

基于游程判别法和 VMD 残差修正的风电功率预测

王瑞¹, 冉锋¹, 逯静^{2†}

(1. 河南理工大学 电气工程与自动化学院, 河南 焦作 454000;

2. 河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

摘要:为了提高风电功率预测准确性,提出一种基于频率游程判别法和变分模态分解(VMD)残差修正的风电功率超短期预测模型.采用变分模态分解将原始风电功率序列分解,得到一系列不同中心频率的子序列,再利用序列之差提取残差序列,残差序列继承原始序列噪声分量与分解被屏蔽的真实分量,呈现波动性大,非线性复杂和不平稳的特点,采用 t-SSA-LSTM 模型并结合天气特征进行预测.利用频率游程判别法把子序列划分为低频分量类和高频分量类:低频分量呈现线性平稳的特点,采用自适应 t 分布麻雀搜索算法(t-SSA)优化自回归滑动平均模型(ARIMA)预测;高频分量具有波动性大且复杂的特点,采用 t-SSA 优化长短时记忆神经网络(LSTM)进行预测.将不同序列的预测结果线性叠加得到风电功率预测结果.将该模型应用于国内某风电发电厂的风电功率预测中,试验结果表明,该模型能有效提高预测精度.

关键词:风电功率;长短时记忆神经网络;自回归滑动平均模型;残差;麻雀搜索算法

中图分类号:TM715; **文献标志码:**A

Wind Power Prediction Based on Run Discriminant Method and VMD Residual Correction

WANG Rui¹, RAN Feng¹, LU Jing^{2†}

(1. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;

2. College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: To improve the accuracy of wind power prediction, an ultra-short-term combination forecasting method based on Frequency Run Length Discriminant and Variational Modal Decomposition (VMD) residual error correction is proposed. Firstly, the original wind power sequence is decomposed by VMD to obtain a series of sub-sequences with different center frequencies, and then the residual sequence is extracted from the difference in the sequences. The residual sequence has the characteristics of large fluctuation, nonlinear complexity, and unsteadiness, which inherits the original sequence noise component and the masked information during decomposition, and the adaptive t-distribution Sparrow Search Algorithm (t-SSA-LSTM) combined with the weather features is used for the prediction. The sub-sequences are divided into two kinds of signals class, namely high and low-frequency sequences, by using the Frequency Run Length Discriminant method. The low-frequency sequences are linear stability

* 收稿日期:2021-07-11

基金项目:河南省科技攻关项目(222102210120), Science and Technology Project of Henan Province (222102210120)

作者简介:王瑞(1977—),男,河北内丘人,河南理工大学副教授,工学博士

† 通信联系人, E-mail: lujing@hpu.com

and the adaptive t-distribution Sparrow Search Algorithm (t-SSA) is used to optimize the autoregressive integrated moving average (ARIMA) model prediction. The characteristics of high-frequency sequences are volatile and complex, and the t-SSA is used to optimize the Long Short-Term Memory (LSTM) neural network for the prediction of high-frequency sequences. Finally, the wind power prediction results are achieved by linearly superimposing the prediction results of different sequences. The proposed model is finally applied to a wind farm in China, and the results show that the model can effectively improve the prediction accuracy.

Key words: wind power; Long Short-Term Memory (LSTM) Network; Autoregressive Integrated Moving Average Model (ARIMA); residual correction; sparrow search algorithm

风力发电作为低碳可再生能源发电方式之一,在电网的发电比例中逐年增加,但风力发电受风力的大小,温度的高低等环境因素的影响,其发电功率具有波动性强与不可控性,大量的风电功率并网容易打破电力系统平衡,危害电力系统的安全运行^[1].由此可见,精准的风电功率预测对电网的优化、调度和稳定经济运行具有重要的价值^[2].

根据时间尺度的不同,风电功率预测时长可划分为超短期、短期、中期和长期.其中,超短期预测的时间为未来4 h,主要用于发电机组的控制,电网可根据超短期预测结果实时调整发电计划及功率控制策略,满足电力系统的发用电平衡;短期预测的时间为从次日开始的未来72 h以内,主要用于电网的安排和调整调度计划,是电力系统制订日发电计划的重要依据;中期预测的时间为未来几周或几个月,主要用于制订风电场季度发电计划及安排大型检修;长期预测的时间跨度一般以年为单位,主要用于预测风电场的年发电量,对风电场选址及风电机组分布等进行评估.随着预测时长的增加,影响因素也难以预测,导致中长期的预测难度提高,近年来短期与超短期预测研究较多.风电功率预测方法有物理法和统计法.物理法一般综合分析风电机组所在的位置和历史功率曲线等因素,然后通过物理法寻找主要影响因素与发电功率之间的非线性关系,实现对风电未来发电功率的预测^[3-5].统计预测法根据采集的风速、风向和功率等的历史数据,采用人工智能方法建立历史风电功率数据与风速等相关因素之间的复杂统计数学信息模型,实现对未来风电功率的预测^[6].统计方法可分为单一预测方法和组合预测方法,其中单一预测方法存在一定的局限性,用于风电功率预测的精度较低^[7].由于风能的波动性、随机

性,导致风电功率数据波动频率高,具有强烈的随机性,并且容易受到噪声的干扰.将原始风电功率的信号进行分解,可以得到几组具有不同频率,振幅较低的分解信号,并且可以有效减小原始数据噪声的影响,对分解得到的信号使用参数优化的预测模型进行预测,再结合误差修正等方法构成的组合预测方法大大改善了风电预测性能.文献[8]提出用变分模态分解(Variational Modal Decomposition, VMD)、排列熵(Permutation Entropy, PE)重构与相关向量机(Relevance Vector Machine, RVM)对风电功率进行短期预测,采用混沌布谷鸟搜索算法提高风电功率预测精度.文献[9]采用一种互补集合经验模态分解法(Complementary Ensemble Empirical Mode Decomposition, CEEMD)、缎蓝园丁鸟优化算法(Satin Bower Birdoptimization Algorithm, SBO)与最小二乘支持向量回归(Least Squares Support Vector Machines, LSSVM)的组合预测模型.文献[10]建立VMD分解与LSTM相结合的光伏功率预测模型,并在模型的基础上添加了双重注意力机制,有效提高了预测精度.文献[11]采用游程检测重构法对CEEMD分解得到的子序列重构为3个具有不同特征的分量,考虑独有的特征选取不同的预测模型.文献[12]利用相空间重构对集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)分解好的子序列进行重构,采用改进引力搜索算法(Improved Gravitational Search Algorithm, IGSA)优化LSSVM,有效提高了模型的预测精度.但是,通过分解对原始序列的处理,分解得到的子序列之和与原始序列之间存在很大的残差,并且子序列重构在一定程度上减少了工作量,重构后的序列可能会出现模态混叠和频率混叠现象.

针对上述问题,本文提出了一种基于频率游程判别法和VMD残差修正的超短期风电组合预测模型.首先,将原始风电功率经VMD方法分解,得到一系列具有不同中心频率和零噪声的子序列,并提取残差序列.其次,采用频率游程判别法将子序列分为高频分量类和低频分量类,根据不同类别分量的变化特征,建立t-SSA-LSTM模型和t-SSA-ARIMA模型预测高频分量与低频分量.残差序列继承了原始序列中未被分解的噪声分量和分解过程中被屏蔽的小信号真实波动分量,波动性大且具有复杂的非线性关系,因此,将其与气象因素特征结合后,采用t-SSA-LSTM进行预测.最后,将不同模型预测出来的结果线性叠加,得到风电功率预测值.本文采用国内某风力发电厂的历史发电功率数据进行仿真,试验结果显示,本文模型能有效提高预测精度.

1 变分模态分解与残差提取

原始风电功率序列受到天气等环境因素的影响,数据呈现振动幅度大、波动频率高等特点,导致直接预测精度低,误差大.因此,需要对原始风电功率数据进行信号分解.由于经验模态分解(Empirical Mode Decomposition, EMD)方法易造成模态混叠、端点效应、分解不充分等问题,故选用鲁棒性强,具有自适应性,且不会出现模态混叠的VMD方法.变分模态分解(VMD)是由Dragomiretskiy等^[13]首次提出.VMD方法是一种准正交完全非递归的模态变分和信号处理方法,该技术的优点在于模态分解的个数 K 可以自行确定,通过迭代寻优的方式不断更新每个模态分量的最优中心频率,然后将原始序列有效分解为固有模态分量(Intrinsic Mode Functions, IMF).VMD方法的核心思想是将信号分解计算转到变分理论框架中,根据预设模态分解个数 K 实现对信号的分解,将原始信号 $X(t)$ 分解成 K 个中心频率为 ω_k 的本征模态函数 $U_k(t)$.VMD分解方法详见文献^[13].

为了使VMD分解充分且不会出现频率混叠,本文使用中心频率均值确定模态个数 K 的大小.原始风电功率数据含有大量的噪声,经过仿真试验分析可知,噪声容忍度 $\gamma = 0.3$ 时,会使得分解得到的序列之和与原始序列相等,但是子序列的噪声含量高,导致预测精度低.为了得到不受原始数据噪声影响的

子序列,保证分解得到子序列具有最好的保真度,应将噪声容忍度 γ 设置为0,但是容易造成分解得到的序列之和与原始发电功率序列之间存在很大的残差.利用序列之差提取残差序列,公式如下:

$$R(t) = X(t) - \sum_{k=1}^K U_k(t) \quad (1)$$

式中: $R(t)$ 为残差序列.

2 频率游程判别法

频率游程判别法是一种根据样本序列的游程数和游程长度来判别样本序列频率波动大小的方法^[14-15].游程是样本序列中连续出现的变量值的次数,是根据比较样本序列的绝对值与绝对平均值的大小,得到一组由两个不同的数构成的游程序列,计算游程数与游程长度,进而对样本序列进行频率波动的判断.

设定各子序列对应的时间序列为 $\{U(t)\}(t = 1, 2, 3, \dots, L)$; L 为时间序列长度;样本绝对平均值为 $U_{\text{avg}} = \frac{1}{L} \sum_{t=1}^L |U(t)|$,根据式(2),得到各个本征模态分量对应的游程序列 $\{Y(t)\}$:

$$Y(t) = \begin{cases} 1, & |U(t)| \geq U_{\text{avg}} \\ 0, & |U(t)| < U_{\text{avg}} \end{cases} \quad (2)$$

从式(2)可以看出,得到的游程序列 $\{Y(t)\}$ 是一组由“0”和“1”构成的序列,长度为 L ;通过式(3)判断准则,进一步判断VMD分解后各序列的波动程度:

$$\begin{cases} 0 < m < L \times 0.1 \\ L_{\text{max}} \geq L \times 0.1 \\ L_{\text{sum}} \geq L \times 0.3 \end{cases} \quad (3)$$

式中: m 为游程个数; L_{max} 为最大游程长度; L_{sum} 为游程长度 ≥ 90 的游程之和.

当游程序列 $\{Y(t)\}$ 满足式(3)中的所有条件,则对应的时间序列 $\{U(t)\}$ 为低频分量;否则时间序列 $\{U(t)\}$ 为高频分量.因此,应将VMD分解后的序列划分为高频分量类与低频分量类.

3 预测模型的搭建

风电功率经过VMD分解以后大大降低了序列内部的复杂程度,分为低频分量类与高频分量类以及残差序列.针对不同分量波动幅度和复杂程度存

在差异性,搭建 ARIMA 预测模型与 LSTM 预测模型,分别对不同分量进行预测.由于预测模型中的超参数合理选择有助于发挥模型的最佳性能,因此,选用 t-SSA 优化算法对模型超参数进行寻优,可在极大程度上发挥模型的最优性能,提升模型的预测精度.

3.1 ARIMA 原理

时间序列预测中,ARIMA 预测模型能够高效简单地捕捉数据内部的线性关系,对波动小、复杂程度低的时间序列有着良好的预测效果,能够完美适应低频分量类波动性小、线性的特征^[16]. ARIMA 由 3 部分组成:自回归模型(Autoregressive Model, AR),差分模型(Integrated Model, IM),移动平均模型(Moving Average Model, MA).模型的主要超参数也根据模型 3 个部分分为自回归项数 p 、滑动平均数 q 和差分次数 d .

模型对平稳的时间序列有更好的预测性能.首先,对序列进行平稳性检验和白噪声检验,只有平稳非白噪声序列才能使用 ARIMA 模型进行预测,对于不平稳的序列需要进行差分,直到平稳.其次,通过赤池信息准则(Akaike Information Criterion, AIC),确定 p 和 q 的最优取值,避免过拟合情况,以模型的误差服从独立正态分布为条件,公式如下:

$$\xi_{AIC} = 2k - 2\ln L \quad (4)$$

式中: ξ_{AIC} 表示不同 p 和 q 值下的赤池信息值,其值越小越好; k 为参数的数量; L 为似然函数.

3.2 LSTM 原理

与低频分量类相比,高频分量类中的序列复杂程度高,波动性强,而残差序列则更甚之.因此,选择有较强的鲁棒性和容错性,预测精度高的 LSTM 预测模型. LSTM 是在循环神经网络(Recurrent Neural Networks, RNN)的结构基础上引入了“门”的概念,解决了数据在预测与传播过程中梯度消失和梯度爆炸的问题,使得模型相对于 RNN 模型具有记忆功能,能够高效地进行时间序列预测^[17]. LSTM 单元结构如图 1 所示.

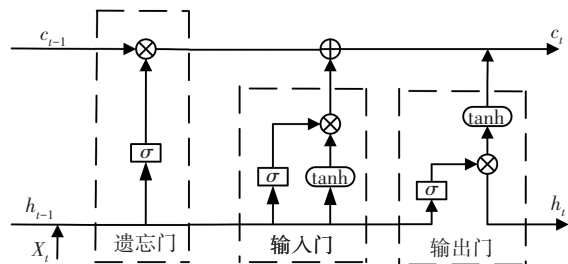


图 1 LSTM 单元结构

Fig.1 LSTM unit structure

其中,遗忘门的功能是决定上一时刻的单元状态 c_{t-1} 有多少保留到当前时刻 c_t ,公式如下:

$$f_t = \sigma(W_f \times [h_{t-1}, X_t] + b_f) \quad (5)$$

输入门的功能是决定当前时刻网络的输入 X_t 有多少保存到单元状态 c_t ,公式如下:

$$i_t = \sigma(W_i \times [h_{t-1}, X_t] + b_i) \quad (6)$$

$$\tilde{c}_t = \tanh(W_c \times [h_{t-1}, X_t] + b_c) \quad (7)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (8)$$

输出门的功能是控制单元状态 c_t 有多少输出到 LSTM 的当前的输出值 o_t ,以及最终的输出结果 h_t ,公式如下:

$$o_t = \sigma(W_o \times [h_{t-1}, X_t] + b_o) \quad (9)$$

$$h_t = o_t \times \tanh(c_t) \quad (10)$$

式中: h_{t-1} 为上一时刻的输出; x_t 为 t 时刻的输入; f_t 表示遗忘门的输出; σ 表示 sigmoid 函数; W 和 b 分别代表权重与偏置; h_t 为当前时刻的输出.

3.3 自适应 t 分布的麻雀搜索算法

麻雀搜索算法(SSA)是受麻雀觅食行为和反捕食行为启发而提出的一种新型群体智能优化算法.在试验中,需要模拟麻雀觅食过程的行为,获得优化问题的最优解.假设在 d 维搜索空间中,存在 N 只麻雀,第 i 只麻雀在 d 维空间中的位置为 $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$;那么,第 i 只麻雀的适应度值可以表示为如下形式:

$$F_i = f(x_{i,1}, x_{i,2}, \dots, x_{i,d}) \quad (11)$$

式中: $f(x)$ 表示适应度函数.在每一次迭代过程中,发现者位置更新公式如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j} \exp(-\frac{i}{\partial t_{\max}}), R_2 < S_T \\ X_{i,j} + QL, R_2 \geq S_T \end{cases} \quad (12)$$

式中: t 为当前迭代次数; t_{\max} 表示最大迭代次数; ∂ 表示 $(0, 1)$ 之间的随机数; Q 是服从正态分布的随机数; L 表示一个每个元素全部为 1 的矩阵. R_2 和 S_T 分别表示预警值和安全值.

剩下的麻雀作为加入者,位置更新公式如下:

$$X_{i,j}^{t+1} = \begin{cases} Q \exp(\frac{X_{i,\text{worst}}^t - X_{i,d}^t}{t^2}), i > \frac{n}{2} \\ X_{i,\text{best}}^{t+1} + \frac{1}{d} \sum_{d=1}^n (\text{rand}\{-1, 1\} \times |X_{i,d}^t - X_{i,\text{best}}^t|), i \leq \frac{n}{2} \end{cases} \quad (13)$$

式中: $X_{i,\text{worst}}^t$ 表示种群第 t 次迭代时麻雀最劣位置; $X_{i,\text{best}}^{t+1}$

表示种群第 $t+1$ 次迭代时麻雀的最佳位置.

侦察麻雀一般占种群的 10%~20%, 位置更新如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^{t+1} + \beta(X_{i,d}^t - X_{\text{best}}^{t+1}), & f_i > f_g \\ X_{i,d}^t + k \times \left(\frac{X_{i,d}^t - X_{\text{best}}^{t+1}}{|f_i - f_w| + e} \right), & f_i \leq f_g \end{cases} \quad (14)$$

式中: β 表示控制步长参数; $k \in [-1, 1]$ 表示麻雀移动的方向; e 是一个极小的常数; f_i 表示当前麻雀个体的适应度值; f_w 表示当前全局最劣适应度值.

自适应 t 分布的麻雀搜索算法是在原有麻雀搜索算法基础上进行改进, 利用自适应 t 分布对麻雀位置进行更新, 如式 (15) 所示.

$$X_i' = X_i + X_i \times t_{\text{iter}} \quad (15)$$

式中: X_i' 为变异后麻雀的位置; X_i 为第 i 只麻雀个体的位置; t_{iter} 是以算法的迭代次数为参数自由度的 t 分布. 该改进算法是麻雀位置更新后, 利用自适应 t 分布对麻雀位置更新, 麻雀的选择采用随机选择的方式, 对比更新前后麻雀, 如果更优则替代之前的麻雀. t -SSA 的优化算法流程如图 2 所示, 其具体步骤如下.

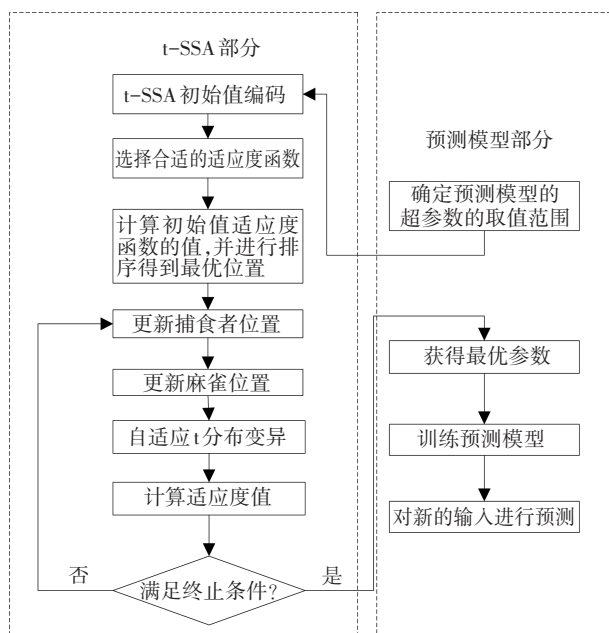


图 2 t-SSA 优化流程图

Fig.2 t-SSA optimization flow chart

Step1: 设置 ARIMA 和 LSTM 待优化的超参数的取值范围. ARIMA 需要优化的参数为 p 和 q ; LSTM 待优化的超参数为学习率、神经元个数和 batch-size.

Step2: 初始化 t -SSA 优化算法的相关参数.

Step3: 设置适应度函数. ARIMA 算法的适应度

函数为 AIC; LSTM 算法的适应度函数为均方根误差函数.

Step4: 随机生成初始麻雀的位置, 每一个位置代表一组参数. 计算每个位置的适应度函数的值, 并进行排序, 得到当前最优位置.

Step5: 更新捕食者位置. 更新麻雀位置.

Step6: 如果 $\text{rand} < p$, 则进行自适应 t 分布变异.

Step7: 求得最优位置. 若满足算法结束条件, 则最优的麻雀位置即为优化选择的参数值; 反之, 返回 Step5 继续迭代.

Step8: 将最后得到的最佳值作为 ARIMA 和 LSTM 的最优参数, 建立训练模型进行预测.

4 预测流程

风电功率预测流程如图 3 所示, 具体步骤如下:

Step1: 利用 VMD 将原始风电功率序列分解为 N 个具有不同中心频率的子序列, 再利用残差原理提取出残差序列.

Step2: 采用频率游程判别法, 把分解得到的子模态序列归类为高频分量类与低频分量类; 为消除不同输入特征的量纲不等, 将其归一化处理, 归一化公式为:

$$\tilde{x}(t) = \frac{x(t) - x_{\min}}{x_{\max} - x_{\min}} \quad (16)$$

式中: $\tilde{x}(t)$ 为某一变量归一化后的数值; x_{\max} 和 x_{\min} 分别为某一变量原始数据的最大值和最小值; $x(t)$ 为变量原始数据.

Step3: 采用 t -SSA 优化 LSTM 和 ARIMA 预测模型的超参数; 针对高频分量具有波动性强且复杂的特点, 选择 t -SSA-LSTM 组合算法进行滚动预测; 根据低频分量呈现出线性、平稳特征, 采用 t -SSA-ARIMA 组合算法滚动预测; 并将预测值反归一化处理.

Step4: 提取出来的残差序列波动性极强, 继承了原始序列中分解后残留的噪声和被屏蔽的信息分量, 其不平稳且复杂. 因此, 将其结合气象因素, 搭建 t -SSA-LSTM 预测模型进行预测, 并将预测值反归一化处理.

Step5: 将 Step3 和 Step4 的预测结果线性叠加, 得到风电功率的预测结果.

Step6: 用均方根误差 (Root Mean Square Error,

RMSE)、平均相对误差均值(Mean Absolute Percentage Error, MAPE)和平均绝对误差(Mean Absolute Error, MAE)指标进行误差分析. 计算公式分别为:

$$r_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i,pre} - y_{i,exp})^2} \quad (17)$$

$$r_{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_{i,pre} - y_{i,exp}|}{y_{i,exp}} \times 100\% \quad (18)$$

$$r_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{i,pre} - y_{i,exp}| \quad (19)$$

式中: n 为预测样本个数; $y_{i,pre}$ 和 $y_{i,exp}$ 分别为未来第 i 个预测样本的预测值和实际值.

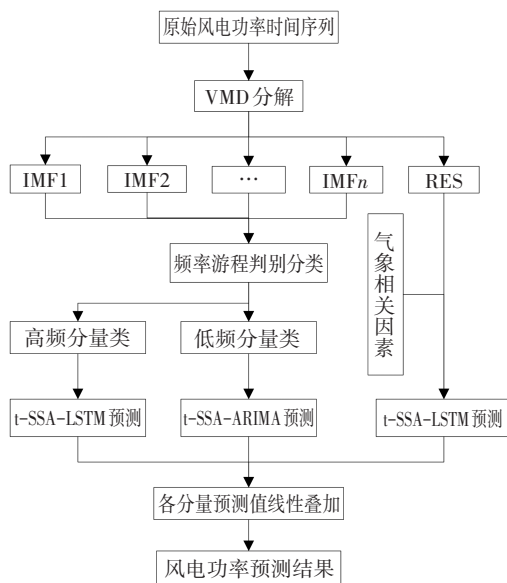


图 3 风电功率预测流程

Fig.3 Wind power prediction process

5 试验与结果分析

试验数据来源于国内某风电场 2020 年实测功率数据, 数据采集时间间隔为 15 min, 每 15 min 记为一个时间点, 每天采集数据 96 个, 即 96 个时间点. 采集的数据包括历史发电功率、温度、压强、湿度以及 10 m、30 m、50 m、70 m 高度的风速.

在仿真试验中, 样本数据的大小对预测精度至关重要. 过少的数据样本无法反映试验数据内部的规律与联系, 难以让预测模型得到充分训练, 导致预测欠拟合, 预测精度下降; 过多的数据样本会导致数据冗余, 训练与寻优困难, 时间过长, 造成过拟合现象和预测精度不达标. 因此, 需要合理地选择试验数据样本的大小, 控制训练时间, 保证预测精度. 不同样本个数的预测时长与误差如表 1 所示.

表 1 不同样本个数的预测时长与误差

Tab.1 Predicted duration and error for different sample days

样本个数	预测所用时间/s	预测误差
960	38	2.697 2
1 440	54	1.457 7
2 880	101	0.984 4
4 320	147	1.397 6
5 760	194	2.373 3

由表 1 可知, 当样本个数为 2 880 时, 不仅预测精度高, 而且训练与预测时长短. 因此, 选取风电场 5 月 1 日—30 日的数据作为模型的训练集, 5 月 31 日作为模型的测试集, 对模型进行验证. 原始风电功率时间序列如图 4 所示.

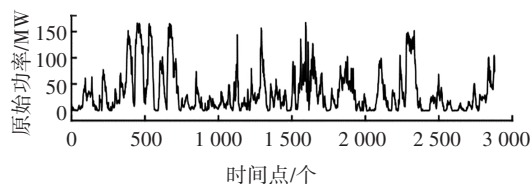


图 4 原始风电功率时间序列

Fig.4 Time series of original wind power

5.1 VMD分解

对原始风电发电功率时间序列进行 VMD 分解时, 需要确定模态的分解个数 K 的大小. 因为, VMD 算法相当于自适应维纳滤波器组, K 值太小, 容易导致分解不够充分; K 值过大, 相邻的两个模态之间的中心频率会相距较近, 易造成频率重叠. 本文通过中心频率的均值来确立模态的分解个数 K . VMD 的参数设置为: 惩罚参数 $\alpha = 2 000$, 初始中心频率 $\omega = 0$, 收敛判据 $r = 10^{-7}$, 噪声容忍度 $\tau = 0$ 以保证分解能更好地排除噪声的影响.

不同的 K 值对应的中心频率均值如表 2 所示. 从表 2 中可以看出, 当 $K=6$ 时, 第 2 个和第 3 个模态的中心频率均值分别为 0.033 7 kHz 和 0.092 5 kHz, 两个模态的中心频率相距较近, 可能出现了模态重叠, 所以模态个数选 5 比较适宜. 利用序列之差将残差序列提取. VMD 分解得到的子序列如图 5(a) 所示.

通过对比分析分解前后的数据可知, 风电功率在原始数据极值处与波动性大的时间点, 经 VMD 分解之后得到的数据更平滑, 证明了 VMD 分解能够降

低原始数据的复杂性和波动性.但是,分解后屏蔽了原始数据中的真实小波动分量,导致分解之后的数据产生数值差异,即残差序列,如图5(b)所示.

表2 不同K值对应的中心频率

Tab.2 Center frequency corresponding to different K

K值	中心频率均值/kHz					
2	0.005 4	0.103 2				
3	0.004 1	0.059 2				
4	0.003 0	0.039 5	0.126 6	0.314 0		
5	0.002 8	0.035 1	0.130 5	0.253 8	0.375 5	
6	0.002 7	0.033 7	0.092 5	0.173 1	0.288 3	0.426 9

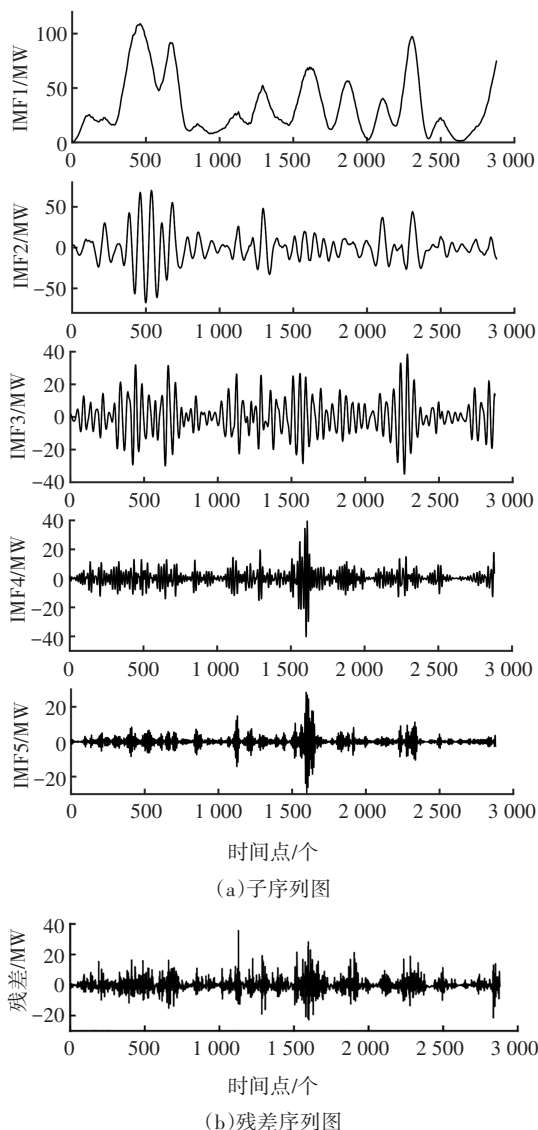


图5 VMD分解结果图

Fig.5 VMD decomposition results

5.2 频率游程判别法分类

不同的分量有着不同的特征,且同一预测模型对不同的特征分量预测效果存在差异,因此采用频率游程判别法,将5个模态分量根据游程的个数、最大游程长度和游程长度 ≥ 90 的游程数之和进行分类,各序列的总长度 $L=2\ 880$,根据第2节给出的判定条件,把IMF1与IMF2划分为低频分量,IMF3~IMF5划分为高频分量.不同分量的游程统计数如表3所示.

表3 不同分量的游程统计数

Tab.3 Runs statistics of different sequences

分量	游程个数 t	最大游程长度 L_{\max}	游程长度 ≥ 90 的游程数之和 L_{sum}
IMF1	14	507	2 770
IMF2	84	336	1 191
IMF3	184	216	480
IMF4	426	234	332
IMF5	610	259	465

5.3 预测结果分析

根据不同序列分量的特点,分别选用合适的方法进行预测.对于波动性小,整体呈现出线性趋势的低频分量类经过一阶差分,非平稳的低频分量序列转化为平稳低频分量,并采用ARIMA预测模型进行滚动预测;采用t-SSA对自回归项数 p 和滑动平均数 q 进行参数寻优,得到IMF1的最优参数为(4,8),IMF2的最优参数为(3,5).由于高频类的信号波动性大,具有复杂的非线性关系.因此,本文选用广泛使用的LSTM神经网络对高频分量类进行时序滚动预测.首先,对高频类的信号进行一阶差分,提升序列的平稳性;然后,对其归一化处理,减小序列的波动性,提高预测精度,采用t-SSA对模型的学习率、神经元个数和batch-size 3个超参数进行寻优,分别得到IMF3的寻优参数为0.009 34、63和1;IMF4的寻优参数为0.009 01、178和8;IMF5的寻优参数为0.008 66、83和1.残差序列继承了原始数据中的大量噪声与分解后被屏蔽的小波动真实分量,整体序列复杂,波动性大,具有强烈的非线性特点,并且难以预测.考虑原始风电功率数据的噪声来源于外界环境因素的变量,采用t-SSA-LSTM模型结合气象因素对残差序列进行滚动预测.LSTM模型超参数的寻优结果分别为0.007 3、135和5.

在对残差序列进行的滚动预测中,输入序列由历史残差序列 $\{R(1), R(2), \dots, R(p-1), R(p)\}$ 和 p 时刻的气象因素组成的序列 | 温度, 湿度, 大气压强, 10 m、30 m、50 m、70 m 高度的风速 | 组成, 预测下一时刻的残差值. 然后, 将预测值反馈到下一时刻的输入端, 再结合 $p+1$ 时刻的气象数据继续预测. 残差序列滚动预测图如图 6 所示.

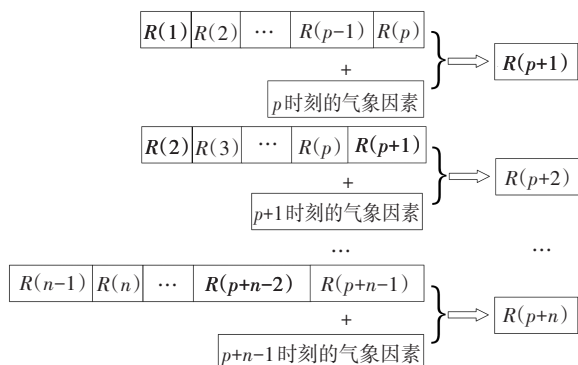


图 6 残差序列滚动预测图

Fig.6 Rolling residual forecast

为了综合评估本文所提出模型的性能, 选用 ARIMA、LSTM、t-SSA-LSTM、VMD-t-SSA-LSTM、不考虑残差序列共 5 种预测模型与本文所提出的“分解-分类-预测-残差修正”模型分别对 5 月 31 日的数据进行预测. 不同模型预测误差如表 4 所示, 不同模型的预测值和实际值对比如图 7 所示.

表 4 不同模型预测误差

Tab.4 Different model prediction errors

预测模型	评价指标		
	r_{RMSE}	r_{MAE}	r_{MAPE}
ARIMA	6.509 1	4.912 5	0.070 8
LSTM	3.471 6	2.982 7	0.052 1
t-SSA-LSTM	3.162 2	2.559 5	0.036 7
VMD-t-SSA-LSTM	2.593 8	2.024 4	0.029 5
不考虑残差序列	1.920 6	1.529 9	0.022 5
考虑残差序列	1.250 6	0.944 6	0.013 9

从图 7 中可以发现, 在功率值相对平稳的时间点附近, ARIMA 的预测效果优于 LSTM 的预测效果; 在波动性大的时间点, ARIMA 的预测性能明显低于 LSTM 的预测性能, 充分验证了 ARIMA 对于平滑稳定的时间点有不错的预测效果; 而 LSTM 能够对波动性大的时间点有不错的抑制效果. 经过对比可以发现, 相比较于经验选择超参数的 LSTM 模型, 经过 t-SSA 优化后的 LSTM 预测模型能够发挥最佳的预测性能, r_{RMSE} 、 r_{MAE} 和 r_{MAPE} 分别下降了 0.309 4、0.369 2、0.015 4, 有效地提升了预测精度.

原始功率的波动性和复杂程度, 导致单一的预测模型在波峰波谷处难以拟合实际功率数据, 采用 VMD 技术对原始功率数据进行分解, 可以有效地降低原始数据的复杂程度. 从图 7 中不难发现, VMD-

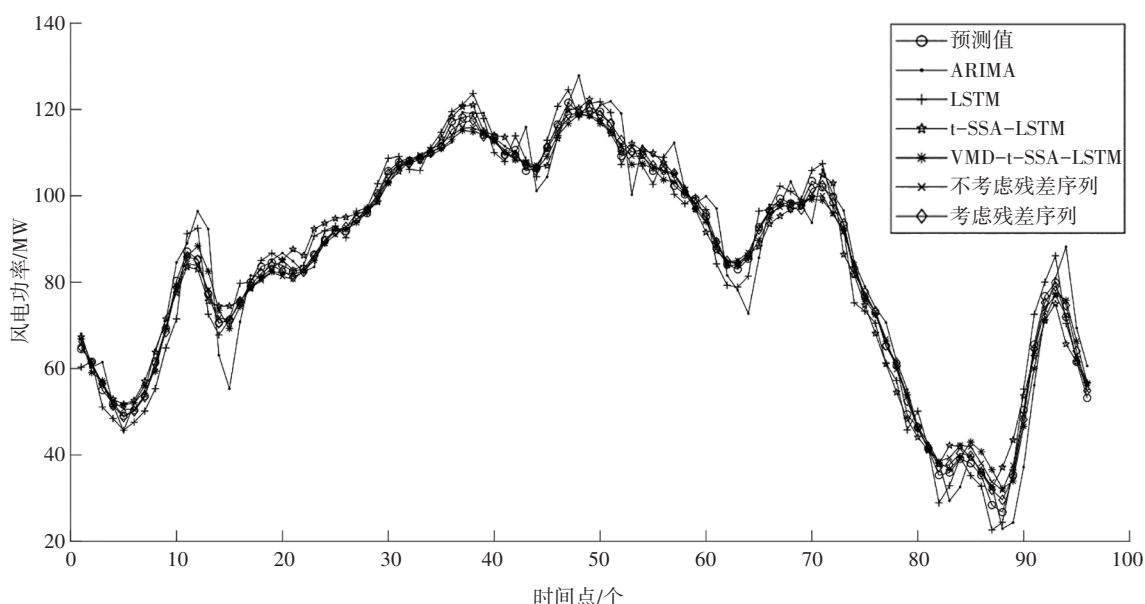


图 7 不同预测模型的预测值和实际值对比

Fig.7 Comparison of predicted values and actual values of different prediction models

t-SSA-LSTM 预测模型相比较于 t-SSA-LSTM 模型, r_{RMSE} 、 r_{MAE} 和 r_{MAPE} 分别下降了 0.568 4、0.532 1、0.007 2, 且在波峰波谷处能够更好地拟合实际功率值. 本文提出的“分解-分类-预测”能够更充分发挥模型特有的属性, 针对低频分量类波动小、平稳线性等特征选择 ARIMA 预测模型; 针对高频分量类波动大、不平稳、强非线性等特征选择 LSTM 预测模型, 并采用 t-SSA 算法对模型参数寻优处理. 其 r_{RMSE} 、 r_{MAE} 和 r_{MAPE} 相比较于 VMD-t-SSA-LSTM 分别下降了 0.673 2、0.494 5、0.007 0. 但是, 对原始功率数据的分解, 屏蔽了原始序列中部分小波真实分量, 在少数波峰波谷点始终与实际功率存在较大的误差. 因此, 将分解屏蔽的小波真实分量提取作为残差序列. 考虑残差序列继承了原始序列的大量噪声信号, 波动性极强, 因此结合气象因素特征选择 t-SSA-LSTM 对其进行预测, r_{RMSE} 、 r_{MAE} 和 r_{MAPE} 相比较于“分解-分类-预测”分别降低了 0.670 0、0.585 3、0.008 6, 充分验证了残差补偿能够有效地提升风电功率预测精度.

总体分析, 本文提出的“分解-分类-预测-残差修正”模型在风电功率预测中, r_{RMSE} 、 r_{MAE} 和 r_{MAPE} 分别为 1.250 6、0.944 6、0.013 9, 有着良好的预测精度.

6 结 论

1) 研究 VMD 分解方法、频率预测判别法以及 t-SSA 优化算法和 ARIMA、LSTM 预测算法, 提出频率游程判别法和 VMD 残差修正的组合预测模型, 旨在提高风电功率预测精度.

2) 调节 VMD 的噪声容忍参数, 对原始风电功率序列进行分解, 得到不受噪声污染的若干子序列, 并用序列之差提取残差序列, 为进一步预测提供基础. 考虑子序列具有不同的波动程度, 采用频率游程判别法将子序列分为高频和低频分量.

3) 使用 t-SSA 优化算法对 LSTM 和 ARIMA 算法的超参数进行寻优, 使用 t-SSA-LSTM 预测高频分量, t-SSA-ARIMA 预测低频分量.

4) 对于残差序列, 采用 t-SSA-LSTM 结合气象因素特征进行滚动预测; 最后线性叠加各模型的预测结果从而得到风电预测结果.

5) 使用国内某风电场 2020 年 5 月的实测数据进行模型性能验证, 选用均方根误差等评价指标进行预测结果评判, 使用 ARIMA、LSTM、t-SSA-LSTM、

VMD-t-SSA-LSTM 和不考虑残差序列的预测模型共 5 种预测模型作为参考, 发现本文预测模型相较于其他的预测模型有相对优越的动态性能, 有着更高的预测精度和更好的稳定性, 且残差的修正补偿对预测精度有很大的提升. 证明了本文所提预测模型的合理性, 对风电功率预测具有一定的借鉴意义.

参考文献

- [1] 雷亚洲. 与风电并网相关的研究课题[J]. 电力系统自动化, 2003, 27(8): 84-89.
LEI Y Z. Studies on wind farm integration into power system[J]. Automation of Electric Power Systems, 2003, 27(8): 84-89. (In Chinese)
- [2] TASCIKARAOGLU A, UZUNOGLU M. A review of combined approaches for prediction of short-term wind speed and power[J]. Renewable and Sustainable Energy Reviews, 2014, 34: 243-254.
- [3] 王聪, 张宏立, 范文慧. 基于投影寻踪主成分分析和耦合模型的风电功率预测[J]. 太阳能学报, 2018, 39(2): 315-323.
WANG C, ZHANG H L, FAN W H. Wind power prediction based on projection pursuit principal component analysis and coupling model[J]. Acta Energetica Solaris Sinica, 2018, 39(2): 315-323. (In Chinese)
- [4] 杨茂, 董昊. 基于数值天气预报风速和蒙特卡洛法的短期风电功率区间预测[J]. 电力系统自动化, 2021, 45(5): 79-85.
YANG M, DONG H. Short-term wind power interval prediction based on wind speed of numerical weather prediction and Monte Carlo method[J]. Automation of Electric Power Systems, 2021, 45(5): 79-85. (In Chinese)
- [5] 胡帅, 向月, 沈晓东, 等. 计及气象因素和风速空间相关性的风电功率预测模型[J]. 电力系统自动化, 2021, 45(7): 28-36.
HU S, XIANG Y, SHEN X D, et al. Wind power prediction model considering meteorological factor and spatial correlation of wind speed[J]. Automation of Electric Power Systems, 2021, 45(7): 28-36. (In Chinese)
- [6] 李俊卿, 李秋佳. 基于 Kriging 和长短期记忆网络的风电功率预测方法[J]. 太阳能学报, 2020, 41(11): 241-247.
LI J Q, LI Q J. Wind power prediction method based on Kriging and LSTM network[J]. Acta Energetica Solaris Sinica, 2020, 41(11): 241-247. (In Chinese)
- [7] 殷豪, 欧祖宏, 陈德, 等. 基于二次模式分解和级联式深度学习的超短期风电功率预测[J]. 电网技术, 2020, 44(2): 445-453.
YIN H, OU Z H, CHEN D, et al. Ultra-short-term wind power prediction based on two-layer mode decomposition and cascaded deep learning[J]. Power System Technology, 2020, 44(2): 445-453. (In Chinese)
- [8] 武小梅, 林翔, 谢旭泉, 等. 基于 VMD-PE 和优化相关向量机

- 的短期风电功率预测[J]. 太阳能学报, 2018, 39(11): 3277-3285.
- WU X M, LIN X, XIE X Q, *et al.* Short-term wind power forecasting based on variational mode decomposition-permutation entropy and optimized relevance vector machine [J]. *Acta Ener-giae Solaris Sinica*, 2018, 39(11): 3277-3285. (In Chinese)
- [9] 周小麟, 童晓阳. 基于 CEEMD-SBO-LSSVR 的超短期风电功率组合预测[J]. 电网技术, 2021, 45(3): 855-864.
- ZHOU X L, TONG X Y. Ultra-short-term wind power combined prediction based on CEEMD-SBO-LSSVR [J]. *Power System Technology*, 2021, 45(3): 855-864. (In Chinese)
- [10] 杨晶显, 张帅, 刘继春, 等. 基于 VMD 和双重注意力机制 LSTM 的短期光伏功率预测[J]. 电力系统自动化, 2021, 45(3): 174-182.
- YANG J X, ZHANG S, LIU J C, *et al.* Short-term photovoltaic power prediction based on variational mode decomposition and long shortterm memory with dual-stage attention mechanism [J]. *Automation of Electric Power Systems*, 2021, 45(3): 174-182. (In Chinese)
- [11] 胡威, 张新燕, 郭永辉, 等. 基于游程检测法重构 CEEMD 的短时风功率预测[J]. 太阳能学报, 2020, 41(11): 317-325.
- HU W, ZHANG X Y, GUO Y H, *et al.* Short-time wind power prediction of CEEMD reconstructed based on run-length detection method [J]. *Acta Ener-giae Solaris Sinica*, 2020, 41(11): 317-325. (In Chinese)
- [12] 江岳春, 杨旭琼, 贺飞, 等. 基于 EEMD-IGSA-LSSVM 的超短期风电功率预测[J]. 湖南大学学报(自然科学版), 2016, 43(10): 70-78.
- JIANG Y C, YANG X Q, HE F, *et al.* Super-short-time wind power forecasting based on EEMD-IGSA-LSSVM [J]. *Journal of Hunan University (Natural Sciences)*, 2016, 43(10): 70-78. (In Chinese)
- [13] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. *IEEE Transactions on Signal Processing*, 2014, 62(3): 531-544.
- [14] 宦娟, 曹伟建, 秦益霖, 等. 基于游程检测法重构集合经验模态的养殖水质溶解氧预测[J]. 农业工程学报, 2018, 34(8): 220-226.
- HUAN J, CAO W J, QIN Y L, *et al.* Dissolved oxygen prediction in aquaculture based on ensemble empirical mode decomposition and reconstruction using run test method [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2018, 34(8): 220-226. (In Chinese)
- [15] 徐青山, 郑维高, 卞海红, 等. 考虑游程检测法重构的 EMD-Elman 风电功率短时组合预测[J]. 太阳能学报, 2015, 36(12): 2852-2859.
- XU Q S, ZHENG W G, BIAN H H, *et al.* Short term forecast of wind power based on empirical mode decomposition and Elman neural network considering runs-test reconstruction [J]. *Acta Ener-giae Solaris Sinica*, 2015, 36(12): 2852-2859. (In Chinese)
- [16] 田中大, 李树江, 王艳红, 等. 基于 ARIMA 与 ESN 的短期风速混合预测模型[J]. 太阳能学报, 2016, 37(6): 1603-1610.
- TIAN Z D, LI S J, WANG Y H, *et al.* Short-term wind speed hybrid prediction model based on ARIMA and esn [J]. *Acta Ener-giae Solaris Sinica*, 2016, 37(6): 1603-1610. (In Chinese)
- [17] 许淼, 刘宏飞, 初凯. 基于 AM-LSTM 模型的共享单车时空需求预测[J]. 湖南大学学报(自然科学版), 2020, 47(12): 77-85.
- XU M, LIU H F, CHU K. Spatiotemporal demand prediction of bike-sharing based on AM-LSTM model [J]. *Journal of Hunan University (Natural Sciences)*, 2020, 47(12): 77-85. (In Chinese)