

基于注意力机制的 CNN-BiGRU 短期 光伏发电功率预测

梁宏涛, 王莹, 刘红菊, 郭超男
(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

摘要: 精确的光伏发电短期预测在微电网智能能源管理系统中起着至关重要的作用; 文章提出一种基于注意力机制的 CNN-BiGRU 短期光伏发电功率预测模型; 其核心思想是通过 CNN 提取光伏数据的空间特征, 把 CNN 提取的这些空间特征送入到 BiGRU 神经网络中, 利用 BiGRU 模型捕捉光伏时序数据集的双向信息流, 学习光伏特征的动态变化规律, 引入 Attention 机制为 CNN-BiGRU 的隐藏层输出赋予权重, 减少因时序过长造成的信息丢失, 并且突出强相关特征的影响, 减少弱相关特征的影响; 在美国俄勒冈州本德市公开数据集上做了验证, 并与 BP 神经网络、GRU、BiGRU、基于 Attention 机制的 BiLSTM 以及基于 Attention 机制的 BiGRU 进行对比, 实验结果表明所提模型在预测精度上更有优越性。

关键词: 光伏出力; Attention 机制; CNN; BiGRU; 短期预测

Short-term PV Output Forecast of BiGRU Based on the Attention Mechanism

LIANG Hongtao, WANG Ying, LIU Hongju, GUO Chaonan
(School of Information Science and Technology, Qingdao University of Science and Technology,
Qingdao 266061, China)

Abstract: The accurate short-term forecast of photovoltaic power generation plays an important role in smart energy management system of micro-grid. A convolutional neural network bidirectional CNN-BiGRU model for short-term photovoltaic (PV) power generation prediction based on the attention mechanism is proposed in this paper. The core idea is to extract the spatial features of photovoltaic data through the CNN and these spatial features extracted by the CNN is sent to the BiGRU neural network. The BiGRU model is used to capture the bidirectional information flow of photovoltaic temporal data set and learn the dynamic change rules of photovoltaic features. The Attention mechanism is introduced to assign the weight to the output of the hidden layer of the CNN-BiGRU, and reduce the information loss caused by the excessively long time sequence, also highlight the influence of strong correlation features, and reduce the influence of weak correlation features, which is verified on the public data set in Bend, Oregon, USA. The experimental results show that the proposed model has better prediction accuracy than BP neural network, GRU, BiGRU, bidirectional long short-term (BiLSTM) based on the attention mechanism and the BiGRU based on the attention mechanism.

Keywords: Photovoltaic output; attention mechanism; CNN; BiGRU; short-term forecast

0 引言

传统的发电系统以化石燃料发电为主。然而, 由于其对环境的负面影响, 电力行业现在聚焦于绿色能源发电系统, 因此, “十四五”规划提出要促进化石能源向可再生能源变革转型^[1]。实现“碳达峰、碳中和”的重要举措是构建以可再生能源为主体的新型电力系统。近几年, 太阳能、风能、潮汐能、生物质能等可再生能源发电被大力部署, 特别是小型和大型光伏电站发展迅速。截止到 2020 年末, 全球光伏累计总装机容量达到 760.4 GW, 其中中国累计装

机容量达到 253.4 GW, 位于世界第一^[2]。由于光伏发电本身具有间歇性、波动性和周期性, 因此, 大量光伏并网会导致发电系统的不稳定性, 可能会导致电力输出突然过剩或下降的情况, 甚至会在电力生产和负荷需求之间的平衡方面产生重大问题。精确的光伏出力预测不仅有助于电网规划、平衡供需、能源管理, 而且有助于调度负荷、电源和电池存储之间的能量^[3]。

目前, 光伏出力预测方法可分为 4 类, 即: 物理方法、统计方法、机器学习 (ML, machine learning)^[4]方法和深

收稿日期: 2022-04-02; 修回日期: 2022-04-07。

基金项目: 国家自然科学基金(61973180); 山东省产教融合研究生联合培养示范基地项目(2020-19)。

作者简介: 梁宏涛(1979-), 男, 山东济宁人, 博士, 副教授, 硕士生导师, 主要从事数据挖掘、智能软件、能源互联网方向的研究。

通讯作者: 王莹(1996-), 女, 硕士研究生, 主要从事智能软件、能源互联网方向的研究。

引用格式: 梁宏涛, 王莹, 刘红菊, 等. 基于注意力机制的 CNN-BiGRU 短期光伏发电功率预测[J]. 计算机测量与控制, 2022, 30(6): 259-265.

度学习 (DL, deep learning) 方法^[5]。物理法不需要训练预测模型, 其中期预测甚至更长的时间尺度会有更好的表现, 但严重依赖光伏的气象数据。与物理方法相比, 统计方法具有相对简单的建模过程; ML 和 DL 更适合短期和中期的预测尺度, 具有较高的预测精度, 但 ML 方法具有扩展性差、复杂度高的缺陷。DL 方法在光伏出力预测中具有灵活性、鲁棒性和较高的准确性。深度神经网络能够自动学习从输入到输出的任意复杂映射, 并支持多个输入和输出。常用 DL 光伏出力预测方法包括卷积神经网络 (CNN, convolutional neural network)^[6-7]、循环神经网络 (RNN, recurrent neural network)^[8-9]、长短期记忆神经网络 (LSTM, long short-term memory)^[10-11]、门控循环单元 (GRU, gated recurrent unit)^[12-13] 网络和深度信念网络 (DBN, deep belief network)^[14-15] 等。

文献 [16] 利用 LSTM 解决“长时间周期依赖”问题, 从而进行光伏出力的中长期预测。文献 [17] 利用相似日匹配算法计算不同时刻气象因子权重, 并对相似样本进行特征提取, 采用广义回归神经网络 (GRNN) 对光伏发电功率进行短期预测, 所提模型与 Bp 神经网络、以及未进行特征提取的 GRNN 相比, 预测精度更高。文献 [18] 对相似日算法进行了改进, 结合 Attention-GRU 机制预测不同天气类型的光伏出力。但以上三篇文献都没有考虑时间序列的双向信息流。文献 [19-20] 都是通过粒子群优化算法 (PSO, particle swarm optimization) 对 DBN 进行参数优化, 只不过前者是通过 Semi Boost 算法对天气进行聚类, 后者是通过选取相似日来预测光伏出力。

本文利用 CNN 提取光伏序列的空间特征, 利用 BiGRU 提取光伏序列的双向时间特征, 由于当光伏序列过长时, 容易造成信息丢失、难以提取高维特征等问题, 引入 Attention 机制, 通过注意力权重合理地改变对信息的外部注意, 放大强相关输入因素、减小弱相关输入因素的占比。以美国俄勒冈州本德市真实光伏出力数据为例, 搭建基于 Attention 机制的 CNN-BiGRU 模型, 将实验结果与 BP、GRU、BiGRU、基于 Attention 机制的 BiLSTM、基于 Attention 机制的 BiGRU 做对比, 选择均方根误差 (RMSE, root mean square error)、决定系数 (R^2 , coefficient of determination) 为评价指标。实验结果表明, 文章所提模型 Attention-CNN-BiGRU 效果更好。

1 基于注意力机制的 CNN-BiGRU 模型及原理

1.1 CNN

1989“卷积”一词被 LeCun 在论述其网络结构时首次提到, 卷积神经网络因此得名^[21]。LeCun 最早将 CNN 用于手写数字识别, 并一直保持其在该问题的霸主地位。CNN 多用于提取图像特征, 其基于卷积核或过滤器的共享权重架构, 这些架构沿着输入特征滑动。经典的 CNN 包含卷积层、激活层、池化层 (汇聚层)、全连接层。具体的对 CNN 两大特点层 (卷积层和池化层) 进行描述。

1.1.1 卷积层

卷积层是 CNN 的核心构建模块, 也是大部分计算发生的地方。卷积层的功能是通过卷积计算对输入层的数据进行特征提取。卷积层对输入应用卷积运算, 将结果传递给下一层, 卷积将其感受野中的所有像素转换为单个值, 卷积层的最终输出是一个向量。卷积层参数包括卷积核大小、填充和步长, 是 CNN 的超参数, 共同决定了卷积层输出特征图的维度。卷积核 (k) 的大小可以一般指定为奇数, 卷积核越大, 可以提取的输入特征越复杂; 填充一般分为有效填充、全填充、相同填充; 步长越大, 相应的输出特征图越小, 在某一程度上完成了降维的操作; 可以通过增加填充和增大步长使得输出的维度达到需求。

1.1.2 池化层

池化层通常在卷积层之后, 用于压缩数据和参数的量, 减少过拟合, 具有特征降维, 保持特征不变性的作用。

卷积计算可用来有效地提取局部区域的特征, CNN 计算的过程是: 在卷积之后会引入非线性激活函数; 池化层的作用是特征选择, 用来减少参数数量, 池化一般分为最大池化和均值池化; 卷积神经网络提取完特征后, 一般会送入全连接层。文章选用 1 维卷积 (模型结构如图 1 所示) 在光伏发电出力预测中提取空间特征。

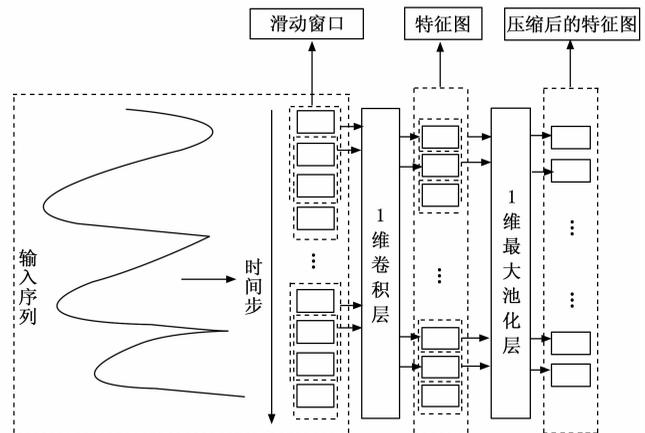


图 1 CNN 模型结构

1.2 GRU

RNN 在有时序信号的领域引起了变革性改变。时序信号的领域指的是语音识别、情感判断、DNA 序列分析、命名实体识别等。RNN 与其他神经网络最大的差别是 RNN 具有记忆功能, 其把前一个时刻网络状态的输出作为当前时刻网络状态的输入。RNN 中的循环核参数时间共享, 循环层提取时间信息。但是 RNN 容易产生梯度消失和梯度爆炸, 因此 LSTM 和 GRU 神经网络作为 RNN 的变体被提出的。

GRU 是在 2014 年由 C. Kyunghyun^[22] 等人提出的。GRU 是 LSTM 进一步简化的一种变体。GRU 在和弦音乐建模、语音信号建模和自然语言处理的某些任务上的表现与 LSTM 相似。LSTM 有 3 个门, 而 GRU 只有两个门。在 LSTM 中, 其是输入门、遗忘门和输出门, 而在 GRU

中, 其是重置门和更新门; 在 LSTM 中, 有两种状态分别是细胞状态和隐藏状态, 而在 GRU 的情况下, 只有一种状态, 即隐藏状态。GRU 由于只有重置门和更新门两个门, 控制要保留多少信息和哪些信息, 所以其参数比 LSTM 少, 更容易实现和计算, 缓解了 LSTM 过拟合的问题。GRU 已被证明在某些更小的数据集上表现出更好的性能。

GRU 的模型结构如图 2 所示。其中, X_t 表示 t 时刻的输入; R_t 表示重置门; Z_t 表示更新门; H_t 表示隐藏状态; \tilde{H}_t 表示候选隐藏状态。重置门和更新门是当前时刻的输入特征 X_t 和上个时刻的 H_{t-1} 的函数, 通过 sigmoid 激活函数, 使门限的范围在 0 到 1 之间。表示候选隐藏状态, 也可表示为现在信息, 其是过去信息 H_{t-1} 过重置门与当前信息共同决定的; H_t 融合了长期记忆和短期记忆。

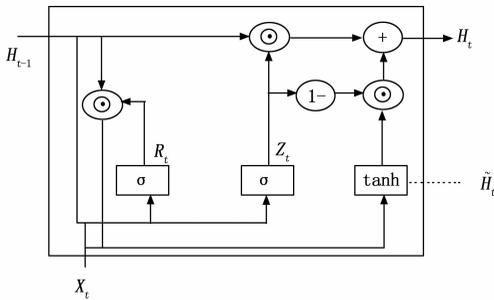


图 2 GRU 的模型结构

根据 GRU 的模型结构, 其计算公式如下所示:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad (1)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad (2)$$

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \quad (3)$$

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t \quad (4)$$

式中, X_t 表示 t 时刻的输入; W_{xr} 、 W_{hr} 、 W_{xz} 、 W_{hz} 、 W_{xh} 、 W_{hh} 为可训练的参数矩阵; R_t 表示重置门; b_r 、 b_z 、 b_h 偏置; Z_t 表示更新门。

1.3 BiGRU

1997 年, M. Schuster 和 K. Paliwal 提出了双向循环神经网络 (BiRNN, bidirectional RNN), 并利用 BiRNN 进行了语音识别实验^[23]。单向 RNN 只能从先前的输入中提取以预测当前状态, 但双向 RNN 会提取未来数据以提高其准确性。BiRNN 在预测的时候, 将前向和后向的值均作为输入。BiRNN 对 BiLSTM 和 BiGRU 同样适用。在 PV 出力预测中, 此时的输出和前一时刻的信息以及后一时刻的信息都密切相关, 所以本文选取 BiGRU 模型, 学习光伏特征的动态变化规律。BiGRU 模型结构如图 3 所示 (实线为前向传播, 虚线为反向传播)。图中 X_i 代表 i 时刻的输入值, h_i 代表 i 时刻隐藏层的值, 而 y_i 代表 i 时刻输出层的值。BiGRU 是 GRU 的改进, 它是由前向传播的 GRU 和反向传播的 GRU 组成。

式中, $\{\vec{h}_1 \rightarrow \vec{h}_2 \rightarrow \vec{h}_3\}$ 表示 BiGRU 的前向传播路径; $\{\leftarrow h_2 \leftarrow h_2 \leftarrow h_3\}$ 表示 BiGRU 的反向传播路径。

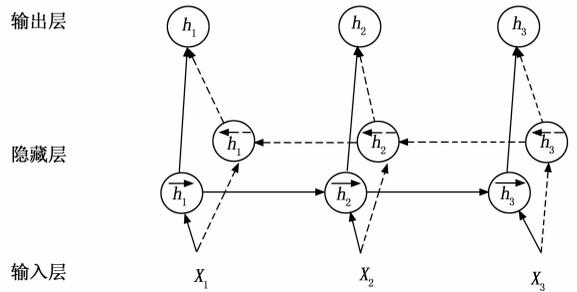


图 3 BiGRU 网络结构

1.4 Attention 机制

Attention 模仿生物观察行为的内部过程, 是一种模仿认知注意力的机制。例如, 人类的视觉在处理一张图片时, 会通过迅速扫描全局图像, 获得注意力焦点, 然后对这一区域投入更多的注意力资源, 并抑制其他无用信息。简单的说 Attention 机制就是从大量信息中快速筛选出高价值信息。2017 年 A. Vaswani 等人^[24]在自然语言领域发表的《Attention is all you need》引起了轰动。注意力的用途包括可微神经计算机中的推理任务、感知器中的多感官数据处理 (声音、图像、视频和文本) 以及转换器中的语言处理。Attention 机制 (模型结构如图 4 所示) 强化输入数据的重要部分, 弱化输入数据的无关部分; 其思想是, 网络应该将更多的注意力放在数据中较小但重要的部分。学习哪一部分数据比其他部分更重要取决于上下文, 并通过梯度下降法进行训练。

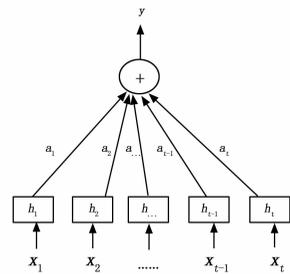


图 4 Attention 机制结构

其中, X_1 、 X_2 、 \dots 、 X_{t-1} 、 X_t 表示光伏数据集的输入特征; h_1 、 h_2 、 \dots 、 h_{t-1} 、 h_t 表示 BiGRU 的隐层状态的输出值, a_1 、 a_2 、 \dots 、 a_{t-1} 、 a_t 表示 Attention 机制对 BiGRU 隐含层输出的注意力权重值。

2 基于 Attention 机制的 CNN-BiGRU 算法

文章提出一种基于 Attention 的 CNN-BiGRU 短期光伏出力预测模型。单一的 BiGRU 算法预测精度不高, 且时序过长时容易造成信息的丢失, 引入 Attention 机制为 BiGRU 的隐藏层输出赋予权重, 来强调关键特征对模型预测的重要性。在此基础上, 引入 CNN 来提取光伏出力数据集的空间特征, 捕捉数据集动态变化规律, 进一步提高光伏发电功率的准确性。Attention 机制的 CNN-BiGRU 的网络模型

结构如图 5 所示。

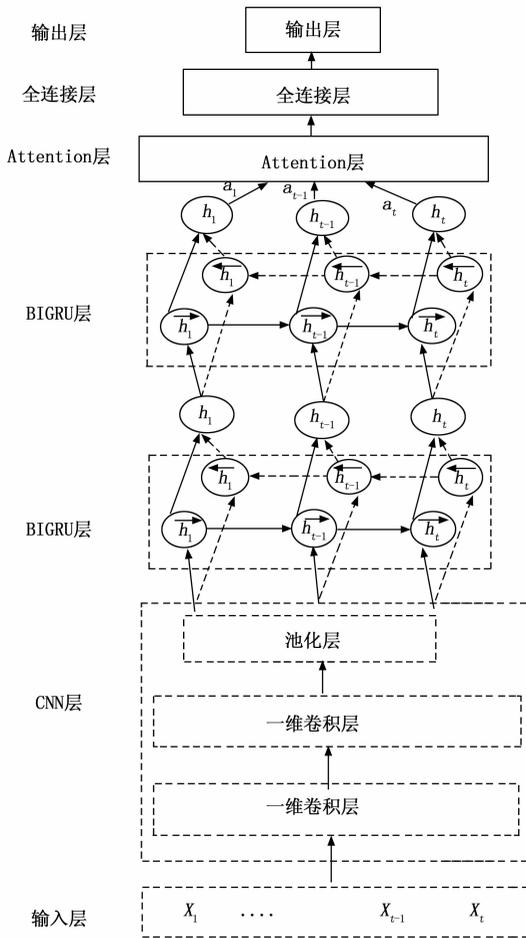


图 5 基于 Attention 的 CNN-BiGRU 机制结构

模型结构具体描述如下：

1) 输入层：把经过相关性分析筛选后得到的光伏发电出力数据集特征作为预测模型的输入，设时间步滑动窗口大小为 T ，经过预处理后在 t 时刻输入到预测模型中的输入序列如公式以下所示：

$$\mathbf{X} = [x_1 \cdots x_{t-1}, x_t \cdots x_T]^T \quad (5)$$

2) CNN 层：CNN 主要用于对光伏序列进行空间特征的提取。由 2 个卷积层、1 个池化层构成。由于光伏序列的特点，2 个卷积层都设计为 1 维卷积。卷积层选取的激活函数为 sigmoid，填充为 same。为保留更多波动信息，池化层选取的最大池化。经过卷积层和池化层进行特征提取，然后送入 BiGRU 神经网络。

3) BiGRU 层：利用前向和反向的 BiGRU 层对 CNN 层提取的特征进行学习。通过搭建两层的 BiGRU，对提取特征充分学习。

4) Attention 层：Attention 层的输入为经过两层 BiGRU 的隐层状态 h_t ，其注意力权重计算公式如下：

$$e_i = \text{vtanh}(W h_i + b) \quad (6)$$

$$a_i = \text{softmax}(e_i) \quad (7)$$

$$C_t = \sum_{i=1}^i \alpha_i h_i \quad (8)$$

式中， e_i 、 a_i 分别表示第 t 时刻光伏数据不同特征所对应的注意力分数和权重， C_t 表示 Attention 层在 t 时刻的输出。

5) 全连接层：全连接层的输入为 Attention 机制的输出。选用 Sigmoid 为激活函数，预测出第 t 时刻的光伏发电出力值，预测公式为：

$$y_t = \text{Sigmoid}(W_o C_t + b_o) \quad (9)$$

式中， y_t 表示为第 t 时刻的光伏发电出力预测值， W_o 为权重矩阵， b_o 为偏置项。

6) 输出层：选用没有激活函数的全连接层作为输出层。

3 数据预处理及相关性分析

为了验证本文提出的基于 Attention 机制的 CNN-BiGRU 短期光伏出力预测模型的性能，采用美国俄勒冈州本德市 (Bend)^[25] 为光伏数据集。光伏组件在早上和晚上的输出功率明显较低，大部分时间接近于零或为零^[26]，所以 Bend 数据集截取 5~19 点的数据，数据集的采样间隔为 5 min，一天共选取 168 个点；Bend 数据集选取的训练集是 2018 年 8~10 月、2019 年 8~10 月，测试集选取的是 2020 年 8 月~10 月。数据集的信息如表 1 所示。

表 1 实验数据集的详细信息

光伏电站名字	BDO(Bend, Oregon)
地址	美国俄勒冈州本德
纬度	44.056 14°N
经度	-121.307 44°W
数据采样间隔	5 min
白天时段	5:00~19:00
输出变量	未来一天的光伏出力
训练集	2018 年 8~10 月、2019 年 8~10 月
测试集	2020 年 8~10 月
评价指标	RMSE、 R^2

文章采用 Python3.8，tensorflow2.7.0 编程实现基于注意力机制的 CNN-BiGRU 短期光伏出力预测。计算环境为 2 GHz 四核 Intel Core i5 处理器、16 GB 内存的计算机。

3.1 数据预处理

数据的准确性是保证预测精度的前提。由于设备检修或网络问题，容易造成数据在采集阶段和传输过程中丢失或异常。且不同数据的量纲不同，范围不同，为进行精确的光伏发电预测，统一量纲和范围是非常有必要的。

3.1.1 缺失值异常值处理

缺失数据一般用 NaN 表示，代表空值，其填充的方法有很多，常用的数据插补方法有均值填充、众位数填充、中位数填充、固定值填充和拉格朗日插值法。文章利用 py-

thon 中 df.isnull() .sum() 函数检测光伏发电数据集的空缺值。如表 2 所示。

表 2 每个特征的空缺值

特征	空缺值
太阳天顶角(SZA)	0
太阳方位角(AZM)	0
地外辐射度(ETR)	0
地外法向辐射度(ETRn)	0
水平面总辐射度(GHI)	336
倾斜面总辐射度(GTI)	339
电流(PV_I)	336
光伏发电功率(PV_Power)	336
电压(PV_DC)	336
风速(Wind Speed)	336
温度(Temperature)	336
组件温度(PanelTemperature)	336

为防止破坏数据的时间特性, 文章选择利用 date-time.time_delta() 函数获取前一天同一时刻的特征值来填充。比如缺失的是 2019-10-26 05:00:00 的 GHI 的值, 用 2019-10-25 05:00:00 的 GHI 值填充; 缺失的是 2019-10-27 08:00:00 的 Wind_Speed 的值, 用 2019-10-26 08:00:00 的 Wind_Speed 值填充。由于光伏发电的间歇性, 其发电功率不可能为负值, 所以当 PV_Power 属性小于 0 时视为异常值, 此时用 0 值替换。

3.1.2 数据归一化与反归一化

在本数据集中, 水平面总辐射度、电流、电压、风速、温度的单位总是不一致的。数据标准化就是消除不同单位特征之间的量纲, 将数据缩放为通用格式, 以分别在 0 和 1 处建立平均值和标准差。常用的归一化的方法有最大-最小值 (Max-Min) 规范化和 0 均值 (z-score) 标准化两种。本文选用 0 均值标准化。

其公式为:

$$x^* = \frac{x - \mu}{\sigma} \quad (10)$$

其中: μ 表示样本的均值, σ 表示为样本的标准差。

由于最终需要的是光伏发电功率的预测值, 并且需要把真实值和预测值做误差分析, 所以需要进行反归一化。其公式如下所示:

$$x = x' * \sigma + \mu \quad (11)$$

3.2 相关性分析

利用皮尔逊 (pearson) 系数对数据集进行输入特征的选择。pearson 系数衡量两组数据之间的相关性, 其本质上是标准化后的协方差, 因此, pearson 系数介于 -1~1 之间。利用 Pearson 系数选取后的输入变量如表 1 所示。其计算公式为:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\left(\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\right) \left(\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}\right)} \quad (12)$$

式中, \bar{X} 、 \bar{Y} 为样本均值; r 为 pearson 系数。

Bend 数据集的相关性分析如表 3 所示。其中电流、水平面总辐射、倾斜总辐射度、地外辐射度的 pearson 系数大于 0.7, 属于强相关; 太阳方位角的 pearson 系数小于 0.1, 属于弱相关。文章在光伏发电功率预测时舍去弱相关, 保留强相关和其他中等相关特征, 所以文章选取的输入变量是太阳天顶角、地外辐照度、地外法向辐照度、水平面总辐射度、倾斜总辐射度、电流、光伏出力、电压、风速、天气温度、组件温度。

表 3 数据集特征的相关系数

特征	pearson 系数
太阳天顶角(SZA)	-0.577
太阳方位角(AZM)	-0.022
地外辐射度(ETR)	0.762
地外法向辐射度(ETRn)	0.595
水平面总辐射度(GHI)	0.995
倾斜面总辐射度(GTI)	0.988
电流(PV_I)	0.999
电压(PV_DC)	0.607
风速(wind_speed)	0.256
温度(Temperature)	0.287
组件温度(PanelTemperature)	0.225

3.3 评价指标

选取的模型评价指标是 RMSE 和 R^2 。RMSE 越小说明模型精确度越高; R^2 越大说明模型拟合度越高。其中 RMSE、 R^2 的公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{\text{predict}} - P_{\text{true}})^2} \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_{\text{true}} - P_{\text{predict}})^2}{(P_{\text{true}} - \bar{P})^2} \quad (14)$$

式中, P_{predict} 、 P_{true} 分别表示为光伏发电出力的预测值、真实值; N 为数据样本数; \bar{P} 为整体样本的平均值。

3.4 优化器及损失函数的选择

Adamax 是 Adam 的变体, 其对学习率的上限提供了一个更简单的范围, 故本文选取 Adamax 作为模型优化器。采用平均绝对误差 (MAE, mean absolute error) 作为模型 loss 函数, 其计算公式为:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{\text{predict}} - P_{\text{true}}| \quad (15)$$

式中, P_{predict} 、 P_{true} 分别为光伏发电出力的预测值和真实值, N 为数据样本数。

4 实验结果分析

文章所提 Attention-CNN-BiGRU 模型与 BP、GRU、BiGRU、Attention-BiLSTM、Attention-BiGRU 做对比 (如表 4、表 5 所示, 其都是各个模型最佳表现时的值)。

表 4 各模型超参数的值

模型	num_epochs	batch_size	lr	h ₁	h ₂	fc
GRU	60	16	0.005	10	10	10
BiGRU	40	16	0.01	10	10	10
Attention-BiLSTM	50	8	0.001	10	10	10
Attention-BiGRU	45	16	0.001	10	10	10
Attention-CNN- BiGRU	120	16	0.002	10	10	20

表 5 不同预测模型的评价指标

模型	RMSE	R ²
BP	0.741 3	-0.439 8
GRU	0.303 6	0.758 4
BiGRU	0.283 4	0.789 6
Attention-BiLSTM	0.274 9	0.801 9
Attention-BiGRU	0.273 3	0.804 3
Attention-CNN- BiGRU	0.272 3	0.805 7

各模型超参数的值如表 4 所示，其中 num_epochs 代表迭代次数、batch_size 代表批数据、lr 代表学习率、fc 代表全连接层节点个数。BiGRU、Attention-BiGRU、Attention-CNN- BiGRU 预测模型中 h₁、h₂ 分别表示第一层、第二层 BiGRU 的隐藏节点个数，GRU 中 h₁、h₂ 分别表示第一层、第二层 GRU 的隐藏节点个数，Attention-BiLSTM 中 hidden1、hidden2 分别表示第一层、第二层 BiLSTM 的隐藏节点个数。Attention-CNN- BiGRU 预测模型中还有两个超参数，分别是 CNN 卷积核数量 filters 取值为 32，CNN 卷积核大小 filter_size 取值为 3。

通过表格数据可看出 BP 神经网络预测效果最差，其 RMSE 高达 0.7 以上，R² 是负值。Attention-CNN-BiGRU 预测模型表现最佳，其 RMSE 为 0.272 3，R² 为 0.805 7。Attention-CNN-BiGRU 与 GRU、BiGRU、Attention-BiLSTM、Attention-BiGRU 相比，RMSE 分别提高了 3.13%，1.11%，0.26%，0.1%；R² 分别提高了 4.73%，1.61%，0.38%，0.14%。

各神经网络模型在 2020 年 8 月 11 日预测值和真实值的拟合情况由图 6 所示。

图 6 (a) 为真实值和 BP 模型预测值的拟合图，可看出 BP 神经网络模型不适合此光伏出力数据集，误差较大；图 (b) 为真实值和 GRU 模型预测值的拟合图，由其可知，GRU 模型相对与 BP 模型极大程度的提高了预测精度；图 (c) 为真实值和 BiGRU 模型预测值的拟合图，相比较于 GRU 模型的预测值在真实值的下方分布，BiGRU 模型的预测值分布在上方；图 (d) 为 Attention-BiLSTM 模型的预测值；图

(e) 为 Attention-BiGRU 模型的预测值，相比较 Attention-BiLSTM 模型，其预测值分布在真实值的两侧；图 (f) 为 Attention-CNN-BiGRU 的模型预测值与真实值的拟合图，相较于其他 5 种预测模型，Attention-CNN-BiGRU 模型的预测值曲线拟合程度是最贴切的。

图 7 展示了利用 Attention-CNN-BiGRU 模型对 2020 年 8 月 11 日到 2020 年 8 月 15 日的光伏发电功率的预测曲线。从图中可看出总体的预测效果，第一天的拟合效果最好。但是当第二天天气状态不好的情况下，Attention-CNN-BiGRU 预测效果不理想。

5 结束语

大规模光伏发电并网给我国电力系统运行的稳定性带来了巨大挑战，因此，光伏发电出力的精确预测至关重要。有效的光伏出力预测降低了储备需求、系统运行成本、市场失衡的风险，提高了系统的稳定性、增加了市场收入；同时也帮助电力运营商做出更有利的决策。

针对短期光伏出力预测研究现状，文章提出了一种基于 Attention 机制的 CNN-BiGRU 光伏出力预测模型。通过 CNN 提取光伏数据集的空间特征，然后送入到与 Attention 机制相结合的 BiGRU 网络，完成对未来一天（5：00~19：00）光伏发电功率预测。这里引入 Attention 机制是为了减少因时序过长造成的信息丢失，对重要信息赋予较高的权重，忽略无关信息，放大所需信息。在美国俄勒冈州本德

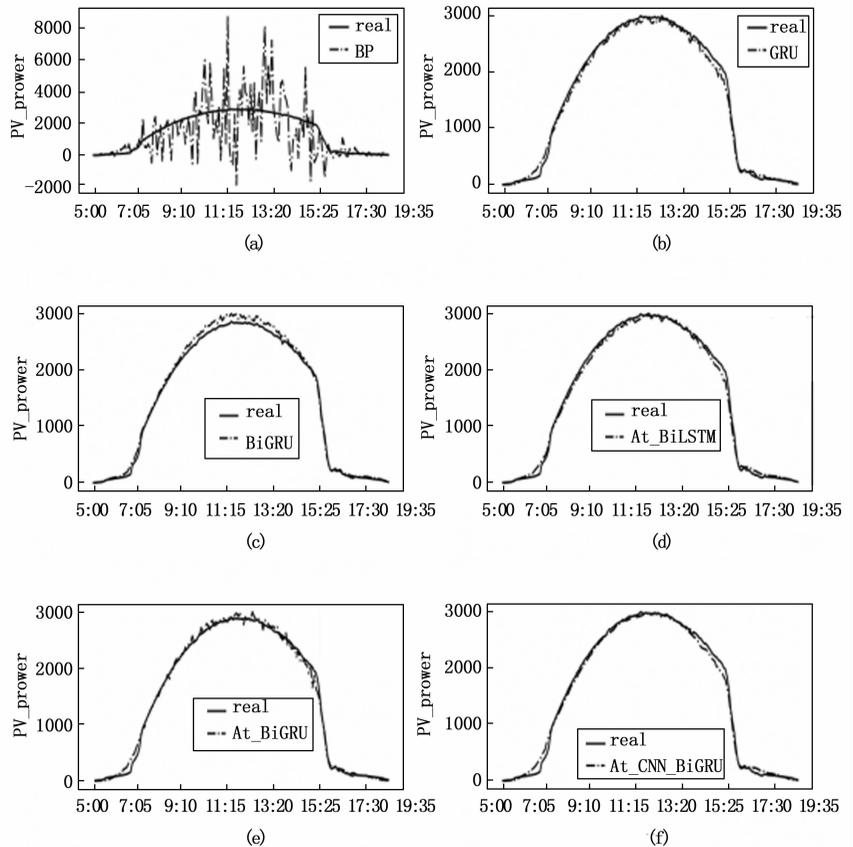


图 6 不同模型的拟合程度

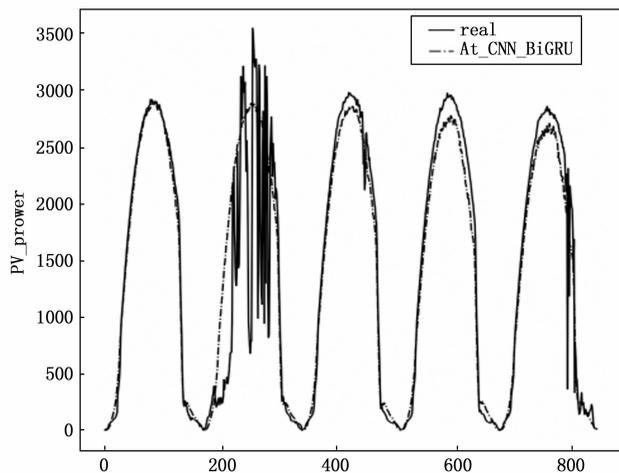


图7 8月11日~8月15日光伏出力预测值

市数据集上做了验证,实验结果表明,BP神经网络预测效果最差,不适合本数据集,Attention-CNN-BiGRU预测模型与GRU、BiGRU、Attention-BiLSTM、Attention-BiGRU预测模型相比RMSE分别提高了3.13%,1.11%,0.26%,0.1%; R^2 分别提高了4.73%,1.61%,0.38%,0.14%。

通过对未来5天的预测,可知在天气状态不好的情况下,所提模型预测精度较低。下一步工作将提高模型在坏天气的预测精度,增加模型的普适性。

参考文献:

[1] 全球能源互联网发展合作组. 中国能源转型“十四五”电力规划研究报告 [EB/OL]. [2020-07-14]. <https://www.geidco.org.cn/html/qqnyhlw/zl20200731/index.html>.

[2] International Energy Agency. Snapshot of Global PV Markets 2021 [EB/OL]. [2021-04-03]. <https://ieapvps.org/snapshot-reports>.

[3] MELLIT A, PAVAN A M, LUGHIV. Deep learning neural networks for short-term photovoltaic power forecasting [J]. *Renewable Energy*, 2021, 172: 276-288

[4] AKHTER M N, MEKHILEF S, MOKHLIS H, et al. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques [J]. *IET Renewable Power Generation*, 2019, 13 (7): 1009-1023.

[5] WANG H, LIU Y, ZHOU B, et al. Taxonomy research of artificial intelligence for deterministic solar power forecasting [J]. *Energy Conversion and Management*, 2020, 214: 112909.

[6] 吕伟杰,方一帆,程泽. 基于FCM-WS-CNN的日前光伏出力预测研究 [J/OL]. *电网技术*: 1-9 [2022-03-07]. <http://DOI:10.13335/j.1000-3673.pst.2020.2246>.

[7] 陈禹帆,温蜜,张凯,等. 基于相似日匹配及TCN-Attention的短期光伏出力预测 [J/OL]. *电测与仪表*: 1-9 [2022-04-21]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200727.1612.024.html>

[8] 刘世岩. 基于深度学习SRNN网络的光伏功率预测 [D]. 西安: 西安理工大学, 2020.

[9] 王鼎. 基于PSO-RNN的短期光伏发电功率预测研究 [D].

北京: 华北电力大学, 2020.

[10] 李丰君,王磊,赵健,等. 基于天气融合和LSTM网络的分布式光伏短期功率预测方法 [J/OL]. *中国电力*: 1-9 [2022-03-07]. <http://kns.cnki.net/kcms/detail/11.3265.TM.20220217.0933.006.html>.

[11] CHAI M, XIA F, HAO S, et al. PV power prediction based on LSTM with adaptive hyperparameter adjustment [J]. *IEEE Access*, 2019, 7: 115473-115486.

[12] JIA P, ZHANG H, LIU X, et al. Short-term photovoltaic power forecasting based on VMD and ISSA-GRU [J]. *IEEE Access*, 2021, 9: 105939-105950.

[13] JAIHUNI M, BASAK J K, KHAN F, et al. A partially amended hybrid Bi-GRU-ARIMA model (PAHM) for predicting solar irradiance in short and very-short terms [J]. *Emerging*, 2020, 13 (2): 435.

[14] HU W, ZHANG X, ZHU L, et al. Short-term photovoltaic power prediction based on similar days and improved SOA-DBN model [J]. *IEEE Access*, 2020, 9: 1958-1971.

[15] 杨茂,王凯旋. 基于CEEMD-DBN模型的光伏出力日前区间预测 [J]. *高电压技术*, 2021, 47 (4): 1156-1164.

[16] 方鹏,高亚栋,潘国兵,等. 基于LSTM神经网络的中长期光伏电站发电量预测方法研究 [J]. *可再生能源*, 2022, 40 (1): 48-54.

[17] 陈世游,左为恒. 基于改进相似样本选取与特征提取的光伏发电功率预测方法 [J]. *计算机测量与控制*, 2019, 27 (2): 165-169, 174.

[18] 刘国海,孙文卿,吴振飞,等. 基Attention-GRU的短期光伏发电功率预测 [J]. *太阳能学报*, 2022, 43 (2): 226-232.

[19] 孙辉,冷建伟. 基于改进的Semi Boost天气聚类的CC-PSO-DBN短期光伏发电预测 [J]. *计算机应用与软件*, 2020, 37 (8): 103-109.

[20] 李正明,梁彩霞,王满商. 基于PSO-DBN神经网络的光伏短期发电出力预测 [J]. *电力系统保护与控制*, 2020, 48 (8): 149-154.

[21] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. *Neural Computation*, 1989, 1 (4): 541-551.

[22] KYUNGHYUN C, BART M V, DZMITRY B, et al. On the properties of neural machine translation: encoder-decoder approaches [J]. *arXiv*: 1409.1259.2014.

[23] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [J]. *IEEE Transactions on Signal Processing*, 1997, 45 (11): 2673-2681.

[24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. *arXiv*, 2017.

[25] Solar Radiation Monitoring Laboratory, University of Oregon. Download solar data [EB/OL]. <http://solardata.uoregon.edu/SolarData.html>.

[26] 周楠,徐潇源,严正,等. 基于宽度学习系统的光伏发电功率超短期预测 [J]. *电力系统自动化*, 2021, 45 (1): 55-64.