

基于自适应滚动匹配预测修正模式的光伏区间预测

梅飞¹, 顾佳琪¹, 裴鑫¹, 郑建勇²

(1. 河海大学 能源与电气学院, 江苏 南京 211100; 2. 东南大学 电气工程学院, 江苏 南京 210096)

摘要:为解决传统点预测方式难以量化光伏发电功率不确定性的问题以及提高预测精度,提出一种基于自适应滚动匹配预测修正模式的光伏区间预测方法。通过结合小波能量的谱聚类方法对历史光伏数据集进行分类,构建不同类别的模型输入和区间输出并采用宽度学习系统进行训练;建立不同类别、不同置信区间、不同预测功率区间的stable误差分布,并结合优化目标函数找出每个预测功率区间的最佳修正分位数点数值;利用滚动匹配预测修正模式进行区间预测。我国无锡某地的2.8 MW光伏电站算例结果表明,所提方法相较于传统的聚类预测方法具有更好的预测效果。

关键词:光伏区间预测;自适应滚动匹配预测修正模式;谱聚类;宽度学习系统;stable误差分布

中图分类号:TM 615

文献标志码:A

DOI:10.16081/j.epae.202201006

0 引言

在碳中和和碳达峰的双碳背景下,光伏能源将大规模接入电网^[1]。由于光伏功率受周围环境因素的影响,如光照强度、温度及湿度等^[2],光伏出力具有随机性与波动性。光伏功率直接接入配电系统将会影响配电系统的安全与稳定性。因此,准确进行光伏功率预测是光伏大规模并网的重要环节^[3]。

光伏预测可分为点预测^[4]和区间预测^[5]。相较于点预测,区间预测能通过不同置信区间对光伏输出功率的概率信息进行分析,量化不确定性因素对光伏预测的影响,更能反映光伏功率变化的趋势,为电网调度、光伏并网提供了有利的参考信息^[6]。

国内外学者对光伏区间预测进行了广泛而深入的研究,主要可分为上下边界理论及概率区间预测2个方面,其中上下边界理论是直接预测光伏功率的上下边界。文献[7]采用神经网络NN(Neural Network)进行点预测,并采用粒子群优化PSO(Particle Swarm Optimization)算法对隐藏层和输出层的权重进行优化调整,直接预测上下边界。文献[8]对预测模型进行改进,采用核极限学习机KELM(Kernel Extreme Learning Machine)替代神经网络模型。除了可利用PSO算法构建算法的上、下区间外,还可利用梯度下降的算法。文献[9]设立长短期记忆LSTM(Long Short Term Memory)神经网络的2个目标损失函数,采用梯度下降训练方法求解上下区间。

概率区间预测分为统计规律的区间预测^[10]、分位数回归^[11]、bootstrap^[12]等。文献[10]统计预测误差的分布情况,利用核密度估计的误差分布拟合

误差模型,构建不同置信区间下的风电预测区间模型。文献[11]采用长短期记忆分位数回归LSTMQR(Long Short Term Memory Quantile Regression)模型直接预测风速区间。

为了提高光伏预测的精度,可从学习率、训练尺度、学习算法等方面进行提升^[13],也可将数据划分为不同类别分别进行预测。文献[14]采用Adaboost改进的K近邻算法对历史数据进行分类,并分别对不同类别数据进行预测。文献[15]对历史光伏出力数据直接进行k-means聚类,并对不同的聚类分别建立预测模型进行误差分析。

对于一段未知的光伏曲线,出力时间、结束时间以及曲线类型均未知,无法直接预测非零光伏功率以及获取曲线类型。此外,传统的区间预测大多建立在点预测的基础上,预测模型相对复杂。综上,本文提出基于自适应滚动匹配预测修正模式的光伏区间预测方法,实现光伏曲线类型自适应匹配、预测和修正。首先,提出结合小波能量的谱聚类方法,对历史光伏数据进行聚类;其次,构建不同类别、不同置信区间下数据的输入和区间输出,采用宽度学习系统BLS(Broad Learning System)进行训练,并建立不同类别、不同置信区间、不同预测功率区间的stable误差分布,找出对应分位数点数值;然后,根据目标函数找出每个预测功率区间最佳修正分位数点数值;最后,采用自适应滚动匹配预测修正模式预测光伏功率,即在滚动的过程中,采用动态时间规整DTW(Dynamic Time Warping)算法进行实时聚类相似度匹配,根据匹配结果预测上下区间的功率范围,并根据功率大小进行实时修正,直到完成最终的预测。

1 结合小波能量的谱聚类

1.1 小波分解

小波包对信号的每一层分解可以滤波出高频信

收稿日期:2021-05-09;修回日期:2021-11-11

基金项目:江苏省重点研发计划资助项目(BE2020027)

Project supported by the Key R&D Program of Jiangsu Province(BE2020027)

号和低频信号。将给定的时域信号 $x(t)$ 分解为 j 层:

$$x(t) = \sum_{i=1}^{2^j} x_i(t) \quad (1)$$

式中: $x_i(t)$ 为信号分解后的第 i 个小波。

小波变换属于线性变换,经过小波变换后的能量保持不变,如式(2)所示。

$$E = \sum_{i=1}^{2^j} E_i = \sum_{i=1}^{2^j} \int_{-\infty}^{+\infty} x_i^2(t) dt \quad (2)$$

式中: E_i 为信号分解后第 i 段的信号能量。

波形的波动变化反映了能量的变化,波形波动越大,能量也就越大。附录 A 图 A1 和图 A2 分别为对晴天和阴天小波进行分解后的结果。由两图可知:晴天的光伏曲线较平稳,发电量也较高;阴天由于受云层遮挡等因素影响,光伏曲线波动较大。当光伏出力波动时,高频小波能量发生明显变化,波动越剧烈,产生的能量越大;而对于平稳类型的光伏出力,由于波动小,产生的能量也小。因此,可借助高频小波能量对平稳曲线和非平稳曲线进行划分。

由于光伏出力曲线多种多样,平稳和波动的出力很难进行界定,因此,本文借助人经验进行判断。为了避免出力大小对能量系数的影响,本文在提取高频小波能量时进行归一化处理。不同光伏曲线的高频小波能量如附录 A 图 A3 所示。经过综合判断,本文将以小波能量 0.01 作为光伏类型划分的界限。图 A4 选取了小波能量最靠近 0.01 的 2 条光伏曲线,由图 A4 可以看出,小波能量为 0.0105 的光伏曲线在前半段有些波形抖动,而小波能量为 0.0098 的光伏曲线整体比较平稳。当然,一定的波动对平稳性的预测以及聚类的精度影响不大。

1.2 结合小波能量的谱聚类

传统的聚类方法采用欧氏距离、余弦距离等进行聚类,很容易将光伏出力相当的波动和平稳类型的光伏出力划分在一起,因此,本文先利用小波能量将平稳类型的光伏出力从样本集中提取出来,再对非平稳的光伏出力进行谱聚类^[16]。同时考虑到每个月的天气因素,如温度,变化不是很大,因此,对非平稳部分按月进行聚类,并计算每个月中每个聚类的平均出力和平均高频小波能量,将其作为最终的特征量,完成最终的聚类。该过程借助底层聚类的思想,并且将高维的光伏曲线降维到 2 维数据更有助于提升聚类效果。为了选取聚类最合适的参数,本文利用 CH(Calinski-Harabaz) 指标选取最佳聚类数量和参数。平均高频小波能量和平均出力分别为:

$$E_{av,j} = \sum_{i=1}^{N_{c,j}} E_{i,j}^h \quad (3)$$

$$P_{av,j} = \frac{1}{mN_{c,j}} \sum_{i=1}^{N_{c,j}} \sum_{t=1}^m P_{t,i,j} \quad (4)$$

式中: $E_{av,j}$ 为聚类 j 的平均高频小波能量; $E_{i,j}^h$ 为聚类

j 的第 i 条曲线的高频小波能量; $N_{c,j}$ 为聚类 j 所含光伏曲线数; $P_{av,j}$ 为聚类 j 的光伏平均出力; m 为某一聚类的日内出力采样数; $P_{t,i,j}$ 为聚类 j 第 i 条曲线 t 时刻的光伏出力。

结合小波能量的谱聚类具体步骤如下:

- 1) 利用高频小波能量提取平稳的光伏曲线;
- 2) 确定光伏聚类的数量范围以及参数范围;
- 3) 采用 CH 指标确定谱聚类的最佳聚类数量以及参数,并对每个月的非平稳光伏曲线进行聚类;
- 4) 计算每个月中每个聚类的平均高频小波能量以及平均出力;
- 5) 将平均高频小波能量以及平均出力作为特征,采用 k -means 进行最终的聚类划分。

2 基于最小绝对收缩和选择算子的 BLS 预测模型

BLS 与传统深层神经网络的不同在于,它是基于随机向量函数连接神经网络 RVFLNN(Random Vector Functional-Link Neural Network) 进行构建的,具有训练参数少、求解速度快等优点^[17]。

BLS 由输入层、特征层、强化层以及输出层组成,如附录 A 图 A5 所示。特征层 Z 由 N_1 个特征窗组成,每个特征窗包含 N_2 个特征节点,用来进行特征提取。强化层 H 仅包含 1 层,由 N_3 个强节点组成。将原始输入数据 X 通过特征层进行数据提取得到一系列特征映射,第 i 组的特征映射如式(5)所示。

$$Z_i = \phi(XW_{e,i} + \beta_{e,i}) \quad i=1, 2, \dots, N_1 \quad (5)$$

式中: Z_i 为第 i 组映射特征; $W_{e,i}$ 为第 i 组特征权重; $\phi(\cdot)$ 为特征函数,本文取线性函数作为特征函数; $\beta_{e,i}$ 为第 i 组的偏置项。

数据的稀疏特性不但有利于减小新生成数据的线性相关度,而且有利于提高数据运算速度和效率以及更有效地表达数据特征。因此,本文基于最小绝对收缩和选择算子 LASSO(Least Absolute Shrinkage and Selection Operator) 对每组的权重进行稀疏重构。

$$W_{e,i} = \underset{W_{e,i}}{\operatorname{argmin}} \left(\|XW_{e,i} - Z_i\| + \lambda \|W_{e,i}\|_1 \right) \quad (6)$$

式中: $\operatorname{argmin} \left(\|XW_{e,i} - Z_i\| + \lambda \|W_{e,i}\|_1 \right)$ 表示使 $\|XW_{e,i} - Z_i\| + \lambda \|W_{e,i}\|_1$ 达到最小值时 $W_{e,i}$ 的值; λ 为正则化参数,用于控制等号右边第二项的大小。式(6)等号右边第一项用于求解线性方程 $Z_i = XW_{e,i}$ 的最小二乘法估计;第二项是 1-范数,用于对稀疏性进行限制。本文通过交叉验证的方式求解每组权重的 LASSO 回归最佳正则化参数。

将稀疏重构后的特征层所有特征节点组合成

$Z=[Z_1, Z_2, \dots, Z_{N_1}]$, 经非线性函数映射到增强层:

$$H=\zeta(ZW_h+\beta_h) \quad (7)$$

式中: W_h 为特征层和增强层相连接的权重; β_h 为特征层和增强层相连接的偏置项; $\zeta(\cdot)$ 为非线性的映射函数, 本文采用双曲正切函数作为映射函数。

最后将特征层和增强层经过权值 W_y 与输出层相连接, 如式(8)所示。

$$Y=[Z|H]W_y=AW_y \quad (8)$$

在BLS网络训练的过程中, $W_{e,i}$ 、 $\beta_{e,i}$ 、 W_h 、 β_h 是随机产生的, 并且在训练的过程中保持不变, 因此, 只需求得 W_y 即可, 如式(9)所示。

$$W_y=\operatorname{argmin}_{W_y} \|AW_y-Y\|=A^+Y \quad (9)$$

式中: A^+ 为矩阵 A 的 Moore-Penrose 广义逆矩阵。

3 自适应滚动匹配预测修正模型

自适应滚动匹配预测修正模型分为自适应滚动匹配模型以及预测修正模型。

3.1 自适应滚动匹配模型

对于一段未知的光伏曲线, 由于无法知晓实际的光伏曲线类型, 因此无法采用合适的模型进行预测。本文以DTW算法^[18]对未知光伏类型和已知光伏类型进行自适应滚动匹配。具体的匹配原则为: 计算待匹配的光伏序列和已知簇的每条光伏曲线的DTW距离, 以平均值最小的原则来判断最为相似的聚类簇, 如式(10)所示。

$$D_{ma,j}=\frac{1}{N_{e,j}}\sum_{i=1}^{N_{e,j}}D_{DTW}(P_n^{un}, P_{n+1,i}) \quad (10)$$

式中: $D_{ma,j}$ 为未知光伏曲线与聚类 j 中的光伏曲线平均DTW距离; P_n^{un} 为未知光伏曲线到 n 时刻出力的时间序列; $P_{n+1,i}$ 为第 i 条已知光伏曲线到 $n+1$ 时刻出力的时间序列; $D_{DTW}(\cdot, \cdot)$ 表示求DTW距离。

自适应滚动匹配模型可以表述为: 时间序列 y_1, y_2, \dots, y_n 为某日前 n 个功率实测数据, 利用DTW算法将该时间序列和不同聚类类别进行相似性匹配, 找出最相似的聚类类别, 在预测下一时刻出力时, 采用该相似簇的预测模型进行预测。到达 $n+1$ 时刻后, 将该时刻实测数据 y_{n+1} 加入原始序列中, 形成新序列 $y_1, y_2, \dots, y_n, y_{n+1}$ 。采用DTW对新序列重新进行相似度匹配, 匹配出与该时间序列最相似的聚类类别, 在 $n+2$ 时刻采用该聚类类别进行预测。在 $n+2$ 时刻预测时需剔除最远的功率点, 以保证数据等维输入, 即采用序列 y_2, y_3, \dots, y_{n+1} 进行预测。按这种滚动匹配预测模式, 完成一整天的发电功率预测。

3.2 预测修正模型

区间预测修正模型是建立在区间预测和误差修正的基础上的。本文预测模型的输出区间数据直接根据置信区间 α 进行构造。光伏输入为预测数据

前 n 个时刻功率数据, 输出数据为 $n+1$ 时刻数据的 $(1-\alpha, 1+\alpha)$, 如式(11)所示。

$$X=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & & \vdots \\ x_N & x_{N+1} & \cdots & x_{N+n-1} \end{bmatrix} \quad Y=(1-\alpha, 1+\alpha)\begin{bmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_{n+N} \end{bmatrix}=\begin{bmatrix} L_1, U_1 \\ L_2, U_2 \\ \vdots \\ L_N, U_N \end{bmatrix} \quad (11)$$

式中: $L_i, U_i (i=1, 2, \dots, N)$ 分别为第 i 个样本下边界值、上边界值, 且有 $L_i=x_{n+i}(1-\alpha)$ 、 $U_i=x_{n+i}(1+\alpha)$, N 为样本数量。

本文采用误差修正的方式修正预测模型, 尽量使预测的功率落入区间内且更偏向于区间中间。

$$\begin{cases} Y_{errup}=Y_{preup}-Y(1+\alpha) \\ Y_{errlow}=Y(1-\alpha)-Y_{prelow} \end{cases} \quad (12)$$

式中: Y_{errup} 、 Y_{errlow} 分别为上、下边界的预测误差; Y_{preup} 、 Y_{prelow} 分别为上、下边界的预测值。

本文以功率间隔划分预测功率区间, 如功率间隔为 0.1 MW, 则将预测功率区间划分为 $[0, 0.1)$ 、 $[0.1, 0.2)$ MW 等区间。为了精细化预测, 分别对预测的上边界以及下边界进行预测功率区间划分, 上、下边界的预测功率区间最终范围取决于上、下边界的预测最大值所在范围。建立不同类别、不同置信区间以及不同上、下边界预测功率区间的误差分布, 确定其分位数功率点, 根据修正优化目标函数确定每个预测区间的最优修正分位数点功率。在本文中, 这些分位点分别为 5% 、 10% 、 \dots 、 95% 。

3.3 区间预测指标及修正优化目标函数

光伏功率的区间准则从可靠性和清晰度2个方面确定。可靠性用实际观测值落入预测区间的概率表示, 该值应尽量大; 清晰度用预测区间宽度表示, 该值应尽量小。本文用3个指标衡量预测的准确性。

1) 预测区间覆盖率PICP(Prediction Interval Coverage Probability)指标。预测区间覆盖率用来评估预测区间的可靠性, 该指标反映实际值落入区间的概率, 如式(13)所示。预测区间覆盖率越大表示越多的实际值落入区间中, 预测效果越好。

$$e_{PICP}=\frac{1}{N}\sum_{i=1}^N\kappa_i \quad (13)$$

式中: e_{PICP} 为预测区间覆盖率; κ_i 为二进制变量, 若实际功率包含于所构建的预测区间内, 则其值为1, 否则为0。

2) 预测区间平均带宽PINAW(Prediction Interval Normalized Average Width)指标。预测区间平均带宽反映了预测清晰度, 如式(14)所示。区间宽度越窄, 区间覆盖率越高, 区间预测效果越好。

$$e_{\text{PINAW}} = \frac{1}{N} \sum_{i=1}^N \frac{U_i - L_i}{y_{\max}} \quad (14)$$

式中: e_{PINAW} 为预测区间平均带宽; y_{\max} 为预测样本中的最大值。

3) 带宽平均偏差 NAD (Normalized Average Deviation) 指标。带宽平均偏差用来评估目标值远离预测区间的程度, 如式(15)所示。带宽平均偏差越小表示未落点离预测区间越近。

$$e_{\text{NAD}} = \frac{1}{N} \sum_{i=1}^N \gamma_i \quad (15)$$

$$\gamma_i = \begin{cases} \frac{L_i - y_i}{\frac{1}{N} \sum_{i=1}^N (U_i - L_i)} & y_i < L_i \\ 0 & y_i \in [L_i, U_i] \\ \frac{y_i - U_i}{\frac{1}{N} \sum_{i=1}^N (U_i - L_i)} & y_i > U_i \end{cases} \quad (16)$$

4) 修正优化目标函数。通过修正优化目标函数来寻找最优的分位数点修正数值。

$$F_{\text{obj}} = (e_{\text{PICP2}} - e_{\text{PICP1}}) / e_{\text{PICP1}} + (e_{\text{PINAW1}} - e_{\text{PINAW2}}) / e_{\text{PINAW1}} + (e_{\text{NAD1}} - e_{\text{NAD2}}) / e_{\text{NAD1}} \quad (17)$$

式中: e_{PICP1} 、 e_{PINAW1} 、 e_{NAD1} 为修正之前的指标; e_{PICP2} 、 e_{PINAW2} 、 e_{NAD2} 为修正之后的指标。

3.4 自适应滚动匹配预测修正模型具体流程

基于自适应滚动匹配修正模式的光伏区间预测方法的流程图如附录 A 图 A6 所示, 具体流程如下。

1) 采用所提聚类方法对历史光伏数据进行聚类。

2) 挑选出每个聚类的 80% 数据作为训练数据, 每个聚类的其余 20% 数据作为测试数据。

3) 对每个聚类的不同置信区间的训练数据进行 BLS 训练, 并确定 BLS 的最优训练参数。

4) 建立不同聚类、不同置信区间以及不同上、下边界预测功率区间的 stable 误差分布, 并通过修正优化目标函数确定每个预测功率区间最优误差修正数值。

5) 采用自适应滚动修正模式对未知光伏功率进行滚动区间预测。用 DTW 滚动匹配最佳的聚类类型, 用 BLS 滚动预测光伏区间, 修正数值根据预测的数值进行上、下边界修正, 直到完成一天的预测。

4 算例分析

为了验证本文所提方法的有效性, 选取我国无锡某 2.8 MW 的光伏电站所测得的全年光伏功率作为数据基础。对 365 d 的光伏曲线剔除不良与缺失的光伏数据, 共得到 345 组数据。每 15 min 取一个采样点, 每天共 96 个点。

4.1 最优滚动步长的确定

不同的滚动步长对区间预测精度有所影响。为

探寻最优滚动步长, 本文计算不同步长下不同置信区间的 3 个指标数值, 并取它们的平均值进行比较, 综合评断最优滚动步长, 结果见附录 A 图 A7。由图可知, 不同步长的预测区间平均带宽相差不大于 0.01, 但预测区间覆盖率约相差 0.15, 带宽平均偏差相差 0.7 左右。相较于其他步长, 步长 3 的预测区间覆盖率平均值最高, 带宽平均偏差平均值最低, 预测区间平均带宽平均值与最优的预测区间平均带宽相差 0.0075, 但整体差距不大。因此, 步长为 3 时指标总体最优, 选取 3 作为最优滚动步长。结合 15 min 的采样时间, 本文以预测前 15、30、45 min 的实测数据作为输入, 开始滚动时刻为 00:15, 采用 00:15、00:30、00:45 时刻的实测数据, 开始预测时刻为 01:00。

4.2 聚类方法的有效性

为验证所提聚类方法的有效性, 将其与 k -means 聚类、未结合小波能量的谱聚类以及无需指定聚类数量的近邻传播聚类 AP (Affinity Propagation) 聚类进行比较。传统的 k -means 聚类和谱聚类采用 CH 指标获得最佳的聚类数量, 并相应确定各聚类的 BLS 最优参数。 k -means 聚类聚成 2 类, 谱聚类聚成 2 类, AP 聚类聚成 9 类, 结合小波能量的谱聚类连同平稳部分一共聚成 4 类。采用 BLS 进行滚动预测, 效果见附录 A 表 A1。由表可知: AP 聚类即使是在带宽较大时, 整体的预测区间覆盖率也较低; 采用 k -means 聚类的预测区间覆盖率略低于所提方法, 但预测区间平均带宽却大于所提方法, 这说明 k -means 聚类牺牲了预测区间平均带宽来提高预测区间覆盖率; 未结合小波能量的谱聚类预测区间平均带宽略大于所提方法, 但预测区间覆盖率平均值较低。总体而言, 结合小波能量的谱聚类的预测区间覆盖率明显高于传统方法, 预测区间平均带宽也明显大于传统方法。

附录 A 图 A8 为 50% 的置信区间下不同聚类方法的光伏曲线预测图, 图 A9 为在滚动预测过程中采用 DTW 算法匹配得到的聚类类型。本文所提聚类方法将光伏曲线分成 4 类, 在刚开始光伏出力为 0 时, 匹配类型默认为 1, 随着光伏开始出力, 由于整体出力情况不清晰, 匹配得到的类型也可能是近似的类型, 有 2 个时刻判断为类型 2, 同时在光伏出力较小时有 4 个出力判断为类型 1, 当光伏轮廓比较清晰时, 整体的类型判断为最终的类型 3。其他聚类方法未能合理化聚类, 导致类型判断的过程中出现反复判断的现象: AP 聚类由于聚成的数量较多, 判断的过程中更换的模型也较多, 导致整体的预测效果较差; 聚成 2 类的谱聚类在进行模型更替的 2 个时刻出现了预测的偏差; k -means 预测效果和所提方法较相似, 但聚成 2 类导致得到的整体区间宽度要大。

4.3 基于 BLS 的区间预测

为验证本文所采用的基于 LASSO 的 BLS 宽度学习算法的有效性, 将其与传统的 BP 神经网络、深度

学习的LSTM神经网络、无需进行梯度训练的KELM进行比较。训练参数如附录A表A2所示,预测结果如表A3所示。由表A3可知,不管是哪种模型,置信区间的增大均使得预测区间平均带宽增大,使落入区间内的实际数值点增多;BP神经网络模型的预测区间平均带宽平均值最大,预测区间覆盖率平均值较低,带宽平均偏差平均值较高;BLS模型的区间平均带宽平均值小于KELM模型和LSTM模型,但预测区间覆盖率平均值却高于LSTM模型和KELM模型0.1左右。选取表A3区间平均带宽较接近的10%~30%的置信区间的预测区间覆盖率情况进行比较,如表1所示。由表可知,BLS模型在低置信区间下的预测精度比其他模型要高出0.1~0.2。

表1 低置信区间下不同预测模型的预测区间覆盖率
Table 1 PICIP of different prediction models under low confidence interval

α	e_{PICP}			
	BP	LSTM	KELM	BLS
0.05	0.124	0.209	0.287	0.412
0.10	0.357	0.363	0.335	0.510
0.15	0.516	0.479	0.473	0.560

附录A图A10为不同方法下不同置信区间的光伏曲线。由图可知:当采用BLS模型预测光伏出力时,各区间有很好的层次划分,所预测的曲线不仅很好地落入区间内,而且更偏于区间中间,即使在10%的置信区间内也有很好的落入效果;当采用BP神经网络模型和KELM模型时,10%的置信区间偏离了预测曲线,在20%~30%的置信区间时曲线才能落入其中;虽然LSTM模型和BLS模型的置信区间层次划分效果大体一致,但在高置信区间时,有些预测点依旧不能落入区间内;当采用BP神经网络模型进行零点预测时,上边界的预测值偏大,整体的区间带宽较大,其他模型的零点预测效果均优于BP神经网络模型。

4.4 误差修正

本文以修正误差方式在缩小带宽的同时进一步提高预测区间覆盖率。本文以0.1 MW间隔进行划分。stable分布是高斯分布的一般化,其具有所有稳定分布的特性,因此本文选用stable分布进行误差拟合。以下边界预测功率为[0.4,0.5) MW时的预测误差为例,附录A图A11为其概率分布图。由图可以看出,stable分布进行误差拟合的效果明显优于其他方法,尤其是对于高功率时的误差以及误差为0时。

在不同类别、不同置信区间、不同功率预测区间的最优分位数点修正功率确定后,在进行滚动预测的过程中,根据所预测的上、下区间的功率大小进行合适的修正功率补偿。经误差修正后的预测区间覆盖率、预测区间平均带宽以及带宽平均偏差见表2。

由表2可知,相较于修正前,虽然修正后的预测

表2 修正前、后的预测精度

Table 2 Prediction accuracy before and after correction

α	修正前			修正后		
	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}
0.05	0.412	0.018	2.846	0.564	0.036	0.905
0.10	0.510	0.042	1.193	0.644	0.068	0.482
0.15	0.560	0.064	0.720	0.749	0.072	0.238
0.20	0.604	0.085	0.488	0.793	0.088	0.218
0.25	0.715	0.106	0.251	0.817	0.109	0.158
0.30	0.751	0.127	0.187	0.843	0.111	0.132
0.35	0.794	0.148	0.145	0.860	0.137	0.086
0.40	0.882	0.170	0.083	0.900	0.151	0.042
0.45	0.897	0.184	0.055	0.931	0.165	0.029
平均值	0.681	0.105	0.663	0.790	0.104	0.254

区间平均带宽平均值只降低了0.001,但是预测区间覆盖率平均值约提高了0.1,带宽平均偏差平均值也降低了0.409。修正后的整体效果有了明显提升。当置信区间低于50%($\alpha < 0.25$)时,预测上边界和下边界的区间偏小,导致整体的区间落入率并不高,通过增大区间宽度可以提高区间的落入率,从而减小小区间偏差。以修正20%($\alpha = 0.10$)的置信区间为例,修正后的预测区间覆盖率为0.644,明显大于40%的置信区间,预测区间平均带宽增加,但低于40%的置信区间,未落入点的带宽平均偏差也低于40%的置信区间。图1为20%的置信区间下修正前、后的光伏区间预测图。由图可知,修正后的区间带宽明显增大,因此,落入区间的点也随之增多,未落入区间的点更接近区间的上、下边界。

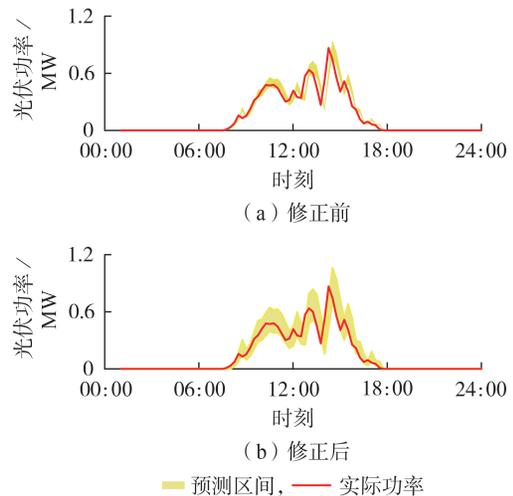


图1 20%的置信区间下修正前、后的光伏区间预测

Fig.1 Photovoltaic interval prediction before and after correction under 20% confidence interval

当置信区间高于50%($\alpha > 0.25$)时,由于上、下带宽较大,因此,落入区间的实际功率点较多,预测区间覆盖率较高,带宽平均偏差较小。通过对误差大小进行修正,减小高功率的区间宽度,增大低功率的区间宽度,提高预测精度。以置信区间为90%为例,修正后的预测区间覆盖率为0.931,比修正前的高,而

修正后的区间带宽为0.165,比修正前的窄。附录A图A12为90%的置信区间下修正前、后的预测图。由图可知:在出力较大的点减小上、下功率区间的出力,减小了整体的区间带宽;在出力较小的点增加出力,提高了小功率区间下的落入率。因此,修正后整体的区间带宽有了明显缩减,落入率有所提高。

4.5 区间预测方法比较

为了验证本文所提出的实时预测修正方法的有效性,将所提方法和传统的基于正态分布估计的概率区间预测和LSTMQR进行对比,结果如附录A表A4所示。由表可以看出:相较于另外2种方法,本文所提方法在更小的区间带宽下有更高的区间预测精度以及更低的区间偏离程度;相较于基于正态分布估计的概率区间预测和LSTMQR,本文所提方法的预测区间平均带宽平均值分别减小了0.031和0.016,但预测区间覆盖率平均值却分别提高了0.128和0.082;基于正态分布估计的概率区间预测效果最差,整体的区间预测精度最低。

图2为3种方法在90%的置信区间下的效果图。由图可知:基于正态分布估计的概率区间预测虽然在低功率下有很好的预测效果,但却牺牲了带宽,在高功率的带宽下预测效果较差;LSTMQR虽然在高功率下有很好的预测效果,曲线基本落入区间内,但上、下带宽较大;本文方法能在减小功率区间带宽的同时,使更多的点落入预测的上、下带宽内。因此,本文方法比传统方法更具精确性与合理性。

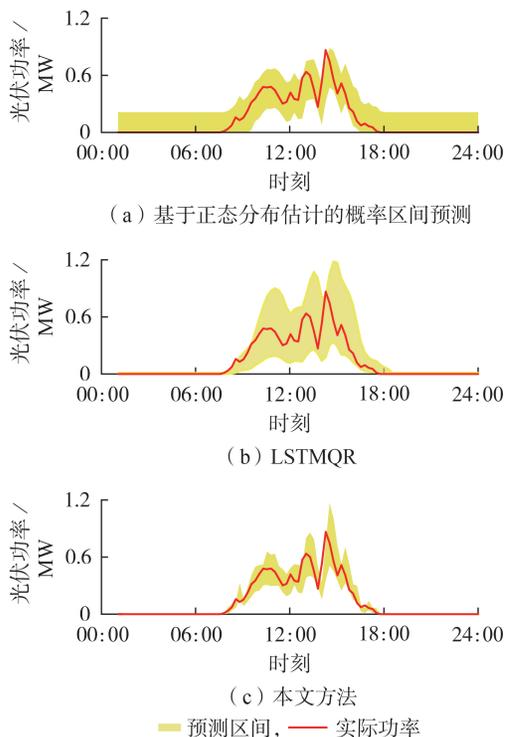


图2 90%的置信区间下不同预测方法对比

Fig.2 Comparison of different prediction methods under 90% confidence interval

5 结论

由于光伏出力的不确定性,光伏概率预测已经成为光伏功率预测的趋势。本文提出一种新颖的光伏区间预测方法,着眼于数据本身,直接构造不同置信区间下的数据集,集聚类、相似度匹配、预测、修正为一体实现对未知光伏曲线的预测,以预测区间覆盖率、预测区间平均带宽、带宽平均偏差3个区间预测指标衡量所提方法的有效性。

1) 本文将区间预测问题简单化,不再基于传统点预测,而是直接构建预测区间,对模型进行训练和测试。

2) 本文采用自适应滚动算法,不仅进行光伏区间的自适应滚动预测和修正,而且提供非等长时间序列类型匹配方法。

3) 本文对聚类方法、预测模型以及最终的效果进行了大量的对比实验。在聚类方面,将AP聚类、 k -means聚类、谱聚类以及本文的聚类方法进行对比。在模型测试方面,将BLS宽度神经网络与BP神经网络、LSTM、KELM进行比较。在最终预测效果方面,与概率预测以及LSTMQR进行对比。结果表明,本文所提方法在3个方面均在更小的预测区间下有更高的预测区间覆盖率以及更低的区间带宽平均偏差。本文所提方法能够用于对未知光伏曲线的测试,且有更高的预测精度。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 李升,卫志农,孙国强,等. 大规模光伏发电并网系统电压稳定分岔研究[J]. 电力自动化设备,2016,36(1):17-23.
LI Sheng, WEI Zhinong, SUN Guoqiang, et al. Voltage stability bifurcation of large-scale grid-connected PV system[J]. Electric Power Automation Equipment, 2016, 36(1): 17-23.
- [2] 姚宏民,杜欣慧,李廷钧,等. 光伏高渗透率下配网消纳能力模拟及电压控制策略研究[J]. 电网技术,2019,43(2):462-469.
YAO Hongmin, DU Xinhui, LI Tingjun, et al. Simulation of consumption capacity and voltage control strategy of distribution network with high penetration of photovoltaics[J]. Power System Technology, 2019, 43(2): 462-469.
- [3] NI Q, ZHUANG S X, SHENG H M, et al. An ensemble prediction intervals approach for short-term PV power forecasting[J]. Solar Energy, 2017, 155: 1072-1083.
- [4] 李乐,刘天琪. 基于近邻传播聚类和回声状态网络的光伏预测[J]. 电力自动化设备,2016,36(7):41-46.
LI Le, LIU Tianqi. PV power forecasting based on AP-ESN[J]. Electric Power Automation Equipment, 2016, 36(7): 41-46.
- [5] 杨锡运,张璜,关文渊,等. 基于滑动分块百分位数Bootstrap法的风电功率概率区间预测[J]. 太阳能学报,2019,40(2):430-437.
YANG Xiyun, ZHANG Huang, GUAN Wenyuan, et al. Probabilistic intervals forecasting of wind power based on moving block percentile Bootstrap method[J]. Acta Energiæ Solaris Sinica, 2019, 40(2): 430-437.
- [6] 李燕青,杜莹莹. 基于双维度顺序填补框架与改进Kohonen天

- 气聚类的光伏发电短期预测[J]. 电力自动化设备, 2019, 39(1): 60-65.
- LI Yanqing, DU Yingying. Short-term photovoltaic power forecasting based on double-dimensional sequential imputation framework and improved Kohonen clustering[J]. Electric Power Automation Equipment, 2019, 39(1): 60-65.
- [7] KHOSRAVI A, NAHAVANDI S, CREIGHTON D, et al. Lower upper bound estimation method for construction of neural network-based prediction intervals[J]. IEEE Transactions on Neural Networks, 2011, 22(3): 337-346.
- [8] 杨锡运, 关文渊, 刘玉奇, 等. 基于粒子群优化的核极限学习机模型的风电功率区间预测方法[J]. 中国电机工程学报, 2015, 35(增刊1): 146-153.
- YANG Xiyun, GUAN Wenyuan, LIU Yuqi, et al. Prediction intervals forecasts of wind power based on PSO-KELM[J]. Proceedings of the CSEE, 2015, 35(Supplement 1): 146-153.
- [9] LI C S, TANG G, XUE X M, et al. The short-term interval prediction of wind power using the deep learning model with gradient descend optimization[J]. Renewable Energy, 2020, 155: 197-211.
- [10] 刘立阳, 孟绍良, 吴军基. 基于风电预测误差区间的动态经济调度[J]. 电力自动化设备, 2016, 36(9): 87-93.
- LIU Liyang, MENG Shaoliang, WU Junji. Dynamic economic dispatch based on wind power forecast error interval[J]. Electric Power Automation Equipment, 2016, 36(9): 87-93.
- [11] 李彬, 彭曙蓉, 彭君哲, 等. 基于深度学习分位数回归模型的风电功率概率密度预测[J]. 电力自动化设备, 2018, 38(9): 15-20.
- LI Bin, PENG Shurong, PENG Junzhe, et al. Wind power probability density forecasting based on deep learning quantile regression model[J]. Electric Power Automation Equipment, 2018, 38(9): 15-20.
- [12] CHEN X, DONG Z Y, MENG K, et al. Electricity price forecasting with extreme learning machine and bootstrapping[J]. IEEE Transactions on Power Systems, 2012, 27(4): 2055-2062.
- [13] ZHANG W J, QUAN H, SRINIVASAN D. An improved quantile regression neural network for probabilistic load forecasting[J]. IEEE Transactions on Smart Grid, 2019, 10(4): 4425-4434.
- [14] 谭津, 邓长虹, 杨威, 等. 微电网光伏发电的Adaboost天气聚类超短期预测方法[J]. 电力系统自动化, 2017, 41(21): 33-39.
- TAN Jin, DENG Changhong, YANG Wei, et al. Ultra-short-term photovoltaic power forecasting in microgrid based on Adaboost clustering[J]. Automation of Electric Power Systems, 2017, 41(21): 33-39.
- [15] BAE K Y, JANG H S, SUNG D K. Hourly solar irradiance prediction based on support vector machine and its error analysis[J]. IEEE Transactions on Power Systems, 2017, 32(2): 935-945.
- [16] 徐杉杉, 朱俊澎, 袁越. 基于CEEMD-SE-MM的中长期风速模拟方法[J]. 电力自动化设备, 2020, 40(2): 69-76.
- XU Shanshan, ZHU Junpeng, YUAN Yue. Mid- and long-term wind speed simulation method based on CEEMD-SE-MM[J]. Electric Power Automation Equipment, 2020, 40(2): 69-76.
- [17] 周楠, 徐潇源, 严正, 等. 基于宽度学习系统的光伏发电功率超短期预测[J]. 电力系统自动化, 2021, 45(1): 55-64.
- ZHOU Nan, XU Xiaoyuan, YAN Zheng, et al. Ultra-short-term forecasting of photovoltaic power generation based on broad learning system[J]. Automation of Electric Power Systems, 2021, 45(1): 55-64.
- [18] 万书亭, 马晓棣, 陈磊, 等. 基于振动信号短时能熵比与DTW的高压断路器状态评估及故障诊断[J]. 高电压技术, 2020, 46(12): 4249-4257.
- WAN Shuting, MA Xiaodi, CHEN Lei, et al. State evaluation and fault diagnosis of high-voltage circuit breaker based on short-time energy entropy ratio of vibration signal and DTW[J]. High Voltage Engineering, 2020, 46(12): 4249-4257.

作者简介:



梅飞

梅飞(1982—),男,安徽滁州人,讲师,博士,主要研究方向为智能电网与分布式能源(E-mail:meifei@hhu.edu.cn);

顾佳琪(1996—),男,江苏南通人,硕士研究生,主要研究方向为综合能源系统、光伏预测(E-mail:gujiaqi@hhu.edu.cn)。

(编辑 王锦秀)

Photovoltaic interval prediction based on adaptive rolling matching prediction correction mode

MEI Fei¹, GU Jiaqi¹, PEI Xin¹, ZHENG Jianyong²

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China;

2. School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: In order to solve the problem that the traditional point prediction method is difficult to quantify the uncertainty of photovoltaic generation power and improve the prediction accuracy, a photovoltaic interval prediction method based on adaptive rolling matching prediction correction mode is proposed. The historical photovoltaic data set is clustered by the spectral clustering combined with wavelet energy, and input and interval output of different clusters are constructed and trained by broad learning system. The stable error distribution of different clusters, different confidence intervals, and different prediction power intervals is established, and the optimal modified quantile point value for each prediction power interval is found combined with the optimization objective function. The rolling matching prediction correction mode is used for interval prediction. The case results of a 2.8 MW photovoltaic power station in a region of Wuxi, China show that the proposed method has better prediction effect compared with the traditional clustering prediction method.

Key words: photovoltaic interval prediction; adaptive rolling matching prediction correction mode; spectral clustering; broad learning system; stable error distribution

附录 A:

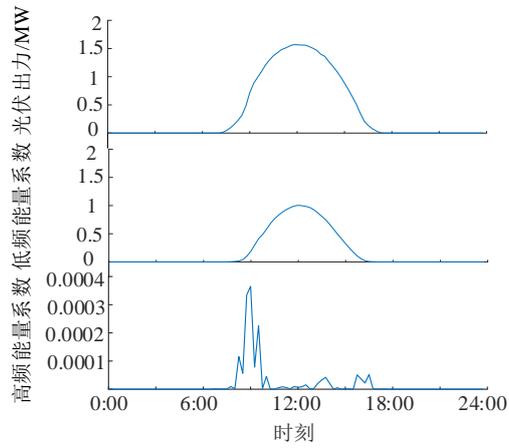


图 A1 晴天出力的高低频小波能量
Fig.A1 High and low frequency wavelet energy on sunny day

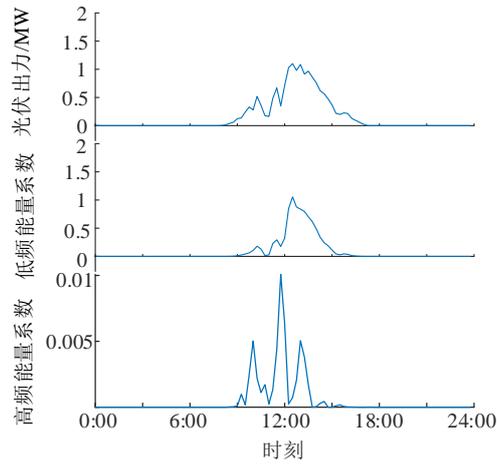


图 A2 阴天出力的高低频小波能量
Fig.A2 High and low frequency wavelet energy on cloudy day

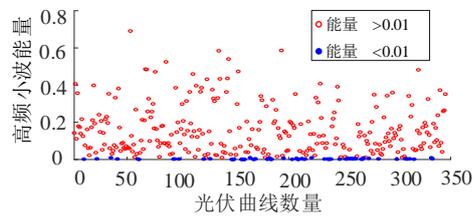


图 A3 不同光伏曲线的高频小波能量
Fig.A3 High frequency wavelet energy of different photovoltaic curves

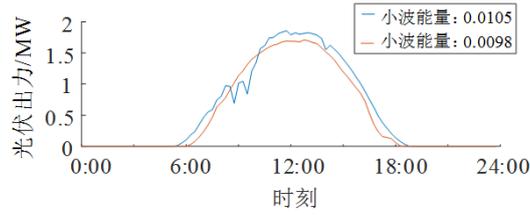


图 A4 不同小波能量下的光伏曲线

Fig.A4 Photovoltaic curves under different wavelet energies

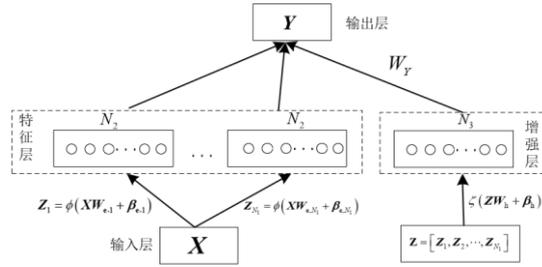


图 A5 BLS 结构图

Fig.A5 Structure diagram of BLS

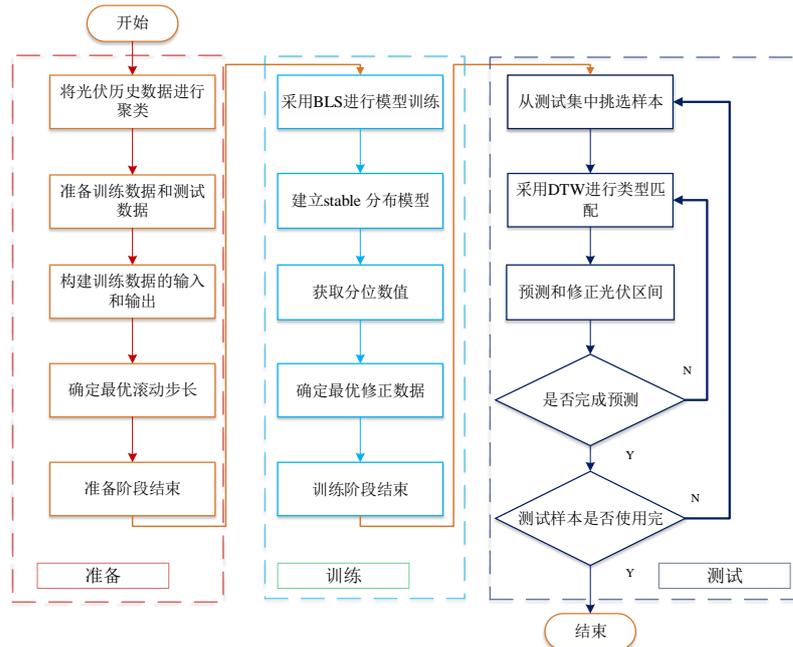


图 A6 基于自适应滚动匹配预测修正模式的光伏区间预测流程图

Fig.A6 Flowchart of photovoltaic interval prediction based on adaptive rolling matching prediction correction mode

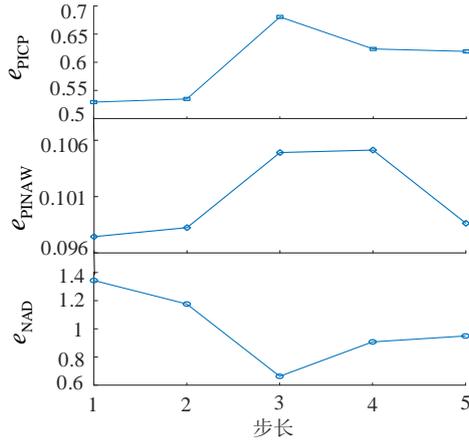
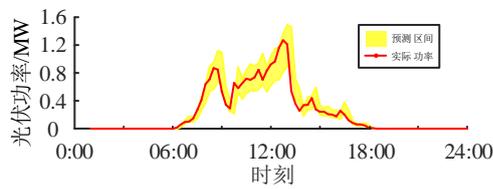


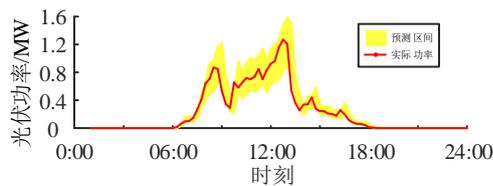
图 A7 不同步长下的平均指标
Fig.A7 Average indexes under different steps

表 A1 不同聚类方法的准确性比较
Table A1 Accuracy comparison among different clustering methods

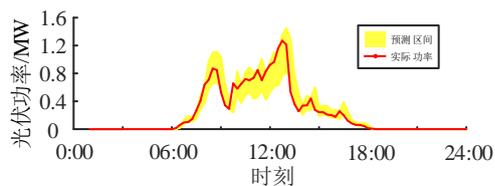
α	AP 聚类			k -means 聚类			谱聚类			所提聚类方法		
	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}
0.05	0.389	0.020	4.142	0.403	0.019	3.449	0.077	0.022	6.392	0.412	0.018	2.846
0.10	0.423	0.050	1.961	0.469	0.047	1.463	0.340	0.043	2.972	0.510	0.042	1.193
0.15	0.466	0.078	1.190	0.517	0.071	0.795	0.415	0.065	1.843	0.560	0.064	0.720
0.20	0.520	0.089	0.798	0.571	0.092	0.510	0.512	0.086	1.285	0.604	0.085	0.488
0.25	0.541	0.116	0.608	0.712	0.105	0.259	0.658	0.108	0.782	0.715	0.106	0.251
0.30	0.646	0.145	0.391	0.745	0.127	0.192	0.698	0.130	0.129	0.751	0.127	0.187
0.35	0.664	0.191	0.322	0.777	0.171	0.101	0.737	0.152	0.107	0.794	0.148	0.145
0.40	0.683	0.212	0.271	0.882	0.196	0.082	0.779	0.173	0.092	0.882	0.170	0.083
0.45	0.700	0.223	0.232	0.893	0.220	0.068	0.813	0.194	0.081	0.897	0.184	0.055
平均值	0.559	0.125	1.102	0.663	0.116	0.769	0.558	0.108	1.520	0.681	0.105	0.663



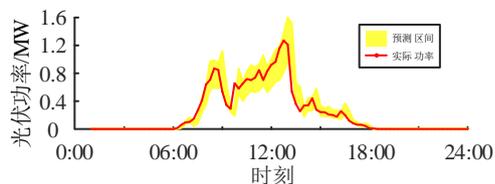
(a) 所提方法



(b) k -means



(c) 谱聚类



(d) AP 聚类

图 A8 50%的置信区间下不同聚类方法的光伏曲线预测图

Fig.A8 Photovoltaic prediction curves of different clustering methods under 50% confidence interval

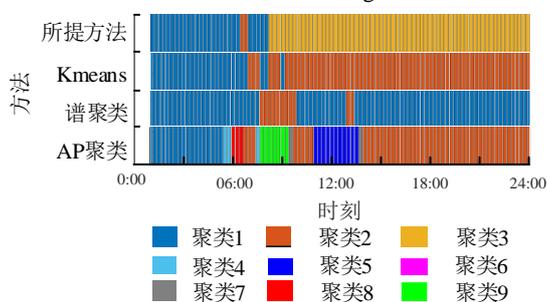


图 A9 不同聚类方法的模型匹配图

Fig.A9 Model matching diagram of different clustering methods

表 A2 不同预测模型的训练参数

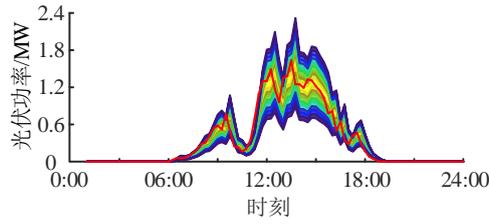
Table A2 Training parameters of different prediction models

模型	参数	聚类 1	聚类 2	聚类 3	聚类 4
BPNN	隐藏层 1 隐藏节点数	17	12	16	17
	隐藏层 2 隐藏节点数	10	5	3	9
	训练次数	500			
	激活函数	tansig, logsig			
	学习率	0.1			
LSTM	隐藏层隐藏节点数	15	10	12	14
	初始学习率	0.005			
	训练 50 次后学习率	0.2			
	求解器	Adam			
KELM	惩罚因子	0.5	0.5	1	1
	核参数	0.5	0.5	0.5	0.5
	隐藏层节点数	500			
BLS	特征窗	4	6	4	4
	特征节点数	2	13	17	9
	强节点数	17	27	29	12

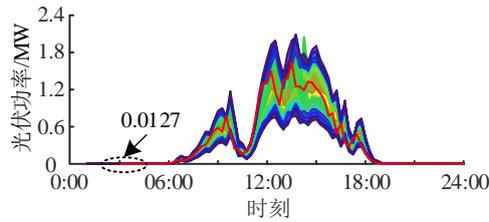
表 A3 不同预测方法下的精确度对比

Table A3 Accuracy comparison among different prediction methods

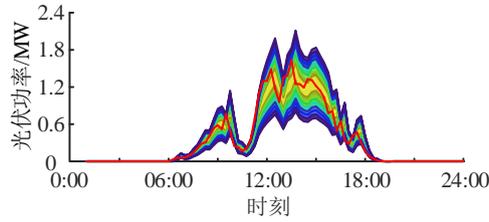
α	BP			LSTM			KELM			BLS		
	e_{PICP}	e_{PINAW}	e_{NAD}									
0.05	0.124	0.019	3.051	0.209	0.020	3.244	0.287	0.021	3.303	0.412	0.018	2.846
0.10	0.357	0.050	1.716	0.363	0.043	1.494	0.335	0.051	1.481	0.510	0.042	1.193
0.15	0.516	0.067	1.567	0.479	0.068	0.949	0.473	0.062	0.972	0.560	0.064	0.720
0.20	0.557	0.082	0.683	0.532	0.088	0.653	0.572	0.089	0.572	0.604	0.085	0.488
0.25	0.591	0.126	0.419	0.572	0.109	0.480	0.610	0.104	0.417	0.715	0.106	0.251
0.30	0.649	0.138	0.282	0.686	0.137	0.275	0.650	0.125	0.319	0.751	0.127	0.187
0.35	0.691	0.155	0.169	0.725	0.156	0.230	0.680	0.145	0.265	0.794	0.148	0.145
0.40	0.724	0.174	0.144	0.834	0.177	0.152	0.776	0.165	0.179	0.882	0.170	0.083
0.45	0.854	0.203	0.123	0.851	0.185	0.119	0.836	0.187	0.128	0.897	0.184	0.055
平均值	0.563	0.113	0.906	0.584	0.109	0.804	0.580	0.105	0.848	0.681	0.105	0.663



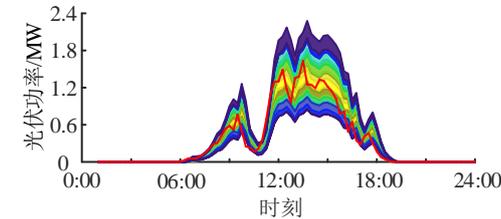
(a) BLS



(b) BPNN



(c) KELM



(d) LSTM

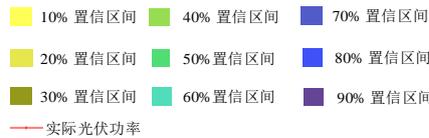
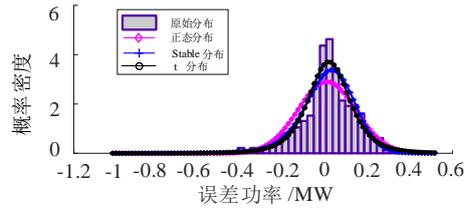
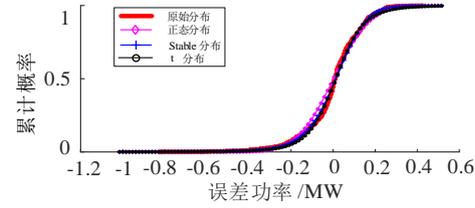


图 A10 不同预测模型下不同置信区间的光伏区间预测

Fig.A10 Photovoltaic interval prediction of different confidence intervals under different prediction models

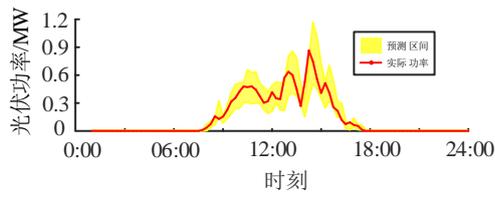


(a) 概率密度

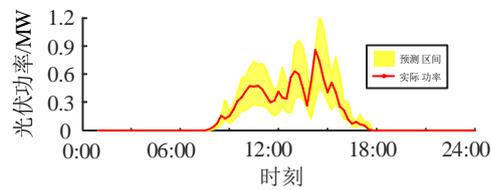


(b) 累计概率

图 A11 下边界预测为 $[0.4, 0.5)$ MW 时的预测误差
Fig.A11 Prediction error when lower boundary is $[0.4, 0.5)$ MW



(a) 修正后



(b) 修正前

图 A12 在 90% 的置信区间下修正前、后的光伏区间预测
Fig.A12 Photovoltaic interval prediction before and after correction under 90% confidence interval

表 A4 不同区间预测方法精确度对比

Table A4 Accuracy comparison of prediction methods under different confidence intervals

α	基于正态分布估计的概率区间预测			LSTMQR			所提方法		
	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}	e_{PICP}	e_{PINAW}	e_{NAD}
0.05	0.425	0.018	2.346	0.440	0.027	1.115	0.564	0.036	0.905
0.10	0.524	0.050	2.016	0.604	0.069	0.601	0.644	0.068	0.482
0.15	0.593	0.052	1.709	0.636	0.072	0.570	0.749	0.072	0.238
0.20	0.625	0.061	1.252	0.684	0.098	0.450	0.793	0.088	0.218
0.25	0.671	0.091	0.776	0.711	0.110	0.150	0.817	0.109	0.158
0.30	0.706	0.121	0.543	0.741	0.121	0.146	0.843	0.111	0.132
0.35	0.748	0.172	0.351	0.807	0.160	0.098	0.860	0.137	0.086
0.40	0.807	0.258	0.189	0.856	0.184	0.047	0.900	0.151	0.042
0.45	0.856	0.394	0.088	0.895	0.241	0.027	0.931	0.165	0.029
平均值	0.662	0.135	1.030	0.708	0.120	0.356	0.790	0.104	0.254