

DOI:10.12158/j.2096-3203.2020.02.008

基于 EMD-LSTM 的光伏发电预测模型

朱玥, 顾洁, 孟璐

(大数据工程技术研究中心(上海交通大学电子信息与电气工程学院), 上海 200240)

摘要:随着能源消费结构的改变,可再生能源发电的消纳比例逐渐上升。文中以光伏发电功率为研究对象,分析了不同天气状态下的发电功率曲线特性及不同气象因素与光伏发电出力的相关性,进而提出了一种经验模态分解-长短期记忆神经网络(EMD-LSTM)方法融合的光伏发电功率预测模型。首先对预处理后的光伏发电功率历史序列进行重构,并对重构后的出力序列进行 EMD 分解,针对分解得到的各子序列分别建立长短期记忆神经网络模型,最后将各子序列预测模型得到的结果叠加得到光伏发电功率预测值。采用国内某地区光伏发电的实际出力数据对模型进行了检验,与滑动平均自回归模型(ARIMA)、支持向量机模型(SVM)、LSTM 等预测模型相比,文中所提出的模型预测误差小,能有效提高光伏发电功率的预测精度。

关键词:光伏发电;出力预测;经验模态分解;长短期记忆神经网络;气象因素

中图分类号: TM744

文献标志码: A

文章编号: 2096-3203(2020)02-0051-08

0 引言

近年来,世界范围内可再生能源的消费占比不断上升,能源消费结构发生巨大改变。我国为了缓解能源资源约束和生态环境压力^[1],大力发展可再生能源发电,并对各地区可再生能源的消纳提出了新的要求^[2]。因此研究可再生能源出力预测对有效利用可再生能源资源,以及保障电网安全可靠运行都有积极意义^[3]。

太阳能和风能是我国可再生能源发电的2种主要能源形式,国内外光伏发电出力的预测方法大致可以分为两类:基于物理基础的方法和基于统计学的方法。前者参照光伏电池的发电原理,利用气象数据与光伏设备参数,建立物理模型直接计算得到光伏发电出力^[4]。物理方法^[5]虽然不需要基于大量光伏发电出力历史数据来训练预测模型,但是较依赖电站具体的地理信息和准确的气象数据,模型抗干扰能力差,鲁棒性不强。统计学方法^[6-9]在预测光伏发电出力时不需要考虑光伏系统参数和复杂的光电转换模型,其具有建模简单等优势。基于统计学的方法包括点预测法、概率预测^[10]等。

文献[11]比较了多种模型的预测能力,结果表明神经网络模型和混合模型具有更高的预测精度。国内对于统计学方法在光伏发电出力预测中的应用研究近年来取得了一定进展。文献[12]提出了以马尔科夫方法修正反向传播神经网络(back-prop-

agation neural network, BPNN)的预测方法,适用于具有较大波动性与随机性的光伏发电输出功率时间序列预测,该方法计算量小,可行性高。文献[13]利用相空间重构技术与支持向量机回归结合的混合模型对光伏发电出力进行了短期预测。文献[14]利用遗传算法-反向传播(genetic algorithm-back propagation, GA-BP)神经网络与模糊C均值聚类的方法找到历史相似日期的历史负荷数据建立光伏发电出力模型。文献[11-14]的研究表明,对光伏发电出力特征有效信息的深入发掘与混合模型的使用,有利于提升预测精度。

文中采用现代统计学方法研究了光伏发电出力的特征,对出力时间序列进行时域-频域分析以提高预测精度,基于实际预测效果较好的长短期记忆神经网络(long-short-term memory, LSTM)模型,引入经验模态分解(empirical mode decomposition, EMD),建立了一种混合光伏发电出力预测模型,提出了一种 EMD-LSTM 的光伏发电预测模型,通过实际算例对模型的预测精度进行了分析与验证。

1 光伏发电出力特性及其影响因素研究

光伏发电的输出功率可采用下式计算^[15]:

$$P_{PV} = \eta SI [1 - 0.005(t_0 + 25)] \quad (1)$$

式中: η 为光能与电能的转化效率; S 为光伏板的面积, m^2 ; I 为太阳辐照度, W/m^2 ; t_0 为光伏装置安装处的环境温度。

式(1)表明,光伏发电的出力与太阳辐照度成强线性关系,随着太阳辐照度的增强,光伏发电功率增大。太阳辐照度的强弱受到多种因素的影响,

收稿日期:2019-09-06;修回日期:2019-10-22

基金项目:国家重点基础研究发展计划(973计划)资助项目(2016YFB0900100);上海市科委重大项目(18DZ1100303)

包括太阳高度、日照时间、地理位置、大气各项指标等。其中,地理位置直接由光伏发电装置的安装位置决定;太阳高度、日照时间则由地球的自转与公转决定,具有很强的周期性;而大气各项指标包括大气质量、湿度等受到天气状况的影响,具有很强的随机性。太阳辐射度的周期性与随机性并存的特性使得光伏发电的出力曲线同时呈现出较强的周期性与随机性。

1.1 不同天气状态下的光伏发电功率曲线

光伏发电装置仅在每日的日出时间到日落时间之间产生光电效应,在北半球,日照时间较长的夏季发电量要多于其他季节。同时,天气的晴雨也会影响到日照强度,因此文中对不同天气状况下发电日出力曲线进行分类分析。国内某地区晴天、阴天、多云、雨雪天气状况典型的光伏发电日出力曲线如图 1 所示。

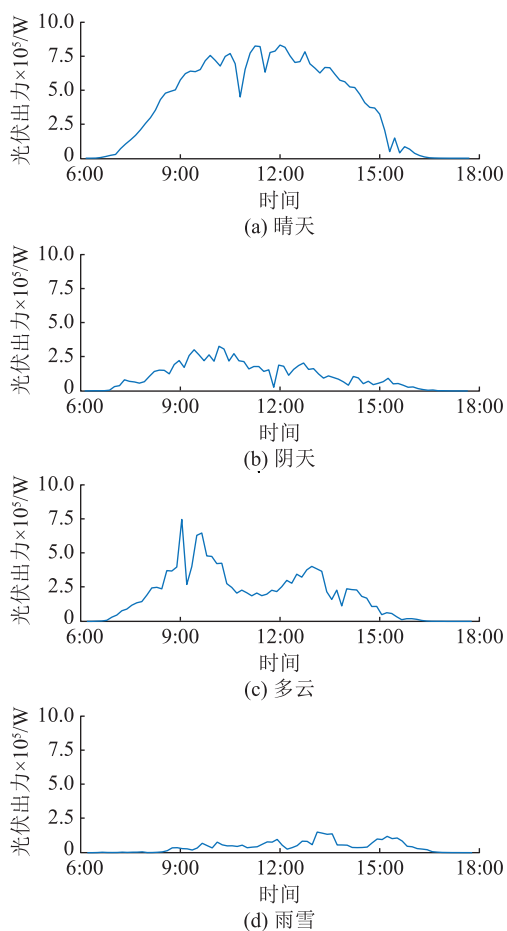


图 1 不同天气光伏发电日出力曲线
Fig.1 Daily output curves of PV under different weather conditions

由图 1 可知:

(1) 天气晴朗时,光伏发电的日出力曲线形态上类似正态分布的曲线,中午 12:00 左右光伏发电的瞬时功率值达到最大值;早晨与傍晚是光伏发电

出力瞬间功率变化率最大时间段,在日出之后 3 h 内瞬间功率值就上涨达到了日最大功率值的 75%,日落前 3 h 内瞬间功率值也同样快速下降至零值。

(2) 阴天时,光伏发电出力曲线的总体形态与晴朗天气类似,同样接近正态分布,但阴天的最大瞬时功率值仅有晴天的 40%。

(3) 多云天气时,光伏发电出力曲线呈现了明显不同的形态,由于云层移动的不确定性,导致正午不再是瞬时功率最大的时间段,同时瞬时功率的波动性与晴天相比也明显增大。

(4) 雨雪天时,与多云天气类似,光伏发电出力随着云层的移动而产生巨大变化,造成了瞬时功率曲线呈现“多峰多谷”的不规则形状。

1.2 气象因素相关性分析

据上文分析发现,光伏出力与气象因素间具有一定的相关性,上述天气状况因素的划分较为粗糙,即使是同一种天气状态下也会因为云层的变化导致光伏板受太阳辐射的面积与强度产生变化。因此,文中收集了与天气状况与云层移动密切相关的 12 种独立指标作为气象因素代表,与光伏出力进行相关性分析,具体见表 1。

表 1 所选取气象因素详细情况

Table 1 Details of selected meteorological factors

变量	对应指标	指标定义
V_1	液态水含量/($\text{kg}\cdot\text{m}^{-2}$)	云层中液态水含量垂直积分
V_2	固态水含量/($\text{kg}\cdot\text{m}^{-2}$)	云层中固态水含量垂直积分
V_3	表面压力/Pa	地表气压
V_4	相对湿度/%	空气中水汽压与饱和水汽压的百分比
V_5	总云量/%	各种高度和类型的云遮蔽天空的成数
V_6	10 m 纬向风速/($\text{m}\cdot\text{s}^{-1}$)	10 m 高度风速纬向分量
V_7	10 m 经向风速/($\text{m}\cdot\text{s}^{-1}$)	10 m 高度风速经向分量
V_8	2 m 气温/K	2 m 高度测得的温度
V_9	表面太阳辐射/($\text{J}\cdot\text{m}^{-1}$)	地球表面接受的太阳辐射总能量
V_{10}	表面热辐射/($\text{J}\cdot\text{m}^{-1}$)	地球表面接受的太阳热辐射能量
V_{11}	顶部净太阳辐射/($\text{J}\cdot\text{m}^{-1}$)	大气顶部净太阳辐射
V_{12}	总降水量/m	天空降落地面的水深度

光伏发电仅在太阳升起到太阳降落的时间段内工作,为了避免这种工作特性导致的光伏出力大量零值数据对分析造成影响,对不同时间点采集的数据分别进行相关性分析。对 PVOutput 数据集中国 Y 地区光伏电站 2018 年全年每日中午 12:00 时刻采集的光伏出力数据与各气象相关因素的散点图如图 2 所示。

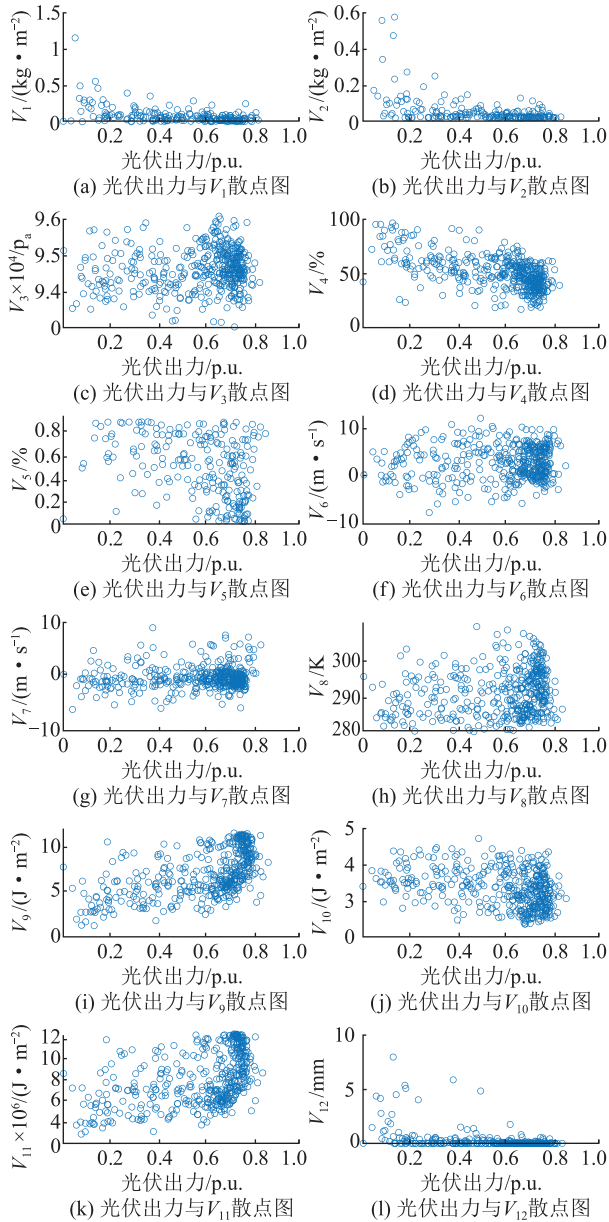


图2 12:00时刻Y地区各气象因素与光伏出力散点图

Fig.2 Scatter diagram of meteorological factors and photovoltaic output of Region Y at 12:00

由图2可知,在12个变量中, V_4, V_9, V_{10}, V_{11} 均体现出了线性相关性, V_3, V_5, V_6, V_7, V_8 体现出了一定的曲线相关性,而 V_1, V_2, V_{12} 由于代表降水情况,存在零值较多的问题,因此在散点图中难以体现对光伏发电出力的影响。 V_8 代表气温变量,在散点图中不同温度下光伏发电的出力分布较为均匀,两者间未表现出很强的线性相关关系。虽然从数学物理模型分析,气温是光伏发电功率的直接影响因素,两者间应该呈现线性关系,但由于温度与太阳辐照共同作用于光伏发电功率,因此对于全年数据的分析很难单独讨论气温对光伏发电功率的影响。

统计学中,检验线性相关性常用的指标包括

Pearson 相关系数^[16]、Spearman 相关系数与 Kendall 相关系数。Pearson 相关系数长于刻画线性相关关系,且含义直观,文中采用 Pearson 相关系数进行线性相关性分析。对变量 $V_3, V_4, V_5, V_6, V_7, V_9, V_{10}, V_{11}$ 与光伏出力(全年数据)进行了 Pearson 相关系数的计算,得到结果见表2。

表2 各变量与光伏出力的 Pearson 相关系数计算结果

Table 2 Pearson coefficient of each variable and PV output

变量	Pearson 相关系数	变量	Pearson 相关系数
V_1	-0.442 4	V_7	0.102 1
V_2	-0.464 2	V_8	0.192 7
V_3	0.144 6	V_9	0.559 1
V_4	-0.570 9	V_{10}	-0.331 9
V_5	-0.552 4	V_{11}	0.470 1
V_6	0.056 4	V_{12}	-0.417 7

对比上表结果,在所有影响因素中,变量 $V_1, V_2, V_4, V_5, V_{10}, V_{12}$ 与光伏发电出力显现出了较为明显的负相关性,其中 V_4 与出力的负相关性最强。 V_4 为相对湿度,相对湿度较大的情况往往是下雨天气,此时光伏发电受到影响出力变小,相关系数结果与物理解释一致。相比其他影响因素, V_9 和 V_{11} 与光伏发电出力间为较强的正相关性,这2个天气变量反映了太阳辐射度,当太阳辐射度大的时候光伏出力较大,这一结论与光伏发电的数学模型也是相吻合的。

综上所述,上述气象因素中与光伏发电出力呈现强正相关性的因素是表面太阳辐射和顶部太阳辐射,与光伏发电呈现强负相关性的因素是相对湿度和总云量,与光伏发电不相关的变量有表面压力以及10 m 经纬向风速。下文将基于此分析结论建立光伏发电功率预测模型。

2 基于 EMD-LSTM 的光伏发电功率预测模型

光伏发电功率的历史时间序列是一系列非平稳非线性的离散数据,难以从其历史数据本身进行规律挖掘,传统线性时序模型方法存在较大的局限性,需要深入分析相关因素对光伏发电功率的影响情况,构建适用于光伏发电功率预测的模型。根据前文的分析,影响到已安装光伏发电设备出力的因素主要为气象因素,具有强波动性与随机性。光伏发电出力除了受到气象因素影响以外,从长期来看,还受到装机容量的影响,因此光伏发电出力时间序列是非平稳非线性的信号,直接对光伏发电出

力曲线进行时间序列预测具有较大困难,因此先对光伏发电出力进行时频分解,对分解后的子序列分别进行时间序列分析与预测,再将子序列预测结果叠加得到最终的预测值。

经过分解后的子序列仍具有一定的非平稳、非线性的特性,为了提高预测的精度,文中结合深度学习的模型与算法研究进展,选用神经网络算法对子序列进行预测,最后通过信号重构得到光伏发电功率预测结果。

2.1 经验模态分解理论

EMD 是一种非常实用的信号时-频分析技术,这种方法于 1998 年被 Huang N.E. 提出^[17]。与传统的信号分解技术,如傅里叶分解及小波分解^[18]不同,EMD 在对信号进行分解时,能够体现信号自身的时域与频域特征,不需要预先选择基函数,是一种自适应的信号分解方法。另外,傅里叶分解技术对信号提取频谱时,需要信号的全部时域信息,对输入信号整体进行变换,难以体现信号频率随时间的变化。而小波分解理论以傅里叶分解理论为基础,是一种滑动窗口的傅里叶分解,因此要求在窗口时间内的信息在频域上是稳定的。运用传统的信号分解方法对非平稳时间序列及非线性信号进行分解时,将导致难以提取信号的全部时-频特性,会产生较大误差。

Huang N.E. 等人提出了用本征模态函数 (intrinsic mode function, IMF) 来表达非平稳非线性时间序列信号中的不同尺度的波动特性:在任意一个时间点,信号都可以由多个本征模态函数的叠加得到。EMD 的分解目的就是将信号分解为一系列 IMF 的组合。符合 EMD 分解要求的 IMF 具有以下 2 个特点:(1) 在信号的全部时长内,IMF 的极值点个数和零点个数必须相等或相差最多不超过 1;(2) 在任意时刻,IMF 的上下包络线都关于时间轴局部对称,即 IMF 函数局部最大值的包络线和局部最小值的包络线的平均值为 0。

因此 EMD 分解是基于 2 个假设条件的:(1) 待分解信号应该至少具有一个最大值和一个最小值;(2) 待分解信号的局部时-频特性由极点值的时间唯一确定。若待分解信号不满足第一个条件但是存在拐点,则可以通过一阶微分求得极值。对于满足假设条件的信号,EMD 按照图 3 中流程对信号序列进行分解。

2.2 LSTM 网络理论

LSTM 由循环神经网络 (recurrent neural network, RNN) 发展而来,目前被广泛应用于时间序列

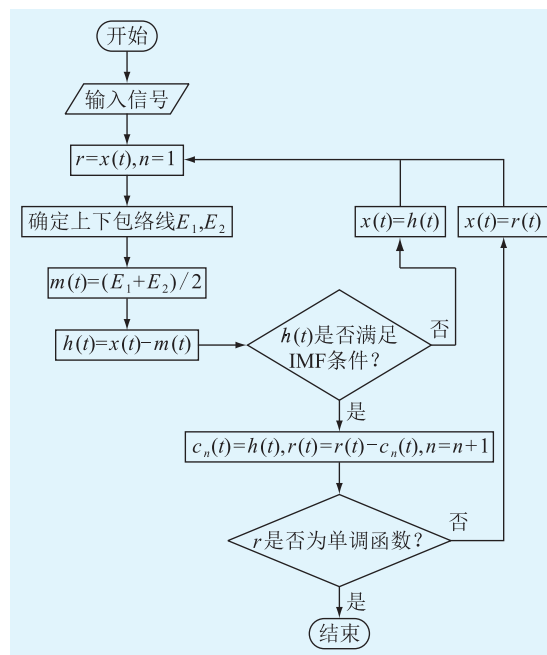


图 3 EMD 分解流程

Fig.3 Flow chart of EMD

预测。RNN 本身具有跨多时间步长传递信息的能力, Siegelman 等在文献[19]中证明了 RNN 具有较强记忆性且图灵完备。在实际应用中, RNN 模型能体现变量点序列之间的时间依赖关系, 使得其对于具有关联性的点序列输入具有较好的预测效果。在传统 RNN 模型的应用过程中, 处理长时间尺度依赖时, 由于求导的链式法则会发生梯度消失或梯度膨胀的问题。为了解决 RNN 模型的这个问题, Hochreiter 和 Schmidhuber 在文献[20]中提出了 LSTM 模型, 在神经网络模型中引入具有递归的隐藏层, 并在隐藏层中运用记忆单元取代传统人工神经元。这些记忆单元包含一个具有权重值为 1 的自连接 4 节点, 也就保证了梯度可以跨越多个时间步长传递, 避免梯度消失或梯度膨胀。

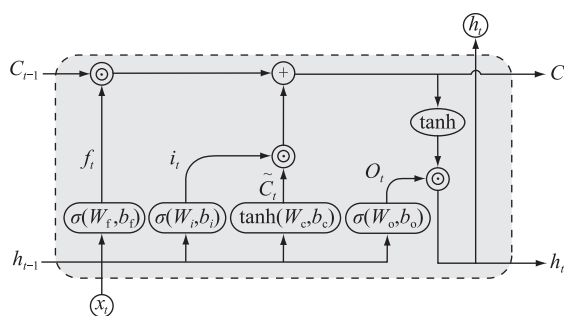


图 4 LSTM 单元结构示意图

Fig.4 Unit structure diagram of LSTM

LSTM 的结构示意如图 4 所示, 其中内部状态 C_t 是 LSTM 的核心, 通过在单元间的传递, 达到将信息跨多步长传递的目的。LSTM 通过门结构来删除

或者增加信息,实现记忆单元的遗忘和记忆功能。一个 LSTM 单元有 3 个门结构,分别是遗忘门(forget gate)、输入门(input gate)及输出门(output gate)。

LSTM 模型在每个时间步长内进行一系列的计算,求取内部状态及其他单元内状态量的值,具体计算如下:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (5)$$

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

$$h_t = o_t \tanh(C_t) \quad (7)$$

式中: σ 为激活函数,一般是 sigmoid 函数; W_f, b_f 分别为遗忘门连接权重及偏移值; W_i, b_i 分别为输入门连接权重及偏移值; W_o, b_o 分别为输出门连接权重及偏移值; \tilde{C}_t 为临时状态。

LSTM 在电力能源预测领域具有以下优势:(1) 能够有效刻画光伏出力时间序列的复杂时间关联性;(2) 能够对不同时间跨度的信息进行传递;(3) 解决了 RNN 模型梯度消失或膨胀的问题。

2.3 基于 EMD-LSTM 的光伏发电功率预测建模流程

结合光伏发电出力的特点以及 EMD 分解和 LSTM 网络的优势,选用 LSTM 模型对分解后的子序列进行预测。基于 EMD-LSTM 的光伏出力预测模型的流程如图 5 所示。

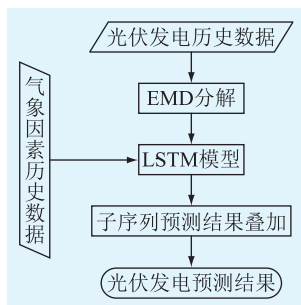


图 5 基于 EMD-LSTM 的光伏出力预测模型的流程

Fig.5 Flow chart of PV output forecasting based on EMD-LSTM

具体步骤:

(1) 首先对影响因素变量进行筛选,并对影响因素历史数据进行标准化处理,以免造成因为量纲不同和数量级不同造成的影响,使得不同度量之间的特征具有可比性,标准化的公式如下:

$$\tilde{x}_i = \frac{x_i - \mu}{\sigma} \quad (8)$$

式中: μ 为样本的均值; σ 为样本的标准差。

(2) 对光伏发电功率历史序列进行 EMD 分解,根据历史数据的具体情况决定分解的 IMF 子序列个数;分析 IMF 子序列特性,去除高频但数值较小的 IMF 子序列,保留其余 IMF 子序列与余量序列;

(3) 利用 LSTM 模型对各个子序列分别进行预测,输入变量为各子序列和对应时间点的气象相关因素数据,输出为各子序列对应的预测结果;

(4) 通过对各子序列预测值的叠加,得到光伏发电功率预测的最终结果。

3 算例分析

文中对 PVOutput 数据集中国 Y 地区光伏发电功率进行预测,所使用的光伏发电历史数据集为 2017-04-01 至 2018-05-31 共 424 d 每日 24 h 出力数据;气象数据来自欧洲中期天气预报中心(European Centre for Medium-range Weather forecasts, ECMWF),以 2017-04-01 至 2018-03-31 的数据进行模型参数训练,采用 2018 年 4 月 1 日至 2018 年 5 月 31 日的数据进行模型检验。

为验证模型的预测精度,采用归一化平均百分误差(normalized average percentage error, MAPE)、均方根误差(root-mean-square error, RMSE)及归一化平均绝对误差(normalized mean absolute error, NMAE)等指标作为评价依据,其具体计算公式如下:

$$\gamma_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{L}_i - L_i}{L_i} \right| \times 100\% \quad (9)$$

$$\gamma_{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (L_i - \hat{L}_i)^2} \quad (10)$$

$$\gamma_{NMAE} = \frac{1}{N} \frac{\sum_{i=1}^N |\hat{L}_i - L_i|}{\max_{1 \leq i \leq N} (L_i) - \min_{1 \leq i \leq N} (L_i)} \quad (11)$$

式中: L_i, \hat{L}_i 分别为测试集的第 i 个数据点出力实际值以及负荷的预测值; N 为测试集样本总数。

3.1 模型参数估计

统计分析表明,光伏发电只在日出与日落时间之间才有功率输出,因此仅对每日 7:00—19:00 之间时间节点进行光伏发电功率预测,去除出力近似为零的数值后,调整后的光伏发电出力序列如图 6 所示。

对光伏出力序列进行 EMD 分解,分解结果见图 7,经过 EMD 分解后一共得到 9 个 IMF 以及 1 个余量函数 res.。从分解结果来看,imf1 和 imf2 是光伏

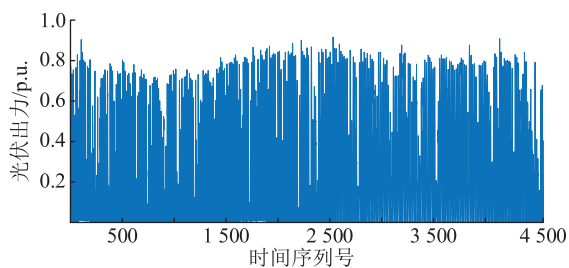


图6 Y地区去除零值后的光伏发电出力序列
(2017-04-01—2018-03-31)

Fig.6 The historical sequence of photovoltaic power generation after removing the zero value in Region Y (2017-04-01—2018-03-31)

发电出力的高频分量,其值很小,代表了光伏发电的随机扰动,对其进行预测没有意义,文中仅针对其余7个IMF子序列以及余量函数序列进行预测。

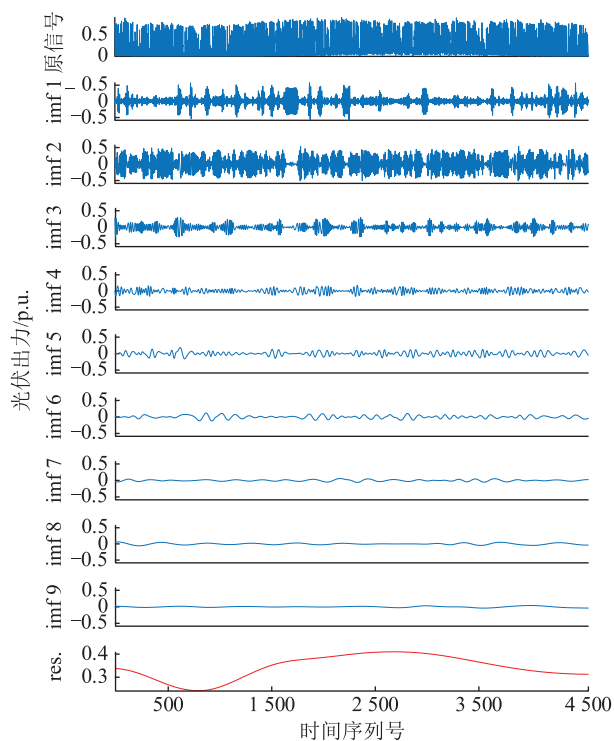


图7 Y地区光伏出力数据 EMD 分解结果
(2017-04-01—2018-03-31)

Fig.7 Result of EMD of PV output in Region Y
(2017-04-01—2018-03-31)

对子序列 imf_i 进行预测时,根据相关变量的维度以及历史数据的输入长度,设置 LSTM 模型的相关参数如下:输入层神经元节点数为 13,输出层的神经元节点数为 1。隐藏层的神经元节点数目通过多次试验根据性能最优原则选择得到。每个子序列的具体参数设置见表 3。

3.2 预测结果分析

对各子序列预测结果进行叠加后得到光伏发电出力预测的最终结果,如图 8 所示。

表 3 LSTM 预测模型参数设置

Table 3 Parameter setting in LSTM model

序列名称	输入层	输出层	隐藏层
imf 3	13	1	12
imf 4	13	1	25
imf 5	13	1	21
imf 6	13	1	18
imf 7	13	1	18
imf 8	13	1	9
imf 9	13	1	24
res.	13	1	16

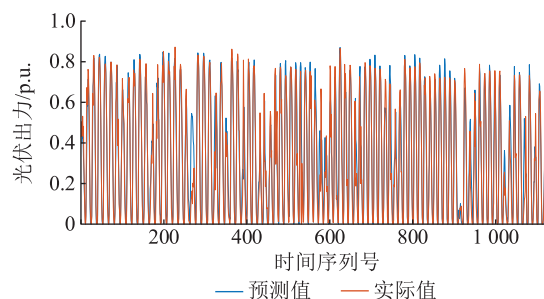


图8 Y地区光伏发电功率预测结果
(2018-04-01—2018-05-31)

Fig.8 Load forecasting result of PV generation in Region Y (2018-04-01—2018-05-31)

为了检验模型预测效果,将文中所提出的预测模型与实际预测系统中常用的自回归滑动平均模型 (autoregressive integrated moving average, ARIMA)、支持向量机模型 (support vector machine, SVM) 以及经典 LSTM 模型和 EMD-SVM 模型进行了对比。其中,ARIMA 模型和 SVM 模型预测分别采用文献[22]和文献[23]中建模步骤,LSTM 模型采用上文提到的方法,EMD-SVM 模型则对 EMD 分解后的子序列采用 SVM 模型进行预测。对 2018-04-03 的日光伏出力曲线的 5 种模型预测结果对比见图 9,具体误差分析如表 4 所示。

表 4 EMD-LSTM 模型与 SVM 模型预测性能比较

Table 4 Comparison of forecasting performance of EMD-LSTM model and SVM model

模型	$\gamma_{MAPE}/\%$	γ_{RMSE}/kW	γ_{NMAE}
ARIMA	32.11	79.41	39.58
SVM	21.68	56.88	24.06
LSTM	19.32	54.97	21.40
EMD+SVM	21.48	56.29	23.81
EMD+LSTM	18.57	54.06	20.61

图 9 的预测曲线对比以及表 4 的 3 种评价预测精度的指标对比显示,文中提出的 EMD-LSTM 模型的预测精度高于其他 4 种模型。该天为多云天气,

光伏发电功率受云层运动影响变化较大,因此 4 种模型都具有一定的误差,但是 EMD-LSTM 模型的预测结果更贴近实际发电功率的变化。从 2018 年 4 月 1 日至 2018 年 5 月 31 日整体预测情况来看,EMD-LSTM 模型的预测误差小,整体预测误差分别降低了 38.3%、14.3%、3.9%、13.6%。

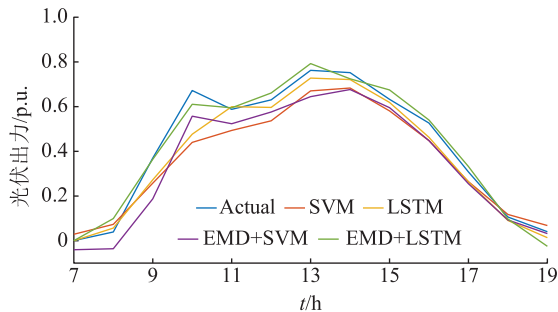


图 9 Y 地区光伏发电功率预测结果 (2018-04-03)

Fig.9 Load forecasting result of PV generation in Region Y (2018-04-03)

4 结论

文中分析了不同天气状态下的光伏发电出力曲线特性,对各气象因素与光伏发电出力序列进行了相关性分析,提出了一种 EMD-LSTM 的光伏发电功率预测模型,主要得到了以下结论:

(1) 气象因素中,与光伏发电正相关性最强的是表面太阳辐射与顶部太阳辐射,负相关性最强的是相对湿度与总云量;

(2) 文中所提出的基于 EMD-LSTM 的光伏发电功率模型的预测精度比 ARIMA 模型、SVM 模型、LSTM 模型、EMD-SVM 模型高。

文中所使用的数据采样频率为每小时一次,数据还不够细化,未来基于更高分辨率的数据,借助所提出的 EMD-LSTM 模型,分析光伏发电出力时间序列的数理特征,得到更精细化的光伏发电出力预测曲线。

参考文献:

- [1] 张孝德. “十三五”经济转型升级新思维:新能源革命引领战略[J]. 国家行政学院学报,2015(2):21-26.
ZHANG Xiaode. New thinking on economic transformation and upgrading during the 13th five-year plan period: new energy revolution leading strategy[J]. Journal of China National School of Administration,2015(2):21-26.
- [2] 国家发展改革委. 清洁能源消纳行动计划(2018—2020年)[J]. 上海建材,2019(2):11-14.
National Development and Reform Commission. The plan of clean energy consumption (2018—2020) [J]. Shanghai Building Materials,2019(2):11-14.
- [3] 于大洋. 可再生能源发电并网协调策略的研究[D]. 山东:

山东大学,2010.

YU Dayang. Studies on coordination strategies of the renewable generation integration [D]. Shandong: Shandong University,2010.

- [4] 赖昌伟. 基于奇异谱分析的短期光伏出力特性分析与预测[D]. 南宁:广西大学,2019.
LAI Changwei. Analysis and prediction of short-term photovoltaic output characteristics based on singular spectrum analysis [D]. Nanning:Guangxi University,2019.
- [5] CAI Tao, DUAN Shanxu, CHEN Changsong. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement [C]//Power Electronics for Distributed Generation Systems (PEDG),2010 2nd IEEE International Symposium on. IEEE,2010.
- [6] HAMMER A, HEINEMANN D, LORENZ E, et al. Short-term forecasting of solar radiation;a statistical approach using satellite data[J]. Solar Energy,67(1-3):139-150.
- [7] LI Ran, LI Guangmin. Photovoltaic power generation output forecasting based on support vector machine regression technique[J]. Electric Power,2008.
- [8] 倪春花,李弘毅,吴在军. 雾霾对光伏发电量的影响分析[J]. 江苏电机工程,2015,34(6):77-81.
NI Chunhua, LI Hongyi, WU Zaijun. Impact analysis of haze on photovoltaic systems power generation [J]. Jiangsu Electrical Engineering,2015,34(6):77-81.
- [9] 金楚,黎嘉明,徐沈智,等. 大规模光伏发电并网概率潮流计算及对电网的影响[J]. 电力工程技术,2017,36(1):1-8.
JIN Chu, LI Jiaming, XU Shenzi, et al. Probabilistic load flow calculation and influence analysis for power grid connected with large scale photovoltaic generation system [J]. Electric Power Engineering Technology,2017,36(1):1-8.
- [10] WAN Can, LIN Jin, SONG Yonghua, et al. Probabilistic forecasting of photovoltaic generation: an efficient statistical approach [J]. IEEE Transactions on Power Systems,2016, PP(99):2471-2472.
- [11] REIKARD G. Predicting solar radiation at high resolutions: a comparison of time series forecasts [J]. Solar Energy,2009,83(3):342-349.
- [12] 姜侨娜,陈中. BP-马尔科夫组合预测方法在光伏发电量预测中的应用[J]. 电力需求侧管理,2011(6):26-29.
JIANG Qiaona, CHEN Zhong. The study of solar power forecasting based on BP-Markov method [J]. Power Demand Side Management,2011(6):26-29.
- [13] 黄磊,舒杰,姜桂秀,等. 基于多维时间序列局部支持向量回归的微网光伏发电预测[J]. 电力系统自动化,2014,38(5):19-24.
HUANG Lei, SHU Jie, JIANG Guixiu, et al. Photovoltaic generation forecast based on multidimensional time-series and local support vector regression in microgrids [J]. Automation of Electric Power Systems,2014,38(5):19-24.
- [14] 耿博,高贞彦,白恒远,等. 结合相似日 GA-BP 神经网络的光伏发电预测[J]. 电力系统及其自动化,2017,29(6):

- 118-123.
GENG Bo, GAO Zhenyan, BAI Hengyuan, et al. PV generation forecasting combined with similar days and GA-BP neural network[J]. Automation of Electric Power Systems, 2017, 29(6): 118-123.
- [15] YUN L. An islanding detection method for photovoltaic power generation system using fluctuation characteristic of PCC harmonic voltage [J]. Advanced Materials Research, 2014, 998-999:574-577.
- [16] O'GRADY N P, ALEXANDER M, DELLINGER E P, et al. Guidelines for the prevention of intravascular catheter-related infections; centers for disease control and prevention [M]. Hospital infection control, 2011.
- [17] NAPOLITANO G, SERINALDI F, SEEL. Impact of EMD decomposition and random initialisation of weights in ANN hind-casting of daily stream flow series: an empirical examination [J]. Journal of Hydrology (Amsterdam), 2011, 406(3-4): 199-214.
- [18] CHINARRO D. System engineering applied to fuenmayor karst aquifer and collins glacier[M]. Springer International Publishing, 2014.
- [19] DU Nan, DAI Hanjun, TRIVEDI R, et al. Recurrent marked temporal point processes: embedding event history to vector [C]//The 22nd ACM SIGKDD International Conference. ACM, 2016.
- [20] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014.
- [21] YUC N, MIROWSKI P, HO T K. A sparse coding approach to household electricity demand forecasting in smart grids [J]. IEEE Transactions on Smart Grid, 2017, 8(2): 738-748.
- [22] 蒋增林, 叶江明, 陈昊. 基于时间序列分析的负荷预测方法的比较研究[J]. 南京工程学院学报(自然科学版), 2018, 16(2): 26-31.
JIANG Zenglin, YE Jiangming, CHEN Hao. A comparative study of load forecasting methods based on time series analysis [J]. Journal of Nanjing Institute of Technology (Natural Science Edition), 2018, 16(2): 26-31.
- [23] 崔杨, 李莉, 陈德荣. 基于最小二乘支持向量机的超短期风电负荷预测[J]. 电气自动化, 2014, 36(5): 35-37.
CUI Yang, LI Li, CHEN Derong. Ultra-short-term wind power load forecast based on least squares SVM [J]. Electrical Automation, 2014, 36(5): 35-37.

作者简介:



朱玥

朱玥(1994),女,硕士在读,研究方向为光伏发电出力预测(E-mail: stray-bird@sjtu.edu.cn);

顾洁(1971),女,博士,副教授,研究方向为电力系统规划、负荷预测、能源大数据、电力市场等;

孟璐(1997),女,硕士在读,研究方向为电力系统规划。

Photovoltaic power generation prediction model based on EMD-LSTM

ZHU Yue, GU Jie, MENG Lu

(Research Center for Big Data Engineering and Technologies(School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University), Shanghai 200240, China)

Abstract: With the change of energy consumption, the consumption proportion of renewable energy power generation rises gradually. Photovoltaic power is chosen as the research object to study the characteristics of power curve under different weather conditions and to analyze the correlation of the efforts of various meteorological factors and photovoltaic power generation. Then by combining Empirical Mode Decomposition (EMD) and Long-short Term Memory (LSTM), a load forecasting model suitable for photovoltaic power generation is proposed. Firstly, the historical sequence of preprocessed photovoltaic power generation is reconstructed. Then, the reconstructed historical time series is decomposed in EMD model, and the decomposed sub-sequences are respectively predicted in the LSTM network. Finally, the results of each sub-sequence are superimposed to obtain the predicted result of photovoltaic power generation. The actual output data of photovoltaic power generation in a certain region in China is used to test the model. Compared with the ARIMA model, SVM model, LSTM model and EMD-SVM model, the model proposed in this paper has lower prediction error and can effectively improve the prediction accuracy of photovoltaic power generation.

Keywords: photovoltaic power generation; output forecasting; empirical mode decomposition; long-short term neural network; meteorological factors

(编辑 钱悦)