frac{\sum_{o \in IS_k(p)} den(o)}{|IS_k(p)|}$, $IS_k(p)$ 是

点 p 影响域内其他点的密度和。

3 实验验证

本文使用 INFLO 对转辙机的油压数据进行数据清洗, 所用到的全部数据均来自于某铁路局电务器材公司提供的转辙机模拟实验平台。

实验台由 ZYJ* 转辙机 (主机) 以及两台副机通过油管相连接组成, 测量精度为 0.1 MPa, 计算机监测系统每隔 0.02 s 采集一次转辙机的油压信号。数据集共有 349 500 条数据, 每条数据记录了转辙机的启动、解锁、转换、锁闭、接通五个阶段共 14 s 的工作状态。采集的数据集可用于预测转辙机的性能, 一些不相关的数据会影响模型的预测精度, 因此需要对数据集进行处理。

3.1 INFLO 算法对转辙机油压数据的清洗

所提出的 INFLO 算法首先从数据集中随机选取 5 000 条数据进行实验, 正常和故障数据比例为 3:1, 数据集绘成的散点图如图 2 所示。其次将数据点整合到二维空间中, 通过相关算法计算局部密度及局部异常因子, 通过分析异常因子的特征确定合适的阈值。然后, 根据选定的阈值进行实验确定合适的 k 值。最后建立模型对异常点进行数据清理。INFLO 方法的流程图如图 3 所示。

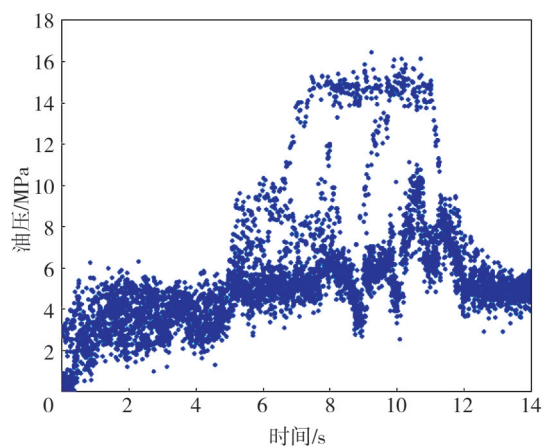


图 2 数据集散点图

Fig. 2 Data distribution point diagram

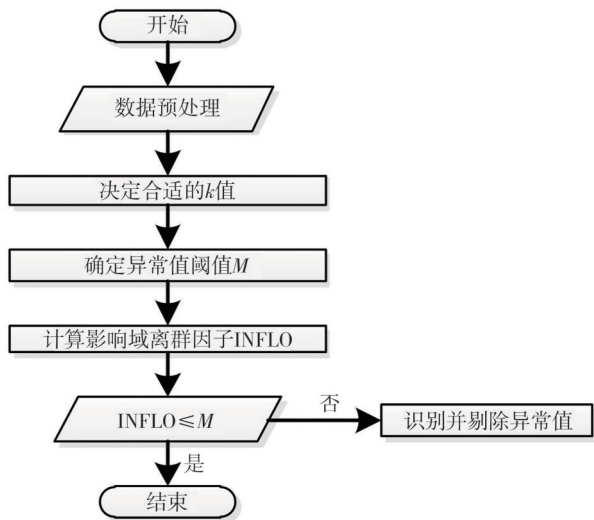


图 3 INFLO方法流程图

Fig. 3 INFLO method flowchart

异常值数量的多少是由影响域离群因子来量化的,所以需要有一个合适的阈值来消除异常值。从图 4 中可以看出,大部分点都集中在 0.5~1.5 附近,表明阈值较小。经过实验对比,将异常值的阈值 M 设为 1.20。

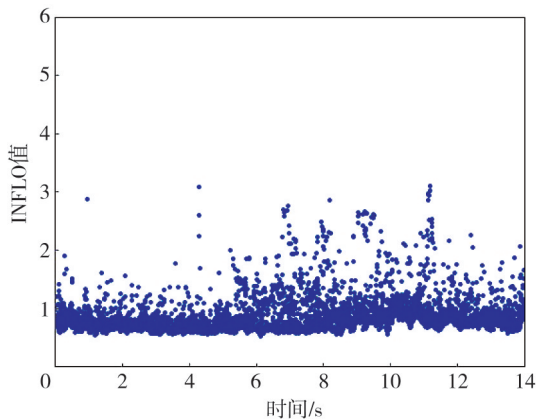


图 4 影响域离群因子分布图

Fig. 4 Distribution diagram of INFLO

通过对 INFLO 算法的分析可以发现, k 最近邻的确定对数据点的局部密度的量化影响很大,所以有必要找到合适的 k 值来确保明显异常点的数量。异常点的数量随 k 值变化的曲线如图 5 所示,可以看出,当 k 值较小时,随着 k 值的增大,异常点的数量整体呈逐渐减少的趋势,当 k 值增大到 8 时,异常点数量的减少趋势明显变缓并逐渐趋于稳定。 k 值越大,计算过程中需要考虑的邻域数据点越多,计算量相应增加,这会导致算法的运行速度变慢,因此设置合适的 k 值为 25。

异常值因子大于阈值的数据点被称为异常点,将异常点剔除,剩下的点即是因子小于阈值的正常

点。利用 INFLO 算法对所有转辙机油压数据进行异常值识别的结果如图 6 所示。由图 6 可以发现,偏差过大的离群值(红色点)都被识别为异常值。

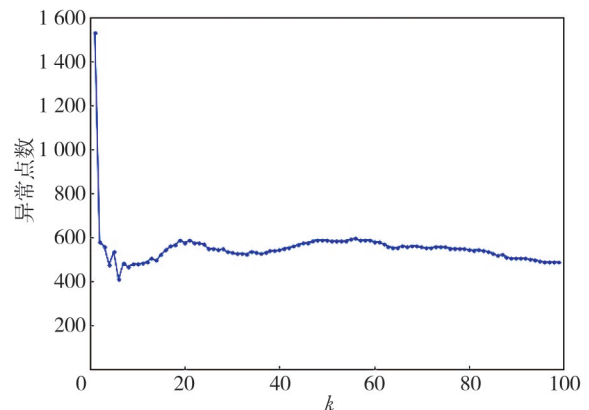


图 5 异常点的数量随 k 值变化的曲线

Fig. 5 Curve of the number of outliers varying with value of k

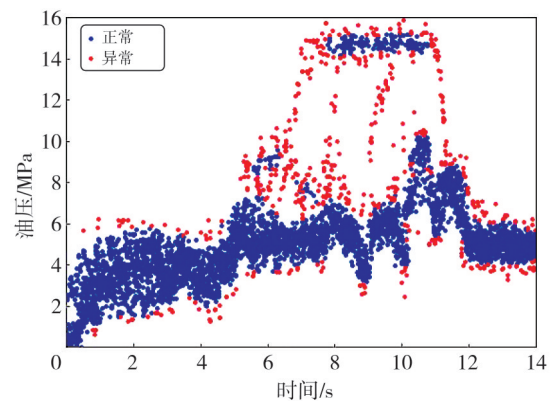


图 6 INFLO 算法异常值识别

Fig. 6 INFLO method for identifying outliers

3.2 对比实验

为了进一步验证该数据处理方法的有效性,使用 4 种方法清洗后的数据集来检测转辙机的异常状态,使用转辙机振动数据集验证该方法的实用性,各清洗方法的比较如表 1 所示。由 Yao 等^[22]可知,清洗的数据点越多,清洗效果越好。由表 1 可以看出,采用本文改进的 INFLO 方法清洗的异常点数量多于其他方法。

表 1 各方法清洗结果

Tab. 1 Cleaning results of each method

清洗方法	总数据量/个	清洗的数据点数/个	清洗数据点的占比/%	清洗后的数据集名称
INFLO 算法	5 000	530	10.60	数据集_1
KNN 算法	5 000	107	2.14	数据集_2
四分位数算法	5 000	446	8.92	数据集_3
LOF 算法	5 000	423	8.46	数据集_4
原始数据	5 000	0	0	数据集_5
LOF 振动	5 000	438	8.76	数据集_6
INFLO 振动	5 000	675	13.5	数据集_7

与图1相比,图6将超出理想状态的油压值峰值部分进行了保留,清洗的目的是清洗影响聚类的相对异常的数据,使数据聚类更明显,保留数据整体特点,为提高故障诊断和预测准确率提供更好的数据状态。图7~图9为采用其他清洗方法的清洗结果,图10~图11为采用振动数据集的清洗结果。

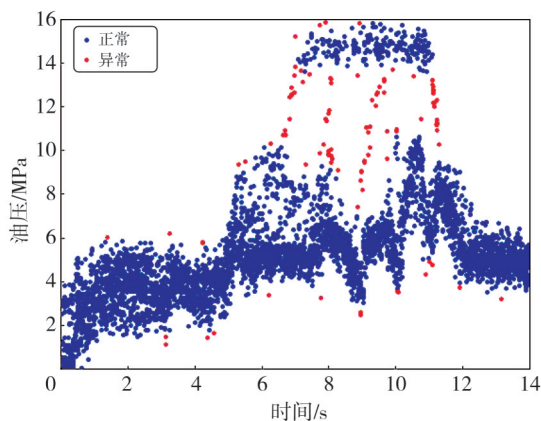


图7 KNN算法的清洗结果

Fig. 7 KNN algorithm cleaning results

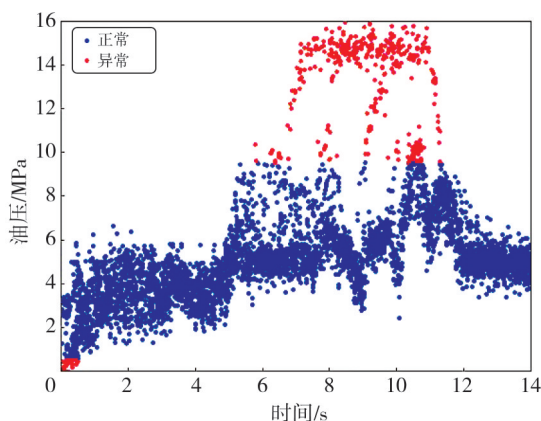


图8 四分位数方法的清洗结果

Fig. 8 Quartile method cleaning results

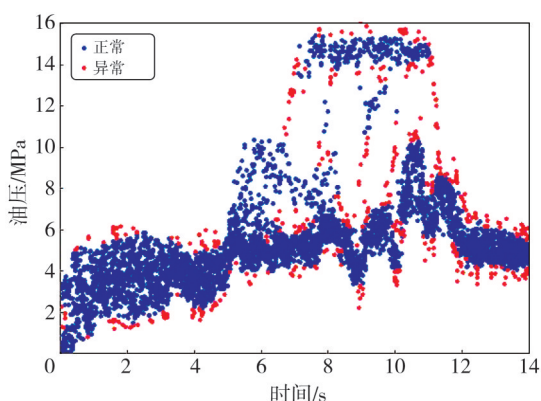


图9 LOF方法的清洗结果

Fig. 9 LOF method cleaning results

由图可知,采用其他清洗方法或多或少存在未清洗的异常数据。通过对比试验表明,所提出的

INFLO方法适用性强,清洗效果好,具有统计学价值。

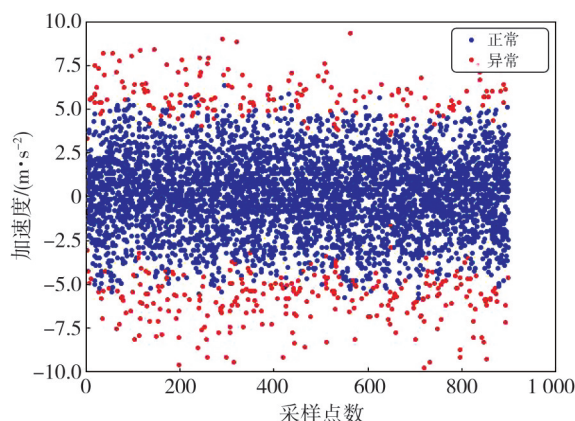


图10 LOF方法(振动)的清洗结果

Fig. 10 LOF method(vibrate) cleaning results

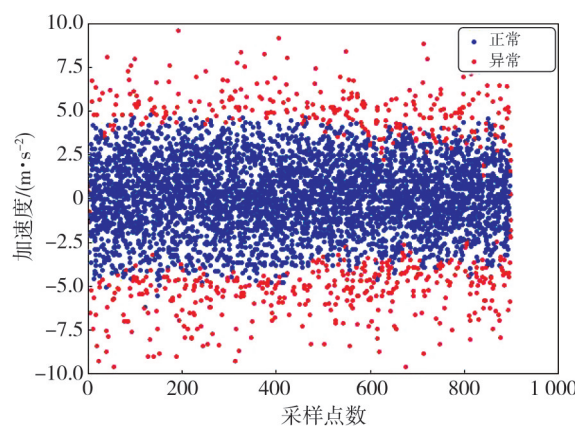


图11 INFLO方法(振动)的清洗结果

Fig. 11 INFLO method(vibrate) cleaning results

3.3 预测检测

LSTM模型能够处理各种不同时间尺度的数据,在数据预处理和特征提取方面的灵活性使它成为时间序列预测中的强大工具。本文所使用的LSTM模型的主要结构是两层LSTM层和全连接层,其中两层LSTM的神经元数分别为128和64,全连通层的输出维数为1。通过模型输出对目标变量的预测,将预测值的残差与实际数据进行了比较。

均方根误差(RMSE)^[23]常被用来评价LSTM的预测性能,也能反映数据集的清洗效果。计算并绘制表1中各方法处理后数据集的RMSE如图12所示。由图可知,所有曲线的变化趋势与原始数据(数据集_5)的变化趋势大致相同,但是INFLO方法的RMSE曲线离原始数据最近,且较好地清理了异常点。结果表明,所提出的数据处

理方法能较好地识别转辙机油压数据的异常值,提高了异常检测的预测性能。

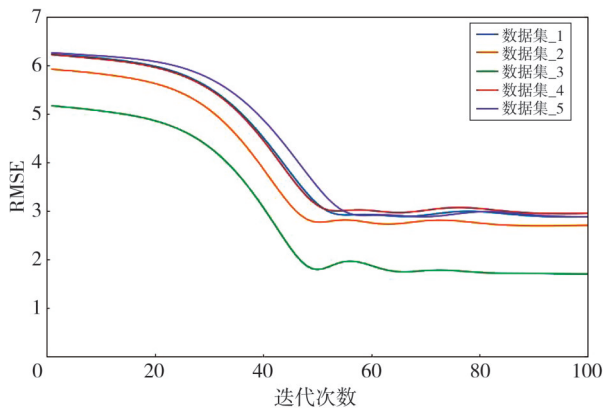


图 12 各方法的 RMSE 曲线

Fig. 12 RMSE curves of various methods

4 结 论

由于转辙机处于复杂的工作环境中,使得实测的油压信号数据集受到不同程度的污染。本文通过对数据的异常因子的分析,改进了 LOF 算法,提出了一种基于密度的数据清洗 INFLO 算法来用于转辙机油压数据的清洗,通过设置相关参数阈值来剔除异常点。然后,建立了长短期记忆(LSTM)模型,通过与 LOF 及几种清洗方法的比较验证了所提出的数据清洗方法的有效性。实验结果表明,所提出的数据清洗方法能较好地清除转辙机油压数据集的异常值并且提高了故障检测和数据的预测性能。数据清洗的结果可以为转辙机运行数据的异常检测和工业实际应用提供依据。

参考文献:

- [1] 杨茂, 孟玲建, 李大勇, 等. 基于类 3σ 准则的光伏功率异常数据识别[J]. 可再生能源, 2018, 36(10): 1443-1448.
YANG Mao, MENG Lingjian, LI Dayong, et al. Identification of abnormal data of photovoltaic power based on class 3σ [J]. Renewable Energy Resources, 2018, 36(10): 1443-1448. (in Chinese)
- [2] 孙建树, 姜渊胜, 陈裕俊. 基于 ARIMA-SVR 的水文时间序列异常值检测[J]. 计算机与数字工程, 2018, 46(2): 225-230.
SUN Jianshu, LOU Yuansheng, CHEN Yujun. Outlier detection of hydrological time series based on ARIMA-SVR model[J]. Computer & Digital Engineering, 2018, 46(2): 225-230. (in Chinese)

- [3] 徐凯, 袁蒋鹏. 粗泛化箱型图法在大中跨径桥梁监测异常值剔除的应用[J]. 广东公路交通, 2024, 50(1): 66-71.
XU Kai, YUAN Jiangpeng. Application of rough generalization box plot in eliminating outlier of bridge monitoring [J]. Guangdong Highway Communications, 2024, 50(1): 66-71. (in Chinese)
- [4] 许璟琳, 彭阳, 余芳强. 基于 k -means 聚类 and 离群点检测算法的医院建筑节能诊断方法[J]. 计算机应用, 2021, 41(S1): 288-292.
XU Jinglin, PENG Yang, YU Fangqiang. Energy-saving diagnosis method for hospital buildings based on k -means clustering and outlier detection algorithm[J]. Journal of Computer Applications, 2021, 41(S1): 288-292. (in Chinese)
- [5] 杨霞, 苟亮, 马倩, 等. 基于离群点检测算法的电力大数据异常值检测[J]. 电子设计工程, 2023, 31(24): 61-64.
YANG Xia, GOU Liang, MA Qian, et al. Outlier detection of electric power big data based on outlier detection algorithm[J]. Electronic Design Engineering, 2023, 31(24): 61-64. (in Chinese)
- [6] WANG B Y, LUO X Y, ZHANG S M. An improved outlier detection algorithm K-LOF based on density[J]. Computing, Performance and Communication Systems, 2017, 2(1): 1-7.
- [7] 贺潇瑞. 基于孤立森林与随机子空间的小干扰稳定评估[J]. 电气应用, 2023, 42(10): 16-20.
HE Xiaorui. Small signal stability assessment based on isolation forest with stochastic subspace identification [J]. Electrotechnical Application, 2023, 42(10): 16-20. (in Chinese)
- [8] 邹春龙, 黄配乐, 王生怀, 等. 基于改进 K-means 聚类定心算法的曲轴轴颈圆度误差评定[J]. 工具技术, 2024, 58(6): 141-150.
ZHOU Chunlong, HUANG Peile, WANG Shenghuai, et al. Crankshaft journal roundness error assessment based on improved K-means clustering centering algorithm [J]. Tool Engineering, 2024, 58(6): 141-150. (in Chinese)
- [9] 王婧怡, 陈胤佳, 袁野, 等. 面向 K-近邻学习模型的高效数据清洗框架[J]. 计算机科学与探索, 2023, 17(9): 2241-2251.
WANG Jingyi, CHEN Yinjia, YUAN Ye, et al. Efficient data cleaning framework for k-nearest neighbor learning models[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(9): 2241-2251. (in Chinese)

- [10] ZHANG S, WAN J. Weight-based method for inside outlier detection[J]. *Optik*, 2018, 154: 145-156.
- [11] 张忠平, 姚春辰, 孙光旭, 等. 基于映射距离比离群因子的离群点检测算法[J]. *计算机集成制造系统*, 2024, 30(5): 1719-1732.
ZHANG Zhongping, YAO Chunchen, SUN Guangxu, et al. Outlier detection algorithm based on mapping distance ratio outlier factor [J]. *Computer Integrated Manufacturing Systems*, 2024, 30(5): 1719-1732. (in Chinese)
- [12] MORRISON R, LIU X, LIN Z. Anomaly detection in wind turbine SCADA data for power curve cleaning [J]. *Renewable Energy*, 2022, 184: 473-486.
- [13] LIN Z, LIU X, COLLU M. Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks[J]. *International Journal of Electrical Power & Energy Systems*, 2020, 118: 105835.
- [14] 黄灶荣, 王春宝, 韦建军, 等. 基于BlazePose和随机森林算法的异常步态检测[J]. *计算技术与自动化*, 2024, 43(2): 62-69.
HUANG Zaorong, WANG Chunbao, WEI Jianjun, et al. Abnormal gait detection based on BlazePose and random forest [J]. *Computing Technology and Automation*, 2024, 43(2): 62-69. (in Chinese)
- [15] 陈泓铭, 杨曦荻, 余玲, 等. 基于优化生成对抗模型的涡轮发动机转子不平衡异常检测方法研究[J]. *兵器装备工程学报*, 2024, 45(6): 104-114.
CHEN Hongming, YANG Xidi, YU Ling, et al. Research on anomaly detection of turbine engine rotor unbalance based on optimized generative adversarial model [J]. *Journal of Ordnance Equipment Engineering*, 2024, 45(6): 104-114. (in Chinese)
- [16] 吴磊, 康英伟. 基于改进粒子群优化长短时记忆神经网络的脱硫系统SO₂预测模型[J]. *热力发电*, 2021, 50(12): 66-73.
WU Lei, KANG Yingwei. Prediction model of SO₂ concentration in desulfurization system based on improved particle swarm optimization LSTM [J]. *Thermal Power Generation*, 2021, 50(12): 66-73. (in Chinese)
- [17] 周棟淞, 唐彰国, 王远强, 等. 基于LSTM的多维度进程异常行为检测[J]. *通信技术*, 2024, 57(6): 598-608.
ZHOU Lansong, TANG Zhangguo, WANG Yuanqiang, et al. Multi-dimensional process abnormal behavior detection based on LSTM [J]. *Communications Technology*, 2024, 57(6): 598-608. (in Chinese)
- [18] 刘云朋, 霍晓丽, 刘智超. 基于深度学习的光纤网络异常数据检测算法[J]. *红外与激光工程*, 2021, 50(6): 288-293.
LIU Yunpeng, HUO Xiaoli, LIU Zhichao. Optical fiber network abnormal data detection algorithm based on deep learning [J]. *Infrared and Laser Engineering*, 2021, 50(6): 288-293. (in Chinese)
- [19] BREUNIG M M, KRIEDEL H P, NG R T, et al. LOF: Identifying density-based local outliers[C]//the 2000 ACM SIGMOD International Conference on Management of Data, 2000: 93-104.
- [20] 孟妮娜, 王正阳, 高晨博, 等. 一种融合局部异常因子的矢量建筑物群聚类方法[J]. *武汉大学学报(信息科学版)*, 2024, 49(4): 562-571.
MENG Nina, WANG Zhengyang, GAO Chenbo, et al. A vector building clustering algorithm based on local outlier factor [J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(4): 562-571. (in Chinese)
- [21] 张喜梅, 解滨, 徐童童, 等. 基于反向K近邻和密度峰值初始化的加权Kmeans聚类入侵检测算法[J]. *南京理工大学学报*, 2023, 47(1): 56-65.
ZHANG Ximei, XIE Bin, XU Tongtong, et al. Intrusion detection algorithm based on weighted Kmeans clustering with reverse K-nearest neighbor and density peak initialization [J]. *Journal of Nanjing University of Science and Technology*, 2023, 47(1): 56-65. (in Chinese)
- [22] YAO Q T, ZHU H W, XIANG L, et al. A novel composed method of cleaning anomaly data for improving state prediction of wind turbine [J]. *Renewable Energy*, 2023, 204: 131-140.
- [23] XIANG L, WANG P, YANG X, et al. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism [J]. *Measurement*, 2021, 175: 109094.