计及历史数据缺失与负荷时间异质的超短期负荷预测

史瑞研，赵永宁，付坤明，刘静文
（中国农业大学信息与电气工程学院，北京市 100083）

摘　要：【目的】受限于实际运行条件，负荷预测往往面临数据质量低下与负荷分布变化的双重挑战。为此，提出一种计及历史数据缺失与负荷时间异质的超短期负荷预测方法。【方法】首先，为重建负荷等缺失时序数据，将卷积-双向长短期记忆神经网络（convolutional neural network-bidirectional long short-term memory，CNN-BiLSTM）嵌入生成对抗插补网络（generative adversarial imputation network，GAIN），捕捉时空依赖关系。训练过程中，为避免模型仅关注可观测值的插补误差，引入随机噪声替代部分可观测值，显式度量上述噪声的生成偏差。其次，为揭示更小时间尺度上的负荷异质性，对训练集预测模型的负荷输入样本进行聚类，识别不同负荷分布模式。最后，结合不同模式样本微调统一预测模型，构建个性化子模型。在线预测时，根据当下负荷输入样本与各模式中心的相似性动态选择子模型，避免对日历、天气等外部信息的依赖。【结果】算例表明，所提缺失数据重建方法较传统方法具有更低重建误差，在此基础上建立的负荷预测模型误差更低。引入个性化预测子模型构建及选取策略，可进一步降低预测误差。【结论】实验结果充分验证了所提方法在实际运行场景下的应用价值。

关键词：负荷预测；历史数据缺失；负荷时间异质；生成对抗；迁移学习
中图分类号： TM74　　**文献标志码：** A　　**文章编号：**1000-7229（2026）06-0123-14
DOI：10.12204/j.issn.1000-7229.2026.06.010

Ultra-Short-Term Load Forecasting Considering Historical Data Missing and Load Temporal Heterogeneity

SHI Ruiyan，ZHAO Yongning，FU Kunming，LIU Jingwen
（College of Information and Electrical Engineering，China Agricultural University，Beijing 100083，China）

ABSTRACT：[**Objective**] Constrained by practical operating conditions，load forecasting often faces the dual challenges of low data quality and variable load distribution. To address this，an ultra-short-term load forecasting method considering historical data missing and load temporal heterogeneity is proposed in this paper. [**Methods**] First，to reconstruct missing time-series data such as load，a Convolutional Neural Network-Bidirectional Long Short-Term Memory（CNN-BiLSTM）neural network is embedded within the Generative Adversarial Imputation Network（GAIN）to capture the spatiotemporal dependencies. During training，to prevent the model from focusing solely on the imputation error of observable values，random noise is introduced to replace partial observable values，enabling explicit measurement of the generation bias corresponding to the aforementioned noise. Second，to reveal load heterogeneity at finer temporal scales，clustering is applied to load input samples for the forecasting model in the training set to identify different load distribution patterns. Finally，a unified forecasting model is fine-tuned according to different pattern-specific samples，constructing personalized sub-models. During online forecasting，the most suitable sub-model is dynamically selected based on the similarity between the current load input sample and the centers of each pattern，eliminating reliance on external information such as calendar and weather data. [**Results**] Case studies demonstrate that the proposed missing data reconstruction method achieves lower reconstruction errors compared with traditional methods. Based on this，the constructed forecasting model yields higher prediction accuracy. Incorporating the construction and selection strategy for personalized forecasting sub-models further decreases forecasting errors. [**Conclusions**] Experimental results confirm the practical engineering value of the proposed method in real-world operational scenarios.

基金项目：国家自然科学基金项目（52207144）

© Editorial Department of Electric Power Construction. OA under CC BY-NC-ND 4.0.

This work is supported by the National Natural Science Foundation of China (No. 52207144).

KEYWORDS: load forecasting; missing historical data; load temporal heterogeneity; generative adversarial; transfer learning

0 引 言

在“双碳”目标驱动下,构建以新能源为主体的新型电力系统成为能源体系转型的关键[1-2]。智能测量装置的广泛应用为负荷预测提供了海量数据支撑,但数据采集与传输过程中的故障或干扰导致数据缺失问题突出[3]。同时,分布式电源和电动汽车等柔性负荷接入加剧了负荷特性的复杂性,显著提升了负荷预测难度[4-5]。超短期负荷预测结果对系统快速响应与实时调度尤为关键[6]。

现有负荷预测方法中,统计模型直观高效但难以处理非线性关系[7-8],机器学习方法具备更强的非线性建模能力与鲁棒性,在负荷预测中表现优异[9-11]。然而,其精度高度依赖数据质量,实际中普遍存在的数据缺失问题会削弱预测可靠性,因此,有效的缺失数据处理方法对提升预测性能至关重要[12]。

现有缺失数据重建方法主要分为统计方法与人工智能方法。统计方法对数据的线性和平稳性要求较高,难以应对大规模时序插补[13-14]。机器学习方法容易忽略数据的时序相关性,难以处理连续数据缺失[15-16]。深度学习方法则通过非线性建模捕捉时序规律,实现缺失数据重建。例如,文献[17]结合多头注意力与图注意力网络实现风电功率缺失数据的高精度填充;文献[18]结合液态神经网络与长短期记忆神经网络提升插补精度。需要指出的是,在时间序列插补中,除时序特性外,关联特征中蕴含的有价值信息同样至关重要。

此外,变分自编码器(variational autoencoder, VAE)[19]、生成矩匹配网络(generative moment matching network, GMMN)[20]和扩散模型[21]等生成式模型也被广泛应用于数据修复领域。但VAE修复结果偏向全局平均,无法还原真实数据的高保真细节;GMMN训练稳定但表达能力较弱,难以精准恢复局部时序和特征关联结构;扩散模型生成精度高但效率低,缺乏实用性。

相比之下,生成对抗网络(generative adversarial network, GAN)在缺失数据修复中表现优异。文献[22]结合GAN与双重语义感知约束,提高了配电网缺失量测数据的插补精度。文献[23]提出基于GAN的堆叠自适应中值迭代插补框架,通过均值递归更新来填补工业生产过程中缺失的时间序列数据。此外,GAN也被应用于电动汽车充电负荷数据[24]、PMU量测数据[25]和智能交通数据[26]、核磁共振数据[27]等的修复。

然而,传统GAN插补方法存在两大缺陷:其一,生成器过度依赖随机噪声,难以充分利用已有数据信息;其二,判别器依赖完整样本训练,但数据缺失场景中往往缺乏连续完整数据。生成对抗插补网络(generative adversarial imputation network, GAIN)通过引入掩码矩阵与提示矩阵克服了上述不足[28],但其在建模训练过程中仅以可观测值的重建误差作为优化目标,难以约束随机噪声对应的生成偏差,限制了其在复杂缺失场景下的重建效果。

影响负荷预测精度的另一关键因素是负荷数据的时间异质性。文献[29]分析了负荷波动特性,并结合节假日、气温等信息判断待预测日类型,进而匹配相应预测模型与特征集,提高了预测精度。文献[30]利用历史用电量识别负荷模式并建立相应模型,预测时结合周、月及节假日等特征选取对应模型。文献[31]利用季节间负荷差异显著、同季节负荷特性相似的规律,提出了新型季节划分方法并构建对应模型,实现日前负荷预测。除此之外,多项研究亦采用“针对不同负荷特性分别构建预测模型以提升预测精度”的研究思路[32-33]。

上述方法在一定程度上提升了预测精度,但其大多基于日、月等较大时间尺度进行负荷模式划分,忽视了小时及以下尺度上的负荷动态差异。此外,在线预测常依赖于日历、气象等外部信息判定负荷类型,而非结合实时负荷特性选取预测模型,在一定程度上依赖人为经验,降低了模型适应性。

负荷模式精细划分后,各子模型可用训练样本减少,同样会对预测精度产生不利影响。迁移学习通过重用预训练模型,有效缓解了上述问题。例如,文献[34]在完整数据集上训练源域模型后,结合特定集群用户行为进行微调,通过迁移学习实现知识转移,从而提升了数据稀缺环境下的预测精度。

综上,针对现有研究在缺失数据重建和负荷时间异质建模方面的不足,本文主要创新点与贡献包括:

1)提出基于噪声正则化的生成对抗插补网络(noise-regularized generative adversarial imputation network, NR-GAIN)。引入随机噪声替代部分可观测

值进行插补模型训练，显式度量上述噪声对应的生成偏差，突破传统方法仅以可观测值的重建误差作为优化目标的局限。

2）分析更小时间尺度上的负荷数据异质性。应用滑动窗口策略和聚类分析算法，在由数据集时间分辨率所决定的最小时间尺度上，有效识别了不同负荷分布模式并提取了对应模式中心，刻画了负荷的短时动态差异。

3）提出一种新颖的个性化预测子模型构建与选取策略。结合不同模式样本微调统一预测模型构建个性化子模型，在线预测时依据当前负荷输入样本与各模式中心的相似性选择对应子模型，而无需依赖日历、天气等外部信息。

1　本文所提超短期负荷预测方法

1.1　基于NR-GAIN的缺失数据插补方法

传统方法通过最小化可观测域上的重建误差进行模型训练，容易出现过拟合可观测样本的问题，且无法对缺失值插补结果进行有效约束。为此，NR-GAIN在训练阶段随机替换部分可观测值为噪声输入，由于该部分随机噪声对应位置上的真实值可以观测，因此在设计模型训练的损失函数时，可以对上述噪声对应的生成偏差进行有效量化，避免模型仅关注可观测值的插补误差。

本文针对缺失数据插补问题，重点在于对GAIN的训练过程进行改进。而框架内部生成器与判别器的基本结构，可以替换为其他模型结构，仅需保证其具备时空特征提取能力。结合卷积神经网络-双向长短期记忆网络（convolutional neural network-bidirectional long short-term memory，CNN-BiLSTM）在捕捉局部时空相关性方面的良好平衡与结构简洁性，本文将其嵌入GAIN框架。

采用滑动时间窗口对原始序列进行分段，并以多通道形式输入多特征数据，来充分利用多维信息。NR-GAIN的训练流程如图1所示。

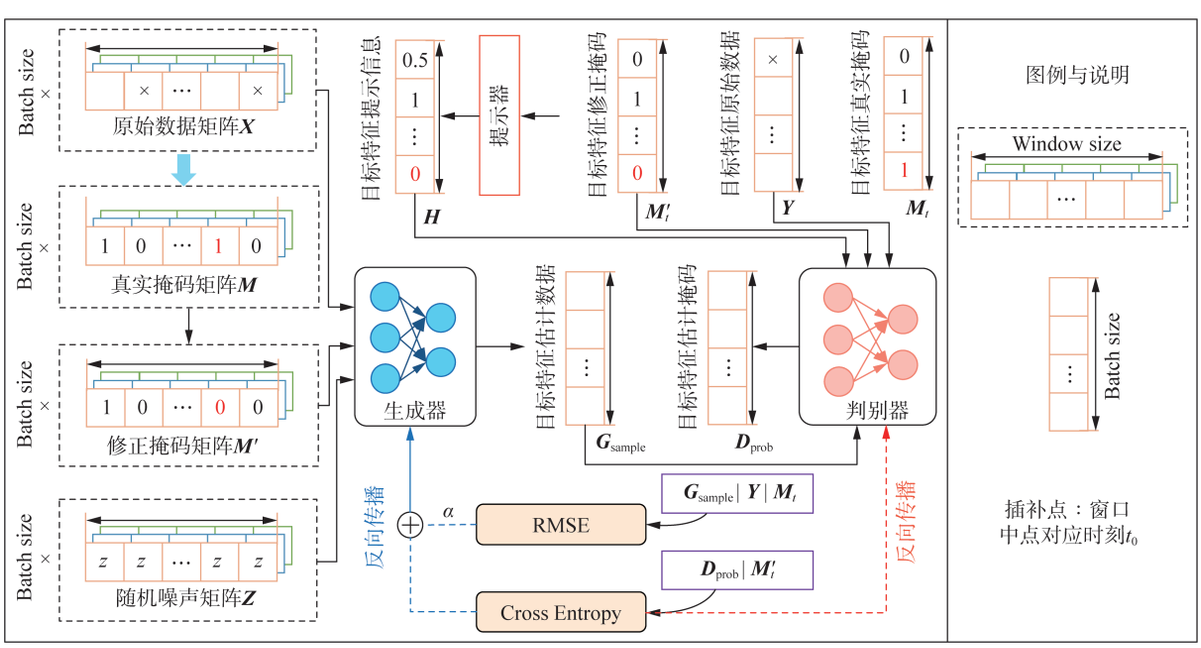


图1　NR-GAIN的训练流程框架

Fig 1　Framework of the NR-GAIN training work flow

1）根据原始不完整数据构造原始数据矩阵$\boldsymbol{X}$及其对应的真实掩码矩阵$\boldsymbol{M}$。矩阵$\boldsymbol{X}$与$\boldsymbol{M}$中各元素x_{ij}与m_{ij}的设定分别如式(1)和式(2)所示。

$$x_{ij}=\begin{cases}x_{ij}, & \text{观测数据}\\ 0, & \text{缺失数据}\end{cases} \tag{1}$$

$$m_{ij}=\begin{cases}1, & x_{ij}=x_{ij}\\ 0, & x_{ij}=0\end{cases} \tag{2}$$

2）调整真实掩码矩阵$\boldsymbol{M}$得修正掩码矩阵$\boldsymbol{M}'$。其元素m'_{ij}的定义如式(3)所示。

$$m'_{ij}=\begin{cases}m_{ij}, & x_{ij}\text{为其他关联特征数据}\\ 1, & x_{ij}\text{为目标特征观测数据且}r\geqslant p\\ 0, & x_{ij}\text{为目标特征观测数据且}r<p\\ 0, & x_{ij}\text{为目标特征缺失数据}\end{cases} \tag{3}$$

式中：超参数p表示将目标特征可观测值替换为随机噪声的比例；$r\sim U(0,1)$。

3）设置目标特征提示信息H，其定义如式(4)所示。

$$\boldsymbol{H}=\boldsymbol{M}'_t\odot\boldsymbol{B}+0.5(1-\boldsymbol{B}) \tag{4}$$

式中：$\boldsymbol{M}_t'$为目标特征修正掩码向量；$\boldsymbol{B}$为随机生成的0-1二值向量，形状与$\boldsymbol{M}_t'$相同，其值为1代表需要提示信息，1占总元素的比例通过超参数β来控制。

4）定义随机噪声矩阵$\boldsymbol{Z}$，其形状与原始数据矩阵$\boldsymbol{X}$相同且元素$z_{ij}\sim U(0,0.01)$。

5）网络核心式为：

$$\boldsymbol{G}_{\text{sample}}=G\left[\boldsymbol{M}'\odot\boldsymbol{X}+(1-\boldsymbol{M}')\odot\boldsymbol{Z},\boldsymbol{M}'\right] \tag{5}$$

$$\hat{\boldsymbol{Y}}=\boldsymbol{M}_t'\odot\boldsymbol{Y}+(1-\boldsymbol{M}_t')\odot\boldsymbol{G}_{\text{sample}} \tag{6}$$

$$D_{\text{prob}}=D\left(\hat{\boldsymbol{Y}},\boldsymbol{H}\right) \tag{7}$$

式中：$\boldsymbol{G}_{\text{sample}}$为生成器生成的目标特征估计数据；$\boldsymbol{Y}$为目标特征原始数据；$\hat{\boldsymbol{Y}}$为目标特征插补数据；$D_{\text{prob}}$为目标特征估计掩码。

6）定义网络损失函数G_{loss}和D_{loss}：

$$G_{\text{loss}}=\frac{\alpha\cdot(\boldsymbol{M}_t\odot\boldsymbol{Y}-\boldsymbol{M}_t\odot\boldsymbol{G}_{\text{sample}})^2}{\text{sum}(\boldsymbol{M}_t)}-\text{mean}\left[(1-\boldsymbol{M}_t')\odot\log D_{\text{prob}}\right] \tag{8}$$

$$D_{\text{loss}}=-\text{mean}\left[\boldsymbol{M}_t'\odot\log D_{\text{prob}}+(1-\boldsymbol{M}_t')\odot\log(1-D_{\text{prob}})\right] \tag{9}$$

式中：$\boldsymbol{M}_t$为目标特征真实掩码；超参数α用于平衡生成器部分的**RMSE**和Cross Entropy损失；mean(·)为求平均值函数；sum(·)为求和函数。

上述步骤最终可实现模型迭代训练优化。

为充分有效利用双向长短期记忆（bidirectional long-short term memory，BiLSTM）网络的双向长时记忆能力，生成器对各窗口中间时刻t_0的目标特征进行填充；当t_0接近序列末端且未来信息不可得时，视其后续数据为完全缺失。为借助批次维度的时序相关性辅助判别器对时序结构进行建模判别，训练时不打乱样本输入顺序。

考虑到不同特征的分布差异，针对各特征单独训练模型以提高重建精度。值得一提的是，在实际重建中，只需结合真实的掩码矩阵，将可观测数据和随机噪声共同输入生成器，完成缺失数据重建。

1.2 基于*K*-medoids的小时间尺度负荷数据异质性分析方法

选用*K*-Medoids对训练集上滑动窗口内的历史负荷输入样本进行聚类，以识别由数据集时间分辨率所决定的最小时间尺度上的负荷数据异质性，即输入样本间的显著分布差异。鉴于本文所用数据集的采样时间间隔为30 min，因此是在0.5 h时间尺度上进行负荷时间异质性分析。此外，为更好地识别历史负荷输入的形态特征，选用余弦相似度作为相似性度量指标[35]。

设$L_i=[l_{i1},l_{i2},\cdots,l_{im}]$与$L_j=[l_{j1},l_{j2},\cdots,l_{jm}]$分别为滑动窗口$i$和$j$内的历史负荷样本，窗口长度为$m$，则两者间的余弦相似度$S(L_i,L_j)$如式(10)所示。

$$S(L_i,L_j)=\frac{L_iL_j}{\left|L_i\right|\left|L_j\right|}=\frac{\sum_{n=1}^{m}l_{in}l_{jn}}{\sqrt{\sum_{n=1}^{m}l_{in}^2}\sqrt{\sum_{n=1}^{m}l_{jn}^2}} \tag{10}$$

式中：$S(L_i,L_j)$取值范围为[0,1]，其值越接近于1，说明L_i和L_j所表征的负荷样本分布越相近。

采用轮廓系数$s(i)$来确定最佳聚类数K。

$$s(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{11}$$

式中：$a(i)$为样本i与其所在簇内其他样本的平均余弦距离；$b(i)$则是将样本i与每个非所属簇内其他样本的平均余弦距离进行比较后得到的最小值；$s(i)$取值范围为[−1,1]，其值越趋于1表明聚类效果越好。

根据最优聚类数目，将预测模型的负荷输入样本划分为K类，每一类样本群体对应一种负荷分布模式，将各群体的中心点视为相应的负荷分布模式中心。

1.3 个性化预测子模型构建与选取策略

通过迁移微调统一预测模型的方式来获得不同负荷分布模式对应的个性化子模型，以兼顾负荷数据异质性并缓解训练样本不足问题。

在线预测的过程中，结合当下真实的负荷特性来选取最恰当的预测子模型。具体来说，实时计算负荷输入样本与各分布模式中心间的余弦相似度，确定其所属分布模式，并选择对应子模型进行预测。摆脱了对日历、天气等外部信息的依赖，避免主观影响。

本文所提个性化预测子模型构建和选取策略的具体实现过程如表1所示。

1.4 本文整体预测流程框架

本文整体预测流程如图2所示。首先，根据具体的不完整历史数据集，获取其对应的真实掩码、修正掩码等相关信息并定义随机噪声，利用NR-GAIN模型实现缺失数据重建。其次，利用完整训练集训练统一预测模型，并对预测模型的负荷输入样本进行聚类，以识别不同负荷分布模式。随后，结合不同分布模式的子训练数据样本，微调统一预测模型，构建适应特定模式的个性化子模型。在线预测阶段，根据当下负荷输入样本与各模式中心的相似性动态选择子模型，实现超短期电力负荷预测。

2 模型结构及其超参数设置

使用Python语言并结合Pytorch深度学习环境进行程序编写，训练模型的PC配置如下：17th Gen Intel(R) Core（TM）i7-14700F 2.10 GHz CPU、16 GB RAM、

表 1　个性化预测子模型构建和选取策略的具体实现过程
Table 1　Specific implementation process of personalized forecasting sub-models construction and selection strategy

Input：训练集 D；各负荷分布模式对应的子训练集$\{D_k, k=1, 2, \cdots\cdots K\}$；各负荷分布模式中心$\{C_k, k=1, 2, \cdots\cdots K\}$；统一预测模型训练参数：$E_u$, η_u, B_u；微调模型训练参数：E_f, η_f, B_f
Output：各负荷分布模式对应的个性化预测子模型参数$\{W_k, k=1, 2, \cdots\cdots K\}$；预测结果 $\hat{y}_t$

Function Unified model training (D, E_u, η_u, B_u):
  Initialize $w = (w_b, w_l)$
  for epoch = 1 …… E_u **do**
    for minibatch $(x, y) \in D$ with size B_u **do**
      $\hat{y} \leftarrow f(x; w_b, w_l)$
      $w \leftarrow w - \eta_u * \nabla w$ Loss $(\hat{y}, y)$ // update Backbone + Linear
    End for
  End for
Return w
Function Fine tuning (D_k, ${w_b}^0$, ${w_l}^0$, E_f, η_f, B_f):
  $w_b = {w_b}^0$ (frozen), $w_l = \text{copy}({w_l}^0)$
  for epoch = 1……E_f **do**
    for minibatch $(x, y) \in D_k$ with size B_f **do**
      $\hat{y} \leftarrow f(x; w_b, w_l)$
      $w_l \leftarrow w_l - \eta_f * \nabla w_l$ Loss $(\hat{y}, y)$ // only update Linear
    End for
  End for
Return (w_b, w_l)
Function Online prediction (x_t, C_k, w_k}):
  $s_k = (x_t \cdot C_k) / (\|x_t\| \|C_k\|)$ for k=1……K
  $k^* = k$ (argmax s_k)
  $\hat{y}_t = f(x_t; w_k^*)$
Return $\hat{y}_t$
Procedure Main (D, D_k, C_k):
  w^0 = Unified model training (D)
  for k = 1……K **do**
    w_k = Fine tuning (D_k, ${w^0}_{wb}$, ${w^0}_{wl}$)
  End for
Return $\{w_k, k=1, 2, \cdots\cdots K\}$, Online prediction
End Procedure Main

NVIDIA GeForce RTX 4060Ti GPU。

2.1　插补模型 NR-GAIN

NR-GAIN 生成器和判别器均由 5 层 1D 卷积层、1 层 BiLSTM 层和 4 层线性层构成，详细结构见表 2。

经多次预实验调优，设置超参数 β 为 0.8，超参数 α 为 2000。此外，当缺失模式为随机缺失与混合缺失时，设置超参数 p 为 0.2，当缺失模式为连续缺失时，设置超参数 p 为 0.4。设置滑动窗口大小为 49，批大小为 256，不打乱样本顺序。采用 Adam 优化器，学习率为 0.000 1，迭代训练 500 轮。

2.2　预测模型 Transformer

Transformer 基于自注意力机制，具备强大的非线性时序特征提取能力。因此，采用 Transformer 编码器部分构建预测模型，其核心结构及详细计算原理可参见文献[36]。

经多次预实验调优，设置滑动窗口大小为 96，预测未来 0~4 h 的负荷，批大小为 128，打乱样本顺序。

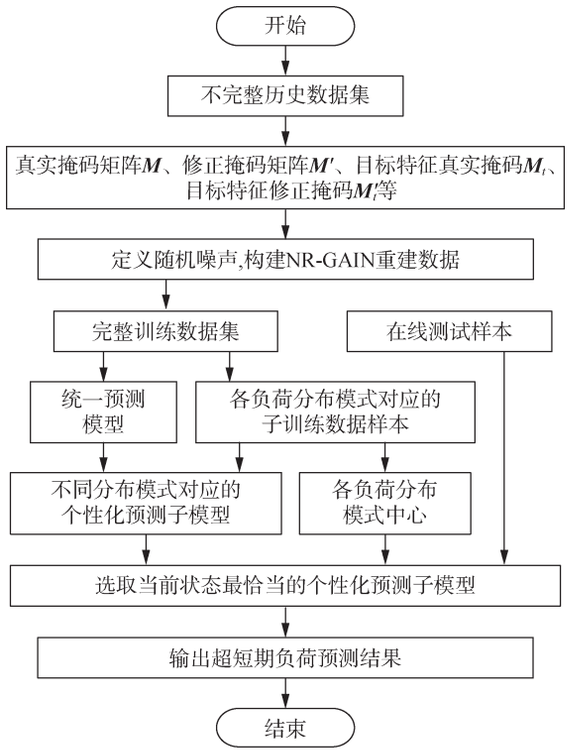


图 2　本文方法的整体预测流程框架

Fig. 2　Overall framework of the method in this paper

表 2　NR-GAIN 网络结构
Table 2　NE-GAIN network structure

名称	参数	生成器部分	判别器部分
5层 1D卷积层	卷积核大小	[1,3,3,3,3]	[1,3,3,3,3]
	滤波器数量	[128,64,32,16,8]	[16,32,64,128,256]
	步长	[1,1,1,1,1]	[1,1,1,1,1]
	零填充	[0,1,1,1,1]	[0,1,1,1,1]
	激活函数	Relu	Relu
BiLSTM层	hidden_size	64	256
前3层全连接层	神经元个数	[64,16,8]	[256,128,64]
	激活函数	Relu	LeakyRelu(0.2)
最后1层全连接层	神经元个数	1	1
	激活函数		Sigmoid

注意力头数为 4、嵌入维度为 32、前馈网络神经元数为 64、编码器模块数为 3、线性层层神经元个数依次为 32×8、32、8。采用 Adam 优化器，统一预测模型和微调模型学习率分别为 0.000 1 和 0.000 05，训练轮数为 200 轮与 50 轮。

3　算例分析

3.1　数据集及其预处理

选用 2008 年 1 月 1 日至 2010 年 6 月 30 日澳大利亚某地区的多维负荷数据。数据集包含干球温度、露点温度、湿球温度、湿度和电力负荷五个特征。采

样间隔为30 min。数据集前80%为训练集，后20%为测试集。为消除量纲差异影响，数据按式(12)进行归一化处理。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

式中：x'为归一化值；x为原始数据；$x_{\max}$为特征序列最大值；$x_{\min}$为特征序列最小值。

3.2 评价指标

3.2.1 缺失数据重建效果评价指标

选取均方根误差(root mean squared error，RMSE)L_{RMSE}、平均误差(mean error，ME)L_{ME}和平均误差率(mean error rate，MER)L_{MER}作为缺失数据重建效果评价指标[37]，三者的具体计算式如下：

$$L_{\mathrm{RMSE}} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(g_i - g_i')^2} \tag{13}$$

$$L_{\mathrm{ME}} = \frac{1}{I}\sum_{i=1}^{I}|g_i - g_i'| \tag{14}$$

$$L_{\mathrm{MER}} = \frac{1}{I}\sum_{i=1}^{I}\frac{|g_i - g_i'|}{g_i} \times 100\% \tag{15}$$

式中：g_i为真实数据；g_i'为重构数据；I为缺失数据的数量。L_{RMSE}、L_{ME}和L_{MER}越小，重构效果越好。

3.2.2 预测效果评价指标

选取均方根误差(root mean squared error，RMSE)e_{RMSE}和平均绝对误差(mean absolute error，MAE)e_{MAE}作为负荷预测效果评价指标[38]，二者的计算式如下：

$$e_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(p_t - \hat{p}_t\right)^2} \tag{16}$$

$$e_{\mathrm{MAE}} = \frac{1}{n}\sum_{t=1}^{n}|p_t - \hat{p}_t| \tag{17}$$

式中：p_t和$\hat{p}_t$分别为t时刻的负荷实测值与预测值；n为预测段的样本量。e_{RMSE}和e_{MAE}越小，预测效果越好。

3.3 实验结果

3.3.1 不同缺失数据重建方法的重构效果对比

为验证所提缺失数据重建方法的优越性，本文在训练集各特征数据上模拟随机缺失、连续缺失以及混合缺失三种数据缺失模式，缺失率设置为50%。后续预测实验在没有特殊说明的情况下，均是在混合缺失50%的情况下进行的。

选取多种典型方法包括：三次样条插值(R1)、(K-nearest neighbor，KNN)(R2)、考虑了关联特征的CNN-BiLSTM(R3)、未考虑关联特征的GAIN(R4)、考虑了关联特征的GAIN(R5)、未考虑关联特征的NR-GAIN(R6)与考虑了关联特征的NR-GAIN，即本文方法(RP)进行对比。

在基于深度学习的插补方法(R3—RP)中，均采用滑动窗口对原始序列进行分段处理。对于气象类特征的重建，关联特征指除当前插补特征之外的其余三项气象特征；而在电力负荷特征的重建中，关联特征还包括小时特征。此外，所有基于GAN的方法生成器与判别器均采用CNN-BiLSTM结构。

表3对比了50%缺失率下不同方法的重构误差。根据表3可知，在任何缺失模式下，RP对各个特征的重构误差均为最低，表明本文方法能够最大程度地重建缺失时序数据，还原原始数据的信息空间。值得注意的是，三种缺失模式下R1对气象特征的重构效果仅次于RP，但在负荷数据上的表现明显较差。主要是由于气象数据波动较小，较好地满足了统计方法对数据线性与平稳性的假设；而负荷受多因素影响，非线性强波动大，传统方法难以应对其大规模缺失。

相同缺失模式下，RP与R5、R6与R4之间的对比，验证了本文所提NR-GAIN相较原有GAIN的优越性。对比RP与R6、R5与R4发现，在捕捉时序特性的基础上同时挖掘关联特征信息可进一步提升重建精度。对比R3与R5发现，仅引入对抗学习机制而不改进训练策略，对重建效果影响有限，部分特征重建效果甚至下降。此外，除连续缺失模式下的湿度特征外，R4在其他情况下的表现均为最差，说明直接将GAIN用于时序数据缺失重建并不可行。

特别注意，本文方法针对连续缺失模式的重构误差最大、随机缺失模式次之、混合缺失模式下的重构误差最小。这是由于连续缺失，长时间段的观测信息中断，破坏了时序依赖结构，使生成器难以利用上下文特征实现有效补全；随机缺失有效保留了局部时序信息，重构误差较小；混合缺失误差最小则是因为相较前两种模式，其在训练过程中为模型提供了最丰富多样的缺失样本模式，使生成器在学习过程中获得了更强的分布泛化能力。

为直观展示真实值与插补值之间的分布差异，在连续缺失50%模式下，以气象特征中的湿度特征与电力负荷特征为例，结合真实数据、本文方法对应的插补数据与仅次于本文方法的对比方法对应的插补数据绘制概率密度图，如图3所示。

从图3可见，RP对应插补数据的分布曲线与真实数据高度吻合，分布平滑性与真实数据最为接近，表明其不仅能够有效降低数值重建误差，还能保持原始数据的统计结构特征，从而为后续预测建模提供更具可靠性的样本数据。

表 3　50% 缺失比例下不同方法对应的重构误差

Table 3　Reconstruction errors corresponding to different methods under 50% missing ratio

特征名称	重构方式	RMSE			ME			MER		
		连续缺失 50%	随机缺失 50%	混合缺失 50%	连续缺失 50%	随机缺失 50%	混合缺失 50%	连续缺失 50%	随机缺失 50%	混合缺失 50%
干球温度	R1	0.418	0.367	0.358	0.272	0.208	0.208	1.571	1.195	1.204
	R2	1.176	1.088	1.017	0.778	0.719	0.669	4.605	4.206	3.875
	R3	0.668	0.677	0.520	0.454	0.389	0.322	2.819	2.402	1.852
	R4	1.276	1.297	1.142	0.976	0.820	0.651	6.312	5.061	3.856
	R5	0.656	0.620	0.500	0.452	0.361	0.280	2.657	2.110	1.636
	R6	0.496	0.425	0.354	0.338	0.272	0.217	1.986	1.580	1.261
	RP	**0.378**	**0.277**	**0.248**	**0.266**	**0.176**	**0.163**	**1.562**	**1.015**	**0.938**
露点温度	R1	0.532	0.480	0.460	0.331	0.258	0.261	4.155	3.656	3.623
	R2	1.375	1.252	1.275	0.914	0.842	0.806	12.556	10.974	10.059
	R3	0.885	1.018	0.828	0.637	0.648	0.482	6.400	8.275	6.651
	R4	0.967	1.399	1.698	0.726	1.004	0.917	8.745	13.443	11.537
	R5	1.167	1.019	0.918	0.924	0.651	0.544	8.195	8.032	6.810
	R6	0.616	0.520	0.500	0.403	0.300	0.314	5.219	4.329	4.347
	RP	**0.431**	**0.371**	**0.357**	**0.283**	**0.235**	**0.224**	**3.440**	**3.350**	**3.067**
湿球温度	R1	0.254	0.224	0.227	0.175	0.136	0.137	1.305	1.016	1.021
	R2	0.744	0.681	0.666	0.522	0.480	0.442	4.037	3.680	3.409
	R3	0.397	0.386	0.429	0.281	0.221	0.276	2.102	1.595	2.321
	R4	0.859	0.988	0.993	0.673	0.712	0.471	5.102	5.404	3.990
	R5	0.369	0.337	0.458	0.264	0.200	0.282	1.997	1.458	2.380
	R6	0.305	0.255	0.256	0.219	0.171	0.165	1.640	1.290	1.317
	RP	**0.225**	**0.189**	**0.168**	**0.165**	**0.134**	**0.113**	**1.233**	**1.028**	**0.845**
湿度	R1	2.835	2.420	2.478	1.903	1.466	1.447	3.193	2.449	2.393
	R2	6.708	6.058	5.805	4.705	4.342	4.038	8.475	7.507	7.009
	R3	11.357	4.552	5.089	8.262	2.921	3.468	15.998	4.916	6.233
	R4	7.163	5.658	8.504	5.487	3.564	5.507	10.678	6.504	10.079
	R5	12.730	5.357	5.410	9.377	3.429	3.352	18.216	6.102	5.050
	R6	3.278	2.678	2.601	2.351	1.752	1.579	3.951	2.905	2.615
	RP	**2.389**	**2.049**	**2.021**	**1.720**	**1.384**	**1.346**	**2.817**	**2.263**	**2.253**
电力负荷	R1	200.005	177.474	193.782	140.982	110.664	114.869	1.637	1.280	1.326
	R2	231.384	196.303	180.379	149.875	132.687	118.422	1.678	1.511	1.337
	R3	210.010	202.263	320.555	126.842	105.369	194.215	1.415	1.195	2.164
	R4	593.771	426.959	614.028	449.013	256.700	384.150	5.240	2.971	4.654
	R5	231.256	204.094	319.129	143.510	109.124	181.626	1.655	1.214	2.072
	R6	158.046	130.371	123.814	113.152	86.184	80.267	1.272	0.969	0.887
	RP	**129.879**	**109.523**	**112.670**	**95.344**	**75.498**	**75.357**	**1.073**	**0.851**	**0.843**

此外，在混合缺失 50% 模式下，以电力负荷为例，随机选取连续时段进行插补结果可视化，如图 4 所示。可见，RP 所得到的插补曲线与真实曲线的拟合度最高，整体表现明显优于其他对比方法。

除 50% 缺失率外，本文还分别在随机缺失和连续缺失两种模式下，设置 10% 与 30% 的缺失率进行实验，不同方法对应的重建误差如表 4 所示。由表 4 可见，RP 方法在上述各实验中针对各特征数据的重建

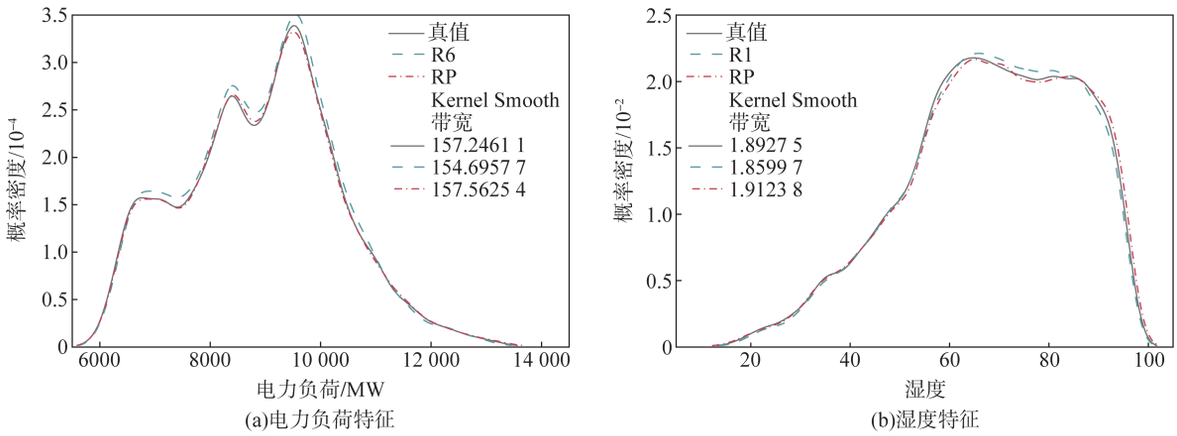


图3 插补数据与真实数据的概率密度分布对比图

Fig. 3 Comparison diagram of probability density distribution between Imputed data and real data

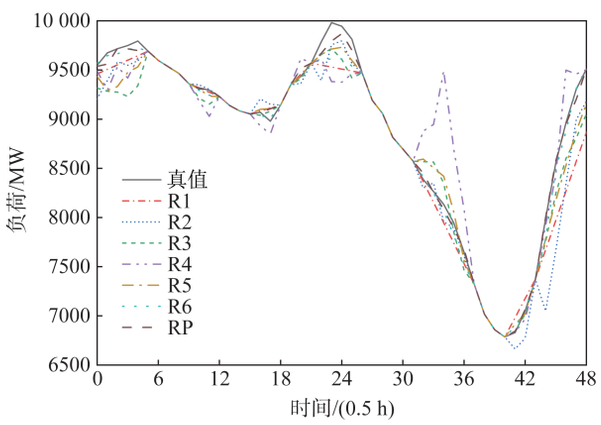


图4 负荷数据插补结果可视化

Fig. 4 Visualization of load data imputation results

误差均为最低，重建效果明显优于相同条件下其他对比方法。这表明本文所提NR-GAIN方法能够很好地应对各缺失率和各种缺失类型。

3.3.2 缺失数据重建对后续预测精度的影响分析

为验证RP能有效应对负荷预测任务中历史数据缺失的挑战，进一步构建负荷预测模型展开实验。R1—R6对应实验分别记为FU1—FU6，本文方法对应实验记为FUP。

表5给出了各实验在测试集上的预测评价指标。结果表明，FUP在超短期各时间步的预测精度均优于其他对比方法。相比其他方法，FUP在0~4 h内平均RMSE下降了3.53%~38.85%，平均MAE下降了5.11%~42.50%。这说明本文方法RP能够高效还原原始数据的信息空间，使预测模型充分学习到负荷及其关联特征中的潜在知识信息，保障高比例数据缺失情况下的负荷预测精度。

表4 其他缺失比例下不同方法对应的重构误差

Table 4 Reconstruction errors corresponding to different methods under other missing ratios

缺失比例	重构方式	评价指标	指标值									
			连续缺失					随机缺失				
			干球温度	露点温度	湿球温度	湿度	电力负荷	干球温度	露点温度	湿球温度	湿度	电力负荷
10%	R1	RMSE	0.387	0.528	0.241	2.743	208.427	0.195	0.294	0.130	1.340	82.799
		ME	0.261	0.337	0.170	1.848	146.297	0.095	0.133	0.068	0.695	58.995
		MER	1.521	3.634	1.279	3.112	1.709	0.537	1.326	0.501	1.143	0.672
	R2	RMSE	0.886	1.268	0.618	5.683	167.538	0.871	1.031	0.570	4.752	146.688
		ME	0.623	0.844	0.451	4.119	117.971	0.597	0.707	0.404	3.490	102.256
		MER	3.705	10.494	3.539	7.333	1.339	3.458	7.405	3.092	5.989	1.155
	R3	RMSE	1.005	1.223	0.474	11.334	285.239	0.247	0.326	0.176	3.201	59.052
		ME	0.552	0.705	0.226	8.303	98.994	0.157	0.224	0.113	2.551	43.290
		MER	3.252	2.651	1.761	11.915	1.077	0.876	2.299	0.809	3.564	0.489
	R4	RMSE	2.538	2.143	1.440	14.745	933.440	0.762	0.755	0.424	3.343	186.704
		ME	2.048	1.766	1.170	11.885	725.831	0.430	0.443	0.254	2.094	113.659
		MER	12.204	17.389	8.375	20.112	8.333	2.486	6.720	1.889	3.375	1.273

续表

缺失比例	重构方式	评价指标	指标值									
			连续缺失					随机缺失				
			干球温度	露点温度	湿球温度	湿度	电力负荷	干球温度	露点温度	湿球温度	湿度	电力负荷
30%	R5	RMSE	1.071	1.459	0.526	8.122	235.388	0.246	0.290	0.146	3.517	59.669
		ME	0.601	0.916	0.271	6.188	98.351	0.153	0.172	0.093	2.829	43.206
		MER	3.402	7.470	1.972	11.285	1.087	0.877	1.854	0.690	3.924	0.485
	R6	RMSE	0.404	0.591	0.283	3.013	149.035	0.245	0.333	0.170	1.207	69.266
		ME	0.281	0.387	0.208	2.074	106.473	0.167	0.179	0.114	0.751	49.090
		MER	1.655	4.297	1.550	3.571	1.210	0.982	1.823	0.830	1.249	0.549
	RP	**RMSE**	**0.213**	**0.223**	**0.117**	**0.981**	**114.907**	**0.107**	**0.102**	**0.058**	**0.599**	**59.457**
		ME	**0.098**	**0.089**	**0.053**	**0.599**	**83.338**	**0.070**	**0.062**	**0.038**	**0.401**	**43.540**
		MER	**0.589**	**0.493**	**0.419**	**0.975**	**0.946**	**0.409**	**0.598**	**0.292**	**0.647**	**0.490**
30%	R1	RMSE	0.410	0.544	0.257	2.897	202.256	0.277	0.360	0.167	1.776	113.252
		ME	0.270	0.334	0.177	1.926	142.131	0.146	0.185	0.096	1.002	75.482
		MER	1.581	4.301	1.327	3.249	1.649	0.829	2.198	0.706	1.666	0.863
	R2	RMSE	1.042	1.248	0.656	5.924	181.939	0.923	1.084	0.596	5.204	154.392
		ME	0.695	0.834	0.466	4.230	123.296	0.637	0.745	0.425	3.754	107.419
		MER	4.072	11.792	3.624	7.583	1.394	3.694	8.768	3.248	6.430	1.220
	R3	RMSE	0.570	1.630	0.391	9.596	294.257	0.351	0.400	0.239	3.890	104.321
		ME	0.343	1.171	0.250	7.202	142.192	0.223	0.243	0.160	2.898	62.116
		MER	2.002	12.410	1.948	12.931	1.575	1.289	3.370	1.151	4.177	0.699
	R4	RMSE	1.369	1.453	0.777	8.857	903.031	0.844	0.828	0.562	4.685	246.342
		ME	1.033	1.157	0.595	7.069	707.984	0.489	0.488	0.335	3.016	144.277
		MER	6.736	11.622	4.749	12.036	8.215	3.007	5.972	2.573	5.529	1.656
	R5	RMSE	0.769	1.124	0.346	13.667	320.474	0.345	0.417	0.208	3.962	105.815
		ME	0.475	0.766	0.231	10.140	148.933	0.204	0.244	0.122	2.935	62.377
		MER	2.929	6.875	1.769	19.734	1.619	1.214	3.701	0.920	4.224	0.699
	R6	RMSE	0.477	0.614	0.312	3.195	175.510	0.323	0.412	0.212	1.874	87.474
		ME	0.330	0.392	0.224	2.196	132.056	0.186	0.254	0.137	1.175	61.280
		MER	1.932	4.995	1.676	3.627	1.491	1.068	2.666	1.028	1.949	0.683
	RP	**RMSE**	**0.254**	**0.329**	**0.181**	**1.649**	**131.853**	**0.189**	**0.221**	**0.116**	**1.221**	**78.921**
		ME	**0.159**	**0.193**	**0.119**	**1.080**	**95.992**	**0.115**	**0.131**	**0.072**	**0.844**	**56.976**
		MER	**0.914**	**2.392**	**0.910**	**1.786**	**1.085**	**0.657**	**1.818**	**0.536**	**1.371**	**0.640**

表 5　不同重构方式下负荷预测评价指标

Table 5　Evaluation metrics for load forecasting under different reconstruction methods　　MW

所用方法	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}
	30 min		1 h		2 h		4 h		0~4 h 平均	
FU1	95.42	72.26	142.50	109.00	204.25	152.72	281.77	202.39	212.65	150.38
FU2	106.52	80.07	135.62	99.63	188.43	135.45	274.04	191.25	205.06	141.11
FU3	114.39	86.74	169.00	127.27	246.87	179.92	335.28	240.27	256.86	179.11
FU4	163.52	124.32	226.84	171.44	316.72	235.48	398.42	299.80	319.06	231.64
FU5	109.51	82.57	156.43	117.54	235.38	172.61	326.06	233.34	247.57	173.22
FU6	101.21	76.15	129.57	96.30	182.84	133.80	276.63	194.99	202.23	140.37
FUP	**89.81**	**67.13**	**120.48**	**89.63**	**177.22**	**128.21**	**268.43**	**186.93**	**195.09**	**133.20**

结合表3，FU6平均预测精度仅次于FUP，得益于其对应数据重建方法R6在负荷数据重构上显著优于R1—R5，所还原的负荷分布更接近真实情况。这表明，挖掘负荷自身的时序特性对预测精度提升具有关键作用。以FU6与FU1为例，尽管R1对气象特征的重构效果优于R6，但其对负荷数据的重建精度远不及R6，导致FU1平均预测精度不及FU6，从侧面验证了负荷自身规律特性对精度提升的贡献度要高于其他关联特征。FU2与FU1的对比也证实了这一点。

与此同时，R3与R5对各个特征的重构效果相当，故二者对应负荷预测实验FU3与FU5表现较为接近，而FU5略优于FU3也是由于R5对负荷数据的重构效果优于R3。相比之下，R4在各个特征上的重构误差均为最大，难以有效还原原始数据的信息结构，导致FU4精度远不及其他对比实验，进一步凸显了高质量数据重建对保障负荷预测精度的重要意义。

随机选取一天的预测结果，对各方法提前4 h的预测绝对误差（预测值与真实值之差的绝对值）进行可视化，如图5所示。结果显示，FUP的预测绝对误差整体上最接近于0。除个别时刻外，FUP的误差始终处于最低水平，表明其预测结果最接近实际负荷。

为分析缺失数据重建对预测精度的影响机制，进一步开展特征重要性分析，采用特征置换法量化不同特征的重建结果对预测性能的贡献度。具体来说，在保持其他特征为RP重建结果的前提下，将某个气象特征的重建值替换为次优方法的重建结果，即R1的重建结果，训练预测模型，观察预测误差的上升幅度。误差上升越大说明该特征的重建对预测越重要。

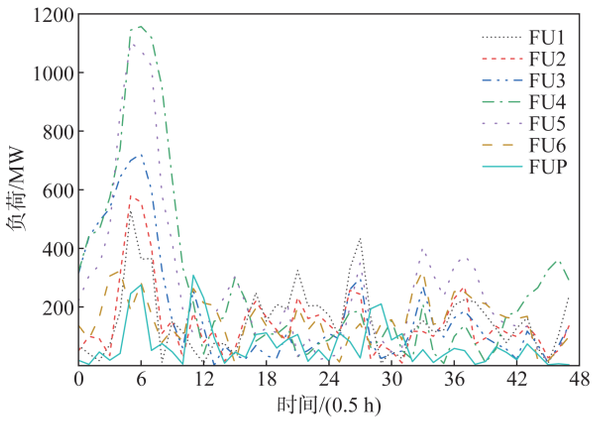


图5 提前4 h预测绝对误差可视化

Fig. 5 Visualization of 4 h-ahead absolute prediction errors

FU_{ODBT}、FU_{ODPT}、FU_{OWBT}、FU_{OHUM}分别代表替换干球温度、露点温度、湿球温度和湿度特征，负荷预测实验后的评价指标，如表6所示。结果表明，上述四个气象特征的重建对保证最终的预测精度均有一定作用。从预测时间步上看，替换各特征后，提前一步的预测误差上升最为明显，因此有效还原各特征的信息空间对保证提前0.5 h的负荷预测精度至关重要。此外，替换干球温度特征后，各个时间步上的预测误差均明显上升，说明还原干球温度的时序特征对于提升负荷预测精度具有关键作用。

表6 不同重建特征替换后对应的负荷预测评价指标

Table 6 Evaluation metrics for load forecasting corresponding to different reconstructed features after replacement MW

替换特征	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}
	30 min		1 h		2 h		4 h		0~4 h平均	
FU_{ODBT}	182.71	139.87	134.66	101.86	196.77	143.94	271.34	190.43	215.71	153.93
FU_{ODPT}	97.76	70.04	122.36	91.28	182.97	133.68	270.42	190.80	199.63	137.71
FU_{OWBT}	90.79	68.72	124.37	92.53	183.28	133.05	269.61	189.48	199.95	137.60
FU_{OHUM}	95.21	72.99	124.89	92.94	184.40	133.75	278.90	193.40	202.53	138.34
FUP	**89.81**	**67.13**	**120.48**	**89.63**	**177.22**	**128.21**	**268.43**	**186.93**	**195.09**	**133.20**

3.3.3 个性化预测子模型构建与选取策略有效性验证

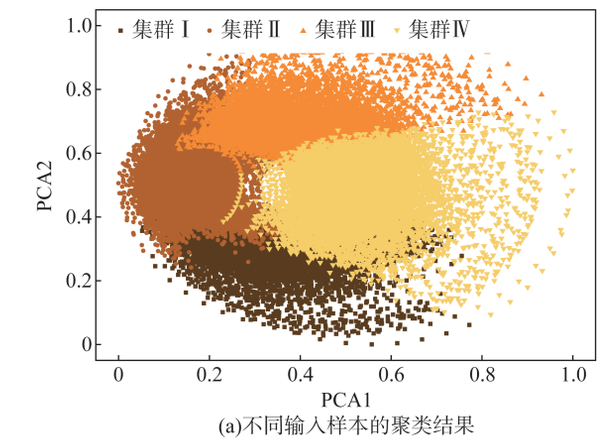
利用RP重建后的完整数据，采用基于余弦相似度的*K*-Medoids算法对滑动窗口内的历史负荷样本进行聚类，以识别相应时间尺度下的负荷数据异质性。由表7可知，最佳聚类数目*K*=4。采用主成分分析对各分布模式的样本群体降维，识别结果及各中心的负荷分布概况见图6。

表7 不同聚类数目下对应的轮廓系数

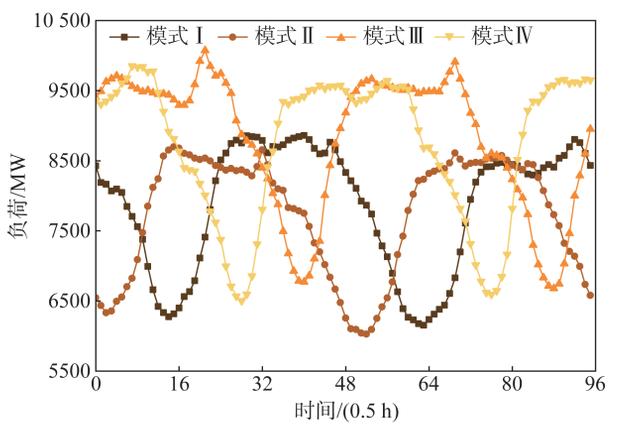
Table 7 Contour coefficients corresponding to different number of clusters

聚类数目*K*	2	3	4	5	6
轮廓系数	0.391 0	0.448 2	**0.463 2**	0.457 3	0.438 0

由图6(a)发现，基于余弦相似度的*K*-medoids算法能够有效聚合具有相似分布模式的负荷样本。结合图6(b)可见，不同群体的负荷样本在幅值形态上



(a)不同输入样本的聚类结果



(b)不同负荷模式中心对应的负荷曲线

图 6　不同输入样本的负荷分布识别概况

Fig. 6　Overview of load distribution recognition for different input samples

差异显著，表明负荷数据本身存在明显的异质性。因此，有必要针对不同负荷分布模式，进一步构建适应特定模式的个性化预测子模型。

基于上述分析，在 3.3.2 节基础上，对 FUP 的线性层进行解冻，并分别结合 4 类分布模式对应样本迁移微调统一预测模型 FUP，构建个性化预测子模型。在线应用阶段，计算负荷输入样本与各分布模式中心之间的余弦相似度以确定其所属分布模式，并选择相应的子模型执行预测。

将应用上述策略的实验 FPRO、针对不同模式独立训练预测模型的实验 FD 及 FUP 进行对比，并引入训练 150 轮与 250 轮的实验 FUP150 和 FUP250 以排除训练轮数影响。不同实验在测试集上的预测评价指标如表 8 所示。

由表 8 可知，FPRO 在各时间步上均优于其他对比方法。具体而言，FD 各子模型在训练阶段可用有效样本不足，难以充分学习到负荷变化的内在规律，导致其预测精度显著偏低。相比 FUP，尽管 FUP150 与 FUP250 在个别时间步的预测误差有所降低，但整体预测效果却表现不足，表明改变预测模型的训练轮数仍无法充分挖掘到历史数据中所蕴含的深层特征信息，限制预测精度的提升。

综上，本文所提个性化预测子模型构建及选取策略可有效保障各子模型对特定分布模式下负荷特性的建模能力，提高整体预测精度。这也从另一角度验证了 RP 在还原数据信息方面的有效性。

表 8　不同方法对应的负荷预测评价指标

Table 8　Evaluation metrics for load forecasting corresponding to different methods　　MW

所用方法	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}
	30 min		1 h		2 h		4 h		0~4 h 平均	
FD	568.59	308.23	555.16	287.58	553.77	325.85	564.27	369.84	555.90	327.15
FUP150	**87.18**	**64.61**	120.68	90.53	178.61	128.83	277.53	195.72	198.38	135.28
FUP250	89.69	67.76	125.23	93.68	180.47	128.44	**266.07**	**184.19**	196.60	133.68
FUP	89.81	67.13	**120.48**	**89.63**	**177.22**	**128.21**	268.43	186.93	**195.09**	**133.20**
FPRO	**80.33**	**59.70**	**115.52**	**84.38**	**173.95**	**124.23**	**262.19**	**182.68**	**190.33**	**128.76**

3.3.4　个性化预测子模型构建与选取策略泛化性验证

为验证本文所提个性化预测子模型构建与选取策略的模型泛化性，分别选取 CNN（D1）、BiLSTM（D2）及 CNN-BiLSTM（D3）在经 RP 重建后的完整数据上完成训练。各模型应用所提策略前后在测试集上的评价指标对比如表 9 所示。

根据表 9，应用了所提个性化预测子模型构建与选取策略后，各模型在多个时间步上的预测精度均有所提升，表明该策略具有良好的模型泛化性。

此外，为验证上述策略在更极端缺失场景下的表现，针对数据集中各特征数据分别设置 60% 连续型缺失，采用本文所提 NR-GAIN 方法重建得到完整数据集，并构建统一预测模型，将该方法命名为 $FUP_{60\%}$。基于重建后的完整数据，对负荷数据异质性进行识别，并应用本文所提个性化预测子模型构建与选取策略，并将该方法命名为 $FPRO_{60\%}$。

表 10 表明，即使面临数据超过半数的 60% 连续缺失场景，在保证数据重建质量的基础上，应用本文所提个性化预测子模型构建与选取策略，仍能进一步提升负荷预测精度。

表 9 不同模型对应的负荷预测评价指标

Table 9 Evaluation metrics for load forecasting corresponding to different models MW

所用方法	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}
	30 min		1 h		2 h		4 h		0~4 h平均	
D1	99.66	77.10	135.74	101.98	192.85	140.37	278.27	199.61	206.98	145.00
D1Pro	**91.79**	**70.12**	**123.60**	**91.44**	**183.15**	**133.16**	**265.00**	**188.18**	**197.11**	**136.48**
D2	88.60	67.31	126.21	95.13	187.91	138.09	279.76	198.75	203.34	140.63
D2Pro	**85.28**	**64.96**	**120.34**	**90.51**	**182.88**	**133.68**	**271.57**	**192.16**	**197.98**	**136.49**
D3	100.27	76.63	129.36	98.49	179.17	132.67	271.04	194.04	196.04	137.91
D3Pro	**89.75**	**68.30**	**120.41**	**90.16**	**174.26**	**128.17**	**265.01**	**187.80**	**191.82**	**133.00**

表 10 连续缺失 60% 场景下不同方法对应的负荷预测评价指标

Table 10 Evaluation metrics for load forecasting corresponding to different methods with continuous 60% missing MW

所用方法	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}	e_{RMSE}	e_{MAE}
	30 min		1 h		2 h		4 h		0~4 h平均	
$FUP_{60\%}$	110.23	84.38	144.29	109.36	196.71	144.27	286.76	199.07	213.75	148.58
$FPRO_{60\%}$	**88.25**	**65.90**	**131.26**	**97.25**	**191.39**	**137.46**	**279.76**	**192.87**	**206.71**	**139.81**

表 9 与表 10 的结果充分验证了本文所提个性化预测子模型构建与选取策略的有效性及良好泛化性，表明其具备较高的工程应用价值。

4 结 论

本文提出一种计及历史数据缺失与负荷时间异质的超短期负荷预测方法，通过算例分析和验证得出以下结论：

1）数据的时序特性以及各关联特征中所蕴含的有价值信息均能作为时间序列数据缺失重建的重要依据。

2）引入随机噪声替代部分可观测值进行插补模型训练，显式度量上述噪声对应的生成偏差，能够避免训练过程模型仅关注可观测值的插补误差，以进一步提升插补精度。

3）负荷数据的波动特性是动态变化的。应用 *K*-medoids，对预测模型滑动窗口内的负荷输入样本进行聚类，能够在更小时间尺度上，分析识别负荷数据的时间异质性。

4）基于负荷输入样本间的异质性构建个性化预测子模型，在线应用时，根据当前负荷输入样本与各模式中心的相似性选择对应子模型，不仅能避免对日历、天气等外部信息的依赖，还能更精准地捕捉负荷波动规律，从而提升预测精度。

本文方法有效缓解了训练数据质量对负荷预测精度的制约，并减弱了负荷数据异质对超短期预测的影响，展现出良好的鲁棒性和工程应用价值。然而，目前关于数据质量低下与时间异质性的挑战仍主要被视为两个相互独立的子问题。未来的研究将致力于从全局视角统筹应对这两类问题，以期进一步提升数据插补与负荷预测的精度。

利益冲突声明（Conflict of Interests）：

所有作者声明不存在利益冲突。

作者贡献声明（Authors' Contributions）：

史瑞研负责研究概念生成、数据管理及分析、实验方法设计、程序设计及验证、实验结果可视化、论文初稿撰写；赵永宁负责研究资金获取、研究课题监管与指导、论文审阅与修订；付坤明参与论文审阅与修订；刘静文参与论文审阅与修订。所有作者均阅读并同意了论文终稿内容。

5 参考文献

[1] GAO D, ZHI Y, RONG X, et al. Mismatch analysis of rooftop photovoltaics supply and farmhouse load: data dimensionality reduction and explicable load pattern mining via hybrid deep learning [J]. Applied Energy, 2025, 377: 124520.

[2] 陶李丹澜，杨明，孙东磊，等. 基于源荷时序耦合特征提取的短期负荷预测[J]. 供用电，2025，42(11)：3-16.
TAO Lidanlan, YANG Ming, SUN Donglei, et al. Short-term load prediction based on source-load time-series coupled features extraction[J]. Distribution & Utilization, 2025, 42(11): 3-16.

[3] SUNDARARAJAN A, KHAN T, MOGHADASI A, et al. Survey on synchrophasor data quality and cybersecurity challenges, and evaluation of their interdependencies [J]. Journal of Modern Power Systems and Clean Energy, 2019, 7(3): 449-467.

[4] 程其云，李峰，罗澍忻，等. 新型电力系统规划方法框架及关键支撑技术[J]. 电网技术，2025，49(6)：2219-2231.

CHENG Qiyun, LI Feng, LUO Shuxin, et al. Research on the planning methodology framework and key supporting technologies for new power systems[J]. Power System Technology, 2025, 49(6): 2219-2231.

[5] 于多，曹燚，王海荣，等. 基于ICEEMDAN-PE和IDBO-Informer组合模型的短期负荷预测[J]. 中国电力，2025，58(6)：19-32.
YU Duo, CAO Yi, WANG Hairong, et al. Short-term load forecasting based on a combined ICEEMDAN-PE and IDBO-informer model[J]. Electric Power, 2025, 58(6): 19-32.

[6] 李加文，孙永辉，王森，等. 计及异常场景数据缺失的负荷超短期预测[J]. 电力系统自动化，2025，49(15)：133-143.
LI Jiawen, SUN Yonghui, WANG Sen, et al. Ultra-short-term load forecasting considering missing data in anomalous scenarios [J]. Automation of Electric Power Systems, 2025, 49(15): 133-143.

[7] 孙玉芹，王亚文，朱威，等. 基于考虑气温影响的门限自回归移动平均模型居民日用电负荷预测[J]. 电力建设，2022，43(9)：117-124.
SUN Yuqin, WANG Yawen, ZHU Wei, et al. Residential daily power load forecasting based on threshold ARMA model considering the influence of temperature [J]. Electric Power Construction, 2022, 43(9): 117-124.

[8] MONJAZEB M R, AMIRI H, MOVAHEDI A. Wholesale electricity price forecasting by quantile regression and Kalman filter method [J]. Energy, 2024, 290: 129925.

[9] LIU J X, LEU J S. LARSI-TPE-XGB: short-term load forecasting by load-adaptive relative strength index and fusion of tree-structured parzen estimator and XGBoost [J]. IEEE Transactions on Power Delivery, 2025, 40(3): 1318-1330.

[10] 胡枭，张泽朕，杨家全，等. 融合路网-气象-日期多特征信息的电动汽车充电负荷预测[J]. 电力建设，2025，46(9)：57-70.
HU Xiao, ZHANG Zezhen, YANG Jiaquan, et al. Electric vehicle charging load forecasting by integrating multi-feature information of road network, meteorology, and date [J]. Electric Power Construction, 2025, 46(9): 57-70.

[11] 吴军英，路欣，刘宏，等. 基于Spearman-GCN-GRU模型的超短期多区域电力负荷预测[J]. 中国电力，2024，57(6)：131-140.
WU Junying, LU Xin, LIU Hong, et al. Ultra-short-term multi-region power load forecasting based on spearman-GCN-GRU model [J]. Electric Power, 2024, 57(6): 131-140.

[12] QIN Y S, RAO J N K, REN Q S. Confidence intervals for marginal parameters under fractional linear regression imputation for missing data[J]. Journal of Multivariate Analysis, 2008, 99(6): 1232-1259.

[13] SHEN X D, ZHAO H X, XIANG Y, et al. Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments[J]. Electric Power Systems Research, 2022, 212: 108247.

[14] SUN M, LAN L, ZHU C G, et al. Cubic spline interpolation with optimal end conditions[J]. Journal of Computational and Applied Mathematics, 2023, 425: 115039.

[15] 余嘉茵，何玉林，崔来中，等. 针对大规模数据的分布一致缺失值插补算法[J]. 清华大学学报(自然科学版)，2023，63(5)：740-753.
YU Jiayin, HE Yulin, CUI Laizhong, et al. Distribution consistency-based missing value imputation algorithm for large-scale data sets[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(5): 740-753.

[16] 贾晓红，石岚，郝玉珠. 基于风电场片区风速的人工智能插补方法对比[J]. 太阳能学报，2025，46(1)：168-175.
JIA Xiaohong, SHI Lan, HAO Yuzhu. Comparison of artificial intelligence interpolation methods based on wind speed of wind farm divided into wind turbine groups[J]. Acta Energiae Solaris Sinica, 2025, 46(1): 168-175.

[17] 卢冠华，余涛，吴毓峰，等. 基于MAGAT的风电场功率缺失数据填充方法[J]. 电网技术，2024，48(8)：3391-3400.
LU Guanhua, YU Tao, WU Yufeng, et al. MAGAT-based method for imputing power loss data in wind farms [J]. Power System Technology, 2024, 48(8): 3391-3400.

[18] MURTAZA S, RAJ S, YUN G Y, et al. Adaptive neural temporal hybridization for missing data imputation in building energy use datasets: an integrated LNN-LSTM weighted model[J]. Journal of Building Engineering, 2025, 112: 113774.

[19] 刘璐璐，惠曾强. 异构不完整数据整合的VAE模型处理与应用[J]. 中国科技信息，2020(14)：74-76.
LIU Lulu, HUI Zengqiang. Processing and application of VAE model for heterogeneous incomplete data integration [J]. China Science and Technology Information, 2020(14): 74-76.

[20] LIAO W L, WANG Y S, WANG Y L, et al. Scenario generation for cooling, heating, and power loads using generative moment matching networks [J]. CSEE Journal of Power and Energy Systems, 2022, 8(6): 1730-1740.

[21] 庞昭辰，刘明，张立宪，等. 基于条件扩散模型的卫星遥测数据缺失值插补方法[J]. 自动化学报，2025，51(10)：2302-2312.
PANG Zhaochen, LIU Ming, ZHANG Lixian, et al. Conditional diffusion model-based imputation method for missing satellite telemetry data[J]. Acta Automatica Sinica, 2025, 51(10): 2302-2312.

[22] 杨玉莲，齐林海，王红，等. 基于生成对抗和双重语义感知的配电网量测数据缺失重构[J]. 电力系统自动化，2020，44(18)：46-54.
YANG Yulian, QI Linhai, WANG Hong, et al. Reconstruction of missing measurement data in distribution network based on generative adversarial network and double semantic perception[J]. Automation of Electric Power Systems, 2020, 44(18): 46-54.

[23] YUAN X F, ZHANG J L, WANG K, et al. Missing data imputation for industrial time series with adaptive Median iteration based on generative adversarial networks [J]. IEEE Sensors Journal, 2024, 24(21): 35081-35091.

[24] 赵厚翔，沈晓东，吕林，等. 基于GAN的负荷数据修复及其在EV短期负荷预测中的应用[J]. 电力系统自动化，2021，45(16)：143-151.
ZHAO Houxiang, SHEN Xiaodong, LÜ Lin, et al. Load data restoration based on generative adversarial network and its application in short-term load forecasting of electric vehicle [J]. Automation of Electric Power Systems, 2021, 45(16): 143-151.

[25] 郭小龙，李子康，刘灏，等. 基于增强生成对抗网络的PMU丢失数据恢复方法[J]. 电网技术，2022，46(6)：2114-2121.
GUO Xiaolong, LI Zikang, LIU Hao, et al. PMU missing data

recovery algorithm based on enhanced generative adversarial network[J]. Power System Technology, 2022, 46(6): 2114-2121.
[26] LUO L, FAN Z Y, CHEN Y M, et al. Cyclic generative adversarial networks with KNN-transformers for missing traffic data completion[J]. Applied Soft Computing, 2024, 167: 112406.
[27] LIN J, WU D D, HUANG L P. Self-supervised bi-directional mapping generative adversarial network for arbitrary-time longitudinal interpolation of missing data [J]. Biomedical Signal Processing and Control, 2025, 105: 107514.
[28] YOON J, JORDON J, SCHAAR M. Gain: missing data imputation using generative adversarial nets [C]//International conference on machine learning. PMLR, 2018: 5689-5698.
[29] 叶林, 宫婷, 宋旭日, 等. 基于波动类型精细划分与聚类的短期负荷预测[J]. 电网技术, 2023, 47(3): 998-1009.
YE Lin, GONG Ting, SONG Xuri, et al. Short-term load forecasting based on fine division and clustering of fluctuation types [J]. Power System Technology, 2023, 47(3): 998-1009.
[30] HNIN S W, KARNJANA J, KOHDA Y, et al. A hybrid K-means and KNN approach for enhanced short-term load forecasting incorporating holiday effects[J]. Energy Reports, 2024, 12: 5942-5959.
[31] SHARMA A, JAIN S K. A novel seasonal segmentation approach for day-ahead load forecasting[J]. Energy, 2022, 257: 124752.
[32] YUAN M, XIE J W, LIU C, et al. Short-term load forecasting for an industrial building based on diverse load patterns [J]. Energy, 2025, 334: 137481.
[33] LIU Q, HOU H J, ZHOU Z K, et al. A novel heat load forecasting method based on CASDA-driven unsupervised clustering [J]. Journal of Building Engineering, 2025, 111: 113457.
[34] GAO Z F, XIE Y F. A parameter-based multi-source transfer learning method for building load forecasting with sparse data scenarios[J]. Energy Reports, 2025, 13: 4936-4947.
[35] 黄宇腾, 侯芳, 周勤, 等. 一种面向需求侧管理的用户负荷形态组合分析方法[J]. 电力系统保护与控制, 2013, 41(13): 20-25.
HUANG Yuteng, HOU Fang, ZHOU Qin, et al. A new combinational electrical load analysis method for demand side management [J]. Power System Protection and Control, 2013, 41 (13): 20-25.
[36] 孟衡, 张涛, 王金, 等. 基于多尺度时空图卷积网络与Transformer融合的多节点短期电力负荷预测方法[J]. 电网技术, 2024, 48(10): 4297-4305.
MENG Heng, ZHANG Tao, WANG Jin, et al. Multi-node short-term power load forecasting method based on multi-scale spatiotemporal graph convolution network and transformer [J]. Power System Technology, 2024, 48(10): 4297-4305.
[37] LI J, LAN D, GONG Y S, et al. Missing traffic data imputation with a conditional diffusion framework [J]. Expert Systems with Applications, 2026, 296: 129140.
[38] SUN J, MA K X, ZHAO H Y. Short-term power load hybrid forecasting using GRU and SCN[J]. Energy and Buildings, 2025, 347: 116372.

收稿日期: 2025-10-21 **修回日期**: 2025-11-09


史瑞研

作者简介:

史瑞研(2001),男,硕士研究生,主要研究方向为电力负荷预测;

赵永宁(1990),男,博士,副教授,通信作者,主要研究方向为新能源发电技术、新能源电力系统调度控制等,E-mail:zyn@cau.edu.cn;

付坤明(2001),男,硕士研究生,主要研究方向为电力系统暂态稳定评估;

刘静文(2003),女,硕士研究生,主要研究方向为新能源发电预测。

(编辑 孙静琳)