

基于流数据的列车在线状态监测方法研究与应用

冯富人

(同济大学 铁道与城市轨道交通研究院, 上海 201804)

摘要: 为实时地监测列车在线状态, 传统的阈值检测方法只能从数值表现上对设备或系统状态做出判断, 而忽略了数据趋势性变化所反映出的列车健康状态变化信息。文章针对列车在线运行过程中产生的流数据提出了一种基于一致性表现的流数据分析方法, 其在离群异常点提供报警的同时, 能够从特征变化趋势中挖掘信息, 同时开发了一套应用于列车多功能车辆总线 MVB 网络的智能维护设备来搭载和验证算法; 基于列车实测数据和故障注入技术在实验室模拟了列车总线网络测试环境, 以列车气制动系统为分析对象对所提方法和所开发的设备进行了测试。结果表明, 其能够有效地对突发性异常进行报警, 在健康状态早期、中期和后期的报警识别率分别为 85.7%, 71.4% 和 57.1%, 且能够较好地识别和监测设备性能的衰退。

关键词: 流数据; 智能维护; 在线监测; MVB 网络; 一致性表现; 列车总线

中图分类号: U279.3

文献标识码: A

文章编号: 2096-5427(2021)01-0070-06

doi:10.13889/j.issn.2096-5427.2021.01.200

Research and Application of an Online Train Condition Monitoring Method Based on Streaming Data

FENG Furen

(Institute of Rail Transit, Tongji University, Shanghai 201804, China)

Abstract: In order to detect the online status of trains in real time, traditional threshold detection methods can only make judgments on the status of equipment or systems based on numerical performance, while ignoring the train health status changes reflected in the trend of data. This paper proposes a consistent performance-based streaming data analysis method for the streaming data generated in the online train operation process. It provides alarms for outliers and abnormal points, and can extract information from the characteristic change trends. At the same time, an intelligent maintenance equipment which applied to the train multi-function vehicle bus MVB network is designed to carry and verify the algorithm. After that, the train bus network test environment is simulated in the laboratory based on the train measured data and fault injection technology, and the train air brake system is used as the analysis object. Tested with the equipment, the results show that it can effectively alarm sudden abnormalities. In the early, middle and late stages of the health state, the alarm recognition rate is 85.7%, 71.4% and 57.1%, and it can better identify and detect equipment performance degradation.

Keywords: streaming data; intelligent maintenance; online monitoring; MVB bus; consistent performance; train bus

0 引言

轨道交通车辆在服役期间会产生巨量的数据, 如

列车轴速、受电弓电压、空气弹簧压力等。随着列车运行, 这些数据会被动态地实时更新, 形成了区别于传统数据库中静态数据的流数据。这些数据反映了列车的状态信息, 对于保障列车安全运营和健康管理而言具有重要意义, 而流数据实时性的特点也令其较历史数据具有更多有用的信息。传统的阈值判断方

收稿日期: 2020-10-27

作者简介: 冯富人(1995—), 男, 硕士研究生, 主要研究方向为列车安全控制。

法过多地关注当前数据流的数值表现，却忽视了流数据变化趋势所隐藏的信息：如果阈值的裕量过大，则服役前期的识别效果差；若阈值的裕量过小，则随着性能的衰退，服役后期的误报率明显提高。

流数据处理方法重在能够实时有效地处理数据流，其要求算法的时间复杂度低，以满足流速要求，主要方法包括滑动窗体技术和关联技术等^[1]。Chang Joong Hyuk^[2]和李国徽^[3]等针对滑动时间窗方法分别进行了研究，以满足其短时流数据中模式信息获取的需要。Seo Bok I, Kim Jae In等^[4]提出了基于流环境的关联算法，能够快速地从提取的数据项中发现关联规则。具体到应用层面，基于内存的分布式流处理框架 Spark Streaming 是现今应用最为广泛的流数据处理平台，其能够较好地实现流数据信息的接收、快速处理和计算需要。基于 Spark 平台，吴海波、施式亮等^[5]建立了瓦斯浓度流数据异常检测系统，提高了瓦斯风险评价时效。聂睿和黄鹏^[6]将历史数据和实时数据进行联合分析，提高了实时采集数据的吞吐量以及准实时分析的计算速度，提高了飞机试飞效率。李欣^[7]实现了对交通流数据的高效处理以及预测，但由于 Spark 平台要求进行分布式计算，对算力要求高，因此难以满足在线单机设备的相关要求。刘光俊、张杜玮等^[8]也只是从合理性角度设计了一车一协议的可配置化方案，为 PHM 系统展示和数据挖掘应用提供了数据基础。基于数据驱动对列车状态进行诊断和健康管理已经得到越来越广泛的应用。秦方方、郑财晖等^[9]针对通用超限报警检测方法存在的不足，提出一种供电系统接地绝缘故障实时诊断与预测方法。潘莹^[10]从数据采集、数据分析和数据应用 3 个层级来实现对地铁车辆的故障预测和健康管理。

鉴于阈值判断方法对设备性能衰退方面分析能力的不足，结合列车流数据的动态性、实时性和海量性特点，本文针对列车运行过程中流数据特征表现异常的实时报警和性能衰退监测问题，提出了一种基于一致性表现的流数据分析方法；为验证该方法，开发了一套应用于多功能车辆总线（multifunction vehicle bus, MVB）的智能设备，其能够满足流数据处理和列车在线状态监测的需要；最后，基于列车实测数据和故障注入技术，对所提方法和所开发的设备进行了验证。

1 基于一致性表现的流数据分析方法

本文提出一种流数据处理方法，其重点在于针对异常特征进行实时报警，同时基于长时间序列内数

据性能的变化规律，对列车设备健康状态进行分析。其基本思想在于某一系统或设备的运行参数或特征应当按设计要求表现出较好的稳定性；对于偏移设计范围的值，则应被视为异常情况。由浴盆曲线可知，随着后期服役时间的增长，设备出现异常的概率大大增加，表现在数据方面，即状态参数明显偏移目标数值，出现大范围波动的情況^[11]。因此可以从聚类的质心迁移规律、离散程度和离群点的偏移量等角度对流数据进行实时分析。

设算法的滑动时间窗宽度为 m ，即一次分析的有序数据集 $D=(a_1, a_2, \dots, a_m)$ 。当有新的数据增加时，将其添加到集合的末尾，并弹出集合中排序第一也就是距今最久的数据点，以保持时间窗的大小不变。不失一般性，以两特征为例进行分析。对数据集 D ，依据质心公式其若有质心 P ，则对任意一点 $a_i(a_i \in D; i=1, 2, \dots, m)$ ，其平均密度 ρ 的计算如下：

$$\rho = \frac{\sum_{i=1}^m \frac{1}{d(a_i, P)}}{m} \quad (1)$$

式中： $d(a_i, P)$ —— a_i 到 P 的欧式距离。

ρ 表征了聚类的离散程度。 ρ 越大，说明数据越集中，数据的一致性越好。若 t 时刻平均密度 ρ_t 小于能接收的最小密度 $\rho_{\min}$ ，即数据离散程度过大，则视设备为需要进行调整和检修。

若新增数据点 a_{m+1} 出现明显偏离聚类情况，即 $d(a_{m+1}, P) > \varepsilon$ ，其中 ε 为聚类最大的邻域范围，则设定该超过聚类邻域的点为离群点，并予以报警，这也就是流数据分析的动态实时报警功能。一般来说， ε 由设备的数值设计冗余决定，其在设备性能优良阶段明显大于聚类的半径。

随时序变化，数据集更新为 $D'=(a_{T+1}, a_{T+2}, \dots, a_{T+m})$ ，其质心由 P 迁移到 P' 。随着数据集的不断更新，根据质心迁移函数 $f(P, t)$ ，在 T 时刻，由质心到设定检修线的距离 $d(P_T, L_{jk})$ 小于设定值 d_{jk} ，则设定为设备的状态难以满足服役要求，需要进行检修和更换。这也就是流数据从性能变化趋势中挖掘的有效信息。

图 1 示出流数据分析方法在不同服役阶段数据聚类表现和实时报警情况的示意。图 1(a) 示出服役状态初期设备的性能表现。可以看到，聚类情况十分明显，聚类半径小，密度大。落在聚类邻域 ε 外的点均被判定为离群点并予以报警。随着服役期的增长，状态过渡到图 1(b) 所示状态，其聚类半径增大，密度减小，且随着整体性能的衰退，聚类的质心发生偏移。因为质心的迁移，早期离群点在此时很可能不

再被判定为离群点，即设备的整体性能出现了衰退现象，该特征表现已成为主流趋势。随着服役时间的继续增长，状态发展至图 1（c）所示情况，可以看到，聚类偏移明显，数据离散程度大，其不满足预期要求的点的出现概率大大提高。

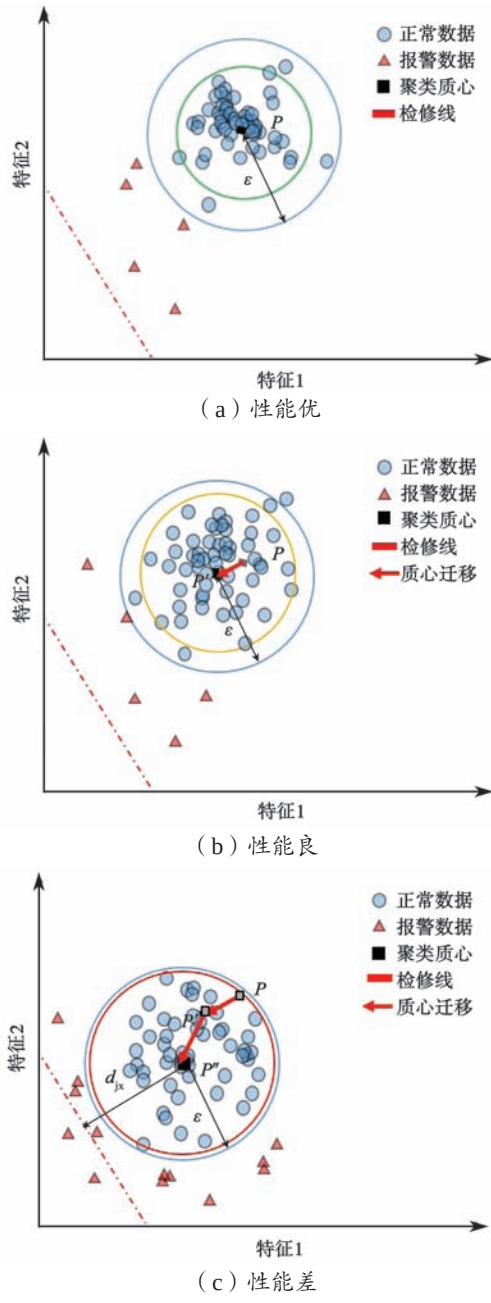


图 1 不同服役阶段数据聚类表现和实时报警情况示意

Fig. 1 Schematic diagrams of data clustering performance and real-time alarm conditions in different service stages

可以利用反映数据数值表现的、到检修线距离 d_{jx} 和反映数据离散程度的最小密度 ρ_{min} 来对设备提出检修策略。预警邻域范围 ϵ 需要针对分析的对象并结合专家经验进行设定，以避免出现过高的误报率，而 ρ_{min} 和 d_{jx} 可以通过对历史数据的训练并结合设备的检

修周期和安全等级等参数进行设定。

本文所提出的流数据分析方法的优势在于：

- (1) 不会因少数异常值影响对性能衰退结果的判断，对设备状态的异常表现只会进行报警，而长时间窗的计算冲淡了少数异常的影响；
- (2) 可以基于不断更新的流数据对数据表现进行实时地动态响应；
- (3) 从数值表现和数据离散程度两个角度对趋势进行分析；
- (4) 计算方式相对容易，较神经网络等方法而言算力少，可满足在线单机测试设备计算能力的要求。

2 基于流数据方法的列车在线状态监测设备开发

流数据处理的关键在于实时准确地获取列车运行状态参数，进行快速分析后，将结果加以反馈。传统安装传感器的方式成本高，布置复杂，且难以多方面获取到列车状态信息来实现多变量联合分析和信息融合。为验证和应用所提流数据分析方法，本文开发了一种适用于 MVB 总线网络的智能监测设备（图 2），其主要包括

- (1) 负责数据处理和算法应用的工控机；
- (2) 负责网络对接的 MVB 网卡网络接口单元 (network interface unit, NIU)；
- (3) 负责数据远程发送的远程数据发送模块 (data transfer unit, DTU) 及供电等外围电路，其能够根据 MVB 协议获取总线过程数据。

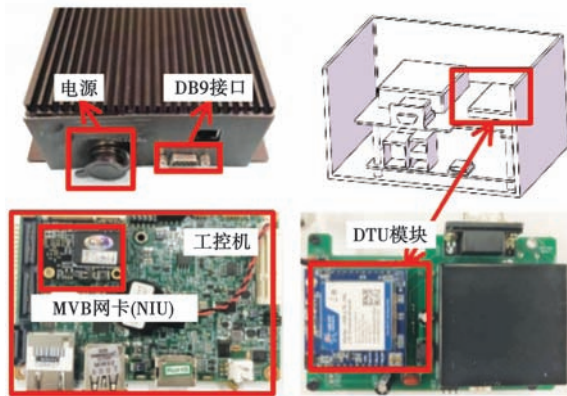


图 2 设备外观与内部结构

Fig. 2 Appearance and internal structure of the equipment

MVB 总线是应用于车厢或者固定编组这一特定范围内的通信网络，其为车厢内诸多功能（如门控、制动、空调、照明等）的自动实现、消息的传送、资源的共享及各设备之间的合理配合提供了可靠、顺畅

的信息交换通道^[12]。智能监测设备作为 MVB 总线网络中的一个节点出现，其安装位置如图 3 所示。

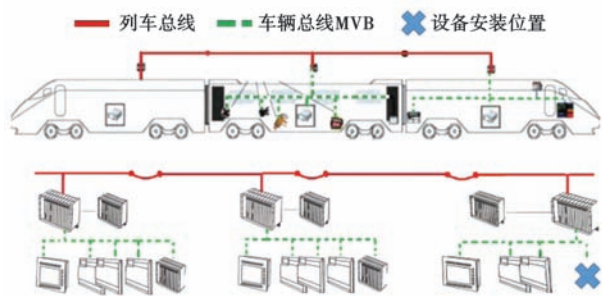


图 3 设备安装位置
Fig. 3 Installation site of the equipment

在得到分析结果后，一方面，智能监测设备将在源端口上利用广播方式通过总线发送给列车网络控制系统（train control and management system, TCMS），以便列车控制系统对实时报警的结果加以研判；另一方面，为更加有效地保存和综合利用列车数据，进一步弥补单机计算能力的不足，该设备将部分加工后的信息通过 DTU 模块并利用 4G 网络发送到云端的服务器中。其传输方式如图 4 所示。



图 4 分析结果的传输方式
Fig. 4 Transmission method of analysis results

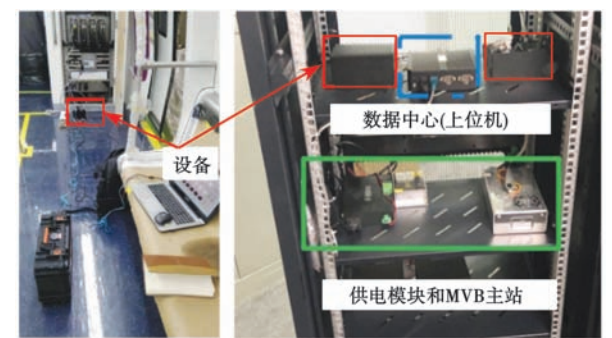
为解决列车运行线路环境中可能存在的网络信号不佳等问题，对设备与服务器接收程序制定了相应的校验 - 重传机制，以保障在轨运行期间数据保存的准确性和完整性。列车智能运维团队一方面可以通过列车通信网络实时监测列车状态；另一方面也可以利用手机随时随地地观察其性能表现、故障状态、运行里程及综合评分等信息，从而有的放矢地指导列车的健康维护。

3 测试验证

为验证本文所提方法和设备的有效性和实用性，在实验室内模拟列车总线 MVB 网络环境，基于列车实测数据并在数据中根据需求注入偶发性异常故障和衰退性故障进行测试验证。

3.1 测试环境

测试数据的来源为某地铁列车的在线测试数据。选用某公司生产的 MVB 主站模拟列车 MVB 网络环境，数据中心（上位机）将保存的实测列车数据利用故障模拟软件注入部分故障或异常后，以串口通信的方式发送给 MVB 主站，主站的 EMD 接口与设备相连接。列车测试与实验室测试环境如图 5 所示。



(a) 列车数据采集环境 (b) 实验室模拟测试环境
图 5 数据的采集与实验室的环境模拟
Fig. 5 Data collection and laboratory environment simulation

MVB 主站具有模拟功能，能够模拟出多个包含源端口和宿端口的虚拟设备来对设备进行测试。数据中心（上位机）在通过故障注入模拟出总线发送的数据后，首先将其转化为 MVB 主站能识别的传输格式，然后根据设备软件所面向的 MVB 协议，将发送数据按照端口类型和端口长度进行配置，配置成功，设备就能直接与模拟设备的对应端口进行信息的接发，也就意味着设备可以直接应用于 MVB 协议对应的列车总线网络中。图 6 示出实验室内 MVB 网络环境示意。

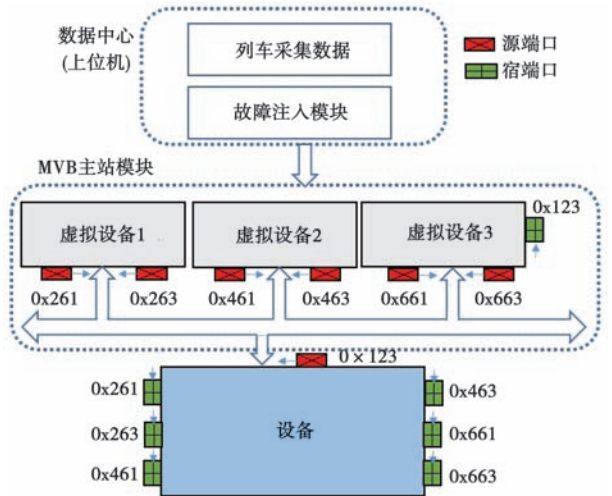


图 6 实验室内 MVB 网络环境示意
Fig. 6 Schematic diagram of MVB network environment in the laboratory

3.2 测试对象及测试内容

本文以列车气制动系统为主要测试分析对象，同

时参考风源系统、轴速控制装置等。列车气制动系统中,制动缸压力作用于基础制动装置,通过闸瓦或制动盘将列车高速运行时的巨大动能转化为热能。利用黑箱理论^[13]对列车气制动系统进行分析,制动缸压力 P 是气制动系统气路部分输出的最后结果,其直接影响制动能力的好坏(图7)。

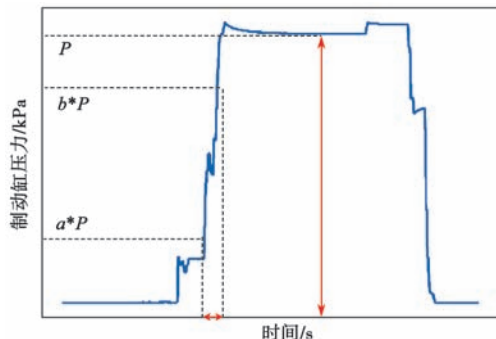


图7 制动缸压力变化与特征选取

Fig. 7 Change of brake cylinder pressure and feature selection

一般而言,选取制动缸压力上升时间(表征制动缸的快速响应能力,其中 a 和 b 为小于1的参数,需要根据制动缸压力的数值表现选择特征)和制动缸的稳态压力(表征制动能力的强弱)为特征进行分析。而这些特征又与列车载荷和制动级位等有关,故分析时需要对其进行标准化处理,将其变为无量纲的值。

测试通过上位机中的故障注入功能与验证仿真平台软件注入故障信息,如图8所示。

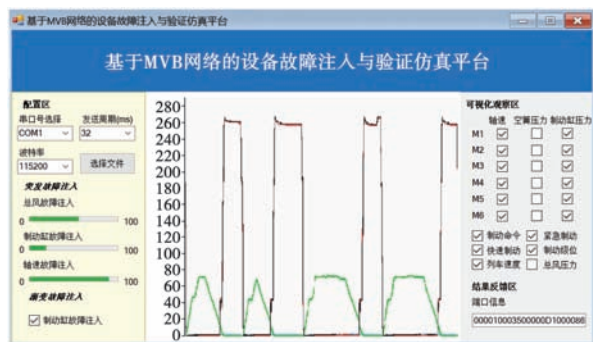


图8 上位机故障注入软件

Fig. 8 Upper computer fault injection software

测试中发送的信息包括制动指令及空簧压力、轴速、总风压力、各轴制动缸压力等物理量信息。测试的内容包括:

(1) 偶发性故障测试。注入添加不同程度的偶发性异常,以检验算法是否能够对其进行识别;

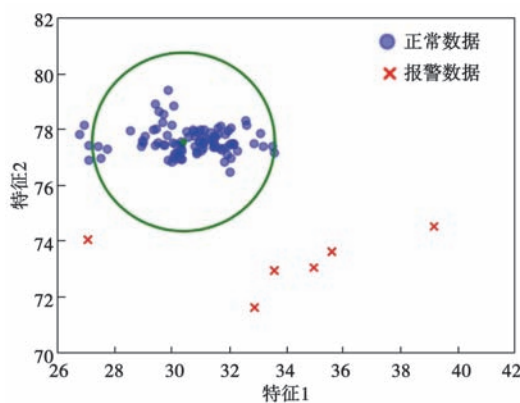
(2) 衰退性故障测试。注入衰退性故障,即依据设定的衰减函数,对数据造成持续性的下降与幅度更大的噪声干扰,观测其质心变化情况及聚类的密度变化情况,假定了一条检修线,观测算法是否能在相

应时刻提出相关建议;

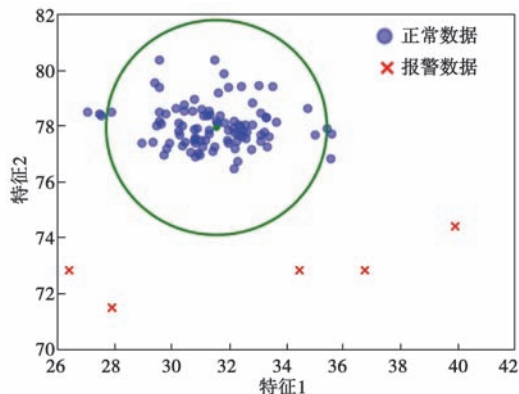
(3) 数据通路测试。对设备的分析结果经总线上传、云端发送和手机界面显示的数据通路进行验证。

3.3 测试结果

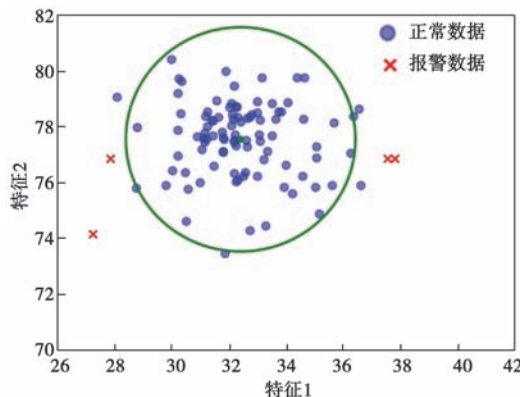
以每次制动为一个分析点,其制动缸压力上升特征和稳态压力作为两个特征(标准化处理后),进行第1~100, 101~200和201~300次制动过程的流数据分析。期间,每一阶段随机注入7次偶发性故障,全程持续注入衰退性故障,利用主站模拟的分析结果接收宿端口对结果进行观察,同时利用云端服务器对分析结果加以记录,结果如图9所示。



(a) 第1~100次制动流数据分析结果



(b) 第101~200次制动流数据分析结果



(c) 第201~300次制动流数据分析结果

图9 试验测试结果

Fig. 9 Results of the test

结果显示，第 1~100 次突发性异常识别率为 85.7%，第 101~200 次突发性异常识别率为 71.4%，第 201~300 次突发性异常识别率为 57.1%。造成识别率降低的主要原因是性能衰退后注入的偶发故障导致数据又偏离到正常的区域带内，致使系统无法正确识别。而质心偏移轨迹为 (30.245 → 31.872 → 32.316, 77.616 → 78.012 → 77.425)，特征 1 变化明显基本满足要求，而特征 2 变化幅度不大，且出现了数据的逆向变化。

由图 9 可以看到，随着偶发性故障和衰退性故障的添加，聚类半径越来越大，数据特征出现较大的浮动和离散，而聚类的质心也发生了迁移。同时，流数据方法对偶发性故障进行了报警，随着性能的逐渐退化，开始予以报警的点因性能的变化将不再成为报警点，满足了动态异常过程识别的需要，而非单纯地根据单次制动过程制动缸压力的数值表现进行对异常的判断，且识别率大于 50%，整体而言满足了设计要求。但也可以看到，质心偏移轨迹仍需进行更为精准的优化与参数校准，从而避免检修预警方式出现问题。

4 结语

为解决列车运行过程中对故障或设备异常表现进行实时报警的问题，同时基于长时间序列内数据性能的变化规律对列车设备健康状态进行分析，本文从数据特征一致性角度出发提出了一种流数据分析方法，研究和设计了一套适用于 MVB 总线的智能监测设备，实现了“列车-设备-手机端”的列车健康状态分析系统，并以制动系统为例，基于实测数据与故障注入技术，在实验室模拟的列车 MVB 网络环境下对所提方法和研制的设备进行了验证。结果表明，本文所提方法和研制的设备能满足列车动态异常过程识别的需要。为了提高方法的精度和异常过程识别的准确性，后续将利用某一特定列车的真实历史数据对方法进行训练和优化；通过对聚类质心迁移曲线和聚类半径变化函数的研究，更好地修正异常过程判别条件。

参考文献：

[1] 孙玉芬, 卢炎生. 流数据挖掘综述 [J]. 计算机科学, 2007(1):1-

5, 11.
SUN Y F, LU Y S. An Overview of Stream Data Mining[J]. Computer Science, 2007(1): 1-5, 11.
[2] CHANG J H, LEE W S. A sliding window technique for open data mining over data streams[J]. The KIPS Transactions: Part D, 2005, 12-D(3): 335-344.
[3] 李国徽, 陈辉. 挖掘数据流任意滑动时间窗口内频繁模式 [J]. 软件学报, 2008(10): 2585-2596.
LI G H, CHEN H. Mining the Frequent Patterns in an Arbitrary Sliding Window over Online Data Streams[J]. Journal of Software, 2008(10): 2585-2596.
[4] SEO B I, KIM J I, HWANG B H. A Method for Frequent Itemsets Mining from Data Stream[J]. The KIPS Transactions: Part D, 2012, 19-D(2): 139-146.
[5] 吴海波, 施式亮, 念其锋. 瓦斯浓度流数据实时异常检测方法 [J]. 计算机与数字工程, 2019,47(5): 1086-1090, 1105.
WU H B, SHI S L, NIAN Q F. Real-time Anomaly Detection for Gas Concentration[J]. Computer and Digital Engineering, 2019, 47(5): 1086-1090, 1105.
[6] 聂睿, 黄鹏. 基于 Spark 的试飞遥测数据准实时分析技术研究 [J]. 中国科技信息, 2020 (5): 51 - 54.
[7] 李欣. Spark 框架下交通流数据高效处理方法及其应用 [J]. 北京大学学报 (自然科学版), 2018, 54(6): 1227-1234.
LI X. Efficient Traffic Flow Data Processing Method and Its Application Based on Spark Framework[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2018, 54(6): 1227-1234.
[8] 刘光俊, 张杜玮, 张珍文. 基于一车一协议配置化 Spark 实时流数据解析的研究与应用 [J]. 通讯世界, 2017(19): 275-276.
[9] 秦方方, 郑财晖, 李学明, 等. 列车供电系统接地绝缘故障实时诊断与预测 [J]. 控制与信息技术, 2020(4): 80-86, 95.
QIN F F, ZHENG C H, LI X M. Real-time Diagnosis and Prediction of Grounding Insulation Fault in Train Power Supply System[J]. Control and Information Technology, 2020(4): 80-86, 95.
[10] 潘莹. 故障预测和健康管理技术在地铁车辆运维中的应用 [J]. 控制与信息技术, 2020(4): 91-95.
PAN Y. Application of Fault Predication and Health Management in Metro Vehicle Operation and Maintenance[J]. Control and Information Technology, 2020(4): 91-95.
[11] 肖文斌. 基于耦合隐马尔科夫模型的滚动轴承故障诊断与性能退化评估研究 [D]. 上海: 上海交通大学, 2011.
[12] 谢维达, 钱存元. 列车通信与网络 [M]. 北京: 中国铁道出版, 2014: 105-121.
[13] 葛祥国, 肖桃华. 浅析黑箱理论 [J]. 中国教育技术装备, 2011(18): 93-94.
GE X G, XIAO T H. Talking about Black Box Theory [J]. China Educational Technology & Equipment, 2011(18): 93-94.