基于典型代表电站和改进SVM的区域 光伏功率短期预测方法

张扬科1,李 刚1,李秀峰2

(1. 大连理工大学 水电与水信息研究所,辽宁 大连 116024;2. 云南电力调度控制中心,云南 昆明 650000)

摘要:准确的区域光伏功率预测作为解决光伏并网消纳和多能互补问题的技术之一受到越来越多的关注,提 出一种基于典型代表电站和改进支持向量机(SVM)的区域光伏功率短期预测方法。通过K-means聚类将同 一地区光伏电站划分到不同汇聚区,使用历史数据和3种数学相关系数计算得到各汇聚区典型代表电站,并 通过4类光伏功率指标分析各典型代表电站与汇聚区的一致性,基于此,以改进SVM代替传统的滚动预报形 成区域功率预测模型。实际算例分析表明,所提方法可提升区域光伏功率短期预测精度。

关键词:K-means聚类;典型代表电站;短期预测;新能源出力;SVM

中图分类号:TM 615

文献标志码:A

DOI:10.16081/j.epae.202108017

0 引言

近年来,我国新能源代表之一的光伏产业发 展迅猛,2019年年底光伏电源装机容量达到2.04× 10⁸ kW,同比增长17.3%。随着光伏渗透率的逐年 提高,单站功率预测已无法满足调度部门协调调度 以及国家对新能源消纳的迫切需求,亟需展开对区 域光伏功率预测的研究。主要体现在以下几个方 面:电网调度与电源规划方面,精确的区域功率预测 有利于调度部门实现对大规模光伏电站的掌控以及 减弱光伏并网对电网的冲击[1],同时将光伏电源纳 入多能源互补协调调度体系可提高多种能源之间的 协同运作能力[2];运行风险方面,单一光伏电站因发 电随机性、波动性[3]无法顺利参与整体协同优化,而 区域出力因汇聚效应可减少系统运行带来的风险, 进而增强电网运行安全性;电力市场方面,2015年 国务院发布的《关于进一步深化电力体制改革的若 干意见》[4]提出,允许拥有分布式电源的用户或微电 网参与电力交易,而分布式电源因"就近建设"的特 点更容易形成光伏汇聚区,因此区域预测对日内现 货市场和电力交易具有重要意义[5];模型适应性方 面,单站预测模型需根据不同电站资料重新构建,这 导致不同电站之间的模型无法共用,而由于相关性 愈强的汇聚区彼此出力相互抵消的能力愈强,对区 域出力波动起到"削峰填谷"的作用[6],从而使总体 出力趋势更加明显,降低了预测难度。

光伏功率预测一直是国内外研究的热点,按预 测模型组成可分为3类:物理成因法,根据光伏电站 的组件参数和地理信息,如光电转换效率、光伏阵列

收稿日期:2021-03-29;修回日期:2021-06-28

基金项目:国家自然科学基金资助项目(51879030)

的安装角度等,结合当地的气象信息,按光伏物理成 因建立预测模型^[7]:统计学法,对光伏出力数据、误差 数据等历史数据进行统计和曲线拟合,寻找适合该 类数据的函数或概率模型构建输入、输出映射关系, 达到预测的目标^[8];人工智能算法,利用人工智能算 法对历史资料进行处理,使输入、输出间形成非线性 关系,进而得到预测结果^[9]。由于汇聚区与单一电 站功率预测在预测原理、数据资料的使用及建模方 式方面的不同,区域功率预测无法"照搬"单站预测 的形式。同时,由于多座单一电站的预测误差积累 和数据量级的影响,区域功率的预测值无法直接采 用多座单一电站预测值进行简单叠加。文献[1]使 用系数矩阵将个别电站预测功率推算到区域功率, 但该方法计算过程较为繁琐:文献[10]利用Pearson 相关系数选择各汇聚区代表电站,但单一数学指标无 法全方位分析该电站对汇聚区的代表性;文献[11] 使用类似滚动预报方法进行功率预测,但传统滚动 预报的误差随时间迅速增加,不是最优的预测方法。

为此,本文提出一种基于典型代表电站和改进 支持向量机(SVM)的区域光伏功率短期预测方法。 首先结合多种数据通过聚类分析形成不同的光伏汇 聚区;然后引入3种相关系数选取各汇聚区中的典型 代表电站,并通过4类评价指标对汇聚区中各电站 进行相关性分析;最后利用改进SVM构建典型代表 电站与汇聚区间的短期功率预测模型,通过典型代 表电站功率预测得到汇聚区总的光伏功率。算例结 果表明,该方法可提高区域光伏功率短期预测精度。

1 光伏汇聚区的建立

光伏汇聚区的划分不仅与出力有直接关系,同时与特性数据和气象数据等内在因素密不可分。文献[12]注重考虑气象因素和发电功率的聚类分析,但是没有考虑光伏出力的特性因素。

Project supported by the National Natural Science Foundation of China(51879030)

为使汇聚区中各电站在数据上更加体现集群内 部电站特征趋同性,在单站出力数据的基础上,添加 光伏出力的相关特性数据,即最大值、最小值、平均 值、峰谷差、变异系数^[13-14]和气象因素,进一步提高 数据系列的独立性。光伏汇聚区划分示意图见图1。



图1 汇聚区划分示意图

Fig.1 Schematic diagram of convergence area division

利用*K*-means聚类方法首先需要确定汇聚区的数量*K*,采用误差平方和最小作为评价函数:

$$\min J = \sum_{i=1}^{K} \sum_{p \in C_i} \left| p - m_i \right|^2 \tag{1}$$

式中:J为各簇中个体与质心差的平方和; C_i 为第i个簇;p为 C_i 中的样本; m_i 为 C_i 的质心,其为 C_i 中所有样本的均值。

如前文所述,将数据系列作为样本集合,通过迭 代计算得出样本集合的K簇,主要计算步骤为:

1)输入样本集合*M*,随机选定*K*个点设为初始 聚类中心,分别将其作为*K*个簇的中心,形成初始聚 类结果,如图1中①—③所示;

2)更新各个簇的中心,计算任一簇中包含的所 有样本数据对应的均值,将其设置为该簇的新中心, 如图1中④所示;

3)计算每个样本与各簇中心的欧氏距离,将其 分配到离其最近的簇,如图1中⑤所示;

4)判断是否满足迭代停止条件(达到循环次数 或者误差平方和达到最小值等),若满足,则聚类完 成,输出最终汇聚区划分结果,如图1中⑥所示,否 则,返回步骤2)。

利用上述步骤可以得到不同类别的划分结果以 及对应评价函数的数值,将数值从大到小排列,找到 下降速度最大的点,则该点数值对应的划分类别和 聚类结果即为最终结果。

2 汇聚区典型代表电站的选取及相关性评价

根据上述方法将区域相关性较强的光伏电站划 分到同一汇聚区,即区域中光伏电站的历史出力与 质心(各电站出力均值)的距离越小,区域相关性越 强,越易形成汇聚区。本文的关键点在于汇聚区典 型代表电站的选取,但常用方法^[10]仅关注 Pearson相 关系数,单一指标无法全面反映某座电站与汇聚区的 代表关系和一致性,为此,引入Spearman相关系数 和Kendall相关系数,结合多种数学相关系数从多个 角度进行综合对比。因此本文中电站典型性与相关 性类似,即为光伏电站与汇聚区的历史功率在较长 时间范围内的一致性,相关系数越大,典型性越强。

2.1 3种数学相关系数

2.1.1 Pearson 相关系数

Pearson相关系数是考察2个事物(经常为变量) 之间的相关程度,定义为:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}$$
(2)

式中: $\rho_{X,Y}$ 为Pearson相关系数; $X \setminus Y$ 分别为不同的变 量集合;cov(X,Y)为X和Y的协方差; σ_X 和 σ_Y 分别为 X和Y的标准差; $E(\cdot)$ 为数学期望。

2.1.2 Spearman 相关系数

将变量集合 $X=\{x_1, x_2, \dots, x_n\}$ (*n*为数据的数量) 中元素按降序或升序重新排列得到 $A=\{a_1, a_2, \dots, a_n\}$ 。将集合 X中每个元素 x_i 在序列 A中的次序记为 $r_g(x_i)$ 。同理按上述方法得到集合 $Y=\{y_1, y_2, \dots, y_n\}$ 中元素次序为 $r_g(y_i)$ 。Spearman 相关系数定义为:

$$r_{s} = 1 - \frac{6\sum d_{i}^{2}}{n(n^{2} - 1)}$$
(3)

$$d_i = r_g(x_i) - r_g(y_i) \tag{4}$$

式中: r_s 为Spearman 相关系数; d_i 为2个次序 $r_g(x_i)$ 与 $r_g(y_i)$ 的差值。

2.1.3 Kendall相关系数

当集合 X 和 Y 中变量对(x_i, y_i)、(x_j, y_j)同时增减 时,认为这2个变量一致;当其增减相反时,认为这2 个变量不一致;当二者相等时,认为这2个变量既不 是一致也不是不一致。

Kendall相关系数定义为:

$$R = \frac{C - D}{\sqrt{(N_3 - N_1)(N_3 - N_2)}}$$
(5)

$$N_1 = \sum_{i=1}^{s} \frac{1}{2} U_i (U_i - 1)$$
(6)

$$N_2 = \sum_{i=1}^{w} \frac{1}{2} V_i (V_i - 1)$$
(7)

$$V_3 = \frac{1}{2}n(n-1)$$
(8)

式中:*R*为Kendall相关系数;*C*、D分别为有一致性、 有不一致性的变量对数;*s*为变量集合*X*中由相同变 量组成的子集合数,*U*_i为其第*i*个子集合所包含的变 量数;*w*为变量集合*Y*中由相同变量所组成的子集合 数,*V*_i为其第*i*个子集合所包含的变量数。

参考文献[15]将汇聚区中各电站的3种相关系数结果按1:1:1的权值进行组合计算,将相关关系



表现最好、相关系数最大的电站作为该汇聚区中的 典型代表电站。

2.2 典型代表电站相关性分析

为进一步证实通过3种相关系数选择出的典型 代表电站与汇聚区出力的一致性,选择以下4类评 价指标对汇聚区中的各电站分别进行分析。

1)季节性。根据历史出力数据,对比不同季节 的典型日出力曲线,观察出力曲线呈现的特性。

2)气象典型日。根据已有的气象信息分别找出 不同季节典型日(晴天、阴天)的出力曲线,观察不同 曲线呈现的特点。

3)装机利用小时数。装机利用小时数可理解为 电站在日内的使用时间,如式(9)所示。

$$H = \left(\sum_{i=1}^{T} P_{i} h_{i}\right) / P_{\rm E} \quad t = 1, 2, \cdots, T$$
(9)

式中:H为装机利用小时数;T为时间段总数; P_t , h_t 分别为第t个时间段的功率和时长; P_E 为装机容量。

4)日内峰谷差及日内峰谷差占装机比重。峰 谷差为日内功率最大值与最小值之差,即:

$$P_{\rm df} = P_{\rm peak} - P_{\rm valley} \tag{10}$$

式中: P_{df} 为日内峰谷差; P_{peak} 为日内功率最大值,即顶峰功率; P_{valley} 为日内功率最小值,即低谷功率。

3 汇聚区出力预测模型

3.1 多输出 SVM 预测模型

SVM是基于结构风险最小化理论和万普尼克--泽范兰杰斯维VC(Vapnik-Chervonenkis dimension) 理论的一种小样本学习理论,使用领域较为广泛^[16]。 传统的SVM无法满足预测多个变量的要求,借鉴滚 动预测思想可实现多步骤滚动预测,虽然滚动预测 原理简单,易于操作,但随着预测次数的增多,误差 会迅速积累,导致最终结果失真,无法保持预测结果 的准确性,因此舍弃该方法。结合SVM本身输入、 输出结构,参考文献[17]对其进行简化,在不改变训 练数据输入值的基础上,依次对输出值对应的各个 特征点进行预测,该方法可在保持SVM优点的情况 下不从底层对SVM进行修改,其构造图如附录A图 A1所示。

3.2 区域功率预测模型

本文基于外推法和统计升尺度法^[10,18],结合多 种数学相关系数,优化典型代表电站及原网络结构, 建立24h内的短期预测模型,模型建立步骤如下。

1)利用汇聚区中各电站的历史出力数据,结合 3种相关系数计算得到典型代表电站。

2)将汇聚区中各电站的历史出力数据进行叠加 形成基础资料,结合典型代表电站的历史出力数 据及SVM形成区域功率预测模型。模型输入为典 型代表电站的历史出力数据,输出为对应日汇聚区 历史出力数据。 3)在模型建立、训练之后,输入预测日典型代表 电站预测值即可得到汇聚区功率预测值。

区域功率预测示意图如图2所示。



图2 区域功率预测示意图

Fig.2 Schematic diagram of regional power forecasting

该方法主要利用历史出力数据进行数据处理, 建模无需气象数据,且避免了常规叠加法导致的误 差累积问题,已有研究^[19]证实了本文思路的正确性。 由于光伏曲线的固有特性,因此更易找到单站功率 与汇聚区功率间的非线性联系,从而建立较准确的 预测模型,同时减少预测误差。为验证本文方法的 有效性,后文利用单机等值法^[20]和叠加法进行对比。

3.3 区域功率预测流程

基于上述原理,本文通过以下步骤得到区域光 伏功率短期预测结果。

1)划分汇聚区。基于*K*-means聚类方法的工作 原理和计算方法,利用历史出力数据、气象数据和特 性数据进行聚类划分,并判断汇聚区是否只有一座 电站,如果是,则直接转至步骤3),组成单站预测模 型并输出结果,否则,继续执行步骤2)。

2)选择典型代表电站和指标评价。基于外推法 和统计升尺度法,计算各汇聚区中各电站3种相关 系数,选择对应汇聚区典型代表电站;用4类出力指 标对汇聚区中各电站在多个时间节点进行相关性分 析,验证典型代表电站的最优性及代表性。

3)建立区域功率短期预测模型。将计算得到的 典型代表电站与汇聚区历史数据构成区域功率预测 模型(具体见3.2节)。

预测流程图如附录A图A2所示。

4 算例分析

4.1 算例简述

本文所采用的算例为云南省大理州与楚雄州 2017年5月至9月有记录的光伏电站,电站具体名 称如附录A表A1所示。

4.2 汇聚区划分

利用已有的数据资料,采用*K*-means聚类方法进行循环计算,表1为大理州类别数与距离和的关系。

由表1可知,类别数为4时距离和下降速度最快,故选择划分类别为4类,具体划分结果为:汇聚区Ⅰ,电站1-1、1-3、1-6;汇聚区Ⅱ,电站1-2、1-5、1-7、1-8;汇聚区Ⅲ,电站4;汇聚区Ⅳ,电站1-9。

208

表1 大理州类别数与距离和的关系

Table 1 Relationship between category number and

distance si	um of	Dali	prefecture
-------------	-------	------	------------

类别数	距离和	类别数	距离和
1	23214.1000	5	178.3750
2	5895.4600	6	83.9246
3	2031.0900	7	34.4408
4	348.3900	8	10.3196

根据计算结果,对汇聚区I中的3座电站和汇聚 区Ⅱ中的4座电站进行区域短期功率预测。同理, 对楚雄州的汇聚区进行类似划分,划分结果为:汇 聚区V,电站2-1、2-5、2-6、2-9;汇聚区VI,电站2-2、 2-3、2-4、2-7、2-8。

4.3 典型代表电站选择及相关性分析

利用3种相关系数选择典型代表电站并利用4 类指标进行相关性分析,以大理州汇聚区 I 为例。

表2为大理州"大佛山+干塘子+西村"汇聚区 I 典型代表电站选取情况。

表2 汇聚区I各电站相关系数

Table 2 Correlation coefficient of each station in Convergence Area I

中 카		相关	系数	
电珀	Pearson	Kendall	Spearman	复合
大佛山	0.922	0.716	0.871	0.836
干塘子	0.926	0.729	0.888	0.848
西村★	0.986	0.863	0.966	0.938

注:★表示该汇聚区典型代表电站。

由表2可知,西村展示出较强的代表性及与汇 聚区的相关关系,其3种单一相关系数均高于汇聚 区中的其余电站。为进一步说明典型代表电站选取 的正确性,利用各指标对电站进行评价,如图3-6 所示,同时利用各指标分别计算汇聚区中各电站与 汇聚区的相关系数,结果如表3所示。由于自身出 力的特殊性,光伏电站在夜晚以及清晨等时间出力 几乎为0,综合考虑后取96点出力数据(以15 min为 间隔)中08:00-18:00共41点的出力序列。

最终发现西村的3种指标均优于其他电站,证



图3 单站与汇聚区季节性对比

Fig.3 Seasonal comparison between single station and convergence area



图4 单站与汇聚区气象典型日对比

Fig.4 Typical meteorological day comparison between single station and convergence area



图 5 单站与汇聚区装机利用小时数对比





图 6 单站与汇聚区日内峰谷差对比

Fig.6 Intraday peak valley difference comparison between single station and convergence area

表3 相关性综合对比

Table 3 Comprehensive comparison of correlation

由 카	Pearson相关系数								
电珀	季节性	气象典型日	装机利用小时数	日内峰谷差					
大佛山	0.9965	0.9245	0.9070	0.7972					
干塘子	0.9946	0.9273	0.8491	0.6011					
西村	0.9991	0.9530	0.8788	0.9403					

明利用3种相关系数选择汇聚区中典型代表电站的 方法较为准确。

4.4 短期出力预测结果

采用本文方法分别计算各电站预测精度,同时 将本文方法和其他方法进行对比,计算结果分别如 表4、表5和图7所示。可知本文方法的预测精度优 于叠加法和单机等值法。

大理州和楚雄州部分汇聚区3种方法的结果对

表4 汇聚区 | 各电站预测精度

Table 4 Forecasting accuracy of each power

station in	Convergence	Area	L
------------	-------------	------	---

		単位:M₩
电站	RMSE	MAE
大佛山	5.48	4.73
干塘子	4.34	3.38
西村	10.14	7.22
汇聚区I	14.65	11.43

表5 汇聚区 I 预测精度对比

Table 5 Forecasting accuracy comparison of







Fig.7 Photovoltaic power comparison between forecasting and actual values in Convergence Area I

比如附录A表A2—A4所示。由表A2、表A3可知: 相关系数较大的典型代表电站在相关性分析的表现 上也优于汇聚区中的其他电站,这说明典型代表电 站与汇聚区在历史功率方面的高度一致性;本文方 法的RMSE、MAE均优于其他2种方法,其中多个汇 聚区的RMSE小于10 MW;虽然典型代表电站的单站 预测误差较小,均在10 MW以下,但将预测数据叠加 后却没有展现出较好的预测效果,因此虽然叠加法的 计算过程较简单,但在精度要求较高时推荐使用其他 方法。由表A4知:除10月7日的汇聚区 I 外,各汇聚 区在不同日期采用本文方法得到的误差最小;10月7 日所有汇聚区采用3种方法得到的预测误差均较 大,这是由于假期或突发事件导致预测值误差增大, 但综合而言本文方法较准确;单机等值法的预测精 度也较高,但是该方法在建模过程中需要的资料远 多于本文方法,这将会在整体预测流程中消耗过多 时间,此外,单机等值法精度提升较困难,而本文方 法由于预测模型的特点,可通过典型代表电站单站 预测技术的提高进一步提升区域功率预测的精度。

5 结论

为应对新能源的快速发展,本文针对汇聚区出 力预测进行优化,根据预测结果主要得到如下结论: 1)采用3种相关系数计算各汇聚区典型代表电站可减少单一指标对选择结果产生的扰动,尤其当 汇聚区各电站的单一相关系数(Pearson相关系数) 较为接近时,使用本文方法可从多方面分析得出最 优结果;

2)使用4类评价指标在较特殊的时间节点对各 电站进行相关性分析,可从侧面验证典型代表电站 选取的正确性;

3)本文方法不仅可以利用典型代表电站与汇聚 区的一致性得到较好的预测效果,还可将单站预测 方法用于典型代表电站,从而组成新的区域功率预 测模型,因此本文方法有较好的鲁棒性与灵活性。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

- [1]陈颖,陈富荣,程序,等.光伏电站群区域功率预测的统计升尺度方法[J].可再生能源,2012,30(11):20-23.
 CHEN Ying, CHEN Furong, CHENG Xu, et al. The regional photovoltaic power forecasting method based on statistical upscaling approach[J]. Renewable Energy Resources, 2012, 30 (11):20-23.
- [2] 程林,张靖,黄仁乐,等.基于多能互补的综合能源系统多场景规划案例分析[J].电力自动化设备,2017,37(6):282-287.
 CHENG Lin,ZHANG Jing,HUANG Renle, et al. Case analysis of multi-scenario planning based on multi-energy complementation for integrated energy system[J]. Electric Power Automation Equipment,2017,37(6):282-287.
- [3] 钟迪,李启明,周贤,等. 多能互补能源综合利用关键技术研究 现状及发展趋势[J]. 热力发电,2018,47(2):1-5,55.
 ZHONG Di,LI Qiming,ZHOU Xian, et al. Research status and development trends for key technologies of multi-energy complementary comprehensive utilization system [J]. Thermal Power Generation,2018,47(2):1-5,55.
- [4] 国务院.关于进一步深化电力体制改革的若干意见[EB / OL].
 (2015-03-22) [2021-03-28]. http://www.china-nengyuan.com/news/91900.html.
- [5] 郑海兴. 分布式区域光伏功率预测探讨[J]. 中国高新区, 2017(24):98-99.
- [6] 李翠萍,卓君武,李军徽,等.光伏发电与风光联合发电系统输出特性分析[J].电网与清洁能源,2017,33(1):95-102.
 LI Cuiping,ZHUO Junwu,LI Junhui, et al. Characteristic analysis of photovoltaic power generation and wind-photovoltaic joint power generation system[J]. Power System and Clean Energy,2017,33(1):95-102.
- [7] MA T, YANG H X, LU L. Solar photovoltaic system modeling and performance prediction[J]. Renewable and Sustainable Energy Reviews, 2014, 36:304-315.
- [8]管霖,陈旭,吕耀棠,等.适用于电网规划的光伏发电概率模型 及其应用[J].电力自动化设备,2017,37(11):1-7.
 GUAN Lin, CHEN Xu,LÜ Yaotang, et al. Probability model of PV generation for power system planning and its application[J]. Electric Power Automation Equipment,2017,37(11): 1-7.
- [9]张程熠,唐雅洁,李永杰,等.适用于小样本的神经网络光伏预测方法[J].电力自动化设备,2017,37(1):101-106,111.
 ZHANG Chengyi, TANG Yajie, LI Yongjie, et al. Photovoltaic power forecast based on neural network with a small number of samples[J]. Electric Power Automation Equipment,2017,37 (1):101-106,111.

- [10] 陈颖,孙荣富,吴志坚,等. 基于统计升尺度方法的区域风电场 群功率预测[J]. 电力系统自动化,2013,37(7):1-5.
 CHEN Ying,SUN Rongfu,WU Zhijian, et al. A regional wind power forecasting method based on statistical upscaling approach[J]. Automation of Electric Power Systems,2013,37(7): 1-5.
- [11] 王继东,宋智林,冉冉. 基于改进支持向量机算法的光伏发电 短期功率滚动预测[J]. 电力系统及其自动化学报,2016,28 (11):9-13.
 WANG Jidong,SONG Zhilin,RAN Ran. Short-term photovoltaic

power generation rolling forecast based on optimized SVM [J]. Proceedings of the CSU-EPSA, 2016, 28(11):9-13.

- [12] 管霖,赵琦,周保荣,等.基于多尺度聚类分析的光伏功率特性 建模及预测应用[J].电力系统自动化,2018,42(15):24-30.
 GUAN Lin,ZHAO Qi,ZHOU Baorong, et al. Multi-scale clustering analysis based modeling of photovoltaic power characteristics and its application in prediction[J]. Automation of Electric Power Systems,2018,42(15):24-30.
- [13] 葛朝晖,张倩茅,齐晓光,等.考虑灵活性需求的可再生能源出 力特性指标体系[J].电力系统及其自动化学报,2018,30(7): 30-37.

GE Zhaohui,ZHANG Qianmao,QI Xiaoguang,et al. Index system for renewable energy output characteristics considering flexible demand[J]. Proceedings of the CSU-EPSA,2018,30 (7):30-37.

- [14] 王建学,张耀,万筱钟.光伏出力特性指标体系和分类典型曲 线研究[J].电力需求侧管理,2017,19(5):8-12.
 WANG Jianxue,ZHANG Yao,WAN Xiaozhong. Research of index system and typical curves on photovoltaic power output characteristics[J]. Power Demand Side Management,2017,19 (5):8-12.
- [15] 郑可轲,牛玉广.大规模新能源发电基地出力特性研究[J]. 太阳能学报,2018,39(9):2591-2598.
 ZHENG Keke,NIU Yuguang. Research on renewable power basement output characteristics[J]. Acta Energiae Solaris Sinica,2018,39(9):2591-2598.
- [16]朱慧婷,杨雪,陈友媛.电动汽车充电负荷预测方法综述[J]. 电力信息与通信技术,2016,14(5):44-47.
 ZHU Huiting, YANG Xue, CHEN Youyuan. Overview of the charging load forecasting methods of plug-in electric vehicles

[J]. Electric Power Information and Communication Technology, 2016, 14(5):44-47.

- [17] 谢景新,程春田,秦颖.基于并行神经网络集成的多步预测方法[J]. 计算机工程与应用,2006,42(29):75-77,80.
 XIE Jingxin,CHENG Chuntian,QIN Ying. Multi-step-ahead prediction based on parallel neural network ensemble[J]. Computer Engineering and Applications,2006,42(29):75-77,80.
- [18] 崔杨,陈正洪,许沛华. 基于机器学习的集群式风光一体短期 功率预测技术[J]. 中国电力,2020,53(3):1-7.
 CUI Yang, CHEN Zhenghong, XU Peihua. Short-term power prediction for wind farm and solar plant clusters based on machine learning method[J]. Electric Power, 2020, 53(3):1-7.
- [19] 张小萌,白恺,柳玉,等. 大规模分布式光伏短期集群功率预测 综述[J]. 华北电力技术,2017(6):1-7.
 ZHANG Xiaomeng,BAI Kai,LIU Yu,et al. Overview for largescale distributed photovoltaic power prediction with clustered method[J]. North China Electric Power,2017(6):1-7.
- [20] 盛万兴,季宇,吴鸣,等.基于改进模糊C均值聚类算法的区域集中式光伏发电系统动态分群建模[J].电网技术,2017,41 (10):3284-3291.

SHENG Wanxing, JI Yu, WU Ming, et al. Dynamic clustering modeling of regional centralized photovoltaic power plant based on improved fuzzy C-means clustering algorithm[J]. Power System Technology, 2017, 41(10): 3284-3291.

作者简介:



张扬科(1993—),男,山东平度人,硕 士研究生,主要研究方向为新能源功率预 测、多能互补协调调度(E-mail:zyk@mail. dlut.edu.cn);

李 刚(1979—),男,吉林辽源人,副 教授,博士,通信作者,主要研究方向为电力 市场及电力系统经济运行等(E-mail:glee@ dlut.edu.cn);

张扬科 dlut.e

李秀峰(1983—),男,云南大理人,高级 工程师,硕士,主要研究方向为电力系统经济运行(E-mail: lixiufeng198362@163.com)。

(编辑 王锦秀)

Short-term forecasting method for regional photovoltaic power based on typical representative power stations and improved SVM

ZHANG Yangke¹, LI Gang¹, LI Xiufeng²

Institute of Hydropower System & Hydroinformatics, Dalian University of Technology, Dalian 116024, China;
 Yunnan Electric Power Dispatching and Control Center, Kunming 650000, China)

Abstract: Accurate regional photovoltaic power forecasting attracts more and more attention since it is one of the techniques for solving problems of photovoltaic grid-connection consumption and multi-energy complementary. A short-term forecasting method for regional photovoltaic power based on typical representative power stations and improved SVM (Support Vector Machine) is proposed. The photovoltaic power stations in the same region are divided into different convergence areas by *K*-means clustering, the typical representative station in each convergence area is calculated by using historical data and three mathematical correlation coefficients, and the consistency of each typical representative station with the convergence area is analyzed through four photovoltaic power indices. On this basis, a regional power forecasting model is formed by substituting the traditional rolling forecasting with the improved SVM. The actual example analysis shows that the proposed method can improve the short-term forecasting accuracy of regional photovoltaic power.

Key words: K-means clustering; typical representative power stations; short-term forecasting; renewable energy output; SVM

附录 A:



图 A2 预测流程图 Fig.A2 Flowchart of forecasting

大理州光伏国	电站	楚雄州光伏电站			
编号	名称	编号	名称		
1-1	大佛山	2-1	班幸		
1-2	干海子	2-2	长冲		
1-3	干塘子	2-3	大古衙		
1-4	老鹰岩	2-4	大庄		
1-5	四角山	2-5	干巴拉		
1-6	西村	2-6	河外		
1-7	岩淜	2-7	天子山		
1-8	杨梅坪	2-8	小西村		
1-9	子牙关	2-9	秀田		

表 A1 大理州和楚雄州光伏电站名称

Table A1 Name of photovoltaic power stations in Dali and Chuxiong prefectures

表 A2 剩余汇聚区各电站相关性分析及单站预测精度

	, ,• • • •	
Lable A / Correlation analysis of each po	ower station in remaining convergence area	and forecasting accuracy of single station
ruble 112 Conclution unarysis of cuch p	swer stution in remaining convergence area	and forecusting accuracy of single station

汇			相	关系数			Pearson				
聚区	电站	Pearson	Kendall	Spearman	复合	季节性	气象典型日	利用小时数	峰谷差	RMSE/MW	MAE/MW
	干海子	0.831	0.524	0.667	0.674	0.995 4	0.951 1	0.917 7	0.792 1	3.91	3.04
	四角山★	0.830	0.656	0.822	0.769	0.998 5	0.982 5	0.872 1	0.900 9	8.41	6.23
п	岩淜	0.836	0.606	0.771	0.737	0.995 7	0.961 0	0.870 8	0.815 7	2.34	1.87
	杨梅坪	0.831	0.598	0.741	0.723	0.997 0	0.893 3	0.937 4	0.818 7	3.38	2.76
	班幸	0.896	0.738	0.905	0.846	0.998 4	0.942 3	0.952 0	0.827 0	6.00	5.18
V	千巴拉	0.857	0.684	0.858	0.800	0.998 7	0.922 7	0.898 3	0.827 8	9.55	7.94
v	河外★	0.908	0.751	0.914	0.858	0.998 9	0.985 7	0.943 2	0.891 4	7.18	5.81
	秀田	0.834	0.669	0.817	0.773	0.998 0	0.921 8	0.806 7	0.687 0	7.89	6.52
	长冲	0.811	0.618	0.806	0.745	0.995 5	0.972 4	0.718 6	0.713 9	3.83	3.17
	大古衙	0.763	0.573	0.765	0.700	0.994 9	0.927 0	0.762 9	0.687 8	4.59	3.76
VI	大庄	0.731	0.492	0.659	0.627	0.996 2	0.896 4	0.721 7	0.575 2	5.71	4.57
	天子山	0.808	0.635	0.820	0.754	0.997 7	0.950 1	0.771 0	0.529 1	4.46	3.78
	小西村★	0.820	0.644	0.834	0.766	0.998 4	0.981 9	0.884 0	0.608 2	3.94	3.24

			里位: MW
汇聚区	方法	RMSE	MAE
	叠加法	10.19	7.56
II	单机等值法	9.15	6.73
	本文方法	9.09	6.76
	叠加法	23.83	20.60
v	单机等值法	24.17	20.48
	本文方法	23.80	20.18
	叠加法	10.01	7.28
VI	单机等值法	10.00	7.37
	本文方法	9.97	7.37

表 A3 汇聚区功率预测精度

Table A3 Power forecasting accuracy of convergence area

表 A4 6 天之内各汇聚区 3 种预测方法预测精度对比

Table A4 Forecasting accuracy comparison among three forecasting methods in each convergence area within six days

												单位:	MW
		10 月	2 日	10 月	3日	10 月	4日	10 月	5日	10 月	6日	10月	7日
汇聚区	方法	RMSE	MAE										
Ι	叠加法	14.65	11.43	19.93	17.33	23.78	18.05	25.05	21.05	33.76	28.24	30.25	27.50
	单机等值法	15.53	12.24	20.54	17.82	25.65	19.67	19.08	14.57	16.32	12.33	54.35	50.09
	本文方法	14.59	11.37	19.84	17.30	22.27	16.06	18.60	13.35	15.79	11.78	49.75	41.54
II	叠加法	10.19	7.56	10.55	8.96	15.17	12.02	10.16	7.91	12.88	10.40	28.02	26.04
	单机等值法	9.15	6.73	10.53	8.92	14.73	11.45	10.22	8.06	12.92	10.58	27.86	25.99
	本文方法	9.09	6.76	10.44	8.61	14.55	11.34	9.95	7.78	11.72	8.97	27.74	25.68
V	叠加法	23.83	20.60	54.81	45.44	26.46	19.28	21.36	17.89	15.36	12.18	46.56	41.54
	单机等值法	24.17	20.48	53.85	44.78	26.63	20.04	22.44	18.67	15.97	12.43	46.04	41.43
	本文方法	23.80	20.18	53.84	44.07	25.70	19.16	20.39	16.60	15.28	11.79	45.88	41.17
VI	叠加法	10.01	7.28	10.73	8.32	10.10	7.88	11.14	8.86	10.33	7.79	23.25	21.12
	单机等值法	10.00	7.37	11.22	8.43	10.53	8.24	10.37	7.93	10.98	8.35	24.95	22.83
	本文方法	9.97	7.37	10.72	8.16	10.03	7.81	10.27	7.48	10.11	7.24	23.03	20.92