



(12) 发明专利申请

(10) 申请公布号 CN 115759389 A

(43) 申请公布日 2023. 03. 07

(21) 申请号 202211423091.0

(22) 申请日 2022.11.15

(71) 申请人 山西格盟中美清洁能源研发中心有限公司

地址 030032 山西省太原市综改示范区太原科技创新城化章北街1号

(72) 发明人 叶泽甫 崔京港 张帅 徐旭东 马跃 阎高伟

(74) 专利代理机构 深圳市兴科达知识产权代理有限公司 44260

专利代理师 潘月仙

(51) Int. Cl.

G06Q 10/04 (2023.01)

G06Q 50/06 (2006.01)

G06F 18/23 (2023.01)

G06F 18/2113 (2014.01)

G06N 3/126 (2023.01)

G06N 5/01 (2023.01)

G06N 20/20 (2023.01)

H02J 3/00 (2019.01)

H02J 3/38 (2012.01)

H02S 40/30 (2006.01)

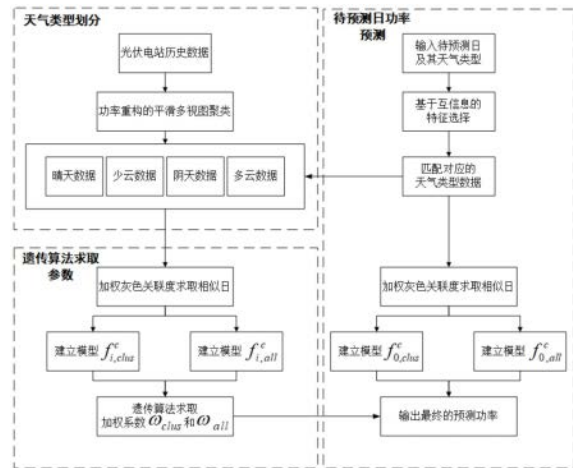
权利要求书4页 说明书14页 附图5页

(54) 发明名称

基于天气类型的相似日组合策略的日前光伏功率预测方法

(57) 摘要

本发明公开了基于天气类型的相似日组合策略的日前光伏功率预测方法,包括以下步骤:(1)对已有的光伏功率历史日数据集进行最小最大归一化处理;(2)计算各气象特征对功率的MIV值;(3)对天气类型一致的历史日进行平滑表示多视图聚类;(4)对各天气类型的功率数据的类内距离进行约束;(5)通过交替迭代求解聚类标签;(6)对各个数据集气象特征基于互信息进行特征选择;(7)对各天气类型下的历史日依次构建类内相似日集和全数据相似日集;(8)求取两模型预测值的加权系数;(9)获取最终的功率预测值;本发明充分应用了真实的功率特征信息,可以显著提高波动程度大的天气类型的功率预测精度。



1. 基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,包括以下步骤:

- (1) 对已有的光伏功率历史日数据集进行最小最大归一化处理;
- (2) 利用步骤(1)处理后的数据计算各气象特征对功率的MIV值,并取其绝对值作为多视图中各气象视图的权重;
- (3) 根据步骤(1)处理后的数据,对天气类型一致的历史日进行平滑表示多视图聚类;
- (4) 对步骤(2)的功率视图进行重构,结合步骤(3)得到的天气类型标签阵对各天气类型的功率数据的类内距离进行约束;
- (5) 结合步骤(4)的功率视图重构和平滑表示多视图聚类,通过交替迭代求解聚类标签,对历史日聚类为晴天、少云、阴天和多云天气类型;
- (6) 对步骤(5)聚类的气象特征进行新特征构建,并依据聚类结果将数据集划分为晴天数据集、少云数据集、阴天数据集、多云数据集和全历史数据集,对各个数据集气象特征基于互信息进行特征选择;
- (7) 结合步骤(6)的数据,按照加权灰色关联度的相似度准则选择相似日,对各天气类型下的历史日依次构建类内相似日集和全数据相似日集;
- (8) 对步骤(7)构建的历史日的两组相似日集分别建立GBDT模型获取预测值,应用遗传算法求取两模型预测值的加权系数;
- (9) 由天气预报获取待预测日天气类型,应用该日的NWP数据获取其类内相似日集和全数据相似日集,利用步骤(8)建立的模型,通过该天气类型下的模型加权系数获取最终的功率预测值。

2. 如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,步骤(1)中的光伏功率历史日数据集包括NWP气象数据和对应时刻的功率数据。

3. 如权利要求2所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(3)具体包括:

将历史日的NWP数据和功率数据写为视图形式 $\{V_k^{n \times l}\}_{k=1}^{q+1}$ ,其中,当 $k=1, \dots, q$ 时, $V_k = [X_{1,k} \dots X_{n,k}]^T$ 表征气象视图; $V_{q+1} = [y_1 \dots y_n]^T$ 表征功率视图;

对特定历史日的表征气象视图、表征功率视图进行重构,描述为:

$$\sum_{i=1}^{q+1} \sum_{j=1}^n r_i \left\| v_{i,j} - \sum_{k=1}^n v_{i,j} z_{jk} \right\|_2^2 = \sum_{i=1}^{q+1} r_i \|V_i - ZV_i\|_F^2 \quad (1)$$

其中, $v_{i,j} \in \mathbb{R}^{1 \times 1}$ 表示 $i$ 视图的日序列 $j$ , $z_{jk}$ 表示历史日 $j$ 和历史日 $k$ 的相似性大小为自表示阵系数; $r_i$ 表示不同视图的权重系数;同时显示地对自表示系数阵应用分组效应约束:

$$\frac{1}{2} \sum_{i=1}^{q+1} \sum_{j=1}^n \sum_{k=1}^n r_i w_{i,jk} \|z_j - z_k\|_2^2 = \sum_{i=1}^{q+1} \text{tr}(ZL_i Z^T),$$

$$\text{s.t. } w_{i,jk} = \begin{cases} 1, & \text{if } v_{i,k} \in \text{Neighbor}(v_{i,j}) \\ 0, & \text{if } v_{i,k} \notin \text{Neighbor}(v_{i,j}) \end{cases}, d_{i,jk} = \begin{cases} \sum_{k=1}^n w_{i,jk}, & i=k \\ 0, & i \neq k \end{cases}, L_i = D_i - W_i \quad (2)$$

定义亲和矩阵 $S = (Z+Z^T)/2$ ,应用谱聚类求解得到天气类型标签阵 $F \in \mathbb{R}^{n \times C}$ , $c$ 为聚类得到的天气类型个数, $F_{ij} = 1$ 表示历史日 $i$ 属于天气类型 $j$ , $F_{ij} = 0$ 表示历史日 $i$ 不属于天气类

型j,且 $\sum_{j=1}^C F_{ij} = 1$

4.如权利要求3所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(4)具体包括:

将功率视图表示为:

$$\sum_{i=1}^n \|v_{q+1,i} - v_{q+1,i} PP^T\|_2^2 = \|V_{q+1} - V_{q+1} PP^T\|_F^2, s.t. P^T P = I \quad (3)$$

其中, $P \in R^{1 \times m}$ 为投影阵;同时对各天气类型的功率数据的类内距离进行约束,具体描述为:

$$\sum_{i=1}^C \sum_{j=1}^{n_i} \|P^T v_{q+1,j}^i - o_i\|_2^2 = \|V_{q+1} P\|_F^2 - tr(\hat{F}^T V_{q+1} PP^T V_{q+1}^T \hat{F}) s.t. o_i = \frac{1}{n_i} \sum_{j=1}^{n_i} P^T v_{q+1,j}^i, \hat{F} = F \text{diag}(\frac{1}{\sqrt{n_1}} \dots \frac{1}{\sqrt{n_c}}) \quad (4)$$

其中, $o_i$ 是属于天气类型 $i$ 的功率序列中心, $n_i$ 是属于天气类型 $i$ 的历史日天数, $\hat{F}$ 是经过系数调整之后的天气类型标签阵。

5.如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(5)具体包括:

确定以下总目标函数:

$$\begin{aligned} \min_{Z,P} Loss = & \min_{Z,P} \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \\ & + \|V_{q+1} - V_{q+1} PP^T\| + \beta (\|V_{q+1} P\|_2^2 - tr(\hat{F}^T V_{q+1} PP^T V_{q+1}^T \hat{F})) \\ \Rightarrow & \min_{Z,P} \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \\ & + P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P \end{aligned} \quad (5)$$

其中, $r_i$ 是各视图的权重系数,功率视图权重系数 $r_{q+1}$ 默认为1,其余视图特征权重采用MIV方式计算得到的MIV绝对值; $\alpha$ 是平滑表示多视图的超参; $\beta$ 是功率重构的超参;

采用交替的方式求解:

固定投影阵P,求解自表示阵Z和天气类型阵F,即对下式进行求解:

$$\min_Z \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \quad (6)$$

对其求导,并设置导数为0,有:

$$\alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} PP^T V_{q+1}^T) Z + Z \sum_{i=1}^q r_i L_i + r_{q+1} L_{q+1} = \alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} PP^T V_{q+1}^T) \quad (7)$$

由该式得到Z,再计算亲和矩阵S,并应用谱聚类算法可以求解天气类型阵F,并得到系数调整后的天气类型标签 $\hat{F}$ ;

固定自表示系数阵Z,求解投影阵P,即对下式进行求解:

$$\min_P P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P \quad (8)$$

解为 $V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1}$ 对应的前m个最小特征值对应的特征向量组成的矩阵。

6.如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(6)具体包括:

新特征构建:计算各气象特征的一阶后向差分和二阶后向差分,挖掘气象特征动态信息,扩充特征集;

基于互信息的特征选择:在给定离散随机变量X和Y后,X的边缘分布为 $p(x)$ ,Y的边缘分布为 $p(y)$ ,两者的联合概率分布为 $p(x,y)$ ;此时X和Y之间的互信息为

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

计算各数据集的特征对功率的互信息大小,并对互信息由大到小排列,选取前u个特征作为该数据集相似日选择和功率预测模型建立的特征。

7.如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(7)的计算过程为:

定义经过特征构建和特征选择后的历史日数据为 $X_i = [X_{i,1} \cdots X_{i,u}] \in \mathbb{R}^{1 \times u}$ ,计算历史日 $X_0$ 和历史日 $X_i$ 的关于特征k的灰色关联系数 $\varepsilon_{ik}$ :

$$\varepsilon_{ik}(t) = \frac{\min_{ik} \min_t \Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}{\Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}, \quad s.t. \quad t = 1, \dots, L, \Delta_{ik}(t) = |X_{0,k}(t) - X_{i,k}(t)| \quad (9)$$

其中, $X_{i,k}(t)$ 表征气象特征k第t时刻的值,分辨系数 $\mu = 0.5$ ;

对应的加权灰色关联度 $\zeta_i$ 的计算式为:

$$\zeta_i = \sum_{k=1}^M w_k \left( \frac{1}{L} \sum_{t=1}^L \varepsilon_{ik}(t) \right) \quad s.t. \quad \sum_{k=1}^M w_k = 1 \quad (10)$$

其中, $w_k$ 为气象特征k对灰度系数的权重,该权重应用CRITIC方法计算得出,采用特征间的对比强度和冲突性来计算,特征k的标准差 $\sigma_k$ 表现对比强度,相关系数 $\rho_{ki}$ 衡量冲突性,计算式为:

$$\begin{cases} C_k = \sigma_k \sum_{i=1}^M (1 - |\rho_{ki}|), k = 1, \dots, u \\ w_k = C_k / \sum_{k=1}^M C_k \end{cases} \quad (11)$$

计算历史日 $X_0$ 和数据集内的所有历史日计算灰色关联度,对其由大到小排列,选取前若干个相似日构成历史日 $X_0$ 在该数据集下的相似日集;

定义类内相似日集和全数据相似日集:定义待预测日 $X_0$ 在同天气类型历史日中选择的相似日数据为类内相似日集 $D_{0,clus}$ ,在全部历史日中选择的相似日数据为全数据相似日集 $D_{0,all}$ 。

8.如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法,其特征在于,所述步骤(8)中:

GBDT模型表示为:

$$f(X) = \sum_{t=1}^T \alpha_t h_t(X) \quad (12)$$

其中,X为输入样本, $h_t(X)$ 为分类回归树,T是GBDT中需要构建的树的数量, $\alpha_t$ 是第t棵树的权重;

GBDT采用前向分布算法,首先确定 $F_0(x)$ 为模型初始值为常数,第m步的模型是:

$$f_m(X) = f_{m-1}(X) + \alpha_m h_m(X) \quad (13)$$

其中,  $F_{m-1}(x)$  为当前模型; 新添加的分类回归树  $h_m(x)$  通过最小化损失函数求得:

$$h_m = \arg \min_h \sum_{i=1}^N L(y_i, f_{m-1}(X_i) + h(X_i)) \quad (14)$$

其中,  $N$  是样本个数, GBDT 采用梯度下降法来求解最优模型, 将损失函数在当前模型  $f_{m-1}(X)$  的负梯度值作为梯度下降的方向:

$$F_m(X) = f_{m-1} - \alpha_m \sum_{i=1}^N \nabla_F L(y_i, f_{m-1}(X_i)) \quad (15)$$

其中,  $\alpha_m$  通过线性搜索求得:

$$\alpha_m = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, f_{m-1}(x_i)) - \alpha \frac{\partial L(y_i, f_{m-1}(X_i))}{\partial f_{m-1}(X_i)} \quad (16)$$

遗传算法求取模型加权系数: 经过特征构建的天气类型  $c$  的气象特征数据记为  $X_i^c \in R^{l \times u}$ , 功率数据记为  $y_i^c \in R^l$ , 由对应的类内相似日集  $D_{i,clus}^c$  和全数据相似日集  $D_{i,a}^c$  分别建立 GBDT 模型  $f_{i,c}^c$  和  $f_{i,all}^c$ , 最终的功率预测值为  $\hat{y}_i = w_{clus}^c f_{i,clus}^c(X_i^c) + w_{all}^c f_{i,all}^c(X_i^c)$ , 其中  $w_{clus}^c$  和  $w_{all}^c$  为该天气类型下两类模型的加权系数; 确定预测目标函数作为遗传算法的适应度函数, 设定种群大小和迭代轮次, 求取对应天气类型下待预测日建立的两模型的加权系数, 具体为:

$$\min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - \hat{y}_i\|_2^2 = \min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - w_{clus}^c f_{i,clus}^c(X_i^c) - w_{all}^c f_{i,all}^c(X_i^c)\|_2^2, s.t. 0 \leq w_{clus}^c, w_{all}^c \leq 1 \quad (17)$$

其中:  $n_c$  为该天气类型的历史日个数。

9. 如权利要求1所述的基于天气类型的相似日组合策略的日前光伏功率预测方法, 其特征在于, 所述步骤(9)具体包括:

获取待预测日天气类型的待预测日气象特征记为  $X_0^c$ , 利用步骤(8)的方式分别类内相似日集和全数据相似日集建立模型  $f_{0,clus}^c$  和  $f_{0,all}^c$ , 应用天气类型  $c$  的加权模型系数, 确定功率预测值  $\hat{y}_0$ :

$$\hat{y}_0 = w_{clus}^c f_{0,clus}^c(X_0^c) + w_{all}^c f_{0,all}^c(X_0^c) \quad (18)。$$

## 基于天气类型的相似日组合策略的日前光伏功率预测方法

### 技术领域

[0001] 本发明涉及光伏功率预测技术领域,具体涉及基于天气类型的相似日组合策略的日前光伏功率预测方法。

### 背景技术

[0002] 工业化导致能源需求不断加剧,太阳能作为一种清洁的可再生能源,是应对能源安全与环境污染矛盾之间的有力手段,是实现绿色低碳发展的关键途径。我国光伏产业发展迅速,2021年新增装机容量达5493万千瓦,截至2021年12月底,光伏发电装机容量约3.1亿千瓦。

[0003] 然而光伏发电功率具有较强的间歇性、随机性和波动性,随着光伏发电在传统电网中装机容量比例上升,为电网能源管理和电力系统调度带来挑战,准确预测光伏功率是保证系统安全运行的关键。传统的光伏发电功率日前预测大多采用确定性估计方法,针对历史数据中不存在天气类型标签的情况,应用历史NWP数值天气预报信息来划分天气类型,建立对应模型完成待预测日天气类型识别和功率预测。但是NWP数据不准确无法对历史日天气类型进行准确划分,导致在功率波动程度大的天气类型下预测精度不高。

### 发明内容

[0004] 针对上述存在的技术不足,本发明的目的是提供基于天气类型的相似日组合策略的日前光伏功率预测方法。

[0005] 为解决上述技术问题,本发明采用如下技术方案:

[0006] 本发明提供基于天气类型的相似日组合策略的日前光伏功率预测方法,包括以下步骤:

[0007] (1) 对已有的光伏功率历史日数据集进行最小最大归一化处理;

[0008] (2) 利用步骤(1)处理后的数据计算各气象特征对功率的MIV值,并取其绝对值作为多视图中各气象视图的权重;

[0009] (3) 根据步骤(1)处理后的数据,对天气类型一致的历史日进行平滑表示多视图聚类;

[0010] (4) 对步骤(2)的功率视图进行重构,结合步骤(3)得到的天气类型标签阵对各天气类型的功率数据的类内距离进行约束;

[0011] (5) 结合步骤(4)的功率视图重构和平滑表示多视图聚类,通过交替迭代求解聚类标签,对历史日聚类为晴天、少云、阴天和多云天气类型;

[0012] (6) 对步骤(5)聚类的气象特征进行新特征构建,并依据聚类结果将数据集划分为晴天数据集、少云数据集、阴天数据集、多云数据集和全历史数据集,对各个数据集气象特征基于互信息进行特征选择;

[0013] (7) 结合步骤(6)的数据,按照加权灰色关联度的相似度准则选择相似日,对各天气类型下的历史日依次构建类内相似日集和全数据相似日集;

[0014] (8) 对步骤(7)构建的历史日的两组相似日集分别建立GBDT模型获取预测值,应用遗传算法求取两模型预测值的加权系数;

[0015] (9) 由天气预报获取待预测日天气类型,应用该日的NWP数据获取其类内相似日集和全数据相似日集,利用步骤(8)建立的模型,通过该天气类型下的模型加权系数获取最终的功率预测值。

[0016] 优选地,步骤(1)中的光伏功率历史日数据集包括NWP气象数据和对应时刻的功率数据。

[0017] 优选地,所述步骤(3)具体包括:

[0018] 将历史日的NWP数据和功率数据写为视图形式  $\{V_k^{n \times l}\}_{k=1}^{q+1}$ , 其中, 当  $k=1, \dots, q$  时,  $V_k =$

$[X_{1,k} \dots X_{n,k}]^T$  表征气象视图;  $V_{q+1} = [y_1 \dots y_n]^T$  表征功率视图;

[0019] 对特定历史日的表征气象视图、表征功率视图进行重构,描述为:

$$[0020] \quad \sum_{i=1}^{q+1} \sum_{j=1}^n r_i \left\| v_{i,j} - \sum_{k=1}^n v_{i,j} z_{jk} \right\|_2^2 = \sum_{i=1}^{q+1} r_i \|V_i - ZV_i\|_F^2 \quad (1)$$

[0021] 其中,  $v_{i,j} \in \mathbb{R}^{1 \times l}$  表示  $i$  视图的日序列  $j$ ,  $z_{jk}$  表示历史日  $j$  和历史日  $k$  的相似性大小为自表示阵系数;  $r_i$  表示不同视图的权重系数; 同时显示地对自表示系数阵进行应用分组效应约束:

$$[0022] \quad \frac{1}{2} \sum_{i=1}^{q+1} \sum_{j=1}^n \sum_{k=1}^n r_i w_{jk}^i \|z_j - z_k\|_2^2 = \sum_{i=1}^{q+1} \text{tr}(ZL_i Z^T),$$

$$s.t. \quad w_{i,jk} = \begin{cases} 1, & \text{if } v_{i,k} \in \text{Neighbor}(v_{i,j}) \\ 0, & \text{if } v_{i,k} \notin \text{Neighbor}(v_{i,j}) \end{cases}, d_{i,jk} = \begin{cases} \sum_{k=1}^n w_{i,jk}, & i=k \\ 0, & i \neq k \end{cases}, L_i = D_i - W_i \quad (2)$$

[0023] 取相似阵  $S = (Z + Z^T) / 2$ , 通过谱聚类算法求解得到天气类型标签阵  $F \in \mathbb{R}^{n \times C}$ ,  $c$  为聚类得到的天气类型个数,  $F_{ij} = 1$  表示历史日  $i$  属于天气类型  $j$ ,  $F_{ij} = 0$  表示历史日  $i$  不属于天气类型  $j$ , 且  $\sum_{j=1}^C F_{ij} = 1$ 。

[0024] 优选地,所述步骤(4)具体包括:

[0025] 将功率视图表示为:

$$[0026] \quad \sum_{i=1}^n \|v_{q+1,i} - v_{q+1,i} P P^T\|_2^2 = \|V_{q+1} - V_{q+1} P P^T\|_F^2, s.t. \quad P^T P = I \quad (3)$$

[0027] 其中,  $P \in \mathbb{R}^{1 \times n}$  为投影阵; 同时对各天气类型的功率数据的类内距离进行约束,具体描述为:

$$[0028] \quad \sum_{i=1}^C \sum_{j=1}^{n_i} \|P^T v_{q+1,j} - o_i\|_2^2 = \|V_{q+1} P\|_F^2 - \text{tr}(\hat{F}^T V_{q+1} P P^T V_{q+1} \hat{F}) s.t. \quad o_i = \frac{1}{n_i} \sum_{j=1}^{n_i} P^T v_{q+1,j}, \hat{F} = F \text{diag}\left(\frac{1}{\sqrt{n_1}} \dots \frac{1}{\sqrt{n_C}}\right) \quad (4)$$

[0029] 其中,  $o_i$  是属于天气类型  $i$  的功率序列中心,  $n_i$  是属于天气类型  $i$  的历史日天数,  $\hat{F}$  是经过系数调整之后的天气类型标签阵。

[0030] 优选地,所述步骤(5)具体包括:

[0031] 确定以下总目标函数:

$$\begin{aligned}
\min_{Z,P} Loss &= \min_{Z,P} \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1}P - ZV_{q+1}P\|_F^2 + tr(ZL_{q+1} Z^T)) \\
&+ \|V_{q+1} - V_{q+1}PP^T\| + \beta (\|V_{q+1}P\|_2^2 - tr(\hat{F}^T V_{q+1}PP^T V_{q+1}^T \hat{F})) \\
[0032] \quad &\Rightarrow \min_{Z,P} \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1}P - ZV_{q+1}P\|_F^2 + tr(ZL_{q+1} Z^T)) \\
&+ P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P
\end{aligned} \tag{5}$$

[0033] 其中,  $r_i$  是各视图的权重系数, 功率视图权重系数  $r_{q+1}$  默认为1, 其余视图特征权重采用MIV方式计算得到的MIV绝对值;  $\alpha$  是平滑表示多视图的超参;  $\beta$  是功率重构的超参;

[0034] 采用交替的方式求解:

[0035] 固定投影阵  $P$ , 求解自表示阵  $Z$ , 即对下式进行求解:

$$\min_Z \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1}P - ZV_{q+1}P\|_F^2 + tr(ZL_{q+1} Z^T)) \tag{6}$$

[0037] 对其求导, 并设置导数为0, 有:

$$\alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} P P^T V_{q+1}^T) Z + Z \sum_{i=1}^q r_i L_i + r_{q+1} L_{q+1} = \alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} P P^T V_{q+1}^T) \tag{7}$$

[0039] 由该式得到  $Z$ , 再计算亲和矩阵  $S$ , 并应用谱聚类算法可以求解天气类型阵  $F$ , 并得到系数调整后的天气类型标签  $\hat{F}$ ; 固定自表示系数阵  $Z$ , 求解投影阵  $P$ , 即对下式进行求解:

$$\min_P P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P \tag{8}$$

[0041] 解为  $V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1}$  对应的前  $m$  个最小特征值对应的特征向量组成的矩阵。

[0042] 优选地, 所述步骤 (6) 具体包括:

[0043] 新特征构建: 计算各气象特征的一阶后向差分和二阶后向差分, 挖掘气象特征动态信息, 扩充特征集;

[0044] 基于互信息的特征选择: 在给定离散随机变量  $X$  和  $Y$  后,  $X$  的边缘分布为  $p(x)$ ,  $Y$  的边缘分布为  $p(y)$ , 两者的联合概率分布为  $p(x, y)$ ; 此时  $X$  和  $Y$  之间的互信息为

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}; \text{ 计算各数据集的特征对功率的互信息大小, 并对互信息}$$

由大到小排列, 选取前  $u$  个特征作为该数据集相似日选择和功率预测模型建立的特征。

[0045] 优选地, 所述步骤 (7) 的计算过程为:

[0046] 定义经过特征构建和特征选择后的历史日数据为  $X_i = [X_{i,1} \cdots X_{i,u}] \in \mathbb{R}^{1 \times u}$ , 计算历史日  $X_0$  和历史日  $X_i$  的关于特征  $k$  的灰色关联系数  $\varepsilon_{ik}$ :

$$\varepsilon_{ik}(t) = \frac{\min_{ik} \min_t \Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}{\Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}, \quad s.t. \quad t = 1, \dots, l, \Delta_{ik}(t) = |X_{0,k}(t) - X_{i,k}(t)| \tag{9}$$

[0048] 其中,  $X_{i,k}(t)$  表征气象特征  $k$  第  $t$  时刻的值, 分辨系数  $\mu = 0.5$ ;

[0049] 对应的加权灰色关联度  $\zeta_i$  的计算式为:

$$\zeta_i = \sum_{k=1}^u w_k \left( \frac{1}{l} \sum_{t=1}^l \varepsilon_{ik}(t) \right) \quad s.t. \quad \sum_{k=1}^u w_k = 1 \tag{10}$$

[0051] 其中,  $w_k$  为气象特征  $k$  对灰度系数的权重, 该权重应用CRITIC方法计算得出, 采用



特征间的对比强度和冲突性来计算,特征k的标准差 $\sigma_k$ 表现对比强度,相关系数 $\rho_{ki}$ 衡量冲突性,计算式为:

$$[0052] \quad \begin{cases} C_k = \sigma_k \sum_{i=1}^u (1 - |\rho_{ki}|), k = 1, \dots, u \\ w_k = C_k / \sum_{k=1}^u C_k \end{cases} \quad (11)$$

[0053] 计算历史日 $d_0$ 和数据集内的所有历史日计算灰色关联度,对其由大到小排列,选取前若干个相似日构成历史日 $d_0$ 在该数据集下的相似日集;

[0054] 定义类内相似日集和全数据相似日集:定义待预测日 $X_0$ 在同天气类型历史日中选择的相似日数据为类内相似日集 $D_{0,clus}$ ,在全部历史日中选择的相似日数据为全数据相似日集 $D_{0,all}$ 。

[0055] 优选地,所述步骤(8)中:

[0056] GBDT模型表示为:

$$[0057] \quad f(X) = \sum_{t=1}^T \alpha_t h_t(X) \quad (12)$$

[0058] 其中, $X$ 为输入样本, $h_t(X)$ 为分类回归树, $T$ 是GBDT中需要构建的树的数量, $\alpha_t$ 是第 $t$ 棵树的权重;

[0059] GBDT采用前向分布算法,首先确定 $F_0(x)$ 为模型初始值为常数,第 $m$ 步的模型是:

$$[0060] \quad f_m(X) = f_{m-1}(X) + \alpha_m h_m(X) \quad (13)$$

[0061] 其中, $F_{m-1}(x)$ 为当前模型;新添加的分类回归树 $h_m(x)$ 通过最小化损失函数求得:

$$[0062] \quad h_m = \arg \min_h \sum_{i=1}^N L(y_i, f_{m-1}(X_i) + h(X_i)) \quad (14)$$

[0063] 其中, $N$ 是样本个数,GBDT采用梯度下降法来求解最优模型,将损失函数在当前模型 $f_{m-1}(X)$ 的负梯度值作为梯度下降的方向:

$$[0064] \quad F_m(X) = f_{m-1} - \alpha_m \sum_{i=1}^N \nabla_F L(y_i, f_{m-1}(X_i)) \quad (15)$$

[0065] 其中, $\alpha_m$ 通过线性搜索求得:

$$[0066] \quad \alpha_m = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, f_{m-1}(x_i)) - \alpha \frac{\partial L(y_i, f_{m-1}(X_i))}{\partial f_{m-1}(X_i)} \quad (16)$$

[0067] 遗传算法求取模型加权系数:经过特征构建的天气类型 $c$ 的气象特征数据记为 $X_i^c \in R^{l \times u}$ ,功率数据记为 $y_i^c \in R^l$ ,由对应的类内相似日集 $D_{i,clus}^c$ 和全数据相似日集 $D_{i,a}^c$ 分别建立GBDT模型 $f_{i,c}^c$ 和 $f_{i,all}^c$ ,最终的功率预测值为 $\hat{y}_i = w_{clus}^c f_{i,clus}^c(X_i^c) + w_{all}^c f_{i,all}^c(X_i^c)$ ,其中 $w_{clus}^c$ 和 $w_{all}^c$ 为该天气类型下两类模型的加权系数;确定预测目标函数作为遗传算法的适应度函数,设定种群大小和迭代轮次,求取对应天气类型下待预测日建立的两模型的加权系数,具体为:

$$[0068] \quad \min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - \hat{y}_i\|_2^2 = \min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - w_{clus}^c f_{i,clus}^c(X_i^c) - w_{all}^c f_{i,all}^c(X_i^c)\|_2^2, s.t. 0 \leq w_{clus}^c, w_{all}^c \leq 1 \quad (17)$$

[0069] 其中: $n_c$ 为该天气类型的历史日个数。

[0070] 优选地,所述步骤(9)具体包括:

[0071] 获取待预测日天气类型的待预测日气象特征记为 $X_0^c$ ,利用步骤(8)的方式分别类

内相似日集和全数据相似日集建立模型  $f_{0,clus}^c$  和  $f_{0,all}^c$ ，应用天气类型c的加权模型系数，确定功率预测值  $\hat{y}_0$ ：

$$[0072] \quad \hat{y}_0 = w_{clus}^c f_{0,clus}^c(X_0^c) + w_{all}^c f_{0,all}^c(X_0^c) \quad (18)。$$

[0073] 本发明的有益效果在于：

[0074] 本发明通过将历史数据中的气象特征和功率特征转换为视图形式，挖掘各个视图中一致的局部结构，并利用重构的功率序列来辅助聚类任务完成，通过超参数的设定调整聚类效果，获得接近于真实天气类型的聚类结果，最后再针对各个聚类后的结果建立组合相似日模型，完成功率预测；与传统的聚类方法相比，本发明充分应用了真实的功率特征信息，其聚类后的结果相较于应用NWP数据聚类的方法更加准确，同时组合相似日的策略可以显著提高波动程度大的天气类型的功率预测精度。

### 附图说明

[0075] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0076] 图1为本发明的流程示意图；

[0077] 图2为本发明应用MIV技术计算各气象视图权重的流程图；

[0078] 图3为本发明应用功率重构的多视图聚类算法的功率曲线图；

[0079] 图4为本发明应用相似日策略后历史日的类内相似日和全数据相似日的功率指标对比图；

[0080] 图5为本发明的一个实施例的天气数据实际值与对应方法预测值的对比示意图。

### 具体实施方式

[0081] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0082] 如图1所示：本实施例提供一种基于天气类型的相似日组合策略的日前光伏功率预测方法，具体实施步骤如下：

[0083] (1) 已有的光伏功率历史日数据集包括NWP气象数据  $\{X_i^{log}\}_{i=1}^n$  和对应时刻的功率数据  $\{y_i^l\}_{i=1}^n$ ，对所有历史日数据进行最小最大归一化方法处理，其中， $X_i = [X_{i,1} \cdots X_{i,q}]$  为历史日i的气象数据，n为历史日个数，l为日序列长度，q为气象特征个数，；

[0084] (2) 应用本领域技术人员所熟知的MIV方法计算各气象特征对功率的MIV值，并取其绝对值作为各气象特征对功率的影响大小，作为多视图技术中各气象视图的权重，MIV技术流程图如图2所示，特别地，新增两噪声特征作比较以选择对功率影响大的气象特征；

[0085] (3) 针对天气类型一致的历史日，其气象特征曲线和功率波动曲线具有相似性，运用本领域技术人员所熟知的平滑表示多视图聚类方法处理，其过程为：将历史日NWP数据和

功率数据改写为视图形式  $\{V_k^{n \times l}\}_{k=1}^{q+1}$ ,  $k=1, \dots, q$  时,  $V_k = [X_{1,k} \ \dots \ X_{n,k}]^T$  表征气象视图, 特别地,  $V_{q+1} = [y_1 \ \dots \ y_n]^T$  表征功率视图;

[0086] 气象特征曲线和功率特征曲线具有相似性, 即特定历史日各特征曲线可以由同类型历史日曲线重构, 可以描述为:

$$[0087] \quad \sum_{i=1}^{q+1} \sum_{j=1}^n r_i \left\| v_{i,j} - \sum_{k=1}^n v_{i,j} z_{jk} \right\|_2^2 = \sum_{i=1}^{q+1} r_i \|V_i - ZV_i\|_F^2 \quad (1)$$

[0088] 其中,  $v_{i,j} \in \mathbb{R}^{1 \times l}$  表示  $i$  视图的日序列  $j$ ,  $z_{jk}$  表示历史日  $j$  和历史日  $k$  的相似性大小为自表示系数;  $r_i$  表示不同视图的权重系数; 同时显示地对自表示系数阵进行应用分组效应约束:

$$[0089] \quad \frac{1}{2} \sum_{i=1}^{q+1} \sum_{j=1}^n \sum_{k=1}^n r_i w_{i,jk} \|z_j - z_k\|_2^2 = \sum_{i=1}^{q+1} \text{tr}(ZL_i Z^T),$$

$$\text{s.t. } w_{i,jk} = \begin{cases} 1, & \text{if } v_{i,k} \in \text{Neighbor}(v_{i,j}) \\ 0, & \text{if } v_{i,k} \notin \text{Neighbor}(v_{i,j}) \end{cases}, d_{i,jk} = \begin{cases} \sum_{k=1}^n w_{i,jk}, & i = k \\ 0, & i \neq k \end{cases}, L_i = D_i - W_i \quad (2)$$

[0090] 分组效应的数学定义有:

[0091] 给定一组数据点  $[x_1 \ x_2 \ \dots \ x_n] \in \mathbb{R}^{d \times n}$  和自表示系数阵  $[z_1 \ z_2 \ \dots \ z_n] \in \mathbb{R}^{n \times n}$ ,  $\forall \|x_i - x_j\| \rightarrow 0$ , 对应的自表示系数阵  $\|z_i - z_j\| \rightarrow 0$ 。拉普拉斯矩阵  $L_i$  形式的约束, 不仅可以满足分组效应的应用条件, 而且共用同一个自表示系数阵  $Z$  挖掘各视图局部流形结构的一致性;

[0092] 取相似阵  $S = (Z + Z^T) / 2$ , 应用本领域技术人员所熟知的谱聚类方法得到天气类型标签阵  $F \in \mathbb{R}^{n \times C}$ ,  $c$  为聚类得到的天气类型个数,  $F_{ij} = 1$  表示历史日  $i$  属于天气类型  $j$ ,  $F_{ij} = 0$  表示历史日  $i$  不属于天气类型  $j$ , 且  $\sum_{j=1}^C F_{ij} = 1$ ;

[0093] (4) 通过对功率视图进行重构, 应用本领域技术人员所熟知的平滑多视图聚类方法得到的聚类标签功率数据对其进行约束, 其过程为:

[0094] 功率视图是真实采集得到的信息是反应天气类型的重要方式, 保持重构后功率数据信息量不变的前提下, 用尽量少的维度提炼其主要信息:

$$[0095] \quad \sum_{i=1}^n \|v_{q+1,i} - v_{q+1,i} P P^T\|_2^2 = \|V_{q+1} - V_{q+1} P P^T\|_F^2, \text{s.t. } P^T P = I \quad (3)$$

[0096] 其中,  $P \in \mathbb{R}^{1 \times m}$  为投影阵; 同时对各天气类型的功率数据的类内距离进行约束, 从而对平滑多视图聚类结果进行改进, 可以描述为:

$$[0097] \quad \sum_{i=1}^C \sum_{j=1}^{n_i} \|P^T v_{q+1,j}^i - o_i\|_2^2 = \|V_{q+1} P\|_F^2 - \text{tr}(\hat{F}^T V_{q+1} P P^T V_{q+1} \hat{F}) \text{s.t. } o_i = \frac{1}{n_i} \sum_{j=1}^{n_i} P^T v_{q+1,j}^i, \hat{F} = F \text{diag}\left(\frac{1}{\sqrt{n_1}} \ \dots \ \frac{1}{\sqrt{n_C}}\right) \quad (4)$$

[0098] 其中,  $o_i$  是属于天气类型  $i$  的功率序列中心,  $n_i$  是属于天气类型  $i$  的历史日天数,  $\hat{F}$  是经过系数调整之后的标签阵;

[0099] (5) 结合功率视图重构和平滑表示多视图聚类, 通过交替迭代求解聚类标签, 对历史日聚类为晴天、少云、阴天和多云天气类型, :

[0100] 有总目标函数：

$$\begin{aligned}
 \min_{Z,P} Loss = & \min_{Z,P} \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \\
 & + \|V_{q+1} - V_{q+1} P P^T\| + \beta (\|V_{q+1} P\|_2^2 - tr(\hat{F}^T V_{q+1} P P^T V_{q+1} \hat{F})) \\
 \Rightarrow \min_{Z,P} & \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \\
 & + P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P
 \end{aligned} \tag{5}$$

[0102] 其中，功率视图权重系数 $r_{q+1}$ 默认为1，其余视图特征权重采用本领域技术人员所熟知的MIV方法计算得到的MIV绝对值； $\alpha$ 是平滑表示多视图技术的超参；

[0103]  $\beta$ 是功率重构的超参；

[0104] 采用交替求解的方式：

[0105] 固定投影阵 $P$ ，求解自表示阵 $Z$ ，即对下式进行求解：

$$\min_Z \sum_{i=1}^q r_i (\alpha \|V_i - ZV_i\|_F^2 + tr(ZL_i Z^T)) + r_{q+1} (\alpha \|V_{q+1} P - ZV_{q+1} P\|_F^2 + tr(ZL_{q+1} Z^T)) \tag{6}$$

[0107] 对其求导，并设置导数为0，有：

$$\alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} P P^T V_{q+1}^T) Z + Z \sum_{i=1}^q r_i L_i + r_{q+1} L_{q+1} = \alpha (\sum_{i=1}^q r_i V_i V_i^T + r_{q+1} V_{q+1} P P^T V_{q+1}^T) \tag{7}$$

[0109] 该形式为Sylvester方程，具有唯一解；求解 $Z$ 后，可以求取亲和矩阵 $S$ ，从而对 $S$ 应用谱聚类算法获取天气类型标签 $F$ ，和系数调整后的天气类型标签 $\hat{F}$ ；

[0110] 固定自表示系数阵 $Z$ ，求解投影阵 $P$ ，即对下式进行求解：

$$\min_P P^T V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1} P \tag{8}$$

[0112] 显然解为 $V_{q+1}^T [(\beta - 1)I - \beta \hat{F} \hat{F}^T] V_{q+1}$ 对应的前 $m$ 个最小特征值对应的特征向量组成的矩阵；

[0113] (6)对气象特征进行新特征构建，并依据聚类结果将数据集划分为晴天数据集、少云数据集、阴天数据集、多云数据集和全历史数据集，对各个数据集气象特征基于互信息进行特征选择：

[0114] 新特征构建：计算各气象特征的一阶后向差分和二阶后向差分，挖掘气象特征动态信息，扩充特征集；

[0115] 基于互信息的特征选择：互信息是表明是否具有依赖关系的一种度量；在给定离散随机变量 $X$ 和 $Y$ 后， $X$ 的边缘分布为 $p(x)$ ， $Y$ 的边缘分布为 $p(y)$ ，两者的联合概率分布为 $p(x, y)$ 。此时 $X$ 和 $Y$ 之间的互信息为

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} ;$$

[0116] 计算各数据集中的特征对功率的互信息大小，并对互信息由大到小排列，选取前 $u$ 个特征作为该数据集相似日选择和功率预测模型建立的特征；

[0117] (7)按照加权灰色关联度的相似度准则选择相似日，对各天气类型下的历史日依次构建类内相似日集和全数据相似日集：

[0118] 加权灰色关联度：判定两历史日是否具有相似性的相似度准则，利用几何形状衡量序列之间关联程度的分析比较方法，灰色关联度越高，待预测日与历史日的相似性越大，

其计算过程为:

[0119] 定义经过特征构建和特征选择后的历史日数据为  $X_i = [X_{i,1} \cdots X_{i,u}] \in \mathbb{R}^{1 \times u}$ , 计算历史日  $X_0$  和历史日  $X_i$  的关于特征  $k$  的灰色关联系数  $\varepsilon_{ik}$ :

$$[0120] \quad \varepsilon_{ik}(t) = \frac{\min_{ik} \min_t \Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}{\Delta_{ik}(t) + \mu \max_{ik} \max_t \Delta_{ik}(t)}, \quad s.t. \quad t = 1, \dots, l, \Delta_{ik}(t) = |X_{0,k}(t) - X_{i,k}(t)| \quad (9)$$

[0121] 其中,  $X_{i,k}(t)$  表征气象特征  $k$  第  $t$  时刻的值, 分辨系数  $\mu = 0.5$ 。那么对应的加权灰色关联度  $\zeta_i$  的计算式为:

$$[0122] \quad \zeta_i = \sum_{k=1}^u w_k \left( \frac{1}{l} \sum_{t=1}^l \varepsilon_{ik}(t) \right) \quad s.t. \quad \sum_{k=1}^u w_k = 1 \quad (10)$$

[0123] 其中,  $w_k$  为气象特征  $k$  对灰色系数的权重, 该权重应用 CRITIC 方法计算得出, 采用特征间的对比强度和冲突性来计算, 特征  $k$  的标准差  $\sigma_k$  表现对比强度, 相关系数  $\rho_{ki}$  衡量冲突性, 计算式为:

$$[0124] \quad \begin{cases} C_k = \sigma_k \sum_{i=1}^u (1 - |\rho_{ki}|), k = 1, \dots, u \\ w_k = C_k / \sum_{k=1}^u C_k \end{cases} \quad (11)$$

[0125] 计算历史日  $d_0$  和数据集内的所有历史日计算灰色关联度, 对其由大到小排列, 选取前若干个相似日构成历史日  $d_0$  在该数据集下的相似日集;

[0126] 类内相似日集和全数据相似日集: 定义待预测日  $X_0$  在同天气类型历史日中选择的相似日数据为类内相似日集  $D_{0,clus}$ , 在全部历史日中选择的相似日数据为全数据相似日集  $D_{0,all}$ ;

[0127] (8) 对历史日的两组相似日集分别建立 GBDT 模型获取预测值, 特定天气类型下的所有历史日组成类内 GBDT 模型预测和 GBDT 全数据模型预测值, 应用遗传算法求取两模型预测值的加权系数:

[0128] GBDT 模型: 梯度提升决策树 (Gradient boosting decision tree, GBDT) 是一种加法模型, 表示为:

$$[0129] \quad f(X) = \sum_{t=1}^T \alpha_t h_t(X) \quad (12)$$

[0130] 其中,  $X$  为输入样本,  $h_t(X)$  为分类回归树 (Classification and regression trees, CART),  $T$  是 GBDT 中需要构建的树的数量,  $\alpha_t$  是第  $t$  棵树的权重。GBDT 采用前向分布算法, 首先确定  $F_0(x)$  为模型初始值, 通常为常数, 第  $m$  步的模型是:

$$[0131] \quad f_m(X) = f_{m-1}(X) + \alpha_m h_m(X) \quad (13)$$

[0132] 其中,  $F_{m-1}(x)$  为当前模型。新添加的分类回归树  $h_m(x)$  通过最小化损失函数求得:

$$[0133] \quad h_m = \arg \min_h \sum_{i=1}^N L(y_i, f_{m-1}(X_i) + h(X_i)) \quad (14)$$

[0134] 其中,  $N$  是样本个数, GBDT 采用梯度下降法来求解最优模型, 将损失函数在当前模型  $f_{m-1}(X)$  的负梯度值作为梯度下降的方向:

$$[0135] \quad F_m(X) = f_{m-1} - \alpha_m \sum_{i=1}^N \nabla_F L(y_i, f_{m-1}(X_i)) \quad (15)$$

[0136] 其中,  $\alpha_m$  通过线性搜索求得:

$$[0137] \quad \alpha_m = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, f_{m-1}(x_i)) - \alpha \frac{\partial L(y_i, f_{m-1}(X_i))}{\partial f_{m-1}(X_i)} \quad (16)$$

[0138] 遗传算法求取模型加权系数: 经过特征构建的天气类型  $c$  的气象特征数据记为  $X_i^c \in R^{I \times u}$ , 功率数据记为  $y_i^c \in R^l$ , 由对应的类内相似日集  $D_{i,clus}^c$  和全数据相似日集  $D_{i,all}^c$  分别建立GBDT模型  $f_{i,c}^c$  和  $f_{i,all}^c$ , 最终的功率预测值为  $\hat{y}_i = w_{clus}^c f_{i,clus}^c(X_i^c) + w_{all}^c f_{i,all}^c(X_i^c)$ , 其中  $w_{clus}^c$  和  $w_{all}^c$  为该天气类型下两类模型的加权系数; 为保证预测效果, 有目标函数为:

$$[0139] \quad \min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - \hat{y}_i\|_2^2 = \min_{w_1^c, w_2^c} \sum_{i=1}^{n_c} \|y_i^c - w_{clus}^c f_{i,clus}^c(X_i^c) - w_{all}^c f_{i,all}^c(X_i^c)\|_2^2, s.t. 0 \leq w_{clus}^c, w_{all}^c \leq 1 \quad (17)$$

[0140]  $n_c$  为该天气类型的历史日个数; 应用上式作为遗传算法的适应度函数, 设定种群大小和迭代轮次, 求取对应天气类型下待预测日建立的两模型的加权系数。(9) 由天气预报获取待预测日天气类型, 应用该日的NWP数据获取其类内相似日集和全数据相似日集, 分别建立模型, 通过该天气类型下的模型加权系数获取最终的功率预测值:

[0141] 图4为相似日组合策略的流程图, 由天气预报等手段获取待预测日天气类型, 待预测日气象特征记为  $X_0^c$ , 分别建立模型  $f_{0,clus}^c$  和  $f_{0,all}^c$ , 应用天气类型  $c$  的加权模型系数, 有功率预测值  $\hat{y}_0$ :

$$[0142] \quad \hat{y}_0 = w_{clus}^c f_{0,clus}^c(X_0^c) + w_{all}^c f_{0,all}^c(X_0^c) \quad (18)$$

[0143] 本实施例结合国能日新第二届光伏功率预测大赛数据来说明基于天气分型的组合相似日策略的光伏功率预测模型应用NWP数据预测功率的性能。

[0144] 光伏发电具有波动性和间歇性, 大规模光伏电站并网运行可能对电力系统的安全稳定经济运行造成影响。为了帮助调度部门统筹安排常规能源和光伏发电的协调配合, 及时调整调度计划, 合理安排电网运行方式, 对光伏电站的输出功率建立合适模型并准确预测。

[0145] 光伏功率历史日数据包含时间、NWP数据和功率; 其中, 时间形式为“年/月/日时分”, 且每15分钟采集一个数据, 一天共96个数据; NWP数据特征包括辐照度、风速、风向、温度、压强和湿度, 均为对应时刻气象特征的预测值, 并非实测值; 功率数据为对应时刻光伏电站功率的实测值; 采用春季数据2016年4月和5月、2017年3月、4月和5月, 2018年3月作为进行聚类 and 功率预测的历史日集, 避免由于不同季节的出力水平不同, 而致使天气类型划分错误。原始特征为NWP数值天气预测数据, 标签为对应时刻的光伏发电功率。通过历史日太阳辐照强度和Hottel模型数据进行比较, 对2018年4月数据的天气类型进行确定, 并选取其中典型天气类型的历史日进行效果对比。对各特征和新建白噪声特征应用本领域技术人员所熟知的MIV方法得到的MIV绝对值如表1所示, 白噪声MIV值符合其与功率无关的特性, 以此为基准, 风速、风向和温度的MIV值虽然相较白噪声大一个量级却比辐照度小一个量级, 故舍去以上三种气象特征, 选取辐照度、压强和湿度作为气象视图; 对应各视图权重设置为  $[r_{辐照度} \ r_{压强} \ r_{湿度} \ r_{功率}] = [0.4096 \ 0.12 \ 0.1104 \ 1]$ 。

[0146] 表1

[0147]	属性	MIV	MIV
	辐照度	+0.4096	0.4096
	风速	-0.0337	0.0337
	风向	+0.0345	0.0345
	温度	+0.0176	0.0176
	压强	+0.1200	0.1200
	湿度	-0.1104	0.1104
	白噪声1	-0.0025	0.0025
	白噪声2	-0.0015	0.0015

[0148] 根据短期预报标准即国标GB/T21984-2017,天空状况被分为晴、少云、多云和阴天,结合功率波动主要由天空中云量变化引起,故天气类型划分为晴、少云、多云和阴天4种;功率重构的平滑多视图聚类法的参数设置为 $\alpha=50, \beta=0.9$ ,功率重构维数 $m=30$ 。

[0149] 表2为使用方法1、方法2和方法3聚类并判定天气类型后的聚类结果性能指标比较,表2的各方法为:Kmeans方法即方法1、只应用了功率特征的平滑表示聚类法即方法2、本实施例提出的基于天气类型的相似日组合策略的日前光伏功率预测方法即方法3,本实施例采用欧式距离 $d$ 、皮尔逊相关系数 $Cor$ 和时间相关系数 $Cort$ 作为性能指标:

[0150] 
$$d = \frac{1}{n_c^2} \sum_{i,j=1}^{n_c} \|y_i^c - y_j^c\|$$

[0151] 
$$Cor = \frac{1}{n_c^2} \sum_{i,j=1}^{n_c} \frac{\sum_{t=1}^l (y_i^c(t) - \bar{y}_i^c) \cdot (y_j^c(t) - \bar{y}_j^c)}{\sqrt{\sum_{t=1}^l (y_i^c(t) - \bar{y}_i^c)^2} \sqrt{\sum_{t=1}^l (y_j^c(t) - \bar{y}_j^c)^2}}$$

[0152] 
$$Cort = \frac{1}{n_c^2} \sum_{i,j=1}^{n_c} \frac{\sum_{t=1}^{l-1} (y_i^c(t+1) - y_i^c(t)) \cdot (y_j^c(t+1) - y_j^c(t))}{\sqrt{\sum_{t=1}^{l-1} (y_i^c(t+1) - y_i^c(t))^2} \sqrt{\sum_{t=1}^{l-1} (y_j^c(t+1) - y_j^c(t))^2}}$$

[0153] 表2

[0154]

天气类型	方法	$d$	$Cor$	$Cort$
晴天	1	0.858	0.971	0.502
	2	0.5	0.989	0.678
	3	<b>0.459</b>	<b>0.991</b>	<b>0.718</b>
少云	1	1.451	0.897	0.177
	2	1.039	0.955	0.288
	3	<b>0.917</b>	<b>0.965</b>	<b>0.342</b>
阴天	1	1.617	0.811	0.099
	2	1.557	0.806	0.095
	3	<b>1.496</b>	<b>0.812</b>	<b>0.104</b>
多云	1	<b>1.408</b>	<b>0.906</b>	<b>0.201</b>
	2	1.422	0.896	0.154
	3	1.483	0.877	0.129

[0155] 式中,  $y_i^c$ 表示天气类型c下历史日i的功率序列,  $\bar{y}_i^c$ 表示表示天气类型c下历史日i的功率序列均值,  $n_c$ 表示天气类型c的历史日个数,  $d$ 越小越好,  $Cor$ 和 $Cort$ 越大越好。

[0156] 表2可以看出,本实施例的方法在晴天、少云、阴天的聚类上具有明显优势,可以清晰地将以上三种天气类型区分开来。多云类型下,复杂的天气情况导致功率序列的皮尔逊系数和时间相关系数在三种方法上的表现都较小,在该天气类型下比较欧氏距离,虽然本实施例的方法欧式距离最大,但相较于其余方法相差不多。综上所述,本实施例的方法在三种方法中具有明显优势,对历史日集标注本发明聚类得到的天气类型,划分为天气类型历史日集  $\{X_i^c, y_i^c\}_{i=1}^{n_c}$ 。应用历史日集验证组合相似日策略的效果,对  $\{X_i^c, y_i^c\}_{i=1}^{n_c}$ 按照7:3的比例划分训练集和测试集。应用训练集通过GA算法寻优求解  $\omega_{clus}$ 和  $\omega_{all}$ 的加权系数,并作为测试集中待预测日  $(X_i^c, y_i^c)$ 寻找相似日的历史日集,构建对应  $D_{i,clus}^c$ 和  $D_{i,all}^c$ ,相似日天数选择为6。

为了比较  $D_{i,clus}^c$ 和  $D_{i,all}^c$ 的差异,对天气类型下  $(X_i^c, y_i^c)$ 的  $D_{i,clus}^c$ 和  $D_{i,all}^c$ 的功率指标进行统计。本实施例定义的功率指标如下,其中  $y'_i$ 为功率序列  $y_i$ 的极值点序列,  $l_e$ 为极值点序列  $y'_i$ 的极值点个数:

[0157] 波动峰值 =  $\max_{1 \leq t \leq l_e - 1} (|y'_i(t+1) - y'_i(t)|)$ ;

[0158] 波动频率 =  $l_e / l$ ;

[0159] 功率和 =  $\sum_{t=1}^l y_i(t)$  ;

[0160] 如图3所示,由于在相似日选择前会进行特征选择,故选择相似日的特征和加权灰色关联度的系数不一定相同,所以  $(X_i^c, y_i^c)$ 的  $D_{i,clus}^c$ 和  $D_{i,all}^c$ 会有较大差异。明显看出各天气类



型下  $D_{i,clus}^c$  和  $D_{i,all}^c$  的功率指标在不同区间集中,晴天和少云类型下尤其明显,阴天和多云类型下集中的区间一致但是分布却不一致,故训练出的模型  $f_{i,clus}^c$  和  $f_{i,all}^c$  具有不同的特点。通过GA求解特定天气类型下  $f_{i,clus}^c$  和  $f_{i,all}^c$  的加权系数  $\omega_{clus}$  和  $\omega_{all}$ ,结合两模型的优点改善预测效果。

[0161] 分析采取的相似日选择方式和  $f_{i,clus}^c$ 、 $f_{i,all}^c$  的加权策略是否有效,对相似日组合策略的各个部分进行移除或添加后的预测效果进行预测效果对比。同时充分考虑策略的稳定性,对历史日集随机划分10次训练集和测试集,取10次评价结果的平均值为最终预测效果。表3为各方法与相似日组合预测模型各部分的关系。本实施例应用均方根误差RMSE、平均绝对误差MAE和决定系数 $R^2$ 作为模型的评价指标,其中,N为功率点个数, $C_{ap}$ 为光伏电场的装机容量:

$$[0162] \quad e_{RMSE} = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2} / C_{ap} \sqrt{N}$$

$$[0163] \quad e_{MAE} = \sum_{i=1}^N |y_i - \hat{y}_i| / C_{ap} N$$

$$[0164] \quad R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}$$

[0165] 表3

[0166]

序号	模型组成
方法a	GBDT
方法b	全数据相似日建模+GBDT
方法c	类内相似日建模+GBDT
方法d	相似日组合建模+GBDT
方法e	特征选取+GBDT
方法f	特征选取+全数据相似日建模+GBDT
方法g	特征选取+类内相似日建模+GBDT
方法h	特征选取+相似日组合建模+GBDT

[0167] 预测效果如表4所示,表3中的方法h为本实施例的方法。通过方法a~方法d与方法e~方法h的对比,基于互信息的特征提取对模型预测的准确性有较大提升。对比方法e和方法h,方法e是对所有历史日建模,包含与待预测日灰色关联度差别较大的日功率数据,导致GBDT建模的预测准确性较方法h差;对比方法f、方法g和方法h,在晴天和少云天气类型下,方法g建立的模型预测效果都接近于方法h的效果,但在功率波动峰值小,波动频率大的阴天天气类型下,方法h相较于方法f和方法g有较大提升,误差评价RMSE分别提升11.0%和8.9%,MAE分别提升10.0%和10.2%;在波动峰值大,波动频率小的多云天气类型下同样有所改善,RMSE上分别提升1.3%和6.1%,MAE上分别提升0.8%和5.8%。相似日组合预测模型在波动程度较小的天气类型中,提升或维持同类型相似日选取策略的性能,对于波动程度较大的天气类型,会改善其预测性能。

[0168] 表4

[0169]

类型	晴天			少云			阴天			多云		
	方法	$R^2$	$e_{RMSE} / \%$	$e_{MAE} / \%$	$R^2$	$e_{RMSE} / \%$	$e_{MAE} / \%$	$R^2$	$e_{RMSE} / \%$	$e_{MAE} / \%$	$R^2$	$e_{RMSE} / \%$
a	0.9466	7.4891	4.3575	0.8320	12.3887	7.9230	-0.928	16.2105	10.0125	0.4169	17.6315	11.7882
b	0.9012	10.0723	6.3230	0.8061	13.0764	7.9027	0.2246	14.7256	8.9489	0.6140	15.7566	9.8732
c	0.9478	7.8904	4.6047	0.8626	11.5583	7.2050	-0.1771	16.2866	10.1863	0.5484	17.1207	10.8170
d	0.9478	7.8904	4.6047	0.8735	11.1227	6.9856	0.4139	16.3860	9.8915	0.6265	16.3671	10.0743
e	0.9848	4.5086	2.5501	0.9357	<b>8.1954</b>	4.6480	0.4558	14.7328	8.3254	0.8155	11.8964	6.9949
f	0.9324	8.5416	4.7589	0.9229	8.7223	5.1283	<b>0.5335</b>	15.2531	8.2863	0.8110	12.3866	7.0562
g	0.9851	4.4515	2.5264	0.9347	8.2314	4.6708	0.4323	14.9136	8.3106	0.8239	11.7796	6.6984
h	<b>0.9863</b>	<b>4.2794</b>	<b>2.451</b>	<b>0.9356</b>	8.203	<b>4.6194</b>	0.4558	<b>13.5764</b>	<b>7.4583</b>	<b>0.8245</b>	<b>11.6306</b>	<b>6.6466</b>

[0170] 如图5所示,在验证相似日组合策略对于聚类后的数据有效后,再验证本实施例的方法是否对特定天气类型的历史日有效。取特定天气历史日,按照相似日组合策略流程对其功率进行预测。为验证效果和其他方法进行对比,包括应用其他回归器在只进行基于互信息的特征选择情况下进行回归器和组合相似日策略的结合的情况下进行回归。为保证效果稳定有效,每种天气类型在2018年4月份选取典型的2天进行回归,评价效果包括 $e_{RMSE}$ 和 $e_{MAE}$ 。如表5,方法A为GBDT回归,方法B为应用组合相似日策略和GBDT回归;方法C为RF(Random Forest)随机森林回归,方法D为组合相似日策略和RF回归。

[0171] 表5

[0172]

天气类型	日期	方法 1		方法 2		方法 3		方法 4	
		$e_{RMSE}$	$e_{MAE}$	$e_{RMSE}$	$e_{MAE}$	$e_{RMSE}$	$e_{MAE}$	$e_{RMSE}$	$e_{MAE}$
晴天	04-02	5.313	3.185	<b>4.156</b>	<b>2.486</b>	5.600	3.513	4.955	3.140
	04-16	6.861	4.309	<b>4.922</b>	<b>3.296</b>	6.283	3.938	4.966	3.020
少云	04-05	8.036	5.062	<b>7.787</b>	4.824	8.362	5.091	7.888	<b>4.762</b>
	04-20	8.048	4.588	<b>7.675</b>	<b>4.352</b>	8.565	4.617	8.574	4.526
阴天	04-12	12.599	6.909	<b>12.364</b>	<b>6.488</b>	16.503	8.734	13.860	7.735
	04-23	14.240	7.662	<b>10.567</b>	<b>5.798</b>	12.701	6.448	10.984	6.057

[0173]

多云	04-11	13.124	7.574	<b>11.760</b>	<b>6.514</b>	12.523	7.305	12.042	6.906
	04-28	13.765	7.043	<b>12.406</b>	<b>6.646</b>	14.204	8.496	13.798	8.088

[0174] 对2018年4月选取特定天气类型历史日的回归效果如表5所示;对比采用了相似日组合策略的方法B、方法C和不应用相似日组合策略的方法A、方法C,采取组合相似日组合策略的模型比仅应用GBDT或RF的效果好;对比方法B和方法A、方法C、方法D,本实施例方法的

预测效果更好。对比表4和表5,各天气类型待预测日在效果上和表4效果一致,反映出本实施例所提功率重构的平滑表示多视图聚类算法的优势。本实施例所提的基于天气类型的相似日组合短期光伏功率预测方法相对于其他方法在阴天和多云情况下预测精度更高。

[0175] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

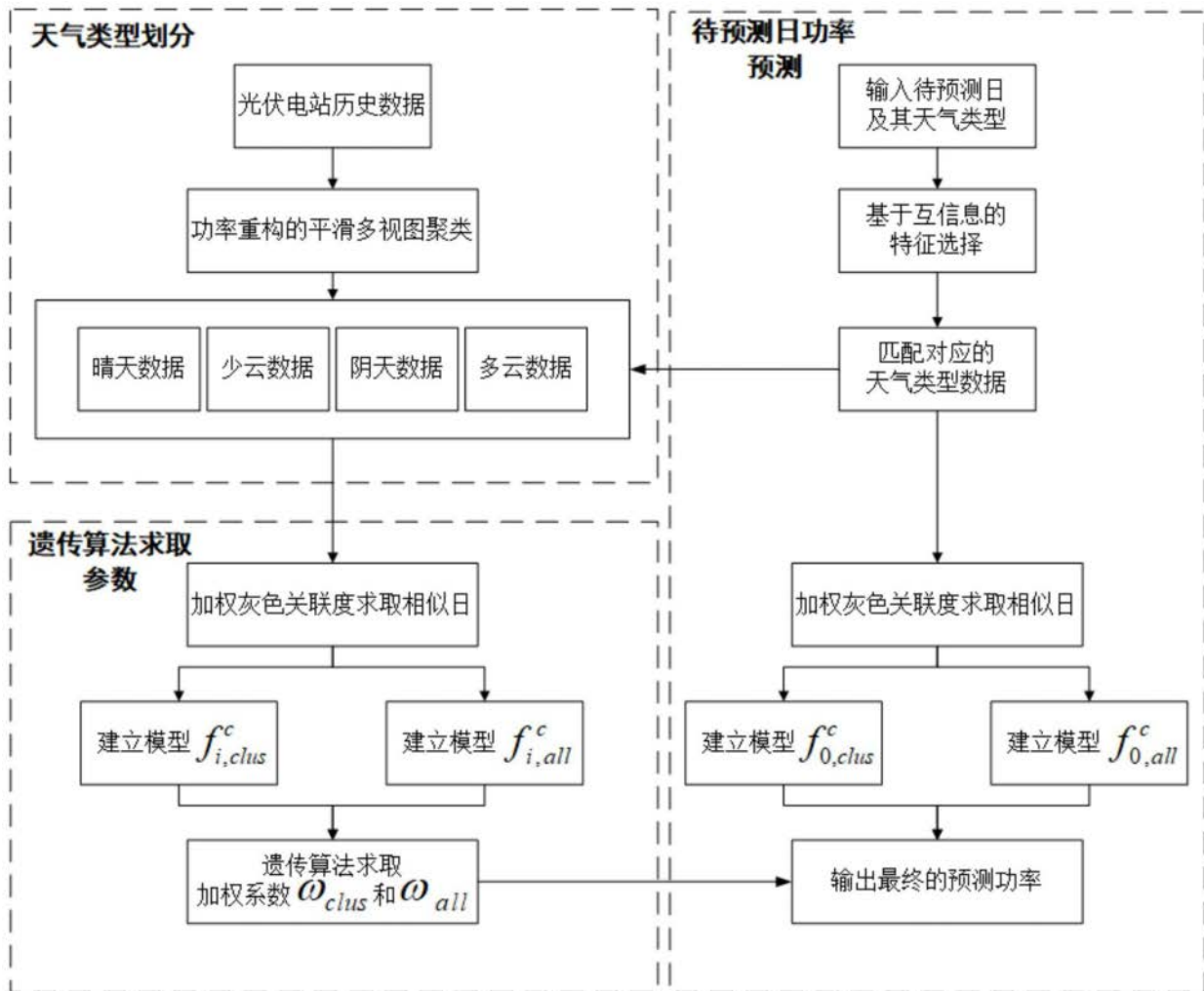


图1

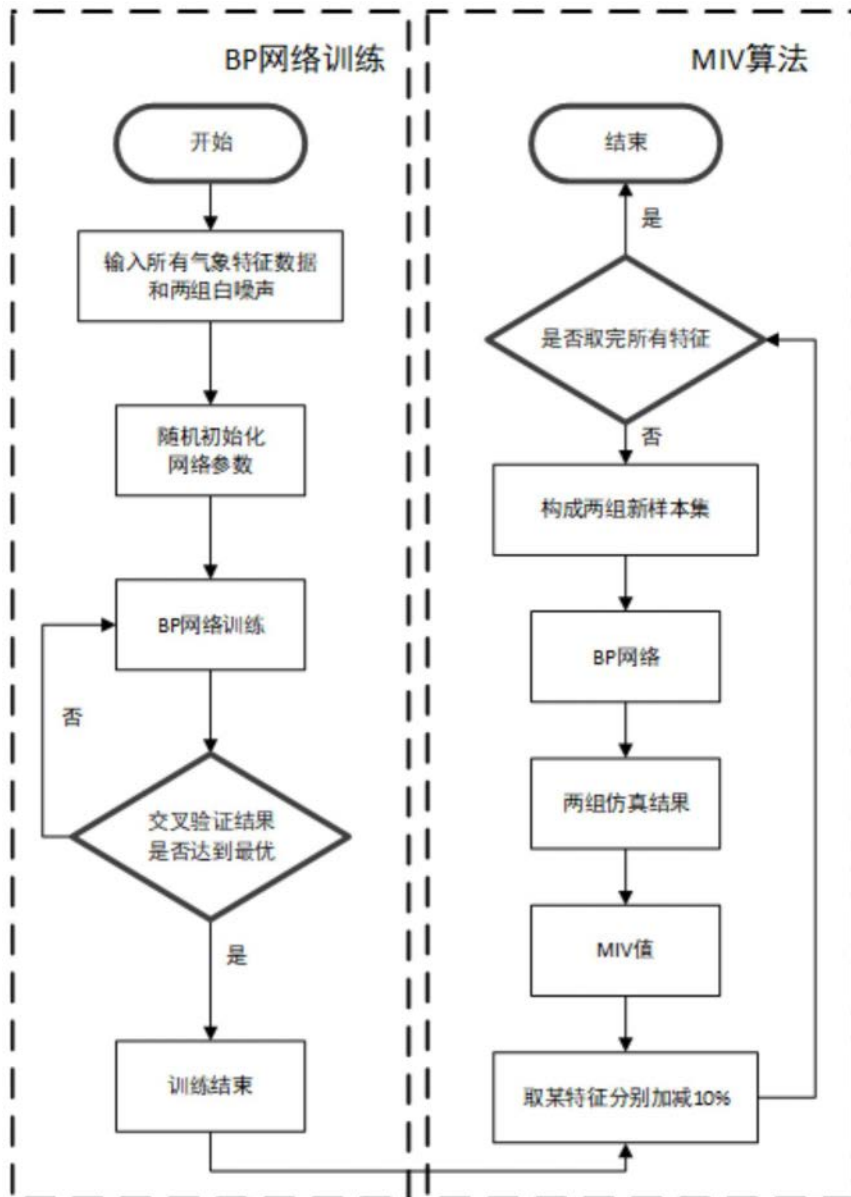


图2

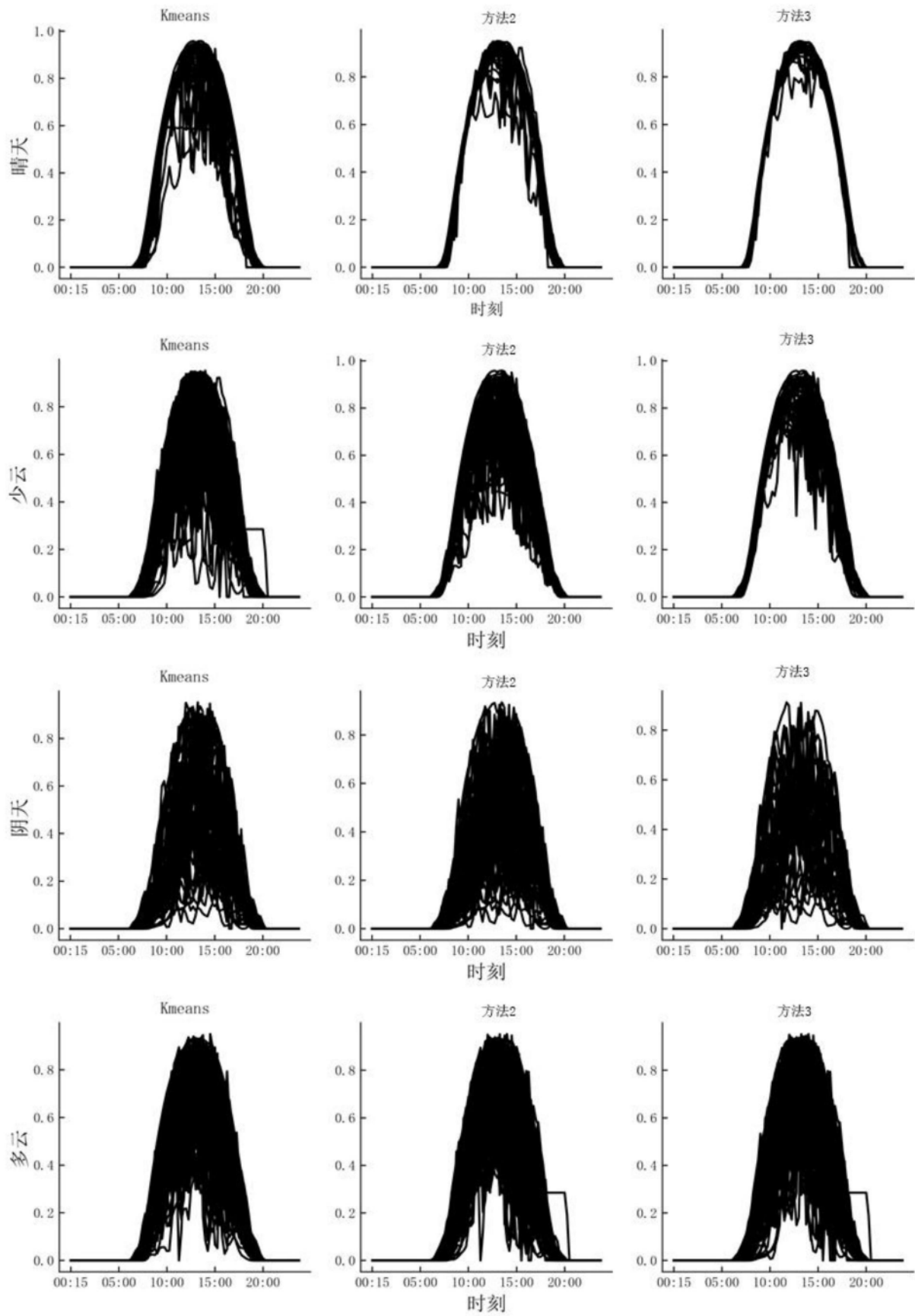


图3

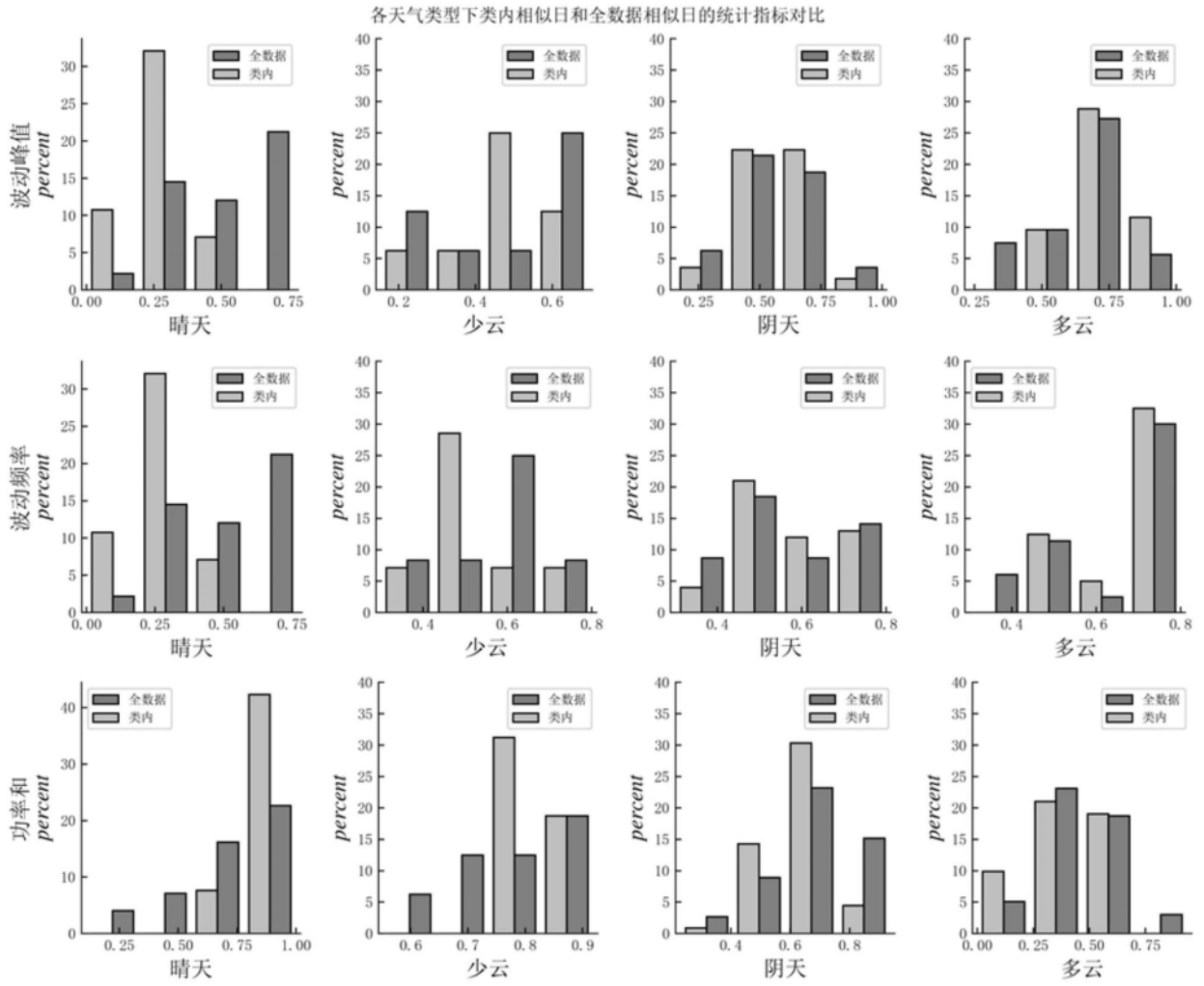


图4

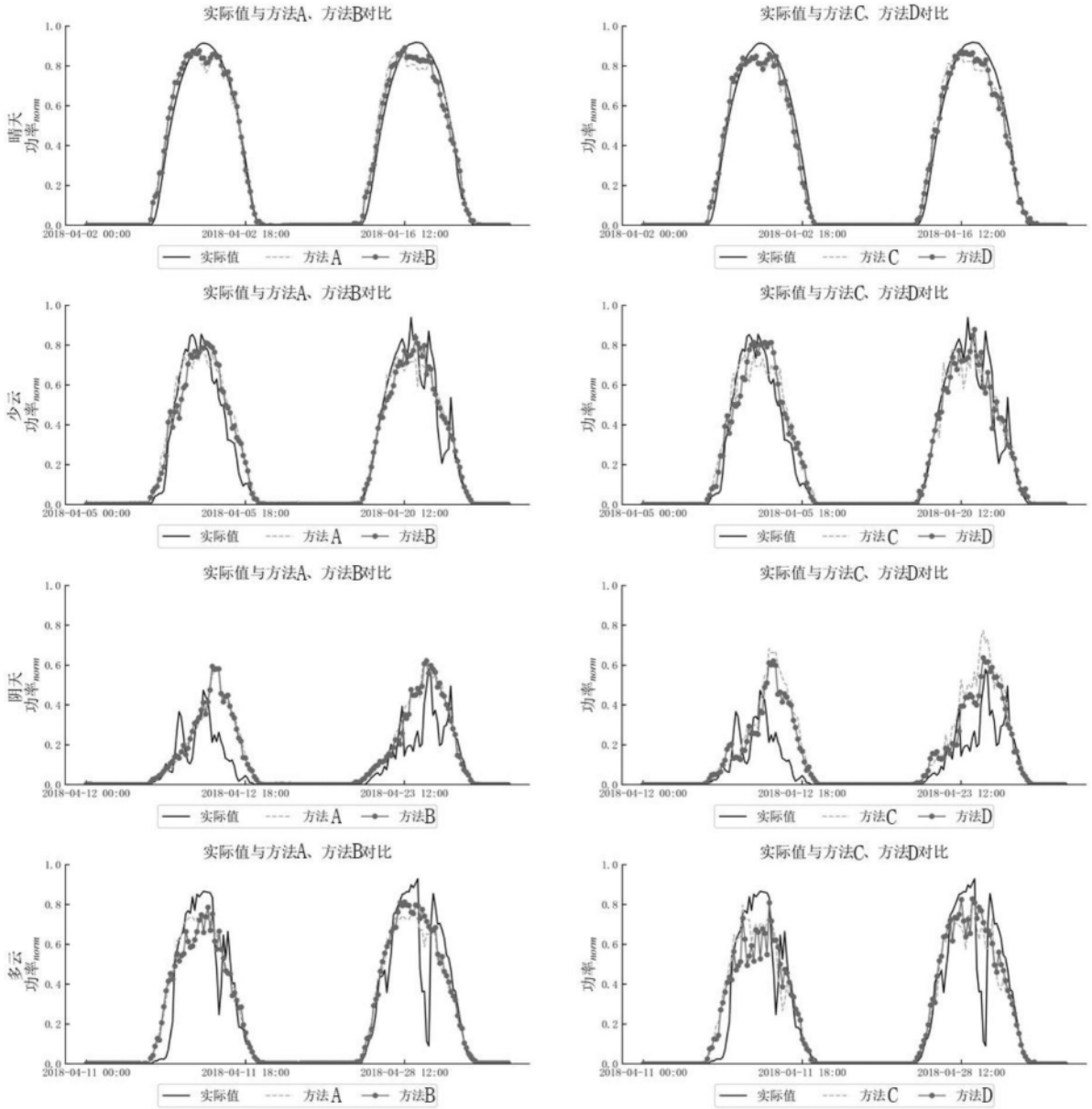


图5