

基于 ARIMA-TCN-LSTM-AM-CatBoost 的 WSN 太阳辐照度预测算法

龙晨¹,徐震¹,文士元²

(1.武汉轻工大学 电气与电子工程学院,武汉 430048;2.成都理工大学 计算机科学与网络安全学院,成都 610059)

摘要:在无线传感器网络中,节点的供能问题是制约其实际应用的重要因素,将太阳能转化成电能为节点供电是一种有效的解决方案.对可收集的太阳能进行预测,有助于提高无线传感器网络的使用寿命.针对太阳能的不稳定性和波动性,提出了一种混合模型来预测太阳能.该模型首先利用自回归差分移动平均模型(ARIMA)对序列中线性成分的敏感性来对太阳辐照度中的线性成分进行提取,保留非线性成分于残差序列中;融合时间卷积网络(TCN)、长短期记忆神经网络(LSTM)和注意力机制(AM),提取光伏数据和由 ARIMA 生成的残差数据中更深层次的时间依赖关系和复杂模式;最后,利用 Catboost 决策树算法对预测结果进行集成和综合分析.实验结果表明,在太阳辐照度预测中,所提出的方案在精度、鲁棒性和泛化能力上相较于其他方法具有明显优势.

关键词:无线传感器网络;太阳辐照度;自回归差分移动平均模型;长短期记忆网络;CatBoost

中图分类号:TN919.2

文献标志码:A

文章编号:1000-2367(2026)03-0076-09

近年来,随着无线通信技术、微电子技术的高速发展以及各种低功耗传感器的兴起,将数据采集和信息处理等多种功能集成在微小体积的传感器上成为研究人员热衷的目标,从而推动了无线传感器网络(wireless sensor networks, WSN)在环境监测^[1]、智能家居^[2]、工业监测^[3]和医疗护理^[4]等领域的广泛应用.

WSN 一般部署在无人或不易接近的环境中,因此无线传感器节点的能耗问题和能量持久性问题常常成为它在实践和理论研究中的核心问题.若从根本上解决节点的能量供给问题,需利用节点附近的环境能源,将其转化为电能为节点补充所需电量^[5],这类环境能源包括太阳能、风能、热能、射频能等.其中太阳能因为自身能量密度大、具有周期性且资源丰富等特点而备受关注.但由于太阳辐射强度会随着时间的变化而变化,且节点的位置及其光伏收集板的朝向也会对能量采集产生较大的影响.环境能源的不确定性为研究能量收集型无线传感器网络的能量预测方案提出了新的挑战.

传统的太阳能预测算法基于统计学原理和随机性原则.KANSAL 等^[6]提出了指数加权移动平均算法(exponential weighted moving average, EWMA),假定一天内某特定时段的能量收集曲线与前几天在相同时段的能量收集曲线相似,由此,将前一天同时段所收获的能量与上个时段预测到的能量进行加权平均得到预测结果.NOH 等^[7]提出了精确太阳能分配法(accurate solar energy allocation, ASEA)在 EWMA 的基础上引入一个新的参数,该参数为前一个时隙的实际值与预测值的比值,用于反映当前的太阳条件.由于 ASEA 仅考虑了前一个时隙的气象状况,使得短期内的气象条件的变化引起较大的预测波动.LI 等^[8]在 EWMA 的基础上提出了天气条件移动平均法(weather-conditioned moving average, WCMA),WCMA 引入了一个新的权重 GAP,以根据当前的天气条件(如温度、湿度、风力等)来动态调整移动平均值的权值,使其更好地反映天气变化的影响,

收稿日期:2024-12-18;**修回日期:**2025-01-10.

基金项目:国家自然科学基金(62072319).

作者简介:龙晨(2000—),女,吉林长春人,武汉轻工大学硕士研究生,研究方向为无线传感器网络与机器学习,E-mail: longchen0328@icloud.com.

通信作者:徐震,E-mail: xuzhen2046@qq.com.

引用本文:龙晨,徐震,文士元.基于 ARIMA-TCN-LSTM-AM-CatBoost 的 WSN 太阳辐照度预测算法[J].河南师范大学学报(自然科学版),2026,54(3):76-84.(Long Chen, Xu Zhen, Wen Shiyuan. Prediction algorithm solar irradiance based on ARIMA-TCN-LSTM-AM-CatBoost model in wireless sensor network[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(3): 76-84. DOI: 10.16366/j.cnki.1000-2367.2024.12.18.0003.)

从而提高预测精度,但对于某些极端天气条件或突发事件,WCMA 仍然无法进行准确预测.DEB 等^[9]提出了分布能量预测模型,模型通过分析能量分布的历史数据和趋势,来预测未来的能量分布情况.该模型考虑不同时间段内的能量使用情况,并结合各种气候条件等影响因素来建立能量分布的预测模型.

近几年,随着新一代人工智能技术的快速发展,传统的预测方法逐渐被支持向量机^[10-11]、决策树^[12-13]、神经网络^[14-15]等人工智能预测方案所取代.SAXENA 等^[11]提出 KNN-SVM 方法进行太阳能预测.该方法将 K-最近邻(K-Nearest neighbors, KNN)机器学习技术与支持向量机(support vector machine, SVM)相结合,通过利用 KNN 的结构多样性和 SVM 的数据多样性来提高预测精度.但该方法存在一定的局限性,在 KNN 算法中, K 值的选择依赖于大量实验,导致计算成本较高.RAHUL 等^[13]使用决策树回归模型预测太阳能发电厂的发电量,该方法的优点在于能够通过选择合适的分裂标准(如平均绝对误差 MAE、均方误差 MSE)和控制决策树的深度,有效处理数据集中的异常值,从而能较为准确地预测太阳能发电量.LIM 等^[14]提出了一种将卷积神经网络(convolutional neural network, CNN)和长短期记忆网络(long short-term memory, LSTM)相结合的稳态光伏发电功率预测算法.CNN 用于识别天气类型, LSTM 能够学习并预测光伏发电电量的长期依赖关系.预测模型可以根据不同的天气类型学习发电量的模式,从而提高预测精度和模型对非平稳天气的适应性,但该模型在天气出现大幅波动时会出现较大误差.AWAIS 等^[15]提出了一种基于时空注意力机制的 LSTM 模型用于太阳能预测,特别是针对高效灌溉系统中的太阳能发电厂功率输出预测.该方法的优点在于通过时空注意力机制,能有效筛选数据中的关键特征,使模型聚焦于对功率输出影响较大的因素,如温度、风速、日照时长等气象变量,从而提升预测精度.

针对单一或传统模型中预测精度较低、预测过程不稳定等问题,本文提出一种组合模型来预测太阳能.该模型将多种模型进行融合,第一层为自回归差分移动平均模型(auto-regressive integrated moving average model, ARIMA);第二层将时间卷积网络(TCN)、长短期记忆神经网络(LSTM)和注意力机制(attention mechanism, AM)融合,构建 TCN-LSTM-AM 模型;第三层为 CatBoost 模型.组合模型将太阳直接辐照度数据传入 ARIMA 模型进行预处理,得到其残差序列,再将残差序列作为预测的特征之一,与皮尔逊相关性分析后选取的特征集一同传入 TCN-LSTM-AM 网络中对太阳能直接辐照度进行训练,最后将训练结果传入 CatBoost 网络进行集成.经实验验证,本文方案有助于提高太阳能的预测精度、鲁棒性和泛化能力.

1 太阳能预测模型

1.1 数据处理

1.1.1 数据来源

实验选用澳大利亚沙漠知识太阳能中心(DKASC)的公开数据集^①,该数据集包含了功率数据和气象数据.选用该地区内 Alice Spring 电站 2023 年 9 月 1 日至 2023 年 9 月 30 日期间的数据进行实验,时间窗口为当地时间 06:00—18:00,时间间隔为 5 min.

1.1.2 归一化处理

现有的光伏数据普遍会出现局部缺失、数据波动异常等问题.为此,本文在保证原始序列连续性的同时,对数据中的局部空缺进行去除,以避免其对神经网络模型的训练产生影响.在此基础上,对多个特征序列进行归一化处理,使各特征序列数值均在 0 到 1 范围内,以避免因各特征之间量纲不同引起的影响,从而提高模型的预测精度.公式如下:

$$y_n = \frac{x_n - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

式中, y_n 是归一化后的数值, x_n 是原始数据, x_{\max} 和 x_{\min} 分别是序列中的最大值和最小值.

1.1.3 皮尔逊相关性分析

太阳辐射强度受多种因素的影响,如果只利用单一的太阳直射辐射数据进行预测,将会产生较大的误

① <https://dkasolarcentre.com.au/locations/alice-springs/>.

差.因此,本文利用数据集提供的直接辐照度、水平辐照度、倾斜辐照度、风向、风速、环境温度和相对湿度等 7 个特征序列进行相关性分析,并将与直接辐照度关联度中等及偏上的特征作为研究对象进行多特征预测.本文选用皮尔逊相关性分析法^[16],以度量两个随机变量间的关系.相关系数在 $[-1,1]$ 内,其绝对值决定特性间的相关程度,如果相关系数的绝对值较大,则说明两个变量之间的关联度较高;绝对值等于 0 意味着某一变量的增加和减少不会影响到其他变量的值.相关系数表示如下:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2)$$

式中, r_{xy} 是变量 x 和 y 的皮尔逊相关系数, \bar{x}, \bar{y} 是 x, y 元素的平均值.

相关系数绝对值与相关程度的关系如表 1 所示,太阳直接辐照度与其他特征的相关系数如表 2 所示.

表 1 相关系数绝对值与相关程度关系

Tab. 1 The relationship between the absolute value of the correlation coefficient and the degree of correlation

相关系数绝对值	[0.0, 0.2)	[0.2, 0.4)	[0.4, 0.6)	[0.6, 0.8)	[0.8, 1.0]
相关程度	极弱相关	弱相关	中等相关	强相关	极强相关

表 2 太阳辐照度与其他特征相关系数

Tab. 2 Correlation coefficient between solar irradiance and other characteristics

影响因素	水平辐照度	倾斜辐照度	风向	风速	环境温度	相对湿度
相关系数	0.879 3	0.883 4	-0.231 7	0.442 3	0.303 6	-0.311 9

综上分析,本文选取水平辐照度、倾斜辐照度和风速作为预测太阳能的主要影响因素.

1.2 预测模型构建

1.2.1 ARIMA 模型

ARIMA 的核心思路是通过时间序列进行差分处理,使数据平稳,而后通过结合时间序列的过去值(自回归部分)和过去预测误差(移动平均部分),来捕捉数据的趋势和季节性变化.ARIMA 主要包括 3 个模块:自回归模型(AR)、差分过程(I)、移动平均模型(MA).3 个模块的组合使得 ARIMA 能把握数据的变化趋势,并能处理瞬时的、突发的或含有高噪声的数据.ARIMA 模型可以表示为:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t, \quad (3)$$

式中, X_t 为 t 时观察值, L 为滞后算子, p 为自回归阶数, q 为偏自回归阶数, d 为差分阶数, ϕ 为自相关系数, θ 为移动平均系数, ϵ_t 为 t 时误差项.考虑到太阳直接辐照度具有非线性和非平稳性,而 ARIMA 模型在处理这类时间序列数据时具有显著的优势,且 ARIMA 模型能够有效捕捉太阳辐照度中的线性关系.采用 ARIMA 模型对太阳辐照度进行预处理,将有助于提升预测的准确性,并为进一步分析太阳辐照度的变化趋势奠定基础.

1.2.2 TCN-LSTM-AM 模型

由于太阳辐射受多种因素的影响,在预测模型中,TCN、LSTM、AM 3 种模型被广泛采用.TCN 能够捕捉多尺度的时间依赖关系,并保持因果性;LSTM 能够捕捉时间序列之间长期依赖关系的能力;AM 能够突出长短期记忆网络中关键时间步的优点.基于太阳辐射的影响因素,本文将 3 种模型进行融合,减少模型的预测误差,提高在复杂情况下模型预测的准确性和稳定性,TCN-LSTM-AM 模型图如图 1 所示.

第一层是对输入数据的格式进行规定(批量尺寸、特征维度、时间步长)的输入层.

第二层是 TCN 层.TCN 是一种用于处理序列数据的神经网络架构^[17],该网络结合了因果卷积和空洞卷积,能够有效地处理长序列并捕捉全局特征,其核心思想是通过卷积操作捕捉时间序列数据中的模式和依赖关系.本文利用 TCN 层结构中的膨胀因果卷积,通过增加卷积核在时间维度上的感受野,能够有效地捕捉较长时间序列中的依赖关系,缓解 LSTM 在处理长序列时的困难,并利用指数线性单元算子增强 TCN 对非线性表达能力,提高训练的稳定性,减少梯度消失或爆炸的问题.TCN 的特征提取原理如图 2 所示.

第三层是 LSTM 层.将 TCN 层进行特征提取后得到的数据传入 LSTM 层中进行训练,利用 LSTM 的记忆单元来维护和更新这些特征的历史状态,进行更深层次的建模和预测.LSTM 是一种循环式神经网络,该网络在 RNN 的基础上增加门结构,有效避免了 RNN 中可能出现的梯度消失和梯度爆炸现象,同时,

LSTM 具有记忆功能,能够从太阳辐照数据中提取出时序变化的信息.LSTM 由若干个记忆块组成,每一个记忆块均有 3 个乘法控制单元^[18];输入门 i_t 、遗忘门 f_t 、输出门 o_t ,表达式如下:

$$i_t = \sigma(\omega_i \cdot [h_{t-1}, x_t] + b_i), \tag{4}$$

$$f_t = \sigma(\omega_f \cdot [h_{t-1}, x_t] + b_f), \tag{5}$$

$$o_t = \sigma(\omega_o \cdot [h_{t-1}, x_t] + b_o), \tag{6}$$

式中, σ 为激活函数, ω 为模型权重参数, b 为模型偏置参数, x_t 为 t 时的输入值, h_t 为 t 时的隐藏状态向量.

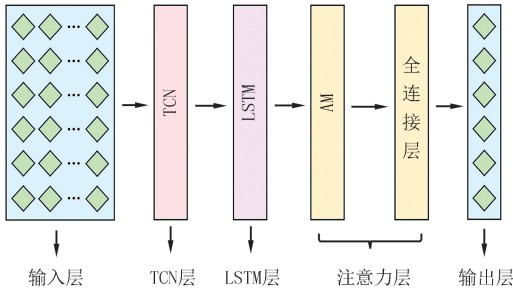


图1 TCN-LSTM-AM模型
Fig.1 TCN-LSTM-AM model

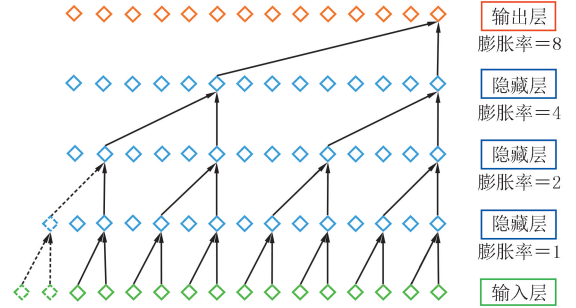


图2 TCN特征提取原理图
Fig.2 Flow chart of TCN feature extraction

第四层是注意力层.注意力机制能够突出 LSTM 中重要的时间步,进而降低模型的预测误差.本文利用时间注意力机制,根据时间步长对输入数据进行操作,将 LSTM 隐藏层的输出向量输入 AM,并利用一个全连接层进行训练,将全连接层的输出结果归一化,输出各隐藏层向量的权值,权值表示各时间步的隐藏状态对于预测结果的重要程度.AM 能够帮助模型更集中地处理重要特征,结合 TCN 和 LSTM 的特征处理能力,可以提高时序预测任务的准确性.

1.2.3 CatBoost 模型

CatBoost 是一种基于决策树的梯度提升算法,它在处理复杂数据特征、类别特征、高效并行计算以及防止过拟合方面具有显著优势.在本文模型中,CatBoost 作为最终的集成器,接收 TCN-LSTM-AM 的输出结果进行再训练,结合各个模型的优势,从而提供更精确的预测.同时,CatBoost 的并行化特性使得它在处理多个复杂子模型的输出时,仍能保证计算的高效性和预测的及时性.相比于其他集成模型,CatBoost 在处理多样化输入特征、抗噪性和稳定性方面表现更优,能够更好地提升整体模型的预测准确性和稳健性,故本文利用 CatBoost 模型进行集成.其基本公式如下:

$$F(x) = F_0(x) + \sum_{m=1}^M \gamma_m T(x; \theta_m), \tag{7}$$

式中, $F(x)$ 为模型的预测输出, M 为树的数量, $T(x; \theta_m)$ 为第 m 棵树的输出,其中 θ_m 代表树的参数, γ_m 为学习率.

在每次迭代中,CatBoost 通过最小化损失函数的负梯度来更新模型,负梯度(n_g)数学公式如下:

$$n_g = - \frac{\partial L(y, F(x))}{\partial F(x)}, \tag{8}$$

式中, $F(x)$ 为模型的预测输出, $L(y, F(x))$ 为损失函数.

1.2.4 ATLAC 模型构建

太阳辐照度的变化是复杂的,存在着线性和非线性变化两种趋势,采用传统的或单一的模型进行预测存在精度不高、拟合程度不良等问题.本文提出 ARIMA-TCN-LSTM-AM-CatBoost(ATLAC)组合模型对太阳辐照度进行预测.首先获取光伏及气象数据,处理数据中的缺失值和不正常的波动值,将其归一化,选取数据的前 85% 作为训练集,后 15% 为测试集.利用 ARIMA 将太阳直接辐照度中的非线性分量以残差的形式分离出来,同时对其他特征进行皮尔逊相关性分析,选取中等相关性及以上的特征,连同残差序列传入 TCN-LSTM-AM 网络进行训练,其中,TCN 通过使用膨胀因果卷积来增加卷积核在时间序列上的感受野,能够有效地捕捉长期的时间依赖关系;LSTM 能够记忆并处理序列中的长期依赖关系,也可以避免传统 RNN 中梯度消失或爆炸的问题;AM 使得模型能够聚焦于输入序列中最重要的部分,提高预测的准确性.TCN-LSTM-

AM 模型能够在 ARIMA 基础预测上,进一步提取深层次的时间依赖关系和复杂模式,生成更加精确的预测结果.最后,利用 CatBoost 模型集成两个模型的输出,对预测结果进行精进.CatBoost 能够优化模型的特征选择和训练,捕捉序列中复杂的非线性关系,显著提升预测的准确性.CatBoost 内置的正则化机制和处理类别特征的方法能够有效减少过拟合的风险,提高模型的泛化能力.模型总体框架图如图 3 所示.

2 实验结果分析

2.1 实验参数

实验环境基于 Windows11 (64 bit),运行内存 32 GB,处理器 13th Gen Intel(R) Core(TM) i9-13905H 2.60 GHz,显卡为 NVIDIA GeForce RTX 4060.实验在 PyCharm Community Edition 2022.2.2 中进行,软件编程环境为 Python 3.7.实验参数设置如附录表 S1 所示.

2.2 模型性能评价指标

本文利用均方根误差(RMSE)、平均绝对百分比误差(MAPE)和决定系数(R^2)评估太阳能预测模型,表达式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}, \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - y'_i}{y_i} \right) \times 100\%, \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y'_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (11)$$

式中, y_i 为实际值, y'_i 为预测值, n 为数据总数, \bar{y} 为平均值. RMSE 和 MAPE 的值越小表明预测的误差越小, R^2 的取值越接近 1 表明预测的结果越好.

2.3 数据说明

图 4 为 2023 年 9 月 1 日至 2023 年 9 月 30 日期间经过数据处理后的 Alice Spring 电站太阳直接辐照度数据,划分数据的前 85% 为训练集,其余 15% 为测试集.从图 4 中可以看到,在连续 360 h 获得的数据中出现了波状的起伏,这表明连续 360 h 所获得的太阳辐射强度在各个时段内有不同的平均值.如图 5 所示,在 2023 年 9 月 10 日 16 时获得的太阳辐射强度是 514 W/m^2 ,与之相比,在 2023 年 9 月 10 日 18 时获得的太阳辐射强度是 0 W/m^2 ,表明该数据是一个周期的、非平稳的随机序列.

2.4 对比分析

2.4.1 不同预训练、集成模型的预测结果

表 3 展现了利用不同的预训练模型和不同的集成模型进行太阳辐照度的预测结果.由表 3 可见,利用 TCN-LSTM-AM 进行预训练并用 CatBoost 进行集成的模型在各项评估指标上都表现出了最优的效果.该组合模型的 RMSE 值最低,仅为 0.063 8,显著优于其他模型;其 MAPE 也最低,为 15.283 1%,显示出在预测准确性上的明显优势.此外, R^2 值达到 0.962 7,这表明 TCN-LSTM-AM 与 CatBoost 结合能够更准确地捕捉数据中的变化趋势.综上,结合 TCN-LSTM-AM 进行预训练并使用 CatBoost 进行集成的模型不仅显著降低了预测误差,提高了预测准确性,还增强了模型的拟合能力.

2.4.2 连续 96 h 的预测结果

本文分别使用 CNN、BiLSTM、KMeans-SVR^[19]、ARIMA-LSTM^[20]、GRU-Xgboost^[21]、GRU-TCN^[22]和 ATLAC 进行太阳辐照度预测,并将 7 个模型的预测结果进行对比分析,以验证 ATLAC 的优越性.表 4

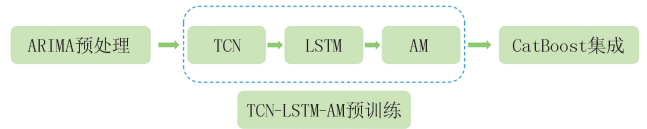


图3 总体框架图

Fig. 3 Overall framework diagram

显示了各模型评价指标的比较结果.从表 4 可以看出,与 CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM、GRU-Xgboost 和 GRU-TCN 模型相比,所提方案的 RMSE 值分别降低了 48.88%、46.83%、42.47%、32.98%、29.74%和 28.07%,说明 ATLAC 对数据的拟合效果更好;MAPE 值分别降低了 55.67%、48.43%、47.91%、52.04%、41.86%和 33.38%,说明该方案的预测结果更准确.此外,ATLAC 的 R^2 值达到 0.962 7,分别比 CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM、GRU-Xgboost 和 GRU-TCN 高 12.31%、10.90%、8.50%、4.98%、4.13%和 3.75%,说明 ATLAC 对目标变量的解释能力更强,能更有效地捕捉数据的变化趋势.

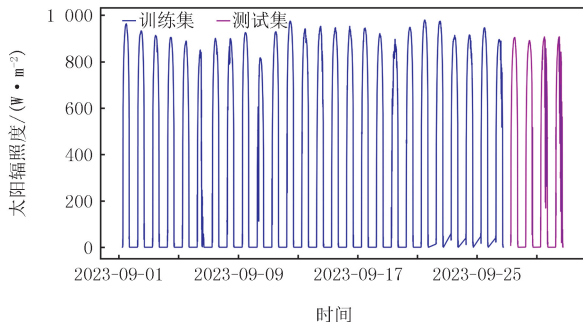


图4 连续360 h的太阳直接辐照度采集数据

Fig. 4 Continuous 360 hours direct solar irradiance acquisition data

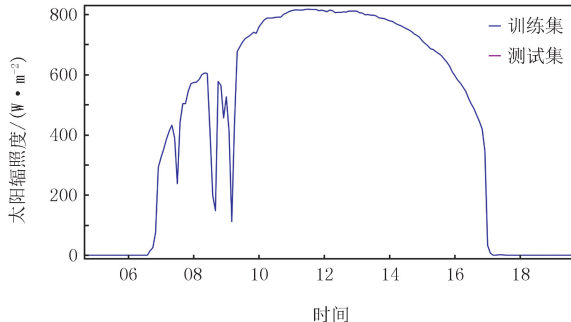


图5 2023年9月10日6时至18时的太阳直接辐照度数据

Fig. 5 Direct solar irradiance data from 6:00 to 18:00 on September 10, 2023

表 3 不同预训练模型和集成模型比较

Tab. 3 Comparison of different pre-training models and integrating models

预处理	集成	RMSE	MAPE/%	R^2	预处理	集成	RMSE	MAPE/%	R^2
None	None	0.111 8	43.015 2	0.885 4	LSTM	CatBoost	0.104 2	25.386 5	0.900 4
None	CatBoost	0.109 4	28.880 8	0.890 3	TCN-LSTM	CatBoost	0.079 2	17.315 4	0.942 6
TCN	TCN	0.112 5	27.182 5	0.884 1	TCN-LSTM-AM	CatBoost	0.063 8	15.283 1	0.962 7
LSTM	LSTM	0.113 3	33.875 1	0.882 4					

表 4 模型预测结果

Tab. 4 Prediction results of the models

模型	RMSE	MAPE/%	R^2	模型	RMSE	MAPE/%	R^2
CNN	0.124 8	34.475 4	0.857 2	GRU-Xgboost	0.090 8	26.287 6	0.924 5
BiLSTM	0.120 0	29.632 9	0.868 1	GRU-TCN	0.088 7	22.940 0	0.927 9
KMeans-SVR	0.110 9	29.338 7	0.887 3	ATLAC	0.063 8	15.283 1	0.962 7
ARIMA-LSTM	0.095 2	31.869 0	0.917 0				

2.4.3 不同天气条件下的预测结果

为了更清晰地展示本文方案的预测效果,从测试数据中选取了 2023 年 9 月 27 日(平稳天气)和 2023 年 9 月 30 日(非平稳天气)的数据,评估本文方案在这两种天气条件下的预测精度.图 6 展示了在平稳天气条件下 7 种模型的预测结果,图 7 展示了在非平稳天气条件下 7 种模型的预测情况.附录表 S2 展示了各模型在平稳天气条件下的误差评估,附录表 S3 展示了各模型在非平稳天气条件下的误差评估.

从图 6 和表 S2 可以看出,所有预测模型在平稳天气下都表现出稳定和准确的性能.与其他模型相比,ATLAC 在 3 个评价指标上具有明显优势.RMSE 分别比 CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM 和 GRU-Xgboost 模型低 64.43%、66.78%、57.28%、28.94%和 13.71%;MAPE 仅为 7.987 3%,分别比 CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM 和 GRU-Xgboost 模型低 53.33%、65.16%、43.19%、60.45%和 5.57%; R^2 分别比 CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM 和 GRU-Xgboost 模型分别高出 7.76%、9.19%、4.91%、1.04%和 0.37%,表明该模型在预测精度方面具有优势.GRU-TCN 模型在平稳天气条件下的预测精度与 ATLAC 模型基本持平.

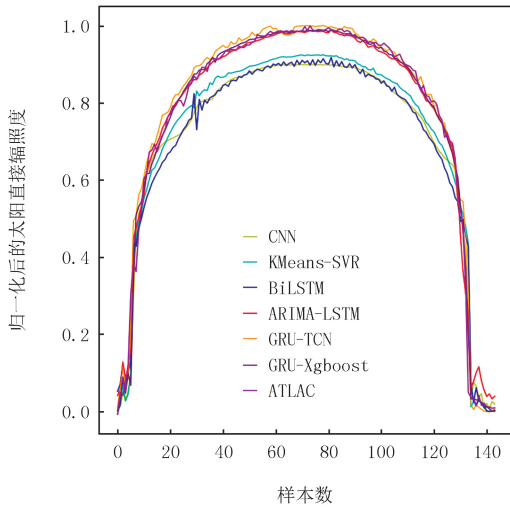


图6 平稳天气下的各模型预测结果

Fig. 6 Prediction results of the models during smooth weather

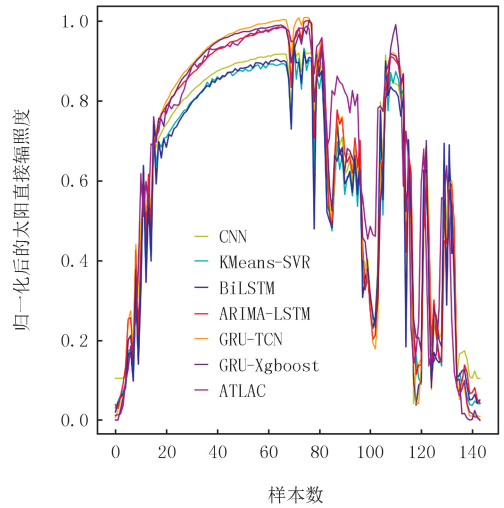


图7 非平稳天气下的各模型预测结果

Fig. 7 Prediction results of the models during non-smooth weather

从图7和表S3中可以看出,在非平稳天气下,模型的预测误差增大,而ATLAC在这种情况下表现突出,与CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM、GRU-Xgboost和GRU-TCN模型相比, $RMSE$ 分别提高了40.96%、36.20%、34.84%、29.73%、25.08%和27.87%, $MAPE$ 分别降低了35.90%、30.67%、37.78%、45.02%、33.71%和28.75%。 R^2 分别比CNN、BiLSTM、KMeans-SVR、ARIMA-LSTM、GRU-Xgboost和GRU-TCN模型高出17.67%、13.27%、12.22%、8.97%、6.71%和8.01%。结果表明,ATLAC对非平稳天气具有更强的预测能力,不仅能准确捕捉数据的趋势和变化,还能在气象条件变化时保持较高的预测精度和稳定性。

2.4.4 时间开销

附录表S4列出了模型的时间开销,预测开销用模型的训练时长和推断时长来表示,推断时长即单次预测时长。由于引入了多个网络,组合模型的训练时长普遍比单一模型长,但在组合模型中,ATLAC的训练时长与其他模型基本持平。表S4显示,ATLAC的训练时长比CNN、BiLSTM和KMeans-SVR分别长26.003 7 s、7.986 5 s和16.462 3 s,而比ARIMA-LSTM、GRU-Xgboost和GRU-TCN的训练时长短3.789 8 s、3.746 6 s和6.553 3 s,推断时长7种算法基本相同。结合表5、表S2、S3、S4分析,本文所提的ATLAC能量预测模型以较小的时间开销为代价,获得了最高的预测精度和鲁棒性。

2.4.5 各模型在其他数据集上的误差评估

为验证ATLAC模型的泛化能力,确保其稳定性与可靠性,选用澳大利亚沙漠知识中心中Katherine电站^[16]的公开数据集进行预测。澳大利亚特殊的地理与气候条件,让该电站的太阳辐照度数据独具特点,与常见实验数据差异显著,能有效丰富模型测试的多样性。选取2023年7月1日至2023年7月31日的数据,此时间段天气状况多样,涵盖晴天、多云及降雨等天气,为模型构建了复杂测试场景,能有效验证模型对复杂天气的适应性。图8为各模型利用该数据进行预测的误差评估图。

从折线图呈现的趋势能够清晰看出,在 $RMSE$ 、 $MAPE$ 和 R^2 这3项关键评估指标的对比中,ATLAC模型脱颖而出。在 $RMSE$ 指标上,ATLAC模型的折线处于最低位置,反映出其预测值与真实值之间的误差波动最小;在 $MAPE$ 指标方面,ATLAC模型同样表现卓越,折线走势显示其预测结果的平均相对误差低于其他模型;在 R^2 指标的折线图上,ATLAC模型的折线位于最高处,表明该模型对数据的拟合优度最佳,整体泛化能力显著优于其他对比模型。

3 结论

针对无线传感器网络中节点供电不足和能源不稳定的问题,本文提出了一种预测太阳辐照度的组合模型,该模型将多种模型进行融合,其中,利用ARIMA对数据进行预处理,ARIMA擅长处理时间序列中的线

性部分和长期趋势,捕捉序列的自相关性和趋势;将 TCN、LSTM、AM 进行融合,对数据进行预训练.TCN 通过卷积运算处理数据,捕捉局部数据的时间依赖性.LSTM 能够有效处理长时间跨度的依赖问题,捕捉数据中的非线性关系,AM 增强了模型对重要时间步的关注,通过分配不同的权重,能够有效捕捉关键时间步的信息,增强模型的解释能力和预测性能;最后利用 CatBoost 决策树算法,集成以上各模型的输出进行最终预测.

这种多模型结合可以捕捉光伏数据中的不同特征,包括短期波动和长期趋势,适应不同时间尺度的变化,从而提供更全面的预测.所采用的模型虽然能有效处理太阳能的不稳定和波动,但在面对极端天气等特殊情况下,仍可能存在一定的局限性.比如,在暴雨、沙尘等恶劣天气下,传感器数据的噪声干扰较大,可能影响模型的预测精度.

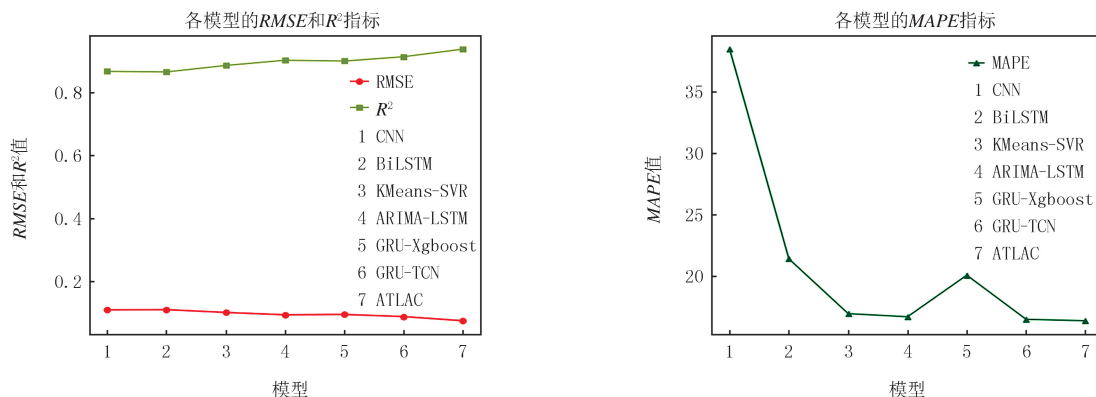


图8 各模型的误差评估

Fig. 8 Error assessment of the models

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.12.18.0003).

参 考 文 献

- [1] ZHAO T C, XU M Y, XIAO X, et al. Recent progress in blue energy harvesting for powering distributed sensors in ocean[J]. *Nano Energy*, 2021, 88: 106199.
- [2] KULURKAR P, DIXIT C K, BHARATHI V C, et al. AI based elderly fall prediction system using wearable sensors: a smart home-care technology with IOT[J]. *Measurement; Sensors*, 2023, 25: 100614.
- [3] BHUSHAN B, SAHOO G. Requirements, protocols, and security challenges in wireless sensor networks: an industrial perspective[M]// *Handbook of Computer Networks and Cyber Security*. Cham: Springer International Publishing, 2020: 683-713.
- [4] GROSSI M. Energy harvesting strategies for wireless sensor networks and mobile devices: a review[J]. *Electronics*, 2021, 10(6): 661.
- [5] BANGOTRA D K, SINGH Y, SELWAL A, et al. An intelligent opportunistic routing algorithm for wireless sensor networks and its application towards e-healthcare[J]. *Sensors*, 2020, 20(14): 3887.
- [6] KANSAL A, HSU J, ZAHEDI S, et al. Power management in energy harvesting sensor networks[J]. *ACM Transactions on Embedded Computing Systems*, 2007, 6(4): 32.
- [7] NOH D K, KANG K. Balanced energy allocation scheme for a solar-powered sensor system and its effects on network-wide performance [J]. *Journal of Computer and System Sciences*, 2011, 77(5): 917-932.
- [8] LI Y B, JIA Z P, XIE S. Energy-prediction scheduler for reconfigurable systems in energy-harvesting environment[J]. *IET Wireless Sensor Systems*, 2014, 4(2): 80-85.
- [9] DEB M, ROY S. Enhanced-pro: a new enhanced solar energy harvested prediction model for wireless sensor networks[J]. *Wireless Personal Communications*, 2021, 117(2): 1103-1121.
- [10] GUO N, GUI W F, CHEN W, et al. Using improved support vector regression to predict the transmitted energy consumption data by distributed wireless sensor network[J]. *EURASIP Journal on Wireless Communications and Networking*, 2020, 2020(1): 120.
- [11] SAXENA N, KUMAR R, RAO Y K S S, et al. Hybrid KNN-SVM machine learning approach for solar power forecasting[J]. *Environmental Challenges*, 2024, 14: 100838.
- [12] 金崇强, 徐震, 王雪山. 基于 Attention 与改进 SCINet 模型的无线传感器网络能量预测与分簇路由算法[J]. *河南师范大学学报(自然科学版)*, 2025, 53(5): 52-60.

- JIN C Q, XU Z, WANG X S. Energy prediction and cluster routing algorithm for wireless sensor networks based on attention and improved SCINet modeling[J]. *Journal of Henan Normal University(Natural Science Edition)*, 2025, 53(5): 52-60.
- [13] RAHUL, GUPTA A, BANSAL A, et al. Solar energy prediction using decision tree regressor[C]//2021 5th International Conference on Intelligent Computing and Control Systems(ICICCS). Madurai: IEEE, 2021: 489-495.
- [14] LIM S C, HUH J H, HONG S H, et al. Solar power forecasting using CNN-LSTM hybrid model[J]. *Energies*, 2022, 15(21): 8233.
- [15] AWAIS M, MAHUM R, ZHANG H, et al. Short-term photovoltaic energy generation for solar powered high efficiency irrigation systems using LSTM with Spatio-temporal attention mechanism[J]. *Scientific Reports*, 2024, 14: 10042.
- [16] MARSHALL G, JONKER L. An introduction to inferential statistics: a review and practical guide[J]. *Radiography*, 2011, 17(1): e1-e6.
- [17] MO J Y, WANG R, CAO M D, et al. A hybrid temporal convolutional network and Prophet model for power load forecasting[J]. *Complex & Intelligent Systems*, 2023, 9(4): 4249-4261.
- [18] FRAIHAT H, ALMBAIDEEN A A, AL-ODIENAT A, et al. Solar radiation forecasting by Pearson correlation using LSTM neural network and ANFIS method: application in the west-central Jordan[J]. *Future Internet*, 2022, 14(3): 79.
- [19] ZHAI W, WU P Y, QI C B, et al. Abnormal data detection method of smart substation based on K-means-SVM[C]//Third International Conference on Electronics, Electrical and Information Engineering(ICEEIE 2023). Xiamen: SPIE, 2023: 45.
- [20] LI C, FANG X Q, YAN Z G, et al. Research on gas concentration prediction based on the ARIMA-LSTM combination model[J]. *Processes*, 2023, 11(1): 174.
- [21] XU Y J, ZHENG S F, ZHU Q L, et al. A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting[J]. *Expert Systems with Applications*, 2024, 254: 124286.
- [22] ELMOUSAID R, DRIQUI N, ELGOURI R, et al. Ultra-short-term global horizontal irradiance forecasting based on a novel and hybrid GRU-TCN model[J]. *Results in Engineering*, 2024, 23: 102817.

Prediction algorithm solar irradiance based on ARIMA-TCN-LSTM-AM-CatBoost model in wireless sensor network

Long Chen¹, Xu Zhen¹, Wen Shiyuan²

(1. Wuhan Polytechnic University, Wuhan 430048, China;
2. Chengdu University of Technology, Chengdu 610059, China)

Abstract: In wireless sensor networks, the energy supply problem of sensor nodes is an important constraint for their practical application, and converting solar energy into electricity to power nodes is an effective solution. The prediction of the harvestable solar energy helps to improve the lifetime of the wireless sensor network. A combination model is proposed to predict solar energy based on its instability and volatility. The model first extracts the linear component of solar irradiance using the auto-regressive differential moving average model(ARIMA), while retaining the nonlinear component in the residual series. Next, temporal convolutional network(TCN), long and short-term memory neural network(LSTM), and attention mechanism(AM) are combined to further extract deeper temporal dependencies and complex patterns in the PV data and the residual data generated by ARIMA. Finally, the CatBoost decision tree algorithm is integrated to synthesize the prediction results. The experimental results show that in solar irradiance prediction, the scheme proposed in this paper has obvious advantages in terms of accuracy, robustness, and generalization ability compared with other methods.

Keywords: wireless sensor network; solar radiation; auto-regressive differential moving average model; long short-term memory network; CatBoost

[责任编辑 陈留院 杨浦]

附录

表 S1 参数设置

Tab. S1 Parameter settings

模型	参数	设置	模型	参数	设置	模型	参数	设置
ARIMA	p, d, q	2, 1, 0	TCN	返回序列	True	CatBoost	目标函数	RMSE
TCN	过滤器	64		丢弃率	30%		迭代次数	10
	卷积核尺寸	2	LSTM	单元数	64		树的深度	3
	膨胀率	[1, 2, 4, 8]		返回序列	True		学习率	0.1
	激活函数	ELU($\alpha=1.0$)		丢弃率	30%			

表 S2 平稳天气条件下 7 个模型的预测误差评估

Tab. S2 Evaluation of prediction errors of five models during smooth weather

天气类型	模型	RMSE	MAPE/%	R^2	天气类型	模型	RMSE	MAPE/%	R^2
平稳天气	CNN	0.084 9	17.113 0	0.918 4	平稳天气	GRU-Xgboost	0.035 0	8.458 0	0.986 1
	BiLSTM	0.090 9	22.922 6	0.906 4		GRU-TCN	0.029 9	9.878 0	0.989 9
	KMeans-SVR	0.070 7	14.060 1	0.943 4		ATLAC	0.030 2	7.987 3	0.989 7
	ARIMA-LSTM	0.042 5	20.196 0	0.979 5					

表 S3 非平稳天气条件下 7 个模型的预测误差评估

Tab. S3 Evaluation of prediction errors of five models during non-smooth weather

天气类型	模型	RMSE	MAPE/%	R^2	天气类型	模型	RMSE	MAPE/%	R^2
非平稳天气	CNN	0.160 9	50.820 3	0.786 5	非平稳天气	GRU-Xgboost	0.126 8	49.137 7	0.867 3
	BiLSTM	0.148 9	46.981 4	0.817 1		GRU-TCN	0.131 7	45.718 7	0.856 9
	KMeans-SVR	0.145 8	52.357 6	0.824 7		ATLAC	0.095 0	32.574 4	0.925 5
	ARIMA-LSTM	0.135 2	59.251 0	0.849 3					

表 S4 模型运行时间的比较

Tab. S4 Comparison of the running time of the models

模型	训练时长/s	推断时长/s	模型	训练时长/s	推断时长/s
CNN	50.174 0	0.205 8	GRU-Xgboost	79.924 3	0.239 6
BiLSTM	68.191 2	0.219 1	GRU-TCN	82.731 0	0.290 0
KMeans-SVR	59.715 4	0.224 3	ATLAC	76.177 7	0.265 6
ARIMA-LSTM	79.967 5	0.291 4			