

DOI:10.19344/j.cnki.issn1671-5276.2024.03.032

# 基于时序数据的电梯数据清洗方法研究

王容

(南京理工大学机械工程学院,江苏南京 210094)

**摘要:**为了对非周期性、非高斯性及间歇性的曳引式电梯数据进行数据清洗,对电梯运行过程中的异常数据进行排查,提出一种改进后的长短期记忆网络的数据清洗模式。在对基于物联网技术使用数据库存储的时序数据进行异常数据的清洗时,提取不等长的时间序列数据进行划分与填充,利用长短时神经网络对其进行建模,进行初期的异常数据检测清洗。在系统中实现电梯故障系统的故障预测、寿命分析、可视化前的数据清洗工作,完成数据优化。

**关键词:**物联网;时间序列;数据管理

**中图分类号:**TP183 **文献标志码:**A **文章编号:**1671-5276(2024)03-0151-04

## Research on Data Cleansing Method of Elevator Based on Time Series Data

WANG Rong

(School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** In order to perform the data cleaning of aperiodic, non-Gaussian and intermittent traction type elevator and troubleshoot the abnormal data in the process of elevator operation, this paper proposes a data cleaning mode of modified long short-term memory network. Based on the IoT technology using a database to store the time-series data of abnormal data cleaning, the time series data of different lengths are extracted for division and filling, the long and short time neural network is used for modeling, and the initial abnormal data detection and cleaning are carried out. Data cleaning and data optimization are completed before the realization of fault prediction, life analysis and visualization of the elevator fault system.

**Keywords:** IoT; time series data; data management

### 0 引言

大数据已经广泛运用于各个行业,电梯物联网正逐步完善。由于所使用的电梯数量多、类型广,造成了电梯品类的管理与维保质量各不相同,经验不同的工作人员对电梯实际故障情况的判断会出现偏差,影响最终故障的解决。因此,加强对电梯的信息化管理、提高电梯设备质量、做好售后服务、保障电梯使用安全是非常重要的。对电梯数据进行监测、数据管理、预警,构建电梯系统成为了必然趋势。

随着物联网的兴起,智能设备和传感器技术得到大量应用,在数据清洗的研究领域,又新增了时序数据的清洗。卢峰等<sup>[1]</sup>使用云计算对电网大数据进行数据清洗,该数据清洗模型针对的是企业内部决策者提供的准确数据。谢智颖等<sup>[2]</sup>针对公交车规律使用长短期记忆(LSTM)神经网络进行了数据清洗来解决公交车到站预测问题,并对各类异常数据实现了数据清洗。XIE等<sup>[3]</sup>比较了

传统的卷积神经网络和长短期记忆网络提取特征,实现对汽车品质的准确评价。沈沛等<sup>[4]</sup>提出了一种基于相似重复数据少的DM-SVM数据清洗,利用时序数据的时序唯一性,将SVM算法的窗口改为自动缩放型,提高了数据清洗的效率。韩红桂等<sup>[5]</sup>使用SVM与粒子算法相结合,对城市污水处理过程的异常数据建立了一种缺失数据补偿模型,提高了数据质量。WANG等<sup>[6]</sup>针对BIM信息集成技术产生的海量数据进行数据清洗,利用RNN-LSTM网络对建筑施工项目中质量数据进行预测,降低施工中的问题率。YANG等<sup>[7]</sup>基于多传感器周期性数据使用LSTM模型进行数据异常分类与定位,提出了条件生成对抗网络来进行数据修复。

为了解决上述问题,本文提出了一种基于电梯时序数据的LSTM异常数据清洗方法。首先,根据时间序列进行整合重组,剔除停顿的数据后,使用长短期记忆深度学习进行数据清洗,最后得到相对干净的电梯数据。

**基金项目:**江苏省研究生科研与实践创新计划项目(SJCX21\_0126)

**作者简介:**王容(1998—),女,江苏扬州人,硕士研究生,研究方向为智能制造、智能优化算法等,920937342@qq.com。

# 1 数据采集

## 1.1 电梯信号数据采集

电梯数据采集主要基于运行过程的各个传感器和电梯监控数据,如温度传感器、加速度传感器等。电梯运行信号采集具有以下特点:1)采样频率高;2)数据量大;3)采集实时性较高;4)电梯机房比较封闭且有较多影响电磁波传输的障碍物。

如图 1 所示,基于物联网的电梯数据管理系统的数据源是由安装在电梯中各个设备组件中的传感器及其电梯机房的数据构成的。多个电梯工作组上的传感器构成整个系统的物联网感知层,并将采集到的信号数据源源不断地通过有线以太网组成的物联网传输层发送到服务器集群中。

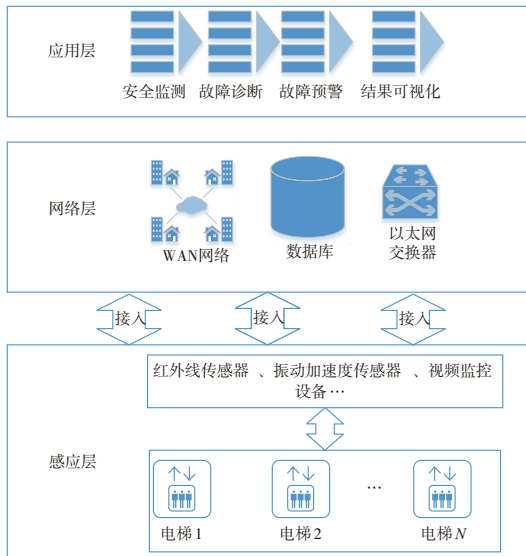


图 1 电梯物联网架构图

为了满足电梯运行数据采集实时性高、数据量大等需求,电梯运行信号采集系统采用有线以太网的数据传输方式。系统主要采集的数据是电梯运行状态信号以及  $x$ 、 $y$ 、 $z$  轴 3 个方向的振动信号数据。系统将采集到的数据进行换算后存入实时数据库中,并进行数据清洗,等待分布式处理集群的后续故障诊断、故障预警等应用实现。

## 1.2 电梯数据存在的问题

电梯数据属于连续、离散作业混杂的数据类型。影响电梯运行的要素很多,如所在位置、机房温度、运行加速度的变化等。由于物联网技术的兴起,许多电梯已经开始步入智能化管控阶段,故障检测与预警系统实际需要的数据是经过数据清

理后获得初值化的多维、大规模数据集,直接使用原始数据会加大计算复杂性,难以保证预测的准确性和可靠性。所以,需要对电梯数据进行清洗,避免故障预测的失真。电梯数据质量问题主要表现在以下几个方面。

1)数据失真和失准。由于电梯所处的环境及电梯使用年限各不相同、设备运维保养情况各不相同等原因,可能造成各类运行数据出现数据失真和失准。

2)数据冗余。冗余数据是指对电梯数据分析预测时没有价值的数,如:同一个时间间隔内数据重复上传的情况。

3)数据错列。当数据采集器出现故障或者控制器发生收录错误时,会出现部分数据与其设置的属性无法对应,从而产生错列问题。

本文针对电梯时序数据的数据失真失准问题进行研究,使用 LSTM 算法对数据的异常值进行探究,对电梯的异常值数据进行清洗,优化数据质量。

## 1.3 数据评估

不同的数据质量问题具有不同的严重性和故障发生的可能性,使用风险评估矩阵来对各类数据的质量问题进行评估,如表 1 所示。

表 1 数据评估混淆矩阵表

| 混淆矩阵 | 检测结果           |                |
|------|----------------|----------------|
|      | 正常             | 异常             |
| 真实情况 | 正常<br>真正常 (TP) | 异常<br>假异常 (TN) |
|      | 异常<br>假异常 (FP) | 正常<br>真异常 (FN) |

对于检测结果可以使用如下公式来进行计算:

$$A = \frac{T_p}{T_p + F_p} \tag{1}$$

$$R = \frac{T_p}{T_p + F_n} \tag{2}$$

式中: $A$  为准确率; $R$  为召回率。

## 2 电梯时序大数据特征

电梯传感器在采集数据时会带有时间序列这一时间要素。电梯的时序数据体现了数据的实时性和连续控件位置变化的属性,反映了其中某部电梯与其所处楼层位置之间的关系特征。如图 2 所示,图中的各个节点代表电梯控制系统整个运

行周期的状态情况。

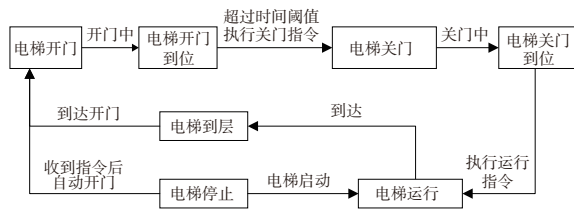


图2 电梯状态转移图

电梯的时间维度是以某一个时刻作为起点,在某段时间内不固定地发生变化,根据电梯状态转移图可知,电梯数据具有一定的规律性、序列性。

### 1) 序列性

电梯的采样频率以一定的时间间隔进行数据采集,上、下楼梯时会有剧烈的抖动,影响  $x$ 、 $y$ 、 $z$  轴上的加速度、按键的楼层、到达的时间、等待序列等。在这个序列中数据是连续的,丢失几个时间点的值,会出现较大的奇异值。

### 2) 规律性

电梯加速度变化存在一定的随机性,但是随时间变化有一定规律性。例如:工作场所内的电梯在工作日的工作时间段日变化规律相近,工作日与休息日的形状不同。从日流量变化图可获得每天的人流量高峰时间与高峰小时流量,这些都具有随机性。但是电梯具有状态转移规律,电梯数据会具有规律性,电梯开门、选择到达楼层、关门、运行,循环往复,一直到最后电梯停止。

## 3 电梯大数据清洗方法

本文基于时序电梯数据的改进长短期记忆神经网络算法(LSTM)建立数据清洗方法。LSTM具有控制遗忘的结构设计,非常适合处理时序任务。相比于传统神经网络RNN,LSTM提出了“遗忘门-输入门-输出门”,对前期的数据进行选择性地记忆与遗忘,提取时间特征对数据实现长期记忆。利用LSTM方法进行异常数据清洗并实现数据补偿的方法如下。

1) 由于电梯数据呈现非周期性、间歇性的特点,所以需要在进行预测前将非周期性的时间序列进行转换,转化为不等长的时间序列片段,将每次暂停在某一楼层的时间段剥离后进行时间序列的重组,去除长时间停留在某层的时间间隔,完成电梯数据的重组排序,如图3所示。

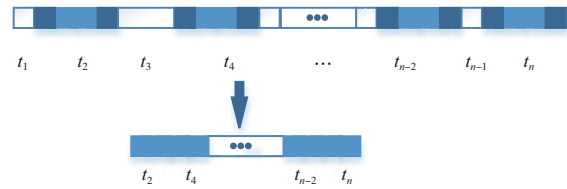


图3 非周期性时间序列片段重组

2) 对电梯数据在使用LSTM算法进行清洗前,需要进行归一化处理,在训练样本上需要求出每个维度的均值和方差,在训练样本上进行归一化处理的公式为:

$$\text{std}(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3)$$

$$Y = \text{std}(x) \times (\max(x) - \min(x)) + \min(x) \quad (4)$$

式中: $Y$ 为归一化后的数据; $x$ 为待处理的数据。其中,对时间进行归一化处理时,提取日期和时间后,再对时间进行转化。

3) 确定模型其他参数。例如:输出层、节点数、优化器等,结合损失函数和观测函数随训练轮次的结果,修改相应的参数信息及选取合适的训练轮次。

4) 通过不同的数据预测模型进行预测及数据分析,异常数据具有很强的突变性,数据分布的方差也会发生很大的变化。因此,将预测值与真实值的误差值进行记录,在阈值内进行数据的清洗与补偿。

## 4 数据清洗结果验证

### 4.1 利用时序重组后的LSTM对数据的预测与判断

本文以某学校内的电梯数据集为例进行分析。该数据集选用电梯加速度数据集,使用加速度传感器记录  $x$  轴、 $y$  轴、 $z$  轴及绝对值加速度,随机采集上下电梯的一段数据,共采集 85 770 组数据。选用了  $x$  轴加速度、 $y$  轴加速度、 $z$  轴加速度、绝对值加速度来作为输入特征,预测绝对值加速度。以电梯上下运行加速度为正常样本数据,以成年人在轿厢内的中等强度跳动模拟异常数据。如图4所示,得到传感器数据,并且标记出异常数据。直接采用LSTM对电梯数据进行异常值清洗会出现较大误差,预测值与真实值的误差阈值较大,异常值数据清洗的准确率较低。将静止时的数据与运行中产生的加速度变化数据进行剔除,完成剔除后对电梯运行过程中的数据进行时间序

列的重组,再根据数据的时间间隔进行序列生成,完成后再对数据进行 LSTM 算法的数据清洗。结果如图 5 所示。预测的驼峰数据变少,数据预测的误差值在 $\pm 0.5$ 之间。

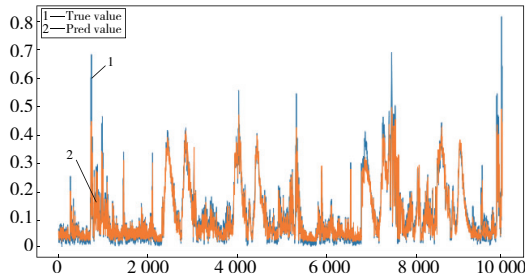


图 4 重组前的 LSTM 预测数据

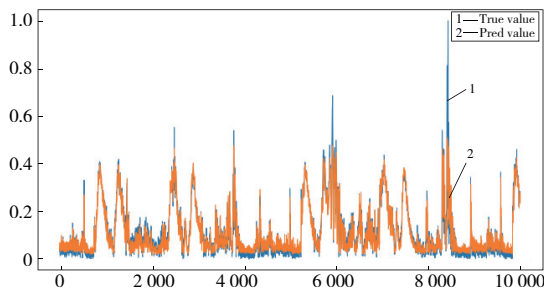


图 5 重组后的 LSTM 预测数据

重组前的训练耗时 210 s,重组后的算法迭代轮数不变,训练耗时约 170 s,单轮迭代时间 2 s。相比于数据重组前,改进后由于数据结构简单,静态数据减少,训练速度提升了 19%。

#### 4.2 数据清洗实验结果

通过比较预测值和真实值,计算两者误差,设定真实值与预测值的误差阈值为 $\pm 0.1$ ,可以成功检测出异常数据的数量,然后比较异常数量并使用评判结果进行评价。数据重组前与数据重组后的 LSTM 数据清洗方法分别对电梯加速度数据进行清洗后的结果对比如表 2 所示。

表 2 不同方法对数据清洗后的结果对比 单位:%

| 对比项 | LSTM | 改进后的 LSTM |
|-----|------|-----------|
| 精度  | 93.4 | 95.9      |
| 查准率 | 91.8 | 92.2      |

通过对比可以发现,电梯数据在进行重组后

再进行异常值清洗,提高了数据清洗的准确率和召回率,说明时间重组后确实可以有效地提高数据的质量。

## 5 结语

本文结合物联网采集的时序电梯大数据对非周期性的电梯数据进行时序重组后,使用 LSTM 神经网络对电梯异常数据进行数据清洗并补偿,并对采集的实际数据进行了验证。电梯数据清洗的后期研究需要结合其他的清洗方式清洗后的数据进行对比分析才能真正地应用到系统预测、可视化界面以及寿命预测分析中,使其真正服务于电梯物联网系统。

## 参考文献:

- [1] 卢峰,吴朝文,陈小龙,等. 基于云计算的电力能源大数据清洗模型构建[J]. 自动化仪表,2022,43(1): 72-76.
- [2] 谢智颖,何原荣,李清泉. 基于时空相关性的公交大数据清洗[J]. 计算机工程与应用,2022,58(1): 113-121.
- [3] XIE L P, LU C H, LIU Z E, et al. The evaluation of automobile interior acceleration sound fused with physiological signal using a hybrid deep neural network[J]. Mechanical Systems and Signal Processing, 2023,184:109675.
- [4] 沈沛,毛海涛,胡文林,等. 面向时序的相似重复数据清洗算法优化[J]. 计算机时代,2022(9):68-72,77.
- [5] 韩红桂,鲁树武,伍小龙,等. 基于改进型 SVM 的城市污水处理过程异常数据清洗方法[J]. 北京工业大学学报,2021,47(9):1011-1020.
- [6] WANG D P, FAN J F, FU H L, et al. Research on optimization of big data construction engineering quality management based on RNN-LSTM[J]. Complexity, 2018:1-16.
- [7] YANG K, DING Y L, JIANG H C, et al. A two-stage data cleansing method for bridge global positioning system monitoring data based on bi-direction long and short term memory anomaly identification and conditional generative adversarial networks data repair[J]. Structural Control and Health Monitoring, 2022,29(9): e2993.1-e2993.19.

收稿日期:2022-10-27