

基于动态 HDBSCAN 的动力电池单体电压不一致性故障识别方法

曾建邦^{1,3}, 胡亦非¹, 黄豪^{1,3}, 周熙雷², 李彬斌^{1,3}, 沈超群^{1,3}, 覃清梅^{1,3}, 单丰武^{1,3}

(1. 交通智能运维技术与装备教育部重点实验室, 江西 南昌 330013; 2. 抚州市东理能源和环境研究院, 江西 抚州 344000; 3. 江西江铃集团新能源汽车有限公司, 江西 南昌 330013)

摘要: 单体电压不一致性故障是电动汽车动力电池系统的典型故障, 不仅会显著降低电池组综合性能, 还会对车辆行车安全构成严重隐患。为此, 本文依托车企大数据监管平台的实际运行数据, 提出一种基于联合特征增强的动态 HDBSCAN 动力电池单体电压不一致性故障识别方法, 充分利用动态 HDBSCAN 算法无需预设聚类数量、可自适应识别多密度层次异常簇等突出优势。首先, 为保留电压整体偏移信息, 强化模型对极端异常值的敏感度, 构建了曼哈顿距离-切比雪夫距离联合特征矩阵。其次, 为提升故障识别的鲁棒性与模型泛化能力, 引入基于 ROC 曲线与 Youden 指数无监督自适应优化的互可达距离阈值判定机制。最后, 随机选取 5 辆存在“单体一致性差”的故障车辆和 1 辆正常车辆进行分析, 结果表明: 相较于现有监管平台以压差阈值为基准的检测方法, 该方法对 5 辆故障车的异常单体识别时间分别提前 9 天 6 小时 10 分钟、19 天 20 小时 42 分钟、12 天 22 小时 57 分钟、12 天 11 小时 3 分钟和 18 天 22 小时 35 分钟; 通过与 OPTICS、DBSCAN 聚类算法的对比实验进一步发现, 本文提出的聚类算法在故障识别时效上表现更优。可见, 本文所提方法在电动汽车动力电池故障诊断领域具有较好的工程应用价值。

关键词: 电动汽车; 动力电池; 故障识别; 单体电压不一致性; 动态 HDBSCAN

中图分类号: U472

文献标志码: A

Dynamic HDBSCAN-based Fault Identification Method for Inconsistent Cell Voltage in Power Batteries

Zeng Jianbang^{1,3}, Hu Yifei¹, Huang Hao^{1,3}, Zhou Xilei², Li Binbin^{1,3}, Shen Chaoqun^{1,3}, Qin Qingmei^{1,3}, Shan Fengwu^{1,3}

(1. Key Laboratory of Transportation Intelligent Operation and Maintenance Technology and Equipment(East China Jiaotong University), Ministry of Education, Nanchang 330013, China; 2. Fuzhou Dongli Institute of Energy and Environment, Fuzhou 344000, Jiangxi, China; 3. New Energy Vehicle Corporation, Jiangxi Jiangling Motors Group, Nanchang 330013, China)

Abstract: The single-cell voltage inconsistency fault is a typical fault in the power battery system of electric vehicles, which not only significantly degrades the overall performance of the battery pack but also poses a serious threat to vehicle safety. To address this issue, this paper proposes a method for identifying single-cell voltage inconsistency faults in power batteries based on a dynamic HDBSCAN algorithm with joint feature enhancement, utilizing actual operational data from a big data supervision platform of an automotive enterprise. The method fully leverages the advantages of the dynamic HDBSCAN algorithm, such as not requiring a preset number of clusters and its ability to adaptively identify abnormal clusters with multiple density levels. First, to preserve the overall voltage deviation information and enhance the model's sensitivity to extreme outliers, a joint feature matrix based on Manhattan distance and Chebyshev distance is constructed. Second, to improve the

收稿日期: 2025-12-26

基金项目: 国家自然科学基金项目(51206171); 江西省重点研发项目(20243BBG71009); 江西省自然科学基金项目(20242BAB25273, 20244BAB28006); 江西省研究生创新专项资金计划(YC2023-S505)

robustness and generalization capability of fault identification, an unsupervised adaptively optimized threshold determination mechanism for the mutual reachability distance is introduced, based on the ROC curve and Youden index. Finally, five faulty vehicles exhibiting "poor cell consistency" and one normal vehicle were randomly selected for analysis. The results indicate that, compared to the existing supervision platform's detection method based on voltage difference thresholds, the proposed method identifies abnormal cells in the five faulty vehicles 9 days 6 hours 10 minutes, 19 days 20 hours 42 minutes, 12 days 22 hours 57 minutes, 12 days 11 hours 3 minutes, and 18 days 22 hours 35 minutes earlier, respectively. Further comparative experiments with OPTICS and DBSCAN clustering algorithms demonstrate that the proposed clustering algorithm achieves superior performance in terms of fault identification timeliness. Consequently, the method presented in this paper exhibits promising engineering application value in the field of electric vehicle power battery fault diagnosis.

Key words: electric vehicles; power batteries; fault identification; cell voltage inconsistency; dynamic HDBSCAN

随着全球能源日趋紧张和环境污染日益加重,绿色出行成为新时代下陆运交通实现节能减排的重要方式。作为电动汽车的心脏部件,动力锂离子电池内单体一致性影响着整车的动力、安全性能^[1]。由于生产误差和运行条件不均匀等因素的影响,单体间存在不一致性。并且随着使用时间延长,该不一致性会被逐渐放大,如果不一致性增加到一定程度,会引起电池包性能恶化甚至引发起火或爆炸等恶性事故^[2]。若能在单体电压偏离初期发现故障单体,则可以通过调整充放电策略或更换故障单体等措施,减少甚至是避免安全事故的发生。因此,研究动力电池单体一致性故障识别方法,对提高电动汽车全生命周期的安全性和可靠性具有显著的现实意义。

目前,国内外学者围绕动力电池故障诊断技术已开展了大量研究工作,主流技术路径可划分为知识驱动、模型驱动与数据驱动三大类^[3]。其中,知识驱动的动力电池故障诊断技术,主要通过人工智能手段从历史运行数据中挖掘隐含规律并构建知识库。但该技术面临两大核心瓶颈:一方面,新能源车企为保障数据安全,对车辆运行数据采取严格保密措施,难以覆盖多场景、大规模的电池故障知识库^[4];另一方面,专家经验与知识体系存在固有局限性,诊断过程易受主观因素干扰,进而影响诊断结果的准确性^[5-6]。

基于模型驱动的故障诊断,其基本思路是建立高精度的电池等效模型用以表征系统运行状态。在已有的研究中,常用的模型有电池等效电路模型、电化学模型、电-热耦合模型、电化学-热耦合模型等^[7]。此类模型能够在一定程度上表征电池的动态特性与内部反应过程,并为电池的状态估计与故障识别提供相应的理论依据。然而动力电池是高度非线性、强耦合等复杂系统,其运行状态受温度、负荷和电流变化以及传感器噪声等多维因素影响,若要得到高精度的模型需要进行大量的实验测试来确定模型参数,而且不同工况还需要进行多轮校核,在模型的确定上会耗费大量时间,并且模型的鲁棒性差^[8-9]。

近年来,基于数据驱动的动力电池故障诊断技术随着机器学习以及大数据技术的迅速发展逐渐走入人们的研究视野,其基于车联网或者电池管理系统获取动力电池历史运行数据,提取故障特征并构建故障诊断算法以实现故障识别,常用算法包括信号处理方法^[10-11]、统计分析方法^[12-13]和机器学习方法^[14-16]。在上述方法中,作为机器学习方法之一的聚类方法,因其能够同时实现数据的离群故障检测且无须对样本进行任何标注,而被广泛应用于动力锂离子电池故障诊断领域。

目前,常见的有 K-means、DBSCAN、OPTICS 等方法。何佳星等^[17]使用主成分分析,通过将时序化的单体电压数据进行降维处理,以简化数据冗余、减少计算负担为前提,通过 K-means 聚类实现对系统过压、欠压、压差告警等故障的正确判别。李一帆等^[18]对故障告警时刻的单体电压数据,直接采用 K-means 聚类,可有效检测异常单体,可为故障排查提供技术支持。Li 等^[19]提出了一种基于 DBSCAN 密度聚类的三元锂离子电池故障诊断方法,通过对电池运行过程中的多维特征数据进行聚类分析,实现异常工况与潜在故障单体的有效识别。近年来,本团队围绕聚类方法在动力电池故障诊断领域的应用开展了一系列研究:为克服传统 K-means 算法需预设聚类数量、对初始中心敏感、难以适应复杂故障模式等方面的不足,提出了基于动态 k 值 K-means++ 聚类的电动汽车动力电池电压不一致故障识别方法,实现了电池单体一致性故障的早期微弱异常辨识及异常单体定位^[20];利用离散弗雷歇距离对时序曲线形态的高敏感性与电压偏差值的幅值量化优势,克服了传统特征对动态波动与渐进式异常捕捉不足的困难,基于电压偏差值与离散弗雷歇距离构建二维特征空间,提出了一种结合 OPTICS 聚类与电压异常指数的电池单体不一致性故障分析方法^[21],该方法既能够实现早期识别,还可以进行异常单体定位并量化故障等级。虽然 K-means、DBSCAN 和 OPTICS 聚类由于结构简单、计算速度快等原因被广泛使用,但 K-means 需要事先人为设定聚类数目 K,且基于簇为球形结构的假设,对于非球形分布或密度差异较大的样本数据难以取得理想效果^[22-24];DBSCAN 对邻域半径参数和最小样本数的选取高度敏感,参数设置具有较强经验性,同时在大规模数据场景下计算复杂度较高,不利于在线实时应用^[25-27];OPTICS 虽然通过可达距离排序刻画数据结构,但在实际应用中参数设置仍然敏感,且其聚类结果解释与异常判定过程往往依赖人为阈值划分,主观性较强,增加了不确定性^[28,29]。

为避免上述问题,本文采用基于层次密度的带噪声应用空间聚类(Hierarchical Density-Based Spatial Clustering of Applications with Noise, HDBSCAN)算法,能够实现层次密度聚类与自适应参数选择^[30]。该算法在 DBSCAN 的基础上,通过引入密度可达的层次聚类结构和聚类稳定性分析,能够自动基于密度层次树确定最佳聚类数量,并识别数据中的稳定聚类结构,从而有效克服了传统 DBSCAN 方法需要手动设置邻域半径和最小样本数参数的局限性。然而,在实车运行过程中,由于电池的电压分布和噪声水平呈动态变化,传统的 HDBSCAN 仍然会存在聚类敏感问题,不能做到故障识别过程中的持续和健壮。为此,本文提出一种基于动力电池实际运行电压场景构建的动态 HDBSCAN 聚类算法,该算法利用车载终端平台采集的历史电压数据,构建融合曼哈顿距离(Manhattan Distance, MD)与切比雪夫距离(Chebyshev Distance, CD)的二维联合特征,兼顾电压整体偏移趋势与极端异常捕捉能力。将特征输入动态 HDBSCAN 算法后,通过优化互可达距离阈值参数,模型能够识别不同密度分布的异常簇结构,有效提升故障检测的准确性和鲁棒性。最后,在相同数据集上,通过与 DBSCAN 和 OPTICS 聚类方法进行对比实验,验证了所提方法在故障识别时效性上具有显著优势。

1 数据预处理及特征提取

为提升新能源汽车质量和运行安全监测,工信部要求建立国家、地方和车企三级监控平台。本文使用的电动汽车运行数据均来自车企监控平台,该平台严格遵循《电动汽车远程服务与管理系统技术规范》(GB/T 32960-2016)国家标准,车载终端以小于 30s 的频率(本文数据来自某车企平台,其数据频率为 20s)通过通信协议发送车辆实时运行数据给车企平台。车企平台再按要求将数据上传至地方和国家平台,

其中与动力电池相关的数据为数据聚类法识别故障提供了重要来源。本文使用的车辆其电池包均为 3P81S 配置,即由 81 个单体电池组成,并采用 3 并 81 串的结构。表 1 展示了本文所用车辆相关电池包主要参数,表 2 展示了与动力电池相关的整车、储能装置、极值和告警数据。

表 1 电池包参数

电池包参数	数值	电池包参数	数值
类型	三元锂 (NCM)	结构	3P81S
能量 (kWh)	31.15	单体额定电压 (V)	4.22
总额定电压 (V)	312	单体截止电压 (V)	3.0

Tab.1 Battery pack parameters

表 2 车企监控平台中与动力电池相关的数据

Tab.2 Data related to power batteries in the vehicle manufacturer monitoring platform

数据项	静态信息	动态信息
整车数据	\	数据采集时间、SOC、充电状态、总电压、总电流、绝缘电阻
极值数据	最高/最低电压电池组序号、单体蓄电池序号、最高/最低温度蓄电池组序号、探针序号	最高电压单体代号、单体电压最高值、最低电压单体代号、单体电压最低值、最高温度 探针号、最高温度值、最低温度探针号、最低温度值
报警数据	\	SOC 低报警、SOC 过高报警、SOC 跳变报警、温度差异报警、电池高温报警、单体电池过压报警、单体电池欠压报警、车载储能装置类型欠压报警、车载储能装置类型过压报警、车载储能装置类型过充报警、电池单体一致性差报警、绝缘报警

1.1 数据预处理方法

数据采集过程中可能存在缺失值、异常值、重复记录及噪声干扰等问题,为保障分析可靠性,必须进行数据预处理,具体清洗方法如下:

1) 重复数据清洗:若相邻两行数据完全相同,极有可能是数据采集过程中出现错误,这种情况可判定为重复记录。对于重复记录,本文选择将后一行数据移除。

2) 缺失数据处理:根据缺失数据的时间分布特征实施不同处理。连续型缺失:当缺失持续 3 个采样周期 ($\geq 1\text{min}$) 时,判定为系统级通信中断,排除该时段数据。离散型缺失:针对单点或双点 ($\leq 40\text{s}$) 缺失情况,则采用相邻单元格平均值插值法对缺失值进行插值处理,以确保数据时序连续性。

3) 异常数据处理:异常数据指的是电池电压、电流、温度等数据出现明显不合理的数值。例如,总电压和电机控制器输入电压超出 300-480V 范围、相邻点电压变化超过 50V; SOC 超出 0%~100%范围、以及在 20%~100%区间内突然出现 0%的点(前后点都在 20%~100%之间);绝缘电阻大于 10000k Ω 。此类异常数值通常是由传感器故障导致的个别测量值不符合常理,故将异常值排除在分析之外,并使用线性回归在时间维度上对数据进行插值补全。

4) 数据滤波处理: Savitzky-Golay 滤波器是一种广泛用于降噪时间序列数据的滤波方法,该方法能在保留信号特征的同时,有效地去除高频噪声^[28]。因此,本文使用 Savitzky-Golay 滤波器对原始电压信号进

行降噪处理。

1.2 特征参数提取方法

在单体电压随时间的变化曲线中，正常单体的电压曲线呈现出较高的一致性，异常单体则会偏离整体趋势，表现为离群点。在早期研究中常采用电压均值、方差、极差、标准差、偏度、峰度等时域特征来刻画电压分布特征^[31-33]，但这类特征无法刻画时间序列的动态演化特性。部分研究通过滑动窗口引入电压变化率等趋势特征增强对动态变化的表征能力^[34-35]，但这些特征易受噪声干扰。针对以上问题，本文提出一种二维特征，将曼哈顿距离与切比雪夫距离进行互补组合，在抑制噪声干扰的同时，能够有效表征电池电压的动态异常与稳态一致性变化。

1) 曼哈顿距离

曼哈顿距离的定义为在规则矩形网格上两点之间的最短路径长度。在二维空间中，给定两点 $a(x_a, y_a)$ 与 $b(x_b, y_b)$ ，曼哈顿距离可表示为

$$d_1(a, b) = |x_a - x_b| + |y_a - y_b| \quad (1)$$

曼哈顿距离能够刻画单体在整个窗口范围内的累计偏差，并在刻画电压差异方面表现出良好的灵敏性与鲁棒性。

切比雪夫距离的定义为在规则矩形网格上两点在所有坐标轴方向上的最大绝对差，在二维空间中，给定两点 $a(x_a, y_a)$ 与 $b(x_b, y_b)$ ，切比雪夫距离定义为

$$d_\infty(a, b) = \max(|x_a - x_b|, |y_a - y_b|) \quad (2)$$

切比雪夫距离强调单体在某一时刻的极端偏离程度，能够在全时间范围内识别出最显著的差异。

在滑动窗口下构建特征向量，设滑动时间窗口的大小为 v ，包含 ω 个采样时刻，动力电池系统由 n 个单体组成。定义单体电压矩阵为

$$U^{(v)} = \begin{bmatrix} U_{1,1} & \cdots & U_{1,n} \\ \vdots & U_{t,i} & \vdots \\ U_{\omega,1} & \cdots & U_{\omega,n} \end{bmatrix} \quad (3)$$

其中， $U_{t,i}$ 表示第 t 个采样时刻 ($t=1, 2, \dots, \omega$) 对应的第 i 个单体 ($i=1, 2, \dots, n$) 的电压值。

为更好的消除系统异常单体运行状态的影响，取窗口内所有单体电池在每个采样时刻的中位数电压曲线作为基准

$$\tilde{U}_t = \text{median}\{U_{t,1}, U_{t,2}, \dots, U_{t,n}\} \quad (4)$$

随后，计算出每个单体在整个窗口范围内的电压相对于基准电压 \tilde{U}_t 的曼哈顿距离与切比雪夫距离。

其中，曼哈顿距离刻画单体电压在整个窗口范围内的累计偏差，计算公式为

$$MD_i^{(v)} = \sum_{t=1}^v |U_{t,i} - \tilde{U}_t| \quad (5)$$

切比雪夫距离刻画单体电压在窗口内的最大瞬时偏差，其计算公式为

$$CD_i^{(v)} = \max_{1 \leq t \leq v} |U_{t,i} - \tilde{U}_t| \quad (6)$$

通过上述计算，可为每个单体构建包含累计偏差与瞬时极差的二维特征向量 $F_i^{(v)} = [MD_i^{(v)}, CD_i^{(v)}]$ ，进而汇总所有单体得到电池簇特征矩阵

$$\mathbf{F}^{(v)} = \begin{bmatrix} MD_1^{(v)}, CD_1^{(v)} \\ MD_2^{(v)}, CD_2^{(v)} \\ \vdots \\ MD_n^{(v)}, CD_n^{(v)} \end{bmatrix} \quad (7)$$

该矩阵为在窗口大小为 v 的情况下曼哈顿距离与切比雪夫距离两类互补的特征参数构造联合特征矩阵，能够在不同时间尺度下刻画单体电池的一致性，前者通过时间窗口内各时刻电压偏差的累计和，强调一段时期内的整体差异程度；后者通过时间窗口内各时刻电压偏差的最大值，突出该时期内的极端最大偏离情况，为后续基于动态 HDBSCAN 的聚类分析和单体不一致性检测提供输入基础。

2 基于动态 HDBSCAN 聚类的故障识别方法

在完成电池电压特征参数提取后，本研究进一步提出了一种动态 HDBSCAN 方法，用于动力电池单体不一致性故障的精准识别。与传统的静态聚类方法不同，本方法在保留 HDBSCAN 密度聚类核心框架的基础上，通过引入时间演化建模、双距离度量融合和动态核心距离平滑三大创新，实现了对电池状态随时间演变过程的深度刻画，具体步骤如下：

Step1: 特征矩阵构建。将 1.2 节中得到的联合特征矩阵 $F^{(v)} \in \mathbb{R}^{n \times m}$ 作为输入数据，其中 n 为单体数量， m 为特征维度，每一行表示单体 i 在时间窗口 v 内的特征向量

$$f_i^{(v)} = [MD_i^{(v)}, CD_i^{(v)}] \quad (8)$$

Step2: 动态核心距离与互可达距离(Mutual Reachability Distance, MRD)计算。基于上述特征，本研究对传统 HDBSCAN 的距离度量进行了改进，通过时间维度的平滑机制和双距离度量融合，构建动态距离度量。首先计算单体之间的联合特征的欧氏距离

$$d(p, q) = \|f_p^{(v)} - f_q^{(v)}\|_2 \quad (9)$$

p, q 代表两个不同的电池单体。

然后计算核心距离。传统核心距离仅基于当前窗口的 k -近邻计算，是静态的局部密度度量，忽略了电池状态在时间序列上的连续性与演化趋势。本研究提出的动态核心距离，通过指数平滑将前一窗口的核心距离信息融入当前计算，实现了对时间演化过程的显式建模

$$c_t(p) = \alpha \cdot cc_t(p) + (1 - \alpha) \cdot c_{t-1}(p) \quad (10)$$

式中, α 为平滑系数, $c_{i-1}(p)$ 为前一窗口的核心距离, $cc_i(p)$ 为当前窗口基于 k -近邻的核心距离, 其计算如下

$$cc_i(p) = \min_{q \in N_k(p)} d(p, q) \quad (11)$$

其中, $N_k(p)$ 是单体 p 的 k -近邻集合。

根据 Brown^[36]和 Makridakis^[37]对指数平滑参数的工程推荐范围 (0.6–0.8), 并结合电池电压序列的动态波动特性, 选取 $\alpha=0.7$ 。通过这种方式, 动态核心距离不仅反映了当前的密度特征, 还继承了历史状态信息, 使其在继承 HDBSCAN 对噪声和非球形聚类鲁棒性的同时, 能够捕捉复杂的时序演化模式。由此定义单体 p 相对于单体 q 的相互可达距离为

$$\text{MRD}(p, q) = \max \{c_i(p), c_i(q), d(p, q)\} \quad (12)$$

MRD 作为改进的相似性度量, 将核心距离与欧氏距离进行融合, 在保持局部几何结构信息的同时, 引入密度约束机制, 从而有效缓解传统距离度量对噪声点和异常点敏感的问题。同时, 由于核心距离的动态平滑特性, MRD 也具备了时间演化的记忆效应, 能够更准确地识别出随时间逐渐显现的不一致性故障。

Step3: 动态 HDBSCAN 聚类。利用 MRD 矩阵构建最小生成树(Minimum Spanning Tree, MST), 并基于边权阈值逐步分解得到层次聚类树。动态 HDBSCAN 通过多参数扫描, 构建最小生成树并考虑不同密度阈值下簇的持久性来定义稳定性,稳定性高的簇更可能对应正常单体群体, 低稳定性或孤立点则被识别为潜在异常, 计算不同簇的稳定性

$$S(C) = \sum_{i \in C} (\lambda_i^{\max} - \lambda_C) \quad (13)$$

式中, 簇 C 是通过 MST 分解得到的层次聚类树中的一个节点, 表示满足一定相似性条件的电池单体集合; λ_i^{\max} 表示单体 i 在簇 C 中的能够保持成员身份的最大保留尺度, 即最后一次作为簇 C 的核心成员时, 该簇对应的 λ , 刻画了单体 i 与簇 C 的关联强度; λ_C 为簇 C 的形成尺度阈值, 计算式如下

$$\lambda_C = \min_{p, q \in C} \lambda(p, q) \quad (14)$$

式中, $\lambda(p, q)$ 为相似性度量, 其值越大, 单体 p 和 q 之间的相似性越高, 其计算式如下

$$\lambda(p, q) = \frac{1}{\text{MRD}(p, q)} \quad (15)$$

Step4: 故障单体识别。聚类完成后, 单体被划分为正常簇与噪声点。为进一步增强判别效果并防止误判, 引入 MRD 阈值法判断异常单体

$$l(i) = \begin{cases} 1, & \overline{\text{MRD}}_i > \delta \\ 0, & \overline{\text{MRD}}_i \leq \delta \end{cases} \quad (16)$$

式中, δ 通过 ROC 曲线与 Youden 指数确定; 单体 i 的平均 MRD 计算如下

$$\overline{\text{MRD}}_i = \frac{1}{n-1} \sum_{j=1, j \neq i}^n \text{MRD}(i, j) \quad (17)$$

其中, 单体 i 与单体 j 之间的 MRD 值为 $\text{MRD}(i, j)$, 其反映了单体 i 与其他单体的平均相对距离。

Step5: 故障识别结果输出。最终输出包括正常单体簇、异常单体标签以及聚类可视化结果。

本文提出方法的流程图如图 1 所示。本模型突破了传统 HDBSCAN 的静态假设, 将时间维度的平滑机制引入核心距离计算, 并提出了动态核心距离与改进的互可达距离, 将历史状态信息与当前特征差异进行有机融合, 形成了一种兼具密度敏感性和时间记忆性的新型距离度量。这种度量方式不仅继承了 HDBSCAN 对噪声和复杂形状聚类的鲁棒性, 还能有效识别出那些在时间序列上逐渐偏离正常群体的异常单体, 显著降低了误报率。此外, 该方法无需预先设定聚类数量, 能够自动识别不同密度的聚类并有效剔除噪声点, 适合处理电池运行数据中复杂多变的一致性特征。同时, 通过 ROC 曲线与 Youden 指数自动确定最优阈值, 避免了人工经验的主观性, 使得模型具备更强的自适应能力和工程可落地性。

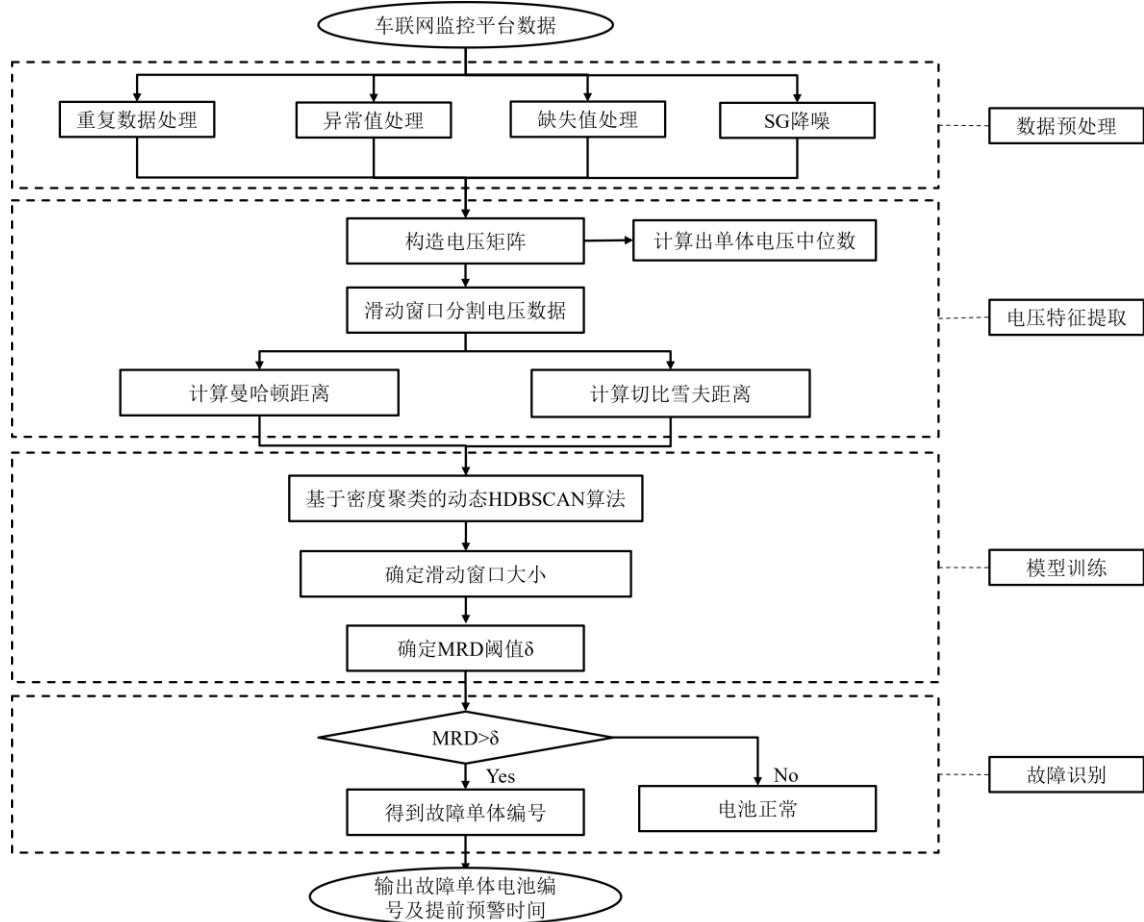


图 1 基于 HDBSCAN 的动力电池单体电压不一致性故障识别方法流程图

Fig.1 Flowchart of the method for identifying inconsistent cell voltage faults in power batteries based on HDBSCAN

3 模型参数训练

完成基于动态 HDBSCAN 的故障诊断模型设计后, 合理选择算法关键参数对结果的稳定可靠性有重要影响。其中, 滑动窗口大小 v , 决定着电压特征的提取时间尺度。过小的窗口, 特征易受到瞬时波动的影

响, 未能平稳反映单体电压变化规律; 反之, 过大的窗口则容易引入冗余信息, 导致本应被局部异常掩盖的特征被平滑掩盖, 即影响异常识别的敏感度。MRD 阈值用于划分正常单体与异常单体, MRD 阈值过大可能会导致异常漏检; MRD 阈值过小, 会导致误报, 严重影响到聚类结果的判断准确度。因此, 本节对窗口长度 v 及 MRD 阈值分别进行了训练和调优, 通过多组对比试验进行统计, 从检测可分性与计算效率中寻求平衡点, 选取比较稳健的参数组合。为兼顾计算效率与样本覆盖性, 本文采用“随机选取”的规则, 从车企监测平台随机下载 50 辆某车型行驶里程超过 10 000 km 的无“单体一致性差报警”的实车运行数据和 12 辆某车型行驶里程超过 10 000 km 的“单体一致性差报警”的实车运行数据, 作为模型训练的数据, 有效支撑模型训练与性能验证。

3.1 滑动窗口大小 v 的确定

滑动窗口是对时序电压数据进行分段特征提取的基本单元。本研究所用动力电池运行数据的采集频率为 1 帧/20 s, 即每帧数据对应 20 s 的采样间隔。因此, 窗口时间长度为滑动窗口大小与 20 s 之积。窗口 v 越大, 对应的时间跨度越长, 所提取的电压特征反映越长时间尺度下的变化趋势; 窗口 v 越小, 则越贴近短时波动特征。

为系统研究窗口长度对检测性能的影响, 本文从上述选取 50 辆无“单体一致性差报警”车辆的数据中随机选取其中 1 辆作为对象进行研究, 导出了其 2024 年 1 月 1 日至 30 日一个月的数据, 经过滤波处理后剩 18 168 条。在此基础上, 设计了基于滑动时间窗口的对比实验。

具体而言, 首先将窗口大小设置在 20 至 200 帧之间, 并以 5 为窗口间隔依次遍历。在每个候选窗口内, 分别提取异常单体的平均 MRD 值和正常单体的最大 MRD 值, 其差值用于衡量异常与正常单体的区分度。同时, 记录各窗口大小对应的计算时间。

为兼顾故障识别性能与计算效率, 本研究引入综合评分机制, 表达式为

$$S_e = \hat{R}_e \cdot \tilde{T}_e \quad (18)$$

式中, \hat{R}_e 为 R_e (e 个窗口的 MRD 比值, 即异常单体平均 MRD/正常单体最大 MRD) 正向归一化处理, 主要原因是: R_e 反映了异常单体与正常单体的区分度, 比值越大表明异常特征越显著, 障识别性能越好, 故对该值进行正向归一化处理, 即 R_e 越大对应归一化值 \hat{R}_e 越大。 \tilde{T}_e 为 T_e (e 个窗口的计算时间) 反向归一化处理, 主要原因是: 计算时间 T_e 是效率指标, 时间越短意味着计算效率越高, 故对该值进行反向归一化处理, 即 T_e 越小对应归一化值 \tilde{T}_e 越大。 \hat{R}_e 和 \tilde{T}_e 归一化表达式分别为

$$\hat{R}_e = \frac{R_e - \min(R_e)}{\max(R_e) - \min(R_e)} \quad (19a)$$

$$\tilde{T}_e = \frac{\max(T_e) - T_e}{\max(T_e) - \min(T_e)} \quad (19b)$$

通过公式 (18) 可知, 当 MRD 比值越大、计算时间越短, 综合评分越高。

首先, 对可比较的窗口集合排查无意义项, 及当某一窗口在所有滑动子窗口中均未检测到异常单体, 则无法计算 MRD 比值, 即这类窗口对判别性能的评估没有信息量。窗口大小 20~40 未检测到异常单体, 故优先排除窗口大小 20~40。

从图 2(a)看运行时间变化趋势, 随着滑动窗口 v 的增大, 模型的运行时间逐步减小。从图 2(b)看 MRD 比值变化趋势, 当窗口从 20 增大到 45 时, 系统由“无法检测异常”跃变为能稳定区分异常与正常单体且比值达到峰值 3.033; 随后随着窗口进一步增大, 比值总体呈缓慢回落。这是由于过短的窗口样本量不足, 瞬时噪声掩盖异常信号从而无法产生异常簇; 滑动窗口大小为 45 时既能累积足够信息以凸显异常, 又不至于把异常信号过度平均化; 而过长窗口会将局部异常在时间上平滑, 从而降低检测灵敏度。综合评分整体呈现随窗口增大而波动下降的趋势, 如图 2(c)所示。具体而言, 当窗口大小从 45 逐步增大至 135 的过程中, 评分由 0.684 4 持续下降, 其间虽有局部小幅回升, 但整体下行态势明显, 表明模型性能随窗口扩大出现显著下滑。当窗口超过 135 后, 评分进入一个波动平台期, 未再出现持续的单边趋势, 但整体水平已明显低于前段窗口范围。

因此, 窗口 45 在检测区分度与计算时间之间取得了最优折中, 不仅产生了最大的 MRD 比值, 且其计算时间约为 5.831 6 s, 在可接受的实时性范围内, 从而获得最高综合评分 0.684 4。

综上所述, 在当前以检测区分度为主、兼顾计算效率的权重分配下, 将滑动时间窗口 v 设为 45, 对应时间长度 15 min。

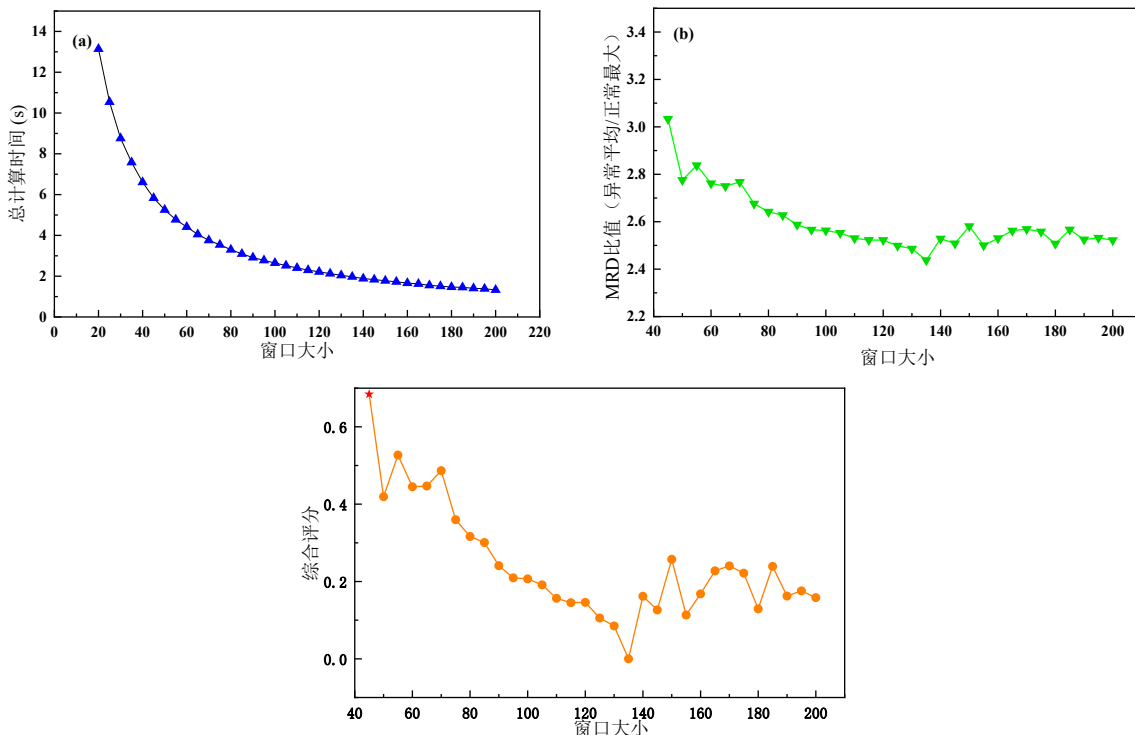


图 2 不同时间窗口下的综合评分. (a)窗口计算时间随窗口大小的变化; (b)异常与正常单体 MRD 差值随窗口大小的变化; (c)综合评分随窗口大小的变化

Fig.2 Comprehensive Scores under Different Time Windows. (a) Variation of Window Calculation Time with Window Size; (b) Variation of MRD Difference between Abnormal and Normal Cells with Window Size; (c) Variation of Comprehensive Score with Window Size

3.2 MRD 阈值的确定

通过设置 MRD 的阈值将异常和正常单体区分开来是进行有效故障识别的关键所在，小于 MRD 阈值的单体会被归为正常单体，反之，则归为异常单体。为了获得科学合理且具有稳健性的阈值，本文将上述随机选取的 50 辆无“单体一致性差报警”的实车运行数据和 12 辆“单体一致性差报警”的实车运行数据通过动态 HDBSCAN 得到各车辆的最大 MRD，统计结果如图 3 所示。

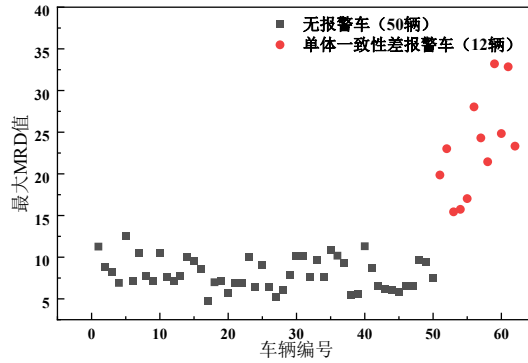


图 3 不同车辆最大 MRD 分布对比图

Fig.3 Comparison of Maximum MRD Distributions Across Different Vehicles

采用 ROC(Receiver Operating Characteristic)曲线结合 Youden 指数的方法确定 MRD 的阈值。本研究以车辆动力电池的最大 MRD 作为区分无报警车与单体一致性差报警车的指标。令车辆编号为 r ($r = 1, 2, \dots, R$)，每辆车的动力电池包含多个单体，单体编号为 i 。对于车辆 r ，其动力电池第 i 个单体的 MRD 值记为 $MRD_{r,i}$ 。则该车辆动力电池的最大 MRD 定义为

$$M_r = \max_i MRD_{r,i} \quad (20)$$

即车辆 r 动力电池内所有单体上的 MRD 最大值。每辆车对应一个类型标签：无报警车或单体一致性差报警车。将类型映射为二值标签：无报警车记为 $y_r = 0$ ，单体一致性差报警车记为 $y_r = 1$ 。定义无报警车集合 $R_N = \{r | y_r = 0\}$ ，单体一致性差报警车集合 $R_F = \{r | y_r = 1\}$ 。则无报警车最大 MRD 可表示为

$$M_N^{\max} = \max_{r \in R_N} M_r \quad (21)$$

单体一致性差报警车最小 MRD 可表示为

$$M_F^{\min} = \min_{r \in R_F} M_r \quad (22)$$

若 $M_N^{\max} < M_F^{\min}$ ，则两类样本的 MRD 分布完全分离，存在可实现完美分类的阈值区间。

为全面评估不同阈值下的分类效果，构建候选阈值集合。首先取所有车辆动力电池 M_r 去重值作为基础阈值；其次为覆盖所有可能影响分类性能的关键点，补充关键边界附近的点： $M_N^{\max} \pm 0.1$ 、 $M_F^{\min} \pm 0.1$ 以及中点 $(M_N^{\max} + M_F^{\min}) / 2$ ，其中偏移量 0.1 是基于数据尺度的经验选择^[38]。合并去重后按升序排列，得到

阈值序列 $\{T_h\}_{h=1}^H$ 。对于每个候选阈值 T 分类规则为：若某车动力电池 $M_r \geq T$ ，则预测为单体一致性差报警车（正类），否则为无报警车（负类）。根据真实标签与预测结果，计算以下指标

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (23)$$

其中，FPR (False Positive Rate, 假正率) 表示实际为正类的样本中被正确预测为正类的比例；TPR (True Positive Rate, 真正率) 表示被错误预测为负类的正类样本数；TP (True Positive) 表示检测正确的单体一致性差报警车辆数量，FN (False Negative) 表示未能检测出的单体一致性差报警车辆数量，FP (False Positive) 表示检测出的无报警车辆数量，TN (True Negative) 表示检测出的无报警车辆数量。ROC 曲线下的面积 (area under curves, AUC) 反映了分类器的区分能力，AUC 值越大，分类性能越好。

Youden 指数为

$$J = TPR - FPR \quad (24)$$

Youden 指数综合反映了分类器的整体性能，取值范围 $[0,1]$ ， $J=1$ 表示完美。

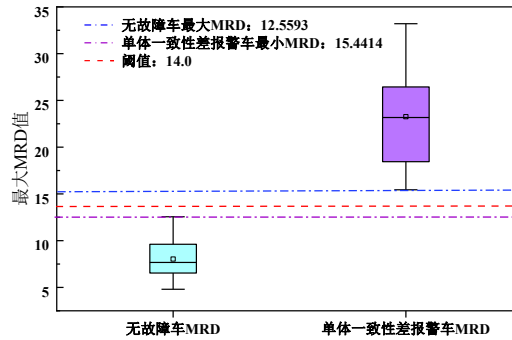


图4 无故障车与单体一致性差报警车最大 MRD 分布对比

Fig.4 Comparison of Maximum MRD Distributions between Fault-Free Vehicles and Vehicles with Poor Cell Consistency Alarms

通过计算各阈值下的 Youden 指数，选择 J 最大值对应的阈值作为最优判别点。首先，根据实验收集的无故障车与单体一致性差报警车的最大 MRD 分布如图 4 所示，进行标签生成，生成 ROC 曲线，对各阈值对应 TPR 与 FPR 进行综合评估并结合 Youden 指数，结果如图 5 和图 6 所示。

结果表明：当阈值低于无故障车辆的最大 MRD 值 12.559 3 时，部分无故障车辆被误判为单体一致性差报警，导致 $FPR > 0$ ，尽管此时所有单体一致性差报警车辆均被识别 ($TPR=1$)，但对应的 Youden 指数小于 1；当阈值在 12.559 3~15.441 4 时， $TPR=1$ 且 $FPR=0$ ，即同时满足 Youden 指数最大值 1 的要求；当阈值在单体一致性差报警车辆的最小 MRD 值 15.441 4 以上时，TPR 开始出现小于 1 的单体一致性差报警车被漏检现象，虽然误判率为 0 ($FPR=0$)，但故障判别能力降低， J 回落。

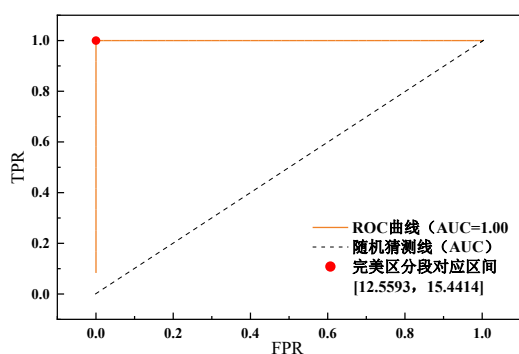


图5 MRD 阈值诊断 ROC 曲线

Fig.5 ROC Curve for MRD Threshold Diagnosis

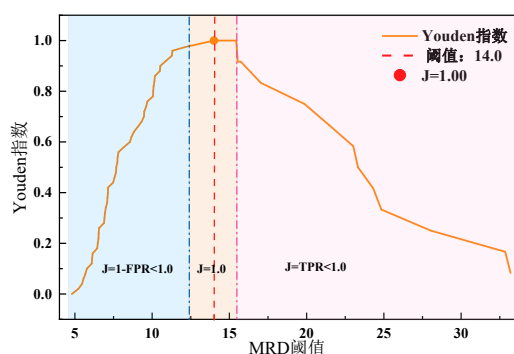


图6 Youden 指数随 MRD 阈值变化曲线

Fig.6 Curve of Youden Index versus MRD Threshold

为了兼顾工程中的阈值不易与边界接近导致对异常数据敏感的问题，本文最终选定了 $\delta = 14.0$ 作为 *MRD* 阈值，该阈值位于无故障车最大 *MRD*(12.559 3)和单体一致性差报警车最小 *MRD*(15.4414)之间，比较稳定且鲁棒性好。相应的判断规则为：车辆最大 *MRD* ≥ 14.0 时，判定为单体一致性差报警车；否则，判定为无故障车。

4 模型应用

4.1 车辆数据概述

本文从发生“单体一致性差报警”的车辆中，随机选取了 5 辆行驶里程超过 10 000 km 的车辆（车辆 1~ 车辆 5），并从正常车辆中随机选取了 1 辆行驶里程超过 10 000 km 的车辆（车辆 6）进行研究。

为了确保分析结果的可靠性，本研究采集了这 6 辆车在过去一个月内的连续运行数据。表 3 详细列出了这 6 辆车的具体信息。考虑到实车运行中电压信号易受电机干扰、传感器误差等噪声影响，本研究先采用 Savitzky-Golay 滤波对这 6 辆车的电压数据进行预处理。滤波后的电压信号如图 7 所示，其中图 7(a)~(f) 分别对应车辆 1~6 的单体电压时序曲线。受日常使用频率差异影响，各车数据量存在差异：车辆 1 为 18 168 条、车辆 2 为 16 107 条、车辆 3 为 30 117 条、车辆 4 为 17 616 条、车辆 5 为 8 556 条、车辆 6 为 23 290 条。

图 7(a)展示了车辆 1 的电池单体电压变化趋势。在后期阶段，46 号电池单体的电压逐渐偏离，在采样点 14 960 时平台告警，进一步表明其与其他单体存在不一致性差异。图 7(b)、7(c)和图 7(d)分别描绘了车辆 2、车辆 3 和车辆 4 的单体电池电压随采样点的变化情况。由图可知，车辆 2 的 71 号单体、车辆 3 的 10 号单体、车辆 4 的 20 号单体自初始阶段起就显现出一定的偏离性，且偏离程度随时间推移逐步加剧，然而平台系统并未在单体偏离初期发出告警，而是分别在采样点 12 410、13 625 和 6 449 处才发出告警信号。该结果印证了现有告警机制存在“延迟性缺陷”，即对初期微小一致性差异的识别灵敏度不足^[39]。图 7(e)显示，车辆 5 的单体电压在后期阶段存在明显异常，尤其是 46 号和 81 号电池单体，其电压逐渐偏离其他单体，并在采样点 4 177 达到平台告警阈值。图 7(f)为车辆 6 电池单体电压变化情况，从其电压曲线可以看出，其在充放电周期各单体都表现出较好的一致性，未出现离群单体。

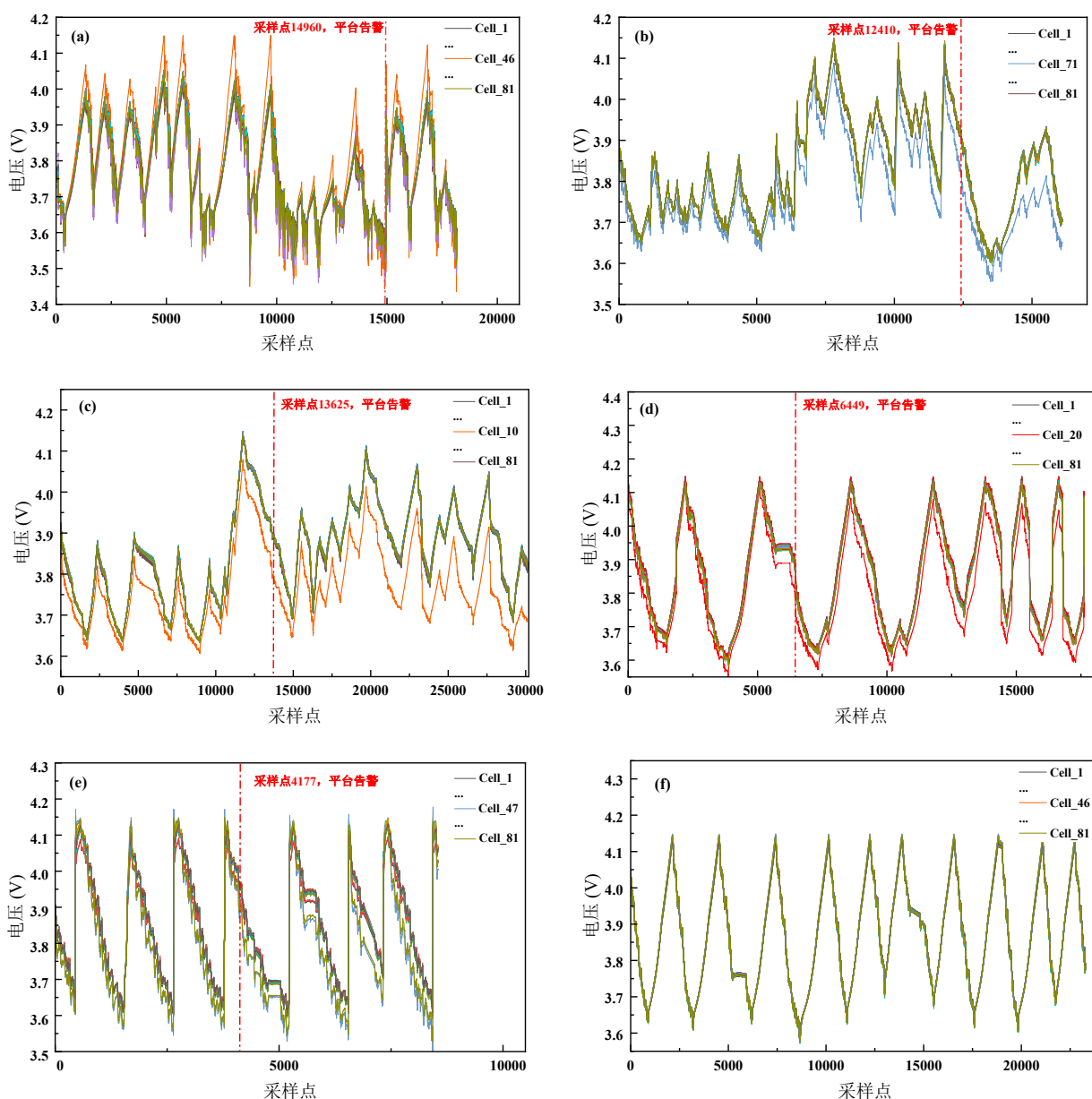


图 7 车辆 1~6 滤波后电压变化曲线。(a)车辆 1; (b)车辆 2; (c)车辆 3; (d)车辆 4; (e)车辆 5; (f)车辆 6

Fig.7 Filtered voltage variation curves for Vehicles 1~6. (a) Vehicle 1; (b) Vehicle 2; (c) Vehicle 3; (d) Vehicle 4; (e) Vehicle 5; (f) Vehicle 6

表 3 车辆相关信息

Tab.3 Vehicle-Related Information

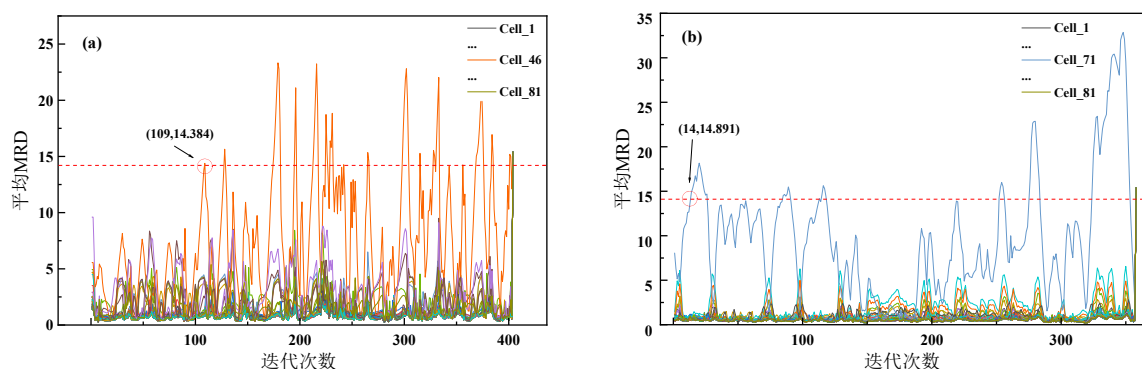
车辆号	电池包结构	能量 /KWh	行驶里程 /km	采样间隔/s	车辆状况	数据选取时间范围
车辆 1	3P81S	31.15	33 995	20	单体一致性差报警	2024-01-01 01:26:35 ~2024-01-30 20:14:21
车辆 2	3P81S	31.15	47 578	20	单体一致性差报警	2023-05-20 03:04:56 ~2024-06-19 18:12:52
车辆 3	3P81S	31.15	44 142	20	单体一致性差报警	2024-07-15 03:04:56 ~2023-08-14 23:55:47
车辆 4	3P81S	31.15	53 317	20	单体一致性差报警	2024-07-01 00:00:09 ~2024-07-30 09:09:15
车辆 5	3P81S	31.15	94 295	20	单体一致性差报警	2024-08-01 02:51:33 ~2024-08-31 23:22:10
车辆 6	3P81S	31.15	16 477	20	无报警	2023-08-01 00:29:15 ~2023-08-31 21:54:33

4.2 故障识别及结果分析

为验证动态 HDBSCAN 模型的有效性,对上述 6 辆车的电池单体进行分析,通过计算每次迭代过程中所有单体的平均 MRD,将其与阈值进行比较,从而判断每个单体是否存在电压不一致性故障。

从图 8(a)的车辆 1 的平均 MRD 值变化曲线可以看到,当单体一致性良好时,所有单体电池的平均 MRD 值均低于阈值,随着一致性越来越差,46 号单体的平均 MRD 逐渐升高,并在第 109 次迭代时超出阈值,对应时间为 2024 年 1 月 18 日 15:44:11,比平台告警提前 9 天 6 小时 10 分钟。其第 109 次迭代的聚类结果图如图 9(a)所示,从图中可以清晰看出 46 号单体明显离群于其他单体,被聚类为异常类别。图 8(b)为车辆 2 的平均 MRD 值变化曲线,结合图 7(b)可以看出,除 71 号单体外,其他单体始终具有良好的一致性。71 号单体的平均 MRD 从初始阶段便展现出一定的偏离性,并在第 14 次迭代时就超出阈值,达到 14.891,对应时间为 2023 年 2 月 22 日 15:22:06,比平台告警提前 19 天 20 小时 42 分钟。图 9(a)为其第 14 次迭代的聚类结果,71 号单体明显离群于其他单体,被聚类为异常类别。图 8(c)和图 8(d)展示了车辆 3 和车辆 4 的平均 MRD 值变化曲线,与车辆 2 类似,除车辆 3 的 10 号单体和车辆 4 的 20 号单体偏离从初始阶段便具有一定的偏离性,其中车辆 3 在第 20 次迭代超出 MRD 阈值,达到 14.201,对应时间为 2024 年 7 月 16 日 8:23:26,比平台告警提前 12 天 22 小时 57 分钟;车辆 4 在第 35 次迭代超出 MRD 阈值,达到 14.730,对应时间为 2024 年 7 月 2 日 20:17:35,比平台告警提前 12 天 11 小时 3 分钟。其他正常单体的平均 MRD 始终远小于 MRD 阈值。结合图 9(c)和图 9(d),车辆 3 的 10 号单体和车辆 4 的 20 号单体也明显离群于正常单体,被聚类为异常类别。图 8(e)为车辆 5 的平均 MRD 值变化曲线,可以看到,车辆 5 存在两个异常单体,分别为 47 号单体和 81 号单体,且在第 11 次迭代首次超出阈值,其中 47 号单体的平均 MRD 达到 24.012,对应的时间为 2024 年 8 月 2 日 22:13:33,比平台告警提前 18 天 22 小时 35 分钟。图 9(e)中,47 号单体和 81 号单体离群于正常单体,且 47 号单体离群更远,与图 8(e)中 47 号单体平均 MRD 值高于 81 号单体形成良好对应。图 8(f)为车辆 6 的平均 MRD 值变化曲线,在完整采样区间内,81 个单体的平均 MRD 均未超出阈值,未检测到异常单体,其聚类结果如图 9(f),未出现异常偏离单体,与实际情况相符。

因此,本模型能够实现正常车辆和单体不一致故障车辆的有效区分,并对电压不一致性故障进行提前识别。



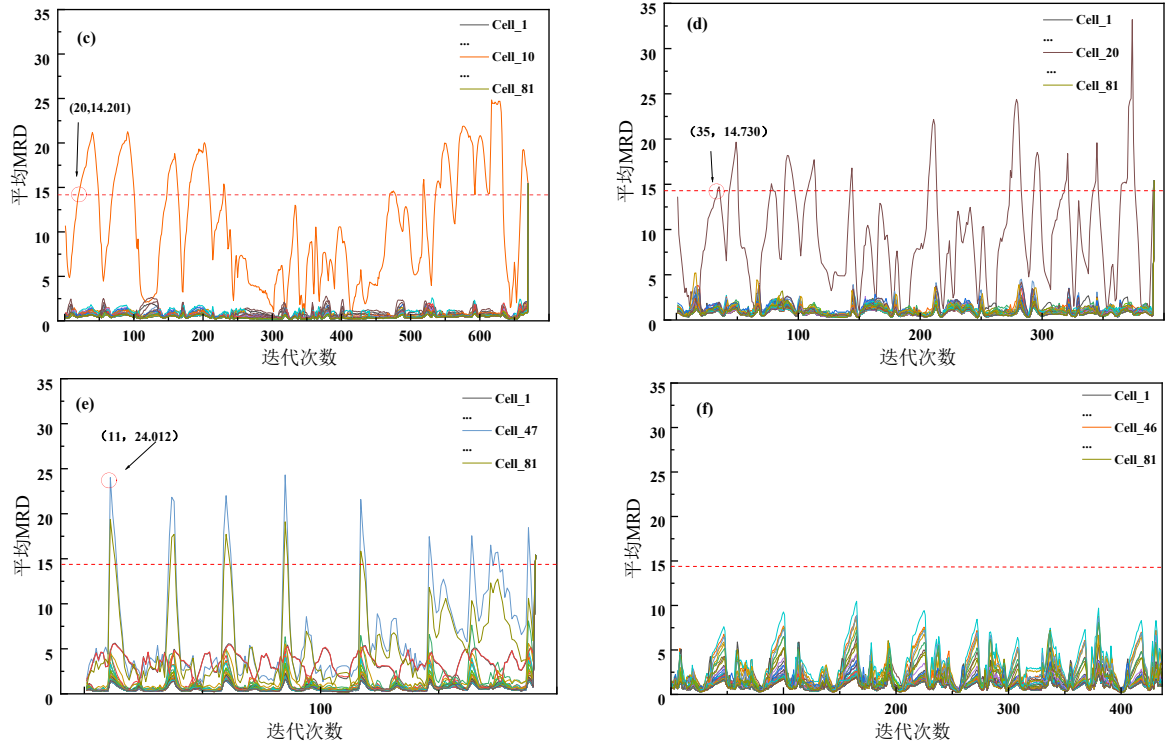
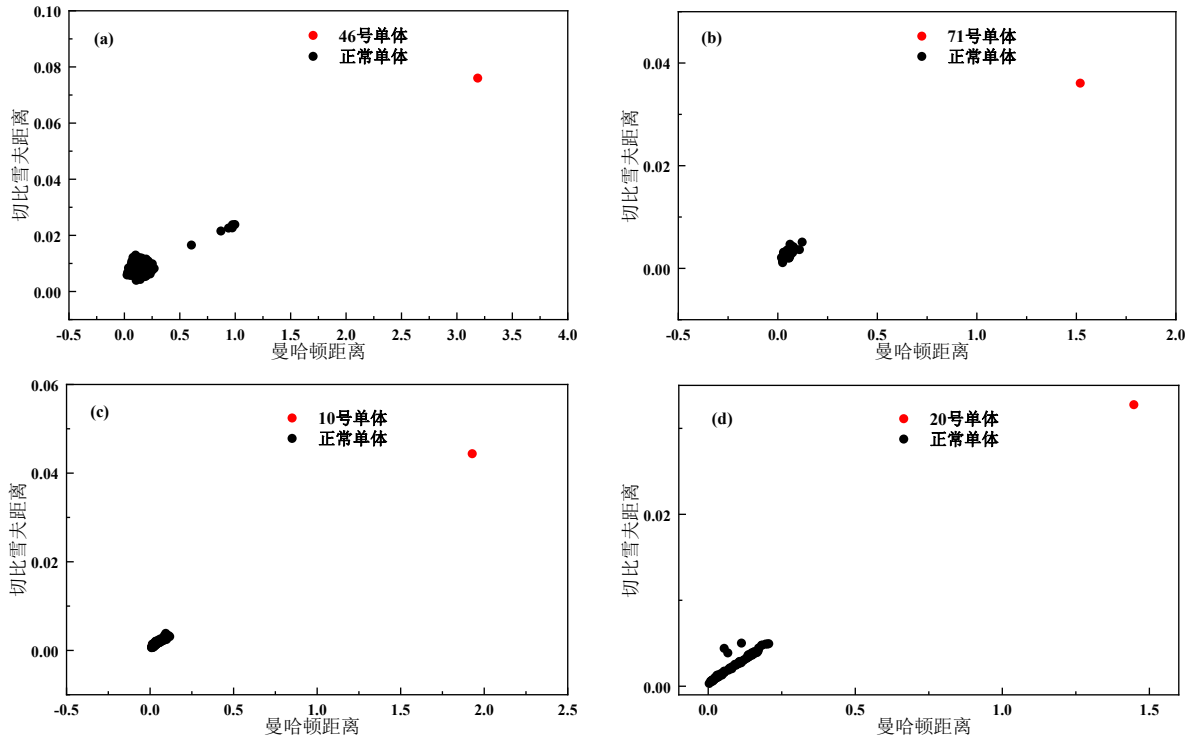


图8 车辆1~6MRD值变化曲线.(a)车辆1; (b)车辆2; (c)车辆3; (d)车辆4; (e)车辆5; (f)车辆6

Fig.8 MRD Value Variation Curves for Vehicles 1-6. (a) Vehicle 1; (b) Vehicle 2; (c) Vehicle 3; (d) Vehicle 4; (e) Vehicle 5;

(f) Vehicle 6



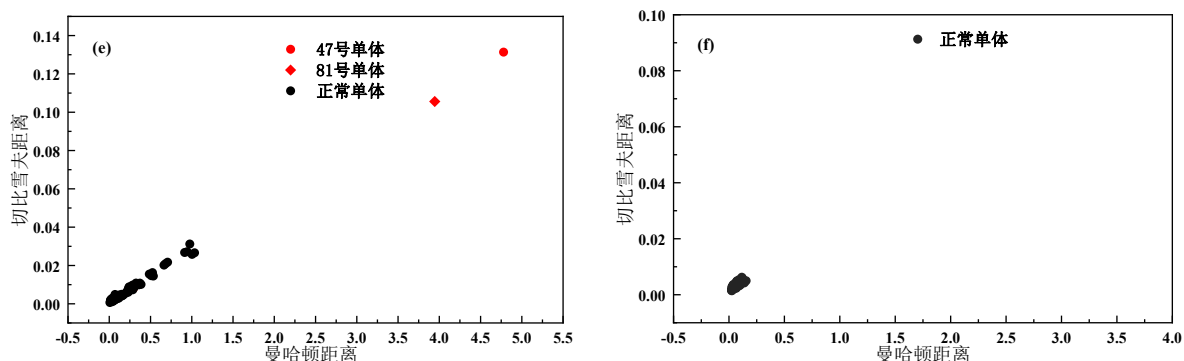


图9 车辆 1-6 动态 HDBSCAN 模型聚类结果。(a)车辆 1; (b)车辆 2; (c)车辆 3; (d)车辆 4; (e)车辆 5; (f)车辆 6

Fig.9 Clustering Results of Dynamic HDBSCAN Model for Vehicles 1-6.(a) Vehicle 1; (b) Vehicle 2; (c) Vehicle 3; (d) Vehicle 4; (e) Vehicle 5; (f) Vehicle 6

4.3 不同模型对比分析

为评估不同动力电池故障识别方法的时效性能差异，本文将所提动态 HDBSCAN 模型，与文献[19]中的 DBSCAN 聚类模型和文献[21]中的 OPTICS 密度聚类模型展开性能对比。在对比算法中，DBSCAN 模型首先基于单体电压计算电压累计偏差和偏差增量，然后使用 DBSCAN 聚类检测电池潜在故障；OPTICS 模型则是计算电压的离散弗雷歇距离和电压偏差值作为故障诊断的特征参数，其次使用 OPTICS 算法对故障特征进行聚类分析，并基于统计确定 rd 阈值 δ^u 为 0.03，实现潜在异常单体的识别。

对比实验同样采用前文提及的 5 台故障车辆数据进行分析，具体对比结果如图 10 和图 11 所示。图 10(a)呈现了 OPTICS 密度聚类模型第 193 次迭代时，车辆 1 首次异常检测结果。从图中可看到，46 号单体显著偏离其他单体，依据模型预设的判定阈值，上述单体被识别为异常单体。图 10(b)为车辆 2 的首次检测结果，对应模型第 76 次迭代过程。图中可见 71 号单体的特征与其他单体存在明显差异，据此被判定为故障单体。图 10(c)记录了车辆 3 的首次异常识别结果，出现于模型第 48 次迭代，图中 10 号单体的偏离显著，按照模型判定规则被归类为故障单体。图 10(d)则为车辆 4 的首次检测结果，对应模型第 32 次迭代阶段，从该图可观察到，20 号单体出现较明显偏离，最终被判定为故障单体。图 10(e)展示了车辆 5 的首次检测结果，对应模型第 32 次迭代阶段，从该图可观察到，20 号单体出现显著偏离，最终被判定为故障单体；但需注意，结合图 7(e)中的电压曲线进一步分析可知，81 号单体也存在一定偏离趋势，却未在本次检测中被判定为异常。

图 11(a)~图 11(e)为车辆 1 至车辆 5 在 DBSCAN 聚类模型中的异常检测效果，车辆 1 在第 237 次迭代时检测出异常，车辆 2 在第 85 次迭代时检测出异常，车辆 3 在第 63 次迭代时检测出异常，车辆 4 在第 121 次迭代时检测出异常，车辆 5 在第 39 次迭代时检测出异常。

结合平台实际数据采集时间，动态 HDBSCAN 模型与 OPTICS 模型、DBSCAN 模型识别异常单体的具体时间差异如表 4 所示。由表中数据可知，本文所提模型对每辆车中电压一致性差的故障单体，均能实现比 OPTICS 模型和 DBSCAN 模型更早的识别。

综上所述，本文提出的动态 HDBSCAN 聚类算法模型，在电池单体电压一致性差故障的识别任务中，相较于 OPTICS 模型和 DBSCAN 模型，具备更快的识别速度与更高的判定精度，更符合实际工程化应用的需求。

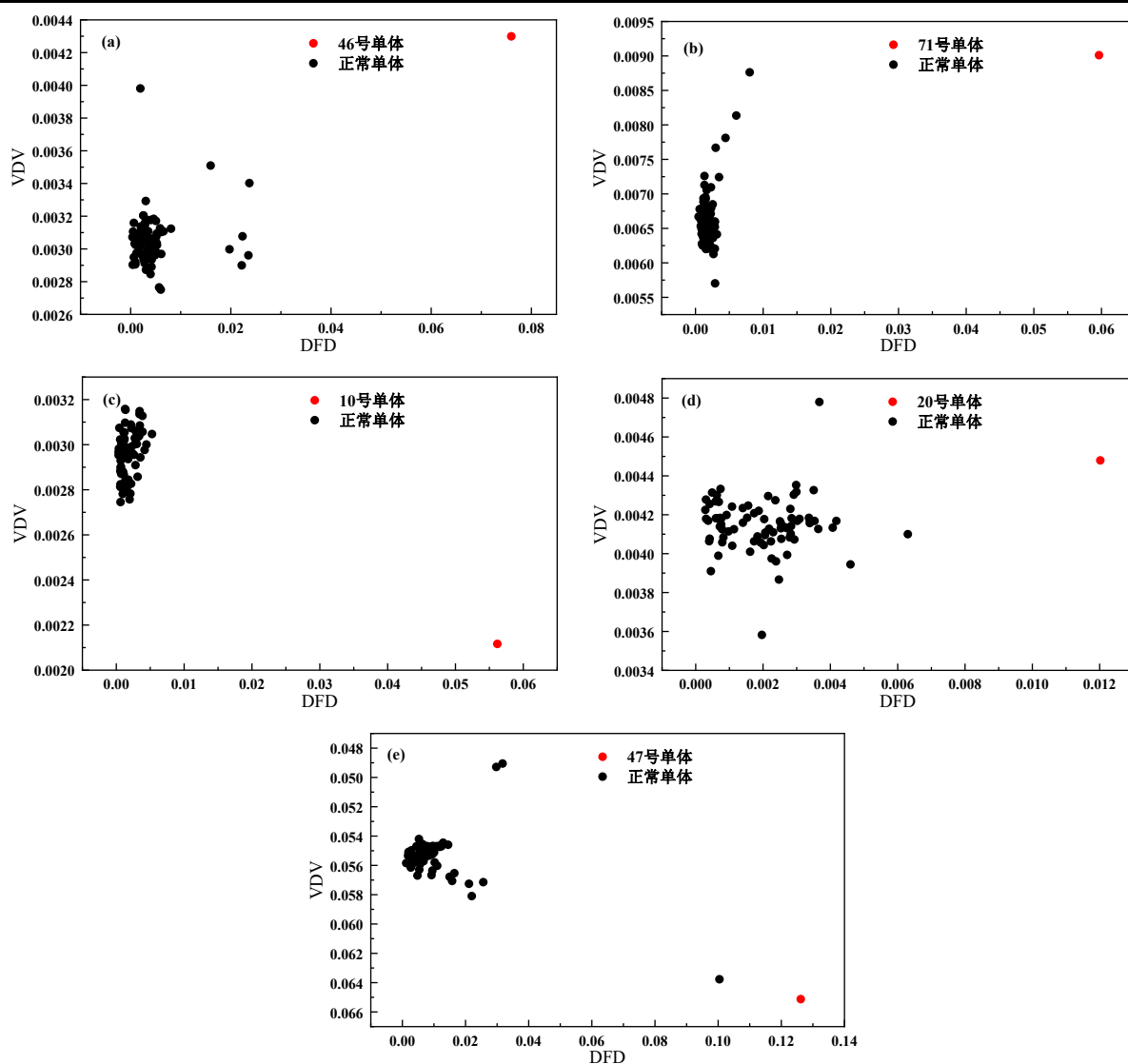
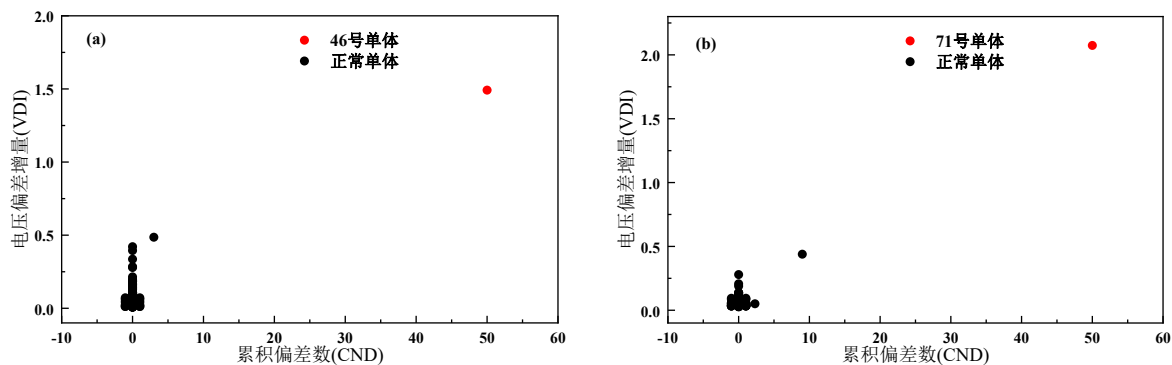


图 10 车辆 1~5 OPTICS 模型聚类结果。(a)车辆 1; (b)车辆 2; (c)车辆 3; (d)车辆 4; (e)车辆 5

Figure 10 Clustering Results of the OPTICS Model for Vehicles 1-5. (a) Vehicle 1; (b) Vehicle 2; (c) Vehicle 3; (d) Vehicle 4; (e) Vehicle 5



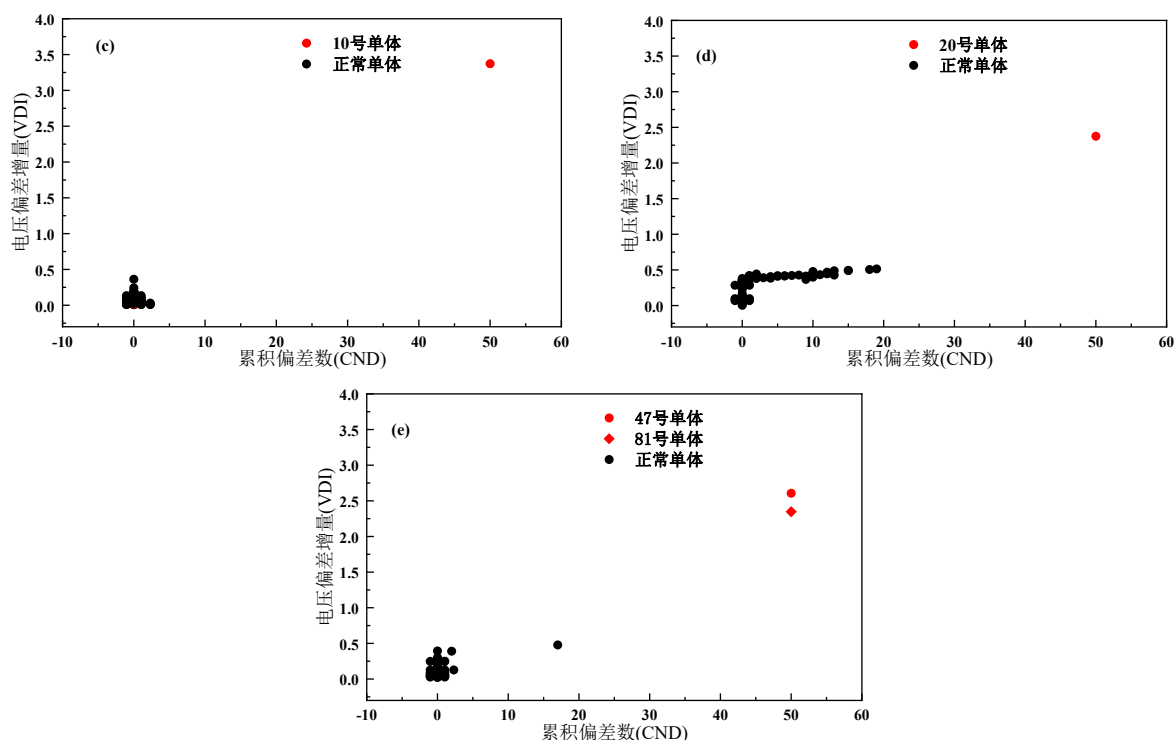


图 11 车辆 1~5 DBSCAN 模型聚类结果。(a)车辆 1; (b)车辆 2; (c)车辆 3; (d)车辆 4; (e)车辆 5

Fig.11 Clustering Results of the DBSCAN Model for Vehicles 1-5. (a) Vehicle 1; (b) Vehicle 2; (c) Vehicle 3; (d) Vehicle 4;

(e) Vehicle 5

表 4 不同型对比情况

Tab.4 Comparison of Different Models

车辆号	第几次迭代识别异常单体			相比于平台提前报警时间		
	本模型	OPTICS	DBSCAN	本模型	OPTICS	DBSCAN
车辆 1	109	193	237	9 天 6 小时 10 分钟	6 天 2 小时 15 分钟	4 天 27 分钟
车辆 2	14	76	85	19 天 20 小时 42 分钟	14 天 19 小时 52 分钟	14 天 4 时 41 分钟
车辆 3	20	48	63	12 天 22 小时 57 分钟	10 天 19 分钟	8 天 15 时 28 分钟
车辆 4	35	97	121	12 天 11 小时 3 分钟	6 天 5 小时 21 分钟	4 天 3 时 43 分钟
车辆 5	11	32	39	18 天 22 小时 35 分钟	10 天 13 小时 40 分钟	9 天 20 时 55 分钟

5 结论

本文针对电动汽车动力电池的单体电压不一致性故障，基于某车企监控平台采集的实际运行数据，提出了一种基于动态 HDBSCAN 的故障识别方法，具体研究结论如下：

1) 提出了一种融合曼哈顿距离和切比雪夫距离的特征参数提取方法，通过构造联合特征矩阵，能够在不同时间尺度下刻画单体电池的一致性，为后续单体电压不一致性检测提供基础。

2) 提出的动态 HDBSCAN 在保留原有 HDBSCAN 聚类框架的基础上，引入了对时间和数据演化过程的建模，使其能够对不断更新的数据集进行增量聚类，显著提升计算效率。在此过程中，算法仍然依赖核心距离与相互可达距离来保证对不同密度分布的适应性，并结合聚类稳定性准则保持聚类结果的鲁棒性，有助于精准识别潜在故障。

3) 通过对模型进行训练, 确定了本模型的滑动窗口 ν 为 45, 并通过 ROC 曲线结合 Youden 指数的方法确定 MRD 的阈值 $\delta=14.0$ 。

4) 通过随机选取 5 辆“单体一致性差”报警的故障车进行实验, 对比发现本文提出的故障识别方法较平台告警分别提前 9 天 6 小时 10 分钟、19 天 20 小时 42 分钟、12 天 22 小时 57 分钟、12 天 11 小时 3 分钟和 18 天 22 小时 35 分钟识别出异常单体, 且相比 OPTICS 模型和 DBSCAN 模型, 本方法可更早识别故障单体。

此外, 本文方法虽然能较早识别出故障单体, 但仅考虑了电压特征, 未考虑容量、内阻及温度等特征参数的不一致性。未来伴随传感器技术的发展, 若能实现内阻与温度的低成本、便捷化、全覆盖在线测量, 相关数据必将被系统纳入模型框架之中, 以进一步增强方法的工程适用性与扩展潜力。

参考文献:

- [1] 程贤福, 马晓冬, 曾建邦, 等. 基于孤立森林算法的动力电池不一致单体识别与预警方法[J]. 华东交通大学学报, 2023, 40(2): 95-102.
CHENG X F, MA X D, ZENG J B, et al. Inconsistent monomer identification and warning method for power batteries based on isolated forests [J]. Journal of East China Jiaotong University, 2023, 40(2): 95-102.
- [2] 戴海峰, 王楠, 魏学哲, 等. 车用动力锂离子电池单体不一致性问题研究综述[J]. 汽车工程, 2014, 36 (02): 181-188+203.
DAI H F, WANG N, WEI X Z, et al. A review on the issue of cell-to-cell variation in automotive power lithium-ion batteries [J]. Automotive Engineering, 2014, 36(02): 181-188+203.
- [3] 孙振宇, 王震坡, 刘鹏, 等. 新能源汽车动力电池系统故障诊断研究综述[J]. 机械工程学报, 2021, 57(14): 87-104.
SUN Z Y, WANG Z P, LIU P, et al. A review of fault diagnosis for new energy vehicle power battery systems [J]. Journal of Mechanical Engineering, 2021, 57(14): 87-104.
- [4] 黄亮亮, 郭拯诗, 何峰, 等. 基于 WPD-GA-BP 的电动汽车动力电池健康状态预测方法 [J/OL]. 华东交通大学学报, 2026, 43 (1): 101-113.
HUANG L L, GUO C S, HE F, et al. A state of health prediction method for electric vehicle power batteries based on WPD-GA-BP [J/OL]. Journal of East China Jiaotong University, 2026, 43 (1): 101-113.
- [5] LIU H X, ZHANG L, LI L W. Progress on the fault diagnosis approach for lithium-ion battery systems: advances, challenges, and prospects [J]. Protection and Control of Modern Power Systems, 2024, 9(5): 16-41.
- [6] ZOU B S, ZHANG L S, XUE X Q, et al. A review on the fault and defect diagnosis of lithium-ion battery for electric vehicles [J]. Energies, 2023, 16(14): 5507.
- [7] 胡晓松, 唐小林. 电动车辆锂离子电池建模方法综述 [J]. 机械工程学报, 2017, 53 (16): 20-31.
HU X S, TANG X L. A review of modeling methods for lithium-ion power batteries in electric vehicles [J]. Journal of Mechanical Engineering, 2017, 53(16): 20-31.
- [8] Fan C X, O'Regan K, Li L Y, et al. Data-driven identification of lithium-ion batteries: A nonlinear equivalent circuit model with diffusion dynamics [J]. Applied Energy, 2022, 321, 1-17.
- [9] WU M Y, WU J, WANG L. Online modeling of the LiFePO₄ power battery based on the data supervisory mechanism [J]. Journal of Energy Storage, 2023, 72 (PB): 108359.
- [10] KANG Y Z, DUAN B, ZHOU Z K, et al. A multi-fault diagnostic method based on an interleaved voltage measurement topology for series connected battery packs [J]. Journal of Power Sources, 2019, 417(31): 132-144.
- [11] XIA B, SHANG Y L, NGUYEN T, et al. A correlation based fault detection method for short circuits in battery packs[J]. Journal of Power Sources, 2017, 337: 1-10.
- [12] SUN Z Y, HAN Y, WANG Z P, et al. Detection of voltage fault in the battery system of electric vehicles using statistical

- analysis [J]. *Applied Energy*, 2022, 307: 118172.
- [13] WANG Z P, HONG J C, LIU P, et al. Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles [J]. *Applied Energy*, 2017, 196: 289-302.
- [14] WANG J, ZHANG S L, HU X. A fault diagnosis method for lithium-ion battery packs using improved RBF neural network [J]. *Frontiers in Energy Research*, 2021, 9: 702139.
- [15] LI D, ZHANG Z S, LIU P, et al. Battery fault diagnosis for electric vehicles based on voltage abnormality by combining the long short-term memory neural network and the equivalent circuit model [J]. *IEEE Transactions on Power Electronics*, 2020, 36(2): 1303-1315.
- [16] 郭煜, 王亦伟, 彭鹏, 等. 基于孤立森林算法的锂离子电池微内短路故障诊断方法[J]. *储能科学与技术*, 2024, 13(11): 4102-4112.
- GUO Y, WANG Y W, PENG P, et al. A fault diagnosis method for micro-short circuit in lithium-ion batteries based on the isolation forest algorithm [J]. *Energy Storage Science and Technology*, 2024, 13(11): 4102-4112.
- [17] 何佳星. 基于数据挖掘的电动汽车动力电池的故障预测研究[D].北方工业大学,2019.
- HE J X. Research on fault prediction of electric vehicle power battery based on data mining [D]. North China University of Technology, 2019.
- [18] 李一帆. 基于数据挖掘的动力电池一致性与健康状态研究[D].福建工程学院,2021.
- LI Y F. Research on power battery consistency and state of health based on data mining [D]. Fujian University of Technology, 2021.
- [19] LI D, ZHANG Z S, LIU P, et al. DBSCAN-based thermal runaway diagnosis of battery systems for electric vehicles [J]. *Energies*, 2019, 12(15): 2977.
- [20] 曾建邦, 张月娅, 张壮, 等. 基于动态 k 值 K-means++聚类的电动汽车动力电池电压不一致故障识别方法[J]. *中国科学: 技术科学*, 2023, 53(01):28-40.
- ZENG J B, ZHANG Y Y, ZHANG Z, et al. A voltage inconsistency fault identification method for electric vehicle power batteries based on dynamic K-value K-means++ Clustering [J]. *Science China: Technological Sciences*, 2023, 53(01): 28-40.
- [21] 单丰武, 汪超仁, 曾建邦, 等. 基于 OPTICS 聚类 and 电压异常指数的动力电池单体不一致性故障程度分析方法[J]. *中国科学: 技术科学*, 2024, 54(08):1611-1624.
- SHAN F W, WANG C R, ZENG J B, et al. An analysis method for the fault severity of power battery cell inconsistency based on optics clustering and voltage abnormality index [J]. *Science China: Technological Sciences*, 2024, 54(08): 1611-1624.
- [22] LONG J W, LIU L P. K*-Means: An efficient clustering algorithm with adaptive decision boundaries [J]. *International Journal of Parallel Programming*, 2024, 53(1).
- [23] HE H L, HE Y X, WANG F, et al. Improved K-means algorithm for clustering non-spherical data [J]. *Expert Systems*, 2022, 39(9).
- [24] LIAN Y H, WANG P, GUO G Y. Analysis model of efficiency and accuracy in big data based on clustering algorithm K-means [C]. *International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC)*, 2023: 433-437.
- [25] ZHANG RT, PENG H, DOU Y T, et al. Automating DBSCAN via deep reinforcement learning [C]. *Proceedings of the 31st acm international conference on information & knowledge management 2022*: 2620-2630.
- [26] ZHANG X L, ZHOU S B. WOA-DBSCAN: Application of whale optimization algorithm in DBSCAN parameter adaption [J]. *IEEE Access*, 2023, 11: 91861-91878.
- [27] MUSTAFA H, LEAL E, GRUENWALD L. An experimental comparison of GPU techniques for DBSCAN clustering [C]. *2019 IEEE International Conference on Big Data (Big Data)*, 2019.
- [28] RAJABI F, MCARTHUR J J. Applying OPTICS with and without PCA for fault detection of fan coil units using building automation system data [J]. *Energy and Buildings*, 2024, 317: 114368.
- [29] MOUTON J P, FERREIRA M, HELBERG A S J. A comparison of clustering algorithms for automatic modulation classification [J]. *Expert Systems with Applications*, 2020, 151: 113317.
- [30] ZHOU H, CHEN X X, MA B Z, et al. Exploring HDBSCAN for effort-aware defect prediction: A density-based clustering method [C]. *International Conference on Frontier Technologies of Information and Computer (ICFTIC)*, 2024: 1583-1586.

- [31] TORRES-CASTRO L, BATES A M, JOHNSON N B, et al. Early detection strategies for lithium-ion battery safety in electric vehicles [J]. ECS Meeting Abstracts, 2024, 2(1): 61–61.
- [32] 刘威,刘伯奇,杨娜.基于随机子空间法和自适应 DBSCAN 的自动模态识别[J].工程力学, 2025, 42:1-12.
LIU W, LIU B Q, YANG N. Automated modal identification based on the stochastic subspace method and adaptive DBSCAN [J]. Engineering Mechanics, 2025, 42: 1-12.
- [33] 宋国兵,常娜娜,侯俊杰,等.基于系统电压分布曲线拟合的后备保护方案[J].电力系统保护与控制,2022,50(16):95-104.
SONG G B, CHANG N N, HOU J J, et al. A backup protection scheme based on system voltage distribution curve fitting [J]. Power System Protection and Control, 2022, 50(16): 95-104.
- [34] YANG Z G, WANG J, HUANG Q J, et al. Anomaly detection algorithm of single variable time-series data based on dynamic parametrization for subsurface fluid data anomaly detection [J]. Geophysical Journal International, 2025, 243(1): 1-12.
- [35] MA Y L, ZHANG C J, XIONG D, et al. Multi-task dynamic graph learning for brain disorder identification with functional MRI [J]. Pattern Recognition, 2025: 111922.
- [36] BROWN R G. Smoothing, forecasting and prediction of discrete time series [M]. Courier Corporation, 2004.
- [37] WHEELWRIGHT S, MAKRIDAKIS S, HYNDMAN R J. Forecasting: methods and applications [M]. John Wiley & Sons, 1998.
- [38] Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated Cutoff Point [J]. Biometrical Journal, 2005, 47(4): 458–472.
- [39] CHEN Z H, ZHAO W F, LI W H, et al. A novel differential voltage segment-based rejection and clustering method for enhancing short and long term consistency in reuse of retired batteries [J]. Sustainable Energy Technologies and Assessments, 2025, 82: 104442.



第一作者: 曾建邦 (1981—), 男, 博士, 副教授/研究员, 研究方向为新能源汽车大数据技术, 动力电池主被动安全技术, 电化学能源 (如锂离子电池、锂空气电池和燃料电池等), 多尺度强化传热传质技术。



通信作者: 单丰武 (1975—), 男, 教授高级工程师, 博士, 硕士生导师, 国家杰出青年科学基金获得者, 享受国务院特殊津贴专家, 江西省双千人才, 研究方向为研究方向为新能源汽车智能网联技术。