新能源发电技术

基于排列熵重构的 EEMD-RVM 短期光伏功率预测

林 翔, 武小梅, 谢海波, 谢旭泉

(广东工业大学自动化学院,广东 广州 510006)

摘 要: 针对光伏功率序列周期性非平稳特性,难以准确预测的问题,提出一种基于排列熵重 构原则的集合经验模态分解(ensemble empirical mode decomposition, EEMD)和相关向量机 (relevance vector machine, RVM)的短期光伏功率预测新方法。仿真结果表明:该方法不仅能够有 效提高短期光伏功率预测的准确性,而且缩短了预测时间,提高了光伏功率预测效率,适用于光 伏功率短期在线预测。

关键词: 集合经验模态分解; 排列熵; 相关向量机; 光伏功率预测 中图分类号: TM615 文献标志码: A 文章编号: 1672-3643(2016)06-0042-07 有效访问地址: http://dx.doi.org/10.3969/j.issn.1672-3643.2016.06.008

EEMD-RVM short-term photovoltaic power forecasting based on permutation entropy reconstruction

LIN Xiang, WU Xiaomei, XIE Haibo, XIE Xuquan

(School of Automation, Guangdong University of Technology, Guangzhou Guangdong 510006, China)

Abstract: It's hard to accurately forecast the periodicity of the photovoltaic power series due to its non-stationarity, puts forward a novel short-term photovoltaic power forecasting approach as ensemble empirical mode decomposition (EEMD) and relevance vector machine (RVM) based on permutation entropy reconstruction. The simulation results illustrate that the method can not only improve the accuracy of short- term photovoltaic power forecasting, but also shorten the forecasting time, and promote the efficiency of photovoltaic power forecasting. The method is suitable for photovoltaic power short-term on-line forecasting.

Keywords: ensemble empirical mode decomposition (EEMD); permutation entropy; relevance vector machine (RVM); photovoltaic power forecast

DOI: 10.3969/j.issn.1672-3643.2016.06.008

收稿日期: 2016-08-26

作者简介:林翔(1991),男,硕士研究生,主要研究方向为负荷预测与新能源发电预测应用研究。

光伏发电作为一种新型的可再生能源,近年 来在国内外获得了迅猛地发展,但由于其出力受 太阳辐射强度等多种气象因素的影响,具有一定 的随机波动性与周期性,因此大规模光伏发电系 统并网运行,会对电网安全稳定造成不利影响,从 而增加电网调度工作的难度。如何更加准确地预 测出光伏系统出力,对不同的光伏出力水平采取 相应的调控措施,是学术界和工程界关注的问 题^[1-2]。因此,对光伏系统短期出力更为准确的预 测显得尤为重要。

1 存在问题及分析

由于光伏功率受自然环境变化的影响呈现出 更强的随机波动性与更快的变化频率^[3],传统方 法常采用神经网络或支持向量机建立的单一预测 模型,由于单一预测方法在预测中会受到自身特 性的限制,无法取得最佳的预测性能,所以多采用 建立组合预测模型来提高整体预测性能,尤其是 以信号分解技术建立组合预测模型为主,主要包 含了小波变换^[4](wavelet transform,WT)、经验模 式分解^[5](empirical mode decomposition,EMD)、集 合经验模式分解(ensemble empirical mode decomposition,EEMD)等。小波分解法存在小波基和分 解尺度选取困难问题,经验模式法克服了两者的 缺陷,但却存在模态混叠现象,EEMD法是在EMD 基础上提出的改进方法,性能较其更为优越。

尽管 EEMD 分解方法在预测建模具有诸多优 点,但若直接对 EEMD 分解所得多个分量进行一 一建模预测,会出现引入多重随机误差与增大预 测模型的计算规模等缺点^[8]:如文献[6]指出, EEMD 分解后得到的子模态数目过多,会导致预 测模型训练时间过长,不适用于短期风电在线预 测。文献[7]直接对经验模式分解所得多个分量 分别进行 ELM 建模预测,虽然能够提高光伏功率 预测精度,但是由于对经验模式分解所得的各个 分量进行逐一建模预测,大大增加了预测模型的 计算规模。文献[8]指出了采用经验模式分解建 立组合预测模型过程中,不仅会增加预测部分的 工作量,更是在分解过程中会引入多重随机误差, 最终降低预测模型的精度。

2 基本数学理论

2.1 EEMD原理

EEMD分解法是针对原始EMD存在的模态混 叠等问题的基础上提出的一种改进的方法^[9-11]。 该方法通过在EMD的基础上加入了高斯白噪声, 并进行多次EMD分解,最后将多次分解结果取平 均值作为最终实验结果。高斯白噪声具有频率均 匀分布的特性,使得信号在不同的尺度上具有连 续性,有效地解决了EMD分解过程造成的混叠现 象,从而改善EMD分解法的分析结果。EEMD具 体分解步骤^[10]如下:

(1)确定 EMD 分解法最大迭代次数 M, 白噪 声信号的幅值系数 k 和 EMD 迭代次数 m。

(2)执行第 m 次 EMD 实验,步骤为:

Step1,在原始时间序列 x(t) 中添加一个随机 产生的白噪声序列 n_w(t)

$$x_m(t) = x(t) + kn_m(t) \tag{1}$$

式中: $x_m(t)$ 一加入噪声后待处理的第m个时间 序列。

Step2,采用EMD分解法将 $x_m(t)$ 分解成为n个 IMF分量 $c_{i,m}(t)$ 和一个剩余分量 $r_{n,m}(t)$,EMD具体分 解步骤见文献[9]。

Step3,每次迭代都在原始时间序列 x(t) 中加 入方均根相等且又不相同的白噪声序列,重复步 骤 Step1和Step2,最终获得 M 组不同的 IMF 分量 和剩余分量的组合。

(3)对 M 次 EMD 分解实验后得到的所有 IMF 分量及剩余分量计算均值,最后分解得到的各 IMF 分量及剩余分量的均值即为 EEMD 最后分解 的结果,表示为

$$\bar{c}_{i}(t) = \sum_{m=1}^{M} c_{i,m}(t) / M \qquad i = 1, 2, \cdots, n$$
(2)

$$\bar{r}_{n}(t) = \sum_{m=1}^{M} r_{i,m}(t) / M$$
(3)

2.2 排列熵原理

熵是衡量系统复杂度的一种定量描述工具,

熵值随系统状态的变化而变化。排列熵(permutation entropy, PE)^[12]作为一种衡量时间序列复杂性 的测度方法,因其抗干扰能力强、计算简便等特点 被广泛应用于序列复杂度和非线性分析中。由于 时间序列的微小变化都会影响排列熵的熵值,因此 其熵值可以作为时间序列复杂度分析的重要指标。

若原始时间序列 {x(i), i = 1, 2, ..., N}, 对该序列 进行相空间重构操作,由此可得重构向量

 $X_{j} = [x(j), x(j+\tau), x(j+(m-1)\tau)]$ (4) 式中: m — 嵌入维数;

 τ —延迟时间;下标 $j=1,2,\cdots,N-(m-1)\tau$ 。

重构向量 *X_j* 可作为矩阵 *X* 的行向量,将矩阵 *X* 的每一行,即各重构向量 *X_j* 重新进行升序排 列,即

 $x[j+(j_1-1)\tau] \le x[j+(j_2-1)\tau] \le \dots \le x[j+(j_m-1)\tau]$ (5) 式中: j_1, j_2, \dots, j_m 一重构向量中各个元素所在列的 索引号。

若存在 $x[k+(j_p-1)\tau]=x[k+(j_q-1)\tau]$,则按 j 值 的大小来进行排序。

因此,对于原始时间序列 { $x(i), i = 1, 2, \dots, N$ } 重 构得到的矩阵 X 中每一行,均可得到一组符号序列 $S(g) = (j_1, j_2, \dots, j_m)$ (6)

式中:g=1,2,...,l。

由此可得,总共有 $l \leq m$ 种不同的排列方式。 计算每一种符号序列出现的概率 $P_1, P_2, \cdots P_l$,显 然 $\sum_{i}^{l} P_g = 1$ 。

所以,参照Shannon熵的形式,定义时间序列的PE为

$$H_p(m) = -\sum_{g=1}^{l} P_g \ln P_g \tag{7}$$

当 $P_g = 1/m$ 时, $H_p(m)$ 达到最大值 1n(m!),因此,可将排列熵 $H_p(m)$ 进行归一化处理,

$$H_{p} = H_{p} / \ln(m!) \tag{8}$$

式中, H_p取值范围为0≤H_p≤1,其值大小用 来判定时间序列随机程度的强弱,值越小,表示该 序列越规则;反之,越接近随机,其变化反映并放 大了时间序列的微小变化。

2.3 相关向量机原理

相关向量机^[13-15]是Tipping提出的一种与支持

向量机相似的稀疏概率模型,在贝叶斯框架下实现样本的训练学习,在先验参数的结构基础上采用主动相关决策模型来移除不相关的点,进而得到稀疏化模型。

若已有训练样本输入集 $\{x_n, t_n\}_{n=1}^N$ 为样本数 量, t 为标量输出, RVM 回归模型可定义为

$$t_i = \sum_{i=1}^{N} w_i K(x, x_i) + \omega_0 + \varepsilon$$
(9)

式中: ε 一服从 $N(0,\sigma^2)$ 分布的各独立样本误差;

权值向量 $\omega = (\omega_0, \omega_1, \dots, \omega_N)^T$;

 ω_0 一权值初值;

 $K(x,x_i)$ 一核函数;

x一相关向量。

假定输出集 t 互相独立,训练样本的似然函数为

$$p(t \mid \boldsymbol{\omega}, \boldsymbol{\sigma}^2) = (2\pi\sigma^2)^{-N/2} \exp(-\frac{1}{2\sigma^2} \|t - \boldsymbol{\Phi}\boldsymbol{\omega}\|^2) \quad (10)$$

式中:
$$t = (t_1, t_2, \dots, t_N)^T$$
;
 $\boldsymbol{\omega} = (\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N)^T$;
 $\boldsymbol{\Phi} = [\boldsymbol{\varphi}(x_1), \boldsymbol{\varphi}(x_2), \dots, \boldsymbol{\varphi}(x_N)]^T$;
 $\boldsymbol{\varphi}(x_i) = [1, K(x_1, x_i),$

$$K(x_2, x_i), \cdots, K(x_N, x_i)]^T$$

为了避免直接使用最大似然方法求解 ω 和方 差 σ^2 而带来的过适应现象,根据贝叶斯理论,引 人先决条件: ω 的分布为零均值的标准正态分 布。同时引入超参数 $a = (\alpha_0, \alpha_1, \dots, \alpha_N)^r$,可得:

$$P(\boldsymbol{\omega} \mid \boldsymbol{\alpha}) = \prod_{i=0}^{N} N(\boldsymbol{\omega}_i \mid \mathbf{0}, \boldsymbol{\alpha}_i^{-1})$$
(11)

式中: α_i 一权值 ω_i 对应的超参数。

由自动相关决定先验理论可知,在经过足够 多的更新后,大部分超参数会趋向于无穷大,若对 每个权值限定先决条件,可使与其对应的权值为 0,便可保证了RVM的高稀疏性^[14]。

根据样本似然函数式(10)与先验分布式 (11),根据贝叶斯理论,便可得权值的后验分布概 率密度为:

$$P(\boldsymbol{\omega} \mid t, \boldsymbol{\alpha}, \sigma^{2}) = (2\pi)^{-\frac{N+1}{2}} \boldsymbol{\psi} \mid -\frac{1}{2} \cdot \exp(-\frac{(\boldsymbol{\omega}-\boldsymbol{\mu})^{T}\boldsymbol{\psi}^{-1}(\boldsymbol{\omega}-\boldsymbol{\mu})}{2})$$
(12)

式中: $\psi = (\sigma^{-2}\Phi T\Phi + \alpha)^{-1}$, $\alpha = \operatorname{diag}(\alpha_0, \alpha_1, \cdots \alpha_N)$,

$$\mu = \sigma^{-2} \psi \Phi^T t$$

最后利用极大似然估计法可得超参数 a 和方 差 σ^2 。

若已有输入值 x^{*},则对应的输出概率分布服 从高斯分布,其相应的预测值为

$$y^* = \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\varphi}(x^*) \tag{13}$$

为了改善相关向量机的泛化能力,采用多项 式核函数与高斯核函数线性组合的方式构造一种 组合核函数,该组合核函数表达式为

$$K(x,x_i) = \mu P(x,x_i) + (1-\mu)G(x,x_i)$$
(14)

$$P(x,x_i) = [(x \cdot x_i) + 1]^2$$
(15)

$$G(x, x_i) = \exp(-\|x - x_i\|^2 / \delta^2)$$
(16)

式中: P(x,x_i)一表示多项式核函数;

 $G(x,x_i)$ 一表示高斯核;

 μ 一权重参数, $\mu \in (0,1)$;

 δ 一高斯核参数。本文采用粒子群算法确 定 μ 与 δ 最优值。

3 改进方案

为了克服传统单一预测方法无法取得最佳的 预测性能以及直接对EEMD分解所得多个分量进 行建模预测,会出现引入多重随机误差与增大预 测模型的计算规模等缺点,采用基于排列熵重构 的EEMD-RVM的光伏预测改进方案来提高预测 性能。

3.1 基于排列熵重构的 EEMD-RVM 预测模型

本文提出一种基于排列熵重构原则的集合经 验模态分解和相关向量机的短期光伏功率预测新 方法。首先,对光伏电站出力的功率时间序列进 行 EEMD 分解,得到多个本征模态函数 IMF 和趋 势项Res,考虑到经EEMD分解后产生的模态分量 数目较多,假如对其逐一建模预测,将会带来较大 工作量,为此本文根据排列熵理论对分解所得模 态和剩余分量进行时间序列复杂度分析,将具有 相近熵值的 EEMD 分解项进行重构,重构得到多 组子序列分量;接着分别对重构得到的多组子序 列分量与气象数据(太阳辐射照度、平均湿度与平 均温度)进行数据归一化处理,作为相关向量机的 训练和测试数据,然后分别应用RVM预测模型进 行24h的日前功率预测,最后将多组子序列分量 的日前预测值自适应叠加,得到目标光伏出力24 h的日前预测功率值。具体光伏预测模型算法流 程如图1所示。



图1 光伏功率预测模型算法流程

3.2 仿真验证

3.2.1 EEMD分解

为了验证本文提出的光伏功率预测模型,研 究对象采用位于美国俄勒冈州的Ashland光伏电 站,其总容量为15 kW^[16],研究时段为2014年5月 1日-31日的(6:00-18:00),由于系统采样频率为每 5 min 1次,将15 min的数据求取平均值,每天12 h, 一日总共49个点。在实际建模过程中,选取5月 31日作为RVM模型的测试集,其它作为模型的测 试集。训练集和测试集数据采用历史输出功率数 据与太阳辐射照度、平均湿度与平均温度的气象 数据。为了提高预测精度,本文采用EEMD方法 对原始光伏功率序列进行分解以降低时间序列的 非平稳性,分解后可获得多组不同时间尺度的模 态分量,EEMD分解结果如图2所示。





由图2可知,光伏出力样本输入向量经过 EEMD分解得到8个子模态和1个余量,分解得到 各个分量波动频率依次减小,相比原始光伏出力 时间序列逐渐表现出明显规律性,非线性非平稳 出力时间序列信号自适应的分解为具有不同时间 特征尺度的平稳分量,降低了不同特征信息之间 的彼此影响和干涉,使得对于平稳分量的预测更 加准确。

3.2.2 排列熵重组

考虑到若利用EEMD分解所得多个分量直接 进行一一建模预测,会出现引入多重随机误差,同 时增加预测模型的计算规模。为降低最终光伏预 测结果的误差,同时减少预测模型的计算规模,采 用排列熵法对多个子模态和余量进行时间序列复 杂度分析,其计算所得的熵值大小用来判定时间 序列随机程度的强弱,值越小,表示该序列越规 则;反之,越大接近随机。排列熵计算结果如图3 所示。



从图3可知,各子模态分量的排列熵的熵值 随子模态分量频率的降低整体呈递减趋势,表明 从高频到低频分量复杂度减小,证明了排列熵方 法的有效性。故可以根据图3排列熵值对EEMD 分解得到的子模态进行重构。重构原则主要是把 排列熵值相近的子模态进行合并重组,熵值相差 较大的子模态单独为一组独立的子序列。由图3 所示:模态4-8的熵值分别为0.947 5,0.839 9, 0.788 5,0.744 7,0.702 2;相邻子模态差值分别为 0.107 6,0.051 4,0.043 8,0.042 5,差值基本都在 0.1以下,说明两两模态生成新模式的概率相近, 可把两者合并成一个新的子序列作为RVM新的 输入,而模态1与模态2之间的差值为0.3,相差过 大需要单独作为一个子序列,同理可得子序列重 构结果如表1所示。

	-AC 1	日民心力重用至月コネ
新序列号		原始模态分量序列号
1		1
2		2
3		3
4		4,5,6,7,8
5		9

表1 各模态分量的重构结果

由表1可知,序列号为1、2、3、9的原始模态无 需进行重组,各自可单独作为一组子序列,分别为 新子序列1、2、3、5,而序列号为4、5、6、7、8的原始 模态,由于排列熵值相近,所以将其进行合并重组 形成新子序列4,子序列重构结果如图4所示。



3.2.3 算例结果及分析

为了验证所述方法的效果,将基于排列熵分 析后重构得到的5个子序列(如图4所示)与历史 气象数据形成预测样本集。首先需要进行样本数 据归一化预处理,把最后一天数据作测试集,其他 作为训练集;然后采用RVM预测算法逐一为5个 子序列建立预测模型,并得到各子序列相应的预 测值,再将各子序列的预测值相叠加就可以得到 最终的预测结果,最佳预测结果如图5所示。



图5 最佳光伏功率预测结果

为定量评价各方法的预测结果与真实值接近 程度,以国际上普遍采用的均方根误差 e_{RMSE} 和绝 对平均百分比误差 e_{MAPE} 作为预测误差评价指标, 定义如下

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$
(17)

$$e_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\%$$
(18)

式中:n、y_i、ŷ_i—分别是样本数,实际值和预测值。 表2 各预测模型预测误差与训练时间比较

预测模型	$e_{\rm RMSE}/{ m kW}$	$e_{_{\mathrm{MAPE}}}$ /%	训练时间/s
RVM	0.386	8.258	114
EEMD-RVM	0.353	6.681	986
本文方法	0.304	6.276	565

为了检验本文方法的有效性及优越性,采用 未考虑排列熵重构原则的RVM与EEMD-RVM两 种预测模型与本文方法进行比较,预测效果以均 方根误差 e_{RMSE}和绝对平均百分比误差 e_{MAPE} 作为 评价标准,预测结果误差与预测模型训练时间比 较见表2。

由图5与表2可知,3种预测模型相比较,本文 方法能够较好地捕捉光伏功率的变化规律,均方 根误差 e_{RMSE} 与绝对平均百分比误差 e_{MAPE} 分别为 0.304 kW和6.276%,两个误差指标都低于其他两 种预测模型,而且本文方法训练时间为565 s,低

《宁夏电力》2016年第6期

于 EEMD-RVM 预测模型所用的训练时间,但又高 于 RVM 预测模型所用的训练时间。从预测误差 与训练时间两个指标整体考虑,本文方法不仅降 低了光伏预测误差,同时缩短了预测模型的训练 时间,优于另外两种预测模型。

4 效果评价

(1)采用排列熵重构的EEMD-RVM预测模型 进行各子模态分量的复杂度分析,将复杂度相近 的子模态相叠加,可以进一步改善预测精度,这是 由于重构后的新子序列局部光伏出力特性更加明 显,降低了EEMD分解过程产生的随机误差。

(2)本文方法预测模型训练时间低于未采用 排列熵重构的EEMD-RVM预测模型,说明了采用 排列熵理论对各子模态分量合并重组,同时降低 了RVM预测模型的计算规模,更有利于RVM模型 的预测。因此,本文所提方法不仅改善了光伏功 率预测精度,而且缩短了预测时间,适用于短期在 线预测,可以较好地推广应用。

5 结论

(1)本文采用的EEMD-RVM组合预测模型提高了整体预测性能,预测效果理想,克服了传统单一预测方法会受到自身特性的限制、无法取得最 佳预测性能的缺点。

(2)采用排列熵进行各子模态分量的复杂度 分析,将复杂度相近的子模态相叠加,可以进一步 改善预测精度,而且重构后的新子序列局部光伏 出力特性更加明显,降低了EEMD分解过程产生 的随机误差。

(3)基于排列熵重构原则的EEMD-RVM的短 期光伏功率预测新方法,可以减少EEMD分解过 程中产生多重随机误差以及预测部分工作量,算 例结果表明:该方法不仅提高了预测精度,而且缩 短了预测时间,适用于光伏短期在线预测,可以较 好地推广应用。

参考文献

- [1] 代倩,段善旭,蔡涛,等. 基于天气类型聚类识别的光 伏系统短期无辐照度发电预测模型研究[J].中国电机 工程学报,2011,31(34):28-35.
- [2] 艾欣, 韩晓男, 孙英云. 大型光伏电站并网特性及其低 碳运行与控制技术[J]. 电网技术, 2013, 37(1):15-23.
- [3] 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(4):140-151.
- [4] 朱红路,李旭,姚建曦,等. 基于小波分析与神经网络的光伏电站功率预测方法[J].太阳能学报,2015,36
 (11):2725-2730.
- [5] 王艳,杨德全.基于EMD-ESNs的光伏系统短期发电 量预测[J].可再生能源,2014,32(10):1436-1440.
- [6] 张学清,梁军,张熙,等.基于样本熵和极端学习机的 超短期风电功率组合预测模型[J].中国电机工程学 报,2013,33(25):33-40.
- [7] 李多,董海鹰,杨立霞. 基于 EMD 与 ELM 的光伏电站 短期功率预测[J]. 可再生能源,2016,34(2):173-177.
- [8] 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
- [9] 茆美琴,龚文剑,张榴晨,等. 基于EEMD-SVM方法 的光伏电站短期出力预测[J].中国电机工程学报, 2013(34):17-24.
- [10] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences, 1998, 454(1971):903-995.
- [11] 范磊, 王越, 梁智, 等. 基于 EEMD-RVM 的短期风功 率区间预测[J]. 广东电力, 2016, 29(2): 14-20.
- [12] 李军,李青. 基于 CEEMDAN-排列熵和泄漏积分 ESN 的中期电力负荷预测研究[J]. 电机与控制学报, 2015,19(8):70-80.
- [13] Tipping M E. Sparse bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research, 2001, 1(3):211-244.
- [14] 李慧杰,刘亚南,卫志农,等. 基于相关向量机的短期风 速预测模型[J]. 电力自动化设备,2013,33(10):28-32.
- [15] Yan J, Liu Y, Han S, et al. Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine[J]. Renewable & Sustainable Energy Reviews, 2013, 27(6):613 - 621.
- [16] 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(16):16-22.