

DOI: 10.19783/j.cnki.pspc.190611

基于 MIE-LSTM 的短期光伏功率预测

吉 铮 格, 李 慧, 刘 思 嘉, 王 丽 婕

(北京信息科技大学自动化学院, 北京 100192)

摘要: 提升精细化的光伏预测技术对电力系统的实时调度运行至关重要。它不仅依赖于预测模型的优劣, 还依赖于训练样本日与预测日的相似程度。提出一种基于 MIE-LSTM 的短期光伏功率预测方法。在建立基于互信息熵 (Mutual Information Entropy, MIE) 的相关性衡量指标基础上, 计算出光伏功率与各气象因素间的互信息熵, 从而对高维气象数据进行降维处理。然后, 利用历史日与预测日多维气象因素间的加权互信息熵筛选出相似日样本。最后, 通过长短期记忆 (Long-short Term Memory, LSTM) 神经网络预测模型训练并建立气象因素与光伏出力之间的映射关系。通过对某实测光伏电站不同天气类型下的发电功率进行预测分析, 验证了新方法能够达到理想的预测精度。

关键词: 光伏功率预测; 数值天气预报; 互信息熵; 相似日; 长短期记忆神经网络

Short-term photovoltaic power forecasting based on MIE-LSTM

Ji Xinge, Li Hui, Liu Sijia, Wang Lijie

(School of Automation, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract: It is crucial for the real-time scheduling operation of the power system to improve the refined photovoltaic power forecasting technology. It depends not only on the superiority and inferiority of the predictive model, but also relies on the similarity between the training sample day and the forecast day. A novel method based on MIE-LSTM is proposed for short-term photovoltaic power forecasting. The correlation metrics is established based on Mutual Information Entropy (MIE), the MIE between photovoltaic power and meteorological factors is calculated to reduce the dimension of high-dimensional meteorological data. Then, the weighted mutual information entropy between the multi-dimensional meteorological factors of historical day and those of predicted day is used to screen out similar day samples. Finally, the mapping relationship between meteorological factors and photovoltaic output is established via training Long-Short Term Memory (LSTM) neural network prediction model. Through forecasting and analyzing the power generation of a photovoltaic power station under different weather types, it is verified that the new combination method can achieve ideal forecasting precision.

This work is supported by National Natural Science Foundation of China (No. 51607009), Beijing Natural Science Foundation (No. 3172015), and Scientific Research Project of Beijing Education Council (No. KM201911232016).

Key words: short-term photovoltaic power forecasting; numerical weather prediction; mutual information entropy; similar day; long-short term memory neural network

0 引言

近年来, 全世界范围内太阳能、风能等清洁能源得到迅速发展, 环境恶化和一次能源过度消耗的

基金项目: 国家自然科学基金项目资助 (51607009); 北京市自然科学基金项目资助 (3172015); 北京市教委科技计划面上项目资助 (KM201911232016)

问题在一定程度上得以改善^[1]。但是, 随着大规模光伏电站的接入, 其随机性和波动性给电力系统安全稳定运行带来威胁。为了保证电力系统安全可靠地运行, 每年都发生大规模的弃光现象。解决这一问题的有效途径就是进行短期光伏功率预测。短期光伏功率预测的精度直接影响了其并网的规模和范围, 也对解决能源消耗和环境恶化之间的矛盾有着重要的意义^[2]。

目前, 短期光伏功率预测方法根据预测模型建立方式的不同可分为物理法^[3]和统计法^[4]。物理方法无需大量历史数据, 利用场站地理信息及精确的组件参数, 建立复杂的组件运行方程及辐照度-功率传递方程。该方法在晴天条件下可以取得较高的预测精度, 但应对复杂天气类型时, 由于建模过程极其复杂且模型参数难以确定, 功率预测精度无法满足要求^[5]。统计法是采用人工神经网络^[6]、支持向量机^[7-8]等智能方法, 挖掘光伏功率与多维气象因素间的非线性映射关系, 但是传统的智能方法存在泛化能力不强和易陷入局部最优等问题, 很难准确描述此类复杂非线性映射关系^[9]。

近年来, 挖掘数值天气预报(Numerical Weather Prediction, NWP)中多维气象因素与光伏功率之间映射关系的深度学习算法^[10]受到广泛关注。文献[11]通过建立双层 ELMAN 神经网络挖掘相关气象因子与光伏功率的映射关系, 验证了该模型具有良好的预测效果。文献[12]采用基于受限玻尔兹曼机的深度置信网络与 BP 神经网络组合, 建立组合深度学习模型, 该方法减少了 BP 网络的训练时间, 避免了其训练过程易陷入局部最优的问题^[13]。Hochreiter 等人^[14]在 1997 年首次提出 LSTM 神经网络, 通过在隐层引入门控结构使其对长时间序列信息具有记忆(选择)和遗忘(剔除)功能, 该方法在短期负荷预测领域有较好的应用^[15-16], 同时其也被验证在长时间序列信息挖掘中的有效性^[17]。

提高短期光伏功率预测精度不仅依赖于预测模型的优劣, 还依赖于模型训练样本与预测日信息的相似程度^[18-19]。文献[20]以 Hottel 模型数据为参考量, 筛选出与预测日天气类型相同、温度欧氏距离较短的历史日作为相似日样本, 得到理想的预测精度。文献[21]采用 K 均值聚类算法筛选出相似日样本, 运算效率较高, 但当预测日天气波动较大时, 构建的预测日 NWP 特征向量复杂程度也随之增大。文献[22]利用各气象因素与光伏功率的秩相关系数, 构建了历史日与预测日的相似度函数, 从而筛选出相似日样本, 该方法在多种天气类型下均取得了较高的预测精度。

互信息是信息论中一种度量随机变量间信息共享程度的方式, 它可以表征一个随机变量中包含另一个随机变量信息量的多少, 或者说是一个随机变量在多大程度上受另一随机变量的影响而减少其自身的不确定性^[23]。目前, 互信息理论已经广泛应用于热力学、气象学和水文学等领域^[24-25]。为了消除互信息量纲的影响, 互信息熵 MIE 的概念被用来量

化随机变量之间的相关性^[26]。文献[27]采用 MIE 对多维气象因素与风功率之间的非线性相关性进行了量化评估, 验证了其在非线性相关性辨识中优于线性相关和秩相关。

本文提出一种基于 MIE-LSTM 的短期光伏功率预测方法。首先, 建立基于 MIE 的相关性衡量指标; 然后, 计算光伏出力与各气象因素间的 MIE 相关系数, 对高维气象数据进行降维处理; 接着, 利用历史日与预测日多维气象因素间的加权 MIE 值筛选出相似日样本; 最后, 通过 LSTM 预测模型训练并建立气象因素与光伏出力之间的映射关系。以某实测光伏电站数据和相关 NWP 数据为例, 对比分析了组合预测方法的精度。结果表明, 本文提出的方法预测精度优于其他组合方法。

1 基于互信息熵(MIE)的相关性衡量指标

定义向量 $\mathbf{X} = [x_1, x_2, \dots, x_n]$, n 表示向量 \mathbf{X} 的序列长度。按式(1)对向量 \mathbf{X} 的元素进行归一化处理:

$$\mathbf{X}_0 = \frac{\mathbf{X} - \min(\mathbf{X})}{\max(\mathbf{X}) - \min(\mathbf{X})} \quad (1)$$

式中, \mathbf{X}_0 表示向量 \mathbf{X} 归一化后的结果。

定义 B 为向量 \mathbf{X}_0 的组距, E_i 为向量 \mathbf{X}_0 的波动区间边缘, 其表达式分别为

$$B = \frac{\max(\mathbf{X}_0) - \min(\mathbf{X}_0)}{n} \quad (2)$$

$$E_i = \begin{cases} -\infty & i = 1 \\ \min(\mathbf{X}_0) + B \times (i - 1) & i = 2, 3, \dots, n \\ +\infty & i = n + 1 \end{cases} \quad (3)$$

以数值边界为分界条件, 根据一维直方图分析法, 按照式(3)将向量 \mathbf{X}_0 分为 n 个波动区间, 定义 X_i 为第 i 个波动区间, 其表达式为

$$X_i \subseteq [E_i, E_{i+1}] \quad i = 1, 2, \dots, n \quad (4)$$

定义 $p(\cdot)$ 为边缘概率密度函数, 计算各波动区间 X_i 的边缘概率密度 $p(X_i)$ 为

$$p(X_i) = \frac{n_i}{n} \quad (5)$$

式中, n_i 表示向量 \mathbf{X}_0 中的元素落入波动区间 X_i 的数据点数。

利用自信息熵 $H(X_i)$ 来表征向量 \mathbf{X} 自身随机波动的不确定度, 即

$$H(X_i) = - \sum_{i=1}^n p(X_i) \log_2^{p(X_i)} \quad (6)$$

根据香农信息论^[28], 将一维向量的自信息熵 $H(X_i)$ 扩展为二维向量的互信息 $I(\mathbf{X}; \mathbf{Y})$, 即

$$I(\mathbf{X};\mathbf{Y}) = -\sum_{i=1}^n \sum_{j=1}^n p(X_i Y_j) \log_2 \left(\frac{p(X_i Y_j)}{p(X_i) p(Y_j)} \right) \quad (7)$$

式中： \mathbf{Y} 表示区别于 \mathbf{X} 的另一个代表气象或功率的向量； $p(Y_j)$ 表示波动区间 Y_j 的边缘概率密度； $p(X_i Y_j)$ 是 \mathbf{X}_0 和 \mathbf{Y}_0 在波动区间 X_i 和 Y_j 的联合概率密度。

$I(\mathbf{X};\mathbf{Y})$ 表征了 \mathbf{X} 和 \mathbf{Y} 共享的信息，用来衡量两个变量中一个对另一个不确定性减少的程度。如果 \mathbf{X} 和 \mathbf{Y} 相互独立，则 \mathbf{X} 不对 \mathbf{Y} 提供任何信息，则它们的互信息为零；反之，如果 \mathbf{X} 和 \mathbf{Y} 互相均为彼此的确定性函数，那么 \mathbf{X} 和 \mathbf{Y} 传递的所有信息均共享，则他们的互信息 $I(\mathbf{X};\mathbf{Y})=1$ 。

为了消除互信息量纲的影响，广义互信息 $I(\mathbf{X};\mathbf{Y})$ 再经过归一化处理，得到MIE相关系数 I_{XY} 为

$$I_{XY} = \frac{I(\mathbf{X};\mathbf{Y})}{\sqrt{H(\mathbf{X})H(\mathbf{Y})}} \quad (8)$$

式中， $H(\mathbf{Y})$ 表示向量 \mathbf{Y} 的自信息熵。

MIE描述了变量间概率分布的统计特性，具有非负特性 $I_{XY} \geq 0$ ，相比于广义的线性相关系数而言，适用范围更广，可以更好地挖掘气象因素与光伏发电功率之间的非线性相关性，有利于进行历史相似日样本的筛选。

2 基于LSTM神经网络的预测模型

循环神经网络RNN(Recurrent Neural Network, RNN)是一种以时间序列数据为输入，按照序列的演进方向进行递归运算的神经网络。它具有特殊的隐层神经元结构，对过往的输入信息具有选择性“记忆”功能。也就是说，当前时刻的输出既考虑了当前时刻的输入，还考虑了之前时刻的隐层输出状态。传统的RNN网络是从单层前馈神经网络出发构建递归连接，在长时间序列的学习过程中，易出现梯度消失和梯度爆炸的现象，导致其无法构建出较为精确的非线性映射关系模型。

LSTM神经网络是在RNN结构的基础上，引入门控单元替代RNN隐层中的神经元，使其对较长时间序列上的信息能够选择性地通过和剔除，改善RNN在长时间序列上记忆能力不足、梯度消失和梯度爆炸的问题，图1为LSTM神经网络隐层神经元的结构图。

图中： x_t 为 t 时刻的输入序列； h_{t-1} 、 h_t 分别为 $t-1$ 和 t 时刻的隐层状态； C_{t-1} 、 C_t 分别为 $t-1$ 和 t 时刻记忆单元的状态； \hat{C}_t 为 t 时刻记忆单元的临时状态，记忆单元是控制时序数据信息传递的核心； i_t 表示 t 时刻门控单元中输入门状态，用来控制 x_t 传递到 C_t 信息的多少； o_t 表示 t 时刻输出门状态，

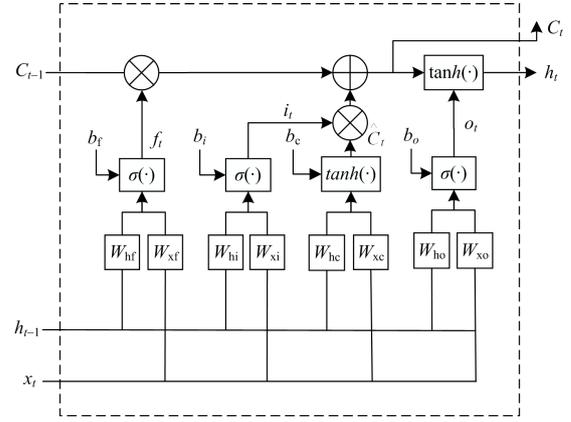


图1 LSTM神经网络隐层神经元的结构图

Fig. 1 Structural diagram of LSTM network hidden layer neurons

用来控制 C_t 传递到 h_t 信息的多少； f_t 表示 t 时刻遗忘门状态，它决定了记忆状态 C_t 保留 C_{t-1} 中信息量的多少。图1中信息的传递过程可表示为式(9)一式(14)。

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (9)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (10)$$

$$\hat{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (11)$$

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \quad (12)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (13)$$

$$h_t = o_t \times \tanh(C_t) \quad (14)$$

式中： σ 和 \tanh 分别为双极性sigmoid激活函数和双曲正切激活函数； W_{xi} 、 W_{hi} 、 b_i 分别为输入层和隐层到输入门的权值向量以及输入门的偏置量； W_{xf} 、 W_{hf} 、 b_f 分别为输入层和隐层到遗忘门的权值向量以及遗忘门的偏置量； W_{xo} 、 W_{ho} 、 b_o 分别为输入层和隐层到输出层的权值向量以及输出层的偏置量； W_{xc} 、 W_{hc} 、 b_c 分别为输入层和隐层到记忆单元的权值向量以及记忆单元的偏置量。

3 预测步骤

3.1 输入-输出数据相关性分析

利用多维的NWP气象数据进行预测之前，首先需要对输入与输出数据间的相关性进行分析，换句话说，就是对高维气象数据进行降维处理。这是因为相关性较弱的NWP数据输入预测模型，不仅会对预测精度造成影响，还会严重影响神经网络的训练及预测效率。

假设各气象变量 X_i ($i=1,2,\dots,5$)分别代表辐照度、温度、云量、风速和湿度，光伏功率表示为 Y 。利用式(1)一式(8)分别计算各气象因素与光伏功率

的 MIE 相关系数 $I_{XY}=[I_{FP}, I_{TP}, I_{CP}, I_{WP}, I_{HP}]$, 其计算结果如表 1 所示。

表 1 各气象因素与光伏功率的 MIE 相关系数

Table 1 MIE coefficients between meteorological factors and photovoltaic power

气象因素	相关系数	I_{XY} 值
辐照度 F	I_{FP}	0.918 0
温度 T	I_{TP}	0.901 7
云量 C	I_{CP}	0.414 3
风速 W	I_{WP}	0.559 7
湿度 H	I_{HP}	0.660 5

由表 1 可知, 根据 MIE 相关系数由大到小来判断, 相关性由高到低依次是辐照度、温度、湿度、风速和云量, 而且前两者的相关系数都大于 0.9。为了更好地挖掘出多维 NWP 输入与光伏功率的映射关系, 本文选取 MIE 相关系数较高的前三种气象因素作为预测模型最终的输入变量, 即辐照度、温度及湿度。

3.2 基于 MIE 的相似日样本筛选

定义历史日代表向量 $N=[N_F, N_T, N_H]$, 预测日代表向量 $M=[M_F, M_T, M_H]$ 。根据式(1)一式(8)计算历史日与预测日各气象数据的 MIE 相关系数为 $I_{NM}=[I_{NF, MF}, I_{NT, MT}, I_{NH, MH}]$, 用来反映历史日样本与预测日气象信息的相似程度。

将第 3.1 节中计算得到的 MIE 相关系数 $I_{XY}=[I_{FP}, I_{TP}, I_{HP}]$ 作为历史日与预测日各气象因素之间相似度的权重因子, 即 $w_{XY}=I_{XY}$; 由式(15)一式(17)可以计算得到历史日与预测日三种气象因素的加权 MIE 之和 I_{WNM} , 用来表征历史日与预测日气象信息的综合相似程度。

$$W_{total} = w_{FP} + w_{TP} + w_{HP} \quad (15)$$

$$W_F = \frac{w_{FP}}{W_{total}}, W_T = \frac{w_{TP}}{W_{total}}, W_H = \frac{w_{HP}}{W_{total}} \quad (16)$$

$$I_{WNM} = W_F \times I_{NF, MF} + W_T \times I_{NT, MT} + W_H \times I_{NH, MH} \quad (17)$$

式中: w_{FP} 、 w_{TP} 、 w_{HP} 分别为辐照度、温度和湿度相对于光伏功率的权重因子; W_{total} 为以上三种气象因素的权重因子之和; W_F 、 W_T 、 W_H 分别为辐照度、温度和湿度所占权重之和的百分比数, 反映了三种气象因素对光伏功率影响的加权系数。

本文采用对 MIE 加权求和的方式来表征历史日与预测日综合的相似程度, 是为了确保筛选出的相似日样本不仅与预测日的辐照度、温度和湿度的综合相似程度最高, 而且三种气象因素对光伏功率的影响程度也保持一致, 有效地避免了错选现象, 比如某日与预测日的湿度 MIE 相关系数 $I_{NH, MH}$ 较高而

辐照度 MIE 相关系数 $I_{NF, MF}$ 较低也被选为相似日的情况, 实际上湿度与光伏功率的 MIE 相关系数 I_{HP} 较低。

3.3 LSTM 预测模型的训练与预测

3.3.1 数据的归一化与反归一化

为了消除各气象数据的量纲不同而对光伏功率预测的影响, 首先要将原始 NWP 数据和光伏功率等按照式(1)进行归一化处理, 将数据压缩到[0,1]区间内。最后, 再将预测输出的功率值进行反归一化处理, 恢复其物理意义。

3.3.2 LSTM 神经网络模型参数设置

根据 3.1 节中输入-输出数据相关性分析可知, 输入层节点数设置为 3。隐层采用单层结构, 神经元个数越多, 训练误差的收敛速度越快, 非线性拟合能力越强, 但是由此带来的模型计算量和复杂程度也随之增大, 造成训练时间过长; 经过反复尝试, 确定隐层神经元个数为 200, 预测效果最佳。输出层仅包含光伏功率输出, 故其节点数为设置为 1。训练次数取为 250 次, 初始学习率 $\eta=0.005$, 每训练 125 次学习率的衰减因子 $\sigma=0.2$ 。

3.3.3 预测误差分析

统一有效的预测精度评价指标有利于不同结果之间的比较, 本文选取常用的两种预测精度指标, 具体定义如下。

1) 均方根误差

$$RMSE\% = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{P_i - \hat{P}_i}{P_{cap}} \right)^2} \times 100\% \quad (18)$$

2) 平均绝对百分比误差

$$MAPE\% = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - \hat{P}_i}{P_i} \right| \times 100\% \quad (19)$$

式中: n 为日光伏功率的采样点数; P_i 和 \hat{P}_i 分别代表 i 时刻光伏功率实际值和预测功率值; P_{cap} 为研究场站的装机总容量。

4 算例分析

为了验证本文提出的预测方法的正确性和有效性, 以某实际光伏电站为研究对象, 装机容量为 20 MW。由于夜晚辐照度几乎为零, 所以本文只考虑每天 7:00—20:00 时段的数据, 采样分辨率为 15 min。选取 8 月的某一日作为预测日, 选取同年 6 月—8 月为相似日样本筛选区间。在相似日样本的筛选过程中, 分别采用 K-means 聚类、欧氏加权法和 MIE 三种方法, 得到与预测日气象信息高度相似

的样本，送入神经网络模型进行训练。预测模型分别采用 BP 神经网络、ELMAN 神经网络和 LSTM 神经网络三种方法，根据预测步骤逐步计算得到该气象条件下的预测结果。

4.1 晴天数据预测

以 8 月的某一晴天数据为预测日样本，采用上述方法进行预测。图 2 为三种相似日筛选方法+BP 法预测的对比结果，两种误差指标的对比如表 2 所示。图 3 为 MIE+三种预测方法的结果对比，两种误差指标的对比如表 3 所示。

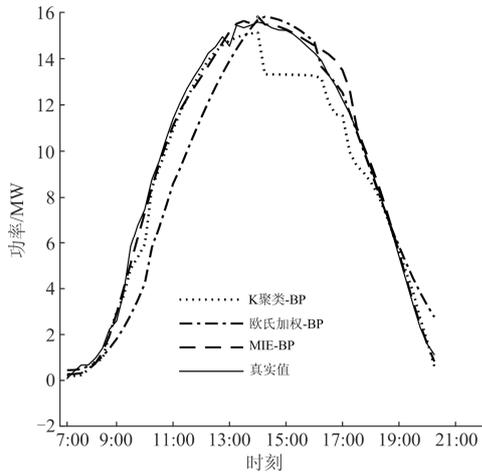


图 2 三种相似日筛选方法+BP 法预测的对比结果

Fig. 2 Comparison of three similar days screening methods and BP method for prediction

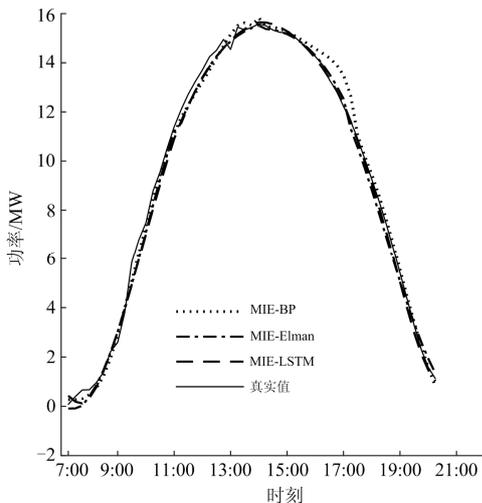


图 3 MIE+三种神经网络法的对比

Fig. 3 Comparison of MIE and three neural network methods

由图 2 可以看出，晴天天气条件下，实际日光伏功率波动较小。由表 2 可知，采用 MIE、K-means 聚类法和欧氏加权法的误差指标 $RMSE\%$ 分别为

2.042 2、4.551 2 和 7.449 3，它们对应的误差指标 $MAPE\%$ 分别为 13.981 3、15.481 5 和 27.981 3。这就说明，采用 MIE 法进行相似日筛选，得到的样本最能接近预测日气象条件及光伏出力，从而获得更高的预测精度。

表 2 三种相似日筛选方法+BP 法预测误差统计表

Table 2 Errors of three similar days screening and BP prediction

误差指标	K 聚类-BP	欧氏加权-BP	MIE-BP
$RMSE\%$	4.551 2	7.449 3	2.042 2
$MAPE\%$	15.481 5	27.656 3	13.981 3

表 3 MIE+三种神经网络法的误差统计表

Table 3 Error statistical of MIE and three neural network methods

误差指标	MIE-BP	MIE-ELMAN	MIE-LSTM
$RMSE\%$	2.042 2	1.731 5	1.681 6
$MAPE\%$	13.981 3	12.937 4	10.893 9

由表 3 可知，利用 MIE 法进行相似日筛选，再采用 LSTM 法、ELMAN 法和 BP 法进行预测得到的误差指标 $RMSE\%$ 分别为 1.681 6、1.731 5 和 2.042 2，它们对应的误差指标 $MAPE\%$ 分别为 10.893 9、12.937 4 和 13.981 3。这也说明，MIE-LSTM 法与其他组合方法相比，在晴天这种天气情况下的光伏功率的预测精度最高。

4.2 阴雨天数据预测

为了验证本文提出的 MIE-LSTM 法在各种气象条件下均能获得较高的预测精度，选 8 月的某一阴雨天作为预测日样本进行预测。图 4 为三种组合预测方法下的预测结果，两种误差指标的对比如表 4 所示。

由图 4 可以看出，阴雨天天气条件下，实际日光伏功率相比晴天波动较大。由表 4 可知，在阴雨天

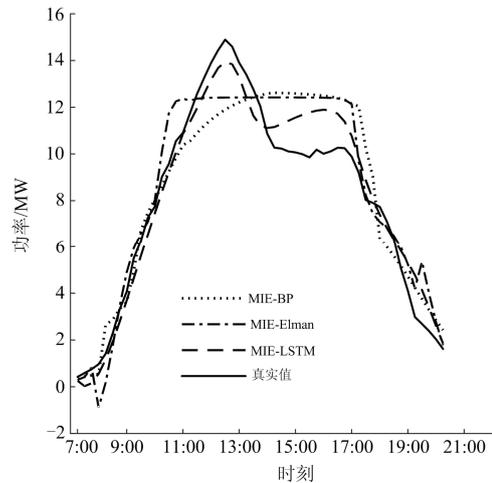


图 4 MIE+三种神经网络法

Fig. 4 MIE and three neural network methods

表 4 阴雨天气类型预测误差统计表
Table 4 Statistical results of forecast errors for rainy weather types

误差指标	MIE-BP	MIE-ELMAN	MIE-LSTM
$RMSE\%$	8.137 9	7.485 0	4.708 2
$MAPE\%$	19.940 5	22.353 0	16.523 1

气条件下, 利用 MIE 法进行相似日筛选, 再采用 LSTM 法、ELMAN 法和 BP 法进行预测得到的误差指标 $RMSE\%$ 分别为 4.708 2、7.485 0 和 8.137 9; 它们对应的误差指标 $MAPE\%$ 分别为 16.523 1、22.353 0 和 19.940 5。这也验证了 MIE-LSTM 法与其他组合方法相比, 得到的预测精度最高。

4.3 复杂强波动天气数据预测

一个预测方法的好坏取决于能否应对恶劣天气情况下的光伏功率预测。为了验证本文提出的 MIE-LSTM 法能较好地预测复杂强波动天气状况下的光伏功率, 选取 8 月某一强波动日作为预测日样本进行预测。图 5 为三种组合预测方法下的预测结果, 两种误差指标的对比如表 5 所示。

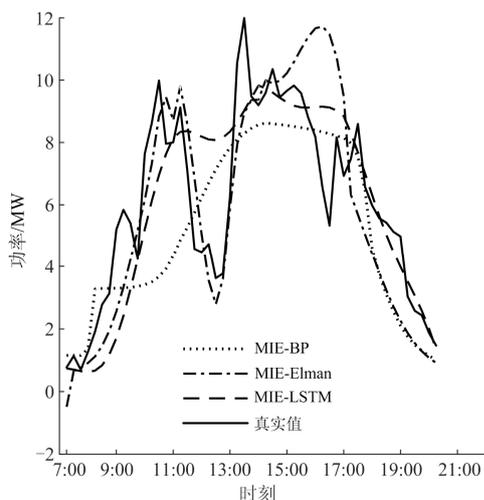


图 5 MIE+三种神经网络法

Fig. 5 MIE and three neural network methods

表 5 复杂强波动天气类型预测误差统计表
Table 5 Statistical results of forecast errors for complex and strong fluctuation weather types

误差指标	MIE-BP	MIE-ELMAN	MIE-LSTM
$RMSE\%$	11.191 5	9.612 6	4.439 1
$MAPE\%$	32.341 4	30.894 1	27.585 4

由图 5 可以看出, 复杂强波动天气下日光功率波动比阴雨天还要大。由表 5 可知, 在复杂强波动天气条件下, 利用 MIE 法进行相似日筛选, 再采用 LSTM 法、ELMAN 法和 BP 法进行预测得到的

误差指标 $RMSE\%$ 分别为 9.439 1、9.612 6 和 11.191 5, 它们对应的误差指标 $MAPE\%$ 分别为 27.585 4、30.894 1 和 32.341 4。这也验证了 MIE-LSTM 法与其他组合方法相比, 得到的预测精度最高。

5 结论

本文以某实测光伏电站为研究对象, 考虑到光伏出力受多种气象因素的影响以及长序列数据预测中保留多元的有效信息并剔除冗余信息以提高预测精度, 提出了一种基于 MIE-LSTM 的短期光伏功率预测方法。通过对晴天、阴雨天及复杂强波动天气等算例进行分析, 结果表明:

1) 利用互信息熵 MIE 的相关性衡量指标, 判断出辐照度、温度和湿度与光伏出力之间的相关性最高, 实现对高维气象数据的降维处理, 从而降低了计算量。

2) MIE 法、K-means 聚类法和欧氏加权法等三种相似日筛选方法, 经 BP 神经网络预测后得到的误差指标 $RMSE\%$ 和 $MAPE\%$ 中, MIE 法的误差最小。

3) LSTM 神经网络、ELMAN 神经网络和 BP 神经网络分别对利用 MIE 法提取的相似日样本进行训练并预测, 无论在晴天、阴雨天及复杂强波动天气下, LSTM 神经网络模型的预测精度为最高。

参考文献

- [1] 唐艳梅, 闫华光, 朱彬若, 等. 基于自适应遗传算法的分布式综合能源系统优化设计[J]. 电网与清洁能源, 2018, 34(2): 30-35, 42.
TANG Yanmei, YAN Huaguang, ZHU Binruo, et al. Optimal design of distributed integrated energy system based on adaptive algorithm[J]. Power System and Clean Energy, 2018, 34(2): 30-35, 42.
- [2] 杨秋霞, 刘同心, 高辰, 等. 计及光伏电站功率预测的电力系统优化分析[J]. 电力系统保护与控制, 2018, 46(2): 117-123.
YANG Qiuxia, LIU Tongxin, GAO Chen, et al. Power system optimization analysis considering power prediction of PV power station[J]. Power System Protection and Control, 2018, 46(2): 117-123.
- [3] 管霖, 赵琦, 周保荣, 等. 基于多尺度聚类分析的光伏功率特性建模及预测应用[J]. 电力系统自动化, 2018, 42(15): 24-30, 232-236.
GUAN Lin, ZHAO Qi, ZHOU Baorong, et al. Multi-scale clustering analysis based modeling of photovoltaic power characteristics and its application in prediction[J].

- Automation of Electric Power Systems, 2018, 42(15): 24-30, 232-236.
- [4] 龚莺飞, 鲁宗相, 乔颖, 等. 光伏功率预测技术[J]. 电力系统自动化, 2016, 40(4): 140-151.
GONG Yingfei, LU Zongxiang, QIAO Ying, et al. An overview of photovoltaic energy system output forecasting technology[J]. Automation of Electric Power Systems, 2016, 40(4): 140-151.
- [5] 丁明, 王磊, 毕锐. 基于改进 BP 神经网络的光伏发电系统输出功率短期预测模型[J]. 电力系统保护与控制, 2012, 40(11): 93-99, 148.
DING Ming, WANG Lei, BI Rui. A short-term prediction model to forecast output power of photovoltaic system based on improved BP neural network[J]. Power System Protection and Control, 2012, 40(11): 93-99, 148.
- [6] CHEN H, ASSALA P D S, CAI Y, et al. Intelligent transient overvoltages location in distribution systems using wavelet packet decomposition and general regression neural networks[J]. IEEE Transactions on Industrial Informatics, 2016, 12(5): 1726-1735.
- [7] 黄予春, 曹成涛, 顾海. 基于 IKFCM 与多模态 SSO 优化 SVR 的光伏发电短期预测[J]. 电力系统保护与控制, 2018, 46(24): 96-103.
HUANG Yuchun, CAO Chengtao, GU Hai. Short-term photovoltaic power generation forecasting scheme based on IKFCM and multi-mode social spider optimization SVR[J]. Power System Protection and Control, 2018, 46(24): 96-103.
- [8] 张雨金, 杨凌帆, 葛双冶, 等. 基于 Kmeans-SVM 的短期光伏发电功率预测[J]. 电力系统保护与控制, 2018, 46(21): 118-124.
ZHANG Yujin, YANG Lingfan, GE Shuangye, et al. Short-term photovoltaic power forecasting based on Kmeans-SVM algorithm and support vector machine[J]. Power System Protection and Control, 2018, 46(21): 118-124.
- [9] LORENZ E, HURKA J, HEINEMANN D, et al. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2009, 2(1): 2-10.
- [10] 张倩, 马愿, 李国丽, 等. 频域分解和深度学习算法在短期负荷及光伏功率预测中的应用[J]. 中国电机工程学报, 2019, 39(8): 2221-2230, 5.
ZHANG Qian, MA Yuan, LI Guoli, et al. Applications of frequency domain decomposition and deep learning algorithms in short-term load and photovoltaic power forecasting[J]. Proceedings of the CSEE, 2019, 39(8): 2221-2230, 5.
- [11] 曹煜祺, 张立梅, 白牧可. 基于双层 Elman 神经网络的光伏发电功率预测[J]. 供用电, 2017, 34(10): 8-13.
CAO Yuqi, ZHANG Limei, BAI Muke. Photovoltaic power prediction based on double Elman neural network[J]. Distribution & Utilization, 2017, 34(10): 8-13.
- [12] 史佳琪, 张建华. 基于深度学习的超短期光伏精细化预测模型研究[J]. 电力建设, 2017, 38(6): 28-35.
SHI Jiaqi, ZHANG Jianhua. Ultra short-term photovoltaic refined forecasting model based on deep learning[J]. Electric Power Construction, 2017, 38(6): 28-35.
- [13] TENG Xianliang, GAO Zonghe, ZHANG Yingyuan, et al. Key technologies and the implementation of wind, PV and storage co-generation monitoring system[J]. Journal of Modern Power Systems and Clean Energy, 2014, 2(2): 104-113.
- [14] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [15] 彭文, 王金睿, 尹山青. 电力市场中基于 Attention-LSTM 的短期负荷预测模型[J]. 电网技术, 2019, 43(5): 1745-1751.
PENG Wen, WANG Jinrui, YIN Shanqing. Short-term load forecasting model based on attention-LSTM in electricity market[J]. Power System Technology, 2019, 43(5): 1745-1751.
- [16] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(8): 131-137.
LU Jixiang, ZHANG Qipei, YANG Zhihong, et al. Short-term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Automation of Electric Power Systems, 2019, 43(8): 131-137.
- [17] 李鹏, 何帅, 韩鹏飞, 等. 基于长短期记忆的实时电价条件下智能电网短期负荷预测[J]. 电网技术, 2018, 42(12): 4045-4052.
LI Peng, HE Shuai, HAN Pengfei, et al. Short-term load forecasting of smart grid based on long-short-term memory recurrent neural networks in condition of real-time electricity price[J]. Power System Technology, 2018, 42(12): 4045-4052.
- [18] 杨锡运, 刘欢, 张彬, 等. 组合权重相似日选取方法及光伏输出功率预测[J]. 电力自动化设备, 2014, 34(9): 118-122.
YANG Xiyun, LIU Huan, ZHANG Bin, et al. Similar day selection based on combined weight and photovoltaic power output forecasting[J]. Electric Power Automation Equipment, 2014, 34(9): 118-122.
- [19] 郭辉, 杨国清, 姚李孝, 等. 基于综合相似日和功率相

- 关性的光伏电站预测功率修正[J]. 电网与清洁能源, 2018, 34(9): 52-58.
- GUO Hui, YANG Guoqing, YAO Lixiao, et al. Correction of predictive power of pv plants based on integrated similar days and power correlations[J]. Power System and Clean Energy, 2018, 34(9): 52-58.
- [20] 叶林, 陈政, 赵永宁, 等. 基于遗传算法—模糊径向基神经网络的光伏发电功率预测模型[J]. 电力系统自动化, 2015, 39(16): 16-22.
- YE Lin, CHEN Zheng, ZHAO Yongning, et al. Photovoltaic power forecasting model based on genetic algorithm and fuzzy radial basis function neural network[J]. Automation of Electric Power Systems, 2015, 39(16): 16-22.
- [21] 王丽婕, 冬雷, 高爽. 基于多位置 NWP 与主成分分析的风电功率短期预测[J]. 电工技术学报, 2015, 30(5): 79-84.
- WANG Lijie, DONG Lei, GAO Shuang. Wind power short-term prediction based on principal component analysis of NWP of multiple locations[J]. Transactions of China Electrotechnical Society, 2015, 30(5): 79-84.
- [22] 冬雷, 周晓, 郝颖, 等. 基于样本双重筛选的光伏发电功率预测[J]. 太阳能学报, 2018, 39(4): 1018-1025.
- DONG Lei, ZHOU Xiao, HAO Ying, et al. Power prediction model of PV power generation based on double screening of samples[J]. Acta Energetica Sinica, 2018, 39(4): 1018-1025.
- [23] 王勇, 黄雄华, 蔡国永. 信息论与编码[M]. 北京: 清华大学出版社, 2013: 28-30.
- [24] 杨正理, 史文, 陈海霞, 等. 大数据背景下采用互信息与随机森林算法的空气质量预测[J]. 环境工程, 2019, 37(3): 180-185.
- YANG Zhengli, SHI Wen, CHEN Haixia, et al. Air quality forecasting with mutual information and random forests based on big data[J]. Chinese Journal of Environmental Engineering, 2019, 37(3): 180-185.
- [25] 舒朝君, 王亚, 崔浩, 等. 基于互信息的配电网多谐波源识别方法[J]. 电力科学与技术学报, 2018, 33(1): 16-21.
- SHU Chaojun, WANG Ya, CUI Hao, et al. Multi-harmonic source identification method for distribution network based on mutual information[J]. Journal of Electric Power Science and Technology, 2018, 33(1): 16-21.
- [26] 丁晶, 王文圣, 赵永龙. 以互信息为基础的广义相关系数[J]. 四川大学学报: 工程科学版, 2002(3): 1-5.
- DING Jing, WANG Wensheng, ZHAO Yonglong. General correlation coefficient between variables based on mutual information[J]. Journal of Sichuan University: Engineering Science Edition, 2002(3): 1-5.
- [27] 史坤鹏, 乔颖, 赵伟, 等. 计及历史数据熵关联信息挖掘的短期风电功率预测[J]. 电力系统自动化, 2017, 41(3): 13-18.
- SHI Kunpeng, QIAO Ying, ZHAO Wei, et al. Short-term wind power prediction based on entropy association information mining of historical data[J]. Automation of Electric Power Systems, 2017, 41(3): 13-18.
- [28] 齐华. 信息理论与编码[M]. 北京: 中国电力出版社, 2014: 9-12.

收稿日期: 2019-05-28; 修回日期: 2019-09-11

作者简介:

吉铎格(1996—), 男, 硕士研究生, 研究方向为新能源发电技术; E-mail: 857651369@qq.com

李慧(1976—), 女, 通信作者, 博士, 副教授, 研究方向为柔性直流输电系统稳定与控制, 新能源发电技术; E-mail: lhbxy@bistu.edu.cn

刘思嘉(1986—), 男, 博士, 讲师, 研究方向为特种电机非线性控制。E-mail: sjliu@bistu.edu.cn

(编辑 葛艳娜)