

贝叶斯优化超参数的空时融合压缩残差网络在 风速区间预测中的研究

伍耘¹, 葛佳敏¹, 王文烨¹, 李小勇², 车亮¹

(1. 湖南大学电气与信息工程学院, 湖南 长沙 410082; 2. 华南师范大学环境学院, 广东 广州 510630)

摘要: 针对风电场规划中风速的高随机性问题, 提出了一种基于小样本空时融合压缩残差网络点预测 (spatio-temporal integration and compression deep residual, STiCDRS)模型。该模型旨在深入挖掘风速序列中的空间和时间特征, 以提升点预测精度。首先, 采用空时融合压缩残差网络点预测模型得到点预测结果。然后, 在此基础上采用新颖的空时融合压缩残差网络区间(STiCDRS-Gaussian process regression, STiCDRS-GPR)预测模型得到风速的区间预测结果, 进而得到更为可靠的风速概率预测结果。该模型采用贝叶斯优化方法进行超参数选择, 确保超参数的高效自动化调优。最后, 使用内蒙古地区风电场的风速数据集, 将 STiCDRS 模型与传统经典模型的预测结果进行对比。实验结果表明, 相较于其他模型, 所提 STiCDRS-GPR 模型在风速预测中具有更高的点预测精度、适宜的预测区间以及可靠的概率预测结果, 充分展示了其在风速预测领域的良好应用潜力。

关键词: 风速预测; 时序卷积网络; STiCDRS 模型; GPR 区间预测; 贝叶斯优化

Wind speed interval prediction using spatio-temporal fusion compressed residual networks with Bayesian optimized hyperparameters

WU Yun¹, GE Jiamin¹, WANG Wenyue¹, LI Xiaoyong², CHE Liang¹

(1. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China;

2. School of Environment, South China Normal University, Guangzhou 510630, China)

Abstract: To address the challenge of high wind speed variability in wind farm planning, a small-sample-based spatio-temporal integration and compression deep residual point prediction model, spatio-temporal integration and compression deep residual (STiCDRS), is proposed. This model is designed to deeply explore the spatial and temporal characteristics within wind speed sequences to enhance the accuracy of point prediction. Initially, the spatio-temporal integration and compression deep residual network is employed to obtain point prediction results. Subsequently, an innovative hybrid model, STiCDRS-Gaussian process regression (STiCDRS-GPR), is introduced to achieve interval prediction results, thereby providing more reliable probabilistic forecasts of wind speed. The model utilizes a Bayesian optimization method for hyperparameter selection, ensuring efficient and automated tuning. Finally, the wind speed dataset from a wind farm in Inner Mongolia is used to compare the prediction results of the STiCDRS model with those of traditional classical models. Experimental results demonstrate that, in comparison to other models, the proposed STiCDRS-GPR model delivers superior point prediction accuracy, appropriate interval predictions, and reliable probabilistic forecasting outcomes, fully showcasing its considerable potential in the domain of wind speed forecasting.

This work is supported by the National Key Research and Development Program of China (No. 2021YFB2601504).

Key words: wind speed prediction; temporal convolutional network; STiCDRS model; GPR interval prediction; Bayesian optimization

0 引言

随着全球对可再生能源需求的增长, 风能因其

可再生和清洁的特性成为替代传统能源的重要资源^[1]。风力发电技术的核心是将风能转化为电能, 而其发电功率直接受风速影响。因此, 精准预测风速显得尤为重要。准确的风速预测不仅有助于优化电网调度, 还能提高风电场容量的评估精度, 从而

为确定风电上网价格提供科学依据, 具有重要的现实应用价值^[2]。然而, 风能的波动性和随机性给风速预测带来极大挑战。

传统的风速预测方法主要依赖于物理驱动模型^[3]或统计方法^[4], 这些方法在处理大规模或复杂数据时效果显著, 但面临以下问题: 1) 模型训练困难, 难以从有限的数据中学习到风速的复杂变化规律; 2) 过拟合风险^[5], 过度拟合训练数据中的随机波动或噪声, 导致在不同数据集上的表现存在显著差异; 3) 预测不确定性增加, 由于数据量的限制, 模型对于风速波动性的预测往往伴随着较高的不确定性。

相较于传统统计方法, 机器学习方法通过特征学习挖掘数据的内在联系, 能更加有效地处理复杂非线性关系^[6], 从而提高预测的准确性和适应性。常用的机器学习方法有高斯过程回归(Gaussian process regression, GPR)^[7]、K 最近邻算法(K-nearest neighbors, KNN)^[8]、随机森林回归(random forest regressor, RFR)^[9]、支持向量机回归(support vector regression, SVR)^[10]等。相较于这些浅层学习模型, 深度学习模型采用了分布式、层次化的特征表示方法, 能自动从数据的最低层到最高层提取固有的抽象特征和隐藏的不变结构^[11]。这种出色的特征提取能力使深度学习模型在风速和风电功率预测领域备受学者关注。在风电领域, 时序卷积网络(temporal convolutional network, TCN)^[12]使用非常广泛, 为处理时序预测任务提供了一个更为高效且稳定的框架^[13]。TCN 已经逐步应用于电力负荷预测、风速预测和光伏预测等领域。如文献[14]提出了一种基于时间卷积网络的短期风速预测方法, 比较结果显示该模型具有较高的预测准确率和最佳的预测效果。

尽管 TCN 在时序数据处理方面表现出色, 但也存在一些不足: 1) 在小样本场景下, 这种深度学习自动特征提取能力受到限制, 且其序列化的数据处理特性导致计算效率低下, 反而增加了参数优化和模型应用的复杂度; 2) 在处理包含大量冗余信息或噪声的序列数据时, 模型会将这些无关信息纳入预测过程, 导致预测精度下降; 3) 模型超参数的设定依赖于先前研究或实践中的经验进行简单试凑, 适应性差、效率低, 这对于处理具有非线性趋势、强随机性或尺度变化的序列数据来说是一个巨大的挑战。

此外, 传统的风速预测研究主要侧重于点预测, 即预测未来特定时刻的风速值。对于小样本数据集而言, 点预测面临的重大问题之一就是高不确定性。鉴于风速的波动性和环境的多变性, 加之数据点数

量有限, 导致预测结果的可信度下降。仅依赖点预测难以提供充分的决策依据, 限制了其在能源系统规划与运营中的实际应用。为了克服这些局限性, 区间预测应运而生, 区间预测通过预测一个可能包含未来风速值的范围, 而非单一数值, 为量化未来风速的不确定性提供了一种有效手段, 从而提升了预测结果的可靠性^[15]。常用的区间预测方法有自举法^[16]、线性回归置信区间^[17]、高斯过程回归^[18]、自回归移动平均(autoregressive moving averages, ARMA)模型^[19]及其变体等。

针对上述问题, 本文提出了一种基于改进 TCN 的 STiCDRS-GPR 预测模型。首先, 提出了一种端到端模型进行风速预测, 将 TCN 中的传统残差结构 1×1 卷积层替换为共享权系数的长短时记忆网络(shared weight long-term memory network, SWLSTM)^[20]; 其次, 在改进的 TCN 特征输出层之后引入了基于注意力机制的软阈值残差结构(deep residual shrinkage network, DRSN); 此外, 本文采用贝叶斯优化方法, 实现模型超参数的高效、自动化调优; 最后, 在改进 STiCDRS 点预测结果的基础上, 通过高斯过程回归生成了区间预测结果。结果表明, 本文所提 STiCDRS-GPR 预测模型, 显著提高了风速预测精度。

1 基于贝叶斯优化的改进 STiCDRS 风速点预测模型

1.1 时序卷积网络

TCN 是一种利用膨胀因果卷积设计的神经网络架构^[12], 后续点预测模型的构建将基于 TCN 模型进行结构重构, 可大量减少模型的参数量, 其最核心的结构包括膨胀因果卷积和残差连接结构, 如图 1 所示。

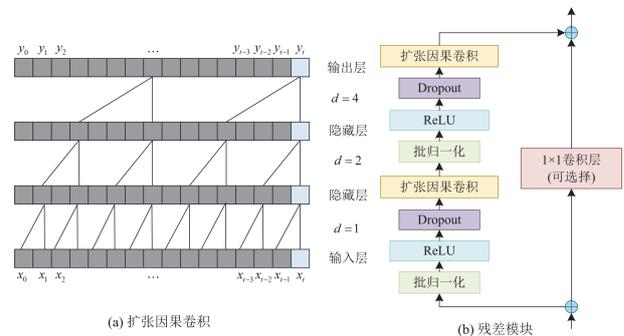


图 1 TCN 结构图

Fig. 1 Structure diagrams of TCN

在一维时间序列的情况下, 因果卷积的计算方式表示为

$$y_t = \sum_{m=1}^k \omega_m \cdot x_{t-m+1} \quad (1)$$

式中: y_t 是输出序列的第 t 个元素; x_{t-m+1} 是输入序列的第 $t-m+1$ 个元素; ω_m 是卷积核的第 m 个权重; k 是卷积核的大小。

膨胀卷积通过在卷积核中引入间隔来扩大卷积核的感受野, 且不增加卷积核的大小。给定一个输入序列 x , 一个大小为 k 的卷积核 ω 编辑器输出和一个膨胀率 d , 则膨胀卷积的计算方式表示为

$$y_t = \sum_{m=1}^k \omega_m \cdot x_t + (m-1) \cdot d \quad (2)$$

式中, d 为膨胀率。

一个残差模块包含两个膨胀因果卷积层, 每层还添加 WeightNorm 和 Dropout 来正则化网络, 另外还加入了一个 1×1 卷积来保持输入输出的尺度相同, 如图 1 所示。具体来说, 一个基本的残差块, 其输入为 x_{in} , 输出为 y_{out} , 那么残差连接的计算方式表示为

$$y_{out} = F(x) + x_{in} \quad (3)$$

式中: $F(x)$ 是残差块中的非线性变换; x_{in} 是输入到该残差块的原始输入。

1.2 深度残差收缩网络

DRSN 是 Zhao M 在 2020 年提出的一种改进的残差结构^[21]。为进一步提高预测精度, DRSN 将会用于进一步改进 TCN 的残差模块, 该模块通过结合 TCN 中的残差连接和基于注意力机制的软阈值残差进行组合滤波, 进一步提升了模型的特征提取和表示能力。

图 2(a) 为传统的残差结构, 图 2(b)、图 2(c) 为 DRSN 的两种变体, 即深度残差共享收缩网络(deep residual shrinkage network with channel-shared thresholds, DRSN-CS)和深度残差逐点收缩网络(deep residual shrinkage network with channel-wise thresholds, DRSN-CW)。相较于 DRSN-CS 模块, DRSN-CW 残差模块所获得的阈值是一个向量, 即每个通道对应一个收缩阈值, 因此本文采用的 DRSN 方法默认为 DRSN-CW。不同于传统残差模块, DRSN 在传统结构的基础上增加了软阈值和注意力机制。

软阈值的主要作用是将绝对值低于某个阈值的特征置零, 即将接近 0 的特征变换为零, 是一个将输入数据向零方向收缩的函数。软阈值化函数为

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau \leq x \leq \tau \\ x + \tau & x < -\tau \end{cases} \quad (4)$$

式中: x 是输入特征; y 是输出特征; τ 是阈值。

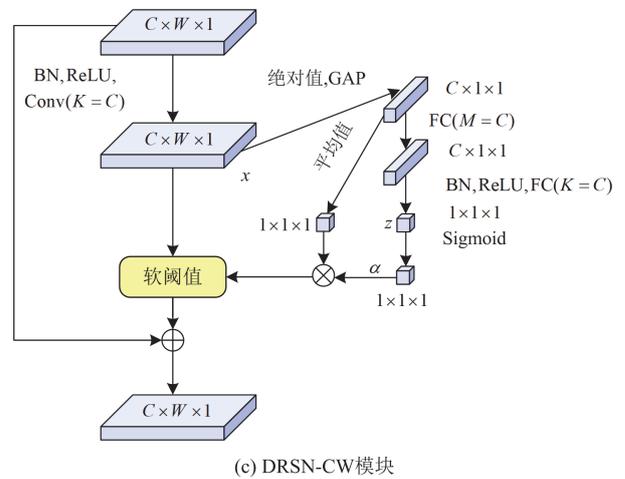
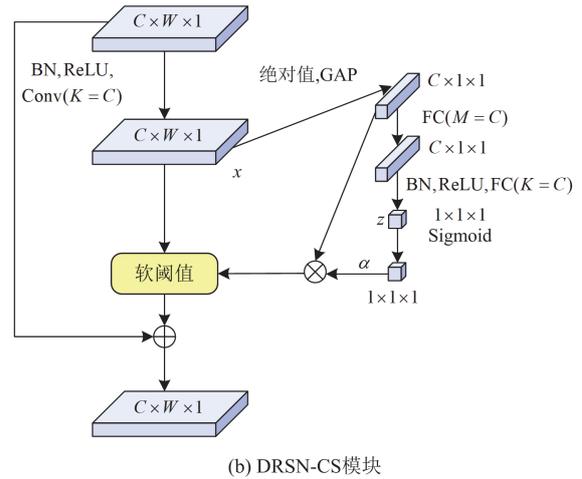
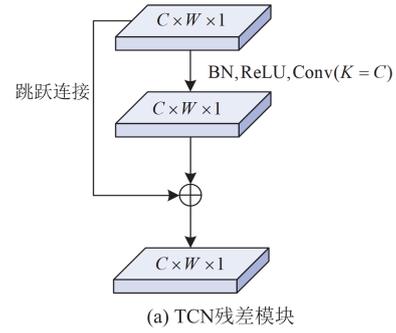


图 2 TCN 残差模块、DRSN-CS 模块与 DRSN-CW 模块结构图

Fig. 2 Structural diagrams of TCN residual module and DRSN-CS module in comparison with DRSN-CW module

压缩和激励网络(squeeze-and-excitation network, SENet)作为注意力机制的主要部分, 使模型能够自动学习一组权重系数, 将学习到的权重加权到特征图中的每个通道, 评估各个通道的重要程度, 从而使模型能够区分重要及冗余特征通道。在 SENet 中,

学习路径通过全局池、全连接层、ReLU 激活函数和全连接层，以 Sigmoid 函数结束。

在深度残差收缩网络中，SE 模块自动学习阈值并结合软阈值去噪，解决了阈值选取的困难。

1.3 贝叶斯优化

贝叶斯优化是一种常用的黑箱函数优化方法^[22]。可以有效解决超参数依赖以往经验简单试凑、适应性差的问题，实现了超参数的高效、自动化调优，贝叶斯优化流程图如图 3 所示。

高斯过程模型假设目标函数在参数空间中的任意点处的取值服从一个高斯分布，分布为

$$f(x) \sim N(\mu(x), k(x, x')) \quad (5)$$

式中： $f(x)$ 是模型输出函数； $\mu(x)$ 是均值函数； $k(x, x')$ 是协方差函数。

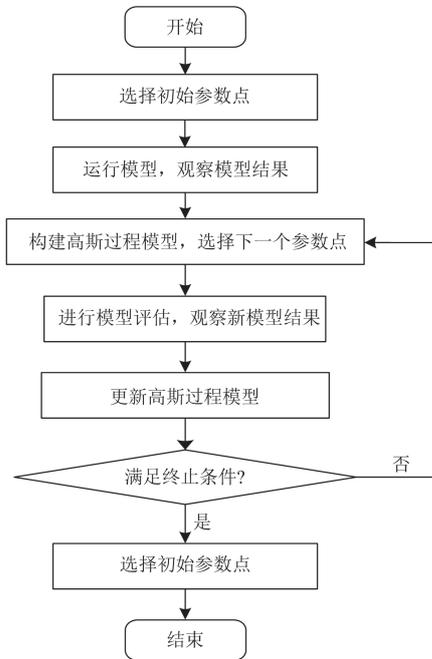


图 3 贝叶斯优化流程图

Fig. 3 Bayesian optimization flowchart

通过观察目标函数的一些样本点，根据已有的观测数据 $D = \{(x_i, y_i)\}_{i=1}^n$ ，利用贝叶斯定理更新模型的参数，得到参数的后验概率分布，表示为

$$p(f(x)|D) \propto p(y|x, D)p(f(x)|D) \quad (6)$$

式中： $p(y|x, D)$ 表示观测数据的似然函数； \propto 表示“正比于”，即左侧与右侧存在比例关系； $p(f(x)|D)$ 表示先前的后验分布。

后验概率反映了参数的不确定性以及在已有观测数据下参数的分布情况。采样策略根据后验概率分布选择下一个最有可能的极值点。贝叶斯优化算法通

过迭代更新后验分布和选择下一个参数点进行评估，逐步优化目标函数，直至达到预设的终止条件。

1.4 点预测模型构建

在 TCN 的传统残差结构中， 1×1 的卷积层存在一些缺陷， 1×1 卷积层只进行通道数的调整，没有明确的记忆能力，可能会导致一些重要的时间序列信息在传播过程中丢失。 1×1 卷积层的感受野非常小，只能捕获局部信息，对于长期依赖关系的建模能力相对较弱，局部感知能力受限。因此，本文提出了使用 SWLSTM^[20] 替代 1×1 卷积网络，得到新的网络 STiC。

然而，STiC 结构仍然有一些局限性：难以处理噪声和冗余的特征信息，风速数据本身的特征无法进行重要程度筛选。针对这两个缺点，本文对初步改进的 STiC 残差模块进一步优化：在改进 STiC 的特征输出层之后加入 DRSN 模块，提出基于空时特征融合和残差注意力软阈值机制的改进深度学习模型，本文将命名为 STiCDRS，用于风速的点预测，其结构如图 4 所示。

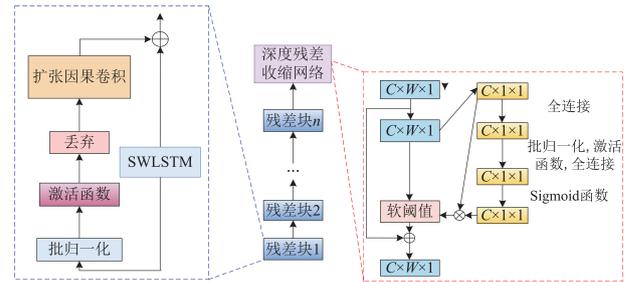


图 4 STiCDRS 结构示意图

Fig. 4 Structural diagrams of the STiCDRS

针对提出的点预测模型，预测过程如图 5 所示。

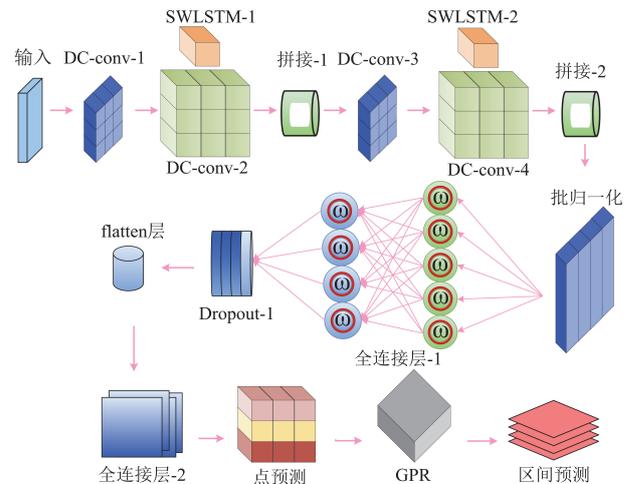


图 5 STiCDRS 点预测流程图

Fig. 5 STiCDRS point prediction flowchart

具体过程为: 输入数据通过第 1 个深度卷积层 (DC-conv-1), 接着同时通过第 2 个深度卷积层 (DC-conv-2) 和第一个共享权重系数长短时记忆网络 (SWLSTM-1), 进一步提取高级特征; 将 DC-conv-1 和 SWLSTM-1 的输出与 DC-conv-2 的输出进行拼接, 拼接后的数据流经第 3 个深度卷积层 (DC-conv-3), 进一步加工特征; 数据依次通过另一个特殊的循环单元 (SWLSTM-2) 和第 4 个深度卷积层 (DC-conv-4), 以增强模型对时间序列数据的处理能力并提取更深层次的特征; 将 SWLSTM-2 和 DC-conv-4 的输出进行拼接, 进一步融合特征信息; 通过批量归一化处理拼接后的数据, 加快训练速度, 提高模型稳定性; 数据流通过全连接层, 将特征进行整合; 经过丢弃层减少过拟合, 将数据展平为一维, 以适配接下来的全连接层; 数据通过另一个全连接层, 最终输出点预测结果。

2 基于贝叶斯优化的改进 STiCDRS-GPR 风速区间预测模型

2.1 GPR 区间预测

GPR 是一种基于贝叶斯理论和统计学习理论的机器学习方法^[23], 针对风速的高不确定性, GPR 不仅预测每个点的预测值, 还可以得到区间预测结果, 直观量化了预测结果的不确定性。

在 GPR 中, 假设目标函数 $g(x)$, 在参数空间中任意点 x 处的取值都符合一个高斯分布, 高斯过程定义为

$$g(x) \sim N(m(x), k(x, x')) \quad (7)$$

式中, $m(x)$ 是均值函数, 表示在位置 x 处的期望值。

2.2 区间预测模型构建

高斯过程回归能够支持模型关于预测的不确定性(置信区间), 直接输出关于单点预测值的概率分布, 但其在单点预测方法上属于机器学习范畴, 非常依赖输入信息的特征工程^[24]。结合深度学习和 GPR 进行区间预测的方法是将 GPR 模型中的参数和核函数设定为深度学习的输入和隐藏层, 并利用深度学习的非线性特性来提高模型的预测精度和泛化能力。同时, 深度学习模型也可以通过学习大量的数据, 自动发现和提取风速数据信息中的隐含特征, 从而提升区间预测性能, 同时满足对于精准性和置信度的需求^[25]。

因此, 获取高精度点预测、高可靠区间预测和概率预测的思路为: 结合 STiCDRS 点预测优势提出 STiCDRS 融合 GPR 的可靠概率预测性能, 即 STiCDRS-GPR 的风速区间预测方法。其中 STiCDRS-GPR 模型的训练和测试过程如图 6 所示。

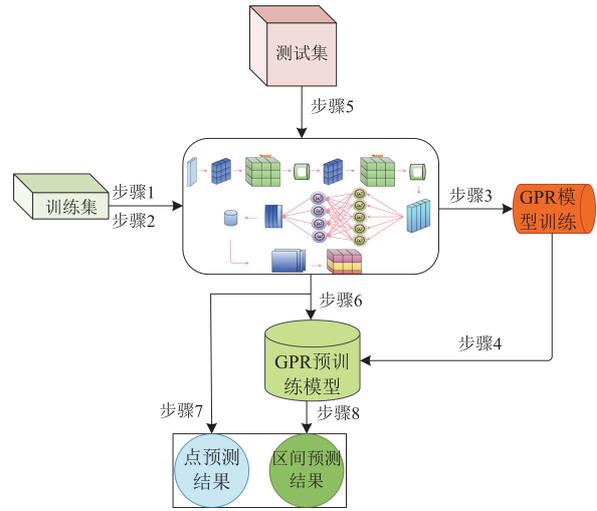


图 6 STiCDRS-GPR 训练和测试流程图

Fig. 6 Flow chart of STiCDRS-GPR training and testing process

该方法的具体步骤如下。

步骤1: 完成STiCDRS模型在单点预测过程的训练;

步骤2: 将训练值输入训练后的STiCDRS模型获得输出;

步骤3: 将步骤2的输出作为GPR的训练样本, 同时采用原训练集对应的标签作为GPR的标签;

步骤4: 通过步骤3的数据训练得到GPR预训练模型;

步骤5: 将测试集输入STiCDRS模型;

步骤6: 得到模型的单点预测值;

步骤7: 将单点预测值输出并进行保存;

步骤8: 将该单点预测值输入 GPR 模型得到区间预测值。

该区间预测模型以STiCDRS模型获得的高精度点预测结果为基础, 因此在第一个点测量值和观测值之间构造GPR模型时, 可获得更可靠的区间预测范围和概率密度函数(probability density function, PDF), 进而得到最终的STiCDRS-GPR模型, 如图7所示。

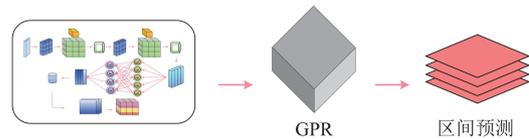


图 7 STiCDRS-GPR 模型的结构示意图

Fig. 7 Structural diagram of STiCDRS-GPR model

2.3 评价指标

点预测是区间预测的基础, 为充分验证所提区间预测模型的有效性, 本研究采用 3 个点评价指标^[26]、

3 个区间评价指标和 1 个计算效率评价指标对模型的性能进行评价,并用相同的评价指标对本文的区间预测模型及与该模型具有类似结构的经典模型进行对比。点预测评价方法主要涉及均方根误差(root mean square error, RMSE) E_{RMSE} 、平均绝对误差(mean absolute error, MAE) E_{MAE} 和皮尔逊相关系数(Pearson's correlation coefficient, R)。设预测值为 $h(x_i)$, 对应时刻真实值为 y_i , 则评价方法计算方式为

$$E_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |h(x_i) - y_i| \quad (8)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (h(x_i) - y_i)^2} \quad (9)$$

$$R = \frac{\sum_{i=1}^N (h(x_i) - \bar{h})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (h(x_i) - \bar{h})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (10)$$

式中: N 为测试样本数; \bar{y} 为目标值的平均值; \bar{h} 为预测值的平均值; $h(x_i)$ 为第 i 个样本的点预测值。

E_{MAE} 和 E_{RMSE} 的值越小表示模型的点预测性能越好, 而 R 取值范围为 $[-1, 1]$, 当 R 值接近 -1 或 1 时, 表示两个变量之间存在强烈的线性关系, 当其接近 0 时, 表示线性关系较弱或不存在。

本研究采用的区间预测性能评价指标包括覆盖率(coverage probability, CP)、综合评价指标和涉及概率预测的概率积分变换(probability integral transform, PIT)^[27]。

覆盖率 p_{CP} 定义为观测值落在置信水平 α 下的预测区间内的概率, 具体计算公式如式(11)所示。

$$p_{\text{CP}} = \frac{C_\alpha}{T_e} \quad (11)$$

式中: C_α 是观测值落在预测区间内的样本数; T_e 是测试集的样本总数。

假设区间足够宽, 很容易覆盖所有的观测值, 导致 p_{CP} 的值为 100% , 进而无法得到关于预测不确定性的有效信息, 使得预测区间没有实用价值。间隔宽度百分比(mean width percentage, MWP) p_{MWP} 为间隔宽度与观察值的比值, 如式(12)^[28]所示。

$$p_{\text{MWP}} = \frac{1}{T_e} \sum_{i=1}^{T_e} \frac{h_{\text{max}_i} - h_{\text{min}_i}}{h(x_i)} \quad (12)$$

式中, h_{max_i} 、 h_{min_i} 分别为第 i 个样本点的预测区间上、下限。

较为理想的区间预测结果应该同时具有较大的 p_{CP} 和较小的 p_{MWP} 。因此通过定义两者的综合评

价指标度量 p_{MC} 进行综合考量, 其计算公式如式(13)所示。

$$p_{\text{MC}} = \frac{p_{\text{MWP}}}{p_{\text{CP}}} \quad (13)$$

通过 p_{MWP} 与 p_{CP} 的比值可得: p_{MC} 越小, 所表示的区间预测性能越高。

可靠性是指预测和观察的统计一致性, i_{PIT} 的累计分布函数图用于评估概率预测的可靠性, 其计算公式如式(14)^[29]所示。

$$i_{\text{PIT}} = \int_{-\infty}^{h(x_i)} p(t) dt \quad (14)$$

式中, $p(t)$ 表示随机变量 t 的概率密度函数。

i_{PIT} 通常可以通过 P-P 图来展示, 即绘制概率积分变换(probability integral transform, PIT)的累计分布函数(cumulative distribution function, CDF)与标准均匀分布的 CDF 的关系图。预测的结果越准确, 所绘制的 P-P 图越接近直线。

平均训练时间(average training time, ATT) t_{ATT} 是模型在训练集上训练 5 次的平均时间, 单位为 s, 其用于衡量模型的训练效率, 训练时间较短的模型在处理大规模数据集时具有更高的实用性和效率。

计算公式为

$$t_{\text{ATT}} = \frac{1}{5} \sum_{n=1}^5 T_n \quad (15)$$

式中, T_n 是第 n 次训练的时间。

3 算例分析

3.1 数据集设置

为验证本文所提方法 STiCDRS 的有效性, 本实验研究中国内蒙古风电场 2016 年 3 月 20 日至 2016 年 5 月 19 日的 4 个数据集, 各风速数据步长为 15 min ^[30], 分别对应本文中的数据集 1、2、3、4。将前 8 个周期点作为输入, 预测当前点。数据集的训练集与测试集划分如图 8 所示。

数据集 1 从 2016 年 4 月 23 日 00:00 到 2016 年 4 月 28 日 07:00, 采样周期为 15 min , 共包含 481 个周期点。按照 8:2 的比例划分训练集和测试集, 取前 385 个周期点为训练集, 后 96 个周期点为测试集。数据集 2 从 2016 年 3 月 20 日 00:00 到 2016 年 3 月 27 日 00:00, 采样周期为 15 min , 共包含 673 个周期点, 按照 8:2 的比例划分训练集和测试集, 取前 538 个周期点为训练集, 后 135 个周期点为测试集。数据集 3 从 2016 年 4 月 11 日 00:00 到 2016 年 4 月 18 日 00:00, 采样周期为 15 min , 共包含 673 个周期点, 按照 8:2 的比例划分训练集和测试集, 取前 538 个周期点为训练集, 后 135 个周

期点为测试集。数据集 4 从 2016 年 5 月 3 日 07:00 到 2016 年 5 月 13 日 07:00, 采样周期为 15 min, 共包含 961 个周期点, 按照 8:2 的比例划分训练集和测试集, 取前 769 个周期点为训练集, 后 162 个周期点为测试集。

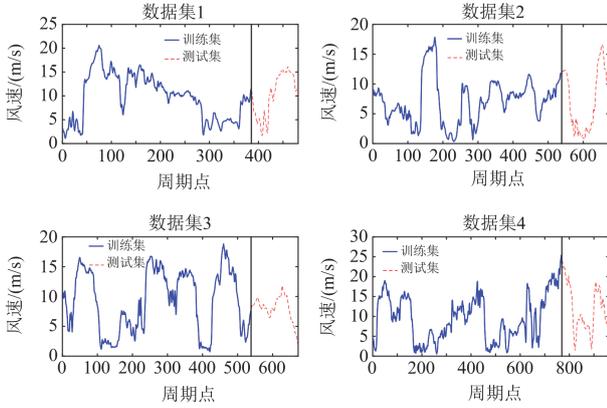


图 8 4 个风速数据集波形图训练集和测试集的划分示意图
Fig. 8 Schematic diagram for dividing the training and testing sets of four wind speed datasets waveform

3.2 参数配置

对于模型在训练过程中的参数设置, 使用 Kaiming 法对 CNN 层和全连接层的权系数进行初始化^[31], 使用正交方法对 LSTM 层的权系数进行初始化, 有利于加快模型的收敛速度。在网络训练过程中通过 Adam 优化器使模型收敛速度更快、鲁棒性更强。随机打乱处理后的数据集并以 8:2 划分为训练集和测试集, 具体设置如表 1 所示。

表 1 CNN、LSTM、CLSTM、TCN、STiCDRS 和 GPR 模型训练参数配置

Table 1 Training parameter configurations of CNN, LSTM, CLSTM, TCN, STiCDRS and GPR models

参数名称	参数配置	参数名称	参数配置
CNN 层初始化方法	Kaiming 法	训练测试比例	8:2
LSTM 层初始化方法	正交法	数据集个数	4
全连接层初始化方法	Kaiming 法	训练轮数	100
优化器方法	Adam	丢弃率	0.5
数据预处理方法	z-score	卷积层层数	2
损失函数	MSE	激活函数	ReLU

后续使用贝叶斯优化方法对卷积核大小、卷积核数量、全连接层的隐藏单元大小、批量大小、学习率、正则化系数进行自适应选择组合。

3.3 模型超参数选取实验

在机器学习和深度学习中, 模型的性能通常受到超参数的影响, 因此选择合适的超参数组合对于模型的性能至关重要。通常采用贝叶斯优化算法自适应地选择超参数组合。贝叶斯优化算法通过不断

地尝试不同的超参数组合, 并根据目标函数的输出结果更新超参数组合的概率模型, 从而使得算法能够自适应地选择出更有可能达到最优值的超参数组合。这种方法与传统的网格搜索或随机搜索相比能更加高效地搜索到最优的超参数组合。另外为了充分说明提出的 STiCDRS 在风速预测方面有效性, 本研究先通过贝叶斯优化对具有相似结构的 TCN 模型的关键参数(表 2 所示)进行组合优化, 寻找最优超参数组合, 然后设置其他模型和 STiCDRS 模型, 使其具有相同的超参数。CNN 模型最优的超参数组合如表 2 所示。

表 2 贝叶斯优化确定的 CNN 模型网络超参数

Table 2 Bayesian optimization-determined hyperparameters of CNN model network

超参数名称	取值
卷积核大小	3
卷积核个数	32
全连接层的隐藏单元大小	128
批量大小	64
学习率	0.005
正则化系数	0.9

设定 CNN 模型网络中卷积核大小寻优范围为 [1, 4], 卷积核个数寻优范围为 [8, 256], 全连接层隐藏单元大小寻优范围为 [64, 512], 批量大小寻优范围为 [8, 256], 学习率寻优范围为 [0.001, 0.01], 正则化系数寻优范围为 [0.1, 0.9]。基于贝叶斯优化对 CNN 模型进行超参数寻优, 计算返回最小损失值, 并记录产生最小损失的超参数组合, 从而得到最终的超参数组合表。

3.4 实验结果对比分析

基于上述参数对本研究选定的 4 个风速数据集进行仿真实验, 得到 CNN、LSTM、CLSTM、TCN、STiCDRS 和 GPR 模型在测试集上的预测值、观测值以及预测区间, 如图 9 所示。

图 10 中对应 4 个数据集的所有模型点预测和区间预测结果评价指标值的雷达图和条形图, 由雷达图可知, STiCDRS 模型在 6 个指标上的分布都优于其他模型, 对于 R、CP 两个正向指标, STiCDRS 模型对应圆的半径最大, 接近标出的 0.99, 可见其具有最优的预测效果, 对于其余 4 个反向指标, STiCDRS 指标最接近圆心, 绝大多数小于 0.3, 同样具有最好的预测效果。条形图中, 红色、黄色柱条代表 2 个正向指标, 在 4 个数据集的表现上, STiCDRS 模型的柱条值全部大于 0.9, 基本上高于其他所有模型, 具有最好的预测效果; 绿色、深蓝色、浅蓝色、灰色柱条代表 4 个反向指标, STiCDRS

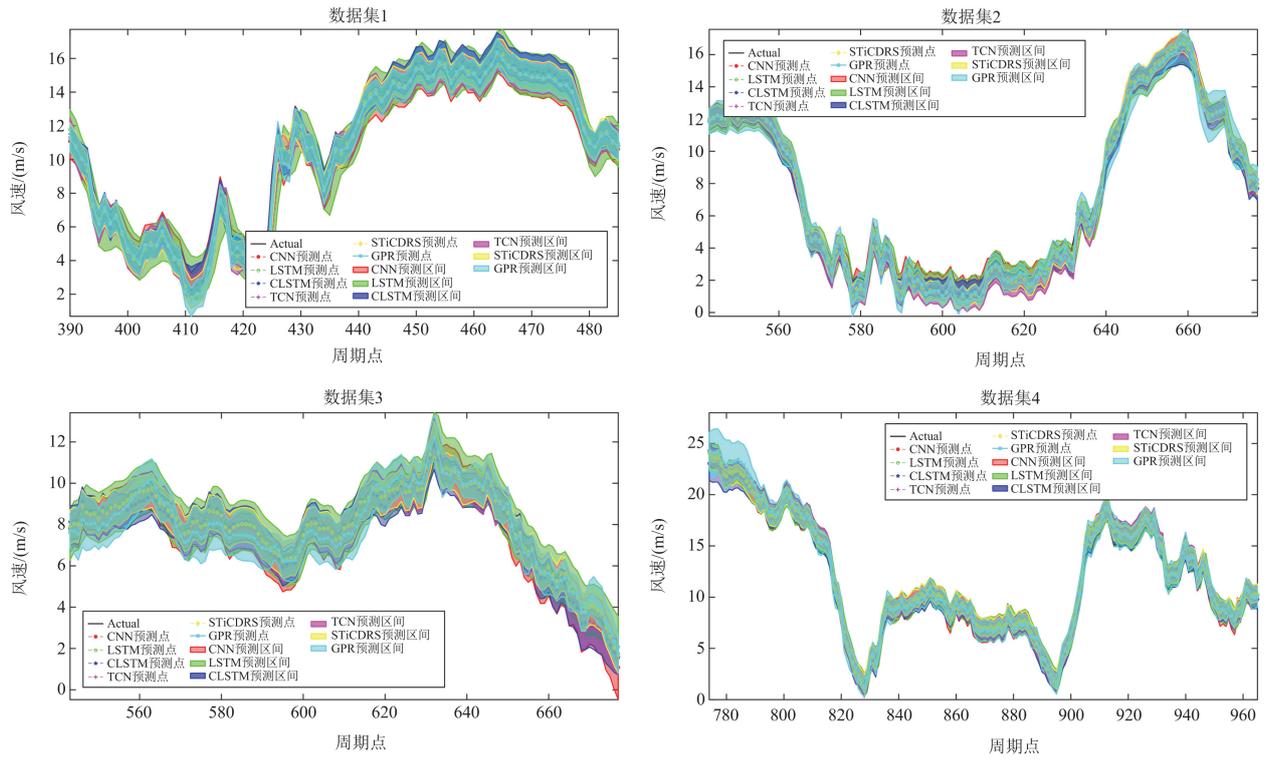


图 9 所有对比模型预测对应 4 个数据集测试集的点预测、观测值和区间预测结果图
Fig. 9 All comparison model predictions correspond to the point prediction, observed value and interval prediction results of the 4 test datasets

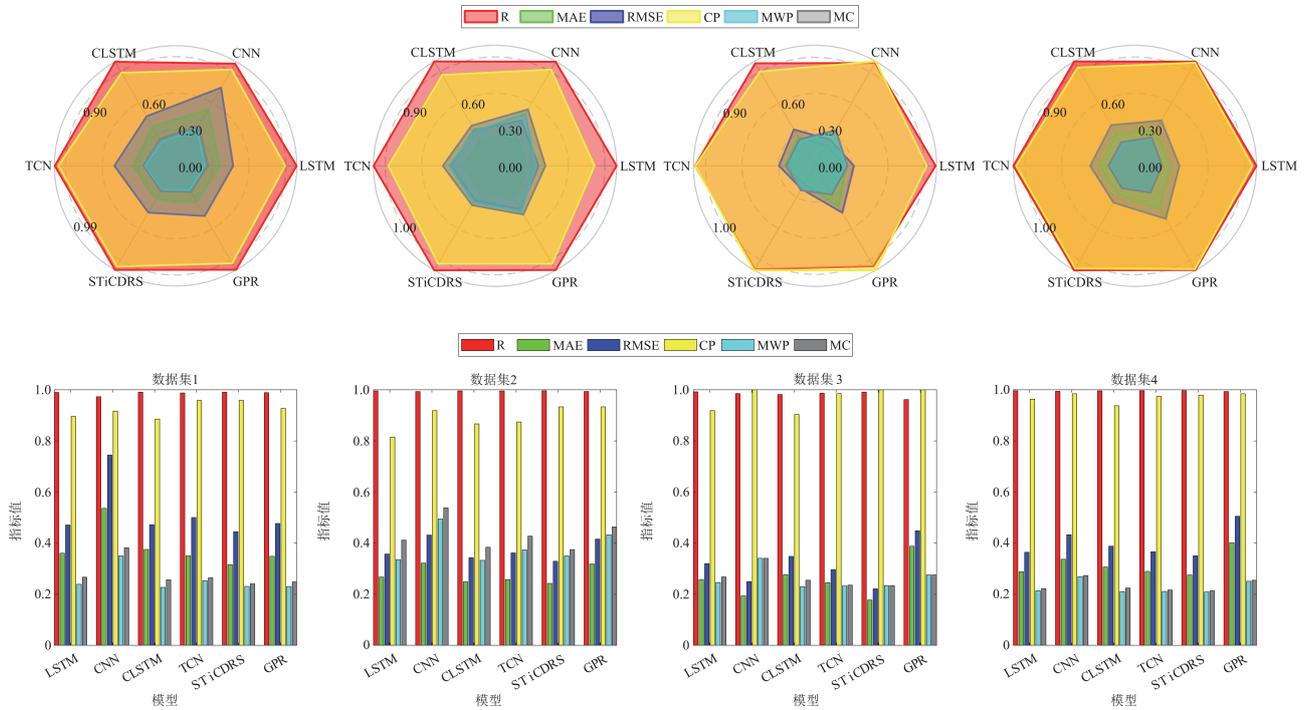


图 10 对应 4 个数据集的所有模型点预测和区间预测结果评价指标值的雷达图和条形图
Fig. 10 Radar and bar charts of evaluation index values for all model point predictions and interval predictions corresponding to four datasets

模型指标值在数据集 1、2、4 上几乎全小于 0.4, 且普遍优于其他模型, 在数据集 3 上的预测效果最好, 几乎接近 0.2, 显著小于其他几个常用模型。

表 3 显示了对应 4 个数据集的 6 种模型点预测和区间预测结果评价指标。其中, 使用 STiCDRS-GPR 模型的点预测评价指标 R、MAE 和 RMSE 和区间预测评价指标 CP、MWP 和 MC 在所有模型中综合结果最优, 相比未采用空时融合的 STiCDRS 结构的 TCN 有明显的提升。比较可得: 采用空间特征提取模块的模型在风速数据集上的预测能力(如 CNN、CLSTM、TCN、STiCDRS)优于仅使用时序

提取模块的 LSTM 模型, 同时融合了空时残差模块的 STiCDRS 模型优于其他基本残差结构模型, 进一步提升了模型的性能, 这说明改进空时融合的残差模块在不明显增加深度学习模型训练时间的条件下, 可以提升模型的精度。所有模型都使用相同的数据集进行试验, 均在相同的随机权重下进行初始化, 因此从整体上看: 1) 模型的耦合提升了对变量之间关系的映射能力; 2) 空时融合的残差结构对模型的预测准确度提升较大; 3) 融合基于注意力机制的软阈值残差进一步提升了模型预测的精度。结果均表现为 STiCDRS-GPR 模型拥有更精准的预测能力。

表 3 对应 4 个数据集的所有模型点预测和区间预测结果评价指标

Table 3 Evaluation metrics for all models' point and interval predictions corresponding to four datasets

数据集	模型指标	CNN	LSTM	CLSTM	TCN	STiCDRS	GPR	数据集	模型指标	CNN	LSTM	CLSTM	TCN	STiCDRS	GPR
数据集 1	R	0.994	0.986	0.995	0.994	0.996	0.994	数据集 3	R	0.996	0.992	0.990	0.993	0.995	0.981
	MAE	0.360	0.536	0.375	0.350	0.316	0.348		MAE	0.258	0.194	0.277	0.246	0.177	0.388
	RMSE	0.470	0.744	0.472	0.500	0.445	0.477		RMSE	0.320	0.250	0.348	0.296	0.221	0.449
	CP	0.895	0.917	0.885	0.958	0.958	0.927		CP	0.919	1.000	0.904	0.985	1.000	1.000
	MWP	0.239	0.350	0.227	0.253	0.231	0.230		MWP	0.246	0.341	0.230	0.232	0.234	0.276
	MC	0.267	0.382	0.256	0.264	0.241	0.248		MC	0.268	0.341	0.255	0.236	0.234	0.276
	ATT	18.3	30.5	20.5	17.1	19.2	1.3		ATT	31.0	41.8	34.3	30.2	33.1	1.7
数据集 2	R	0.997	0.996	0.997	0.997	0.998	0.996	数据集 4	R	0.998	0.997	0.998	0.998	0.999	0.996
	MAE	0.267	0.322	0.248	0.258	0.243	0.267		MAE	0.288	0.337	0.307	0.289	0.275	0.402
	RMSE	0.357	0.432	0.343	0.362	0.329	0.357		RMSE	0.364	0.433	0.388	0.366	0.350	0.505
	CP	0.815	0.919	0.867	0.874	0.933	0.815		CP	0.964	0.984	0.938	0.974	0.979	0.984
	MWP	0.336	0.495	0.333	0.374	0.350	0.336		MWP	0.214	0.269	0.210	0.211	0.209	0.251
	MC	0.412	0.538	0.384	0.428	0.375	0.412		MC	0.222	0.273	0.224	0.217	0.213	0.255
	ATT	31.2	42.1	36.5	31.1	33.8	1.5		ATT	53.1	58.8	53.3	50.2	51.8	1.9

图 11 显示测试集分别对应 CNN、LSTM、CLSTM、TCN、STiCDRS 和 GPR 模型点预测的 PIT 展示结果 P-P 图, 所有模型得到的点预测结果 P-P 图在对角线两旁呈现均匀分布, 并且均匀覆盖在 [0,1] 之间, 说明了点预测 PIT 结果服从均匀分布, 同时表明了预测结果较为可靠。与此同时, 本文中所有模型在 4 个数据集上的 PIT 变换点分布均通过了 Kolmogorov-Smirnov 检验(5%显著性水平), 表明模型所预测的概率密度函数与真实分布较为一致, 具有良好的校准效果。其中, STiCDRS-GPR 模型的概率结果在 4 个数据集中均紧密地分布在对角线两旁, 说明 STiCDRS-GPR 模型能够更有效地捕捉风速的概率特征。

4 结论

本文在时间卷积网络模型的基础上, 通过引进

DRSN进行模型结构的优化, 得到改进的深度学习点预测模型STiCDRS, 同时结合风速数据所具有的空时特征, 提出了一种融合深度学习点预测和GPR区间预测模型的混合区间预测框架STiCDRS-GPR, 通过贝叶斯优化算法自适应地选择超参数组合。并利用内蒙古 4 组实测风速数据进行仿真实验验证。通过对比实验得出以下结论:

1) STiCDRS 获得了最高的点预测精度, 可以挖掘数据间的非线性关系, 为区间预测提供更为精确的点预测结果;

2) STiCDRS-GPR 得到了最为合适的预测置信带, 即它通过最小的宽度覆盖最多的观测点, 对不同数据中的噪声和不确定性具有更强的鲁棒性, 同时 GPR 区间预测模型对于模型的置信带具有较好的解释性;

3) STiCDRS-GPR 在保持较高预测性能的同时,

得到了最为可靠的概率预测结果，其各项指标表现最优且置信带覆盖率高，能够较好地捕捉实际数据的分布情况。

总之，本文提出的 STiCDRS-GPR 模型在短期风速预测的准确性、不确定性和可靠性方面具有优异的预测效果。

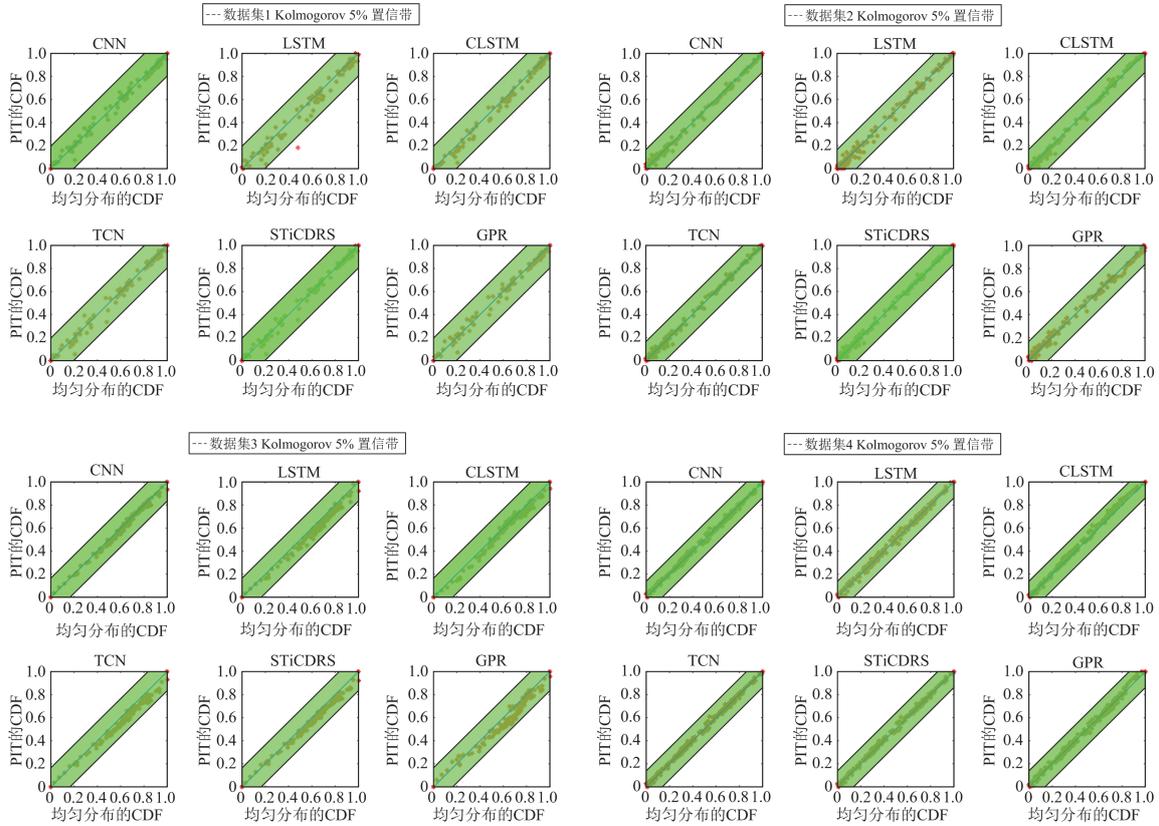


图 11 对应 4 个数据集的所有模型点预测结果的 P-P 图

Fig. 11 P-P plot of all models' point prediction results corresponding to four datasets

参考文献

[1] ZHANG Zhendong, QIN Hui, LIU Yongqi, et al. Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction[J]. Energy Conversion and Management, 2019, 192: 37-51.

[2] 负保记, 赵文, 王建学, 等. 低碳矿区综合能源系统经济运行优化研究[J]. 电力系统保护与控制, 2024, 52(6): 177-187.

YUN Baoqi, ZHAO Wen, WANG Jianxue, et al. Review of research on operation optimization of integrated energy system[J]. Power System Protection and Control, 2024, 52(6): 177-187.

[3] MAO Yang, GUO Yunfeng, HUANG Yutong. Wind power ultra-short-term prediction method based on NWP wind speed correction and double clustering division of transitional weather process[J]. Energy, 2023, 282: 128947.

[4] SAJID S, SALKUTI S R. Short term wind speed forecasting using time series techniques[J]. Energy Sources Part A-Recovery Utilization and Environmental Effects, 2022, 44(4): 9861-9881.

[5] 王延峰, 曹育晗, 孙军伟. 基于多策略改进金豺算法优化 LSTM 的短期电力负荷预测[J]. 电力系统保护与

控制, 2024, 52(14): 95-102.

WANG Yanfeng, CAO Yuhua, SUN Junwei. Short-term power load forecasting based on multi-strategy improved golden jackal algorithm-optimized LSTM[J]. Power System Protection and Control, 2024, 52(14): 95-102.

[6] BABU C N, REDDY B E. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data[J]. Applied Soft Computing, 2014, 23: 27-38.

[7] 刘升伟, 王星华, 鲁迪, 等. 基于改进高斯过程回归的短期负荷概率区间预测方法[J]. 电力系统保护与控制, 2020, 48(1): 18-25.

LIU Shengwei, WANG Xinghua, LU Di, et al. Electric load probabilistic interval prediction method based on improved Gaussian process regression[J]. Power System Protection and Control, 2020, 48(1): 18-25.

[8] DONG Yunxuan, MA Xuejiao, FU Tonglin. Electrical load forecasting: a deep learning approach based on K-nearest neighbors[J]. Applied Soft Computing, 2021, 99: 106900.

[9] FAN Guofeng, ZHANG Liuzhen, YU Meng, et al. Applications of random forest in multivariable response surface for short-term load forecasting[J]. International Journal of Electrical Power & Energy Systems, 2022, 139:

- 108073.
- [10] HONG Weichang. Hybrid evolutionary algorithms in a SVR-based electric load forecasting model[J]. International Journal of Electrical Power & Energy Systems, 2009, 31(7-8): 409-417.
- [11] IBRAHIM Y, OMER F B, DURSUN D. Deep-learning-based short-term electricity load forecasting: a real case application[J]. Engineering Applications of Artificial Intelligence, 2022, 109: 104645.
- [12] BAI Shaojie, KOLTER J Z, VLADLEN K. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv, 2018, 1803: 01271.
- [13] FAN Jin, ZHANG Ke, HUANG Yipan, et al. Parallel spatio-temporal attention-based TCN for multivariate time series prediction[J]. Neural Computing and Applications, 2023, 35(18): 13109-13118.
- [14] GONG Mingju, YAN Changcheng, XU Wei, et al. Short-term wind power forecasting model based on temporal convolutional network and Informer[J]. Energy, 2023, 283: 129171.
- [15] LI Jingrui, WANG Jiyang, LI Zhiwu. A novel combined forecasting system based on advanced optimization algorithm—a study on optimal interval prediction of wind speed[J]. Energy, 2023, 264: 126179.
- [16] BOZORG M, BRACALE A, CARPITA M, et al. Bayesian bootstrapping in real-time probabilistic photovoltaic power forecasting[J]. Solar Energy, 2021, 225: 577-590.
- [17] DA Silva A P A, MOULIN L S. Confidence intervals for neural network based short-term load forecasting[J]. IEEE Transactions on Power Systems, 2000, 15(4): 1191-1196.
- [18] YADAV A, BARETH R, KOCHAR M, et al. Gaussian process regression-based load forecasting model[J]. IET Generation, Transmission & Distribution, 2024, 18(5): 899-910.
- [19] ZAMAN U, TEIMOURZADEH H, SANGANI ELOAS H, et al. Wind speed forecasting using ARMA and neural network models[C] // 2021 IEEE Electrical Power and Energy Conference (EPEC), November 30, Toronto, ON, Canada: 243-248.
- [20] 唐振浩, 赵赓楠, 曹生现, 等. 基于 SWLSTM 算法的超短期风向预测[J]. 中国电机工程学报, 2019, 39(15): 4459-4468.
TANG Zhenhao, ZHAO Gengnan, CAO Shengxian, et al. Very short-term wind direction prediction via self-turning wavelet long-short term memory neural network[J]. Proceedings of the CSEE, 2019, 39(15): 4459-4468.
- [21] ZHAO Minghang, ZHONG Shisheng, FU Xuyun, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681-4690.
- [22] SULTANA N, HOSSAIN S M Z, ALMUHAINI S H, et al. Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand[J]. Energies, 2022, 15(9): 3425.
- [23] YADAV A, BARETH R, KOCHAR M, et al. Gaussian process regression-based load forecasting model[J]. IET Generation, Transmission & Distribution, 2024, 18(5): 899-910.
- [24] 王耀健, 顾洁, 温洪林, 等. 基于在线高斯过程回归的短期风电功率概率预测[J]. 电力系统自动化, 2024, 48(11): 197-207.
WANG Yaojian, GU Jie, WEN Honglin, et al. Short-term wind power probability prediction based on online Gaussian process regression[J]. Automation of Electric Power Systems, 2024, 48(11): 197-207.
- [25] QIU Wenzhi, ZHANG Wenyu, WANG Gang, et al. Combined wind speed forecasting model based on secondary decomposition and quantile regression closed-form continuous-time neural network[J]. International Journal of Green Energy, 2024, 21(8): 1793-1814.
- [26] 杨茂, 张书天, 王勃. 基于因果正则化极限学习机的风电功率短期预测方法[J]. 电力系统保护与控制, 2024, 52(11): 127-136.
YANG Mao, ZHANG Shutian, WANG Bo. Short-term wind power forecasting method based on a causal regularized extreme learning machine[J]. Power System Protection and Control, 2024, 52(11): 127-136.
- [27] WANG J, GAO J, WEI D. Electric load prediction based on a novel combined interval forecasting system[J]. Applied Energy, 2022, 322: 119420.
- [28] 胡文慧, 苏欣, 姜林, 等. 考虑预测误差时空相依性的短期风电功率概率预测[J]. 南方电网技术, 2023, 17(2): 137-144.
HU Wenhui, SU Xin, JIANG Lin, et al. Probabilistic prediction of short-term wind power considering spatio-temporal dependence of prediction error[J]. China Southern Power Grid Technology, 2023, 17(2): 137-144.
- [29] 马德征, 王聪, 王小荣, 等. 基于混合深度学习模型的风速区间预测研究[J]. 太阳能学报, 2023, 44(3): 139-146.
MA Caizheng, WANG Cong, WANG Xiaorong, et al. Research on wind speed interval prediction based on hybrid deep learning model[J]. Acta Energetica Solaris Sinica, 2023, 44(3): 139-146.
- [30] PEI Shaoqian, QIN Hui, ZHANG Zhendong, et al. Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network[J]. Energy Conversion and Management, 2019, 196: 779-792.
- [31] 金阳忻, 徐永进, 黄小琼. 基于卷积神经网络的电能表测得值置信区间研究[J]. 中国电机工程学报, 2021, 41(19): 6630-6641.
JIN Yangxin, XU Yongjin, HUANG Xiaoqiong. Confidence coefficient calculation model of the energy meter measurement based on convolutional neural network[J]. Proceedings of the CSEE, 2021, 41(19): 6630-6641.

收稿日期: 2024-05-14; 修回日期: 2024-10-15

作者简介:

伍耘(2001—), 女, 硕士研究生, 研究方向为风电场风速、配网负荷预测; E-mail: 765570783@qq.com

车亮(1982—), 男, 通信作者, 博士, 教授, 研究方向为决策智能技术在电力系统中的应用。E-mail: cheliang@hnu.edu.cn

(编辑 许威)