

# 基于密度峰值聚类及 GRNN 神经网络的光伏发电功率预测方法

姚宏民, 杜欣慧, 秦文萍

(太原理工大学电气与动力工程学院, 太原 030024)

**摘 要:** 提出一种基于密度峰值聚类及广义回归神经网络(general regression neural network, GRNN)的光伏发电功率短期预测方法。首先分析气象因素特征, 计算统计指标构成天气特征向量; 其次, 为克服传统  $K$  均值聚类算法易陷入局部最优, 需人为指定聚类中心等问题, 采用密度峰值聚类算法对历史气象数据聚类识别; 最终针对每一类簇构建 GRNN 预测模型。仿真结果表明, 该文所述方法具有较高的预测精度, 且在小样本环境下仍有良好的适用性。

**关键词:** 聚类分析; 神经网络; 预测; 大数据; 光伏发电

**中图分类号:** TM615

**文献标志码:** A

## 0 引 言

由于光伏发电受辐照度、温度等天气因素的影响, 其输出功率具有波动性、间歇性和随机性的特点<sup>[1]</sup>, 如果没能准确预测光伏发电功率, 会对现有电力系统的功率平衡和安全稳定运行带来威胁<sup>[2]</sup>。因此, 精确预测光伏发电功率的能力被确定为大规模光伏一体化的最主要挑战之一<sup>[3]</sup>。

光伏发电功率预测按建模方式分为物理法和统计法<sup>[4]</sup>。物理法通过物理方程对光伏系统建模来实现预测, 需要详细的辐照度数据; 统计法采用统计和机器学习等方法预测发电功率<sup>[5]</sup>。由于统计法相较于物理法无需考虑光伏组件参数, 已被证明比物理法能够更好的适用于光伏发电功率预测<sup>[6]</sup>。文献[7]提出根据天气类型将数据集分类, 用不同的数据集对相应类型的预测日进行预测, 但历史气象数据并未包含天气类型, 需根据历史数据人为定义其所属类型; 文献[8]运用基于马尔科夫链的方法, 但该方法对于非晴空和天气状况剧烈变化下的预测精度不高。文献[9]引入天气类型指数, 作为 BP 神经网络输入的一部分进行预测, 但 BP 神经网络存在着学习速度慢、容易陷入局部最小值等缺陷。文献[10]使用数据挖掘的方法对约

减过的气象因素进行聚类, 对每一类建立神经网络, 但  $K$  均值聚类算法需要人为指定聚类数、聚类中心, 其聚类效果很大程度取决于聚类个数以及聚类中心的选取。

针对以上问题, 本文提出基于密度峰值聚类的 GRNN 神经网络光伏发电功率预测。以澳大利亚某太阳能技术示范区近 1 年的光伏发电历史数据和站测气象数据为基础, 分析数据特征及相关因素, 构建天气特征向量, 采用能自动确定聚类数及聚类中心的密度峰值聚类算法对历史数据聚类, 分别针对每一类簇建立 GRNN 预测模型, 将仿真计算结果与传统方法进行对比。实验结果表明, 本文方法在达到较高精度的同时还对不同天气类型及较少样本环境具有很好的适用性。

## 1 数据预处理及分析

### 1.1 数据集预处理

本文采用的出力数据来自位于澳大利亚的某太阳能技术示范区, 收集 2017 年的出力数据和气象数据<sup>[11]</sup>, 并剔除 25 d 的异常值, 选取每日 00:00~24:00 的出力数据和气象数据。原始出力数据时间间隔为 5 min, 本文将其处理成间隔为 1 h 并归一化到区间 $[0, 1]$ 。

收稿日期: 2018-03-30

基金项目: 国网山西省电力公司科技项目

通信作者: 杜欣慧(1965—), 女, 博士、教授, 主要从事电力系统运行方面的研究。duxinhui211@163.com

## 1.2 出力数据分析

采用时间序列的加法分解模型分析光伏系统出力数据的特点,如图 1 将原始数据分解为长期趋势因素、季节变动因素和随机变动因素。可看出虽然每日系统出力有不同程度的波动,但总体趋势从凌晨的最低点上升到正午达到峰值后开始下降。全年光伏系统的最大出力出现在夏季(澳大利亚每年的 10 月份至次年 2 月份),而冬季(每年的 5~8 月份)的出力水平偏低。随机变动因素有时会非常大,这是天气状况剧烈变化所导致的。

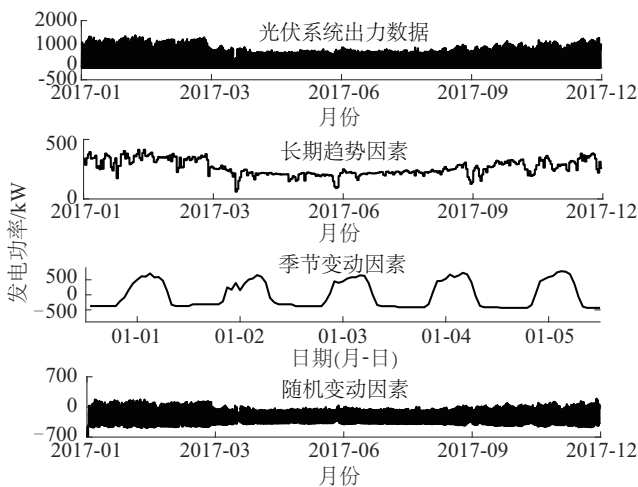


图 1 光伏系统出力数据的时间序列分解

Fig. 1 Decomposition of PV power data

表 1 气象因素与输出功率之间的距离分析结果

Table 1 Distance analysis between weather factors and power

气象因素	$w_s$	$T_a$	$I_r$	$w_d$	$R_d$	$W_{mx}$	$p_a$
$P$	<b>0.47</b>	<b>0.48</b>	<b>0.95</b>	-0.16	-0.03	<b>0.53</b>	-0.11

注:表中粗体为正相关且相关程度较大者。

对于风速的刻画,采用导数来作为描述局部变化特征的指标<sup>[2]</sup>。将归一化后的风速值与平均值差对时间的 1 阶、2 阶和 3 阶差分分别表示为  $w_d'$ 、 $w_d''$  和  $w_d'''$ 。如图 2 所示,其中横轴  $\Delta t$  为时间变化率。可见  $w_d'''$  对

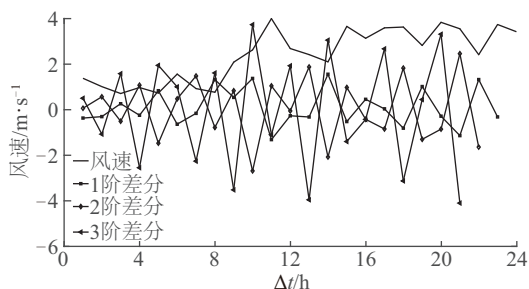


图 2 风速及其统计指标变化

Fig. 2 Changes in wind speed and statistical indicators

此外,不同的天气状况下系统出力也有很大差别。晴天的出力水平远大于多云天气和雨天,且晴天的出力曲线更加平滑,雨天和多云天气的系统出力曲线波动性较大。相似天气状况下系统出力具有较高的相似度,因此,可对原始气象数据进行无监督的天气类型聚类识别,深度挖掘历史气象数据,将相似度高的历史日分为一类,分别针对每一天气类型建立预测模型来提高预测精度。

## 2 天气特征向量的构造

以 2017 年站测气象数据为例,对数据预处理后,分别计算风速 ( $w_s$ )、温度 ( $T_a$ )、辐照度 ( $I_r$ )、风向 ( $w_d$ )、日降雨量 ( $R_d$ )、最大风速 ( $w_{mx}$ )、大气压强 ( $p_a$ ) 与输出功率 ( $P$ ) 的距离如表 1 所示。由表 1 得出风速、温度、辐照度、最大风速与输出功率的距离最小,相关性最高。如果将未作处理的多维历史数据序列作为预测因子会导致模型输入维数太高,还引入了许多冗余信息,给建模造成较大的困难<sup>[12]</sup>。同时,这 4 种气象因子均有很强的波动性,其中风速在 24 h 内的变化有很大的随机性。因此,可通过对气象因素的特征分析,计算其统计指标作为代替多维时间序列的天气特征量,既能充分的提取历史数据中的有效信息,还可避免维数太高带来的问题。

风速的变化相较于  $w_d'$ 、 $w_d''$  有更高的敏感度,故选取 3 阶差分最大值  $w_{dmx}'''$  作为特征因子,能够表征风速随机变化的局部特征。

辐照度与输出功率有着最大的相关系数,起决定性的作用,也应考虑日实时辐照度与平均值差的 3 阶差分最大值  $I_{mx}'''$ 。故选取日辐照度最大值  $I_{mx}$ 、3 阶差分最大值  $I_{mx}'''$  作为特征因子。

对于温度和最大风速选取每日环境温度平均值  $T_m$ 、每日环境温度最大值  $T_{mx}$ 、每日最大风速平均值  $w_{mxm}$  作为特征因子。

综上,共选取 6 维特征因子组成天气特征向量,即:

$$W = \{w_{dmx}''', I_{mx}, I_{mx}''', T_m, T_{mx}, w_{mxm}\} \quad (1)$$

式中,  $\mathbf{W}$  ——天气特征向量;  $\omega_{\text{dmx}}'''$  ——风速 3 阶差分最大值;  $w_{\text{mxm}}$  ——每日最大平均风速;  $I_{\text{mx}}$  ——每日辐照度最大值;  $I_{\text{mx}}'''$  ——每日辐照度 3 阶差分最大值;  $T_m$  ——每日环境温度平均值;  $T_{\text{mx}}$  ——每日环境温度最大值。

### 3 密度峰值聚类算法

密度峰值聚类算法是于 2014 年发表的一种新型聚类算法。该算法完全基于数据点间的距离,能够识别非球状类簇,相比于传统的  $K$  均值算法,可自动确定聚类中心及聚类数,能够快速搜索并找到数据点中的密度峰值<sup>[13]</sup>。

#### 3.1 算法原理

对于聚类中心的定义是密度峰值聚类算法的核心,因此本文着重介绍算法对于聚类中心的刻画。针对聚类中心,考虑 2 点假设:1)聚类中心的自身密度大。即它被局部密度小于它的邻居所包围;2)聚类中心与局部密度比它大的数据点间的距离相对更大。

设待聚类的数据集  $X = \{x_1, x_2, \dots, x_n\}$ , 对应的下标集为  $I_x = \{1, 2, \dots, N\}$ , 基于数据点间的距离  $d_{ij}$ , 对其中每一数据点  $x_i \in X$ , 计算 2 个重要参数:局部密度  $\rho_i$  和距离  $\delta_i$ 。

局部密度  $\rho_i$  采用高斯核函数定义<sup>[14]</sup>, 即:

$$\rho_i = \sum_{j \in I_s} e^{-\left(\frac{d_{ij}}{d_c}\right)^2} \quad (2)$$

式中,  $d_{ij}$  ——数据点  $x_i$  和  $x_j$  间的距离;  $d_c$  ——截断距离。从定义, 即式(2)可看出局部密度  $\rho_i$  表示的是数据集  $X$  中与数据点  $x_i$  距离小于  $d_c$  的点的个数。其中截断距离  $d_c$  为人为选取, 对于大型数据集, 算法对于  $d_c$  的选取具有鲁棒性<sup>[13]</sup>。

高斯核函数为连续值, 不同数据点具有相同的局部密度值的概率很小, 将所有数据点的局部密度集  $\{\rho_i\}_{i=1}^N$  的一个降序排列下标序表示为  $\{q_i\}_{i=1}^N$ , 满足:

$$\rho_{q_1} \geq \rho_{q_2} \geq \dots \geq \rho_{q_N} \quad (3)$$

定义距离  $\delta_i$  为:

$$\delta_{q_i} = \begin{cases} \max_{j < i} \{d_{q_i, q_j}\}, & i \geq 2; \\ \max_{j \geq 2} \{\delta_{q_j}\}, & i = 1 \end{cases}, \quad i \neq j \quad (4)$$

对于每一个数据点  $x_i \in X$ , 都可通过计算得到 ( $\rho_i$ ,  $\delta_i$ )。将所有数据点在二维坐标平面图中画出即为决策图。可直观地看出同时具有较大的  $\rho$  值和  $\delta$  值的数据点为聚类中心。

为进一步突显聚类中心, 减少聚类中心选择中的人为干预, 将每个数据点的  $\rho_i$  和  $\delta_i$  分别进行归一化, 定义指标:

$$\gamma_i = \rho_i \delta_i, \quad i = I_s \quad (5)$$

对  $\{\gamma_i\}_{i=1}^N$  进行降序排列, 以  $\gamma$  为纵坐标, 以数据点下标为横坐标建立直角坐标系, 非聚类中心的  $\gamma$  值比较平滑, 聚类中心的值与非聚类中心的  $\gamma$  值存在明显的跃变。

在聚类中心确定后, 剩下的每个数据点被分配到与它最近的具有更高密度的点所属的类簇中。

#### 3.2 聚类算法的实现

##### 3.2.1 距离 $d_{ij}$ 的选取

距离  $d_{ij}$  用于评价不同样本间的差异度<sup>[15]</sup>。由于待聚类天气特征量均为统计指标, 以数值的方式体现, 因此只需考虑不同样本间统计指标在空间距离上的数值差异度, 引入欧氏距离最为合适, 其表达式为:

$$d_{ij} = \left[ \sum_{k=1}^n (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \quad (6)$$

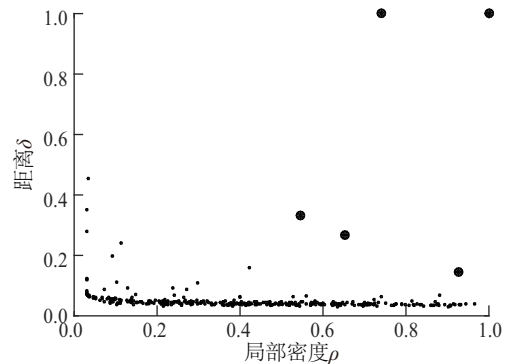
式中,  $x_{ik}$ 、 $x_{jk}$  ——样本  $x_i$  和  $x_j$  的第  $k$  维元素。

##### 3.2.2 截断距离 $d_c$ 的选取

基于若干数据集的测试, 一个合适的  $d_c$  使得每个数据点的平均邻居个数约为数据点总数的 1%~2%<sup>[12]</sup>。本文首先计算样本间的欧氏距离得到  $N$  个距离值, 然后对其按从小到大排序为  $d_1 \leq d_2 \leq \dots \leq d_N$ 。取截断距离  $d_c = d_m$ , 下标  $m = \lceil 0.02N \rceil$ ,  $\lceil \cdot \rceil$  为取整函数。

#### 3.3 聚类决策图分析

按第 2 节所述方法得到天气特征向量  $\mathbf{W}$ , 聚类得到结果如图 3。在图 3a 中有 5 个点同时具有相对较大的  $\rho$  值和  $\delta$  值, 在图 3b 中可看出这 5 个点与其他点在  $\gamma \approx 0.13$  处有明显跃变。可作为聚类中心点, 并得到聚类数  $\eta = 5$ 。



a. 决策图

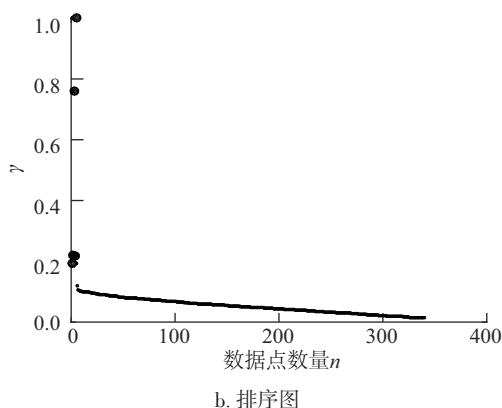


图3 聚类结果图

Fig. 3 Diagrams of clustering results

各类簇的日平均发电功率曲线如图4所示,聚类中心点的天气特征向量见表2。类簇1的风速3阶差分最大值  $w_{\text{dms}}'''$  和辐照度3阶差分最大值  $I_{\text{ms}}'''$  均较小,而辐照度最大值  $I_{\text{mx}}$ 、平均温度  $T_{\text{m}}$  和温度最大值  $T_{\text{mx}}$  均较大,具有晴天、风速波动小而无云层遮挡的理想天气条件,其相应的日平均发电功率也是5类中的最大值。而类簇4辐照度最大值  $I_{\text{mx}}$  和平均温度  $T_{\text{m}}$  均为所有类簇中的最小,具备温度低、阴雨天且风速波动较大的特征,对应的日平均发电功率最低。

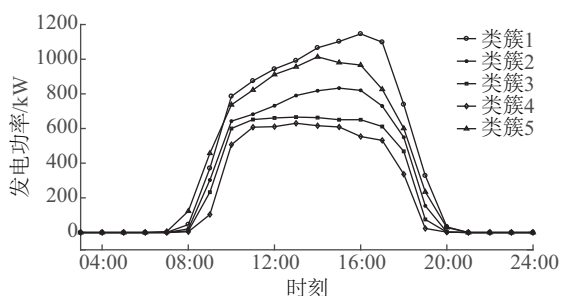


图4 聚类中心日平均发电功率

Fig. 4 Daily average PV power of cluster center

表2 聚类中心天气特征向量

Table 2 Weather feature vector of cluster center

参数	$w_{\text{dms}}'''$	$I_{\text{mx}}$	$I_{\text{ms}}'''$	$T_{\text{m}}$	$T_{\text{mx}}$	$w_{\text{msm}}$
类簇1	0.847	1064.925	0.144	28.919	35.907	4.387
类簇2	0.966	910.447	0.832	20.955	28.533	3.529
类簇3	2.036	794.427	0.281	16.025	24.987	2.579
类簇4	1.416	696.650	0.213	12.220	19.450	4.246
类簇5	1.144	1107.286	0.863	26.920	36.407	4.263

## 4 广义回归神经网络

广义回归神经网络自身具有很强的非线性映射能

力以及高度的容错性和鲁棒性<sup>[16]</sup>。在样本数据较少时也有很好的预测效果,在光伏发电功率短期及超短期预测有着很好的适用性。

### 4.1 神经网络结构

本文所用 GRNN 神经网络结构分为输入层、模式层、求和层和输出层共4层。输入层设6个神经元,分别对应天气特征向量的6个因子。输出层设24个神经元,输出预测日24 h光伏发电功率预测值。

### 4.2 神经网络训练

#### 4.2.1 K折交叉验证

由于部分类簇的样本数量较少,本文采用交叉验证的方法训练神经网络。根据每一类簇的具体样本数确定K折交叉验证,将样本分割为K个子样本,轮流将一个子样本作为测试集,其余K-1个子样本作为训练集,重复K次。分别建立神经网络,求得误差最小的作为最优输入、输出。

#### 4.2.2 SPREAD的最优值选择

SPREAD为径向基函数的扩展系数。作为调节GRNN神经网络的关键参数,其合理的取值对预测结果会产生直接影响。SPREAD值应该足够大,这样使神经元能对输入向量所覆盖的区域都能产生相应,但太大的SPREAD值会导致数值计算上的困难,同时更大的SPREAD值会使网络在数据样本的逼近结果越光滑,误差也相应越大。因此,本文在设计神经网络时,为更严格的对数据进行拟合,循环选取SPREAD,以确定最优值。

#### 4.2.3 实例验证

以类簇4为例,共92个样本,进行4折交叉验证,在每次交叉验证过程中SPREAD分别从0.1取到2.0,步长设为0.1。以均方误差MSE作为网络输出结果的评价指标。交叉验证训练结果如表3。由表3可看出,4折交叉验证训练在第1次交叉验证、SPREAD取0.3时MSE为最小值。故对类簇4,采用第一次交叉验证时所用训练集、SPREAD取0.3,构建的GRNN神经网络预测效果最好。

## 5 预测结果分析与模型对比

以澳大利亚某太阳能技术示范区2017年全年的实际运行数据为基础,对未来24 h系统发电功率预测。分别与K均值+BP、K均值+GRNN的预测模型进行对比。



表3 类簇4交叉验证训练结果  
Table 3 Cross verification results of cluster 4

SPREAD	MSE			
	第1次	第2次	第3次	第4次
0.1	2892.99	2698.25	3895.89	4197.31
0.2	2211.93	2299.42	3647.49	3856.50
0.3	<b>1902.16</b>	2053.66	3282.38	3589.59
0.4	1961.52	2037.80	2886.71	3479.46
0.5	2104.49	2117.50	2620.24	3407.51
0.6	2237.83	2206.61	2470.74	3392.93
0.7	2355.04	2283.67	2360.59	3500.01
0.8	2468.03	2338.59	2271.41	3742.66
0.9	2573.50	2373.91	2214.88	4093.69
1.0	2666.53	2396.73	2200.01	4495.89
1.1	2746.98	2413.42	2227.21	4889.55
1.2	2817.08	2428.18	2290.94	5239.46
1.3	2879.17	2443.24	2382.01	5535.39
1.4	2935.01	2459.55	2489.93	5780.57
1.5	2985.81	2477.27	2605.14	5982.80
1.6	3032.37	2496.19	2720.34	6150.19
1.7	3075.23	2515.91	2830.75	6289.74
1.8	3114.77	2536.04	2933.74	6407.09
1.9	3151.28	2556.20	3028.15	6506.66
2.0	3185.02	2576.09	3113.81	6591.85

注:表中粗体为正相关且相关程度较大者。

### 5.1 预测结果分析

为评价预测结果,本文采用平均绝对误差百分比(MAPE)和方均根误差(RMSE)来衡量系统的预测性能。

它们分别定义为:

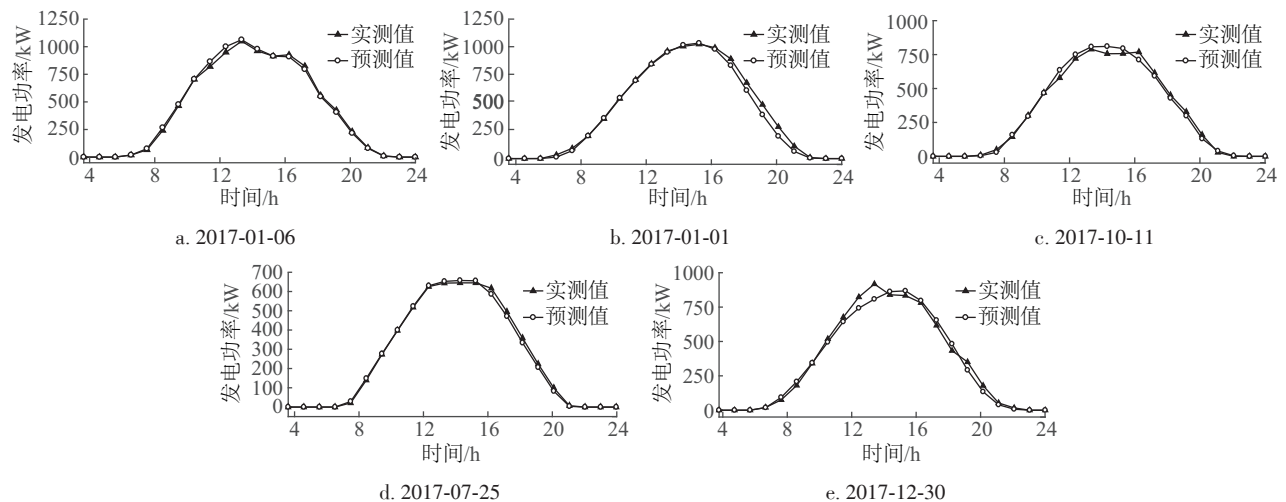


图5 预测结果与实测对比

Fig. 5 Comparison of forecast results with measurements

$$\varepsilon_{MAPE} = \frac{1}{M} \sum_{i=1}^M \left| \frac{P_p(i) - P_m(i)}{P_m(i)} \right| \times 100\% \quad (7)$$

$$\varepsilon_{RMSE} = \frac{\sqrt{\frac{1}{M} \sum_{i=1}^M [P_p(i) - P_m(i)]^2}}{\sqrt{\frac{1}{M} \sum_{i=1}^M P_m^2(i)}} \times 100\% \quad (8)$$

式中,  $P_p(i)$ ——系统发电功率预测值;  $P_m(i)$ ——系统发电功率实际测量值。  $M$  取 24, 为一天的实际预测点个数。预测误差统计指标结果见表 4, 可看出对于聚类所得的 5 个类簇分别建立的 GRNN 神经网络均具有较高的预测精度, 不同类簇预测模型间还存在着预测效果上的差异。从各类簇中分别随机取 2017 年 1 月 6 日(类簇 1)、2017 年 1 月 1 日(类簇 2)、2017 年 10 月 11 日(类簇 3)、2017 年 7 月 25 日(类簇 4)、2017 年 12 月 30 日(类簇 5)做光伏发电功率预测曲线与实测曲线对比图如图 5 所示, 可看出预测模型对不同类簇均有较高的预测精度。但由于短期发电功率预测的局限性, 无法对云层实时运动遮挡阳光导致的发电功率剧烈波动做出响应, 如 10 月 11 日午时的波动、12 月 30 日 13:00 的波动, 1 月 1 日 14:00 后天气晴转多云后实际出力的下降等现象, 均出现预测偏差。

表4 各类簇的预测效果

Table 4 Prediction effect of various clusters

类簇	平均绝对误差百分比 $\varepsilon_{MAPE} / \%$	方均根误差 $\varepsilon_{RMSE} / \%$
1	10.54	5.32
2	10.97	6.14
3	12.54	6.70
4	10.47	5.29
5	13.10	7.99

## 5.2 预测模型对比

用相同的数据集,分别使用  $K$  均值+BP 神经网络的预测方法、 $K$  均值+GRNN 神经网络的预测方法进行建模预测,同样采用  $MAPE$  和  $RMSE$  作为衡量预测效果的指标。

对于  $K$  均值聚类算法的聚类数采用 DB 评价指标来确定,其值越小,聚类效果越好。由图可得将聚类数目确定为 4 最优。

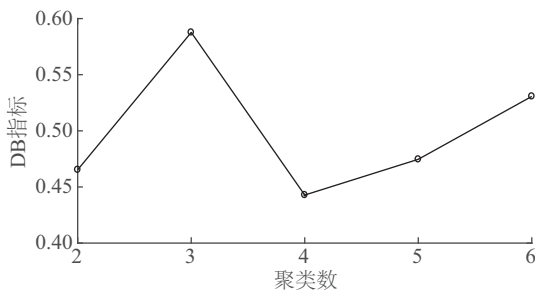


图6 不同聚类数的DB评价指标

Fig. 6 DB evaluation index of different clustering numbers

各模型预测结果指标对比如表 5 所示,用 3 种不同的预测模型对 2017 年 10 月 11 日的光伏发电功率进行预测,预测结果如图 7 所示。由指标对比表和预测曲线对比图可看出,本文方法相较于传统  $K$  均值聚类与 BP 神经网络结合的方法具有更高的预测精度。

表5 各模型预测结果

Table 5 Prediction result of different models

预测模型	平均绝对误差 百分比 $\varepsilon_{MAPE}/\%$	方均根误差 $\varepsilon_{RMSE}/\%$
密度峰值+GRNN	12.54	6.70
$K$ 均值+GRNN	17.67	12.79
$K$ 均值+BP	23.58	16.83

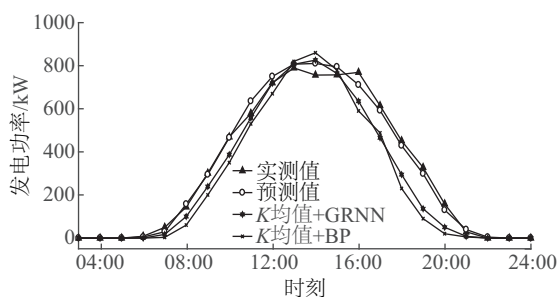


图7 2017年10月11日不同模型预测曲线

Fig. 7 Prediction PV power curve of different models

## 6 结论

本文首先对一年的光伏系统发电功率数据进行分

析,提出分类构建预测模型的基本思想。然后对气象数据进行相关性分析,采集特征因子构建天气特征向量,采用密度峰值聚类的方法将天气特征量聚类,分别针对每一类簇构建 GRNN 神经网络发电功率预测模型,对仿真结果分析得到以下结论:

1)密度峰值聚类可有效地运用于历史天气聚类,克服传统  $K$  均值聚类算法需要人为指定类别数和聚类中心的缺点,作为发电功率预测前的天气聚类有很好的适用性。

2)根据不同类簇样本数的不同选择  $K$  折交叉验证训练模型,循环选  $SPREAD$  值,最终取得最优值构建 GRNN 神经网络,相较于随机选取训练样本的训练方法具有更高的预测精度,对于小样本也有较好的适用性。

## [参考文献]

- [1] 戚军, 张晓峰, 张有兵, 等. 考虑阴影影响的光伏阵列仿真算法研究[J]. 中国电机工程学报, 2012, 32(32): 131-138.  
QI J, ZHANG X F, ZHANG Y B, et al. Study on simulation algorithm of PV array considering shade effect [J]. Proceedings of the CSEE, 2012, 32(32): 131-138.
- [2] 王飞, 米增强, 杨奇逊, 等. 基于神经网络与关联数据的光伏电站发电功率预测方法[J]. 太阳能学报, 2012, 33(7): 1171-1177.  
WANG F, MI Z Q, YANG Q X, et al. Power forecasting approach of PV plant based on ANN and relevant data [J]. Acta energiae solaris sinica, 2012, 33(7): 1171-1177.
- [3] HOYER-KLICK C, BEYER H G, DUMORTIER D, et al. Management and Exploitation of solar resource knowledge [C]//8th EuroSun Conference of ISES Europe, Graz, Austria, 2010: 159-162.
- [4] 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.  
GONG Y F, LU Z X, QIAO Y, et al. An overview of photovoltaic energy system output forecasting technology [J]. Automation of electric power systems, 2016(4): 140-151.
- [5] ANTONANZAS J, OSORIO N, ESCOBAR R, et al. Review of photovoltaic power forecasting[J]. Solar energy, 2016, 136: 78-111.
- [6] GRADITI G, FERLITO S, ADINOLFI G. Comparison of photovoltaic plant power production prediction methods using a large measured dataset [J]. Renewable

- energy, 2016, 90: 513-519.
- [7] BOUZERDOUM M, MELLIT A, PAVAN A M. A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant[J]. Solar energy, 2013, 98(4): 226-235.
- [8] 丁明, 徐宁舟. 基于马尔可夫链的光伏发电系统输出功率短期预测方法[J]. 电网技术, 2011(1): 152-157.
- DING M, XU N Z. A method to forecast short-term output power of photovoltaic generation system based on Markov chain[J]. Power system technology, 2011(1): 152-157.
- [9] 袁晓玲, 施俊华, 徐杰彦. 计及天气类型指数的光伏发电短期出力预测[J]. 中国电机工程学报, 2013, 33(34): 57-64.
- YUAN X L, SHI J H, XU J Y. Short-term power forecasting for photovoltaic generation considering weather type index[J]. Proceedings of the CSEE, 2013, 33(34): 57-64.
- [10] 程泽, 李思宇, 韩丽洁, 等. 基于数据挖掘的光伏阵列发电预测方法研究[J]. 太阳能学报, 2017, 38(3): 726-733.
- CHENG Z, LI S Y, HAN L J, et al. PV power generation forecast based on data mining method[J]. Acta energiae solaris sinica, 2017, 38(3): 726-733.
- [11] The Desert Knowledge Australia Solar Centre. The Yulara PV plant [DB/OL]. <http://dkasolarcentre.com.au/>, 2018-3-28/2018-2-20.
- [12] 廖卫强, 张认成, 俞万能, 等. 基于相似样本及PCA的光伏输出功率预测[J]. 太阳能学报, 2016, 37(9): 2377-2385.
- LIAO W Q, ZHANG R C, YU W N, et al. Prediction of output power of photovoltaic based on similar samples and principal component analysis[J]. Acta energiae solaris sinica, 2016, 37(9): 2377-2385.
- [13] RODRIGUEZ A, LAIO A. Machine learning, clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492.
- [14] 皮果提. 发表在 Science 上的一种新聚类算法[EB/OL]. <http://blog.csdn.net/itplus/article/details/38926837>.
- [15] 刘兴杰, 岑添云, 郑文书, 等. 基于模糊粗糙集与改进聚类的神经网络风速预测[J]. 中国电机工程学报, 2014, 34(19): 3162-3169.
- LIU X J, CEN T Y, ZHENG W S, et al. Neural network wind speed prediction based on fuzzy rough set and improved clustering[J]. Proceedings of the CSEE, 2014, 34(19): 3162-3169.
- [16] 董聪, 柳进, 刘广一, 等. 基于GRNN神经网络的光伏发电预测方法: CN 104050517 A[P]. 2014.
- DONG C, LIU J, LIU G Y, et al. Prediction of photovoltaic power generation based on GRNN neural network: CN 104050517 A[P]. 2014.

## PV POWER FORECASTING APPROACH BASED ON DENSITY PEAKS CLUSTERING AND GENERAL REGRESSION NEURAL NETWORK

Yao Hongmin, Du Xinhui, Qin Wenping

(College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract:** A short-term forecasting method for PV power based on density peaks clustering and General Regression Neural Network is proposed. Firstly, analyzing the characteristics of weather factors, extracting statistical indicators to form the weather feature vector; then, using the density peaks algorithm to cluster the historical meteorological data to overcome the K-means algorithm, which is easy to fall into the local optimum and artificially specify the cluster center, etc. Simulation results show that the proposed method improves the accuracy of prediction and still has good applicability in small sample environments.

**Keywords:** cluster analysis; neural networks; prediction; big data; photovoltaic