[光伏发电]

基于相似日选取的光伏发电功率预测方法研究

文/上官诚江[1]　陈丽霞[2]
1. 福建信息职业技术学院；2. 国网福建省电力有限公司福州供电公司

从光伏电站出力的影响因素入手，深入研究光伏发电功率预测方法。结合某光伏电站2021年实测功率数据和天气数据，对影响光伏发电功率的各项因素用皮尔逊相关系数进行相关性分析，得到相关性最大的四个主要影响因素。然后以这四个影响因素为聚类特征完成模糊C均值聚类，并依此提出一种新的相似日选取的指标χ_i，依据仿真算例验证了该指标的可行性。

传统能源发电的功率是确定且可预期的，而光伏发电功率最显著的特点就是不确定性。随着光伏发电的大规模并网，极大地增加了电力系统调度的难度。为了解决这个问题，需要在电力系统内配置一定的旋转备用容量进行平衡，但是这会进一步加大电力系统的运营成本，不利于光伏发电的利用与推广。

光伏发电功率预测就是根据某些参数对光伏发电未来一段时间内的功率情况进行预测，将具有不确定性、不可控制的电能转变为可预测的电能，再通过能量管理系统（Energy Manage System，EMS）变为可以控制的电能。准确地预测光伏发电功率，可以降低电网里备用的容量及因此产生的各项费用，为系统的调度、实时运行与控制给予理论上的支持，较为有效地减轻光伏发电并网对电网的冲击，消除一部分令人棘手的弃光现象，获得更为可观的经济社会效益。

国内外有诸多针对光伏发电功率预测的相关研究。文献[4]选取太阳辐照时间、辐照强度及气温三个因素，利用灰色关联系数法计算待预测日与历史样本里某一日的总关联度，并依此进行相似日选取。

文献[5]提出以灰色关联分析法结合余弦相似度指标组成综合相似度指标，并进行相似日选取和云自适应粒子群优化（CAPSO）算法优化脉中神经网络（SNN）的发电功率预测模型。

文献[6]选取太阳辐照强度、辐照时间和气温三个因素，利用欧氏距离量化待预测日与历史样本里某一日的总体区别，并依此进行相似日选取。

结合某光伏电站2021年实测功率数据和天气数据，对影响光伏发电功率的各项因素用皮尔逊相关系数进行相关性分析，得到相关性最大的四个主要影响因素。然后以这四个影响因素为聚类特征完成模糊C均值聚类，并依此提出一种新的相似日选取的指标χ_i。

光伏发电功率影响因素分析

排除光伏电站自身固有的属性，引起光伏发电功率变化的主要是外部气象环境因素。从物理意义上讲，光伏组件表面接收的太阳总辐射量是影响其发电的最为直接、关键的因素。除此之外，环境温度和天气类型等气象因素也给光伏发电功率带来不同程度的影响。众多因素影响着光伏发电功率，而它们的作用机理大相径庭，彼此也并不独立，体现

了光伏发电功率复杂和维数高的特点，给建模工作带来相当大的难度。

研究光伏发电功率时，如果将上述因素全部考虑在内，将会导致输入变量的冗余，增加学习时间，降低学习效率，在一定程度上降低系统的鲁棒性。为了简化光伏发电功率预测模型，提高模型预测的效率与准确度，必须严谨地对光伏发电功率与各类因素之间的相关性进行量化研究。

相关系数是统计学上用来表征变量间相关关系密切程度的一个指标，若按照研究的对象、关系的性质和变量的个数等因素进行划分将有多种分类方式。本文用皮尔逊相关系数来说明两个变量之间的相关关系的密切程度，数学计算式为

$$r=\frac{\sum_{i=1}^{N}(X(i)-\bar{X})(Y(i)-\bar{Y})}{\sqrt{\sum_{i=1}^{N}(X(i)-\bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y(i)-\bar{Y})^2}}=\frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{cov}(X,X)}\sqrt{\mathrm{cov}(Y,Y)}} \tag{1}$$

式中，X与Y是两个变量序列；i为元素的编号；N为元素的总数目。

绝对值越大，表示两个序列之间的相关关系越密切，且满足$-1\leqslant r\leqslant 1$。

结合某光伏电站2021年实测功率数据和天气数据进行相关性分析。探讨辐照度与光伏发电功率的相关性，用日照时数近似替代。考虑到天气预报提供的气象数据的种类，主要用相关系数探讨天气类型、气压、气温、相对湿度、风速及日照时数与光伏发电功率的相关性。

统计过程中，由于天气类型以文字形式预报，为了方便进行相关性分析，对天气类型进行量化。天气类型为晴的赋值为4，多云的赋值为3，阴的赋值为2，其余各类雨雪雾等赋值为1。类似的，季节为春、秋的赋值为2，夏的赋值为3，冬的赋值为1（季节以3—5月为春，6—8月为夏，9—11月为秋，12—2月为冬）。

光伏发电功率与季节类型、天气类型、平均气压、平均气温、平均相对湿度、平均风速和日照时数的相关系数统计见表1。

由表1可知，光伏发电功率与日照时数以及天气类型为高度正相关系数（均大于0.8），与平均相对湿度表现为较高的负相关，与平均气温、季节类型为低度正相关，与平均风速为低度负相关，与平均气压的相关性很小。

表1　光伏发电功率与各类影响因素的相关系数

影响因素	季节类型	天气类型	平均本站气压/(0.1MPa)	平均相对湿度(1%)	平均风速/(0.1m/s)	平均气温/(0.1℃)	日照时数/(0.1 h)
相关系数	0.155	0.832	−0.052	−0.571	−0.171	0.223	0.874

选取相关性最大的四个因素（日照时数、天气类型、平均相对湿度和平均气温）作为预测模型的输入，且以此为聚类特征进行模糊C均值聚类，并得出相似日选取指标。

相似日选取

光伏发电功率的波动性使待预测日和历史样本的关联度较小，限制了各类光伏发电功率预测方法的预测效率与预测准确度。若预测前先以待预测日的主要影响因素为特征，从以往有数据记录的天数中选取出相似日，然后用它们组成样本再完成预测，预测效率及准确度会得到一定的提高。基于模糊C均值聚类算法，提出一种新的相似日选取的指标χ_i。

1. 模糊C均值聚类算法

模糊C均值聚类算法（Fuzzy-C-Means，FCM）是一种理论完善且应用广泛的聚类算法，核心是把数据集划分为多个簇，属于相同簇的数据彼此相似程度非常大；不属于相同簇的数据彼此相似程度非常小。

模糊C均值聚类算法是一种迭代求解的算法，聚类计算的目标表达式为

$$J(\boldsymbol{U},\boldsymbol{V})=\sum_{i=1}^{M}\sum_{c=1}^{C}(\mu_c(i))^m\left\|X_i-V_c\right\|^2,\ 1\leqslant m<\infty \tag{2}$$

式中，$\boldsymbol{U}$为隶属度矩阵；$\boldsymbol{V}$为各个聚类中心位置向量；$J(\boldsymbol{U},\boldsymbol{V})$为目标函数，聚类就是让它达到最小值的过程，得到的$\boldsymbol{U}$、$\boldsymbol{V}$为最终的结果；$M$为样本数据点个数；$C$为聚类中心个数；$m$为权重系数；$X_i$ $(i=1,2,\cdots,M)$为样本中每个数据点的聚类特征；V_c为第c $(c=1,2,\cdots,C)$个聚类中心的位置向量；$\mu_c(i)$为第i个数据点对于第c个聚类中心的隶属度，且$\mu_c(i)$

满足如下约束

$$\begin{cases}\sum_{c=1}^{C}\mu_c(i)=1\\ 0\leqslant\mu_c(i)\leqslant 1\end{cases} \tag{3}$$

模糊C均值聚类的主要流程为：

1）给各个参数赋予合适的初值，并且随机选取C个初始的聚类中心。

给M个数据点的隶属度矩阵$\boldsymbol{U}$赋初值

$$\boldsymbol{U}^{(g)}=[\mu_c(i)^{(g)}]\in\boldsymbol{R}_{C\times N} \tag{4}$$

式中，g是迭代次数，其初始值为0；$\boldsymbol{R}_{C\times N}$为$C\times N$阶的有理数矩阵。

2）第g+1次的聚类c的中心位置向量为

$$\boldsymbol{V}_c=\frac{\sum_{i=1}^{M}(\mu_c(i)^{(g)})^m\boldsymbol{X}_i}{\sum_{i=1}^{M}(\mu_c(i)^{(g)})^m},1\leqslant c\leqslant C \tag{5}$$

3）在上一步的基础上计算出各个数据点$X_i(i=1,2,\cdots,M)$的隶属度为

$$\mu_c(i)^{(k)}=\frac{1}{\sum_{j=1}^{C}\left[\frac{\|\boldsymbol{X}_i-\boldsymbol{V}_c\|}{\|\boldsymbol{X}_i-\boldsymbol{V}_j\|}\right]^{\frac{2}{m-1}}} \tag{6}$$

4）判断迭代的结束条件为

$$\left\|\mu_c(i)^{(g)}-\mu_c(i)^{(g-1)}\right\|<\varepsilon \tag{7}$$

式中，ε是迭代终止条件界限。

如果式（7）成立，就停止迭代并输出聚类中心与隶属度矩阵；如果式（7）不成立，就回到计算步骤3）继续进行迭代计算。式（5）与式（6）是对式（2）求取最小值后得到的表达式。

聚类对象是光伏发电功率，聚类特征选择了相关性较大的日照时数、天气类型、平均相对湿度和平均气温四个因素。

2. 相似日的选取

从需要预测的某一天的天气预报信息中提取相关数据，构成第M+1个数据点的聚类特征X_{M+1}；将所有M+1个数据点进行模糊C均值聚类，并获得M+1个数据点的隶属度$\mu_c(i)(i=1,2,\cdots,M+1)$和$C$个聚类的中心位置向量$V_c(c=1,2,\cdots,C)$。

如果第i个数据点和第M+1个数据点属于同一聚类中心，为了充分利用隶属度函数量化它们聚类特征之间的差异，可设相似日选取指标χ_i为

$$\chi_i=\frac{1}{C}\sum_{c=1}^{C}[\mu_c(i)-\mu_c(M+1)]^2 \tag{8}$$

如果第i个数据点和第M+1个数据点不属于同一聚类中心，为了量化它们聚类特征之间的差异，可设相似日选取指标χ_i为

$$\chi_i=1 \tag{9}$$

相似日选取指标χ_i量化了第i个数据点与第M+1个数据点聚类特征的总体区别，χ_i越小，表明它们之间的相似度越高。相似日选取在所有M个数据点中按χ_i从小到大进行选取。

算例分析

依据提出的相似日选取指标，在进行模糊C均值聚类的基础上选取前10名相似日，并将前10名相似日与随机选取的10日对应的数据作为同一神经网络训练样本进行训练，再将训练完的神经网络分别对预测日进行光伏发电功率预测。

1. 相似日的选取

对某装机容量200 kW的光伏电站2021年实测功率数据和天气数据进行处理。选取2021年12月31日为待预测日。取聚类特征为日照时数、天气类型、平均相对湿度、平均气温及聚类样本总M+1=365，聚类中心数C=4，并在此基础上进行聚类。聚类后第一类样本数为74，第二类为92，第三类为97，第四类为102。计算前364天每天的相似日选取指标，并按χ_i从小到大进行排序，前10名相似日及随机选取的10日对比结果见表2。

表2　前10名相似日与随机选取的10日对比结果

前10名相似日	指标	排序	随机选取的日期
2020/3/14	0.000 1	1	2020/10/9
2020/11/27	0.000 1	2	2020/5/22
2020/3/25	0.000 2	3	2020/1/19
2020/12/14	0.000 3	4	2020/7/15
2020/11/30	0.000 4	5	2020/7/8
2020/12/2	0.001 3	6	2020/3/15
2020/2/17	0.001 8	7	2020/1/21
2020/12/15	0.002 0	8	2020/4/7
2020/2/3	0.002 2	9	2020/12/1
2020/4/20	0.002 2	10	2020/11/3

由表2可知，前10名相似日的日期多在冬季、春季的起始以及秋季的末尾，与待预测日较为接

近，而随机选取日的日期四季分布较为平均。

用表2的数据作为训练样本，分别对同一个BP神经网络进行训练，然后利用训练完的两个BP神经网络分别对待预测日进行光伏发电功率预测。随机选取日预测值、相似日预测值与实际功率值如图1所示，随机选取日与相似日的预测绝对误差如图2所示，随机选取日与相似日的百分比误差（MAPE）、均方根误差（RMSE）和希尔不等系数（TIC）见表3。

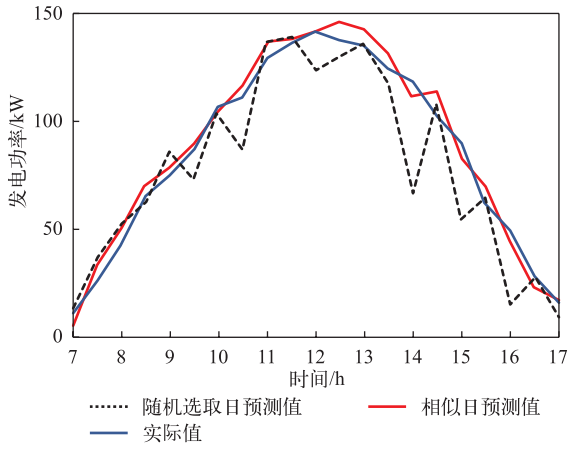

图1　随机选取日预测值、相似日预测值与实际功率值

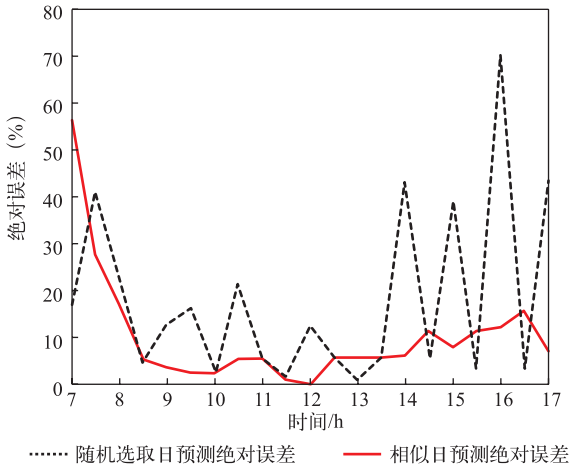

图2　随机选取日与相似日的预测绝对误差

表3　随机选取日与相似日的MAPE、RMSE和TIC

日期	MAPE	RMSE	TIC
随机选取日	18.194	17.953	0.098
相似日	10.288	6.091	0.032

由图1、图2和表3可知，随机选取日预测MAPE较大，为18.194%，而提出的相似日的预测MAPE下降很多，为10.288%；随机选取日预测值与实际功率值在很多点相差较大，而提出的相似日预测值与实际功率值在绝大部分点都较为接近。

相似日选取指标χ_i，以日照时数、天气类型、平均相对湿度和平均气温四个主要因素为聚类特征进行模糊C均值聚类，充分发挥作为聚类结果的隶属度函数，量化待预测日与历史样本第i个数据点聚类特征的总体区别，具有一定的实用性。

结束语

光伏发电功率大小受外界因素影响具有不确定性，给电力系统的稳定运行带来不容忽视的影响。文章基于皮尔逊相关系数对影响光伏发电功率的各项因素进行分析，得出相关性最大的日照时数、天气类型、平均相对湿度和平均气温四个因素。以这四个因素作为聚类特征用模糊C均值聚类算法进行聚类，并以此为基础提出一种新的相似日选取的指标χ_i。通过算例验证，提出的新模型预测效率和准确度高，具有一定的实用性和可行性。EA

（收稿日期：2022-03-23）

参考文献

[1] 李旭．基于典型日出力特性分析的光伏电站功率预测研究[D]．北京：华北电力大学，2016．

[2] Godfrey Boyle．Renewable electricity and the grid[M]．London：Earthscan Publications Ltd，2007．

[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，2016，40(4)：140-151．

[4] 王晓兰，葛鹏江．基于相似日和径向基函数神经网络的光伏阵列输出功率预测[J]．电力自动化设备，2013，33(1)：100-103．

[5] 陈通，孙国强，卫志农，等．基于相似日和CAPSO-SNN的光伏发电功率预测[J]．电力自动化设备，2017，37(3)：6．

[6] 李建红，陈国平，葛鹏江，等．基于相似日理论的光伏发电系统输出功率预测[J]．华东电力，2012，40(1)：5．

[7] 翟学，郭紫昱，曲名新，等．分布式光伏发电系统接线方案优化[J]．电气应用，2020，39(12)：87-93.