



# (12)发明专利申请

(10)申请公布号 CN 106251001 A

(43)申请公布日 2016.12.21

(21)申请号 201610567777.5

(22)申请日 2016.07.18

(71)申请人 南京工程学院

地址 211167 江苏省南京市江宁科学园弘景大道1号

(72)发明人 杨志超 陆文伟 葛乐 马寿虎  
陆文涛 顾佳易 王蒙

(74)专利代理机构 南京纵横知识产权代理有限公司 32224

代理人 董建林

(51)Int.Cl.

G06Q 10/04(2012.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

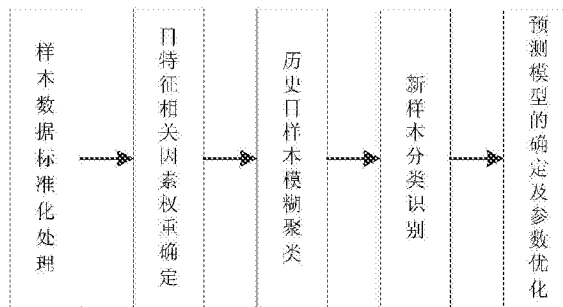
权利要求书2页 说明书8页 附图3页

## (54)发明名称

一种基于改进模糊聚类算法的光伏功率预测方法

## (57)摘要

本发明公开了一种基于改进模糊聚类算法的光伏功率预测方法,为了减小相似日选取误差和提高光伏发电功率预测精度,首先,利用改进模糊聚类算法对历史日样本进行了聚类分析,然后,利用预测日所属类别的历史数据和预测日的气象数据建立了BP神经网络发电预测模型。以实际数据对所提模型进行了验证,并与传统的基于相似日选取的光伏功率预测模型进行了对比,结果表明该模型具有更高的预测精度。



1. 一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,包括如下步骤:

- 1) 采集样本数据,并进行无量纲化处理以及标准化处理;
- 2) 根据步骤1)的样本数据,确定日特征相关因素的权重;

3) 历史日样本模糊聚类,计算衡量本分类对象 $x_i$ 与 $x_j$ 之间的相似系数 $r_{ij}$ ,确定论域 $U$ 上的模糊关系矩阵:

设 $S(X, Y)$ 表示样本 $X$ 和样本 $Y$ 之间的相似度,且 $S(X, Y) = \alpha D(X, Y) + \beta R(X, Y)$ ,

$$D(X, Y) = 1 - \sqrt{\frac{1}{m} \sum_{i=1}^m (\sigma_i x_i - \sigma_i y_i)^2}, \text{模糊矩阵 } R(X, Y) = \frac{\sum_{i=1}^m \sigma_i^2 x_i y_i}{\sqrt{\sum_{i=1}^m \sigma_i^2 x_i^2 \sum_{i=1}^m \sigma_i^2 y_i^2}}, \text{式中, } x_i \text{ 和 } y_i \text{ 分别是}$$

样本 $X$ 和样本 $Y$ 属性因子的数值,且在 $[0, 1]$ 之间, $\sigma_i$ 为样本中第 $i$ 个属性因子的权值( $i=1, 2, \dots, m$ ), $\alpha$ 和 $\beta$ 为 $D(X, Y)$ 和 $R(X, Y)$ 权重系数, $\alpha + \beta = 1$ ;

将得到的模糊矩阵 $R$ 通过逐次平方法求出其传递闭包阵 $R^* = t(R)$ ,依据经验值取适当的阈值 $\lambda$ ,对 $R^*$ 进行截割;

$$\text{对 } \forall \lambda \in [0, 1], \text{记模糊矩阵 } R \text{ 的 } \lambda\text{-截距阵为 } t(R)_\lambda = (r_{ij}^*), \text{其中, } r_{ij}^* = \begin{cases} 1, & r_{ij} \geq \lambda \\ 0, & r_{ij} < \lambda \end{cases}; \text{按不同的}$$

$\lambda$ 值进行分类,得到不同的聚类结果;

$$4) \text{对新样本进行分类识别,首先根据聚类情况,按下式求出各类的聚类中心: } \bar{x}_{ik} = \frac{1}{p} \sum_{j=1}^p x_{jk},$$

$k=1, 2, \dots, m$ ,式中,第 $i$ 类包括 $p$ 个样本; $m$ 为样本特征指标的数目;然后按步骤3)的方法进行新样本的分类识别;

5) 确定预测模型并进行参数优化。

2. 根据权利要求1所述的一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,所述步骤2)具体内容如下:

设有自变量 $x_1, x_2, \dots, x_m$ 和因变量 $y$ ,因变量序列为 $Y = [y_1, y_2, \dots, y_n]$ ,自变量序列为

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}, \text{自变量 } x_k \text{ 对因变量 } y \text{ 的直接途径系数 } E_{x_k \rightarrow y} = b_k \sqrt{\sum_{j=1}^n (x_{jk} - \bar{x}) / (y_j - \bar{y})},$$

自变量 $x_k$ 通过自变量 $x_{k+1}$ ,对因变量 $y$ 的间接途径系数为 $E_{x_k \rightarrow x_{k+1} \rightarrow y} = r_{k(k+1)} E_{x_k \rightarrow y}$ ,自变量 $x_{k+1}$ 通过自变量 $x_k$ 对因变量 $y$ 的间接途径系数为 $E_{x_{k+1} \rightarrow x_k \rightarrow y} = r_{k(k+1)} E_{x_k \rightarrow y}$ ,式中, $b_k$ 为偏回归系数, $r_{k(k+1)}$ 为自变量 $x_k$ 和自变量 $x_{k+1}$ 的单相关系数;

根据上述得到的各自变量的直接途径系数,求得各自变量的对因变量的影响权重 $\sigma_i$  ( $i=1, 2, \dots, k$ )为

$$\sigma_i = \frac{|E_{x_i \rightarrow y}|}{\sum_{i=1}^k |E_{x_i \rightarrow y}|}.$$

3. 根据权利要求1所述的一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,所述步骤3)中 $\alpha = 0.5, \beta = 0.5$ 。

4. 根据权利要求1所述的一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,所述步骤3)中的最佳分类阈值 $\lambda$ 利用F-统计量进行确定。

5. 根据权利要求1所述的一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,所述步骤5)中模型采用BP网络模型,包括输入层、隐含层和输出层;具体建模方式如下:

51) 输入量和输出量:

设 $t$ 为预测日,经过分类识别后 $t \in M_i, (i=1, 2, \dots, n)$ ,式中, $n$ 为历史日样本经过模糊聚类后的分类数;

将 $M_i$ 类中包含的日样本按与预测日 $t$ 的相似度由大到小排序结果为 $t_1, t_2, t_3, \dots, t_d$ ,式中, $d$ 为 $M_i$ 类中包含的日样本数;

选取与预测日 $t$ 的相似度最大的前六日的光伏阵列输出功率序列组成的学习样本作为BP网络的输入矩阵为

$$\begin{bmatrix} P_1(t-6) & P_1(t-5) & \cdots & P_1(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ P_2(t-6) & P_2(t-5) & \cdots & P_2(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ P_{24}(t-6) & P_{24}(t-5) & \cdots & P_{24}(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \end{bmatrix}, \text{式中, } t \text{ 为预测日;}$$

$t-x (1 \leq x \leq 6)$ 为与预测日相似度最大的前 $x$ 天; $P_y (1 \leq y \leq 24)$ 为相似日第 $y$  h光伏阵列的功率值; $R_1(t), R_2(t), \dots, R_6(t)$ 分别为取量化值第 $t$ 天的日最高温度、日最低温度、日平均温度、日相对湿度、日平均风速、天气类型;

BP网络的输出量为预测日的24h光伏阵列功率值,即 $[P_1(t) P_2(t) \cdots P_{24}(t)]$ ;

52) 隐含层:

隐含层节点数的确定采用经验公式 $H = \sqrt{I+O} + a$ ,式中, $H$ 为隐含层节点数; $I$ 为输入层节点数; $O$ 为输出层节点数; $a$ 为1~10之间的常数;

53) 转移函数:

该网络隐含层的转移函数采用双曲正切型S函数,即 $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ ,输出层的转移函数采用线性函数,即 $f(x) = x$ ;

54) BP神经网络的参数优化:

通过差分进化算法进行参数优化;获得最优个体后,将最优个体解码后初始化BP神经网络参数,然后利用样本训练BP神经网络,利用训练好的BP神经网络进行功率预测。

## 一种基于改进模糊聚类算法的光伏功率预测方法

### 技术领域

[0001] 本发明涉及一种基于改进模糊聚类算法的光伏功率预测方法,属于光伏发电技术领域。

### 背景技术

[0002] 光伏发电系统的输出功率受到太阳辐照度、温度、相对湿度、风速等众多气象因素的影响,具有很大的波动性和随机性。大规模光伏并网会对电力系统的安全运行和电网调度造成较大影响。光伏阵列发电功率预测是解决此问题的关键技术之一,有助于电网调度部门及时地调整调度计划,合理安排电网运行方式,对提高电力系统运行的安全性和稳定性具有重要意义。

[0003] 目前,光伏发电功率预测可以归纳为两类:间接预测和直接预测。间接预测法先利用历史太阳辐照强度数据和天气预报信息对地表的太阳辐照强度进行预测,然后根据光伏发电系统的出力模型得到光伏系统的输出功率。与间接预测相比,直接预测利用光伏发电系统历史输出功率数据和天气预报信息直接预测其输出功率,建模比较简单且不需要安装太阳辐射强度测量装置。

[0004] 由于在具有相似气象条件的情况下,光伏阵列的输出功率曲线具有一定的相似度,因此可以通过选取相似日进行功率预测。通过计算日特征向量的关联度和欧式距离来选取相似日。这两种方法均没有考虑不同气象因素对光伏阵列发电功率影响的重要性差异。另外由于关联度仅反映样本间的线性关系,即样本在空间曲线形状的相似性,而欧氏距离仅反映各样本在空间距离上的远近,实际情况经常有样本特征组成的曲线形状很相似而数值却存在较大差别,或者是样本间各分量的总欧氏距离很接近但样本曲线形状差异很大。这两种相似度函数都不足以反映日气象数据的本质特征,因此如果仅以关联度或欧氏距离来选取相似日,容易造成较大误差,进而影响光伏发电功率的预测精度。

### 发明内容

[0005] 为解决现有技术的不足,本发明的目的在于提供一种基于改进模糊聚类算法的光伏功率预测方法,能够减小相似日选取误差和提高光伏发电功率预测精度。

[0006] 为了实现上述目标,本发明采用如下的技术方案:

[0007] 一种基于改进模糊聚类算法的光伏功率预测方法,其特征是,包括如下步骤:

[0008] 1)采集样本数据,并进行无量纲化处理以及标准化处理;

[0009] 2)根据步骤1)的样本数据,确定日特征相关因素的权重;

[0010] 3)历史日样本模糊聚类,计算衡量本分类对象 $x_i$ 与 $x_j$ 之间的相似系数 $r_{ij}$ ,确定论域 $U$ 上的模糊关系矩阵:

[0011] 设 $S(X, Y)$ 表示样本 $X$ 和样本 $Y$ 之间的相似度,且 $S(X, Y) = \alpha D(X, Y) + \beta R(X, Y)$ ,

$$D(X, Y) = 1 - \sqrt{\frac{1}{m} \sum_{i=1}^m (\sigma_i x_i - \sigma_i y_i)^2}, \text{模糊矩阵 } R(X, Y) = \frac{\sum_{i=1}^m \sigma_i^2 x_i y_i}{\sqrt{\sum_{i=1}^m \sigma_i^2 x_i^2 \sum_{i=1}^m \sigma_i^2 y_i^2}}, \text{式中, } x_i \text{ 和 } y_i \text{ 分别是}$$

样本X和样本Y属性因子的数值,且在 $[0, 1]$ 之间, $\sigma_i$ 为样本中第*i*个属性因子的权值( $i=1, 2, \dots, m$ ), $\alpha$ 和 $\beta$ 为 $D(X, Y)$ 和 $R(X, Y)$ 权重系数, $\alpha+\beta=1$ ;

[0012] 将得到的模糊矩阵R通过逐次平方法求出其传递闭包阵 $R^* = t(R)$ ,依据经验值取适当的阈值 $\lambda$ ,对 $R^*$ 进行截割;

[0013] 对 $\forall \lambda \in [0, 1]$ ,记模糊矩阵R的 $\lambda$ -截距阵为 $t(R)_\lambda = (r_{ij}^*)$ ,其中,  $r_{ij}^* = \begin{cases} 1, & r_{ij}^* \geq \lambda \\ 0, & r_{ij}^* < \lambda \end{cases}$ ;

[0014] 按不同的 $\lambda$ 值进行分类,得到不同的聚类结果;

[0015] 4)对新样本进行分类识别,首先根据聚类情况,按下式求出各类的聚类中心:

$$\bar{x}_{ik} = \frac{1}{p} \sum_{j=1}^p x_{jk}, k=1, 2, \dots, m, \text{式中,第 } i \text{ 类包括 } p \text{ 个样本; } m \text{ 为样本特征指标的数目;然后按步骤}$$

3)的方法进行新样本的分类识别;

[0016] 5)确定预测模型并进行参数优化。

[0017] 进一步地,所述步骤2)具体内容如下:

[0018] 设有自变量 $x_1, x_2, \dots, x_m$ 和因变量 $y$ ,因变量序列为 $Y = [y_1, y_2, \dots, y_n]$ ,自变量序列

$$\text{为 } X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}, \text{自变量 } x_k \text{ 对因变量 } y \text{ 的直接途径系数 } E_{x_k \rightarrow y} = b_k \sqrt{\sum_{j=1}^n (x_{jk} - \bar{x}) / (y_j - \bar{y})},$$

自变量 $x_k$ 通过自变量 $x_{k+1}$ ,对因变量 $y$ 的间接途径系数为 $E_{x_k \rightarrow x_{k+1} \rightarrow y} = r_{k(k+1)} E_{x_k \rightarrow y}$ ,自变量 $x_{k+1}$ 通过自变量 $x_k$ 对因变量 $y$ 的间接途径系数为 $E_{x_{k+1} \rightarrow x_k \rightarrow y} = r_{k(k+1)} E_{x_{k+1} \rightarrow y}$ ,式中, $b_k$ 为偏回归系数, $r_{k(k+1)}$ 为自变量 $x_k$ 和自变量 $x_{k+1}$ 的单相关系数;

[0019] 根据上述得到的各自变量的直接途径系数,求得各自变量对因变量的影响权重

$$\sigma_i (i=1, 2, \dots, k) \text{ 为 } \sigma_i = \frac{|E_{x_i \rightarrow y}|}{\sum_{i=1}^k |E_{x_i \rightarrow y}|}.$$

[0020] 进一步地,所述步骤3)中 $\alpha=0.5, \beta=0.5$ 。

[0021] 进一步地,所述步骤3)中的最佳分类阈值 $\lambda$ 利用F-统计量进行确定。

[0022] 进一步地,所述步骤5)中模型采用BP网络模型,包括输入层、隐含层和输出层;具体建模方式如下:

[0023] 51)输入量和输出量:

[0024] 设 $t$ 为预测日,经过分类识别后 $t \in M_i, (i=1, 2, \dots, n)$ ,式中, $n$ 为历史日样本经过模糊聚类后的分类数;

[0025] 将 $M_i$ 类中包含的日样本按与预测日 $t$ 的相似度由大到小排序结果为 $t_1, t_2, t_3, \dots, t_d$ ,式中, $d$ 为 $M_i$ 类中包含的日样本数。

[0026] 选取与预测日t的相似度最大的前六日的光伏阵列输出功率序列组成的学习样本作为BP网络的输入矩阵为

$$[0027] \begin{bmatrix} P_1(t-6) & P_1(t-5) & \cdots & P_1(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ P_2(t-6) & P_2(t-5) & \cdots & P_2(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ P_{24}(t-6) & P_{24}(t-5) & \cdots & P_{24}(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \end{bmatrix}, \text{式中, } t \text{ 为预测日;}$$

[0028]  $t-x(1 \leq x \leq 6)$  为与预测日相似度最大的前x天;  $P_y(1 \leq y \leq 24)$  为相似日第yh光伏阵列的功率值;  $R_1(t), R_2(t), \dots, R_6(t)$  分别为取量化值第t天的日最高温度、日最低温度、日平均温度、日相对湿度、日平均风速、天气类型;

[0029] BP网络的输出量为预测日的24h光伏阵列功率值, 即  $[P_1(t) P_2(t) \cdots P_{24}(t)]$ ;

[0030] 52) 隐含层:

[0031] 隐含层节点数的确定采用经验公式  $H = \sqrt{I+O} + a$ , 式中, H为隐含层节点数; I为输入层节点数; O为输出层节点数; a为1~10之间的常数;

[0032] 53) 转移函数:

[0033] 该网络隐含层的转移函数采用双曲正切型S函数, 即  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , 输出层的转移函数采用线性函数, 即  $f(x) = x$ ;

[0034] 54) BP神经网络的参数优化:

[0035] 通过差分进化算法进行参数优化; 获得最优个体后, 将最优个体解码后初始化BP神经网络参数, 然后利用样本训练BP神经网络, 利用训练好的BP神经网络进行功率预测。

[0036] 本发明所达到的有益效果: 本方法考虑了不同气象因素对光伏阵列输出功率的影响程度, 构造了既考虑值相似程度又考虑行相似程度的相似度统计量, 选取的预测日的相似样本能够更好地反映预测日的输出功率特性, 并且利用差分进化算法对构建的BP神经网络进行了参数优化, 可有效提高预测模型的预测精度, 具有一定的实用性及可行性。

## 附图说明

[0037] 图1是基于改进模糊聚类算法的光伏功率预测方法流程图;

[0038] 图2是差分进化算法优化BP神经网络流程图;

[0039] 图3是本发明所提方法光伏功率预测结果曲线;

[0040] 图4是采用改进相似度模糊聚类, 参数未优化的BP神经网络光伏功率预测结果曲线;

[0041] 图5是采用关联度选择相似日, 参数未优化的BP神经网络光伏功率预测结果曲线。

## 具体实施方式

[0042] 下面结合附图对本发明作进一步描述。以下实施例仅用于更加清楚地说明本发明的技术方案, 而不能以此来限制本发明的保护范围。

[0043] 相较于传统方法流程图, 传统的聚类分析中, 采用的相似程度统计量主要分为距离系数和相似系数, 二者均没有考虑样本各个属性因子之间的重要度差异。另外, 距离系数统计能反映样本间的值相似程度, 相似系数统计反映样本间的形相似程度, 均具有

一定的局限。为了能够反映样本属性因子间的重要度差异和综合考虑样本间值贴近和形贴近,本发明提出一种基于改进模糊聚类算法的光伏功率预测方法。下文以2012年10月6日~10月8日内蒙古某地的光伏电站发电功率预测为例,给出其功率预测方法。

[0044] 具体实施过程如下:

[0045] 1)样本数据标准化处理

[0046] 由于各个指标的量纲和数量级都不同,直接用原始数据进行计算,可能会突显某些数量级大的特性指标对分类的作用。所以,必须对原始数据进行无量纲化处理。

[0047] 2)日特征相关因素权重确定

[0048] 通过光伏发电量与气象因子间的相关性分析得到影响光伏阵列发电量的气象因子主要有大气温度、太阳辐照度、相对湿度、风速、天气类型。由于气象部门提供的气象信息缺乏光照辐射数据,因此选择日最高温度,日最低温度,日平均温度,日相对湿度,日平均风速,天气类型作为影响光伏日发电量的日特征相关因素。

[0049] 在计算日特征向量的相似度时,如果采用平均权重,这种平均权重存在局部相似倾向,在测度值离散的情况下,由点相似测度值大的点决定总体的相似度,造成计算误差。本发明采用通径分析方法确定各个日特征相关因素中的权重系数。

[0050] 通径分析方法通过计算自变量和因变量的直接通径系数和间接通径系数来表示自变量与因变量之间的关系。通径系数能有效地表示相关变量间原因对结果的直接影响或间接影响效应,从而区分因子的相对重要性及其关系。

[0051] 对于一般的多元线性回归分析,设有自变量 $x_1, x_2, \dots, x_k$ 和因变量 $y$ 。因变量序列为

$$Y = [y_1, y_2, \dots, y_n], \text{自变量序列为 } X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}。$$

[0052] 则自变量 $x_k$ 对因变量 $y$ 的直接通径系数为: $E_{x_k \rightarrow y} = b_k \sqrt{\sum_{j=1}^n (x_{jk} - \bar{x}) / (y_j - \bar{y})}$ ,自变量 $x_k$ 通过自变量 $x_{k+1}$ ,对因变量 $y$ 的间接通径系数为 $E_{x_k \rightarrow x_{k+1} \rightarrow y} = r_{k(k+1)} E_{x_{k+1} \rightarrow y}$ 。

[0053] 反过来,自变量 $x_{k+1}$ ,通过自变量 $x_k$ 对因变量 $y$ 的间接通径系数为: $E_{x_{k+1} \rightarrow x_k \rightarrow y} = r_{k(k+1)} E_{x_k \rightarrow y}$ ,式中: $b_k$ 为偏回归系数; $r_{k(k+1)}$ 为自变量 $x_k$ 和自变量 $x_{k+1}$ 的单相关系数。

[0054] 根据上述得到的各自变量的直接通径系数,便可求得各自变量的对因变量的影响

$$\text{权重 } \sigma_i (i=1, 2, \dots, k) \text{ 为: } \sigma_i = \frac{|E_{x_i \rightarrow y}|}{\sum_{i=1}^k |E_{x_i \rightarrow y}|}。$$

[0055] 以内蒙古某光伏并网发电系统2012年发电数据和当地气象站同期数据为样本,选取日特征相关因素:日最高温度( $T_{max}$ ),日最低温度( $T_{min}$ ),日平均温度( $T_{avg}$ ),日相对湿度(RH),日平均风速( $V$ ),天气类型( $W$ )为自变量,以光伏阵列日发电量( $G$ )为因变量,将各因素进行量化并映射至 $[0, 1]$ 区间,然后进行通径分析,分析结果见表1。

[0056] 表1日特征相关因素与光伏阵列发电量之间的通径分析结果

气象因子	直接途径	间接途径					
		$T_{max}$	$T_{min}$	$T_{avg}$	$RH$	$V$	$W$
[0057]	$T_{max}$	0.1481	-0.3281	1.0053	0.0407	-0.1385	0.0987
	$T_{min}$	0.0542	-0.0346	0.9934	0.0435	0.1218	-0.0134
	$T_{avg}$	0.2477	-0.0419	-0.3949	0.0407	0.0184	-0.0581
	$RH$	-0.1238	0.0025	0.0734	-0.1727	-0.1534	0.0277
[0058]	$V$	0.0443	0.0105	-0.0882	0.0335	0.0658	0.1257
	$W$	1.0708	-0.0159	0.0207	-0.2247	-0.0253	0.2673

[0059] 由途径系数可知,天气类型对光伏阵列发电量的影响最大,日平均风速对光伏阵列发电量的影响最小,可以得到日特征相关因素对光伏发电量的影响权重为: $\sigma_{T_{max}}=0.088, \sigma_{T_{min}}=0.032, \sigma_{T_{avg}}=0.147, \sigma_{RH}=0.073, \sigma_V=0.036, \sigma_W=0.624$ 。

[0060] 3)历史日样本模糊聚类:

[0061] 模糊聚类是根据客观事物间的不同特征、亲疏程度和相似性等关系并通过建立模糊相似关系对客观事物进行分类的数学方法。

[0062] 本发明选用模糊聚类分析中基于模糊等价关系的传递闭包法。

[0063] 设论域U为n个待分类样本,即 $U=\{x_1, x_2, \dots, x_n\}$ ,且每个样本都有m个指标表示其性状,即 $x_i=\{x_{i1}, x_{i2}, \dots, x_{im}\} (i=1, 2, \dots, n)$ 。

[0064] 由此可得到论域的原始数据矩阵为 $U = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$ 。

[0065] 计算衡量本分类对象 $x_i$ 与 $x_j$ 之间的相似系数 $r_{ij}$ ,从而确定论域U上的模糊关系矩阵。为了能够反映样本属性因子间的重要度差异和综合考虑样本间的值贴近和形贴近,本发明提出了一种改进的相似度统计量 $S(X, Y)$ 。

[0066] 设 $S(X, Y)$ 表示样本X和样本Y之间的相似度: $S(X, Y) = \alpha D(X, Y) + \beta R(X, Y)$ ,

$$D(X, Y) = 1 - \sqrt{\frac{1}{m} \sum_{i=1}^m (\sigma_i x_i - \sigma_i y_i)^2}, \quad R(X, Y) = \frac{\sum_{i=1}^m \sigma_i^2 x_i y_i}{\sqrt{\sum_{i=1}^m \sigma_i^2 x_i^2 \sum_{i=1}^m \sigma_i^2 y_i^2}}$$

式中, $x_i$ 和 $y_i$ 分别是样本X和

样本Y属性因子的数值,且在 $[0, 1]$ 之间; $\sigma_i$ 为样本中第i个属性因子的权值( $i=1, 2, \dots, n$ ); $\alpha$ 和 $\beta$ 为 $D(X, Y)$ 和 $R(X, Y)$ 权重系数, $\alpha + \beta = 1$ (本发明中取 $\alpha = 0.5, \beta = 0.5$ )。

[0067] 将得到的模糊矩阵R通过逐次平方法求出其传递闭包阵 $R^* = t(R)$ ,取适当的阈值 $\lambda$ ,对 $R^*$ 进行截割,对任意 $\lambda \in [0, 1]$ ,记为 $t(R)_\lambda = (r_{ij}^*)$ 称为R的 $\lambda$ -截距阵。其中:



$$\lambda r_{ij}^* = \begin{cases} 1, & r_{ij}^* \geq \lambda \\ 0, & r_{ij}^* < \lambda \end{cases}.$$

[0068] 按不同的 $\lambda$ 值进行分类,得到不同的聚类结果。最佳分类阈值 $\lambda$ 可以利用F-统计量进行确定。

[0069] 4)新样本分类识别

[0070] 将样本进行分类之后,有新的样本时可以通过模式识别确定新样本的类别。首先

根据聚类情况,按下式求出各类的聚类中心: $\bar{x}_{ik} = \frac{1}{p} \sum_{j=1}^p x_{jk}$ ,  $k=1, 2, \dots, m$ , 式中:第 $i$ 类包括 $p$ 个样本; $m$ 为样本特征指标的数目。

[0071] 然后,按下式计算新样本与上述各聚类中心的相似度,以相似度最大的类别作为新样本的类别。

[0072] 设 $S(X, Y)$ 表示样本 $X$ 和样本 $Y$ 之间的相似度; $S(X, Y) = \alpha D(X, Y) + \beta R(X, Y)$ ,

$$D(X, Y) = 1 - \sqrt{\frac{1}{m} \sum_{i=1}^m (\sigma_i x_i - \sigma_i y_i)^2}, \quad R(X, Y) = \frac{\sum_{i=1}^m \sigma_i^2 x_i y_i}{\sqrt{\sum_{i=1}^m \sigma_i^2 x_i^2 \sum_{i=1}^m \sigma_i^2 y_i^2}},$$

式中: $x_i$ 和 $y_i$ 分别是样本 $X$

和样本 $Y$ 属性因子的数值,且在 $[0, 1]$ 之间; $\sigma_i$ 为样本中第 $i$ 个属性因子的权值( $i=1, 2, \dots, n$ ); $\alpha$ 和 $\beta$ 为 $D(X, Y)$ 和 $R(X, Y)$ 权重系数, $\alpha + \beta = 1$ (本发明中取 $\alpha = 0.5, \beta = 0.5$ )。

[0073] 5)预测模型的确定及参数优化

[0074] 本发明建立的BP网络模型包括输入层,隐含层和输出层三层。

[0075] (1)输入量和输出量

[0076] 设 $t$ 为预测日,经过分类识别后: $t \in M_i$ , ( $i=1, 2, \dots, n$ ), 式中: $n$ 为历史日样本经过模糊聚类后的分类数。将 $M_i$ 类中包含的日样本按与预测日 $t$ 的相似度由大到小排序结果为 $t_1, t_2, t_3, \dots, t_d$ , 式中: $d$ 为 $M_i$ 类中包含的日样本数。

[0077] 选取与预测日 $t$ 的相似度最大的前六日的光伏阵列输出功率序列组成的学习样本作为BP网络的输入矩阵为

$$[0078] \begin{bmatrix} P_1(t-6) & P_1(t-5) & \cdots & P_1(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ P_2(t-6) & P_2(t-5) & \cdots & P_2(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ P_{24}(t-6) & P_{24}(t-5) & \cdots & P_{24}(t-1) & R_1(t) & R_2(t) & \cdots & R_6(t) \end{bmatrix},$$

式中: $t$ 为预测日;

[0079]  $t-x$  ( $1 \leq x \leq 6$ )为与预测日相似度最大的前 $x$ 天; $P_y$  ( $1 \leq y \leq 24$ )为相似日第 $y$ h光伏阵列的功率值; $R_1(t), R_2(t), \dots, R_6(t)$ 分别为取量化值第 $t$ 天的日最高温度、日最低温度、日平均温度、日相对湿度、日平均风速、天气类型(已取为量化值)。

[0080] BP网络的输出量为预测日的24h光伏阵列功率值,即 $[P_1(t) P_2(t) \cdots P_{24}(t)]$ 。

[0081] (2)隐含层

[0082] 隐含层节点数的确定采用经验公式 $H = \sqrt{I+O} + a$ , 式中: $H$ 为隐含层节点数; $I$ 为输入层节点数; $O$ 为输出层节点数; $a$ 为 $1 \sim 10$ 之间的常数,计算出来的值经过四舍五入取整,用该公式确定的隐含层节点数为9。

## [0083] (3)转移函数

[0084] 该网络隐含层的转移函数采用双曲正切型S函数,即 $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ ,输出层的转移函数采用线性函数,即 $f(x) = x$ 。

## [0085] (4)BP神经网络的参数优化

[0086] BP神经网络通过具有简单处理能力的神经元的复合作用使网络具有复杂的非线性映射能力,但该神经网络模型存在着局部最小化和收敛速度慢的缺点。差分进化算法作为一种随机的并行直接搜索算法,通过群体内个体间的合作与竞争产生的群体智能指导优化搜索,同遗传进化算法相比降低了复杂性,具有记忆个体最优解和种群内信息共享的特点,可弥补BP神经网络连接权值和阈值选择上的随机性缺陷,在不影响BP神经网络泛化的映射能力前提下,可以使BP神经网络具有更快的收敛性以及较强的学习能力。

[0087] 差分进化算法优化BP神经网络参数的过程如图2所示。

[0088] 获得最优个体后,将最优个体解码后初始化BP神经网络参数,然后利用样本训练BP神经网络,利用训练好的BP神经网络进行功率预测。

[0089] 2012年10月6日~10月8日的气象信息见表2。首先,将预测日(10月6日~10月8日)之前的历史日气象特征因素(日最高温度,日最低温度,日平均温度,日相对湿度,日平均风速,天气类型)进行量化并映射至 $[0, 1]$ 区间后形成聚类样本。然后,利用改进的相似度统计量建立模糊关系相似矩阵,构造传递闭包 $R^*$ ,根据F-统计量法确定最佳分类阈值 $\lambda$ 为0.6452,将聚类样本分为了六类。最后,通过新样本分类识别根据预测日气象信息确定其所属类别,该类包含样本即为BP神经网络的训练样本,利用差分进化算法对BP神经网络的参数进行寻优,然后对BP网络进行训练。最后,利用训练好的网络对预测日的输出功率进行预测,预测结果如图3所示。

## [0090] 表2气象信息表

日期	天气类型	最高温度 /°C	最低温 度/°C	平均温 度/°C	平均风 速/(m/s)	相对湿 度/%
[0091] 2012-10-06	晴	19	3	10	5	23
2012-10-07	晴转多云	17	6	9	4	31
[0092] 2012-10-08	多云转小雨	17	6	7	5	86

[0093] 采用为经过差分进化算法进行参数优化的BP神经网络进行预测,预测结果如图4所示。

[0094] 通过计算日特征向量的关联度选择预测日的相似日,然后构造BP网络进行预测,预测结果如图5所示。

[0095] 三种预测方法的预测误差如表3所示。

## [0096] 表3预测误差比较

时间	绝对平均误差 $e_{DMAPE}/\%$			均方根误差 $e_{RMSE}/\%$		
	A	B	C	A	B	C
[0097] 2012-10-06	3.68	5.04	5.12	1.42	1.58	2.64
2012-10-07	9.72	11.98	14.34	2.41	3.06	4.11
2012-10-08	11.06	13.67	19.62	3.28	5.14	6.37
总误差	8.15	10.23	14.04	2.31	3.22	4.35

[0098] 注:A为本发明所建立的模型;B为采用改进相似度模糊聚类,构建的BP神经网络未进行参数优化;C为计算日特征向量的关联度选择预测日的相似日,构建的BP神经网络未进行参数优化。

[0099] 
$$e_{DMAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_i - \hat{P}_i}{P_i} \right| \times 100\%, \quad e_{RMSE} = \frac{1}{P} \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \hat{P}_i)^2}$$
 ,式中: $e_{DMAPE}$ 和 $e_{RMSE}$ 分别

为绝对平均误差、均方根误差; $P_i$ 为光伏阵列输出功率实测值; $\hat{P}_i$ 为输出功率预测值; $N$ 为预测样本数目; $P$ 为光伏阵列额定输出功率。

[0100] 通过对图3~图5和表3的分析,可知:

[0101] (1)本发明所提出的光伏功率预测方法比传统的基于相似日选取的光伏功率预测方法具有更高的预测精度,预测精度能够提高2%~5%。

[0102] (2)当天气类型为晴天时,本发明方法与传统相似日选取方法都有较高的预测精度,当天气状况较差时,本发明所提方法预测精度要明显优于传统相似日选取方法。

[0103] (3)经过差分进化优化后的BP神经网络预测模型比原始的BP神经网络预测模型具有更高的预测精度,预测精度能够提高约2%。

[0104] (4)由于本发明方法考虑了不同气象因素对光伏阵列输出功率的影响程度,构造了既考虑值相似程度又考虑行相似程度的相似度统计量,选取的预测日的相似样本能够更好地反映预测日的输出功率特性。

[0105] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明技术原理的前提下,还可以做出若干改进和变形,这些改进和变形也应视为本发明的保护范围。

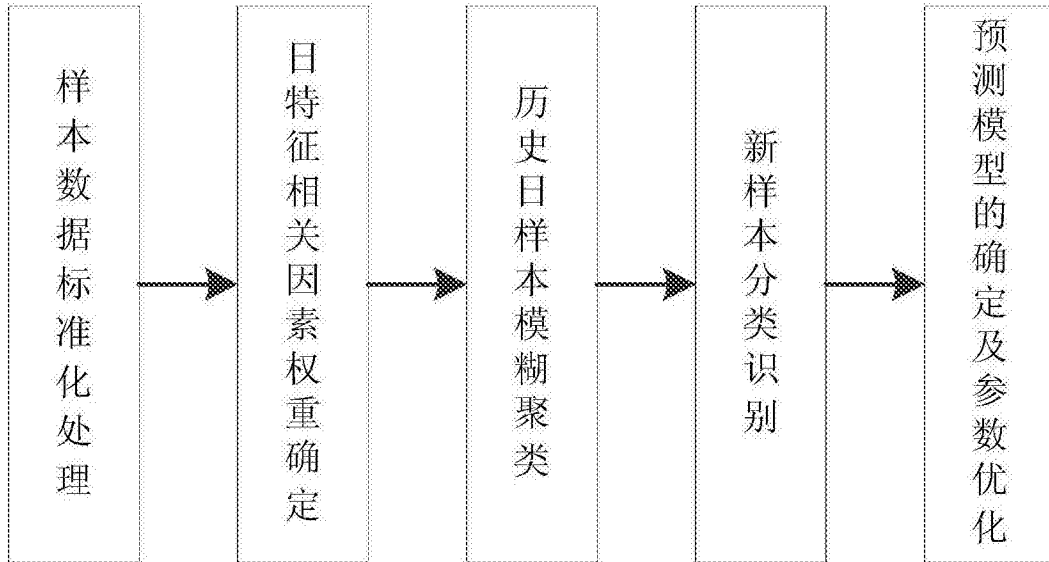


图1

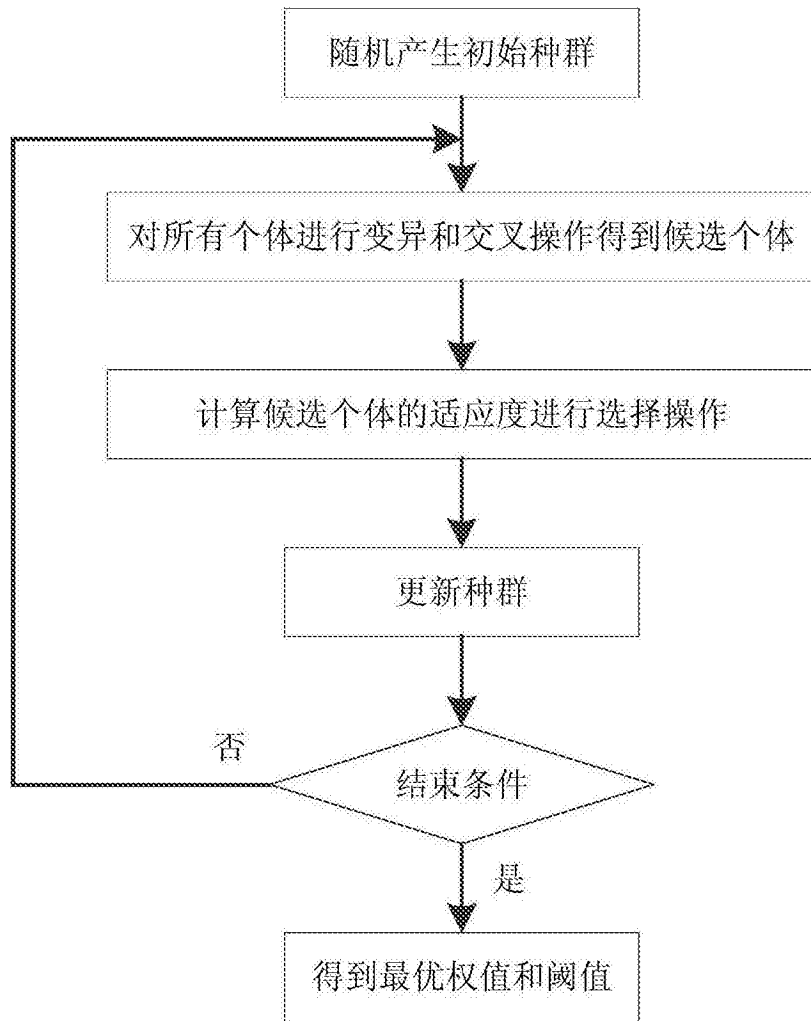


图2

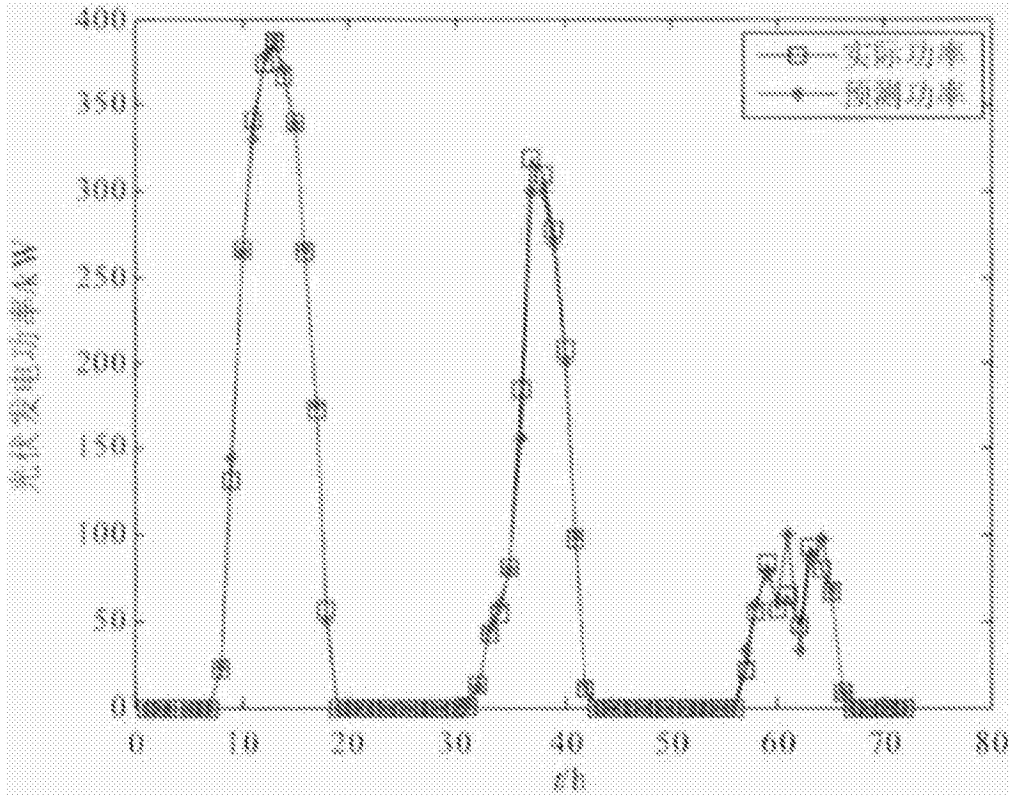


图3

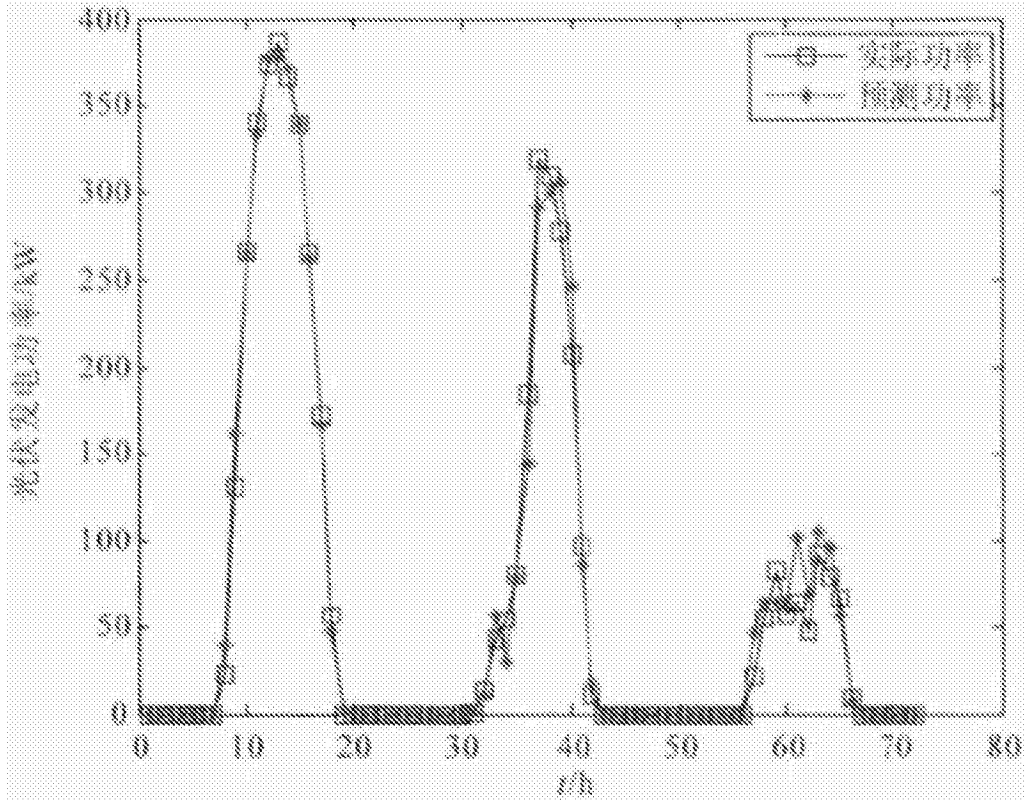


图4

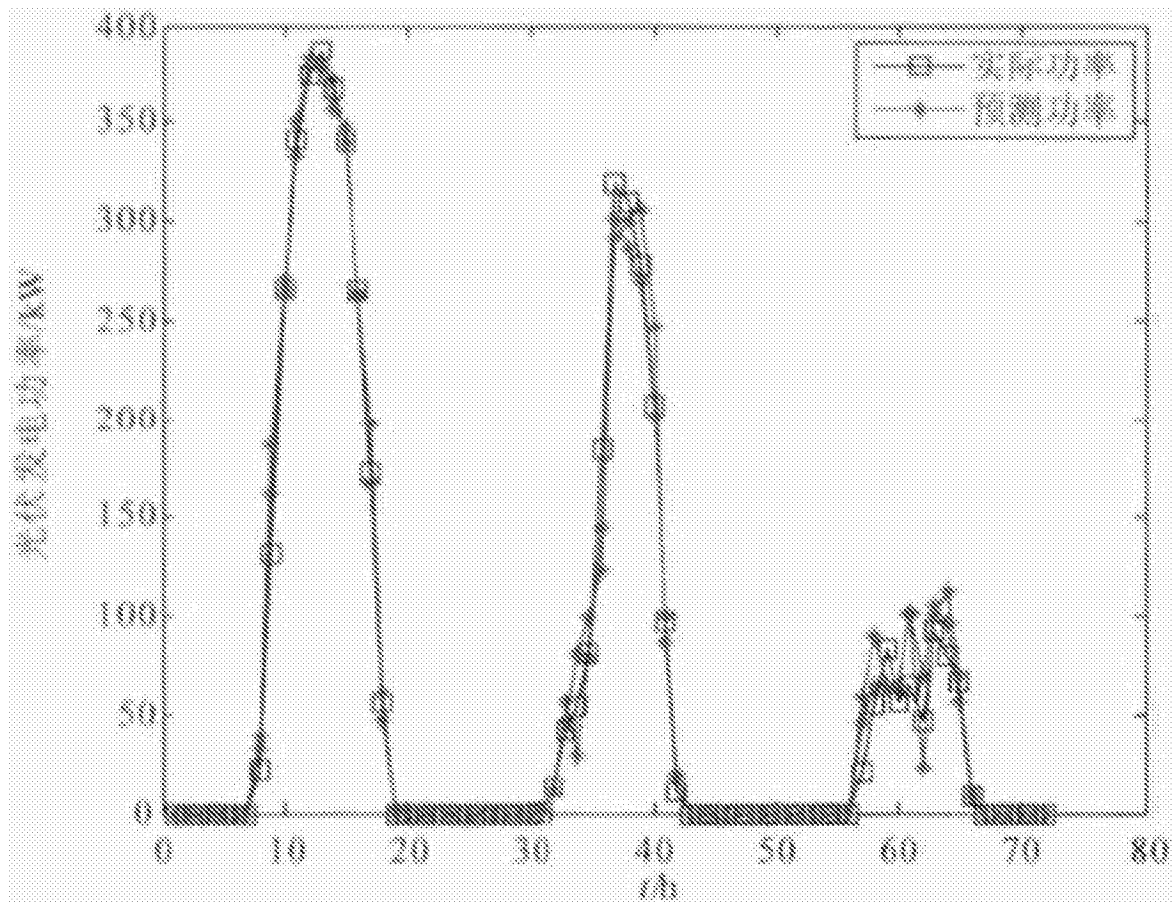


图5