

【电子与信息科学 / Electronics and Information Science】

基于混合模型的非侵入式负荷监测数据的生成

肖勇^{1,2}, 谈竹奎³, 钱斌^{1,2}, 张俊玮³, 罗奕^{1,2}, 张帆^{1,2}, 黄军力^{1,2}

1) 南方电网科学研究院有限责任公司用电与计量技术研究所, 广东广州510663; 2) 广东省电网智能量测与先进计量企业重点实验室, 广东广州510663; 3) 贵州电网有限责任公司, 贵阳贵州550002

摘要: 非侵入式负荷监测(non-intrusive load monitoring, NILM)是一种无需进入每个用电器内部系统, 仅在用户总线入口处安装监测设备的技术. 在开展NILM技术研究时, 往往需要收集大规模的用户负荷数据来证明所提出方法的普适性, 此需求不可避免地带来了繁重的数据收集与整理负担. 为克服该挑战, 设计了一种结合周期信号频率不变变换(frequency invariant transformation for periodic signals, FIT-PS)原理与时间序列生成对抗网络(time series generative adversarial networks, TimeGAN)的混合模型, 记为FIT-PS-TimeGAN. 针对全球家庭与工业瞬态能量数据集(worldwide household and industry transient energy dataset, WHITED)中的空调、微波炉、吸尘器、冰箱和热水壶5种电器, 运用FIT-PS对负荷数据集进行切割和拼接, 构建TimeGAN不同状态下的训练集和测试集. 评估测试集的效果发现, 生成的波形数据与真实数据表现出高度一致性. 进一步采用FIT-PS对训练得到的生成数据进行截取和拼接, 生成满足测试需求的完整的单负荷波形和多负荷波形. 对这些生成的波形与相同状态下的真实数据进行对比, 结果显示两者吻合度很高. 与自回归模型和生成对抗网络(generative adversarial network, GAN)模型相比, FIT-PS-TimeGAN模型在生成数据的性能方面表现更优. 研究结果表明, FIT-PS-TimeGAN混合模型能够有效生成符合标准电器运行规律的波形和场景数据.

关键词: 电力系统及其自动化; 人工智能; 非侵入式负荷监测; 数据生成方法; 周期信号频率不变变换; 时间序列生成对抗网络

中图分类号: TM93; TP18

文献标志码: A

DOI: 10.3724/SP.J.1249.2025.01085

Generation method of non-intrusive load monitoring data based on hybrid model

XIAO Yong^{1,2}, TAN Zhukui³, QIAN Bin^{1,2}, ZHANG Junwei³, LUO Yi^{1,2},
ZHANG Fan^{1,2}, and HUANG Junli^{1,2}

1) Institute of Electricity Consumption and Metering Technology, China Southern Power Grid Science Research Institute Co. Ltd.,
Guangzhou 510663, Guangdong Province, P. R. China

2) Guangdong Provincial Key Laboratory of Intelligent Measurement and Advanced Metering of Power Grid, Guangzhou 510663,
Guangdong Province, P. R. China

3) Guizhou Power Grid Co. Ltd., Guiyang 550002, Guizhou Province, P. R. China

Abstract: Non-intrusive load monitoring (NILM) is a technique that does not require accessing the internal system of each electrical device to monitor user's equipments, but only to monitor them at the entrance of user's bus. During the investigation of NILM techniques, it is often necessary to collect extensive user load data to confirm the

Received: 2024-06-12; **Accepted:** 2024-11-03; **Online (CNKI):** 2024-11-27

Foundation: Science and Technology Project of China Southern Power Grid Corporation (GZKJXM20222417); Science and Technology Planning Project of Guangdong Province (2021B1212050014)

Corresponding author: Senior engineer LUO Yi (luoyi_csg@outlook.com)

Citation: XIAO Yong, TAN Zhukui, QIAN Bin, et al. Generation method of non-intrusive load monitoring data based on hybrid model [J]. Journal of Shenzhen University Science and Engineering, 2025, 42(1): 85-93. (in Chinese)



applicability of proposed methods. This requirement inevitably entails a significant burden of data collection and organization. In order to overcome this challenge, we proposed a hybrid approach that combines the principle of frequency invariant transformation for periodic signals (FIT-PS) with time series generative adversarial networks (TimeGAN), denoted as FIT-PS-TimeGAN. Using a Worldwide Household and Industry Transient Energy Dataset (WHITED), we focused on five appliances: air conditioner, microwave oven, Hoover, refrigerator and kettle. FIT-PS was employed to segment and stitch the load data aiming to construct training and testing sets for TimeGAN under different states. The validation results on effectiveness of testing sets demonstrated high consistency between the generated waveforms and the real data. Subsequently, FIT-PS was applied to intercept and splice the training data to generate complete single-load waveforms and multi-load waveforms that are able to meet the testing requirements. These generated waveforms were compared with real data in the same state and the comparative results showed that a favorable agreement between the generated and real data. In addition, compared with the other two generation models (autoregressive model and GAN model), FIT-PS-TimeGAN outperforms better in terms of data generation performance. In summary, the FIT-PS-TimeGAN hybrid model is capable of effectively generating waveforms and scenarios that comply with the operational principles of standard appliances.

Key words: power system and automation; artificial intelligence; non-intrusive load monitoring; data generation method; frequency invariant transformation for periodic signal; time series generative adversarial network

非侵入式负荷监测(non-intrusive load monitoring, NILM)技术核心在于能否在大量多样化的数据集上展现出普适性和鲁棒性,这对测试数据集的规模性和场景代表性都提出了新的要求.为降低数据库构建的负担,同时适应快速的技术迭代需求,开源数据共享平台^[1]和NILM合成数据技术^[2]逐渐成为研究界的宝贵资源.一方面,通过创建现有数据的转换版本来人为增加训练数据规模可提高模型的泛化能力^[3];另一方面,基于真实负荷特征模拟而来的用户负荷数据,能够在保护个人隐私的同时,为算法测试和优化提供丰富的“试验田”.

学者们深入研究了众多非侵入式的数据合成技术,如基于周期信号频率不变变换(frequency invariant transformation for periodic signals, FIT-PS)的线性叠加^[4]、时间序列相似性^[5]和背景噪音融入^[6]等,实现了从有限的原始数据中产生大量具有多样性的新样本,增强了模型的学习能力和泛化性能.为解决训练样本分布不均衡的问题,线性插值^[7]、辅助分类器的生成对抗网络^[8]和去噪扩散概率模型^[9]等被用来对训练样本进行扩充.此外,考虑到采样数据连续性缺失的问题,杨挺等^[10]提出基于典范多因子分解的正则化低秩张量补全的量测数据缺失修复方法.然而,新增样本仅被用于训练,并未用于测试验证,文献也未明确指明这些样本是否适用于评估模型的泛化性能.考虑到电力消费行为的复杂性,有学者通过结合仿真工具生成具有实际意义的合成智能电表数据^[11-14],且这些合成数据已被

证明适用于训练具有鲁棒性的NILM模型.此外,非侵入式的合成数据技术通常基于自回归模型^[15]和生成对抗网络^[16-18](generative adversarial network, GAN)等先进的人工智能技术,能够学习原始数据的潜在分布,并据此生成新的、但与真实数据难以区分的样本.

综上所述,数据合成技术已在生成NILM训练数据方面得到了广泛应用,但用于评估模型泛化能力的测试集仍局限于有限的真实数据集.为此,本研究设计并实施了一种FIT-PS-TimeGAN混合模型.首先,遵循FIT-PS设计原则,从实际数据库中抽取不同状况下的电流采样信号,构建了TimeGAN模型的训练集和测试集.然后,验证TimeGAN模型训练生成的波形数据与实际数据的一致性.最后,对整个负荷场景的运行状况进行划分,再截取和拼接TimeGAN模型训练生成的波形数据,生成连续的负荷运行波形,并对其与实际负荷场景的吻合度进行测试.

1 模型设计

为验证NILM在各种环境下的广泛适用性,本研究采用混合模型生成一套合成数据集,以加强负荷细分、能耗管理和需求响应等功能实现^[19-20].采用FIT-PS提取电压和电流采样波形,以此作为时间序列生成对抗网络(time series GAN, TimeGAN)的训练与测试数据集.随后,对TimeGAN生成的新

波形进行截取和融合, 构建能反映不同时间段及多种模式组合的单负荷波形与多负荷波形数据集.

1.1 FIT-PS

信号分解方法 FIT-PS 的特点在于分解过程不依赖于信号的主频率, 就能够有效构建出无周期性振荡的信号表征. 该技术以电压信号的每个周期起始点为基准, 确立周期内后续采样点的位置, 进而实现电流信号在不同周期的分割^[21-22]. 由于电网的频率不是恒定的 50 Hz 或 60 Hz, 因此需要对当前信号进行电源频率不变表示. 图 1 以电冰箱的采样信号为例, 展示了非恒定频率的两个相邻整周波以及经过内插后得到的新的电流波形. 其中, 周期数为 2; 每个周期内的采样点数为 50. 电冰箱的采样频率为 2.5 kHz, 信号频率为 50 Hz. 该方法不仅确保了功率、相位、低频与高频成分等关键信息的完整保留, 还促进了稳态与瞬态的有效区分.

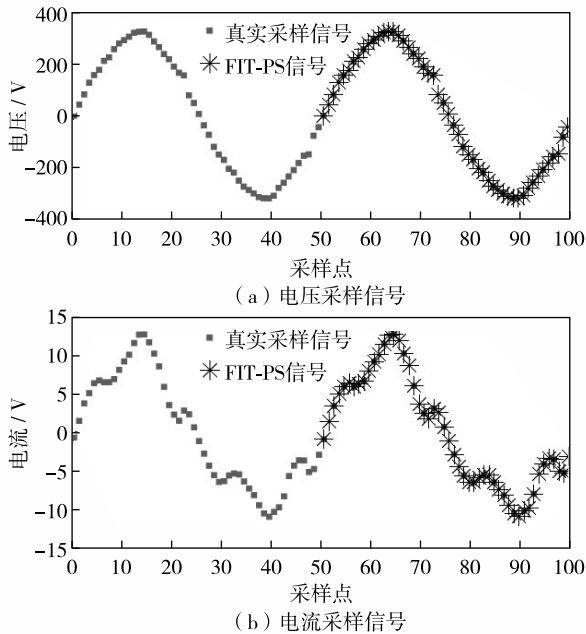


图1 FIT-PS处理电冰箱采样信号的基本过程

Fig. 1 (Color online) FIT-PS basic process for processing refrigerator sample signals. (a) Voltage sample signal and (b) current sample signal. (The black square represents the real sampling signal, and the symbol * is FIT-PS signal.)

1.2 TimeGAN 模型

基于 GAN 的框架的 TimeGAN 模型^[23]能够高效生成多个领域中的真实时间序列数据. 它融入了监督损失的概念, 通过整合无监督的 GAN 与有监督的自回归模型^[24], 促进模型借助原始数据的监督信息, 有效捕获了数据内的时序条件分布. TimeGAN

模型采用闭环模式进行训练, 其框架如图 2, 图中箭头表示各个网络层的数据传输方向.

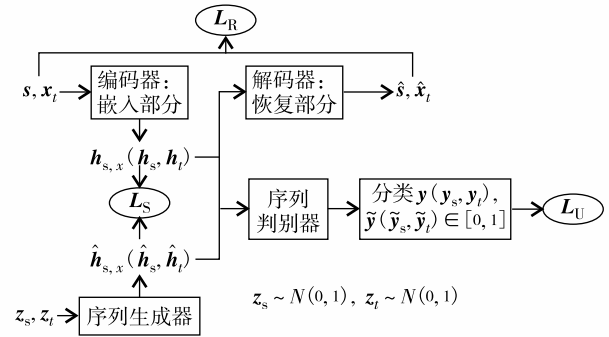


图2 TimeGAN模型框架

Fig. 2 Framework diagram of the TimeGAN model.

TimeGAN 框架的输入一部分是由静态特征向量 s 和时间特征向量 x_t 组成, 另一部分是从已知分布中提取的静态随机特征向量和时间随机特征向量的元组 $\{z_s, z_t\}$. 仅依靠序列判别器的二元对抗反馈可能不足以激励序列生成器捕捉数据中的阶跃条件分布^[27]. 因此, 引入额外的损失来进一步约束学习. 设 L_R 和 L_S 分别为有监督的数据重构和潜在分布损失函数, L_U 为无监督的生成对抗网络损失函数,

$$L_R = E \left[\|s - \hat{s}\|_2 + \sum_{t=1}^N \|x_t - \hat{x}_t\|_2 \right] \quad (1)$$

$$L_U = E \left[\ln(1 - \hat{y}_s) + \sum_{t=1}^N \ln(1 - \hat{y}_t) \right] + E \left[\ln y_s + \sum_{t=1}^N \ln y_t \right] \quad (2)$$

$$L_S = E \left[\sum_{t=1}^N \|h_t - \hat{h}_{s,x}\|_2 \right] \quad (3)$$

其中, E 为求期望操作; $\hat{s}$ 和 $\hat{x}_t$ 分别为解码器重建的静态特征向量和时间特征向量; y_s 和 y_t 分别为对应于 s 和 x_t 的真实的电器标签; $\hat{y}_s$ 和 $\hat{y}_t$ 分别为对应于 s 和 x_t 的判别器的预测概率; h_s 和 h_t 分别对应于编码器产生的关于静态特征和时间特征的隐藏状态, $h_{s,x}$ 为 h_s 和 h_t 的联合表示; $\hat{h}_{s,x}$ 为序列生成器产生的随机特征隐藏状态.

通过赋予 L_S 不同的权重系数 λ 和 η , 从而将其与 L_R 和 L_U 进行组合和线性叠加, 并反复迭代寻找综合最小化损失, 其表达式为

$$L_{\text{overall}} = \min(\lambda L_S + L_R) + \min(\eta L_S + L_U) \quad (4)$$

该结合方式使 TimeGAN 可以同时训练编码(特征向量)、生成(潜在表征)和迭代(跨时间). 此外, 由于对抗学习发生在低维潜在空间且监督损失对生

成器的逐步动态也有约束作用. 因此, TimeGAN 的训练难度不会增加.

1.3 FIT-PS-TimeGAN 混合模型

电压和电流数据在实测/录波时, 其频率不是固定不变的. 为捕捉其时序特征和完整保留电气特

性, 根据 FIT-PS 基本原理对 TimeGAN 的输入和输出波形进行信号同步, 并利用 TimeGAN 对不同状态和时段的信号进行数据生成. 最后, 利用评价指标对不同维度信号的相似性和真实性进行评价. FIT-PS-TimeGAN 模型数据生成流程如图 3.

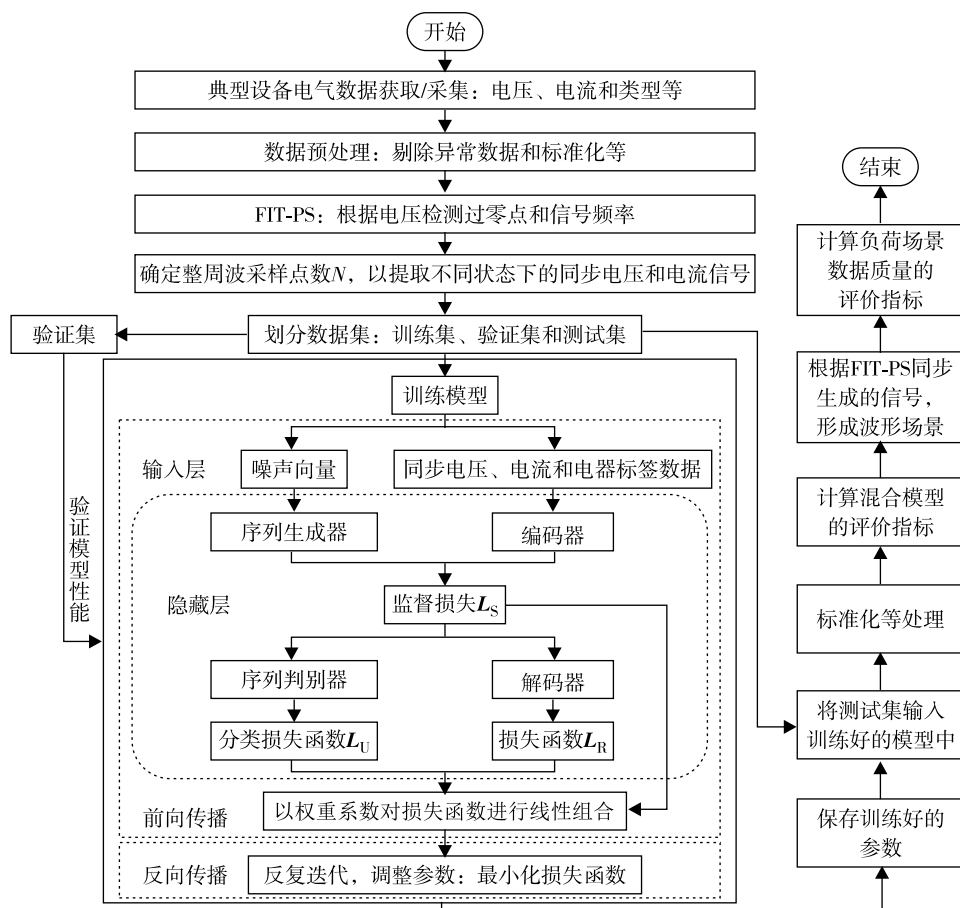


图3 FIT-PS-TimeGAN 模型数据生成流程图

Fig. 3 Flowchart of hybrid model (FIT-PS-TimeGAN) data generation.

2 实验设计与验证

2.1 实验整体设计

2.1.1 实验数据

实验使用全球家庭与工业瞬态能量数据集 (worldwide household and industry transient energy dataset, WHITED)^[28]. 此数据集涵盖了全球各地家庭及小型工业环境中多种开关负荷单独运行时的电压和电流高频数据, 广泛应用于单一负荷辨识的模式识别算法的训练集和测试集, 包含了 47 种电器类型. 本研究选择空调、微波炉、吸尘器、冰箱和热水壶 5 种电器作为实验样本, 原因主要是: ① 日常家庭使用频率较高且能耗显著; ② 运行模式各

具特色, 热水壶为电阻型负荷空调和冰箱属于感性负载, 微波炉既有阻性负载又有感性负载特性, 吸尘器一般为感性负载, 部分高端型号也会表现出容性负载.

2.1.2 整体设计

首先, 依据波形所处状态从 WHITED 数据库中提取特定长度的电压和电流采样数据, 借鉴 FIT-PS 理论, 通过电压信号的过零点确立周期长度, 并选择适宜的周期点数, 对采样数据进行最近邻插值以获取固定采样点数的样本数据. 其次, 将样本数据输入 TimeGAN 模型, 为用电电器的每种状态合成 50 个新的多样的一维数据, 最终形成相应的 {噪声序列, 稳态序列, 非稳态序列} 数据集. 然后, 从前述数据库中随机选取与特定应用电器状态相对应

的一维时间序列, 依据电压过零点以及预先设定的周期点数, 对这些一维时间序列的长度进行剪裁与拼接, 并确保拼接处的相位一致性. 最后, 根据数据库中真实样本的波形场景时序将新生成的波形数据进行组合与叠加, 验证两种波形场景之间的匹配度, 包括: ① 验证TimeGAN模型合成的波形数据与真实样本集中的波形数据是否具有 consistency; ② 将波形数据进行拼接后, 分析生成数据的“负荷场景质量”.

2.2 评价指标

考虑到生成样本的品质、多样性及其与真实数据的接近程度, 以及所需验证的两组数据各自代表时序数据的短期与长期特征, 本研究选取决定系数、均方根误差、回归曲线优劣标准和角余弦系数来判断TimeGAN模型所合成的波形与真实样本集波形的一致性. 同时, 采用整个波形场景的总电量之比、峰值信噪比、结构相似性指标, 来评估生成数据的负荷场景质量.

2.2.1 混合模型性能的评价指标

1) 决定系数(R^2)常用作线性/非线性回归的拟合效果评估指标, 其计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \mu_{y_i})^2} \quad (5)$$

其中, N 为样本数; y_i 和 $\hat{y}_i$ 分别为样本真实值和合成值, 均为 N 维空间中的向量; μ_{y_i} 为 y_i 的均值. R^2 越接近 1, 代表真实值与合成值由自变量引起的变动相关程度越高.

2) 用均方根误差 (root mean squared error, RMSE) 和对称平均绝对百分比误差 (symmetric mean absolute percentage error, SMAPE) 衡量数据间的绝对误差, 二者表达式分别为

$$\sigma_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

$$\eta_{\text{SMAPE}} = \frac{1}{N} \sum_{i=1}^N \frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \times 100\% \quad (7)$$

3) y_i 与 $\hat{y}_i$ 的角余弦系数为

$$\delta_{\text{FR}} = \cos \theta = \frac{\sum_{i=1}^N y_i \cdot \hat{y}_i}{\sqrt{\sum_{i=1}^N (y_i)^2} \sqrt{\sum_{i=1}^N (\hat{y}_i)^2}} \quad (8)$$

y_i 与 $\hat{y}_i$ 越相似, θ 越小, 则认为混合模型的效

果越好.

2.2.2 负荷场景数据质量的评价指标

1) 负荷特征是确定负荷类别的依据, 也是实现NILM的关键因素. 本研究将运行过程中的有功功率进行累计得到电能之比 E_{Ratio} 作为生成数据的评价指标,

$$E_{\text{Ratio}} = \sum_{i=1}^N E_{y_i} / E_{\hat{y}_i} \quad (9)$$

其中, E_{y_i} 和 $E_{\hat{y}_i}$ 分别为真实波形和合成波形累计的能量.

2) 采用峰值信噪比 (peak signal-to-noise ratio, PSNR) 来衡量最大值信号和背景噪音之间的质量参考值, 其计算公式为

$$\zeta_{\text{PSNR}} = 10 \lg \left[\frac{\max(y_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \right] \quad (10)$$

当 $\zeta_{\text{PSNR}} \leq 20$ 时, 认为生成数据质量不可接受; 当 $20 < \zeta_{\text{PSNR}} \leq 30$ 时, 认为生成数据质量差; 当 $30 < \zeta_{\text{PSNR}} \leq 40$ 时, 表示生成数据质量是好的; 当 $\zeta_{\text{PSNR}} > 40$ 时, 说明生成数据质量极好.

3) 结构相似性 (structural similarity, SSIM) 是从结构和对比度等方面来衡量生成数据与真实数据之间的相似性, 取值范围为 $[0, 1]$, 取值越大代表越相似, 由亮度、对比度和结构 3 个因素的组合而得^[29], 计算公式为

$$\xi_{\text{SSIM}} = \frac{(2\mu_{y_i} \mu_{\hat{y}_i} + c_1)(2\sigma_{y_i \hat{y}_i} + c_2)}{(\mu_{y_i}^2 + \mu_{\hat{y}_i}^2 + c_1)(\sigma_{y_i}^2 + \sigma_{\hat{y}_i}^2 + c_2)} \quad (11)$$

其中, 均值 μ 为亮度估计; 协方差 σ 为结构相似性的度量; 标准差 σ^2 为对比度估计; c_1 和 c_2 是用来维持稳定的常数, 本研究取 $c_1 = 1 \times 10^{-4}$, $c_2 = 9 \times 10^{-4}$.

2.3 实验验证

2.3.1 单负荷波形仿真实验

基于每 20 ms 和每 1 s 的采样频率, 分别提取出非稳态和稳态特征的全周波电压和电流序列; 针对同一波形的不同状态, 分别随机抽取 50 个波形长度为 2 500 的时间序列矩阵, 并加入 $\mu = 0$ 、 $\sigma = 1$ 的正态分布随机矩阵, 其维度为 $2\,500 \times 50$. 在构建模型的过程中, 将时间序列的前 80% 作为训练集, 后 20% 作为测试集. TimeGAN 模型参数见表 1.

图 4 给出了从生成的微波炉波形数据中随机抽取的 1 组时间序列下的真实波形与生成波形的对比结果. 由图 4 可见, 无论是在稳态过程还是非稳态

表1 TimeGAN 的模型参数设置

Table 1 Parameters of the TimeGAN model

参数	取值	参数	取值
时间步长	50	权重系数 η	10
隐层维度	64	学习率	0.001
隐层层数	3	批处理维度	16
权重系数 λ	0.1	迭代次数	500

过程中, 真实波形与生成波形的变化趋势十分接近, 相似性高, 证明所提方法能够实现高效的波形生成. 另外还注意到, 一个完整的周期波形并非一条平滑的连续曲线, 而是有细小的波动, 且与实际波形相比, 生成波形的波动幅度更小, 噪声的干扰更小, 展现出更加平滑和整洁的特性.

表2列出了5种典型电器在真实与生成状态下稳态过程中和非稳态过程中的平均波形评价指标. 由表2可见, 5种电器的 R^2 和 δ_{FR} 值都大于0.98, 表明生成的波形与真实波形具有较高的吻合度; σ_{RMSE} 和 η_{SMAPE} 值普遍较小, 最大值为0.3515, 证明生成的波形普遍没有异常值, 且整体的误差也比较小, 表明基于FIT-PS-TimeGAN模型生成数据和构造场景是可行的.

FIT-PS-TimeGAN模型不仅能生成周期极短时

表2 真实和生成的平均波形的评价指标

Table 2 Evaluation metrics for real and generated average waveforms

典型设备	R^2		σ_{RMSE}		η_{SMAPE}		δ_{FR}	
	非稳态	稳态	非稳态	稳态	非稳态	稳态	非稳态	稳态
空调	0.974 6	0.999 2	0.199 5	0.196 9	0.227 5	0.199 3	0.987 2	0.999 6
微波炉	0.999 5	0.999 7	0.195 3	0.197 5	0.066 1	0.072 5	0.999 8	0.999 9
吸尘器	0.985 5	0.987 9	0.230 8	0.200 6	0.227 5	0.217 4	0.995 1	0.997 6
热水壶	0.996 3	0.996 6	0.195 4	0.211 5	0.351 5	0.123 1	0.998 2	0.998 3
冰箱	0.998 6	0.999 0	0.203 1	0.193 2	0.150 5	0.131 6	0.999 3	0.999 3

在不同的工作状态下, 电流波形的赋值和形态等呈现显著差异. 为生成新的、对应特定状态的波形数据, 本研究采用TimeGAN对各工作状态下的电流波形数据进行训练. 通过分析各状态的持续时间, 从相应状态的合成数据库中随机选取与该时间段匹配的波形数据进行融合, 进而构建出与实际运行状态及时间一致的波形场景, 并评估两种波形场景的一致性程度. 图5展示了微波炉与采用上述方法构造出的电流波形的原始电流波形的对比结果, 构造出的电流波形整体上表现得更为平滑. 图5(c)放大绘制了图5(a)和(b)矩形框处的电流波形, 该

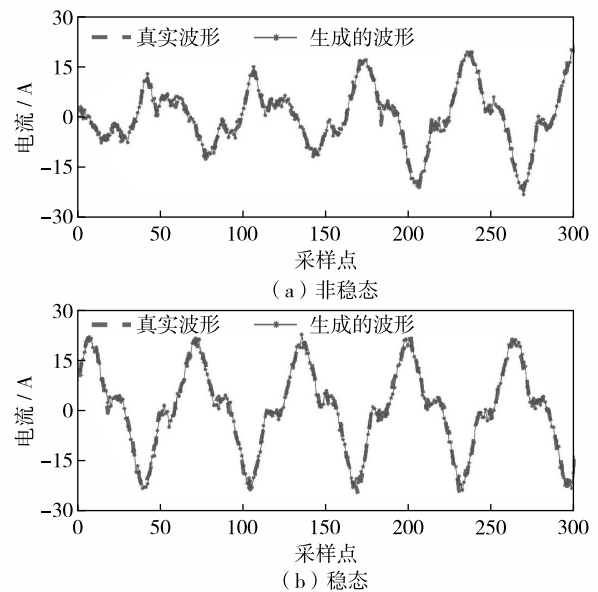


图4 微波炉在非稳态和稳态下的真实和生成的波形对比

Fig. 4 (Color online) Comparison of real and generated waveforms for the microwave under the condition of (a) non-steady state and (b) steady state. (The dashed line represents the actual waveform and the solid dots represent the generated waveform data.)

间跨度的波形数据, 以运用于简单开关负荷的数据生成, 还可以拟合生成长时间段的动态波形, 进一步对连续变化负荷进行波形数据生成.

结果与图4所示现象吻合. 可见FIT-PS-TimeGAN模型有效实现了对稳态与瞬态的分离和构造, 进而为电流波形的分析提供了新的视角和方法.

图5(a)给出了频率为50 Hz的微波炉实测电流波形图, 其中, 采样频率为38 400 Hz, 运行时长在6 s内. 电流波形可大致分为3个部分, 分别对应微波炉3种不同的工作状态. ① 待机状态(0 ~ 2.0 s). 此时电流成分主要为线路噪声, 干扰在一定程度上会使电流发生畸变, 降低负荷识别的准确度. ② 启动瞬间状态(2.0 ~ 2.6 s). 此时电流成分主要为脉冲电流. 启动和关闭是一个瞬时状态, 可

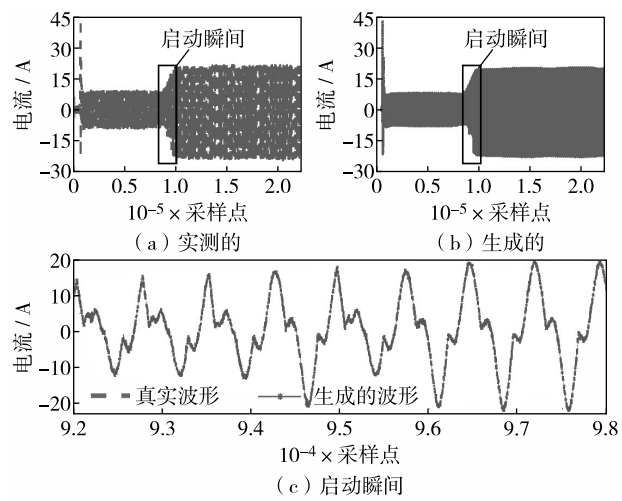


图5 微波炉在实测的、FIT-PS-TimeGAN混合模型生成的和启动瞬间状态下的电流波形

Fig. 5 (Color online) Current waveforms of the microwave in (a) real sampled, (b) FIT-PS-TimeGAN hybrid model generated and (c) start-up instant state. (The dashed line represents the actual waveform and the solid dots represent the generated waveform data.)

通过阈值或其他特征来识别. ③ 正常运行时的稳定状态. 稳态电流是识别典型电器类别的关键. 表 3 列出来了 5 种典型电器原始和构造的电流波形场景的评价指标结果.

由表 3 可见, 5 种电器的 E_{Ratio} 值都接近 1, 表明所构造的波形场景能够反映出真实波形的主要信息, 如有功率、低频和高频. 5 种电器的 ζ_{PSNR} 值都大于 30, 表明混合模型构造的生成数据和构造场景的质量都较好. 5 种电器的 ξ_{SSIM} 值都较接近 1, 证明了生成数据与真实数据之间在结构和对比度等方面是相似的. 综上所述, 基于 FIT-PS-TimeGAN 混合模型生成数据和构造场景的性能是良好的.

表3 不同电器真实和构造的电流波形场景的评价指标
Table 3 Evaluation metrics for real and constructed waveform scenes

典型设备	E_{Ratio}	ζ_{PSNR}	ξ_{SSIM}
空调	0.999 5	33.545 4	0.976 9
微波炉	0.999 5	32.928 9	0.877 0
吸尘器	0.999 8	32.421 4	0.888 0
热水壶	0.999 7	39.371 0	0.994 4
冰箱	1.000 6	33.235 3	0.929 7

此外, 为证明混合模型的生成效果, 采用自回归模型和 GAN 作为对比模型, 仍选择微波炉作为训练样本, 得到各模型平均波形评价指标如表 4.

表4 不同模型真实和生成的平均波形的评价指标对比
Table 4 Evaluation metrics of real and generated average waveforms for different models

模型	R^2	σ_{RMSE}	η_{SMAPE}	δ_{FR}
自回归	0.978 8	0.258 1	0.190 5	0.981 3
GAN	0.986 6	0.223 7	0.141 5	0.989 0
FIT-PS-TimeGAN	0.999 5	0.195 3	0.066 1	0.999 8

由表 4 可见, 采用自回归模型生成的数据的一致性小于 GAN 模型和 FIT-PS-TimeGAN 混合模型. 混合模型与 GAN 模型最大的差异在于后者常用于图像数据生成, 聚焦于数据的局部信息, 且后者信号与信号之间的拼接未进行很好的处理, 使得与前者相比结果较差. 而自回归模型是将联合概率拆成了条件概率累乘的形式, 难以捕捉时序中的长期依赖关系数据, 导致其性能不佳.

2.3.2 多负荷波形仿真实验

多负荷对应的状态较单负荷更多, 且增加了单负荷的叠加. 图 6 给出了频率为 50 Hz 的 5 种设备先后启动和关闭时的实测录波数据. 其中, 采样率为 6 400 Hz. 所有测试均在符合实际家庭环境并考虑不同采样率下的实验室内进行.

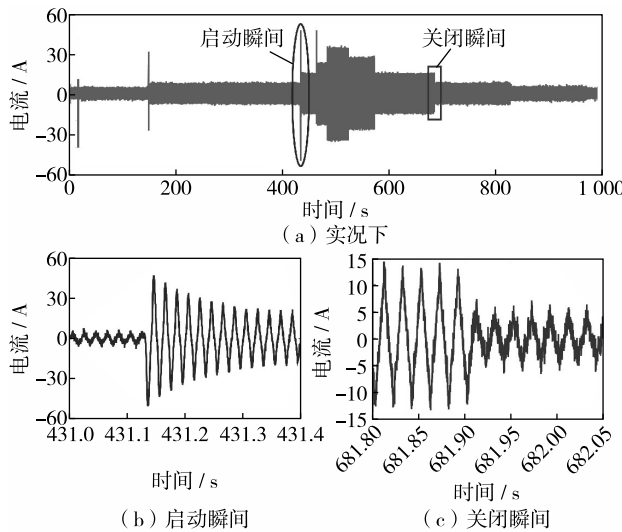


图6 实况下 5 种典型设备的电流波形与启动瞬间和关闭瞬间的吸尘器电流波形

Fig. 6 (Color online) (a) Current waveforms of five typical devices under real conditions. Current waveforms of the vacuum cleaner (b) at the moment of start-up and (c) at the moment of shut-down.

由图 6 可见, 在长达 989.94 s 的运行时间内, 时序电流波形可大致分为 5 个部分: ① 待机状态; ② 启动瞬间状态; ③ 正常运行时的稳定状态; ④

关闭瞬间状态; ⑤ 用电器逐一关闭后其他用电器正常运行时的稳定状态. 图 S1(请扫描论文末页右下角二维码查看)给出了以空调为例, 从 5 个状态下随机提取的 3 周波的电流波形.

由于构造的 5 个典型设备电流波形数据量过大, 限于篇幅, 本研究仅从整体上观测它们的波形状态, 不再进行可视化分析. 评价指标 $E_{\text{Ratio}} = 0.999\ 9$, $\zeta_{\text{PSNR}} = 39.153\ 7$, $\xi_{\text{SSIM}} = 0.997\ 5$, 表明构造的波形场景与真实场景相似的, 证明所提出混合模型的可行性和有效性.

结 语

提出一种新的数据合成方案——FIT-PS-TimeGAN 混合模型, 融合了 FIT-PS 与 TimeGAN 的优势. 应用 FIT-PS 提取并保留采样波形的关键特征, 利用 TimeGAN 创造出与训练样本数据分布类似但又具有一定差异性的新数据序列. 该混合模型有效地生成了高品质的非侵入式波形数据, 从根本上解决了数据获取的难题. 下一步, 可以考虑各种设备之间的异同点以及其他预处理手段, 以减少同一类型设备波形数据的生成, 从而降低混合模型的执行时间.

基金项目: 中国南方电网有限责任公司定向科技项目 (GZKJXM20222417); 广东省科技计划资助项目 (2021B1212050014)

作者简介: 肖勇(xiaoyong@csg.cn), 南方电网科学研究院有限责任公司正高级工程师. 研究方向: 电能计量.

引 文: 肖勇, 谈竹奎, 钱斌, 等. 基于混合模型的非侵入式负荷监测数据的生成[J]. 深圳大学学报理工版, 2025, 42(1): 85-93.

参考文献 / References:

- [1] IQBAL H K, MALIK F H, MUHAMMAD A, et al. A critical review of state-of-the-art non-intrusive load monitoring datasets [J]. *Electric Power Systems Research*, 2021, 192: 106921.
- [2] 栾文鹏, 韦尊, 刘博, 等. 非侵入式负荷监测算法的测试与评价方法[J]. *电网技术*, 2022, 46(11): 4568-4579.
LUAN Wenpeng, WEI Zun, LIU Bo, et al. Testing and performance evaluation of nonintrusive load monitoring algorithms [J]. *Power System Technology*, 2022, 46(11): 4568-4579. (in Chinese)
- [3] FRANCOU J, CALOGINE D, CHAU O, et al. Expanding variety of non-intrusive load monitoring training data: introducing and benchmarking a novel data augmentation technique [J]. *Sustainable Energy, Grids and Networks*,

2023, 35: 101142.

- [4] HELD P, WEIBHAAR D, ABDESLAM D O, et al. Generation of new simulation scenarios for NILM based on real data sets using high-resolution current waveforms [C]// *The 45th Annual Conference of the IEEE Industrial Electronics Society*. Piscataway, USA: IEEE, 2019: 5319-5324.
- [5] WU Haiping, LIU Hui. Non-intrusive load transient identification based on multivariate LSTM neural network and time series data augmentation [J]. *Sustainable Energy, Grids and Networks*, 2021, 27: 100490.
- [6] 刘睿迪, 汪震. 基于低频功率数据处理的负荷分解方法[J]. *能源工程*, 2021(6): 79-84.
LIU Ruidi, WANG Zhen. Load decomposition method based on low frequency power data processing [J]. *Energy Engineering*, 2021 (6): 79-84. (in Chinese)
- [7] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique [J]. *Journal of Artificial Intelligence Research*, 2002, 16(1): 321-357.
- [8] 裴星, 尹仕红, 张之涵, 等. 基于改进生成对抗网络的非侵入式负荷预测样本不均衡的改善方法[J]. *电器与能效管理技术*, 2022(3): 23-29, 38.
QIU Xing, YIN Shihong, ZHANG Zhihan, et al. Improving method of non-intrusive load forecasting sample imbalance based on improved generative confrontation network [J]. *Electrical & Energy Management Technology*, 2022 (3): 23-29, 38. (in Chinese)
- [9] 邢海青, 郭瑞峰, 杨浙川, 等. 基于数据扩充与无阈值递归图的非侵入式负荷识别方法[J]. *浙江电力*, 2024, 43(6): 88-100.
XING Haiqing, GUO Ruifeng, YANG Zhechuan, et al. A non-intrusive load identification method based on data augmentation and threshold-free recurrence plot [J]. *Zhejiang Electric Power*, 2024, 43(6): 88-100. (in Chinese)
- [10] 杨挺, 叶芷杉, 徐嘉成, 等. 基于低秩张量补全的非侵入式负荷监测缺失数据修复方法[J]. *电网技术*, 2024, 48(1): 394-404.
YANG Ting, YE Zhishan, XU Jiacheng, et al. Non-intrusive load monitoring missing data recovery based on low-rank tensor completion [J]. *Power System Technology*, 2024, 48(1): 394-404. (in Chinese)
- [11] 王谱宇, 丁一帆, 陈鉴祥, 等. 基于动态谐波导纳参数的非侵入式负荷监测数据模拟生成方法[J/OL]. *中国电机工程学报*. (2023-10-26) [2024-03-10]. <http://kns.cnki.net/kcms/detail/11.2107.TM.20231026.1452.012.html>.
- [12] KLEMENJAK C, KOVATSCH C, HEROLD M, et al. A

- synthetic energy dataset for non-intrusive load monitoring in households [J]. *Scientific Data*, 2020, 7(1): 108.
- [13] DELFOSSE A, HEBRAIL G, ZERROUG A. Deep learning applied to NILM: is data augmentation worth for energy disaggregation? [C]// ECAI 2020. Amsterdam, Netherlands: IOS Press, 2020: 2972-2977.
- [14] MEISER M, DUPPE B, ZINNIKUS I. Generation of meaningful synthetic sensor data: evaluated with a reliable transferability methodology [J]. *Energy and AI*, 2024, 15: 100308.
- [15] DONG Xianguang, GUO Liang, ZHANG Zhi, et al. Simulation data generation method based on improved cross-entropy algorithm for the prediction of user electricity consumption [C]// The 7th International Forum on Electrical Engineering and Automation (IFEEA). Piscataway, USA: IEEE, 2020: 943-948.
- [16] 李洋, 肖泽青, 聂松松, 等. 生成对抗网络及其在新能源数据质量中的应用研究综述[J]. *南方电网技术*, 2020, 14(2): 25-33.
Li Yang, XIAO Zeging, NIE Songsong, et al. Review of research on generative adversarial network and its application in new energy data quality [J]. *Southern Power System Technology*, 2020, 14(2): 25-33. (in Chinese)
- [17] 杨桐. 非侵入式电力负荷监测的数据生成方法研究[D]. 长春: 吉林大学, 2022.
YANG Tong. Research on data generation method of non-intrusive load monitoring [D]. Changchun: Jilin University, 2022. (in Chinese)
- [18] 薛展豪, 陈力, 林志颖, 等. 基于双向长短期记忆生成对抗网络的电力系统次同步振荡数据生成方法[J]. *智慧电力*, 2024, 52(5): 60-66.
XUE Zhanhao, CHEN Li, LIN Zhiying, et al. Subsynchronous oscillation data generation method of power system based on bidirectional long shortterm memory generative adversarial network [J]. *Smart Power*, 2024, 52(5): 60-66. (in Chinese)
- [19] 何恒靖, 王灏, 肖勇, 等. 基于Bi-LSTM算法的非侵入式负荷监测模型[J]. *南方电网技术*, 2019, 13(2): 20-26.
HE Hengjing, WANG Hao, XIAO Yong, et al. Non-intrusive load monitoring model based on Bi-LSTM algorithm [J]. *Southern Power System Technology*, 2019, 13(2): 20-26. (in Chinese)
- [20] 谈竹奎, 曾鸣. 基于负荷调节能力模型的实时需求响应理论[J]. *南方电网技术*, 2022, 16(1): 58-66.
TAN Zhukui, ZENG Ming. Real-time demand response theory based on load regulation capability model [J]. *Southern Power System Technology*, 2022, 16(1): 58-66. (in Chinese)
- [21] HELD P, LAASCH F, ABDESLAM D, et al. Frequency invariant transformation of periodic signals (FIT-PS) for signal representation in NILM [C]// The 42nd Annual Conference of the IEEE Industrial Electronics Society. Piscataway, USA: IEEE, 2016: 5149-5154.
- [22] HELD P, MAUCH S, SALEH A, et al. Frequency invariant transformation of periodic signals (FIT-PS) for classification in NILM [J]. *IEEE Transactions on Smart Grid*, 2019, 10(5): 5556-5563.
- [23] YOON J, JARRETT D, VAN DER SCHAAR M. Time-series generative adversarial networks [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., 2019: 5508-5518.
- [24] MAO Xudong, LI Qing, XIE Haoran, et al. Least squares generative adversarial networks [C]// IEEE International Conference on Computer Vision (ICCV). Piscataway, USA: IEEE, 2017: 2813-2821.
- [25] LIU Huidong, GU Xianfeng, SAMARAS D. Wasserstein GAN with quadratic transport cost [C]// IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 4831-4840.
- [26] 罗军帅, 蔡远利, 姜浩楠. 基于时间序列的生成对抗网络综述[C]// 第24届中国系统仿真技术及其应用学术年会 (CCSSTA24th 2023) 论文集. 中国安徽合肥: 中国科学技术大学, 2023: 116-120.
LUO Junshuai, CAI Yuanli, JIANG Haonan. Overview of generative adversarial networks based on time series [C]// Proceedings of the 24th China Conference on System Simulation Technology and Its Applications (CCSSTA24th 2023). Hefei, China: University of Science and Technology of China, 2023: 116-120. (in Chinese)
- [27] 贾梦瑶, 王玉玮, 宋明浩. 基于时间生成对抗网络的风电随机场景预测[J]. *智慧电力*, 2023, 51(8): 59-66.
JIA Mengyao, WANG Yuwei, SONG Minghao. Stochastic scenario prediction of wind power based on time generative adversarial networks [J]. *Smart Power*, 2023, 51(8): 59-66. (in Chinese)
- [28] KAHL M, HAQ A U, KRIECHBAUMER T, et al. WHITED: a worldwide household and industry transient energy data set [EB/OL]. [2023-10-15]. http://nilmworkshop.org/2016/proceedings/Poster_ID18.pdf.
- [29] 李正, 刘薇, 张凯兵. 基于机器学习的遥感图像超分辨率综述[J]. *计算机工程与应用*, 2021, 57(13): 8-17.
LI Zheng, LIU Wei, ZHANG Kaibing. Survey of remote sensing image super-resolution based on machine learning [J]. *Computer Engineering and Applications*, 2021, 57(13): 8-17. (in Chinese)

【中文责编: 英子; 英文责编: 木柯】



补充材料