

计及历史气象数据的短期风速预测

史宇伟, 潘学萍

(河海大学 能源与电气学院, 江苏 南京 210098)

摘要: 提出一种计及历史气象数据的短期风速预测方法。首先将历史风速数据和历史气象数据作为风速预测的原始输入, 采用混合特征选择(HFS)方法对输入向量进行删选, 选取与预测风速强相关的变量, 生成预测模型的输入特征集; 然后运用异方差高斯过程回归(HGP)模型进行建模, 该模型能体现风速的随机性。根据其实测风速数据进行提前 1 h 风速预测, 结果表明所提方法能提高风速预测精度。

关键词: 风电; 风速; 预测; 气象数据; 混合特征选择; 异方差高斯过程; 模型

中图分类号: TM 614

文献标识码: A

DOI: 10.3969/j.issn.1006-6047.2014.10.012

0 引言

风电作为当前发展最快的可再生能源发电形式, 得到了世界各国的广泛重视^[1]。随着风电规模的扩大, 当风电穿透功率超过一定数值后, 风电的接入可能影响电能质量和电力系统的安全稳定运行^[2-3]。如何提高风电场风速或风功率的预测精度是合理利用风电资源的关键问题。

短期风速预测的时间分辨率一般为 1 h。对风速或风电功率的准确预测, 可以减少电力系统运行成本和旋转备用, 提高风电穿透功率极限, 增强风电场在电力市场中的竞争能力^[4]。然而由于风速的随机性和间歇性, 准确预测风速难度较大, 目前风电场风速预测误差一般在 25%~40%^[5]。

目前已有的风速预测方法可分为基于物理模型的方法和基于历史数据的预测方法两大类。物理模型法采用天气预报数据进行预测, 由于气象预报数据更新频率低, 适用于中期风速预测。基于历史数据的预测方法较多, 主要有空间相关法^[6]、时间序列法^[7]、人工神经网络法^[8]、支持向量机法^[9]等。由于风速与气象因素(如风向、气温、气压等)密切相关, 在历史数据的输入选择上, 越来越多的学者提出历史风速数据和气象数据相结合的方法, 进行风速或风功率预测^[10-11]。如何对多种输入数据进行最佳的删选, 对提高预测精度有重要作用, 然而这方面的研究很少。文献[12-13]在预测模型的输入变量选择中, 采用主成分分析(PCA)法对输入特征变量进行提取, 减少了输入变量的个数。然而主成分分析法是通过数学变换方式进行特征提取, 不能计及原始数据的物理特性。因此, 本文提出采用混合特征选择算法, 从包含历史风速数据和历史气象数据的输入向量中选取与预测风速强相关的变量, 生成预测模型的输入特征集。该方法能够剔除与预测风速不相

关变量和冗余变量。

在获得预测模型的输入特征集后, 本文进一步采用异方差高斯过程(GP)进行风速预测。高斯过程具有容易实现、灵活的非参数推断及预测输出具有概率意义等优点。它对处理高维数、小样本、非线性等复杂分类和回归问题具有很好的适应性, 且泛化能力强^[14-15]。但标准高斯过程假定噪声为高斯分布, 且整个数据集中噪声方差为常值。而实际风速具有较强的随机性及波动性, 假设噪声方差为常数常带来较大的预测误差。异方差高斯过程是标准高斯过程的一种扩展^[15], 它首先运用高斯过程对无噪声输出进行建模, 然后对噪声再运用高斯过程进行建模, 该方法能体现出风速的随机性及波动性。

1 混合特征选择

风速不仅与历史风速数据有关, 还与气象信息密切相关。本文将历史风速数据和历史气象数据作为预测模型的原始候选输入变量集 Q , 因历史气象数据收集工作的限制, 本文仅考虑风向、温度、气压这 3 类气象因素。原始候选输入变量集 Q 的数据结构如下:

$$\{S(t-1), S(t-2), \dots, S(t-N_s), \\ D(t-1), D(t-2), \dots, D(t-N_D), \\ T(t-1), T(t-2), \dots, T(t-N_T), \\ P(t-1), P(t-2), \dots, P(t-N_P)\} \quad (1)$$

其中, $S(t-i)$ 、 $D(t-i)$ 、 $T(t-i)$ 和 $P(t-i)$ 分别为 $t-i$ 时刻的风速、风向、温度和气压; N_s 、 N_D 、 N_T 和 N_P 分别为风速、风向、温度和气压历史数据的个数。

原始输入变量集 Q 中包含历史风速数据和历史气象数据, Q 维数较高, 含有大量与待预测风速不相关的变量和冗余变量。这里采用混合特征选择法选择与待预测风速的强相关变量, 步骤如下: ①运用互信息法从原始输入特征集 Q 中选取出候选特征变量集 G ; ②结合预测模型, 利用改进的自适应遗传算法

选取出预测精度高的输入变量集,作为最终的输入特征集 X 。

1.1 基于互信息法的输入特征初选

互信息^[16]表示 2 个变量共同拥有信息的含量,可用来衡量 2 个变量间相互依赖的程度。给定 2 个离散量 U 和 V ,设它们的边缘概率分布分别为 $p(u)$ 和 $p(v)$,联合概率分布为 $p(u,v)$,则互信息 $I(U;V)$ 定义为:

$$I(U;V) = \sum_{u \in U} \sum_{v \in V} p(u,v) \lg \frac{p(u,v)}{p(u)p(v)} \quad (2)$$

互信息也可用信息熵表示:

$$I(U;V) = H(U) - H(U|V) \quad (3)$$

$$H(U) = - \sum_{u \in U} p(u) \lg p(u) \quad (4)$$

$$H(U|V) = - \sum_{u \in U} \sum_{v \in V} p(u,v) \lg p(u|v) \quad (5)$$

其中, $H(U)$ 表示 U 的信息熵; $H(U|V)$ 表示 U 关于 V 的条件熵。当 U 和 V 完全无关或相互独立时,它们的互信息为 0,二者不存在任何依赖关系。互信息的估计以概率密度估计为基础,这里采用文献[16]的 k 近邻法计算互信息。

输入特征变量的选择既要考虑与输出变量的相关性,还要考虑与已选特征变量之间的冗余性。本文采用文献[17]中改进的基于互信息特征选择的 MMIFS (Modified Mutual Information-based Feature Selection) 算法作为评价标准,具体如下:

$$G(f_i) = I(f_i; y) - \frac{\beta}{|G|} \sum_{f_s \in G} I(f_i; f_s) \quad (6)$$

其中, β 为经验参数,用来权衡相关性与冗余性,一般取 $\beta=0.5$; $|G|$ 表示已经选择的特征集的大小。

输入特征粗选的基本步骤如下。

a. 初始化: Q 为包含所有候选变量的原始输入特征集, G 为空集。

b. 计算 Q 集中每个特征 f_i 与输出变量 y 的互信息 $I(f_i; y)$ 。

c. 选择最大化 $I(f_i; y)$ 的特征 f_i , 使得 Q 集中剔除特征 f_i , G 集中增加特征 f_i 。

d. 重复以下步骤直至 $|G|=g$: 对任意 $f_i \in Q, f_s \in G$, 计算所有特征对的互信息 $I(f_i; f_s)$; 选择使式(6)最大的特征 f_i , 使 Q 集中剔除特征 f_i , G 集中增加特征 f_i 。

e. 输出候选输入特征集 G 。

1.2 基于改进自适应遗传算法的输入特征优选

根据上述的互信息法得到候选输入特征集 G 后,进一步考虑风速预测模型输入特征变量的不同组合,对一个可能的输入特征组合采用改进自适应遗传算法进行优劣评价,得到最优的输入特征集 X 。改进自适应遗传算法的优劣评价步骤如下。

a. 编码: 采用二进制编码,遗传算法的个体长度为特征向量维度 L ,即候选特征集 G 的大小。其中第 i 位为 1 表示选择特征 f_i , 为 0 表示不选择特征 f_i 。

b. 初始种群: 充分利用互信息法中特征选择的先验知识,将互信息法中相关性大的特征初始化遗传算法中的部分种群。设种群大小为 P ,个体长度为 L ,自定义整数 M 和 N ,其中 $1 \leq M \leq P, 1 \leq N \leq L$,具体方法为:取互信息法得到的特征集 G 中最好的 N 个特征,即特征集 G 中前 N 个特征;初始化种群中 M 个个体,使得到的最好的 N 个特征设置为 1,其余特征随机设置为 0 或 1;随机初始化剩余的 $P-M$ 个个体。

c. 适应度函数: 综合考虑风速预测精度和输入特征集的大小,适应度函数如式(7)所示。

$$M(x_i) = \text{error}(x_i) \quad (7)$$

其中, $\text{error}(x_i)$ 为用个体 x_i 对应的特征子集进行风速预测时的预测误差,文中采用异方差高斯过程进行预测,预测误差采用均方根误差。

d. 遗传算子。

选择操作采用轮盘赌的比例选择法;交叉操作采用单点交叉;变异操作采用基本位变异。交叉概率和变异概率^[18]是随个体适应度值的变化而变化,适应度值好(即适应度函数较小)的个体受到“保护”,适应度值差(即适应度函数较大)的个体遭到“破坏”。交叉概率 p_c 和变异概率 p_m 形式如下:

$$p_c = \begin{cases} p_{c1} - \frac{(p_{c1} - p_{c2})(f_{\text{avg}} - f')}{f_{\text{avg}} - f_{\text{min}}} & f' \leq f_{\text{avg}} \\ p_{c1} & f' > f_{\text{avg}} \end{cases} \quad (8)$$

$$p_m = \begin{cases} p_{m1} - \frac{(p_{m1} - p_{m2})(f_{\text{avg}} - f)}{f_{\text{avg}} - f_{\text{min}}} & f \leq f_{\text{avg}} \\ p_{m1} & f > f_{\text{avg}} \end{cases} \quad (9)$$

其中, f_{min} 为最小个体适应度; f_{avg} 为平均个体适应度; f' 为进行交叉的 2 个个体中较小的适应度; f 为变异个体的适应度。

e. 精英保留策略。

为保证每一代优良个体不被破坏,采用精英保留策略:如果下一代群体的最佳个体适应度小于当前群体最佳个体适应度,则将当前群体最佳个体直接复制到下一代,替代下一代中适应度最差的个体。

f. 终止条件。

终止条件为达到最大迭代次数或连续 4 次最优个体保持不变。

通过上述输入特征集的初选和优选,获得与预测风速强相关的输入特征集 X ,将其作为输入,进一步采用异方差高斯过程回归模型进行风速预测。

2 异方差高斯过程回归

2.1 异方差高斯过程

将数据集 $D = \{(x_i, y_i) \mid (i = 1, 2, \dots, n)\}$ 作为训练

集,其中, $\mathbf{x}_i$ 为 d 维输入向量, y_i 为输出值, n 为训练样本数。

异方差高斯过程设输入向量 $\mathbf{x}_i$ 和目标输出值 y_i 符合如下形式:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \quad (10)$$

其中, $f(\mathbf{x}_i)$ 为潜在函数,其先验概率分布 $p(f(\mathbf{x}_i))$ 为高斯分布;噪声 ε_i 为均值为零、方差为 σ_i^2 的高斯分布,记为 $\varepsilon_i \sim N(0, \sigma_i^2)$ 。与高斯过程模型不同,异方差高斯过程中 σ_i^2 不为常数,噪声方差 σ_i^2 满足函数关系 $\sigma_i^2 = r(\mathbf{x}_i)$ 。

由高斯过程理论可知,给定 n_* 个测试数据集 $D_* = \{(\mathbf{x}_{*i}, y_{*i}) | (i=1, 2, \dots, n_*)\}$, 输入向量 $\mathbf{X}_*$ 的预测分布 $p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y})$ 具有如下高斯分布形式:

$$p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}) \sim N(\mathbf{m}_*, \Sigma_*) \quad (11)$$

其中,均值向量 $\mathbf{m}_*$ 和协方差矩阵 Σ_* 的形式如下:

$$\mathbf{m}_* = E[\mathbf{y}_*] = \mathbf{K}_*(\mathbf{K} + \Omega)^{-1} \mathbf{y} \quad (12)$$

$$\Sigma_* = \text{var}[\mathbf{y}_*] = \mathbf{K}_{**} + \Omega_* - \mathbf{K}_*(\mathbf{K} + \Omega)^{-1} \mathbf{K}^T \quad (13)$$

其中, $\mathbf{K} \in \mathbb{R}^{n \times n}$, $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$; $\mathbf{K}_* \in \mathbb{R}^{n_* \times n}$, $K_{*ij} = k(\mathbf{x}_{*i}, \mathbf{x}_j)$; $\mathbf{K}_{**} \in \mathbb{R}^{n_* \times n_*}$, $K_{**ij} = k(\mathbf{x}_{*i}, \mathbf{x}_{*j})$; $\Omega = \text{diag}(\mathbf{r})$, $\mathbf{r} = (r(\mathbf{x}_1), r(\mathbf{x}_2), \dots, r(\mathbf{x}_n))^T$; $\Omega_* = \text{diag}(\mathbf{r}_*)$, $\mathbf{r}_* = (r(\mathbf{x}_{*1}), r(\mathbf{x}_{*2}), \dots, r(\mathbf{x}_{*n_*}))^T$ 。

这里采用 ISO 型平方指数协方差函数,即:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_i^2 \exp \left[-\frac{1}{2\lambda^2} (\mathbf{x}_i - \mathbf{x}_j)^2 \right] \quad (14)$$

其中, λ 为关联性测度超参数,其值越大,表示输入输出相关性越小; σ_i^2 为核函数的信号方差,用来控制局部相关性的程度。这些参数称为模型的超参数,因均为正数,可表示为 $\theta = (\ln \sigma_i, \ln \lambda)$ 。

异方差高斯过程由 2 个同方差的高斯过程组成:第一个高斯过程模型 G_1 对无噪声目标输出进行建模,第二个高斯过程 G_2 对各噪声的对数进行建模,表示为 $\mathbf{z}(\mathbf{x}_i) = \lg(r(\mathbf{x}_i))$, 该 $\mathbf{z}$ 过程模型中采用超参数为 θ_z 的协方差函数 k_z 。

异方差高斯过程模型中,噪声方差为潜在的独立变量,因此输入向量 $\mathbf{X}_*$ 的预测分布 $p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y})$ 变为如下形式:

$$p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}) = \iint p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}, \mathbf{z}, \mathbf{z}_*) \times p(\mathbf{z}, \mathbf{z}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}) d\mathbf{z} d\mathbf{z}_* \quad (15)$$

给定 $(\mathbf{z}, \mathbf{z}_*)$, $p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}, \mathbf{z}, \mathbf{z}_*)$ 满足高斯分布,其均值和方差可由式(12)、(13)求得,但 $p(\mathbf{z}, \mathbf{z}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y})$ 的求解较为困难。文献[15]给出了近似预测分布如下:

$$p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}) \approx p(y_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{z}}_*) \quad (16)$$

其中, $(\hat{\mathbf{z}}, \hat{\mathbf{z}}_*) = \arg \max_{(\mathbf{z}, \mathbf{z}_*)} p(\mathbf{z}, \mathbf{z}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{y})$ 。

2.2 异方差高斯过程训练

异方差高斯过程模型中, $\mathbf{y}$ 过程中的超参数 θ 和 $\mathbf{z}$ 过程中超参数 θ_z 的选取,对预测结果有着重要

的影响。

异方差高斯过程的训练步骤如下。

a. 数据集 $D = \{(\mathbf{x}_i, y_i) | (i=1, 2, \dots, n)\}$, 运用标准高斯过程 G_1 对其进行建模,根据输入向量 $\mathbf{x}_i$ 预测输出值 y 。

b. 根据模型 G_1 , 估计训练数据的噪声水平,即 $z_i = \lg \{ \text{var}[y_i, G(\mathbf{x}_i, \mathbf{X}, \mathbf{y})] \}$, 根据文献[15], $\text{var}[y_i, G(\mathbf{x}_i, \mathbf{X}, \mathbf{y})]$ 近似取为:

$$\text{var}[y_i, G(\mathbf{x}_i, \mathbf{X}, \mathbf{y})] \approx s^{-1} \sum_{j=1}^s 0.5(y_i - y_{ij}) \quad (17)$$

其中, y_{ij} 为当前高斯过程模型在输入向量 $\mathbf{x}_i$ 的预测样本值; s 为样本的大小。从而形成新的数据集 $D' = \{(\mathbf{x}_i, z_i) | (i=1, 2, \dots, n)\}$ 。

c. 运用标准高斯过程模型 G_2 对数据集 D' 进行建模。

d. 运用组合的高斯过程模型 G_3 对数据集 D 进行建模,其中,用模型 G_2 预测噪声水平 r_{io} 。

e. 如果 G_3 不收敛,设 G_1 与 G_3 相同,重新返回步骤 b。

异方差高斯过程模型训练的过程就是,根据当前的噪声模型和相应数据得出噪声水平,然后得出改进的数据集,进而计算模型 G_3 的最大似然参数,重复该过程,直到模型收敛。

3 计及气象数据的风速预测流程

计及气象数据的短期风速预测流程如图 1 所示。

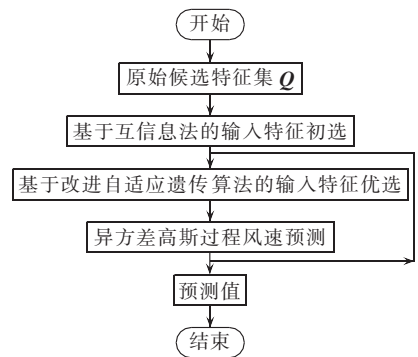


图 1 风速预测流程图

Fig.1 Flowchart of wind speed forecasting

风速预测精度用平均绝对误差(MAE)和平均绝对百分比误差(MAPE)来评估,分别如式(18)和式(19)所示。

$$E_{MAE} = \frac{1}{N_u} \sum_{i=1}^{N_u} |y_i - \hat{y}_i| \quad (18)$$

$$E_{MAPE} = \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\% \quad (19)$$

其中, y_i 为实际风速值; $\hat{y}_i$ 为预测风速值; N_u 为预测值的个数。

4 算例分析

以某风电场实测的 2012 年小时风速数据和对应的实测小时气象数据(风向、气温、气压)为例。

依次选取春季、夏季、秋季和冬季的 3 月 21 日、6 月 21 日、10 月 21 日、12 月 21 日作为预测日,利用各预测日前 20 d 数据进行建模,进行提前 1 h 预测。以预测秋季 10 月 21 日为例,10 月 1 日到 10 月 20 日共 20 d 的历史风速序列、风向序列、温度序列和气压序列分别见图 2—5。

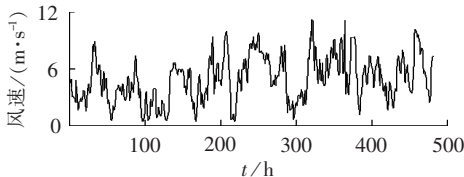


图 2 风速序列
Fig.2 Wind speed series

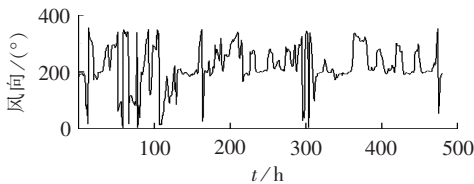


图 3 风向序列
Fig.3 Wind direction series

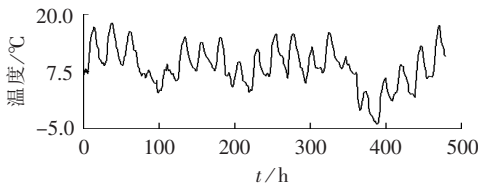


图 4 温度序列
Fig.4 Temperature series

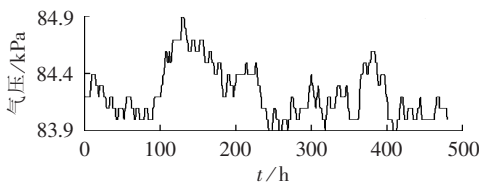


图 5 气压序列
Fig.5 Pressure series

取 $N_s=N_D=N_T=N_P=10$,得原始候选输入特征集 Q 。基于互信息法的输入特征选择方法,取 $\beta=0.5, g=12$,初选得候选输入向量集 G 为 $[S(t-1), D(t-2), S(t-2), T(t-1), S(t-4), D(t-1), S(t-3), D(t-6), S(t-5), P(t-2), D(t-4), T(t-9)]$ 。

进一步基于改进自适应遗传算法进行输入特征优选。这里取 $P=20, L=12, M=8, N=6, p_{c1}=0.8, p_{c2}=0.6, p_{m1}=0.1, p_{m2}=0.01$ 。输入特征选择中,把 20 d 的输入数据分为训练集和测试集,随机选取 80% 的数据作为训练集,剩下的 20% 数据作为测试集,根

据遗传算法从候选特征集 G 中选取预测模型的最优输入特征集 X 为 $[S(t-1), D(t-2), S(t-2), T(t-1), S(t-4), D(t-1), S(t-3), D(t-6)]$,即候选输入特征集 G 中的前 8 个。

最后,将所得到的最优输入特征集 X 作为模型输入,基于异方差高斯过程进行提前 1 h 风速预测,可得 10 月 21 日风速值如图 6 所示。

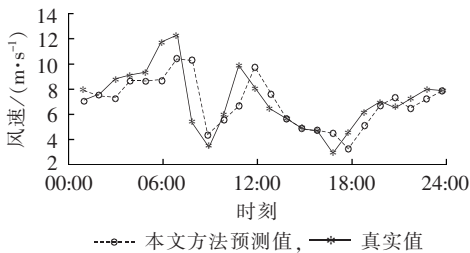


图 6 风速预测结果
Fig.6 Results of wind speed forecasting

由图 6 可以看出,本文方法预测值的变化趋势与真实值基本一致,但对于个别风速变化很大的时刻,误差仍较大。

针对上述算例,本文方法与不计及气象数据时的 ARIMA 模型、不计及气象数据时的 HGP 模型、计及气象数据的主成分分析法和 HGP 模型的预测误差见表 1。

表 1 预测误差
Table 1 Forecast error

季节	预测误差	不计及气象数据的 ARIMA 模型	不计及气象数据的 HGP 模型	计及气象数据的 PCA 和 HGP 模型	本文方法
春季	MAE	0.8103	0.8083	0.7836	0.7586
	MAPE/%	11.69	11.64	11.01	10.81
夏季	MAE	1.7655	1.6867	1.5675	1.4309
	MAPE/%	45.42	40.43	38.66	36.32
秋季	MAE	1.2726	1.2025	1.2008	1.1801
	MAPE/%	20.47	18.52	18.27	17.65
冬季	MAE	0.8095	0.7976	0.7243	0.6356
	MAPE/%	0.3669	35.98	29.74	26.91

由表 1 可以看出,基于 HGP 模型的风速预测精度要高于 ARIMA 模型;计及气象数据的预测精度要高于不计及气象数据的预测精度;本文提出的混合特征选择 HGP 模型比主成分分析 HGP 模型的风速预测精度高。

进一步选取春季、夏季、秋季及冬季的 3 月 21 日至 3 月 27 日、6 月 21 日至 6 月 27 日、10 月 21 日至 10 月 27 日、12 月 21 日至 12 月 27 日的数据作为待预测风速,利用各预测日前 20 d 数据进行建模,进行提前 1 h 预测。图 7 所示为 10 月 21 日至 10 月 27 日的预测值。

本文方法与不计及气象数据时的 ARIMA 模型、不计及气象数据时的 HGP 模型、计及气象数据的主

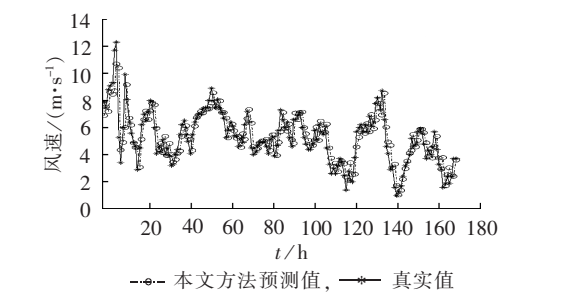


图 7 风速预测结果

Fig.7 Results of wind speed forecasting

成分分析法和 HGP 模型的预测误差见表 2。

表 2 预测误差
Table 2 Forecast error

季节	预测误差	不计及气象数据的 ARIMA 模型	不计及气象数据的 HGP 模型	计及气象数据的 PCA 和 HGP 模型	本文方法
春季	MAE	0.9618	0.8944	0.8753	0.8634
	MAPE/%	23.23	22.59	21.90	20.08
夏季	MAE	1.5554	1.5285	1.4467	1.4294
	MAPE/%	30.65	29.48	28.79	28.02
秋季	MAE	0.7641	0.7507	0.7492	0.7487
	MAPE/%	17.90	17.18	17.04	16.91
冬季	MAE	0.9295	0.8990	0.8314	0.7611
	MAPE/%	39.69	36.18	35.83	31.27

由表 2 可以看出,在增加了待预测的风速数据,仍可得出与表 1 一致的结论,即:基于 HGP 模型的风速预测精度要高于 ARIMA 模型;计及气象数据的预测精度要高于不计及气象数据的预测精度;本文提出的混合特征选择 HGP 模型比主成分分析 HGP 模型的风速预测精度高。

夏季的风速预测绝对误差较大,但相对误差却偏小,这是因为夏季的平均风速较高;冬季则相反。因此在分析风速预测误差时,应同时计及绝对误差和相对误差指标。

5 结论

本文考虑气象数据对风速的影响,将历史风速数据和气象数据作为原始输入向量。针对原始输入的数据量较大,增加了预测时间和预测的难度,为此本文提出采用混合特征选择方法,一方面提取与风速预测序列具有强相关性的特征变量,剔除不相关变量和冗余变量;另一方面还减少了原始的输入数据。从而通过混合特征选择方法选取适合本次风速预测的最佳输入向量。

文中进一步采用异方差高斯过程回归模型进行了短期风速预测。通过不同季节的单预测日和多预测日的预测结果表明:计及历史气象数据的预测方法比不计及历史气象数据的预测方法具有更高的预测精度;计及气象数据时,采用混合特征选择的预处理比主成分分析法具有更高的预测精度。

由于风速的随机性强,因此风速预测很难精确。本文的贡献在于:在计及历史气象数据的预测时,采用合适的输入特征选择有助于提高风速的预测精度。

参考文献:

[1] 李晶,宋家骅,王伟胜. 大型变速恒频风力发电机组建模与仿真[J]. 中国电机工程学报,2004,24(6):100-105.
LI Jing,SONG Jiahua,WANG Weisheng. Modeling and dynamic simulation of variable speed wind turbine with large capacity [J]. Proceedings of the CSEE,2004,24(6):100-105.

[2] MELAM B,AMIT J,VENKATA S. Wind speed forecasting:present status [C]//2010 International Conference on Power System Technology. Hangzhou,China:IEEE,2010:1-6.

[3] BARTHELMIE R J,MURRAY F,PRYOR S C. The economic benefit of short term forecasting for wind energy in the UK electricity market[J]. Energy Policy,2008,36(5):1687-1696.

[4] FABBRI A,ROMAN T G S,ABBAD J R,et al. Assessment of the cost associated with generation prediction errors in a liberalized electricity market[J]. IEEE Trans on Power Systems,2006,21(2): 965-972.

[5] 杨秀媛,肖洋,陈树勇. 风电场风速和发电功率预测研究[J]. 中国电机工程学报,2005,25(11):1-5.
YANG Xiuyuan,XIAO Yang,CHEN Shuyong. Wind speed and generated power forecasting in wind farm[J]. Proceedings of the CSEE,2005,25(11):1-5.

[6] 李文良,卫志农,孙国强,等. 基于改进空间相关法和径向基神经网络的风电场短期风速分时预测模型[J]. 电力自动化设备,2009, 29(6):89-92.
LI Wenliang,WEI Zhinong,SUN Guoqiang,et al. Multi-interval wind speed forecast model based on improved spatial correlation and RBF neural network[J]. Electric Power Automation Equipment,2009,29(6):89-92.

[7] 丁明,张立军,吴义纯. 基于时间序列分析的风电场风速预测模型[J]. 电力自动化设备,2005,25(8):32-34.
DING Ming,ZHANG Lijun,WU Yichun. Wind speed forecast model for wind farms based on time series analysis[J]. Electric Power Automation Equipment,2005,25(8):32-34.

[8] KARINIOTAKIS G,STAVRAKAKIS G,NOGARET E. Wind power forecasting using advanced neural network models[J]. IEEE Trans on Energy Conversion,1996,11(4):762-767.

[9] 杨洪,古世甫,崔明东,等. 基于遗传优化的最小二乘支持向量机风电场风速短期预测[J]. 电力系统保护与控制,2011,39(11): 44-61.
YANG Hong,GU Shifu,CUI Mingdong,et al. Forecast of short-term wind speed in wind farms based on GA optimized LS-SVM [J]. Power System Protection and Control,2011,39(11):44-61.

[10] FAN S,LIAO J R,YOKOYAMA R,et al. Forecasting the wind generation using a two-stage network based on meteorological information[J]. IEEE Trans on Energy Conversion,2009,24(2): 474-482.

[11] HAQUE A U,MANDAL P,MENG J,et al. A novel hybrid approach based on wavelet transform and fuzzy ARTMAP network for predicting wind farm power production [C]//Industry Applications Society Annual Meeting(IAS). Las Vegas,USA:IEEE, 2012:1-8.

[12] 刘宝英,杨仁刚. 基于主成分分析的最小二乘支持向量机短期

- 负荷预测模型[J]. 电力自动化设备, 2008, 28(11): 13-17.
- LIU Baoying, YANG Rengang. Short-term load forecasting model based on LS-SVM with PCA[J]. Electric Power Automation Equipment, 2008, 28(11): 13-17.
- [13] HE Dong, LIU Ruiye. Ultra-short-term wind power prediction using ANN ensemble based on PCA[C]//2012 7th International Power Electronics Motion Control Conference. Haerbin, China: IEEE, 2012: 2108-2112.
- [14] SEEGER M. Gaussian processes for machine learning[J]. International Journal of Neural System, 2004, 14(2): 69-106.
- [15] KERSTING K, PLAGEMANN C, PFAFF P, et al. Most likely heteroscedastic gaussian process regression[C]//Proceedings of the 24th Annual International Conference on Machine Learning. Corvallis, USA: IEEE, 2007: 393-400.
- [16] KRASKOV A, STOGBAUER H, GRASSBERGER P. Estimating mutual information[J]. Physical Review E, 2004, 69(6): 066138.
- [17] FATEMEH A, MOHAMMADMAHDI R Y, CARO L, et al. Mutual information-based feature selection for intrusion detection systems[J]. Journal of Network and Computer Applications, 2011, 34(4): 1184-1199.
- [18] 任子武, 伞冶. 自适应遗传算法的改进及在系统辨识中的应用研究[J]. 系统仿真学报, 2006, 18(1): 41-43, 66.
- REN Ziwu, SAN Ye. Improved adaptive genetic algorithm and its application research in parameter identification[J]. Journal of System Simulation, 2006, 18(1): 41-43, 66.

作者简介:



史宇伟

史宇伟(1988—),男,江苏苏州人,硕士,研究方向为风电场短期风速预测(E-mail: shiyuwei007@hotmail.com);

潘学萍(1972—),女,安徽天长人,教授,博士,研究方向为电力系统分析与控制(E-mail: xueping_pan@163.com)。

Short-term wind speed forecasting considering historical meteorological data

SHI Yuwei, PAN Xueping

(School of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China)

Abstract: A method of short-term wind speed forecasting considering the historical meteorological data is proposed. With the historical wind speed data and historical meteorological data as its original inputs, the HFS (Hybrid Feature Selection) method is applied to select the variables strongly relevant to the wind speed and to generate the input characteristic set of the forecast model. The HGP (Heteroscedastic Gaussian Process) model is then applied to build the forecast model for expressing the randomness of wind speed. A one-hour wind speed forecasting is carried out based on practical wind speed data and results show that the proposed method improves the accuracy of wind speed forecasting.

Key words: wind power; wind speed; forecasting; meteorological data; hybrid feature selection; heteroscedastic Gaussian process; models

(上接第74页 continued from page 74)

Scenario clustering method for extracting TTC operating rules of transmission corridor with wind power

LIU Ruofan¹, LIU Junyong¹, LIU Youbo¹, YANG Jiashi², GOU Jing¹, XU Lixiong¹,
JIANG Changjiang¹, CHEN Chen¹, BAZARGAN Masoud³

(1. School of Electrical Engineering and Information, Sichuan University, Chengdu 610065, China;

2. State Grid Sichuan Electric Power Company, Chengdu 610061, China;

3. ALSTOM Grid Research & Technology Centre, Stafford ST17 4LX, UK)

Abstract: A scenario clustering method for extracting TTC (Total Transfer Capability) operating rules of transmission corridor integrated with wind power is proposed based on the formulation of conventional operating rules. The scenario clustering is carried out according to the output power of wind farm and the load level and the TTC risk calculated for a certain scenario is taken as the reference of dispatch. The data mining technique is applied in the formation of rules, including sample generation, feature selection and rule extraction. The TTC rule base containing the wind power information is generated by integrating the dispatch information of all scenarios. Case simulation considering transient security is carried out for IEEE 39-bus system and results show that, the TTC dispatch information extracted by the proposed method reflects properly the relationship between power grid operation state and TTC.

Key words: wind power; TTC; scenario clustering; operating rule; data mining; risks