中国科学: 技术科学 2026 年 第 56 卷 第 3 期: 697 ~ 708
SCIENTIA SINICA Technologica techcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文 厦门大学建校105周年航空宇航科学与技术学科专刊

基于不完备张量表示学习的锂离子动力电池健康状态预测

吴昊[1], 侯佳雯[1], 包学研[1], 曾念寅[2], 罗辛[1*], 王子栋[3]

1. 西南大学计算机与信息科学学院软件学院, 重庆 400715
2. 厦门大学航空航天学院, 厦门 361005
3. 山东科技大学电气与自动化工程学院, 青岛 266590
* E-mail: luoxin@swu.edu.cn

收稿日期: 2025-11-20; 接受日期: 2026-01-29; 网络版发表日期: 2026-03-09
国家自然科学基金(批准号: 62302402)资助项目

摘要 锂离子动力电池健康状态(state of health, SOH)预测是电池健康管理的核心, 准确的锂电池健康状态预测对确保电池设施的可靠性和安全性具有重要意义. 数据驱动的SOH预测方法因其灵活性高且无需对电池电化学机理深入描绘, 已被广泛应用于锂离子动力电池SOH预测. 然而, 当前数据驱动的SOH预测方法通常建立在完备的电池性能数据(如电流、电压、阻抗)基础上, 但在实际工况中传感器故障等不可避免因素导致获取的电池性能数据完整性、准确性和一致性变差, 对精准预测电池SOH造成严重挑战. 因此, 为在不完备电池性能数据条件下实现精准的SOH预测, 本文提出一种基于不完备张量表示学习的锂离子动力电池SOH预测方法. 首先, 在张量表示学习模块, 将电池性能数据建模为"电池单元×采样时间×监测指标×循环次数"四阶不完备张量, 并构建不完备张量表示学习模型将不完备电池性能数据表示为完备的特征数据; 然后, 在特征提取模块, 通过引入3个因子矩阵与循环次数-监测指标特征张量进行模态乘以提取有效特征; 最后, 在性能数据预测模块, 采用Transformer预测未来$|K|$次循环电池性能数据, 从而完成精准的电池SOH预测. 在广泛使用的两个电池性能数据集上实验结果表明, 所提出方法能基于不完备的历史监测性能数据精准地预测电池在未来循环中的性能数据, 从而实现准确的SOH预测.

关键词 锂离子动力电池, 健康状态预测, 不完备张量, 表示学习

1 引言

随着新能源汽车与储能系统的迅速普及, 锂离子电池由于其能量密度高、自放电率低、循环寿命长等特点逐渐成为应用最为普遍的动力电池类型. 相应地, 锂离子动力电池的安全性和可靠性已成为能源装备领域的核心关注点[1,2]. 电池健康状态(state of health, SOH)是衡量电池性能退化程度的关键指标, 准确预测SOH对于预防电池突发故障、延长电池寿命、优化电池管理策略具有重要意义[3,4]. 在实际工况下, 电池在

引用格式: 吴昊, 侯佳雯, 包学研, 等. 基于不完备张量表示学习的锂离子动力电池健康状态预测. 中国科学: 技术科学, 2026, 56: 697–708
Wu H, Hou J W, Bao X Y, et al. State-of-health prediction of lithium-ion power batteries via incomplete tensor representation learning (in Chinese). Sci Sin Tech, 2026, 56: 697–708, doi: 10.1360/SST-2025-0365

© 2026 《中国科学》杂志社

长期循环中会受到容量衰减、内阻增加和活性物质损耗等多重退化机制影响，其健康状态呈现显著的时变性与非线性特征. 因此，构建高精度的SOH预测方法对推动动力电池智能运维具有重要的工程和科学意义.

目前，国内外锂离子动力电池SOH预测方法主要包括实验法、模型法和数据驱动法[5,6]. 实验法利用试验设备对电池进行测试, 通过相关的测试参数(如电压和电流)，获得标准老化特性参数(如容量和内阻)来估计电池SOH，但其需要严格的试验条件以及复杂的试验设备，测试周期较长；基于模型的方法通过建立电池模型(如等效电路模型)并根据直接测量的数据来识别模型参数，但其通常依赖于精确的物理方程和参数辨识. 随着大数据及人工智能技术的快速发展，数据驱动方法已成为当前SOH预测的主流，其基于监测的性能数据(如电流、电压、阻抗和温度)通过建立学习模型能实现锂离子动力电池SOH预测. 例如, 基于域对抗迁移学习的方法[7]、面向多场景的多级估计方法[8]、融合分层卷积的多通道交互密集Transformer方法[9]、策略特征分析方法[10]和多健康数据融合方法[11]等.

虽然数据驱动的SOH预测方法能从历史数据中学习电池退化规律，但其普遍依赖于完备且高质量的电池性能数据. 然而在实际应用中，电池性能数据常出现采样分辨率不一致、不同循环段缺失、传感器故障导致断点缺失以及不同工况数据采集不充分等问题, 导致获取的数据呈现明显的不完备性[12]. 如果忽略数据缺失模式，直接使用插值或删减会破坏其多维结构特征，从而影响SOH预测的准确性和稳定性[13,14]. 因此，现有数据驱动方法难以充分利用真实运行场景中的不完备性能数据.

针对上述问题，本文提出一种基于不完备张量表示学习的锂离子动力电池SOH预测方法，其构建了基于Tucker2分解的不完备张量表示模型，将不完备电池性能数据表示为适配下游预测任务的完备特征，实现了任务导向的不完备数据表示学习. 具体地，该预测方法包括不完备张量表示学习模块、特征提取模块和性能数据预测模块. 首先, 不完备的电池性能数据被建模为“电池单元×采样时间×监测指标×循环次数”四阶不完备张量$\mathbf{X}$，并基于Tucker2分解原理构建四阶不完备张量表示学习模型以将不完备电池性能数据表示为完备的电池单元特征矩阵$\mathbf{A}$、采样时间特征矩阵$\mathbf{B}$和四阶循环次数-监测指标特征张量$\mathbf{G}$；然后，在特征提取模块, 通过引入3个因子矩阵$\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$分别在3个模态通过张量模态乘运算提取循环次数-监测指标特征张量中的有效特征; 最后, 在性能数据预测模块, 以$|I|$次循环有效特征数据为输入，采用Transformer预测未来$|K|$次循环对应的循环次数-监测指标特征张量，并结合电池单元特征矩阵$\mathbf{A}$和采样时间特征矩阵$\mathbf{B}$获得未来$|K|$次循环电池性能数据张量$\mathbf{Y}$，通过预测的性能数据(如容量和内阻)即实现对电池SOH的精准预测. 因此, 本文贡献如下:

(1) 设计了一个面向电池SOH预测的不完备张量表示学习方法, 其基于Tucker2分解获得了适配下游预测任务的电池性能数据特征;

(2) 提出了一个高精度的锂离子动力电池SOH预测框架，通过完备的电池性能特征数据与Transformer模型实现了电池SOH准确预测.

在广泛使用的锂离子动力电池数据集实验表明, 本文提出的方法能准确预测未来循环中锂离子动力电池的性能数据, 从而实现精准的电池SOH预测.

2 基本概念

2.1 张量Tucker2分解

对于三阶张量$\mathbf{Y}^{|I|\times|J|\times|K|}$，Tucker2分解将$\mathbf{Y}$分解为一个三阶核张量$\mathbf{O}^{|M|\times|N|\times|K|}$和两个因子矩阵$\mathbf{U}^{|I|\times|M|}$和$\mathbf{V}^{|J|\times|N|}$, 图1展示了三阶张量Tucker2分解过程，其计算过程可表示为如下公式[15]:

$$\mathbf{Y} \approx \mathbf{O} \times_1 \mathbf{U} \times_2 \mathbf{V}, \tag{1}$$

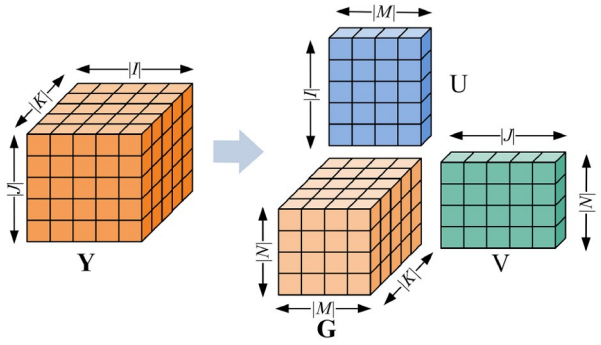


图 1 (网络版彩图)三阶张量Tucker2分解

Figure 1 (Color online) Third-order tensor Tucker2 decomposition.

其中$\times_1$和$\times_2$分别表示张量模态-1乘积和模态-2乘积.

2.2　电池SOH

电池SOH不能通过传感器直接获取，但可以用容量、内阻、功率和剩余循环次数等不同的电池性能参数来定义SOH[16]，由此，公式(2)和(3)展示了使用容量和内阻定义的SOH,

$$SOH_C = \frac{C_i}{C_0} \times 100\%, \tag{2}$$

$$SOH_R = \frac{R_{EOL} - R_i}{R_{EOL} - R_0} \times 100\%, \tag{3}$$

其中SOH_C和SOH_R分别表示以容量和内阻定义的健康状态；C_i和C_0分别表示当前最大可用容量和初始最大可用容量；R_i，R_0和R_{EOL}分别表示当前内阻、初始内阻和寿命截止时内阻.

3　基于不完备张量表示学习的SOH预测

图2展示了提出方法的整体框架图, 其包括不完备张量表示学习模块，特征提取模块和性能数据预测模块, 具体地:

(1) 在不完备张量表示学习模块, 首先历史不完备电池性能数据被建模为“电池单元×采样时间×监测指标×循环次数”四阶不完备张量$\mathbf{X}^{|I|\times|J|\times|S|\times|T|}$，然后仅基于张量已知数据构建不完备张量表示学习模型将电池性能数据表示为完备的电池单元特征矩阵$\mathbf{A}^{|I|\times|M|}$、采样时间特征矩阵$\mathbf{B}^{|J|\times|N|}$和四阶循环次数-监测指标特征张量$\mathbf{G}^{|M|\times|N|\times|S|\times|T|}$.

(2) 在特征提取模块, 将四阶特征张量$\mathbf{G}$沿循环次数模态展开成维度为$|M|\times|N|\times|S|$的三阶特征张量序列$\{\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_{|T|}\}$，然后引入3个因子矩阵$\mathbf{D}^{|M|\times|r1|}$, $\mathbf{E}^{|N|\times|r2|}$, $\mathbf{F}^{|S|\times|r3|}$分别在$\mathbf{G}_t$ ($t = 1, 2, \cdots, T$) 3个模态进行张量模态乘运算以提取每个循环的有效特征，由此得到张量维度为$|r1|\times|r2|\times|r3|$的三阶张量序列$\{\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_{|T|}\}$.

(3) 在性能数据预测模块, 首先将$\mathbf{H}_t$展开成向量并通过变换作为Transformer预测器的输入，然后通过Transformer预测输出三阶张量序列$\{\hat{\mathbf{H}}_{|T|+1}, \hat{\mathbf{H}}_{|T|+2}, \cdots, \hat{\mathbf{H}}_{|T|+|K|}\}$，最后采用因子矩阵$\mathbf{D}$，$\mathbf{E}$和$\mathbf{F}$通过张量模态乘运算获得三阶特征张量序列$\{\hat{\mathbf{G}}_{|T|+1}, \hat{\mathbf{G}}_{|T|+2}, \cdots, \hat{\mathbf{G}}_{|T|+|K|}\}$，最后结合特征矩阵$\mathbf{A}$和$\mathbf{B}$获得未来$|K|$次循环的电池性能数据张量$\mathbf{X}^{|I|\times|J|\times|S|\times|K|}$，通过预测的性能数据(如容量和内阻)即可完成对电池SOH的预测. 下面将详细介绍每个模块.

3.1　不完备张量表示学习

由于传感器故障等因素，监测获取的电池性能数据(包括电压、电流、温度、容量和内阻等)通常是不完备的, 并且这些性能数据间还存在耦合关系. 因张量能有效建模关系数据间的关联关系，保存数据内部的时空结构特性[17]. 因此，为了有效刻画不完备性能数据间的耦合关系，本文采用四阶不完备张量建模这些数据. 由此，面向$|I|$个电池，针对监测的$|S|$个性能指标，每个电池在一次循环中采集$|J|$次监测数据，经过$|T|$次

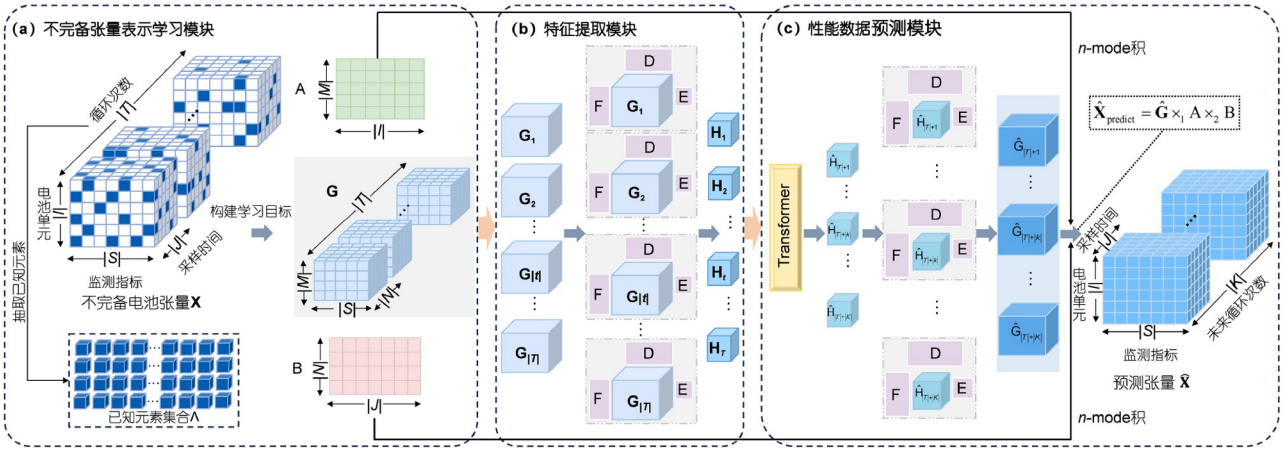


图 2　(网络版彩图)基于不完备张量表示学习的SOH预测

Figure 2　(Color online) State-of-health prediction via incomplete tensor representation learning.

循环后监测的性能数据可建模为“电池单元×采样时间×监测指标×循环次数”四阶不完备张量$\mathbf{X}^{|I|\times|J|\times|S|\times|T|}$. 由于$\mathbf{X}$的不完备特征, 若将其直接作为预测器的输入则需要对缺失的数据进行插补(如均值插值和中值插值), 但这些插补数据在后续模型学习过程中会面临引入系统性偏差、产生伪特征、破坏原数据的结构相关性等问题[18,19]. 数据密度导向建模规则是一种广泛用于构建不完备数据学习模型的模型构建范式[20], 面向不完备的电池性能数据张量$\mathbf{X}$, 数据密度导向建模主要体现为构建的表示学习模型仅定义在不完备张量的已知数据集合, 并且模型参数在学习过程也是仅以张量中的已知数据作为训练样本, 由此极大提高了模型的泛化性与可扩展性. 因此, 本文仅基于张量中的已知数据构建面向电池性能数据不完备张量的表示学习模型, 通过将不完备性能数据表示为完备特征数据, 从而使其能应用于下游任务.

相比于广泛使用的CP分解, Tucker2分解具备对张量不同模态重要性的区别处理能力, 能够有选择地、独立地对不同的模态进行处理. 因此, 面向电池SOH预测任务, 为了获得所有电池单元在每个循环中所有监测指标的数据特征, 本文采用Tucker2分解构建表示学习模型, 通过使用广泛采用的欧式距离和L_2范数正则, 表示学习目标函数如下:

$$\begin{aligned}\varepsilon &= \sum_{x_{ijst}\in\Lambda}\left(\left\|\mathbf{X}-\mathbf{G}\times_1\mathbf{A}\times_2\mathbf{B}\right\|_F^2+\lambda\left(\|\mathbf{G}\|_F^2+\|\mathbf{A}\|_F^2+\|\mathbf{B}\|_F^2\right)\right)\\ &= \sum_{x_{ijst}\in\Lambda}\left(\left(x_{ijst}-\sum_{m=1}^{|M|}\sum_{n=1}^{|N|}g_{mnst}a_{im}b_{jn}\right)^2\right.\\ &\quad\left.+\lambda\left(\sum_{m=1}^{|M|}\sum_{n=1}^{|N|}g_{mnst}^2+\sum_{m=1}^{|M|}a_{im}^2+\sum_{n=1}^{|N|}b_{jn}^2\right)\right).\end{aligned}\tag{4}$$

其中Λ表示张量$\mathbf{X}$中的已知数据集合; g_{mnst}, a_{im}和b_{jn}分别代表四阶循环次数-监测指标特征张量$\mathbf{G}^{|M|\times|N|\times|S|\times|T|}$、电池单元特征矩阵$\mathbf{A}^{|I|\times|M|}$和采样时间特征矩阵$\mathbf{B}^{|J|\times|N|}$中的单个元素; λ表示正则化系数. 通过使用Adam算法求解目标函数可获得$\mathbf{G}$, $\mathbf{A}$和$\mathbf{B}$.

3.2 循环次数-监测指标特征提取

通过不完备张量表示学习获得了完备的四阶循环次数-监测指标特征张量$\mathbf{G}^{|M|\times|N|\times|S|\times|T|}$, 其表征了采样的所有电池单元在不同历史循环的性能数据特征. 由于性能数据间的耦合关系, 学习的特征中通常存在冗余信息. 因此, 为减少冗余特征对预测任务造成的扰动, 需要提取特征张量$\mathbf{G}$中的有效信息.

因此, 针对四阶循环次数-监测指标特征张量$\mathbf{G}$, 为实现细粒度的特征提取, 首先将$\mathbf{G}$沿循环次数模态展开成$|T|$个维度为$|M|\times|N|\times|S|$的三阶特征张量$\mathbf{G}_t$ (t = 1, 2, ⋯, $|T|$), 由此可获得表征每个历史循环中性能数据特征的三阶张量序列$\{\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_{|T|}\}$; 然后3个因子矩阵$\mathbf{D}^{|M|\times|r1|}$, $\mathbf{E}^{|N|\times|r2|}$, $\mathbf{F}^{|S|\times|r3|}$被引入分别提取$\mathbf{G}_t$中的有效特征, 提取过程如下所示:

$$\mathbf{H}_t=\mathbf{G}_t\times_1\mathbf{D}\times_2\mathbf{E}\times_3\mathbf{F}.\tag{5}$$

由此, 可获得维度为$|r1|\times|r2|\times|r3|$的三阶有效特征张量序列$\{\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_{|T|}\}$, 其中$|r1| < |M|$, $|r2| < |N|$, $|r3| < |S|$.

为进一步融合有效特征并降低预测器的计算花费, 一个编码因子矩阵$\mathbf{P}^{|d|\times|r1r2r3|}$被用于将三阶有效特征张量$\mathbf{H}_t$ (t = 1, 2, ⋯, $|T|$)映射到低维空间, 映射过程如下所示:

$$\mathbf{z}_t=\mathbf{P}\mathrm{vec}(\mathbf{H}_t),\tag{6}$$

其中vec(·)表示将$\mathbf{H}_t$展开成模为$|r1r2r3|$的向量. 通过式(6)可获得预测器的输入向量序列$\{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{|T|}\}$.

3.3 性能数据预测

电池容量衰减和性能演化具有明显的长时序依赖特性, Transformer通过自注意力机制可以在一个时间步同时关注整个历史序列, 将远距离循环之间的相关性直接建模, 且自注意力可以根据任务需求自适应地分配权重, 挖掘不同循环、不同特征间的关键关联. 为实现预测未来$|K|$次循环的性能数据, 首先将有效特征张量序列的低维编码向量序列$\{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{|T|}\}$作为Transformer预测器的输入, 预测输出长度为$|K|$的向量序列$\{\hat{\mathbf{z}}_{|T|+1}, \hat{\mathbf{z}}_{|T|+2}, \cdots, \hat{\mathbf{z}}_{|T|+|k|}\}$. 通过引入一个解码因子矩阵$\mathbf{Q}^{|r1r2r3|\times|d|}$将预测输出向量序列映射为预测特征张量$\{\hat{\mathbf{H}}_{|T|+1}, \hat{\mathbf{H}}_{|T|+2}, \cdots, \hat{\mathbf{H}}_{|T|+|K|}\}$, 解码映射过程表示为

$$\hat{\mathbf{H}}_k=\mathrm{fold}\left(\mathbf{Q}\hat{\mathbf{z}}_k\right),\tag{7}$$

其中fold(·)表示将模为$|r1r2r3|$的向量折叠成维度为$|r1|\times|r2|\times|r3|$的预测特征张量.

进一步, 为获得在未来$|K|$次循环的循环次数-监测指标特征张量$\hat{\mathbf{G}}^{|M|\times|N|\times|S|\times|K|}$, 先将$\hat{\mathbf{H}}_k$ (k = $|T|$+1, $|T|$+2, ⋯,

$|T|+|K|$)与因子矩阵$\mathbf{D}$, $\mathbf{E}$和$\mathbf{F}$进行模态乘运算, 获得$|K|$个维度为$|M|\times|N|\times|S|$的三阶预测特征张量$\hat{\mathbf{G}}_k$ ($k = |T|+1, |T|+2, \cdots, |T|+|K|$), 再从循环次数模态将序列$\{\hat{\mathbf{G}}_{|T|+1}, \hat{\mathbf{G}}_{|T|+2}, \cdots, \hat{\mathbf{G}}_{|T|+|K|}\}$折叠可获得预测的循环次数-监测指标特征张量$\hat{\mathbf{G}}$. 最后, 通过将$\hat{\mathbf{G}}$与电池单元特征矩阵$\mathbf{A}$和采样时间特征矩阵$\mathbf{B}$模态乘, 可获得未来$|K|$次循环的性能数据张量$\mathbf{X}^{|I|\times|J|\times|S|\times|K|}$, 其计算过程表示为如下公式:

$$\hat{\mathbf{X}} = \hat{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B}. \tag{8}$$

相应地, 预测的第i个电池在未来第k次循环的第j次采样中监测的第s个性能指标值为

$$\hat{x}_{ijsk} = \sum_{m=1}^{|M|}\sum_{n=1}^{|N|} \hat{g}_{mnsk} a_{im} b_{jn}. \tag{9}$$

4 实验结果分析

4.1 数据集

本研究在广泛使用的丰田电池数据集[1)]和华中科技大学电池退化数据集[2)]进行实验. 本文选取丰田数据中的124个电池单元、每次循环200次采样点、7个电池性能监测指标、100次充放电循环, 因此可构成四阶张量$\mathbf{X}^{124\times200\times7\times100}$. 华中科技大学电池退化数据集包含77个商用LFP/石墨锂离子电芯, 在恒温30°C条件下采用相同的充电协议, 本文选取了其中循环寿命相似的25个电池单元、每次循环100次采样, 3个电池性能监测指标, 1500次充放电循环, 因此可构成四阶张量$\mathbf{X}^{25\times100\times3\times1500}$. 为了模拟实际测量采样过程中数据丢失的情况, 通过随机掩去40%, 30%, 20%, 10%的已知性能数据, 在两个数据集上分别构建4个测试数据集D1.1~1.4和D2.1~2.4, 来评估所提出模型的性能. 实验统一采用时间维度(充放电循环数)进行切分: 训练集占70%, 验证集占10%, 测试集占20%, 其中验证集用来确认最优参数组合.

4.2 实验设置

在实验中, 所有数据采用监测指标最大最小归一化到[0, 1]区间. 在两个数据集上的张量分解秩分别设置为$M = 24, 12$和$N = 16, 50$. 核张量序列维度对应时间长度$|T|$. Transformer的隐层维度设为512, 注意力头数为8, 编码器与解码器层数均为2, 前馈层宽度2048. 历史窗口长度设为$L = 25$, 预测步长为$H = 5$, 步进间隔为1. 训练批大小为64. 优化器采用Adam, 学习率1×10^{-3}训练总轮数100轮所有实验在单张NVIDIA GeForce RTX 5090 Laptop GPU上进行.

为保证与公开实现及对比方法的公平性, 所有模型在相同的训练、验证、测试划分、相同的滑动窗口长度与预测步长下进行训练, 归一化统计量与缺失掩码均保持一致.

4.3 评价指标

为了衡量多步预测整体效果, 本文采用均方误差(mean squared error, MSE)和平均绝对误差(mean absolute error, MAE)两个指标作为主要的评价指标. 给定的观测索引集合Ω, 记$\hat{x}_{ijts}$为模型在原始物理空间中的预测值, x_{ijts}为原始电池数据的观测值, 则MSE与MAE分别定义为

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{|\Omega|}\sum_{(i,j,t,s)\in\Lambda}\left(x_{ijsk}-\hat{x}_{ijsk}\right)^2, \\ \mathrm{MAE} &= \frac{1}{|\Omega|}\sum_{(i,j,t,s)\in\Lambda}\left|x_{ijsk}-\hat{x}_{ijsk}\right|. \end{aligned} \tag{10}$$

4.4 对比方法

为全面评估所提出模型的有效性,本文选取了近年来代表性的多变量时间序列预测模型作为基线,包括MSD-Mixer(M2)[21], GLAFF+iTransformer(M3)[22], FilterNet(M4)[23], TimeSter(M5)[24], TQNet(M6)[25], CycleNet(M7)[26]以及两个张量分解模型结合预测器的两阶段对比方法SGCP+Transformer(M8)[27]和GSNTD+Transformer(M9)[28]. 其中, M2~M7覆盖从简单线性结构到复杂深度网络的不同层次, 主要面向完备或近似完备序列的直接预测, M8和M9则首先对不完备四阶电池张量执行张量分解以恢复隐特征结构, 再将补全后的数据输入Transformer进行预测, 代表先补全后预测的典型方法. 各模型均在本文构造的四阶电池张量上进行训练与测试, 并在输入输出维度与滑动窗口设置上与本文方法保持一致. 需要注意的是, 本文将不

1) https://data.matr.io/1/projects/5c48dd2bc625d700019f3204
2) https://data.mendeley.com/datasets/nsc7hnsg4s/2

完备张量的缺失值用均值填充来作为对比模型输入.

4.5 预测性能比较实验结果

本文提出的模型M1和对比模型(M2~M9)分别在不同缺失率的电池数据集上进行了对比试验, 性能数据预测精度结果展示在图3. 从这些结果可以看到M1在四个数据集上整体均优于对比模型M2~M9. 例如在D1.1, M1的MSE为0.0724, 明显低于所有基线中最优的0.0813, 误差相对降低约11%; 对应的MAE由最优基线的0.0147降至0.0094, 降幅约36%. 当数据密度提高到80%时, M1的MSE进一步降至0.0442, 而最优基线为0.0593, 相对降低约25%; MAE则由0.0118降到0.0075, 降幅达到36%.

在D2, 同样可以观察到一致趋势. 在D2.1, M1的MSE为0.0996, 优于最优基线0.116, 相对降低约14.5%; MAE相比最优基线0.0199降至0.0182, 降幅约8.5%. 当数据密度提升至70%时, M1的MSE为0.0840, 相比最优基线0.1025相对降低约18.0%; MAE为0.0164, 相比最优基线0.0168仍有约2.4%的改进. 在D2.3和D2.4, M1的MSE分别为0.0742与0.0635, 相比最优基线仍保持约2.6%~4.5%的降低; MAE则分别相对最优基线降低约8.8%与9.8%.

综合上述结果可以看出, 通过将不完备的电池性能数据建模为"电池单元×采样时间×监测指标×循环次数"张量, 并执行表示学习能有效获取数据特征, 基于学习的有效特征执行性能数据预测能避免数据插补导致的预测性能损失.

4.6 消融实验

为验证表示学习模块与预测器对模型性能的影响, 本文设计两个消融实验验证模型的性能.

4.6.1 不完备张量表示学习影响

当不采用张量表示学习时, 循环次数-监测指标特征张量**G**, 电池单元特征矩阵**A**和采样时间特征矩阵**B**采用随机初始化方式设置. 图4给出了完整模型与无表示学习在四个数据集上的MSE与MAE对比结果. 可以看到, 在两个图中完整模型对应的曲线始终位于无表示学习曲线之下, 说明引入表示学习后, 模型的预测误差都得到了一致的降低. 随着数据缺失率的降低, 两种方法的MSE和MAE都呈下降趋势, 但完整模型的下降幅度更为明显. 上述结果表明, 通过对不完备张量

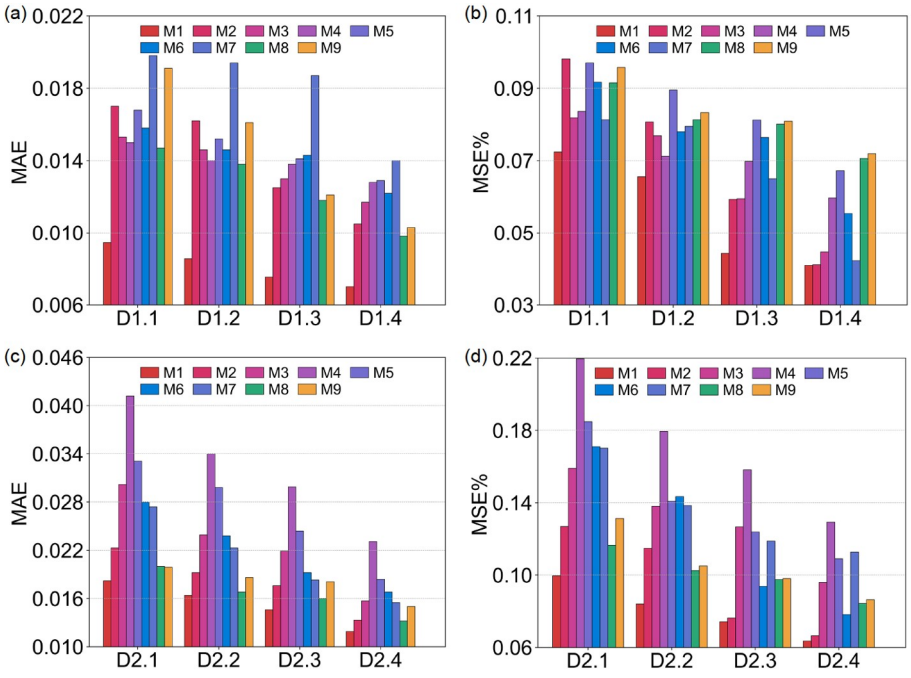


图 3 (网络版彩图)不同模型在D1.1~D2.4上的MAE和MSE对比结果

Figure 3 (Color online) Comparison of MAE and MSE results for different models on datasets D1.1–D2.4.

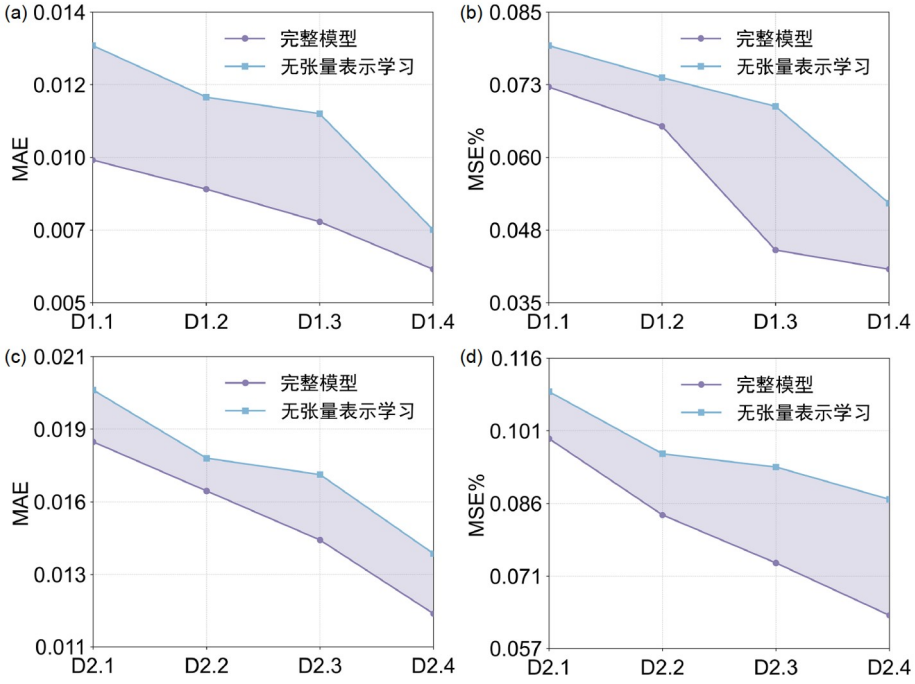


图 4　(网络版彩图)完整模型与无表示学习在数据集D1.1~D2.4上的MSE与MAE对比结果

Figure 4　(Color online) Comparison of MSE and MAE between the complete model and without representation learning on datasets D1.1–D2.4.

表示学习，可以更充分地挖掘“电池单元×采样时间×监测指标×循环次数”中的有效信息，从而在不同数据集和不同观测率下均显著提升多步预测的精度.

4.6.2　不同预测器性能比较

通过将Transformer预测器替换为RNN，GRU，LSTM和ARIMA预测器，测试不同预测器下模型的预测性能. 图5展示了在4个数据集上，将Transformer替换为RNN，GRU，LSTM与ARIMA后，模型MAE与MSE指标上的性能变化. 结果表明Transformer预测器在所有数据集上的表现均最优，并且相较于其他预测器，其优势在不同缺失率、不同电池组之间均保持一致. 具体地，使用RNN时，在4个数据集上均出现较为明显的性能退化，MSE与MAE均有显著上升，说明纯循环结构难以同时捕捉长程依赖和多通道间的复杂耦合. GRU与LSTM的误差整体上介于RNN与Transformer之间，这表明门控循环结构在中短期预测上仍具有一定竞争力，但在面对高阶张量和长时间窗口时，其预测能力仍不及自注意力结构. ARIMA在所有数据集的表现均显著劣于深度学习模型，尤其是在多通道场景下MSE和MAE均远高于其他方法，且对缺失较为敏感，说明线性自回归方法难以捕捉多维电池数据中复杂的时序依赖性.

4.6.3　秩的敏感性分析

为确定张量表示学习模块的秩参数配置，本节在D1与D2数据集上分别进行了多组秩组合的对比实验，并综合考虑预测性能与模型复杂度之间的权衡. 为避免在监测指标维度引入额外的信息损失，本文在两个数据集上均令$r3$等于S，从而仅在其余两个模态上进行低秩压缩. 需要说明的是，本文选取D1.2，D1.3，以及D2.2，D2.3作为对比场景，是因为它们分别对应20%与30%的缺失率设置，该缺失率在电池监测数据的实际采集过程中较为典型同时也与现有电池健康状态预测研究中常用的缺失率保持一致.

在D1上分别测试了($r1$, $r2$, $r3$)的多种组合，结果如表1所示. 可以看出，随着秩增大，预测误差总体下降，但模型参数量同步增长. 综合误差与复杂度，在参数规模可接受的前提下取得了较优的误差水平，因此本文在D1的后续实验中统一采用($r1$, $r2$, $r3$) = (8, 12, 7). 在D2上同样进行了多组秩组合对比，结果见表2. 结果表明，($r1$, $r2$, $r3$) = (9, 25, 3)在D2.2与D2.3上均获得更

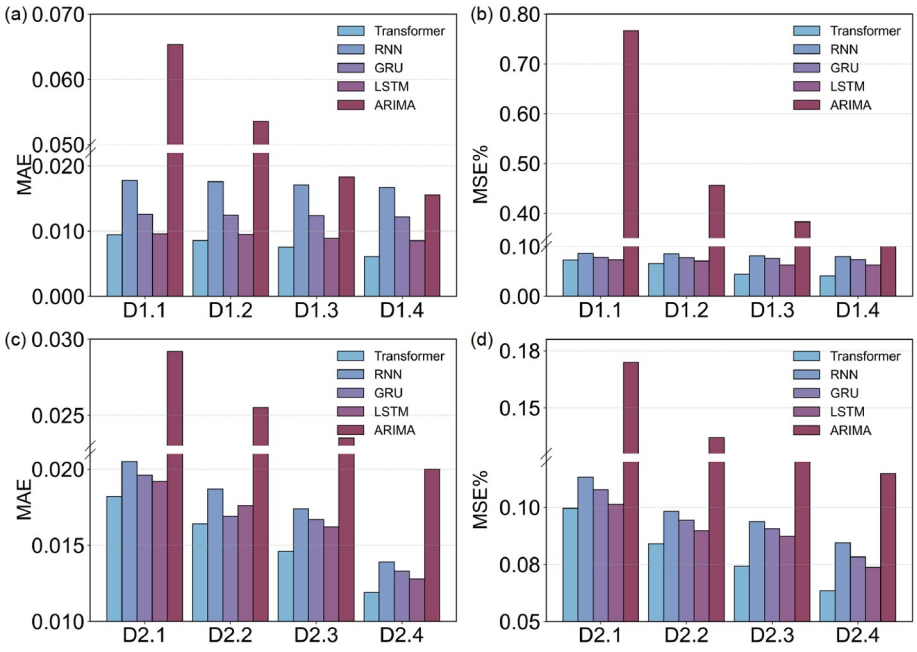


图 5 (网络版彩图)不同预测器预测精度

Figure 5 (Color online) Prediction accuracy with different predictors.

表 1 D1 数据集不同秩组合的预测误差与参数规模对比

Table 1 Comparison of prediction errors and parameter sizes under different rank settings on D1

($r1$, $r2$, $r3$)	参数量	D1.2		D1.3	
		MSE (%)	MAE	MSE (%)	MAE
(4, 6, 3)	7709229	0.2058	0.0321	0.1906	0.0297
(8, 12, 7)	8324465	0.0656	0.0085	0.0442	0.0075
(12, 18, 7)	9185673	0.0480	0.0074	0.0436	0.0072

表 2 D2数据集不同秩组合的预测误差与参数规模对比

Table 2 Comparison of prediction errors and parameter sizes under different rank settings on D2

($r1$, $r2$, $r3$)	参数量	D2.2		D2.3	
		MSE (%)	MAE	MSE (%)	MAE
(6, 12, 1)	131784832	0.2411	0.0430	0.2331	0.0419
(9, 25, 3)	133021110	0.0840	0.0164	0.0742	0.0146
(12, 36, 3)	134294164	0.0882	0.0171	0.0849	0.0165

低的预测误差，与更高秩设置相比，其误差优势或相当，同时参数量增幅较小，因此本文在D2的实验中采用($r1$, $r2$, $r3$) = (9, 25, 3)作为默认配置.

4.7 SOH预测对比

为更直观地展示各模型在SOH预测任务中的动态跟踪能力, 图6与7展示了在D1与D2上的SOH预测曲线对比结果. 从图6可以观察到, 真实SOH曲线在该区间内整体变化较缓, 各方法的差异主要体现在对局部波动的拟合偏差与稳定性上. 其中, M1的预测曲线与真实曲线整体贴合度更高, 能够较稳定地跟踪SOH的细微变化. 相比之下, 部分直接预测基线(M2~M7)出现

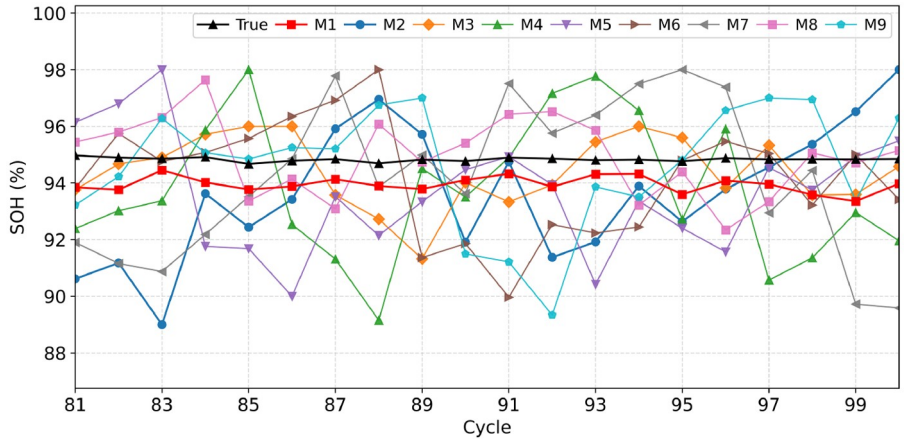


图 6　(网络版彩图) D1数据集上的SOH预测曲线对比
Figure 6　(Color online) SOH prediction curves of different models on D1.

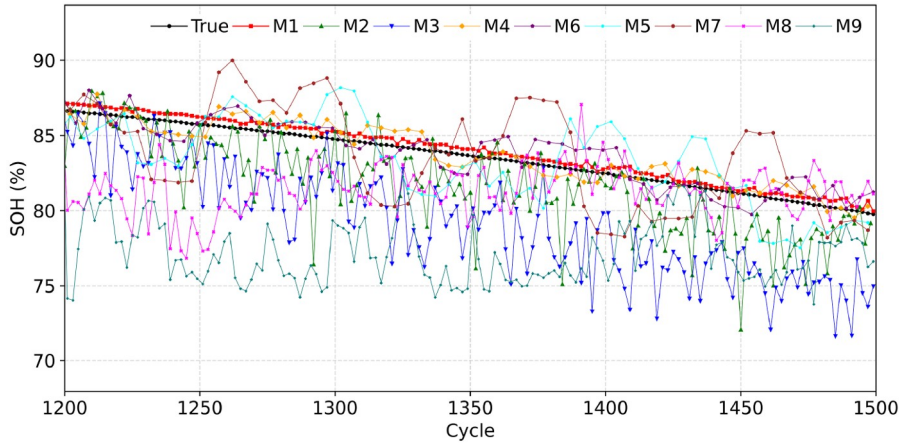


图 7　(网络版彩图) D2数据集上的 SOH 预测曲线对比
Figure 7　(Color online) SOH prediction curves of different models on D2.

更明显的上下波动与局部偏离，而两阶段方法(M8, M9)在局部区间内仍可能产生由补全误差传递导致的偏差. 这表明, 在缺失与噪声共存的条件下, SOH预测的稳定性不仅依赖预测器结构，还受上游表示的显著影响.

图7展示了在D2的SOH预测结果. 可以看到真实SOH呈现明显的非线性下降趋势，后期退化加速阶段的预测难度显著提升. 与对比模型相比，M1在后期退化加速阶段仍保持相对稳定的趋势一致性，而部分基线方法的预测点分散度更大，存在不同程度的系统性高估、低估或局部散点漂移现象.

综合两个数据集上各模型的SOH预测曲线, M1在局部细节拟合与全局退化趋势刻画方面均表现出更好的鲁棒性与一致性，更适用于面向SOH的连续多步预测场景.

5　结论与展望

为解决数据驱动方法在锂离子动力电池SOH预测无法有效使用不完备的电池性能数据的问题，本文提出了一个基于不完备张量表示学习的锂离子动力电池SOH预测方法. 该方法构建张量表示学习模块将不完

备的电池性能数据表示为完备的特征数据，并通过特征提取将循环次数-监测指标特征张量数据输入Transformer预测器以预测未来$|K|$次循环的电池性能数据，从而实现对电池SOH的精准预测. 为验证所提方法的有效性，本文在两个公开数据集的实验结果表明在不完备性能数据条件下，相比于其他先进方法本文所提方法在预测未来$|K|$次循环的性能数据时获得了最高的预测精度，因此能够实现精准的锂离子动力电池SOH预测. 在下一步工作中，将进一步考虑性能数据中的异常值和噪声以构建鲁棒的表示学习模型，另一方面，研究不同的张量分解框架，以构建更强学习能力的表示模型.

参考文献

1 Wu X, Chen J, Wang D, et al. Filter-Based Co-Estimation of State-of-Charge and State-of-Health for Lithium-Ion Batteries. IEEE Trans Ind Electron, 2025, 72: 12942–12951

2 Wang S, Ou K, Zhang W, et al. A state-of-charge and state-of-health joint estimation method of lithium-ion battery based on temperature-dependent extended kalman filter and deep learning. IEEE Trans Ind Electron, 2024, 72: 570–579

3 Alsuwian T, Ansari S, Ammirrul Atiqi Mohd Zainuri M, et al. A review of expert hybrid and co-estimation techniques for SOH and RUL estimation in battery management system with electric vehicle application. Expert Syst Appl, 2024, 246: 123123

4 Shu X, Yang H, Chen Z, et al. Meta learning based state of health estimation of lithium-ion batteries with small sampling retraining. IEEE Trans Ind Electron, 2025, 1–13

5 Shen L, Li J, Meng L, et al. Transfer learning-based state of charge and state of health estimation for Li-ion batteries: A review. IEEE Trans Transp Electrific, 2023, 10: 1465–1481

6 Jin J X, Yu R X, Liu G, et al. Research progress on state-of-health estimation methods for lithium-ion batteries (in Chinese). J Electr Eng, 2024, 19: 33–48 [金建新, 虞儒新, 刘刚, 等. 锂离子电池健康状态估算方法研究进展. 电气工程学报, 2024, 19: 33–48]

7 Ye Z, Yu J. State-of-health estimation for lithium-ion batteries using domain adversarial transfer learning. IEEE Trans Power Electron, 2021, 37: 3528–3543

8 Wei Z, Ruan H, Li Y, et al. Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging. IEEE Trans Power Electron, 2022, 37: 7432–7442

9 Zhao Y, Guo H, Guo X. Lithium-ion battery state-of-health estimation method based on hierarchical convolution and multichannel interactive dense transformer network. IEEE Trans Instrum Meas, 2025, 74: 1–14

10 Wang Y, Zhou Q, Zhang F, et al. Enhancing state of health estimation of lithium-ion batteries through strategic feature analysis. IEEE Trans Transp Electrific, 2025, 11: 11556–11569

11 Wei Z, Sun X, Wang J, et al. State of health estimation of lithium-ion batteries based on multihealth features fusion and improved group method of data handling. IEEE Trans Instrum Meas, 2024, 73: 1–15

12 Tang J, Li Y, Wang S, et al. Data-driven state of health estimation method of lithium-ion batteries for partial charging curves. IEEE Trans Energy Convers, 2024, 39: 2230–2243

13 Liu X, Hu Z, Mao L, et al. Adaptive state-of-health estimation for lithium-ion battery with partially unlabeled and incomplete charge curves. IEEE Trans Transp Electrific, 2025, 11: 6165–6176

14 Yan F Y, Li W Z, Yang W W, et al. Prediction of fuel cell performance based on Bagging neural network ensemble model (in Chinese). Sci Sin Tech, 2019, 49: 391–401 [闫飞宇, 李伟卓, 杨卫卫, 等. 基于Bagging神经网络集成的燃料电池性能预测方法. 中国科学: 技术科学, 2019, 49: 391–401]

15 Wu H, Wang Q, Luo X, et al. Learning accurate representation to nonstandard tensors via a mode-aware tucker network. IEEE Trans Knowl Data Eng, 2025, 37: 7272–7285

16 Wang Z P, Wang Q S, Liu P, et al. Review on techniques for power battery state of health estimation driven by big data methods (in Chinese). J Mech Eng, 2023, 59: 151–168 [王震坡, 王秋诗, 刘鹏, 等. 大数据驱动的动力电池健康状态估计方法综述. 机械工程学报, 2023, 59: 151–168]

17 Liu Y, Liu J, Long Z, et al. Tensor Computation for Data Analysis. Cham: Springer, 2022

18 Wang Q, Wu H, Luo X. A convolution bias-incorporated nonnegative latent factorization of tensors model for accurate representation learning to

dynamic directed graphs. IEEE Trans Syst Man Cybern Syst, 2025, 55: 8902–8914
19 Wan D, Xiangji Huang J. Missing value treatments for machine learning-based misbehavior detection systems: Survey, evaluation, and challenges. IEEE Trans Intell Transp Syst, 2025, 26: 12798–12818
20 Liao X, Wu H, Luo X. A novel tensor causal convolution network model for highly-accurate representation to spatio-temporal data. IEEE Trans Automat Sci Eng, 2025, 22: 19525–19537
21 Zhong S, Song S, Zhuo W, et al. A multi-scale decomposition MLP-mixer for time series analysis. Proc VLDB Endow, 2024, 17: 1723–1736
22 Wang C, Qi Q, Wang J, et al. Rethinking the power of timestamps for robust time series forecasting: A global-local fusion perspective. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS 2024). Cambridge: MIT Press, 2024. 22206–22232
23 Yi K, Fei J, Zhang Q, et al. FilterNet: Harnessing frequency filters for time series forecasting. In: Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2024. 55115–55140
24 Zeng C, Tian Y, Zheng G, et al. How much can time-related features enhance time series forecasting, arXiv: 2412.01557
25 Lin, S, Chen, H, Wu, H, et al. Temporal query network for efficient multivariate time series forecasting. In: Proceedings of the 42nd International Conference on Machine Learning. New York: ACM, 2025. 37797–37814
26 Lin S, Lin W, Hu X, et al. CycleNet: Enhancing time series forecasting through modeling periodic patterns. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS 2024). Cambridge: MIT Press, 2024. 106315–106345
27 Liu Q, Lu L, Chen Z. Non-negative Tucker decomposition with graph regularization and smooth constraint for clustering. Pattern Recognition, 2024, 148: 110207
28 Che H, Pan B, Leung M F, et al. Tensor factorization with sparse and graph regularization for fake news detection on social networks. IEEE Trans Comput Soc Syst, 2024, 11: 4888–4898

State-of-health prediction of lithium-ion power batteries via incomplete tensor representation learning

WU Hao[1], HOU JiaWen[1], BAO XueYan[1], ZENG NianYin[2], LUO Xin[1] & WANG ZiDong[3]

[1] *College of Computer and Information Science, School of Software, Southwest University, Chongqing 400715, China*
[2] *School of Aerospace Engineering, Xiamen University, Xiamen 361005, China*
[3] *College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China*

State of health (SOH) prediction of lithium-ion power batteries is central to battery health management and is essential for ensuring the reliability and safety of battery-operated systems. Data-driven SOH prediction methods have been widely adopted due to their high flexibility and the absence of a need for detailed electrochemical modeling. However, existing data-driven approaches typically rely on complete battery performance data (e.g., current, voltage, and impedance). In real-world operating conditions, unavoidable issues such as sensor malfunctions often lead to incomplete, noisy, or inconsistent performance data, which severely challenge accurate SOH prediction. To address this problem and achieve reliable SOH prediction under incomplete performance data, this paper proposes a lithium-ion power battery SOH prediction method based on incomplete tensor representation learning. First, in the tensor representation learning module, battery performance data are modeled as a fourth-order incomplete tensor with the modes "battery cell × sampling time × monitoring indicator × cycle index", and an incomplete tensor representation learning model is constructed to obtain complete feature representations. Second, in the feature extraction module, three factor matrices and a cycle-indicator feature tensor are introduced, and informative features are extracted through mode-wise tensor-matrix multiplication. Finally, in the performance prediction module, a Transformer-based model is employed to predict battery performance over the next $|K|$ cycles, enabling accurate SOH prediction. Experimental results on two widely used battery performance datasets show that the proposed method can accurately predict battery performance data in future cycles based on incomplete historical monitoring data, thereby achieving accurate SOH prediction.

lithium-ion power batteries, state-of-health prediction, incomplete tensor, representation learning

doi: 10.1360/SST-2025-0365