

一种动力电池稀疏大数据的稠密化还原方法

技术领域

[0001] 本发明主要涉及电池相关技术领域,具体是一种动力电池稀疏大数据的稠密化还原方法。

背景技术

[0002] 动力电池是电动汽车的核心组件之一,其性能对于电动汽车的动力性、经济性及安全有着举足轻重的作用。为了保障动力电池安全、高效的工作,电池管理系统应运而生,其功能主要包括数据采集、状态估计、均衡管理和信息管理等。然而,受到车载电池管理系统计算能力的限制,一些高精度的复杂算法无法应用于实车电池管理系统上。

[0003] 为了解决此类问题,随着大数据和车联网技术的发展,云端电池管理系统的概念被提出,即将动力电池数据传输到云端,借助云端算力实现电池管理系统的诸多功能。然而由于车联网技术存在数据传输带宽能力不足的问题,无法将获取的所有数据上传,因此为了避免数据传输过程中产生数据损失,通常会降低传输频率。这也导致上传到云端的电池数据存在信息缺失问题,无法充分满足动力电池管理系统中多种算法的数据需求。

发明内容

[0004] 为解决目前技术的不足,本发明结合现有技术,从实际应用出发,提供一种动力电池稀疏大数据的稠密化还原方法,采用带有外推器的掩码自编码器(Extra MAE)算法学习电池数据的时间动态特性,对稀疏数据进行还原。

[0005] 本发明的技术方案如下:

[0006] 一种动力电池稀疏大数据的稠密化还原方法,包括如下步骤:

[0007] S1、对电池稠密参考数据集进行归一化处理,将数据集各参数数值放缩到 $[0, 1]$,使得各参数数值处于同一数量级;

[0008] S2、训练过程中,将经归一化处理的电池稠密参考数据输入带有外推器的掩码自编码器(Extra MAE)模型中,模型结构上分为编码器、外推器和解码器,首先,Extra MAE将对电池数据集进行规则掩码操作,随后,编码器仅对未掩码的数据进行编码操作,将其映射到特征空间,外推器根据未掩码数据的潜在表征外推出全部数据的潜在表征,最终解码器根据全部数据的潜在表征映射出稀疏还原数据;

[0009] S3、Extra MAE模型在训练过程中根据所有补丁的损失而不断迭代,利用未被掩码的数据还原稠密参考数据,学习所训练电池数据的时间动态特性,最终产生满足精度要求的稀疏还原数据;

[0010] S4、Extra MAE模型训练完成后,可直接将大数据监控平台获取的电池稀疏特征数据作为输入,以生成相应的稀疏还原数据。此时,不再需要掩码操作。

[0011] 进一步,所述电池数据集包括电压数据集、温度数据集以及电流数据集中的至少一种。

[0012] 进一步,步骤S1中,采用的归一化公式为:

$$[0013] \quad x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

[0014] 其中, x_i 为数据集中某参数某时刻的值, x_{\max} 表示该参数在数据集中的最大值, x_{\min} 表示该参数在数据集中的最小值。

[0015] 进一步, Extra MAE模型中, 编码器由门控循环单元 (GRU) 和全连接 (FC) 层构成, 外推器采用FC层, 解码器由GRU和FC层组成;

[0016] 通过调节隐藏层单元数量和隐藏层层数不断优化模型, 平衡计算量与计算速度。

[0017] 进一步, Extra MAE模型的输入为包含d个特征的多元时间序列 $X = [x_1, x_2, \dots, x_d] \in \mathbb{R}^{d \times L}$, 其中每个特征的长度为L, 多元时间序列将被切片为T个规则不重叠补丁, 以补丁代替时间点作为输入的最小单位;

[0018] 第j个补丁 $P_j \in \mathbb{R}^{d \times l}$ 由 $l = \frac{L}{T}$ 个连续观测值 $P_j = [x_{(j-1)l+1}, x_{(j-1)l+2}, \dots, x_{jl+1}]$ 组成, 其中 $j=1, 2, \dots, T$ 。

[0019] 进一步, 步骤S2中掩码操作的具体方法包括:

[0020] 按照选定的掩码比例, 对数据中的m个补丁进行掩码操作, 其中被掩码的补丁的索引集为 $M = \{j_1, j_2, \dots, j_m\} \subset \{1, 2, \dots, T\}$, 未被掩码补丁的索引集为 $N = \{1, 2, \dots, T\} \setminus M = \{k_1, k_2, \dots, k_n\}$, 其中 $M+N=T$; 将不可见的被掩码补丁 $[P_{j_1}, P_{j_2}, \dots, P_{j_m}]$ 表示为 P_M , 将可见的未被掩码的补丁 $[P_{k_1}, P_{k_2}, \dots, P_{k_n}]$ 表示为 P_N , 被掩码的补丁 $P_j \in \mathbb{R}^{d \times l}$ 中的所有元素将被整体去除, 其中 $j \in M = \{j_1, j_2, \dots, j_m\}$ 。

[0021] 进一步, Extra MAE模型中, 编码器仅作用于未被掩码的补丁 $P_N \in \mathbb{R}^{d \times (n \cdot l)}$, 并将其映射到特征空间 $H_N \in \mathbb{R}^{h \times (n \cdot l)}$, 其中h为隐空间的维度;

[0022] 用函数 $E: \prod_{k=1}^{n \cdot l} (\mathbb{R}^d)_k \rightarrow \prod_{k=1}^{n \cdot l} (\mathbb{R}^h)_k$ 来表示从未掩码补丁 P_N 到其潜在表征 H_N 的编码器:

$$[0023] \quad H_N = E(P_N)。$$

[0024] 进一步, Extra MAE模型中, 外推器根据未掩码补丁的潜在表征恢复掩码补丁的潜在表征, 将外推器表示为从可见未掩码相邻表征中外推不可见掩码潜在表征的函数I:

$$[0025] \quad \prod_{k=1}^{n \cdot l} (\mathbb{R}^h)_k \rightarrow \prod_{k=1}^{T \cdot l} (\mathbb{R}^h)_k,$$

$$[0026] \quad \tilde{H} = I(H_N)。$$

[0027] 进一步, Extra MAE模型中, 解码器从外推的完整潜在表征中重建原始信号, 将解码器表示为函数 $D: \prod_{k=1}^{T \cdot l} (\mathbb{R}^h)_k \rightarrow \prod_{k=1}^{T \cdot l} (\mathbb{R}^d)_k$, 它将潜在表征 $\tilde{H} \in \mathbb{R}^{h \times (n \cdot l)}$ 映射为合成的时间序列

$$\hat{X} \in \mathbb{R}^{h \times (T \cdot l)},$$

$$[0028] \quad \hat{X} = D(\tilde{H})。$$

[0029] 进一步, 步骤S3中, 训练过程中, 首先对原始时间序列X中的一些补丁进行规则掩码, 然后ExtraMAE利用未被掩码的补丁 P_N 生成重建的时间序列 \hat{X} , 并计算重建时间序列 \hat{X} 在 $\mathbb{R}^{d \times L}$ 中与原始时间序列X之间的均方误差MSE, 模型重建序列的损失函数如下式所示:

$$[0030] \quad L_{recon} = \frac{1}{L} (X - \hat{X})^2$$

[0031] 在不断的迭代过程中,模型将学习电压数据的时间动态特性,从而使得稀疏还原数据与稠密参考数据具有相似的时间分布,同时,模型生成数据结果中将以数据未掩码部分的真实值代替未被掩码位置的生成数据,以提高生成数据的精度。

[0032] 模型仅在训练过程中需要进行掩码操作,模型训练完成后,可直接以大数据监控平台的稀疏特征数据作为输入,同时,模型生成数据结果中也将以稀疏数据的真实值代替对应位置的生成数据,以提高生成数据的精度,并最终生成稀疏还原数据。

[0033] 本发明的有益效果:

[0034] 本发明采用Extra MAE算法,该算法是一种可扩展的时间序列生成模型,其结构上包括编码器、外推器和解码器,编码器采用门控循环单元 (GRU) 和全连接 (FC) 层构成,外推器采用FC层,解码器采用GRU和FC层组成。对于不同任务,可通过调节隐藏层单元数量和隐藏层层数不断优化模型,平衡计算量与计算速度,在精确度方面,本发明产生的稀疏还原数据在电流、电压和温度三种电池外特性数据均具有较高的精度;在有效性方面,基于稀疏还原数据的SOC估计精度高于基于稀疏数据和基于稠密参考数据的SOC估计精度。本发明应用于云端稀疏电池特征数据处理的潜力得到证明。

附图说明

[0035] 图1以电压还原为例的Extra MAE结构图;

[0036] 图2Extra MAE方法稀疏数据还原效果;

[0037] 图3HWFET工况下稀疏数据和稀疏还原数据SOC估计效果对比;

[0038] 图4UDDS工况下稀疏数据和稀疏还原数据SOC估计效果对比;

具体实施方式

[0039] 结合附图和具体实施例,对本发明作进一步说明。应理解,这些实施例仅用于说明本发明而不适用于限制本发明的范围。此外应理解,在阅读了本发明讲授的内容之后,本领域技术人员可以对本发明作各种改动或修改,这些等价形式同样落于本申请所限定的范围。

[0040] 本实施例提供一种动力电池稀疏大数据的稠密化还原方法。本方法的主要步骤如下。

[0041] S1、对电池稠密参考数据集进行归一化处理,即将数据集各参数数值放缩到[0, 1],使得各参数数值处于同一数量级,加快模型收敛速度,采用的归一化公式为:

$$[0042] \quad x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

[0043] 其中, x_i 为数据集中某参数某时刻的值, x_{\max} 表示该参数在数据集中的最大值, x_{\min} 表示该参数在数据集中的最小值。

[0044] S2、数据处理采用Extra MAE算法,该算法是一种可扩展的时间序列生成模型,其结构上包括编码器、外推器和解码器。

[0045] 此处以电池电压数据举例,如图1所示,图中最左侧为稠密参考数据,浅灰色点表示需要进行掩码的数据,将该部分数据进行规则掩码操作。编码器仅对未掩码的数据进行

编码操作,将其映射到特征空间,外推器将根据未掩码数据的潜在表征外推出全部数据的潜在表征,最终解码器将根据全部数据的潜在表征映射出稀疏还原数据。

[0046] S3、Extra MAE模型在训练过程中根据所有补丁的损失而不断迭代,根据未被掩码的数据还原稠密参考数据,学习所训练电池数据的时间动态特性,最终产生满足精度要求的稀疏还原数据。Extra MAE是一种简单的、可扩展的时间序列生成模型,本发明中,编码器采用门控循环单元(GRU)和全连接(FC)层构成,外推器采用FC层,解码器采用GRU和FC层组成。对于不同任务,可通过调节隐藏层单元数量和隐藏层层数不断优化模型,平衡计算量与计算速度。

[0047] ExtraMAE计算所有补丁的损失,整体重构损失保证了重构后的时间序列保持了原始时间序列的连续性。但是因为未被掩码数据在编码、外推、解码的过程中同样会产生误差,而该数据在实际应用中是可见的,所以生成数据结果中将以数据未掩码部分的真实值代替未被掩码位置的生成数据,以提高生成数据的精度。

[0048] S4、Extra MAE模型训练完成后,可直接将大数据监控平台获取的电池稀疏特征数据作为输入,以生成相应的稀疏还原数据。此时,不再需要掩码操作。

[0049] 1、在本实施例提供的动力电池稀疏大数据的稠密化还原方法中,可以是电压稀疏数据、电流稀疏数据以及温度稀疏数据。本实施例以电压稀疏数据还原举例,证明本方法的有效性。

[0050] 对于特定的时间序列还原任务,模型的输入为包含d个特征的多元时间序列 $X = [x_1, x_2, \dots, x_d] \in \mathbb{R}^{d \times L}$,其中每个特征的长度为L。为减小掩码解码等操作的计算量,多元时间序列将被切片为T个规则不重叠补丁,以补丁代替时间点作为输入的最小单位。其中第j个补丁 $P_j \in \mathbb{R}^{d \times l}$ 由 $l = \frac{L}{T}$ 个连续观测值 $P_j = [x_{(j-1)l+1}, x_{(j-1)l+2}, \dots, x_{jl+1}]$ 组成,其中 $j = 1, 2, \dots, T$ 。在如图1所示电池电压稀疏数据还原任务中,输入序列 $X = U_i \in \mathbb{R}^{1 \times 12}$ 。另外,由于云端平台接受到的稀疏数据为每秒1个,为了保证后续工作中被掩码补丁与未被掩码补丁维度的一致,取补丁大小为1。

[0051] 随后,按照选定的掩码比例,对数据中的m个补丁进行掩码操作。其中被掩码的补丁的索引集为 $M = \{j_1, j_2, \dots, j_m\} \subset \{1, 2, \dots, T\}$,未被掩码补丁的索引集为 $N = \{1, 2, \dots, T\} \setminus M = \{k_1, k_2, \dots, k_n\}$,其中 $M + N = T$ 。为简单起见,将不可见的被掩码补丁 $[P_{j_1}, P_{j_2}, \dots, P_{j_m}]$ 表示为 P_M ,将可见的未被掩码的补丁 $[P_{k_1}, P_{k_2}, \dots, P_{k_n}]$ 表示为 P_N 。被掩码的补丁 $P_j \in \mathbb{R}^{d \times l}$ 中的所有元素将被整体去除,其中 $j \in M = \{j_1, j_2, \dots, j_m\}$ 。在如图1所示的电池电压稀疏数据还原任务中,随机掩码 $m = 8$ 个补丁,将剩下的 $n = 4$ 个补丁拼接从而得到 P_N ,被掩码补丁的索引集 $M = \{2, 3, 5, 6, 8, 9, 11, 12\}$,未被掩码补丁的索引集 $N = \{1, 4, 7, 10\}$ 。

[0052] 编码器仅作用于未被掩码的补丁 $P_N \in \mathbb{R}^{d \times (n \cdot l)}$,并将其映射到特征空间 $H_N \in \mathbb{R}^{h \times (n \cdot l)}$,其中h为隐空间的维度。在如图1所示的电池电压稀疏数据还原任务中,编码器将 $P_N = [P_1, P_4, P_7, P_{10}]$ 映射到潜在表征空间 $(H_N)_{4 \times 3}$ 。如式(1)所示,用函数 $E: \prod_{k=1}^{n \cdot l} (\mathbb{R}^d)_k \rightarrow \prod_{k=1}^{n \cdot l} (\mathbb{R}^h)_k$ 来表示从未掩码补丁 P_N 到其潜在表征 H_N 的编码器。

[0053] $H_N = E(P_N)$ (1)

[0054] 外推器根据未掩码补丁的潜在表征恢复掩码补丁的潜在表征,如式(2)所示,将外

推器表示为从可见未掩码相邻表征中外推不可见掩码潜在表征的函数 $I: \prod_{k=1}^{n \times l} (R^h)_k \rightarrow \prod_{k=1}^{T \times l} (R^h)_k$ 。在图1所示的电池电压稀疏数据还原任务中,外推器从未被遮盖补丁的潜在表征 $(H_N)_{4 \times 3}$ 推断出所有补丁的潜在表征 $(\tilde{H})_{12 \times 3}$ 。

$$[0055] \quad \tilde{H} = I(H_N) \quad (2)$$

[0056] 解码器从外推的完整潜在表征中重建原始信号。如式(3)所示,将解码器表示为函数 $D: \prod_{k=1}^{T \times l} (R^h)_k \rightarrow \prod_{k=1}^{T \times l} (R^d)_k$, 它将潜在表征 $\tilde{H} \in \mathbb{R}^{h \times (n \times l)}$ 映射为合成的时间序列 $\hat{X} \in \mathbb{R}^{h \times (T \times l)}$ 。在图1所示的电池电压稀疏数据还原任务中,解码器从外推器输出的所有补丁的潜在表征 $(\tilde{H})_{12 \times 3}$ 映射出稀疏还原数据 $\hat{X} = U_i \in \mathbb{R}^{1 \times 12}$

$$[0057] \quad \hat{X} = D(\tilde{H}) \quad (3)$$

[0058] 本发明中,编码器、外推器和掩码器为由GRU和全连接层组成的神经网络结构。仍以图1所示电池电压稀疏数据还原任务为例,Extra MAE各模块组成如表1所示。

[0059] 表1Extra MAE各模块结构(以电压稀疏数据还原为例)

[0060]	模块	第一层	第二层
	编码器	GRU(1,5,3)	FC(5, 5)
	外推器	FC(4, 12)	FC(5, 5)
	解码器	GRU(5,5,3)	FC(5, 1)

[0061] 训练过程中,首先对原始时间序列X中的一些补丁进行规则掩码,然后ExtraMAE利用未被掩码的补丁 P_N 生成重建的时间序列 \hat{X} ,并计算重建时间序列 \hat{X} 在 $\mathbb{R}^{d \times L}$ 中与原始时间序列X之间的均方误差MSE。模型重建序列的损失函数如式(4)所示:

$$[0062] \quad L_{recon} = \frac{1}{L} (X - \hat{X})^2 \quad (4)$$

[0063] 在不断的迭代过程中,模型将学习电压数据的时间动态特性,从而使得稀疏还原数据与稠密参考数据具有相似的时间分布。同时,模型生成数据结果中将以数据未掩码部分的真实值代替未被掩码位置的生成数据,以提高生成数据的精度。

[0064] 2、稀疏还原数据精确度和有效性验证

[0065] 基于松下电池公开数据集,验证本发明数据还原能力和稀疏还原数据于SOC估计中的结果以证明其精度和有效性。

[0066] 松下电池公开数据集记录于松下NCR18650PF锂离子电池,该电池标称容量为2.32Ah。整个测试在由Digatron Firing Circuit生产的测试设备上完成,利用温度控制箱,分别在25℃、10℃和0℃环境下对所选电池在不同动态行驶工况下进行测试。具体实验过程如下:

[0067] 首先将温控箱设置为所选温度,然后静置3小时,保证电池内部温度维持在所选温度。随后,采用1C恒流和4.2V恒压对电池进行充电,当充电电流低于50mA时,停止充电过程。然后温控箱设置为所选温度,静置3小时以使电池内部温度稳定。随后进行动态驾驶循环测试,包括HWFET、HPPC、LA92、US06、UDDS以及混合工况Cycle4和NN。其中,NN工况由LA92工况

与US06工况的放电片段混合,Cycle4由US06,HWFET,UDDS,LA92的放电片段混合而成。

[0068] 以松下数据集中HPPC、Cycle4、LA92、US06四个动态驾驶循环对应的数据构成SOC估计训练数据集,以NN循环对应的数据作为验证数据集,以HWFET循环和UDDS循环对应的数据作为测试数据集。所有数据集均包含电池在25℃、10℃和0℃下的信息。

[0069] 模型输入为 $X = (U_i, I_i, T_i) \in \mathbb{R}^{3 \times 20}$,其中20为所选序列长度。掩码率设置为80%,即在20个时间点中,被掩码补丁数为16,未被掩码补丁数为4,模拟将0.2Hz的数据还原为1Hz的数据。相应的,编码器、外推器和解码器的结构参数设置如表2所示。

[0070] 表2Extra MAE结构设置

[0071]	模块	第一层	第二层
	编码器	GRU(3, 12, 3)	FC(12, 12)
	外推器	FC(4, 20)	FC(12, 12)
	解码器	GRU(12, 12, 3)	FC(12, 3)

[0072] 模型经过50000epoch的训练,数据还原效果如图2所示。电流、电压和温度的稀疏还原数据均与原稠密参考数据具有相似的分布。其在HWFET工况和UDDS工况下的误差指标分别如表3和表4所示。

[0073] 表3HWFET工况下数据还原评价指标

[0074]	特征及单位	25℃			10℃			0℃		
		MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
	电压(V)	0.019	0.028	0.404	0.022	0.033	0.281	0.018	0.030	0.391
	电流(A)	0.435	0.651	5.781	0.499	0.725	5.965	0.310	0.455	3.314
	温度(℃)	0.121	0.157	0.734	0.062	0.099	0.471	0.085	0.118	0.465

[0075] 表4UDDS工况下数据还原评价指标

[0076]	特征及单位	25℃			10℃			0℃		
		MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
	电压(V)	0.013	0.021	0.398	0.017	0.029	0.395	0.018	0.032	0.347
	电流(A)	0.366	0.595	4.832	0.370	0.584	4.655	0.266	0.442	2.973
	温度(℃)	0.091	0.122	0.641	0.061	0.097	0.643	0.105	0.141	1.822

[0077] 从表中可以看出,在HWFET工况下,所还原电压数据的MAE误差在0.022V以内,所还原电流数据的MAE误差在0.499A以内,所还原温度数据的MAE误差在0.121℃以内。在UDDS工况下,所还原电压数据的MAE误差在0.018V以内,所还原电流数据的MAE误差在0.370A以内。所还原温度数据的MAE误差在0.105℃以内。从整体上来看,还原数据的精确度较高。

[0078] 为了验证数据还原模型在电池状态估计任务中的有效性,在SOC估计任务中采用CNN-GRU模型对不同类型数据进行对比验证。将此前在训练集、验证集和测试集得到的稀疏还原数据继续作为SOC估计模型的训练集、验证集和测试集,将稀疏数据和稠密参考数据按照前述同样的方式进行数据集划分。CNN-GRU模型的参数设置如表5所示,设置batch大小为64,时间窗口大小为120,学习率为 9×10^{-3} 。

[0079] 表5CNN-GRU模型参数设置

[0080]	网络层	参数
	CNN 层	(1, 8, (8, 3))
	CNN 层	(8, 32, 8)
	RNN 层	(106,32,2)
	FC 层	(32, 8)
	FC 层	(8, 1)

[0081] 经过1000个epoch的训练,得到如图3和图4所示的两种工况稀疏数据和稀疏还原数据的SOC估计结果,误差指标分别如表6和表7所示。

[0082] 表6HWFET工况下三种数据误差指标

[0083]	模型输入	25℃			10℃			0℃		
		MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
		(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
	稀疏数据	2.26	2.69	5.64	3.98	9.63	85.76	1.50	2.36	7.38
	稠密参考数据	2.33	2.90	6.68	4.26	4.88	10.00	1.91	2.19	13.94
	稀疏还原数据	0.39	0.51	2.36	0.92	1.15	3.49	0.29	0.39	1.70

[0084] 表7UDDS工况下三种数据误差指标

[0085]	模型输入	25℃			10℃			0℃		
		MAE	RMSE	MAX	MAE	RMSE	MAX	MAE	RMSE	MAX
		(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
	稀疏数据	0.90	1.20	6.84	1.12	1.52	5.88	1.70	1.98	4.15
	稠密参考数据	0.78	1.08	4.05	1.22	1.67	6.64	0.86	1.12	4.54
	稀疏还原数据	0.34	0.42	1.95	0.33	0.50	2.66	0.41	0.48	1.98

[0086] 从表6和表7中可以看出,在HWFET工况下,稀疏数据相较于稠密数据拥有更好的SOC估计表现,分析原因是稠密数据相比于稀疏数据存在数据冗余,在模型拟合中出现过拟合现象。相比于稀疏数据和稠密数据,稀疏还原数据在两种工况三种温度下均有较好的表现。相较于稀疏数据,稀疏还原数据填补了其缺失的数据;相比于稠密数据,稀疏还原数据相当于在其中加入了噪声,进行了数据增强。该结果证明了本发明在SOC估计方面的有效性以及应用于云端电池管理平台的潜力。

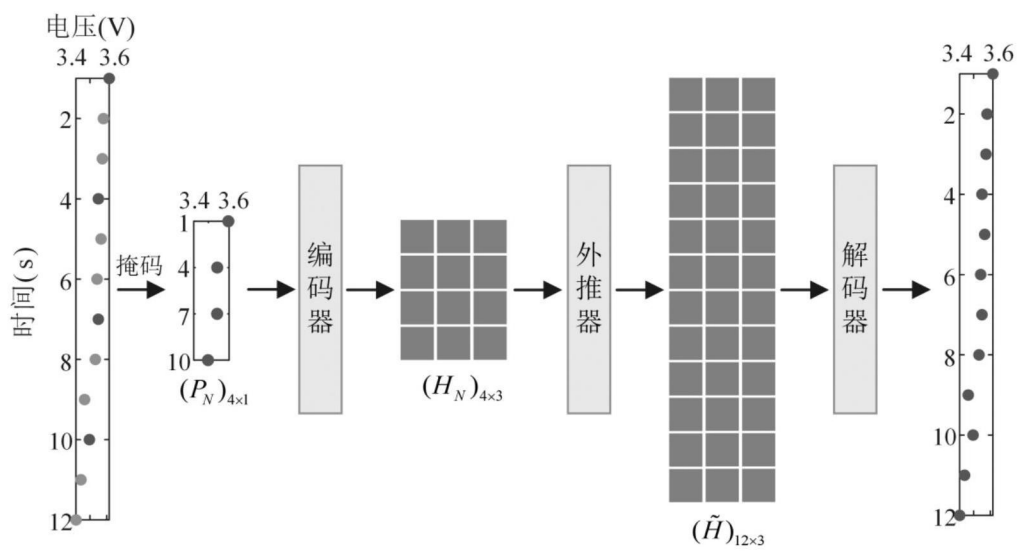


图1

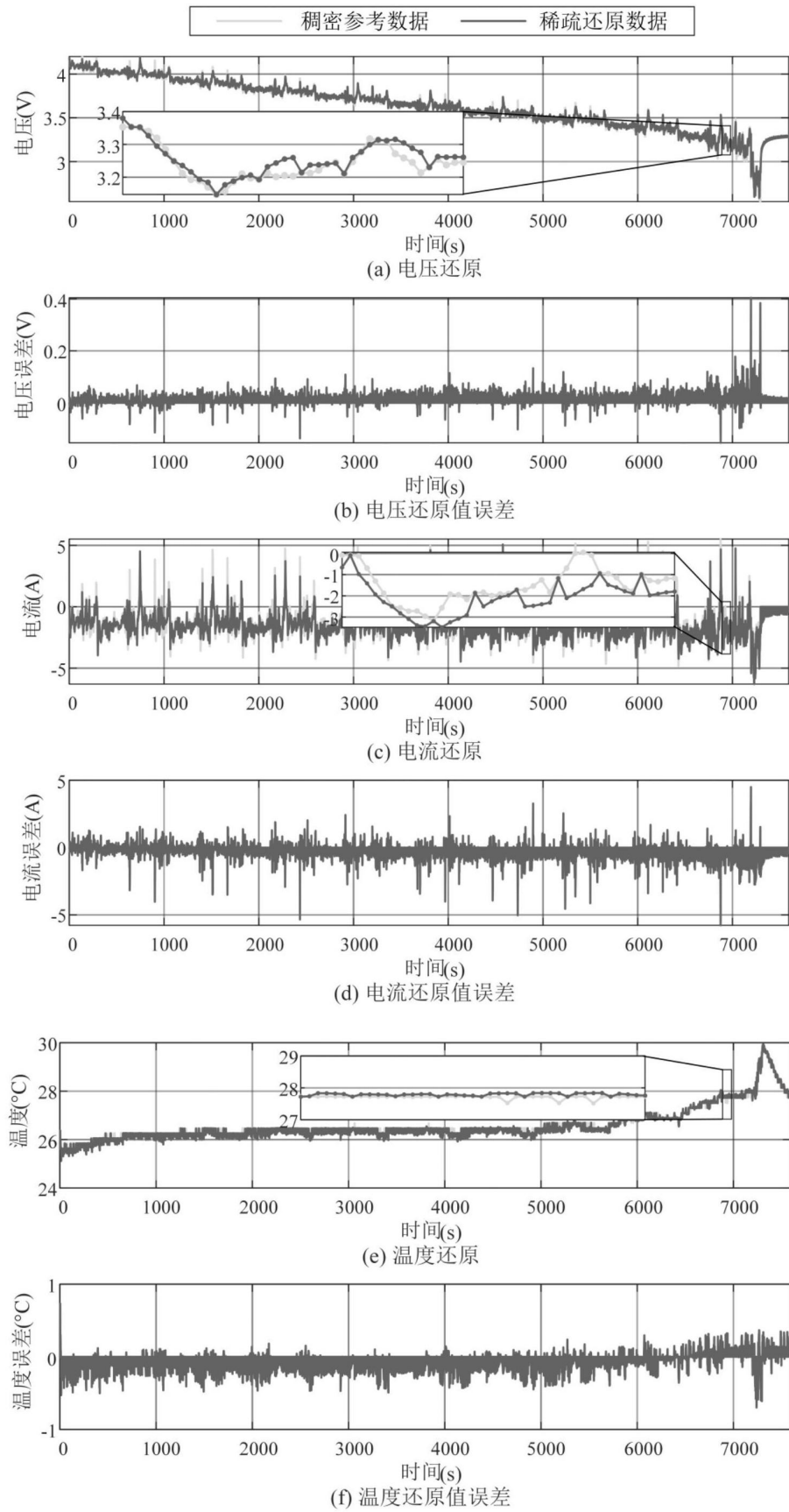


图2

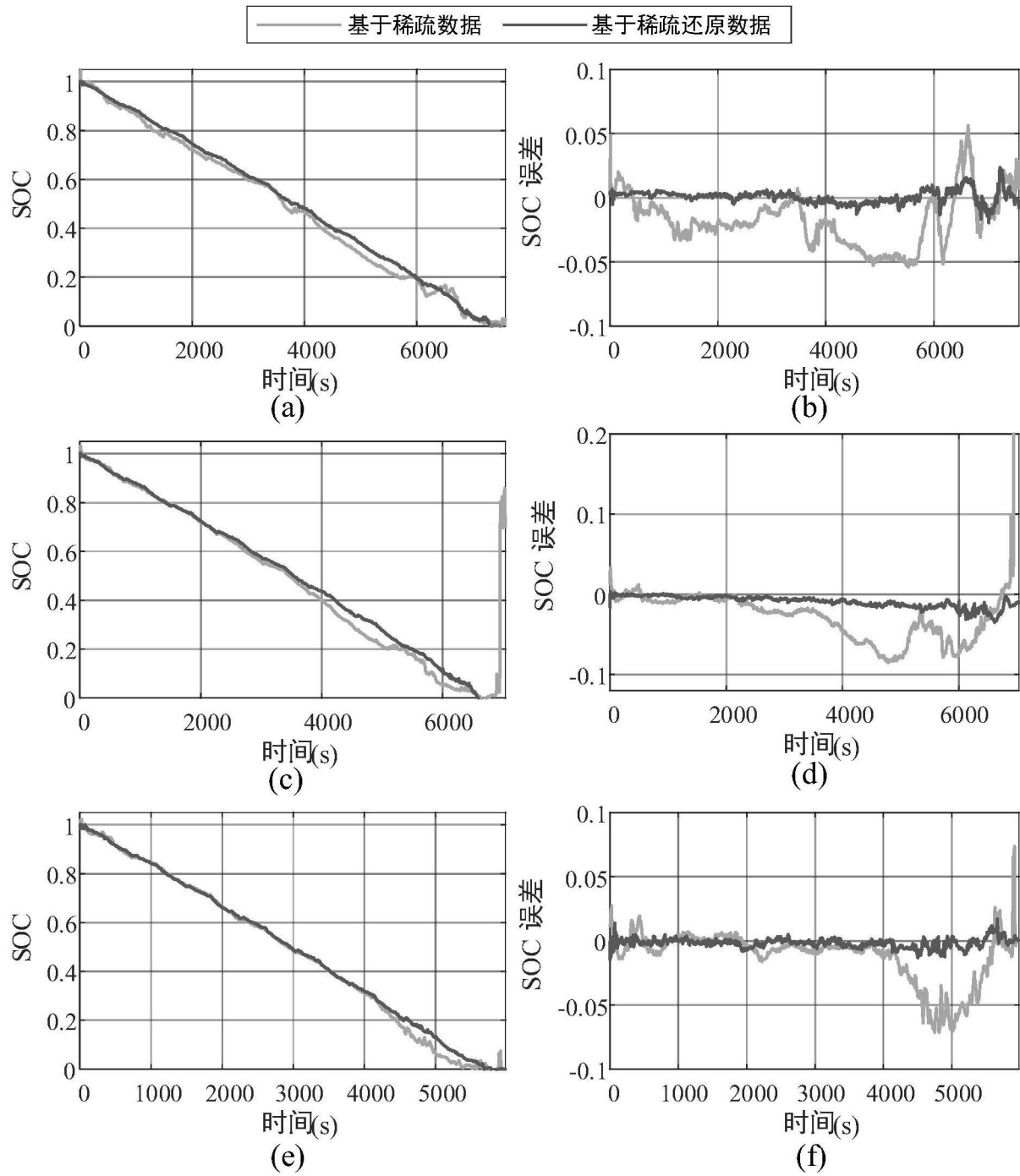


图3

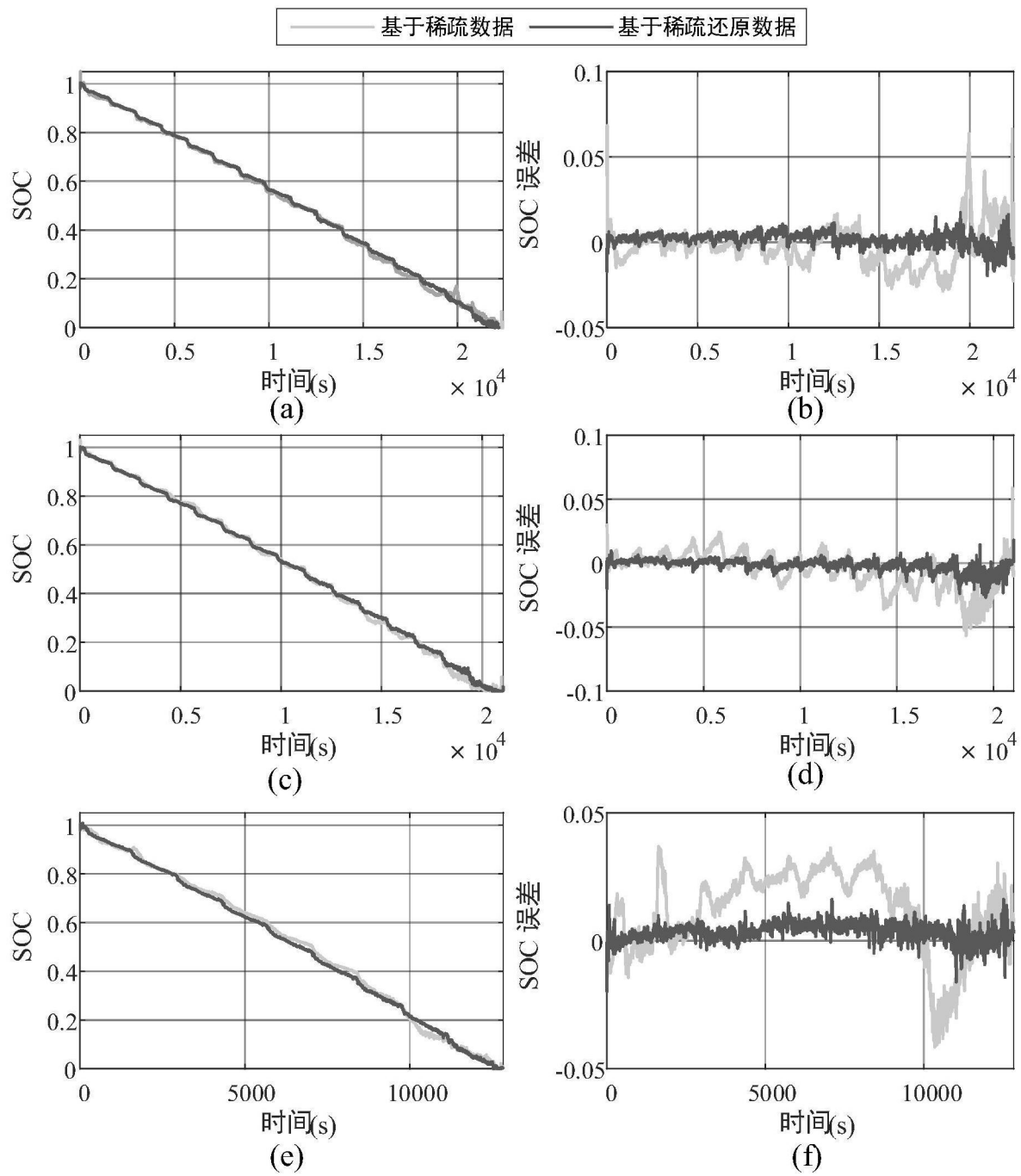


图4