基于聚类和 LSTM 的光伏功率日前逐时鲁棒预测

刘兴霖[1,2],黄 超[1,2],王 龙[1,2],罗 熊[1,2]
(1. 北京科技大学 顺德研究生院,广东 佛山 528399;
2. 北京科技大学 计算机与通信工程学院,北京 100083)

摘 要:太阳能作为具有高可用性且用之不竭的清洁能源,被认为是最有前途的能源替代品之一。光伏是最广泛使用的太阳能技术。然而,由于太阳能的间歇性,光伏发电具有不确定性。随着全球光伏装机容量的不断提升,光伏功率预测的准确性对于电网管理和电力调度至关重要。该文提出一种基于 K-means 聚类分析和长短期记忆神经网络(long-short-term memory,LSTM)的光伏发电功率日前逐时鲁棒预测方法。首先采用 K-means 算法以日前天气预报数据为特征将光伏数据分为晴空天气类型和阴雨天气类型,再针对相应类型数据建立基于长短期记忆神经网络算法的预测模型。同时,为增强预测模型的鲁棒性,选择具有强鲁棒性的 Huber 损失函数用于模型训练,并选择计算简单且收敛速度快的鲸鱼优化算法对 Huber 损失函数中的超参数进行优化。将所提出的预测方法与其他方法进行预测性能的比较,结果表明,提出的方法获得了较高的预测精度。

关键词:光伏发电预测;长短期记忆神经网络;K-means 聚类;Huber 损失函数;鲸鱼优化算法
中图分类号:TP183 **文献标识码**:A **文章编号**:1673-629X(2023)03-0120-07
doi:10.3969/j.issn.1673-629X.2023.03.018

Clustering and LSTM-based Robust Day-ahead Hourly Forecasting of Photovoltaic Power

LIU Xing-lin[1,2],HUANG Chao[1,2],WANG Long[1,2],LUO Xiong[1,2]
(1. Shunde Graduate School,University of Science and Technology Beijing,Foshan 528399,China;
2. School of Computer and Communication Engineering,University of Science and Technology Beijing,
Beijing 100083,China)

Abstract:Due to its inexhaustible nature and friendliness to environment,solar energy is considered to be one of the most promising energy alternatives to fossil fuels. Photovoltaics (PV) is the most widely used technology to make use of solar energy. However,PV power generation is uncertain due to the intermittent nature of solar energy. With the increasing installation of PV power plants,accurate forecasting of PV power generation is critical to grid management and power dispatch. We propose a robust day-ahead hourly PV power generation forecasting method based on K-means clustering algorithm and long-short-term memory (LSTM) neural network. The K-means algorithm is firstly used to classify PV data into clear sky weather type and rainy weather type based on day-ahead forecasting of weather variables,and then LSTM-based deep learning forecasting models are developed for the corresponding types of data. In order to enhance the robustness of the forecasting model,the Huber loss function is selected for model training,and the whale optimization algorithm (WOA) is selected to optimize the hyperparameter in the Huber loss function. The forecasting performance of the proposed method is compared with benchmarks. The results show that the proposed method can achieve higher forecasting accuracy.

Key words:photovoltaics power generation forecasting;long-short-term memory neural network;K-means clustering;Huber loss function;whale optimization algorithm

0 引 言

随着煤炭、石油、天然气等化石燃料的不断消耗,环境、能源问题成为世界关注的焦点。太阳能作为具有高可用性且用之不竭的清洁能源,被认为是最有前

收稿日期:2022-05-25 **修回日期**:2022-09-27
基金项目:国家自然科学基金(62002016);广东省基础与应用基础研究基金(2019A1515111165,2020A1515110431);佛山市人民政府科技创新专项基金(BK20BF010,BK21BF001)
作者简介:刘兴霖(1998-),男,硕士研究生,研究方向为数据挖掘;通信作者:黄 超(1988-),男,博士,副教授,CCF 会员(A5866M),研究方向为数据挖掘、计算智能和能源信息学。

途的能源替代品之一,各国都十分重视太阳能产业的发展,特别是光伏产业[1]。建设光伏系统已经成为可持续发展的重要内容[2]。国际能源署发布的 2020 年全球光伏报告显示,截止 2020 年底,全球累计光伏装机容量 760.4 GW,中国、欧盟和美国的新增光伏装机容量分别以 48.2 GW、19.6 GW 和 19.2 GW 的规模位列前三[3]。然而,具有高波动性和间歇性的太阳能会给电力系统的管理带来巨大挑战[4-5]。因此,随着全球光伏装机容量的不断提升,光伏功率预测的准确性对于电网管理和电力调度至关重要[6]。

预测方法中,有两类常用的方法,即物理法和统计法。物理法的优势在于不需要历史运行数据,但在对复杂天气的抗干扰能力方面有一定缺陷[7-8]。统计法包括线性算法如自回归等,还包括非线性算法如人工神经网络[9]、支持向量机(support vector machine, SVM)[10]、高斯过程回归(Gaussian process regression, GPR)[11]、深度学习(deep learning, DL)[12-14]等。深度学习具有较强的抗干扰性,能够更好地挖掘出数据之间的关系。而深度学习模型中的长短期记忆(long-short-term memory, LSTM)神经网络[15-16]是对循环神经网络(recurrent neural network, RNN)[17]的优化,引入了“记忆块”,可以使得 LSTM 对历史数据进行充分的学习。K-means 算法是一种经典的聚类算法,文献[18]通过 K-means 算法对训练样本集进行聚类分析,在聚类得到的各类别数据上分别训练支持向量机,并根据预测样本的类别选择对应的支持向量机进行发电功率预测。

基于神经网络的预测模型通常以均方误差(mean square error, MSE)为损失函数训练模型[19]。以 MSE 为损失函数训练模型易于收敛,但 MSE 对离群点敏感。而光伏发电功率高度依赖于太阳辐射值,可在数分钟内剧烈变化,较易产生离群点。为此,以通过降低对离群点惩罚程度的 Huber 损失函数训练模型,可使得模型对离群点更加鲁棒[20]。在此基础上,可采用智能优化算法对 Huber 损失函数中的超参数进行优化。而高效的全局搜索技术才能更好地解决参数寻优问题。鲸鱼优化算法[21](whale optimization algorithm, WOA)是一种新型启发式优化算法,并且已被用于优化各个领域。相比于遗传算法等传统进化算法,WOA 具有计算简单和收敛速度快等优点。因此,该文选择 WOA 作为 Huber 损失函数的优化算法。

基于以上分析,该文提出基于 K-means 聚类分析和 LSTM 算法相结合的光伏发电功率日前逐时鲁棒预测模型。首先,基于 K-means 算法以天气预报数据为特征将光伏数据进行分类,再针对每一类数据分别建立基于 LSTM 算法的预测模型。为了提升模型的鲁棒性,使用 Huber 损失函数,并结合 WOA 训练模型。为验证提出的预测模型的有效性,以 GEFCom2014 能源预测竞赛[22]中的光伏数据开展实例研究。研究结果表明:(1)与 MSE 和一般的 Huber 损失函数相比,经 WOA 优化的 Huber 损失函数可有效提升模型的预测精度;(2)与传统的 BP 神经网络、LSTM、Autoformer[23]、时间融合 Transformers (temporal fusion transformers, TFT)[24]及基于决策树算法的梯度提升框架模型 LightGBM[25]相比,K-means 与 LSTM 相结合的预测模型可进一步提升预测精度。

1　算　法

1.1　K-means 算法

一般来说,天气状况分为两种类型:如晴空天气与阴雨天气,利用 K-means 聚类算法分析太阳辐照度、温度、降雨量等环境因素,按照晴空天气与阴雨天气各自聚类,以实现将数据集按照不同天气类型分类。该算法将样本根据相似度聚集到 k 个聚簇当中,最终实现各个簇内相似度高,簇间相似度低[26]。步骤如下:

步骤一:从数据集中随机选择 k 个样本数据,并作为初始聚类中心 $\{\mu_1,\mu_2,\cdots,\mu_k\}$;

步骤二:计算剩余样本到每一个初始中心点的欧氏距离,选择距离最近的初始聚类中心形成 k 簇。距离计算公式如式(1)所示:

$$d=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2 \tag{1}$$

式中,x 为样本空间中的样本;μ_i 为簇 C_i 的质心。

步骤三:对各个簇重新计算聚类中心。聚类中心计算公式如式(2)所示:

$$\mu_i=\frac{1}{C_i}\sum_{x\in C_i}x \tag{2}$$

最后重复步骤二和步骤三直到条件被满足或者达到最大迭代次数而终止,终止条件为:

$$|\mu_{n+1}-\mu_n|\leqslant\varepsilon \tag{3}$$

式中,ε 为阈值条件。

1.2　LSTM 算法

LSTM 神经网络由输入层、隐含层和输出层组成,其中隐含层包含输入门、遗忘门、输出门和具有特殊记忆细胞的神经单元,其单元结构如图 1 所示。

在 LSTM 单元中,首先,遗忘门的输入为当前的输入 X_t 和上一时刻的输出 H_{t-1},将这些输入通过 sigmoid 激活函数把数值压缩到 0 ~ 1 之间。当信息乘以 0 时,代表全部保留;当信息乘以 1 时,代表全部舍弃,以此判断哪些信息需要被抛弃并重置记忆细胞。其次,输入门通过 sigmoid 激活函数控制需要送入记

忆细胞的信息,并结合输入门和候选值 $\tilde{C}_t$ 更新细胞状态 C_t。输出门通过 sigmoid 激活函数控制记忆细胞中需要被输出的信息。最终,结合输出门和通过 tanh 层的新细胞状态 C_t,得到输出值 H_t。各变量之间的计算公式如下:

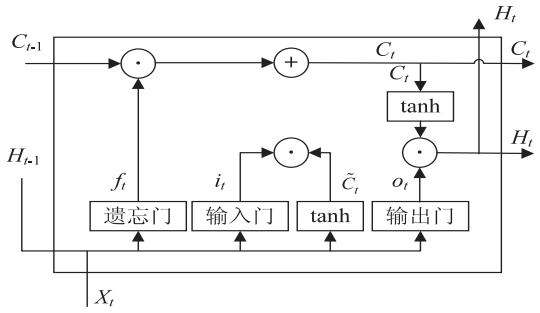

图 1 LSTM 单元结构

$$f_t = \sigma(\boldsymbol{W}_f X_t + U_f H_{t-1} + b_f) \tag{4}$$

$$i_t = \sigma(\boldsymbol{W}_i X_t + U_i H_{t-1} + b_i) \tag{5}$$

$$o_t = \sigma(\boldsymbol{W}_o X_t + U_o H_{t-1} + b_o) \tag{6}$$

$$\tilde{C}_t = \tanh(\boldsymbol{W}_c X_t + U_c H_{t-1} + b_c) \tag{7}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{8}$$

$$H_t = o_t \tanh(C_t) \tag{9}$$

式中,f_t 为遗忘门;i_t 为输入门;o_t 为输出门;C 为记忆细胞状态的向量值;$\boldsymbol{W}$ 为隐藏单元的输入权重矩阵;$\boldsymbol{U}$ 为输出权重矩阵;b 为偏置向量;下标 $t-1$ 和 t 分别代表不同的时间步长;σ 为 sigmoid 函数。

1.3 鲸鱼优化算法

在 WOA 过程中,每条鲸鱼都有两种行为。一种是包围猎物,鲸鱼会迅速包围猎物,螺旋形缩小圈。另一种是利用泡沫网,鲸鱼发现猎物后,排出气泡迫使猎物聚集。在每一个迭代过程中,鲸鱼会在这两种行为之间随机选择进行捕猎。为了模拟这些行为,根据随机概率 q,在每一代鲸鱼中以 50% 的概率选择两种行为之一,并定义为:

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{A}\,|\vec{C}\,\overrightarrow{X^*} - \vec{X}(t)\,| & q < 0.5 \\ |\overrightarrow{X^*}(t) - \vec{X}(t)\,|e^{bL}\cos(2\pi L) + \overrightarrow{X^*}(t) & q \geqslant 0.5 \end{cases} \tag{10}$$

式中,t 表示当前迭代;$\vec{X}$ 表示当前鲸鱼的位置向量;$\overrightarrow{X^*}$ 为当前鲸鱼的最佳位置;$\vec{A}$ 和 $\vec{C}$ 表示实现鲸鱼包围螺旋收缩的系数向量;b 为常数;L 表示在[0,1]区间内的随机数。

2 光伏发电预测模型

2.1 预测模型结构

该文研究基于日前逐时天气预报信息的光伏功率预测方法。主要天气预报信息包括:总液态水、总冰水、表面压力、1 000 毫巴时的相对湿度、总云量、10 米水平风分量、10 米垂直风分量、2 米温度、地面太阳辐射总量、地面散热总量、大气顶部的净太阳辐射总量、总降水量。这些变量都是基于卫星数据的预测量。

针对光伏功率特性,该文提出 K-means 和 LSTM 相结合的光伏发电功率日前逐时预测模型,即先用 K-means 算法以天气预报信息为特征将数据进行分类,再针对不同类的数据建立基于 LSTM 算法的预测模型。整体流程如图 2 所示,对关键步骤的描述如下:

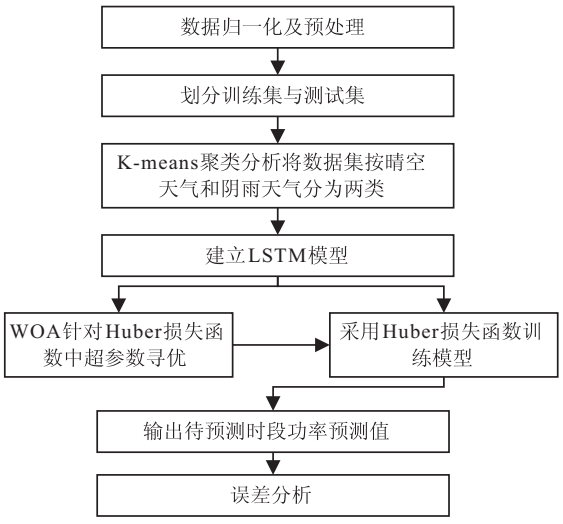

图 2 光伏输出功率的预测过程

步骤一:数据预处理,如:数据归一化以及累积变量的预处理,并划分训练集和测试集。

步骤二:天气聚类分析。从样本数据集中总结并提炼出能够明显表现天气类型的特征数据,如单日日间的太阳辐照度和降雨量等。利用 K-means 聚类算法和提取后的特征数据,并根据晴空天气和阴雨天气设置 $k=2$,将训练集和测试集分为晴空天气训练集、晴空天气测试集、阴雨天气训练集和阴雨天气测试集。

步骤三:建立基于 LSTM 算法的预测模型。模型输入为时间和日前逐时天气预报信息;模型的输入层神经元个数为输入数据维度;考虑到模型拟合效果和训练时间,隐含层的层数设定为 3 层,并设定 Dropout 值以防止过拟合;输出层的前一层为全连接层;模型输出为预测时刻的发电功率。

步骤四:采用 Huber 损失函数(详见 2.2 节)训练 LSTM 神经网络,同时引入 WOA,针对各个模型中的 Huber 损失函数,优化得到对应的超参数 δ,并将采用优化后所得到的结果与优化前和使用 MSE 的结果进行对比。

步骤五:算例验证与实验结果分析。将 K-means 聚类分析后的训练集和测试集,结合 LSTM 进行训练和预测,得到预测结果后,将其评价指标与 BP、LSTM、

Autoformer、TFT 及 LightGBM 算法进行比较。

2.2　损失函数

采用 Huber 损失函数用于预测模型的训练。Huber 损失函数是一种用于回归问题的带参损失函数,与常用的 L1、L2 损失函数相比,降低了对离群点的惩罚程度,并结合 MSE 和 MAE 的优点,对异常点更加鲁棒,以此来提高模型的鲁棒性。Huber 损失函数如式(11)所示。当 $|p_i - P_i| \leqslant \delta$ 时,Huber 损失等价为 MSE;当 $|p_i - P_i| > \delta$ 时,Huber 损失趋向于 MAE。

$$\text{Huber} = \begin{cases} \dfrac{1}{2}(p_i - P_i)^2 & |p_i - P_i| \leqslant \delta \\ \delta |p_i - P_i| - \dfrac{1}{2}\delta^2 & |p_i - P_i| > \delta \end{cases} \tag{11}$$

式中,p_i 为光伏发电功率实际值;P_i 为光伏发电功率预测值;δ 为阈值,用来判断模型应如何处理异常值。

3　实验研究

文中采用的实验环境为 Windows10 操作系统,在 Pycharm2021.1 的 anaconda 环境下使用 python 进行编程,版本号为 3.6,并搭建 TensorFlow 框架,版本号为 2.4。

3.1　数据预处理

所选数据集来自于 2014 年 GEFCom2014 能源预测竞赛。选择该竞赛中光伏预测方向的数据集,该数据集取自澳大利亚某个地区的三座太阳能发电厂,时间段为 2012 年 4 月 1 日至 2014 年 7 月 1 日,数据时间分辨率为 1 h。在 K-means 聚类分类和建立 LSTM 神经网络之前对所有特征进行归一化处理,使得各数值统一取值为 0 到 1 之间。此外,光伏发电功率在原始数据中已做归一化处理,所以在后续的评价指标计算时,得到的均为归一化后的值。

数据预处理中首先要处理的是四个累积变量,即地面太阳辐射总量、地面散热下降总量、大气顶部的净太阳辐射总量和总降水量。这四个变量在每一时刻的数据都是单日内从零时起到某时的累积量,所以需要将这四个累积变量的各个时刻都处理为每小时的增量。处理后的四个新变量命名为每小时地面太阳辐射量、每小时地面散热量、每小时大气顶部的净太阳辐射量和每小时降水量。

其次要处理的是时间戳,将时间信息字符串处理为可以直接输入到模型中的年、月、日、小时四个变量。原始数据采用的时间是世界标准时间,但由于当地时间与预期的太阳辐射进程是匹配的,所以在后续实验中根据当地时间来处理更为便捷。最后,由于需要对三个电厂的所有数据进行实验,但电厂 2 和电厂 3 中存在部分无效数据。在剔除它们后,将 2012 年 4 月 1 日至 2014 年 3 月 30 日内的数据设定为训练集,将 2014 年 4 月 1 日至 2014 年 7 月 1 日内的数据设定为测试集。最终训练集以及测试集划分后的天数如表 1 所示。

3.2　聚类模型实验研究

可用于天气聚类分析的信息包括日内 24 个时间点的气象变量,数据维度较高。为提升天气分类的准确度,天气预报数据中选出或衍生出更具代表意义的特征。通过分析特征与光伏发电量的相关性,选择以下气象变量用于天气聚类分析:单日最高地面太阳辐射量、单日最高净太阳辐射量、单日日间最高降水量、单日日间最高液态水量、单日日间最高冰水量、单日日间平均云量,其中单日日间设定的时间段为 6:00-18:00。结合 K-means 聚类算法将光伏发电日的天气类型分为两类,即晴空天气和阴雨天气。三个电厂的训练集和测试集天气聚类分析结果如表 1 所示。

表 1　各电厂训练集和测试集的划分和聚类分析结果

电厂/天	训练集			测试集		
	晴空天气	阴雨天气	总数	晴空天气	阴雨天气	总数
电厂 1	508	222	730	55	35	90
电厂 2	514	190	704	58	24	82
电厂 3	523	196	719	56	34	90

3.3　LSTM 预测模型实验研究

根据 3.2 中的聚类结果,针对晴空天气类型训练子集和阴雨天气类型训练子集以及完整训练集分别训练 LSTM 预测模型,并依次命名为 LSTM1、LSTM2 和 LSTM3。根据 2.1 中的 LSTM 网络结构,基于 Python 的 TensorFlow 框架中的 keras. layers. LSTM 搭建 LSTM 神经网络。预测模型的输入包括时间与气象特征,共 13 个:小时、总液态水、总冰水、表面压力、1 000 毫巴时的相对湿度、总云量、10 米水平风分量、10 米垂直风分量、2 米的温度、每小时地面太阳辐射、每小时地面散热下降量、每小时大气顶部的净太阳辐射量和每小时降水量。同时,三个 LSTM 模型参数设置为:第一层 LSTM 神经元个数及其 dropout 均为 50 和 0.4;而第二层的神经元个数以及 dropout 分别设置为 LSTM1:100、0.2,LSTM2:80、0.4,LSTM3:100、0.2;第三层神经元个数均为 50;batch_size 均为 32;最后的 Dense 层神经元个数为 1。

在 LSTM 模型的训练阶段,选择 Huber 损失函数并采用 WOA 优化其超参数 δ 来提高模型的性能。由于该文预测任务中的光伏发电功率数据为归一化值,可将需要优化的 δ 的搜索范围设定为[0.000 01,1],

然后进行迭代优化。WOA 参数设置为:种群规模为 80,最大迭代次数为 500。同时,与使用 MSE 和一般 Huber 损失函数获得的结果进行对比。

3.4 评价指标

以平均绝对误差(mean absolute error,MAE)、均方根误差(root mean square error,RMSE)、决定系数 R^2 为指标进行误差计算,计算公式为:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i-P_i)^2} \tag{12}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|p_i-P_i| \tag{13}$$

$$R^2=\frac{\sum_{i=1}^{N}(P_i-\overline{p_i})^2}{\sum_{i=1}^{N}(p_i-\overline{p_i})^2} \tag{14}$$

式中,N 为预测样本数量;p_i 为光伏发电功率实际值;P_i 为光伏发电功率预测值;$\overline{p_i}$ 为光伏发电功率平均值。

4 预测结果分析

为展示损失函数对光伏功率预测性能的影响,分别采用 MSE 损失函数、一般 Huber 损失函数和 WOA 优化后的 Huber 损失函数来训练 LSTM1、LSTM2 和 LSTM3。表 2 比较了基于不同损失函数训练的模型在测试集上的预计精度(RMSE 值)。

表 2 采用不同损失函数时的 RMSE 误差结果对比

电厂	LSTM 模型	测试集类型	损失函数/RMSE 值		
			MSE	Huber	WOA+Huber
电厂 1	LSTM1	晴空天气	0.068 36	0.068 40	0.067 79
	LSTM2	阴雨天气	0.067 33	0.065 80	0.064 68
	LSTM3	晴空天气	0.068 30	0.067 09	0.066 13
		阴雨天气	0.070 33	0.068 59	0.067 52
		所有天气	0.069 18	0.068 38	0.067 62
电厂 2	LSTM1	晴空天气	0.072 19	0.070 52	0.068 91
	LSTM2	阴雨天气	0.059 47	0.058 54	0.057 11
	LSTM3	晴空天气	0.069 50	0.068 94	0.066 92
		阴雨天气	0.061 89	0.060 65	0.059 88
		所有天气	0.067 28	0.066 03	0.065 89
电厂 3	LSTM1	晴空天气	0.069 66	0.068 74	0.066 53
	LSTM2	阴雨天气	0.074 82	0.073 47	0.071 74
	LSTM3	晴空天气	0.067 85	0.066 24	0.065 13
		阴雨天气	0.076 44	0.076 56	0.075 55
		所有天气	0.072 73	0.071 92	0.070 88

表 2 表明,与 MSE 损失函数相比,采用一般 Huber 损失函数训练 LSTM 神经网络,大多数情况下都能获得 1 % 左右的精度提升。而相比于一般的 Huber 损失函数,采用 WOA 优化后的 Huber 损失函数能够更加适应模型,并且 RMSE 值也降低了 1% ~2%。由此可以得出对于一般的光伏功率预测问题,不仅一般的 Huber 损失函数能够通过提升模型的鲁棒性,得到比 MSE 更加精确的预测结果,采用 WOA 优化后的 Huber 损失函数可以得到优于前两者的实验精度。说明光伏数据中离群点的存在会影响模型的精度,并且再经过 WOA 的优化,可以使得 Huber 损失函数更加贴合模型,提高其学习能力。为此,后续实验结果分析都基于 WOA 优化后 Huber 损失函数训练的预测模型。

表 3 比较了各预测模型在测试集上对不同天气类型的预测精度。测试结果表明:在晴空天气下,LSTM3 模型比 LSTM1 模型的表现更优,即选用完整训练集训练预测模型可提高晴空天气条件下的预测精度;在阴雨天气条件下,LSTM2 模型比 LSTM3 模型的表现更好,即选用阴雨天气训练子集训练预测模型可提高阴雨条件下的预测精度。

表 3 不同天气条件下的预测性能

电厂	测试集类型	LSTM 模型	RMSE	MAE	R^2
电厂 1	晴空天气	LSTM1	0.067 79	0.032 28	0.913 0
		LSTM3	0.066 13	0.030 82	0.917 3
	阴雨天气	LSTM2	0.064 68	0.029 66	0.854 1
		LSTM3	0.067 52	0.031 22	0.801 3
电厂 2	晴空天气	LSTM1	0.068 91	0.034 34	0.902 7
		LSTM3	0.066 92	0.033 77	0.922 2
	阴雨天气	LSTM2	0.057 11	0.028 97	0.830 1
		LSTM3	0.059 88	0.028 13	0.812 4
电厂 3	晴空天气	LSTM1	0.066 53	0.031 78	0.923 6
		LSTM3	0.065 13	0.028 95	0.927 9
	阴雨天气	LSTM2	0.071 74	0.032 13	0.774 3
		LSTM3	0.075 55	0.034 52	0.786 9

图 3 展示了 2014 年 4 月 23、24 日晴空天气下三个电厂的 LSTM3 预测结果;图 4 展示了 2014 年 5 月 4、5 日阴雨天气下三个电厂连续五天的 LSTM2 预测结果。所使用的数据集为两年以上的数据,由于 K-means 分类为晴空天气和阴雨天气两类数据时存在分类不够准确的情况,如:在分类后的晴空天气数据集中存在 2% 左右的阴雨天气数据。所以在晴空天气下,使用完整训练集得到的模型能够对小部分的阴雨天气有着更好的拟合效果。同理,分类后的阴雨天气训练

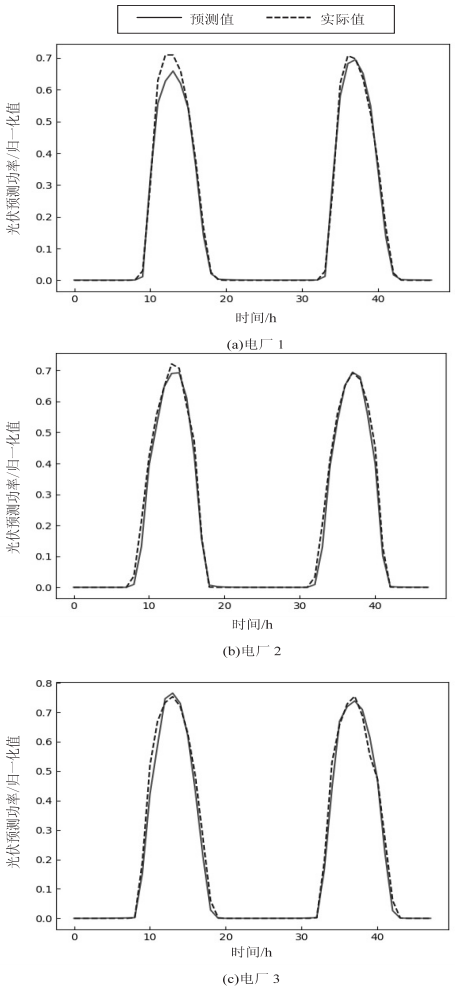

图 3　三个电厂在晴空天气条件下连续五天的光伏功率预测结果

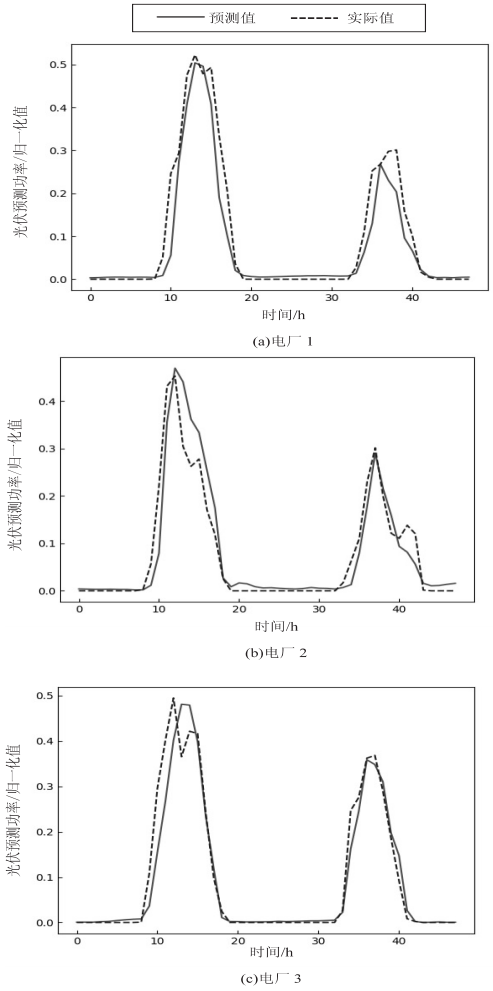

图 4　三个电厂在阴雨天气条件下连续五天的光伏功率预测结果

子集与测试子集也存在小部分晴天数据,但数据中都仍以阴雨天气为主,所以相比于以晴空天气为主的未分类训练集,采用分类后的阴雨天气训练子集可使测试数据达到更好的拟合效果。但又限于阴雨天气条件下的功率预测会受到更加复杂的因素影响,如单日内波动较大的降雨量、降雪量、温度、总云量等,其总体误差均比晴空天气条件下的偏大。

以上分析表明,基于 K-means 聚类分析结果,LSTM2 与 LSTM3 结合即 LSTM2 用于聚类分析后的阴雨天气预测和 LSTM3 用于聚类分析后的晴空天气预测两种方法的结合可提高整体的预测精度。表 4 中展示了基于 K-means 和结合后的 LSTM 模型方法在整个测试集上的预测性能。

为进一步验证提出的 K-means 天气聚类分析与 LSTM 预测模型相结合对光伏发电功率预测的有效性,比较了所提出的预测方法与传统 BP 和 LSTM 算法以及 Autoformer、TFT 和 LightGBM 算法,结果如表 4 所示。所有对比方法的训练均基于未进行天气分类的完整训练集,预测结果均为相应模型在整个测试集上的表现。

表 4　不同预测方法预测性能比较

电厂	模型	RMSE	MAE
电厂 1	K-means+LSTM	0.065 24	0.030 37
	Autoformer	0.065 94	0.030 12
	TFT	0.066 29	0.031 08
	LightGBM	0.070 23	0.030 86
	LSTM	0.067 62	0.031 43
	BP	0.076 79	0.034 48
电厂 2	K-means+LSTM	0.061 28	0.030 75
	Autoformer	0.064 37	0.031 47
	TFT	0.063 35	0.031 82
	LightGBM	0.066 90	0.030 97
	LSTM	0.065 89	0.032 44
	BP	0.075 53	0.034 92
电厂 3	K-means+LSTM	0.067 70	0.030 19
	Autoformer	0.068 52	0.031 48
	TFT	0.068 24	0.031 15
	LightGBM	0.069 58	0.031 24
	LSTM	0.070 88	0.031 59
	BP	0.078 74	0.033 89

根据表 4 给出的结果,可以看出与其他算法相比,除 Autoformer 与文中模型在电厂 1 中的误差比较接近,提出的预测模型的评价指标在三个电厂的数据集中均为最优。与传统的 LSTM 相比,提出的模型在电厂 1、电厂 2、电厂 3 中的平均 RMSE 与 MAE 分别降低了 4.86% 与 4.33%;与 Autoformer 算法相比,提出的模型在三个电厂中的平均 RMSE 与 MAE 分别降低了 2.35% 、1.85%;与 TFT 算法相比,文中模型的平均 RMSE 与 MAE 分别降低了 1.88% 与 2.91%;与 LightGBM 算法相比,文中模型的平均 RMSE 与 MAE 分别降低了 6.06% 与 1.89%。结果证明了文中模型和数据处理方法的可行性与有效性,并且在预测精度和稳定性方面均表现出较好的性能。

5 结束语

该文提出基于 K-means 的天气聚类分析和 LSTM 相结合的光伏发电功率日前逐时预测方法,并从损失函数和训练集的选择上做出研究分析,结果表明:

(1)在光伏发电预测模型中损失函数的选择上,Huber 损失函数相比于 MSE 具有更强的鲁棒性,能更好地处理光伏数据中的离群点,在此基础之上,选择鲸鱼优化算法对 Huber 损失函数中的 δ 实现进一步的优化,从而能有效地提高光伏预测模型的准确度,总体可将预测精度提高 2% 左右。

(2)基于 K-means 算法将数据分类为晴空天气类型和阴雨天气类型,在聚类得到的各类别训练集上分别训练 LSTM,并根据测试集的类别选择对应的 LSTM 进行光伏发电功率的预测。研究结果表明,提出的 K-means 与 LSTM 相结合的预测方法比单独的 LSTM 以及 LightGBM 的预测效果更好,预测精度提升在 4% 左右,同时对比 Autoformer 算法和 TFT 算法,预测精度提升约 2%。

(3)基于 K-means 算法对天气进行聚类分析时,存在对一小部分天气数据分类错误的情况。在后续的研究工作中,可结合其他机器学习算法,提高分类的准确率,进一步提高预测精度。

参考文献:

[1] DUDA J,KUSA R,PIETRUSZKO S,et al. Development of roadmap for photovoltaic solar technologies and market in Poland[J]. Energies,2021,15(1):1-25.

[2] 张 悦. 浅谈光伏产业的发展建议[J]. 能源与节能,2021(3):18-19.

[3] ZHANG J. Solar PV market research and industry competition report[J]. IOP Conference Series:Earth and Environmental Science,2021,632(3):032047.

[4] ZHANG F,WANG X H,WU M Y,et al. Optimization design of uncertain parameters for improving the stability of photovoltaic system[J]. Journal of Power Sources,2022,521:471-492.

[5] SENS L,NEULING U,KALTSCHMITT M. Capital expenditure and levelized cost of electricity of photovoltaic plants and wind turbines-Development by 2050[J]. Renewable Energy,2022,185(C):525-537.

[6] KOEHLER C, STEINER A, SAINT-DRENAN Y,et al. Critical weather situations for renewable energies-part B:low stratus risk for solar power[J]. Renewable Energy,2017,101(C):794-803.

[7] ZAZOUM B. Solar photovoltaic power prediction using different machine learning methods[J]. Energy Reports,2022,8(1):19-25.

[8] KUSZNIER J,WOJTKOWSKI W. IoT solutions for maintenance and evaluation of photovoltaic systems[J]. Energies,2021,14(24):1-24.

[9] 王 巍. 基于人工神经网络和模拟集成的短期光伏发电预测[J]. 可再生能源,2019,37(5):670-675.

[10] GONZÁLEZ C, MARULANDA J, RESTREPO C,et al. Hardware-in-the-loop to test an MPPT technique of solar photovoltaic system:a support vector machine approach[J]. Sustainability,2021,13(6):1-16.

[11] NAJIBI F,APOSTOLOPOULOU D,ALONSO E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast[J]. International Journal of Electrical Power and Energy Systems,2021,130:106916.

[12] 叶 林,裴 铭,路 朋,等. 基于天气分型的短期光伏功率组合预测方法[J]. 电力系统自动化,2021,45(1):44-54.

[13] 徐先峰,刘阿慧,陈雨露,等. 基于气象因素充分挖掘的 BiLSTM 光伏发电短期功率预测[J]. 计算机系统应用,2020,29(7):205-211.

[14] 宋绍剑,李博涵. 基于 LSTM 网络的光伏发电功率短期预测方法的研究[J]. 可再生能源,2021,39(5):594-602.

[15] CORTEZ B,CARRERA B,KIM Y J,et al. An architecture for emergency event prediction using LSTM recurrent neural networks[J]. Expert Systems with Applications,2018,97:315-324.

[16] KUMAR D,MATHUR H D,BHANOT S,et al. Forecasting of solar and wind power using LSTM RNN for load frequency control in isolated microgrid[J]. International Journal of Modelling and Simulation,2021,41(4):311-323.

[17] BEIGI M,BEIGI H H,TORKI M,et al. Forecasting of power output of a PVPS based on meteorological data using RNN approaches[J]. Sustainability,2022,14(5):1-12.

[18] WENG G P,PEI C X,REN J R,et al. Photovoltaic output prediction of regional energy Internet based on LSTM algorithm[J]. Journal of Physics:Conference Series,2021,1732

(下转第 138 页)