

基于深度学习协同持续学习方法的风电功率预测验证

刘莘轶¹, 王梓涛², 王洋², 徐加陵², 刘单珂², 于立军²

(1. 上海交通大学机械与动力工程学院, 上海 200240; 2. 上海交通大学智慧能源创新学院, 上海 200240)

摘要: 随着风电装机容量和并网比例持续提升, 预测模型必须在动态环境中频繁更新, 但增量学习易遗忘且批量学习计算开销巨大。本文提出了一种信息增强记忆感知突触(IE-MAS)协同持续学习框架, 通过特异性缩放、任务相似度调控和知识蒸馏三重机制协同抑制遗忘并平衡新旧知识。以新疆某 200 MW 风电场为实验对象, 分别在长短期记忆网络(LSTM)、门控循环单元(GRU)、Transformer 和时间卷积网络(TCN)4种神经网络中验证 IE-MAS 的性能。结果表明: 在每轮仅引入 10% 新增数据的条件下, IE-MAS 使 LSTM 的均方根误差(RMSE)、平均绝对误差(MAE)分别降低约 1.7% 和 2.5%, 拟合系数 R^2 提升约 1.9%; 在跨模型对比中, 与需累积所有历史数据的批量训练相比, IE-MAS 在各架构上表现相当或略优, 且单轮次训练时间平均压缩约 70%, 为超短期风电预测及其他时序持续学习场景提供了兼顾效率和精度的实时更新方案。

关键词: 风力发电; 功率预测; 持续学习; 深度神经网络

中图分类号: TK221

文献标识码: A

DOI:10.16146/j.cnki.rndlge.2025.11.014

[引用本文格式] 刘莘轶, 王梓涛, 王洋, 等. 基于深度学习协同持续学习方法的风电功率预测验证[J]. 热能动力工程, 2025, 40(11): 119-127. LIU Xinyi, WANG Zitao, WANG Yang, et al. Validation of wind power prediction based on a deep learning-driven collaborative continual learning method[J]. Journal of Engineering for Thermal Energy and Power, 2025, 40(11): 119-127.

Validation of Wind Power Prediction based on a Deep Learning-Driven Collaborative Continual Learning Method

LIU Xinyi¹, WANG Zitao², WANG Yang², XU Jialing², LIU Shanke², YU Lijun²

(1. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China, Post Code: 200240;

2. College of Smart Energy, Shanghai Jiao Tong University, Shanghai, China, Post Code: 200240)

Abstract: As the installed capacity and grid-integration ratio of wind power continue to rise, prediction models must be updated frequently in dynamic environments. However, incremental learning suffers from catastrophic forgetting, while batch retraining incurs prohibitive computational costs. In this paper, an information-enhanced memory aware synapses (IE-MAS) collaborative continual learning framework, that leverages three synergistic mechanisms including specificity scaling, task similarity regulation and knowledge distillation, was proposed to suppress forgetting and balance new versus retained knowledge. Taking a 200 MW wind farm in Xinjiang as experimental object, IE-MAS performance was evaluated across four sequence-modeling architectures, such as long short term memory (LSTM), gated recurrent unit (GRU), Transformer, and temporal convolutional network (TCN). The result shows that under the constraint of introducing only 10% new data per update, IE-MAS reduces the LSTM's root mean square error (RMSE) and mean absolute error (MAE) by approximately 1.7% and 2.5%, respectively, while boosting R^2 by about 1.9%. In cross-model comparisons, IE-MAS performs on par with—or slightly better

收稿日期: 2025-04-25; 修订日期: 2025-06-14

基金项目: 国家电力投资集团有限公司-上海交通大学“未来能源计划联合基金”(WLN-Y-ZD-2021-010)

Fund-supported Project: "Future Energy Plan" Joint Fund of State Power Investment Corporation Limited and Shanghai Jiao Tong University (WLN-Y-ZD-2021-010)

作者简介: 刘莘轶(1999-), 男, 上海交通大学博士研究生。

通信作者: 于立军(1969-), 男, 上海交通大学教授。

than—batch training that accumulates all historical data, and cuts per-round training time by roughly 70% on average. This real-time update strategy offers a compelling balance of efficiency and accuracy for ultra-short-term wind power prediction and other sequential continual learning scenarios.

Key words: wind power generation, power prediction, continual learning, deep neural networks

引 言

随着全球能源结构加速向可再生能源转型,凭借成熟的开发体系与规模化应用优势,风电已成为构建新型电力系统的战略支撑。然而,风电机组固有的强波动、间歇与随机特性^[1],在大规模并网场景下对电网安全稳定运行构成严峻考验,使得高精度风电功率预测系统成为破解并网困局的关键技术手段^[2]。作为电网调度中枢的核心决策依据,功率预测技术不仅支撑着发电计划的制定与电网调度的决策,更是国家能源局“两个细则”(《电力并网运行管理规定》^[3]和《电力辅助服务管理办法》^[4])等政策文件重点部署的实施方向,其技术演进已成为影响新能源消纳效率和双碳进程的关键性因素。

风电功率预测技术的研究主要集中在基于气象与发电数据训练预测模型的统计方法上,研究方法经历了从传统统计模型到时间序列模型的逐步发展(如线性回归、马尔可夫链、ARIMA 等)^[5],随后引入了机器学习方法(如支持向量机、随机森林、SVR 等)^[6-7],并最终发展到深度学习模型(如 RNN、CNN、LSTM、Transformer 等)^[8-9]的广泛应用。目前针对时序性的深度学习神经网络应用最为广泛,包括 LSTM、GRU、Transformer 以及 TCN 等。陈峰等人^[10]构建了基于贝叶斯优化的 LSTM 预测模型,模型利用绳索算法进行特征选择,并采用贝叶斯优化对 LSTM 网络的超参数进行调优,与传统的 BP 神经网络、SVM 等方法相比,基于贝叶斯优化的 LSTM 预测模型具有更高的预测精度。刘新宇等人^[11]提出了一种基于贝叶斯优化的变分模态分解(VMD)–GRU 组合预测模型,采用 GRU 对每个 VMD 分解子序列进行独立预测,通过简化的门控机制捕捉功率数据的时序依赖关系,适用于具有高度波动性的风电功率数据。这些方法充分证明时序性的神经网络在提高预测精度方面取得了显著进展。

当前,大多数研究仍采用基于静态数据集的全量训练方法,未能有效解决数据的时效性和持续更新需求。在环境动态变化及场站算力资源受限的场

景下,传统全量训练方法计算成本高、迭代周期长,导致模型难以实现快速适应性调整。作为一种新兴的技术方法,持续学习仅利用少量新数据对模型进行快速更新,从而显著降低了对计算资源的要求,并加速了模型训练过程^[12]。Li 等人^[13]基于建筑的用电数据,采用多种持续学习策略,评估了持续学习策略在短期建筑能耗预测中的表现,在训练效率上明显优于累积学习。然而,在深度神经网络的应用中持续学习策略在学习新任务时常常“覆盖”旧任务知识,面临“灾难性遗忘”问题^[14],这一问题在风电功率预测等长期运行的任务中尤为突出,对持续学习方法的应用构成了严峻挑战。严璐晗等人^[15]基于神经网络对光伏功率进行时序建模,并引入弹性权重固化(EWC)与在线弹性权重固化(Online-EWC)算法缓解增量学习中的灾难性遗忘问题,表明对参数变化施加约束或惩罚是保持旧知识重要参数稳定的一种有效策略。

针对持续学习中灾难性遗忘问题,本文提出一种融合特异性缩放、任务相似度调控和知识蒸馏技术的信息增强记忆感知突触(Information-Enhanced Memory Aware Synapses, IE-MAS)方法,以增强有关键知识在模型中的稳定性并降低新任务学习过程中的知识覆盖风险,并将 IE-MAS 在 LSTM、GRU、Transformer 和 TCN 4 种处理时序数据的深度神经网络中开展分析与验证。通过 MAE、RMSE、 R^2 及计算时间指标评估基于 IE-MAS 的持续学习策略在提升预测性能与资源优化配置方面的优势。

1 持续学习模型理论

1.1 基于信息增强策略的 MAS 改进

记忆感知突触(Memory Aware Synapses, MAS)^[16]作为一种在持续学习中经验证有效的策略,其核心思想是直接利用梯度信息计算模型参数的重要性,进而构建统一的正则项来约束参数在新任务中的更新,从而较为准确地反映参数对任务输出的实际贡献。相较于 EWC, MAS 直接利用梯度计算,避免了费舍尔信息矩阵近似和独立正则项的复杂性,更精准地捕捉参数贡献。与突触智能相比,

MAS 不依赖梯度波动,减少了权重估计误差,从而提升了模型的适应性。基于这些优势,本研究在 MAS 策略中进一步引入了特异性缩放、任务相似度调控以及知识蒸馏 3 个关键改进点,构建了 IE-MAS 算法,如图 1 所示。其核心是在持续学习过程中平衡新旧任务之间的参数更新,以降低灾难性遗忘风险并提升模型泛化性能。

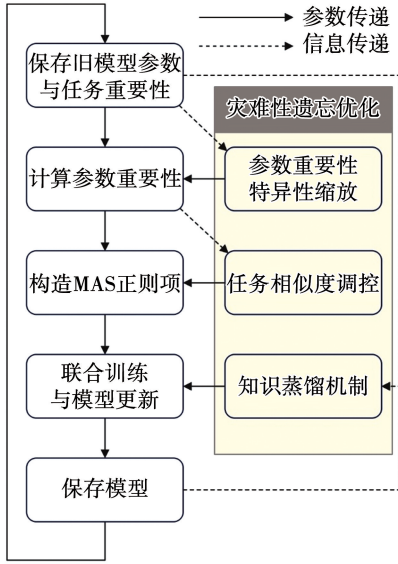


图 1 IE-MAS 算法流程图

Fig. 1 Flowchart of IE-MAS algorithm

基于信息增强策略的 MAS 改进具体流程如下:

流程 1:旧模型参数保存

在上一任务(或历史任务)训练结束后,将模型参数 θ_{old} 以及上一任务计算得到的参数重要性 Ω_{old} 进行快照保存,为后续新任务训练提供对比信息,防止对旧任务中的关键知识产生过大偏离。

流程 2:计算参数重要性

在新任务数据上,采用前向传播和反向传播的方法,使用输出的欧几里得范数作为指标,计算每个参数梯度的绝对值。对全部批次求平均后得到每个参数的重要性 Ω_{new} 。

$$\Omega_{new}(i) = \frac{1}{N} \sum_{n=1}^N \left| \frac{\partial \|f(x_{(n)})\|}{\partial \theta_i} \right| \quad (1)$$

式中: $\Omega_{new}(i)$ —在新任务上计算出的第 i 个模型参数的重要性; N —计算时所用样本的总数; θ_i —第 i 个可训练参数向量; $\|\cdot\|$ —向量的欧几里得范数; x_n —第 n 个样本的输入特征向量。

流程 3:特异性缩放

利用保存的旧任务参数重要性 Ω_{old} ,针对在旧

任务中已经具有较高重要性的参数,对新任务的重要性 Ω_{new} 进行二次放大。对每个参数 i ,若 $\Omega_{old}(i)$ 超过该任务下的均值,则将 $\Omega_{new}(i)$ 按规则进行自适应缩放。

$$\hat{\Omega}_i = \Omega_{new}(i) \times [1 + \alpha \cdot I\{\Omega_{old}(i) > \text{mean}(\Omega_{old})\}] \quad (2)$$

式中: $\Omega_{new}(i)$ —旧任务上第 i 个参数的重要性; α —预设缩放因子,用于增强参数的重要性; $\{\cdot\}$ —指示函数,当内部条件成立时其值为 1,否则为 0; $\hat{\Omega}_i$ —缩放后的重要性。

流程 4:任务相似度调控

通过计算新任务和旧任务的重要性向量之间的余弦相似度,能够衡量两者在参数空间中的相似性。基于该相似度值,动态调整 MAS 正则化强度的超参数 λ ,调整公式为:

$$\lambda' = \lambda \times \left(C - \frac{\langle \Omega_{old}, \Omega_{new} \rangle}{\|\Omega_{old}\| \cdot \|\Omega_{new}\|} \right) \quad (3)$$

式中: λ —MAS 基础权重(未调控前); λ' —调控后权重; C —基准常量,调控正则化强度基线值。其中,基准常量 C 调控正则化强度基线值。

当新任务与旧任务的相似度较高时,调整后的 λ' 会减小,从而使得正则化约束相对柔和,允许模型在新任务中拥有更多自由度;当相似度较低时, λ' 增大,以增强对旧任务知识的保护,抑制灾难性遗忘,平衡学习过程中知识迁移与固化的矛盾。

流程 5:构造 IE-MAS 正则项

利用经过自适应缩放后的参数重要性 Ω 以及调整后的 λ' ,构造正则项:

$$L_{MAS} = \lambda' \sum_i \hat{\Omega}_i (\theta_i - \theta_{old,i})^2 \quad (4)$$

式中: L_{MAS} —最终构造的 IE-MAS 正则化损失; $\theta_{old,i}$ —上一任务训练结束后保存的第 i 个参数的值。

流程 6:引入知识蒸馏机制

为缓解持续学习中的灾难性遗忘问题,引入基于教师-学生架构的知识蒸馏策略。设旧模型参数为 θ_{old} ,新模型参数为 θ_{new} ,定义知识蒸馏损失函数为:

$$L_{KD} = \frac{1}{N} \sum_{i=1}^N \|f_{\theta_{new}}(x_n) - f_{\theta_{old}}(x_n)\|^2 \quad (5)$$

式中: L_{KD} —知识蒸馏的损失; N —批次样本数; $f_{\theta}(\cdot)$ —模型输出层激活值。

损失函数通过最小化新旧模型在特征空间中的欧几里得范数距离,使新模型继承旧模型的泛化表

征能力。

流程 7:联合训练与模型更新

将知识蒸馏损失与 MAS 正则项和任务损失(如 L_{task})按照权重进行加权结合 L_{total} :

$$L_{total} = L_{task} + L_{KD} + L_{MAS} \tag{6}$$

式中: L_{total} —用于联合训练和更新模型的总损失;
 L_{task} —当前新任务的损失,例如预测误差。

利用联合损失函数 L_{total} 对新任务数据进行训练,反向传播更新模型参数。

流程 8:保存模型及更新信息

训练结束后,保存更新后的模型参数以及本任务中计算得到的重要性信息,为下一轮持续学习提供参考。

1.2 基于时序性神经网络的训练模型

为验证 IE-MAS 在基于神经网络的风电功率预测持续学习场景中的性能增益,本文选取了 4 种在时序建模领域有代表性的深度神经网络结构:长短时记忆网络^[17] (Long Short Term Memory, LSTM)、门控循环单元^[18] (Gated Recurrent Unit, GRU)、Transformer^[19] 及时序卷积网络^[20] (Temporal Convolutional Network, TCN)。

LSTM 借助输入门、遗忘门和输出门对信息流进行层层筛选,并通过独立的细胞状态通道保存和更新长期记忆,有效克服了传统循环神经网络(Recurrent Neural Network, RNN)在长时序数据上梯度衰减的问题。GRU 通过将遗忘门与输入门合并为单一的更新门,并保留重置门,以更加紧凑的结构完成与 LSTM 相似的记忆功能。在参数量和计算成本大幅减少的同时,GRU 仍能高效捕获中程依赖。

Transformer 完全基于自注意力机制,摒弃了递归运算结构,实现了对序列中任意两点间依赖的并行建模。多头注意力可在不同子空间并行提取多尺度时序特征,具有并行化优势并可以对长距离依赖进行灵活捕捉。

TCN 利用因果卷积确保模型输出仅依赖于历史信息,并通过膨胀卷积以指数级方式扩展感受野,从而在保持并行计算效率的同时捕获更长的时序依赖。结合残差连接,TCN 能够构建更深层次的网络结构,并在训练中缓解梯度退化。

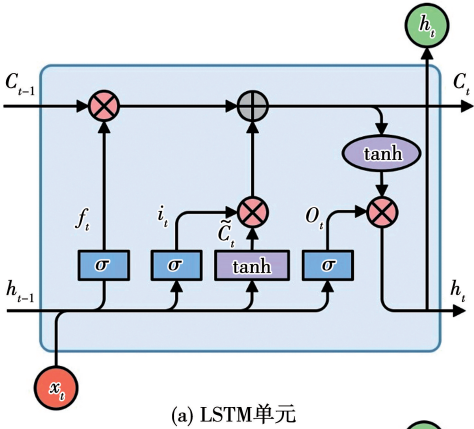
表 1 汇总了上述 4 种模型在关键信息传递或特征提取阶段的核心数学表达式,图 2 展示了 4 种模型的关键结构。以上模型在参数更新机制、长程依

赖捕捉能力以及并行计算效率等方面各有侧重,通过对比分析来揭示 IE-MAS 策略在不同网络架构下对灾难性遗忘的缓解效果及泛化性能提升效果。

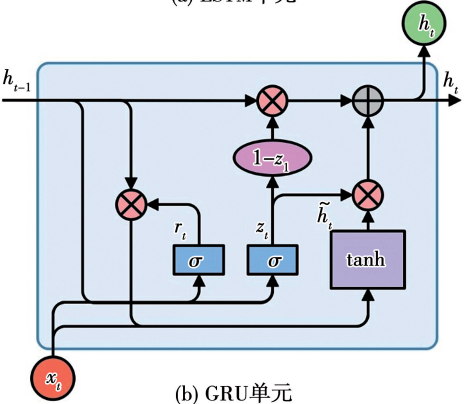
表 1 用于时序建模的神经网络核心数学表达式^[17-20]

Tab. 1 Core mathematical formulations of deep neural network architectures for sequence modeling^[17-20]

模型	功能模块	数学表达式
LSTM	输入门	$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$
	遗忘门	$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$
	输出门	$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$
	候选状态	$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$
	细胞状态更新	$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$
GRU	隐藏状态更新	$h_t = o_t \odot \tanh(c_t)$
	更新门	$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z)$
	重置门	$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$
	候选隐藏态	$\tilde{h}_t = \tanh(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_h)$
Transformer	隐藏状态更新	$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1}$
	自注意力机制	$Attention = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$ $Q = XW_Q, K = XW_K, V = XW_V$
TCN	膨胀因果卷积	$y_t = \sum_{i=0}^{k-1} W_i x_{t-di}$



(a) LSTM单元



(b) GRU单元

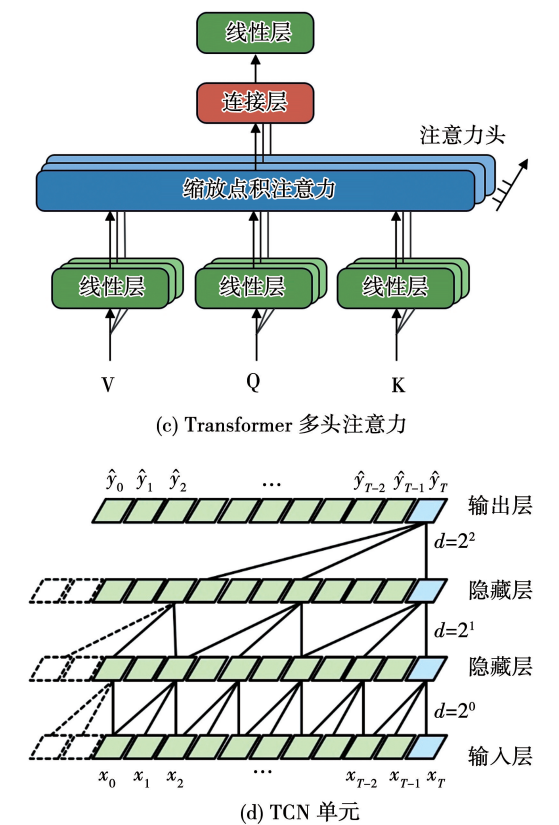


图 2 用于时序建模的深度学习结构示意图^[17-20]

Fig. 2 Schematic diagrams of deep neural network architectures for sequence modeling^[17-20]

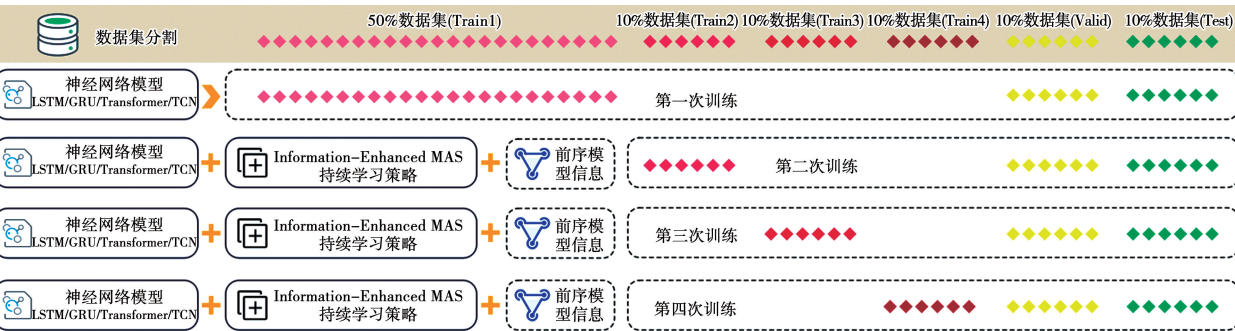


图 3 持续学习流程框架图

Fig. 3 Framework diagram of continual learning process

在每次增量训练结束后,最新的网络权重与更新后参数重要性将作为下一轮迭代的“旧模型”基准,不断循环上述增量流程。最终,在保留验证集和测试集基础上对各轮模型进行评估,量化 IEMAS 策略在不同时序网络结构下对风电功率预测性能和知识保持能力的提升效果。

2 持续学习模型框架

图 3 为 IEMAS 持续学习框架的完整流程,核心在于以迭代增量的方式兼顾历史知识的稳固和新数据的适应反馈。数据集划分模块将全量时序数据首先划分为首条 50% 的训练集(Train₁)和剩余 50% 的 5 个 10% 子集(依次对应 Train₂、Train₃、Train₄、验证集 Valid、测试集 Test)。在第一轮训练中,使用 Train₁、Valid 和 Test 进行模型初始化和初次验证,LSTM、GRU、Transformer 以及 TCN 均在该批量数据上完成基本参数学习,并计算出初始参数重要性分布。

进入第二轮至第四轮的增量训练时,系统首先将新的 10% 数据块并入预处理流程(时间对齐、标准化等),以生成适合网络输入的增量样本。此后,增量数据与先前阶段的“前序模型信息”(即上轮训练所得的模型权重与参数重要性快照)一并输入到 IE-MAS 模块,该模块首先对新任务数据重新评估参数敏感度,继而通过特异性缩放和任务相似度调控对重要性权重进行自适应调整,并生成新的 MAS 正则项。在此基础上,模型通过联合损失(含原始预测误差、MAS 正则以及可选的知识蒸馏项)对新增样本进行微调,完成第二次、第三次和第四次的增量优化。

3 实验结果与分析

3.1 数据处理

本文以新疆哈密东南地区某 200 MW 风电场为研究对象,该场由 80 台单机容量 2.5 MW 的风力机组成。所采集数据覆盖 2022 年 1 月 22 日至 2023 年 10 月 15 日期间,时间分辨率为 15 min,包含逐台

风力机的风速、风向、有功功率及无功功率。首先通过计算各台风力机有功功率与整场总功率的皮尔逊相关系数,筛选出相关性最高的 3 台风力机,并提取其点位风速与风向作为特征。为消除量纲差异并降低异常值影响,所有变量均采用 z-score 标准化,其计算公式如下:

z_i = (x_i - μ) / σ (7)

式中:z_i—原始样本值经过标准化处理后得到的结果;x_i—原始样本值;μ、σ—变量在全部样本上的均值与标准差。

3.2 评价指标

本文选取平均绝对误差 (MAE)、均方根误差 (RMSE) 和拟合系数(R^2) 3 种常用指标以评估预测性能,定义如下:

MAE = 1/N ∑ |y_i - ŷ_i| (8)

RMSE = √(1/N ∑ (y_i - ŷ_i)^2) (9)

R^2 = 1 - ∑ (y_i - ŷ_i)^2 / ∑ (y_i - ȳ)^2 (10)

式中:y_i、ŷ_i—第 i 个样本的真实值与预测值;ȳ—真实值的样本均值;N—样本总数。

3.3 持续学习方法对比

为评估不同训练策略在风电超短期功率预测中的性能,本文在统一的 LSTM 模型框架下比较了批量训练、EWC、MAS 和 IE-MAS 持续学习策略。其中,批量训练 (Batch) 一次性使用全跨度数据进行参数更新,通过多轮迭代学习全局模式;EWC 策略在损失函数中引入基于费舍尔信息矩阵的二次型正则项,用以抑制对关键参数的剧烈更新;MAS 策略通过梯度敏感度估计参数重要性,并对增量更新施加加权约束以强化历史知识保持;IE-MAS 在 MAS 基础上融入信息增强策略,在约束关键参数的同时向模型注入历史知识特征,进一步优化新旧知识的平衡。模型以时刻 t 及其前 15 个周期(t-15, ..., t)的风速、风向与功率输出为输入,预测随后 16 个周期(t+1, ..., t+16)的功率,严格满足“4 h 内、15 min 分辨率”的业务规范,不同策略的功率预测曲线的直观对比,如图 4 所示。其中,轮次 1 为采用初始 50% 数据训练的基准模型;轮次 2~4 则是在此基础上,每轮以 10% 的数据增量进行序贯训练后的模型状态,所有模型均在固定的测试集上进行性能评估。

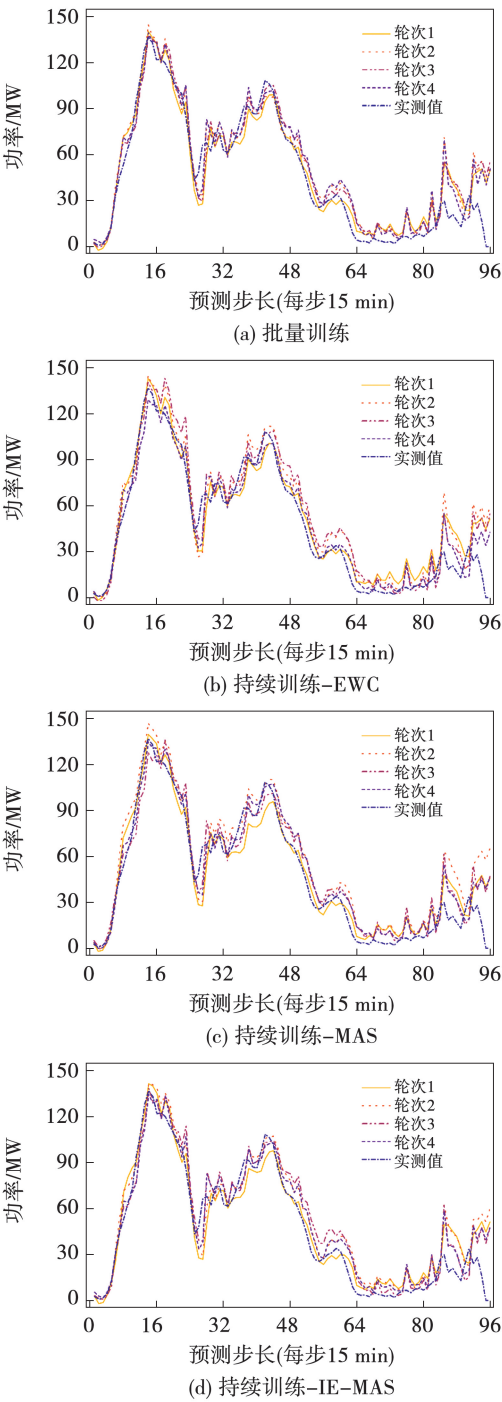


图 4 不同训练策略预测功率曲线
Fig.4 Predicted power curves under different training strategies

表 2 和图 5 分别呈现了各策略在 MAE、RMSE 及 R^2 指标的定量评价结果。量化结果表明,IE-MAS 在四轮增量中始终保持最优性能,其 RMSE 从 37.592 降至 36.953 (提升约 1.70%)、R^2 从 0.636 提升至 0.648 (提升约 1.89%)、MAE 从 26.459 降至 25.805 (提升约 2.47%),指标变化均呈单调优化趋势。

表 2 不同训练策略迭代的评价指标对比

Tab.2 Comparison of evaluation metrics across iterations for different training strategies

评价指标	训练策略	迭代轮次			
		1	2	3	4
RMSE	Batch	37.648	37.396	37.281	37.001
	EWC	37.591	38.158	38.105	37.692
	MAS	37.740	37.949	37.819	37.630
	IE-MAS	37.592	37.382	37.279	36.953
R^2	Batch	0.635	0.640	0.642	0.647
	EWC	0.636	0.625	0.626	0.634
	MAS	0.633	0.629	0.632	0.635
	MAS	0.636	0.640	0.642	0.648
MAE	Batch	26.493	25.957	25.855	25.828
	EWC	26.541	26.457	26.542	27.384
	MAS	26.609	26.326	26.631	27.230
	IE-MAS	26.459	25.891	25.834	25.805

批量训练同样实现了 3 个指标的协同改善,但数值略逊于 IE-MAS;相比之下,EWC 和 MAS 在第二轮增量后出现了全方位的性能回退。其中,EWC 的 RMSE 从 37.591 上升至 38.158,MAE 从 26.541 回落至 26.457, R^2 从 0.635 99 下降至 0.624 92; MAS 则在相同阶段亦出现 RMSE 上升 0.56%,MAE 上升 0.64% 及 R^2 下降 0.62% 的恶化,表明纯参数正则策略在应对新数据时存在局限,反映了仅靠参数正则无法完全抑制灾难性遗忘。而 IEMAS 得益于上述三重机制的协同作用,不仅从参数层面阻断遗忘,更在表征与输出层面同步修正偏移,实现多维度指标的优化,避免了 EWC 与 MAS 策略中常见的性能回弹。

EWC 与 MAS 均以最小化 MSE 为训练目标,通过费舍尔信息或梯度敏感度对关键参数加以约束,虽能迅速压低极端误差,但缺乏对模型输出分布与表征空间的全局一致性维护。当第二批增量数据介入时,上述策略在局部参数层面固化了旧知识,反而导致大批中等偏差样本的预测出现系统性偏移,致使 MAE 出现遗忘效应,而 R^2 、RMSE 也被少数未完全抑制的极端误差同步拉升。

当新增数据量增大时,EWC 仅保持费舍尔信息高的参数更新,MAS 只惩罚高敏感度参数,因而大量低敏感度方向的参数漂移,使中等误差样本预测偏离原分布,导致 MAE 反弹。同时,二者均未维护旧样本的输出分布,旧知识虽然在参数层面保留,但在表征与输出层面已出现漂移。IE-MAS 则从参数、特征与输出 3 个维度同步抑制偏移,实现了 RMSE 与 MAE 的双重优化。

3.4 多维时序性神经网络对比验证

表 3 汇总了 4 种模型分别采用批量训练与 IE-MAS 框架,在经历三次增量学习阶段(轮次 2 ~ 4)后的平均性能指标(RMSE、 R^2 和 MAE)。对于 LSTM、GRU 和 Transformer,IE-MAS 在仅使用每轮 10% 新增数据的情况下,与批量训练(累积利用全部历史数据)的性能基本持平,平均 RMSE 改进幅度小于 0.2%,MAE 变化也在 0.1% 以内, R^2 同步保持或微幅提升;而在 TCN 架构上,IE-MAS 表现出更优越的性能,其平均 RMSE 从 37.15 降至 36.40(降幅约 2.02%),MAE 从 26.38 降至 25.66(降幅约 2.73%), R^2 从 0.64 提升至 0.66(升幅约 2.20%)。

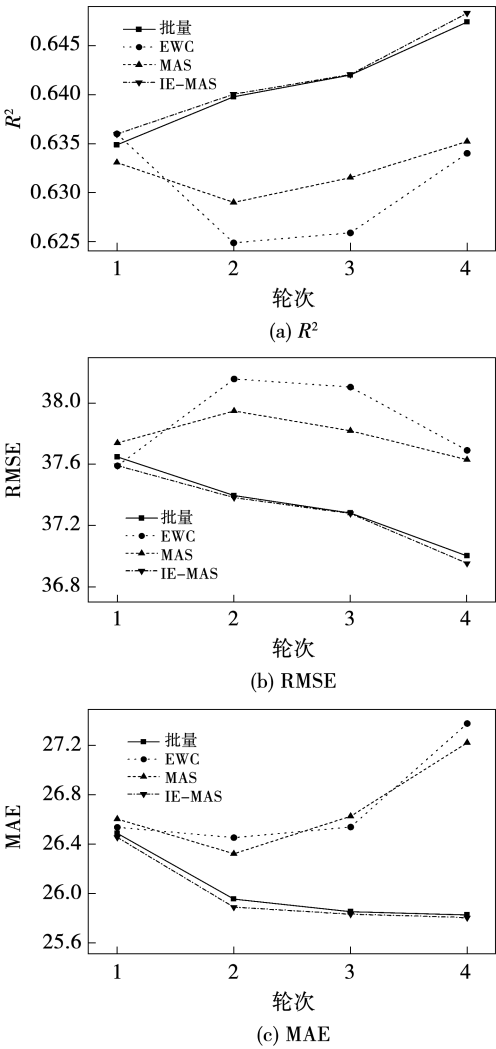


图 5 不同训练策略迭代的评价指标曲线

Fig.5 Curves of evaluation metrics across iterations for different training strategies

表 3 不同时序性神经网络的持续学习评价指标对比

Tab.3 Comparison of evaluation metrics for continual learning across different sequence-based neural network

算法模型	训练策略	评价指标		
		RMSE	R^2	MAE
LSTM	Batch	37.226	0.643	25.880
	IE-MAS	37.205	0.643	25.843
GRU	Batch	37.405	0.640	26.498
	IE-MAS	37.345	0.641	26.492
Transformer	Batch	38.055	0.627	26.454
	IE-MAS	37.998	0.628	26.439
TCN	Batch	37.150	0.644	26.378
	IE-MAS	36.403	0.659	25.657

相较于 LSTM、GRU 和 Transformer,TCN 的核心优势在于,通过扩张卷积和残差连接捕捉长程依赖,但这也使得其对卷积核参数的微小漂移极为敏感。IE-MAS 能够对关键卷积核施加更强保护,并利用知识蒸馏对输出分布进行对齐,从而在最少增量数据下最大程度地保留多层时域特征,因而在长期依赖建模上获得了比其他网络更显著的性能提升。

以上结果表明,IE-MAS 仅利用每轮新增的 10% 数据就能在各类时序网络中抵御灾难性遗忘,并保持甚至超越需要重复训练全部历史样本的批量训练策略的性能。对比中,批量训练必须累积使用 60% ~ 80% 的所有先前数据才能实现相似效果,而 IE-MAS 则通过多维度信息增强手段,从参数空间、特征表征到输出分布 3 个维度同步抑制遗忘,实现了对极端误差和平均误差的双重优化,并提升了模型整体拟合度。

图 6 为在第三个增量阶段上,4 种时序网络在批量训练与 IE-MAS 策略下的平均单轮训练时间及相对缩减比例。LSTM 的平均耗时由 3.38 s 降至 1.17 s,节省 65.4%;GRU 由 3.28 s 降至 1.03 s,节省 68.7%;Transformer 由 6.66 s 降至 1.88 s,节省 71.9%;TCN 由 6.63 s 降至 1.94 s,节省 70.7%。以上结果表明,IE-MAS 通过在每轮仅引入 10% 新增数据的前提下,结合增强机制不仅有效抑制了灾难性遗忘,保持批量训练的预测性能,并且将训练开销平均减少了约 65% ~ 72%。在需要频繁在线更新的超短期风电功率预测场景中,IE-MAS 的速度与算力需求优势为部署实时增量学习提供保障。

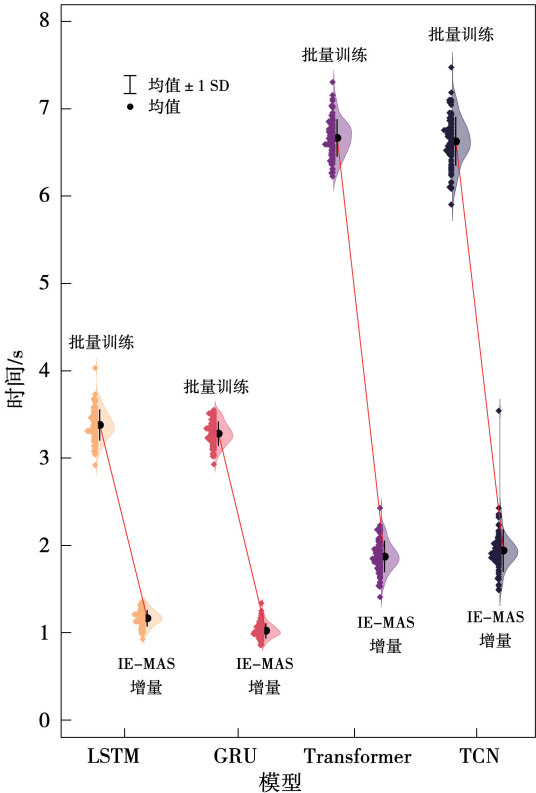


图 6 不同时序性神经网络的持续学习训练时间对比

Fig.6 Comparison of continual learning training time across different sequence-based neural network

4 结 论

- (1) 在每轮仅引入 10% 新增数据的条件下, LSTM 上 RMSE 下降约 1.7%,MAE 下降约 2.5%, R^2 提升约 1.9%。
- (2) 跨模型对比 (LSTM、GRU、Transformer、TCN) 场景下,IE-MAS 的预测精度略优于全量批量训练。
- (3) 单轮训练耗时压缩至原来的约 28% ~ 35%,显著减少在线更新的算力开销。该方法为超短期风电预测及其他时序持续学习场景提供了兼顾效率和精度的实时更新方案。

参考文献:

[1] 杜 刚,赵冬梅,刘 鑫.计及风电不确定性优化调度研究综述[J].中国电机工程学报,2023,43(7):2608-2627.
DU Gang,ZHAO Dongmei,LIU Xin. Research review on optimal scheduling considering wind power uncertainty[J]. Proceedings of the CSEE,2023,43(7):2608-2627.

[2] 朱琼锋,李家腾,乔 骥,等.人工智能技术在新能源功率预测的应用及展望[J].中国电机工程学报,2023,43(8):

3027 – 3048.

ZHU Qiongfeng, LI Jiateng, QIAO Ji, et al. Application and prospect of artificial intelligence technology in renewable energy forecasting[J]. Proceedings of the CSEE, 2023, 43(8): 3027 – 3048.

[3] 国家能源局. 国家能源局关于印发《电力并网运行管理规定》的通知 [EB/OL]. zfxgk. nea. gov. cn/2021 – 12/21/c _ 1310391369. htm.

National Energy Administration. Notice on issuing the " regulations on the administration of grid-connected power operation " [EB/OL]. zfxgk. nea. gov. cn/2021 – 12/21/c _ 131039-1369. htm.

[4] 国家能源局. 国家能源局关于印发《电力辅助服务管理办法》的通知 [EB/OL]. zfxgk. nea. gov. cn/2021 – 12/21/c _ 1310391161. htm.

National Energy Administration. Notice on issuing the " measures for the administration of power ancillary services " [EB/OL]. zfxgk. nea. gov. cn/2021 – 12/21/c_1310391161. htm.

[5] 张 璐, 杨培宏, 刘景霞. 基于 GA-LSSVM 与 ARIMA 组合的短期风电功率预测[J]. 电子技术, 2022, 51(10): 115 – 117.

ZHANG Lu, YANG Peihong, LIU Jingxia. Prediction of short-term wind power based on combination of GA-LSSVM and ARIMA[J]. Electronic Technology, 2022, 51(10): 115 – 117.

[6] 谢 波, 高建宇, 张惠娟, 等. 短期风电功率预测中的 IOFA-SVM 算法实现[J]. 电子测量技术, 2021, 44(12): 63 – 69.

XIE Bo, GAO Jianyu, ZHANG Huijuan, et al. Improved optimal foraging algorithm for support vector machine of short-term wind power prediction [J]. Electronic Measurement Technology, 2021, 44(12): 63 – 69.

[7] 彭嘉宁, 徐鹤勇. 基于随机森林和支持向量回归的风力发电预测算法[J]. 热能动力工程, 2024, 39(5): 143 – 149.

PENG Jianing, XU Heyong. Wind power prediction algorithm based on random forest and support vector regression[J]. Journal of Engineering for Thermal Energy and Power, 2024, 39(5): 143 – 149.

[8] 戴朝辉, 陈 昊, 刘莘轶, 等. 基于 K-medoids-GBDT-PSO-LSTM 组合模型的短期光伏功率预测[J]. 太阳能学报, 2025, 46(1): 654 – 661.

DAI Zhaohui, CHEN Hao, LIU Xinyi, et al. Short-term photovoltaic power prediction based on K-medoids-GBDT-PSO-LSTM combined model[J]. Acta Energiæ Solaris Sinica, 2025, 46(1): 654 – 661.

[9] 张建辉, 滕 婕, 李秀慧, 等. 基于气象特征和改进 Transformer 的光伏功率短期预测[J]. 热能动力工程, 2024, 39(8): 146 – 154.

ZHANG Jianhui, TENG Jie, LI Xiuhu, et al. Short-term prediction of photovoltaic power based on meteorological features and improved transformer[J]. Journal of Engineering for Thermal Energy and Power, 2024, 39(8): 146 – 154.

[10] 陈 峰, 余 轶, 徐敬友, 等. 基于 Bayes-LSTM 网络的风电出力预测方法[J]. 电力系统保护与控制, 2023, 51(6): 170 – 178.

CHEN Feng, YU Yi, XU Jingyou, et al. Prediction method of wind power output based on a Bayes-LSTM network[J]. Power System Protection and Control, 2023, 51(6): 170 – 178.

[11] 刘新宇, 蒲欣雨, 李继方, 等. 基于贝叶斯优化的 VMD-GRU 短期风电功率预测[J]. 电力系统保护与控制, 2023, 51(21): 158 – 165.

LIU Xinyu, PU Xinyu, LI Jifang, et al. Short-term wind power prediction of a VMD-GRU based on Bayesian optimization[J]. Power System Protection and Control, 2023, 51(21): 158 – 165.

[12] WANG L, ZHANG X, SU H, et al. A comprehensive survey of continual learning: Theory, method and application [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5362 – 5383.

[13] LI A, ZHANG C, XIAO F, et al. Large-scale comparison and demonstration of continual learning for adaptive data-driven building energy prediction[J]. Applied Energy, 2023, 347: 121481.

[14] 袁 坤, 张秀华, 溥 江, 等. 非平稳数据流下的持续学习灾难性遗忘问题求解策略综述[J]. 计算机应用研究, 2023, 40(5): 1292 – 1302.

YUAN Kun, ZHANG Xiuhua, PU Jiang, et al. Review of catastrophic forgetting problems solving strategies for continual learning under non-stationary data streams [J]. Application Research of Computers, 2023, 40(5): 1292 – 1302.

[15] 严璐晗, 林培杰, 程树英, 等. 基于增量学习的 CNN-LSTM 光伏功率预测[J]. 电气技术, 2024, 25(5): 31 – 40.

YAN Luhan, LIN Peijie, CHENG Shuying, et al. CNN-LSTM photovoltaic power prediction based on incremental learning[J]. Electrical Engineering, 2024, 25(5): 31 – 40.

[16] ALJUNDI R, BABILONI F, ELHOSEINY M, et al. Memory aware synapses: Learning what (not) to forget [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 139 – 154.

[17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735 – 1780.

[18] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. (2014 – 12 – 11) [2025 – 04 – 25]. https: // doi. org/10. 48550/arXiv. 1412. 3555.

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // International Conference on Neural Information Processing Systems, Long Beach, 2017: 5998 – 6008.

[20] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. (2018 – 03 – 04) [2025 – 04 – 25]. https: // doi. org/10. 48550/arXiv. 1803. 01271.