

# 基于相似日和交叉熵理论的光伏发电功率组合预测

季顺祥<sup>1,2</sup>, 王琦<sup>1,2</sup>, 姚阳<sup>1,2</sup>, 陈佳浩<sup>1,2</sup>, 刘瑾<sup>1,2</sup>

(1.南京师范大学南瑞电气与自动化学院,江苏 南京 210042)

(2.南京师范大学江苏省气电互联综合能源工程实验室,江苏 南京 210023)

**[摘要]** 为进一步提高光伏发电功率预测精度,提出一种基于相似日和交叉熵理论的光伏发电短期功率组合预测方法。首先采用模糊 C 均值聚类方法对历史样本数据分类,并提出一种基于隶属度的指标来选取相似日。然后采用最小二乘支持向量机、时间序列法和 BP 神经网络法分别预测光伏发电功率,通过交叉熵算法动态设置各预测时刻下单一方法的权重值,建立光伏发电功率的组合预测模型。算例结果表明,所提方法能够动态识别单一预测方法包含的信息量,能确定更加合理的权重值,从而提高光伏发电功率的预测精度。

**[关键词]** 光伏发电,组合预测,相似日,隶属度,交叉熵

**[中图分类号]** TM615 **[文献标志码]** A **[文章编号]** 1672-1292(2018)02-0019-10

## Photovoltaic Power Generation Combination Forecasting Based on Similar Days and Cross Entropy Theory

Ji Shunxiang<sup>1,2</sup>, Wang Qi<sup>1,2</sup>, Yao Yang<sup>1,2</sup>, Chen Jiahao<sup>1,2</sup>, Liu Jin<sup>1,2</sup>

(1.School of NARI Electrical and Automation, Nanjing Normal University, Nanjing 210042, China)

(2.Jiangsu Key Laboratory of Gas and Electricity Interconnection Integrated Energy, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** In order to further improve the photovoltaic (PV) power forecasting accuracy, a short-term combination forecasting model based on similar days and cross entropy theory is proposed. Firstly, the fuzzy C-means clustering method is used to classify the historical samples, and a selection index based on membership degree is proposed to select similar days. Then, the LSSVM, ARMA and BP neural network are used to predict the PV power. The weights of three single forecasting methods are dynamically set by the cross entropy algorithm, and the short-term combination forecasting model of PV power is established. The results show that this method can dynamically identify the information of single methods and obtain appropriate weights. As a result, the forecasting accuracy of PV power can be improved.

**Key words:** photovoltaic (PV) power generation, combination forecasting, similar days, membership, cross entropy

随着新能源的不断开发与利用,以太阳能、风能为代表的大规模可再生能源并网发电已成为新型电力系统必然的发展趋势<sup>[1]</sup>。受太阳辐照强度、环境温度和湿度等气象因素的影响,光伏发电功率具有随机性、间歇性与波动性,从而使得大规模光伏发电并网对电网造成不良影响。如果能够准确地预测光伏发电功率,则对电网的调度与光伏电站的运行控制有着重大意义<sup>[2-4]</sup>。

目前,光伏发电功率预测主要有两种方法:直接预测法和间接预测法<sup>[5]</sup>。间接法是根据气象预报的历史数据对光照强度、温度进行预测,然后通过相关公式或算法计算得出光伏电站出力预测值。文献[6]依靠欧洲中尺度天气预报中心对太阳辐射预测模型进行改进以获取具体预报时间内的辐射值,最终得到光伏出力值,但该方法过于依赖准确的气象预报信息。直接法是利用光伏电站的历史出力数据以及相关的天气预报值,对其进行预测。文献[7]采用 BP 神经网络方法对光伏输出功率进行预测,虽然考虑了天气因素,但预测精度还有待进一步提高。文献[8]建立时间序列法(auto-regressive and moving average, ARMA)模型和马尔科夫链预测模型对光伏电站短期出力预测,但是该模型只适用于晴天,对复杂天气类型的预测并不适用。文献[9]建立了基于径向基神经网络的光伏预测模型,结合 K 均值聚类和遗传算法来优化径向

收稿日期:2018-01-18.

基金项目:江苏省研究生科研与实践创新计划项目(KYCX17\_1078).

通讯联系人:王琦,博士,副教授,研究方向:可再生能源发电技术. E-mail:wangqi@njnu.edu.cn

基神经网络,预测精度较传统的模糊神经网络预测模型更高,但预测误差仍可进一步降低.文献[10]利用 LSSVM 建立光伏发电功率预测模型,分析了太阳辐射量、气温、季节以及天气类型等因素对光伏阵列发电的影响,预测光伏电站输出功率,但该方法的预测精度受模型参数的影响.

上述预测方法均采用单一的预测方法,由于单一预测方法具有一定的局限性,预测精度仍可进一步提高.因此,适当地采取组合预测方法能够更好地提高预测精度.文献[11]提出了改进灰色 BP 神经网络的光伏出力组合预测方法,采用模糊 C 均值对历史出力数据统计分类,来选取相似日作为预测模型训练样本,但没有明确的相似日选取指标.文献[12]采用熵权法建立光伏输出功率组合预测模型,但其设置的单一预测方法的权重值固定不变,未能实时动态反映光伏出力的变化,具有一定的局限性.为此,本文提出一种基于相似日和交叉熵理论的光伏发电短期功率组合预测模型.首先采用模糊 C 均值聚类方法对历史数据统计分类,并在此基础上提出一种基于隶属度的指标来选取相似日,以获取更佳的预测模型训练样本.再采用 LSSVM、ARMA 和 BP 神经网络法 3 种单一方法分别对待预测日的发电功率进行预测.然后通过交叉熵算法动态设置各预测时刻点下单一方法的权重值,建立了基于相似日和交叉熵理论的光伏发电短期组合预测模型.通过算例分析,并分别与基于误差平方和的组合模型以及基于相关系数的组合模型进行对比,验证了本文所提交叉熵组合预测模型的正确性与优越性.

## 1 相似日的选取

### 1.1 影响光伏发电功率的因素分析

光伏发电功率与诸多因素有关,主要包括太阳辐照度、温度、湿度、风速、天气类型等,光伏阵列输出功率计算如下<sup>[13]</sup>:

$$P_s = \eta SI [1 - 0.005(t_0 + 25)]. \quad (1)$$

式中, $\eta$  为光伏电池转换效率(%), $S$  为光伏阵列面积( $\text{m}^2$ ), $I$  为太阳辐照强度( $\text{kW}/\text{m}^2$ ), $t_0$  为工作环境温度( $^{\circ}\text{C}$ ).在短期的发电功率预测中,光伏发电系统的阵列面积  $S$  和转换效率  $\eta$  是不变的.因此,由式(1)可知发电功率主要受到太阳辐照强度和环境温度 2 个因素的影响.其中,由于季节和天气类型的不同,太阳辐照度也有较大差异,从而光伏阵列发电功率也有很大差别.考虑到气象观测站一般难以提供太阳辐照度数据,本文以日照时数来代替,选取日照时数、日平均温度、日平均湿度、日平均风速作为影响光伏发电功率的特征量,并采用模糊 C 均值算法对样本进行分类.

### 1.2 相似日选取原理

考虑到常规聚类方法只是将样本数据粗略分类,实际并不精确.为进一步获取相似度更高的数据样本,在模糊 C 均值聚类(fuzzy-C-means, FCM)方法的基础上,提出一种基于隶属度的指标来选取相似日.具体选取过程如下:

#### (1) 确定聚类样本数据集

首先确定待预测日并且获取待预测日的气象特征.对于样本总数  $M$ ,将待预测日气象数据列入历史样本数据集中,则样本总数为  $(M+1)$ .

FCM 聚类计算目标表达式如下<sup>[14]</sup>:

$$J(U, V) = \sum_{i=1}^{M+1} \sum_{r=1}^N (\mu_r(i)^m X_i - V_r^2), 1 \leq m \leq \infty. \quad (2)$$

式中, $U$  是隶属度矩阵, $V$  是各个聚类中心位置向量, $J(U, V)$  是目标函数, $N$  是聚类中心数, $m$  是权重系数, $X_i (i \in \{1, 2, \dots, M+1\})$  是样本里每个数据的聚类特征, $V_r$  是第  $r$  个聚类中心的位置向量, $\mu_r(i)$  是第  $i$  个数据点对于第  $r$  个聚类中心的隶属度.

#### (2) 确定聚类特征与聚类中心数

选取日照时数、日平均温度、日平均湿度和日平均风速作为聚类特征,将历史日分为 4 种基本类型:晴天、多云、阴天和小雨,即聚类中心  $N=4$ .

#### (3) 隶属度矩阵 $U$ 赋初值

$$U^g = [\mu_r(i)^{(g)}] \in \mathbf{R}_{(M+1) \times N}. \quad (3)$$

式中, $g$  为迭代次数, $\mathbf{R}_{(M+1) \times N}$  为有理数矩阵.

(4)更新聚类中心位置向量

第  $g+1$  次的聚类  $r$  的中心位置向量是

$$V_r = \frac{\sum_{i=1}^{M+1} (\mu_r(i)^{(g)})^m X_i}{\sum_{i=1}^{M+1} (\mu_r(i)^{(g)})^m}, \quad 1 \leq r \leq N. \quad (4)$$

(5)隶属度计算

各个数据点  $X_i$  的隶属度为

$$\mu_r(i)^{(k)} = \frac{1}{\sum_{j=1}^N \left[ \frac{\|X_i - V_r\|}{\|X_i - V_j\|} \right]^{\frac{2}{m-1}}}. \quad (5)$$

(6)迭代结束判断

根据迭代结束条件进行判断是否收敛,迭代结束条件为

$$\|\mu_r(i)^{(g)} - \mu_r(i)^{(g-1)}\| < \varepsilon. \quad (6)$$

式中,  $\varepsilon > 0$  为迭代计算结束条件.

(7)计算相似度并选取相似日

将与待预测日属于同一聚类中心的历史日作为选取样本,其中第  $i$  个历史日与待预测日的相似度为

$$\chi_i = \left| \frac{\mu_r(M+1) - \mu_r(i)}{\mu_r(M+1)} \right|. \quad (7)$$

式中,  $\chi_i$  越小则该历史日与待预测日的相似度就越高. 这样选取避免了再去采用其他相似性指标计算的繁琐,选取过程变得更加简易. 这里,相似日选取按  $\chi_i$  从小到大进行选取,选取前  $t$  个历史日数据作为预测样本,其中  $t$  取 6.

## 2 单一预测方法

### 2.1 单一预测方法的选择

光伏发电功率与太阳辐照度、天气类型、季节类型、气温和湿度等诸多因素有关,难以用确定的数学模型加以描述. 常用的单一预测方法有 ARMA、BP 神经网络、LSSVM 等,其中 ARMA 为线性模型,可以较好地对数据的整体趋势分析预测. BP 神经网络和 LSSVM 模型为非线性模型,具有较强的非线性学习能力并且能通过样本数据训练学习得到其内在关系. 为此,选取上述 3 种单一方法来预测光伏发电功率.

### 2.2 ARMA 模型

ARMA 模型的主要思想是当一个时间序列是依赖于时间的一族变量时,构成该时序的单个序列值虽然具有不确定性,但整个序列的变化却有一定的规律性,可以用相应的数学模型近似描述. 设为零均值平稳序列,若其满足模型

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (8)$$

则称  $x_t$  为阶数为  $(p, q)$  的自回归滑动平均序列,简记为 ARMA  $(p, q)$  序列. 其中,  $\varepsilon_t$  为白噪声序列,  $\phi_t$  和  $\theta_t (t \in \mathbf{Z})$  为权重. 本文 ARMA 模型的输入变量为相似日内各预测时刻的光伏发电功率值,其预测过程主要包括序列平稳性检验、模型参数估计和模型定阶. 平稳性采用 ADF 单位根检验,模型参数采用最小二乘法估计,模型阶数根据 AIC 准则确定.

### 2.3 BP 神经网络

BP 神经网络是一种误差反向传播训练的多层前馈神经网络,具有良好的自组织学习能力,它可以实现从输入到输出的任意非线性映射. 其网络预测模型主要是通过输入信号正向传播和误差信号反向传播来实现训练过程的,能够对大规模数据进行并行处理,具有一定的鲁棒性和容错性, BP 神经网络基本结构如图 1 所示,是输入层和隐含层节点之间的连接权值,是隐含层和输出层节点之间的连接权值. 这里,选取相似度较高的前 6 日数据作为训练样本.

## 2.4 LSSVM

LSSVM 是由 Suykens 等人在原来支持向量机基础上,继承其优点并对其改进的一种机器学习方法. LSSVM 具有泛化能力更强和解决小样本、非线性、高维数以及局部极小等问题的优点.

LSSVM 实际主要用来解决函数二次优化问题,其目标函数可以描述为

$$\min \left( \frac{\| \mathbf{w} \|^2}{2} + \gamma \sum_{k=1}^N e_k^2 \right), \quad (9)$$

$$\text{s.t. } \mathbf{y}_k = \mathbf{w}^T \phi(\mathbf{x}_k) + b + e_k.$$

式中,  $\mathbf{w}$  为加权向量,  $\gamma$  为惩罚因子,  $e_k$  为误差变量,  $b$  为常量,  $\mathbf{x}_k$  和  $\mathbf{y}_k$  分别为输入和输出向量.

选取径向基函数作为核函数,表示为

$$\psi(\mathbf{x}, \mathbf{x}_k) = \exp \left( -\frac{\| \mathbf{x} - \mathbf{x}_k \|^2}{2\sigma^2} \right), \quad (10)$$

则 LSSVM 的估计函数为

$$\hat{y}(\mathbf{x}) = \sum_{k=1}^N \alpha_k \psi(\mathbf{x}, \mathbf{x}_k) + b. \quad (11)$$

考虑到惩罚因子  $\gamma$  和核参数  $\sigma^2$  对模型性能影响最大,本文采用交叉验证来对参数寻优. LSSVM 模型的输入变量为相似日的光伏发电功率、相似日的日照时长、平均温度、最高温度和最低温度. 预测日的日照时长、平均温度、最高温度和最低温度.

## 3 交叉熵组合预测模型

### 3.1 交叉熵理论

交叉熵(cross entropy, CE)用于度量两个随机向量之间信息量的差异,也可用于描述两个概率分布间的差异性. 交叉熵在同一概率空间两个概率测度和的定义分为以下两种情况<sup>[15]</sup>:

离散情况:

$$\begin{cases} D(\mathbf{f} \parallel \mathbf{g}) = \sum_{i=1}^n f_i \log_c \frac{f_i}{g_i}, \\ \sum_{i=1}^n f_i = 1, \sum_{i=1}^n g_i = 1. \end{cases} \quad (12)$$

连续情况:

$$\begin{cases} D(\mathbf{f} \parallel \mathbf{g}) = \int_a^b \mathbf{f}(x) \log_c \frac{\mathbf{f}(x)}{\mathbf{g}(x)} dx, \\ \int_a^b \mathbf{f}(x) dx = 1, \int_a^b \mathbf{g}(x) dx = 1. \end{cases} \quad (13)$$

式中,  $\mathbf{f}$  和  $\mathbf{g}$  在式(12)和(13)中分别表示概率矢量和概率密度函数.

性质:  $D(\mathbf{f} \parallel \mathbf{g}) \geq 0$ , 当且仅当  $\mathbf{f} = \mathbf{g}$  时, 等号成立, 并且  $D(\mathbf{f} \parallel \mathbf{g}) \neq D(\mathbf{g} \parallel \mathbf{f})$ .

交叉熵算法作为一种随机优化方法,最先用于模拟小概率事件,后来用来扩展到求解最优化问题. 交叉熵方法目前已被应用于解决多目标优化、组合预测和机器学习等领域的相关实际问题. 在本文所提交叉熵的组合预测模型中,交叉熵表示两个概率密度分布的相互支持程度,即单一预测方法对整体支持程度的大小.

### 3.2 模型建立步骤

基于相似日和交叉熵理论的光伏发电短期功率组合预测模型建立具体步骤如下:

(1) 对待预测日内各预测点的功率进行预测

对于预测时间段  $T$  (时间间隔为 0.5 h) 的光伏发电功率,分别采用  $M$  种算法进行预测,预测样本采用第 1 节相似日选取方法获取,则第  $m$  ( $m \in \{1, 2, \dots, M\}$ ) 种算法在预测时间段内第  $t$  ( $t \in \{1, 2, \dots, T\}$ ) 点的

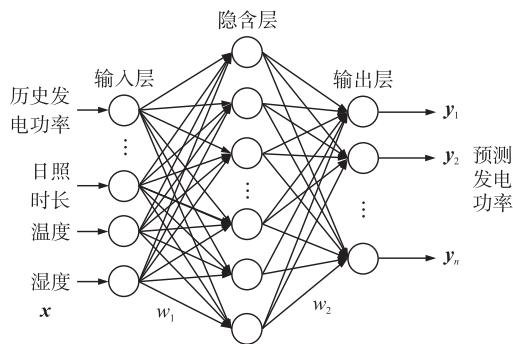


图 1 BP 神经网络结构

Fig. 1 The structure of BP neural network

光伏发电功率为  $P_{mt}$ .

(2) 定义光伏发电功率概率分布函数

设时刻的光伏发电功率服从正态分布,  $g_m(t)$  表示为第  $m$  种算法所得到的时刻的光伏发电功率分布函数, 可描述为

$$g_m(t) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(t-\mu_m)^2}{2\sigma_m^2}}. \quad (14)$$

式中,  $\mu_m$  为平均值,  $\sigma_m^2$  为方差.

(3) 计算光伏发电功率分布函数的数字特征

选取与待预测日相似度较高的个相似日, 则第  $k$  ( $k \in \{1, 2, \dots, K\}$ , 其中  $K$  为相似日样本总数) 个相似日  $t$  时刻点的光伏发电功率值为  $P_{kt}$ . 对于第  $m$  种算法预测所得的光伏发电功率  $P_{mt}$ , 可看作该光伏发电功率概率分布函数的均值  $\mu_m$ . 又可根据所选取的相似日的光伏发电功率计算样本方差

$$\begin{cases} \mu_m = P_{mt}, \\ \sigma_m^2 = \frac{1}{K} \sum_{k=1}^K (P_{kt} - P_{mt})^2. \end{cases} \quad (15)$$

可得到种算法在预测时间段  $T$  内时刻光伏功率的概率密度函数.

设  $g(t)$  为融合各影响因素后的组合光伏发电功率的概率密度分布函数,  $g_m(t)$  表示为第  $m$  种算法所得到的某时刻光伏发电功率分布函数, 满足

$$\begin{cases} \sum_{m=1}^M w_m = 1, \\ g(t) = \sum_{m=1}^M w_m g_m(t). \end{cases} \quad (16)$$

则上式所求的时间段  $T$  内时刻组合光伏发电功率值为  $g(t)$  的数学期望  $\mu$ , 样本方差为

$$\sigma^2 = \frac{1}{K} \sum_{k=1}^K \left( \sum_{m=1}^M w_m (P_{kt} - P_{mt}) \right)^2. \quad (17)$$

(4) 建立支持向量并给出交叉熵目标函数

建立支持向量

$$S = (S_1, S_2, \dots, S_M). \quad (18)$$

由交叉熵定义可设

$$S_m = D[g(t) \parallel g_m(t)] = \int g(t) \ln \frac{g(t)}{g_m(t)} dt, \quad (19)$$

$$\text{s.t. } \int g(t) dt = 1. \quad (20)$$

式中,  $S_m$  用来表示  $g(t)$  与  $g_m(t)$  之间的相互支持程度, 支持程度越高, 其值越小.

记

$$A_m = \frac{S_m}{\sum_{m=1}^M S_m}. \quad (21)$$

为了使权重  $w_m$  更好地反应不同信息源之间的相互支持程度, 设

$$w_m = \frac{1/A_m}{\sum_{m=1}^M 1/A_m} = \frac{1}{1 + \sum_{i, m=1, i \neq m}^M A_m/A_i}, \quad (22)$$

则建立最小交叉熵的目标函数为

$$\begin{cases} F = \min \left( \sum_{m=1}^M D[g(t) \parallel g_m(t)] \right), \\ \text{s.t. } \sum_{m=1}^M w_m = 1, 0 \leq w_m \leq 1, m = 1, 2, \dots, M. \end{cases} \quad (23)$$



上式中  $g(t)$  与  $g_m(t)$  均服从正态分布.

#### (5) 求解最小交叉熵的目标函数

由上述可知,目标函数与权重值  $w_m$  有关,其优化问题实际上是求解的极小值问题. 针对这里的非线性规划问题求解,本文通过 MATLAB 中 Fmincon 函数编程求解,从而获得各时刻下单一方法的权重值,具体求解过程这里不再赘述.

#### (6) 计算各预测时刻的光伏发电组合功率值

根据步骤(5)所求各预测时刻下单一方法的权重值,逐一计算出交叉熵模型的组合预测功率值,最终获得预测时间段  $T$  内所有点的光伏发电功率值.

## 4 算例分析

### 4.1 实验方案设计

为验证本文所提基于相似日和交叉熵理论的光伏发电短期功率组合预测模型的正确性与有效性,以江苏某光伏电站发电系统为例,额定装机容量为 180 kW,该光伏并网系统结构如图 2 所示,部分实验平台图如图 3 所示. 选取 2013 年 1 月 1 日至 12 月 31 日的运行数据作为样本集,气象数据来自当地气象观测站,选择 3 种单一预测方法和 3 种不同的组合预测方法来分别对待预测日光伏发电功率进行预测. 为更好对比各模型的预测精度,3 种单一预测方法的训练样本均采用本文所提的相似日选取方法. 在运行数据样本中代表性地选择 12 月 10 日(晴天)、12 月 12 日(多云)、12 月 15 日(阴天)和 12 月 17 日(小雨)4 种不同的天气类型作为预测日,预测区间为 7:00~17:00,时间间隔为 30 min,各预测模型的学习算法程序均采用 MATLAB R2014a 编程软件实现,最后将预测值与真实值进行对比分析.

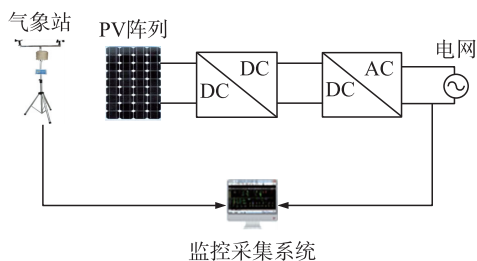


图 2 光伏并网系统结构图

Fig. 2 Structure of grid-connected PV system



图 3 光伏电站部分实验平台

Fig. 3 Part of PV system experimental platform

### 4.2 评价指标选取

为更准确地评价各预测模型的有效性,本文选用平均绝对百分比误差(mean absolute percentage error, MAPE)和希尔不等系数(theil inequality coefficient, TIC)对各预测模型进行对比分析. 其中,TIC 值越小,表明模型预测性能越好,且 MAPE 和 TIC 的计算公式分别为

$$e_{\text{MAPE}} = \left| \frac{y_i - y_{fi}}{y_i} \right|, \quad (24)$$

$$e_{\text{TIC}} = \frac{\sqrt{\sum_{i=1}^n (y_{fi} - y_i)^2 / n}}{\sqrt{\sum_{i=1}^n y_{fi}^2 / n} + \sqrt{\sum_{i=1}^n y_i^2 / n}}. \quad (25)$$

式中,  $y_i$  为实际光伏发电功率,  $y_{fi}$  为预测光伏发电功率,  $n$  为预测点数.

### 4.3 相似日选取结果

采用本文所提的相似日选取方法,以 12 月 12 日(多云)为例,相似日选取结果如表 1 所示. 选取的相似日大多集中在冬季,且在日照时间、平均湿度和平均温度等方面都十分接近. 为验证本文相似日选取方法的合理性,采用文献[16]的方法进行对比分析,选择 LSSVM 为预测模型,预测结果如图 4 所示.

表 1 相似日选取结果

Table 1 Selection results of similar days

相似日	$\chi_i$	排序
3 月 2 日	0.05	1
12 月 1 日	0.08	2
12 月 9 日	0.12	3
12 月 13 日	0.13	4
3 月 11 日	0.25	5
12 月 29 日	0.27	6

由图 4 可以看出,本文所提相似日选取方法的预测值更加接近实际值。

表 2 为不同相似日选取方法的评价指标比较,由表 2 可以看出本文所提相似日选取方法比文献 [16] 中的欧式距离选取法更加有效。

表 2 不同相似日选取方法比较

Table 2 Comparison of different similar days selection methods		
选取方法	评价指标	LSSVM
本文方法	$e_{MAPE}$	10.50%
	$e_{TIC}$	0.034 4
文献[16]方法	$e_{MAPE}$	16.16%
	$e_{TIC}$	0.054 3

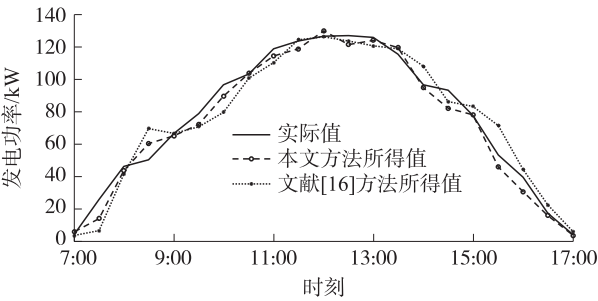


图 4 不同相似日选取方法的预测结果

Fig. 4 Forecasting results of different similar days methods

4.4 单一模型的预测

图 5 是单一预测方法在不同天气类型下的预测结果. 其中,组合模型 1 为采用基于误差平方和最小确定权重的组合预测模型,组合模型 2 为采用基于相关系数最大确定权重的组合预测模型,组合模型 3 则为本文所提基于相似日和交叉熵理论的组合预测模型. 由图 5 可知,3 种单一方法均能较好地预测光伏发电功率,但预测精度仍可进一步提高. 图 5(a)和(b)分别为 12 月 10 日(晴天)和 12 月 12 日(多云)的单一方法预测结果,由于全年运行样本数据中高相似度的训练样本较多,预测所得结果较为理想. 图 5(c)和(d)分别为 12 月 15 日(阴天)和 12 月 17 日(小雨)的预测结果,由于阴雨天气下相似度较高预测训练样本过少且天气变化情况较为复杂,预测结果并不理想。

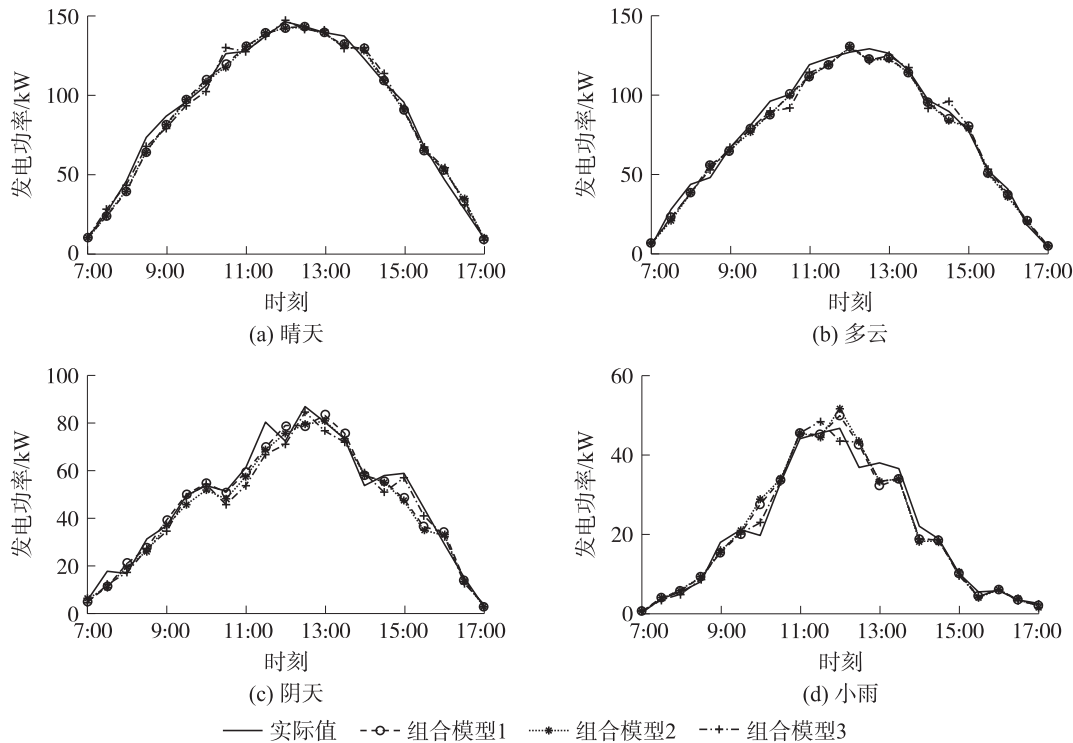


图 5 单一模型的预测结果

Fig. 5 Forecasting results of single models

表 3 给出了不同预测日下单一预测方法的误差统计. 由表 3 可以看出,在 3 种单一预测方法中,LSSVM 表现最佳,晴天的 MAPE 值可达 8.77%,TIC 值也比其余两种方法更小,有着较高的预测精度. 在阴天和小雨的天气下,由于光伏发电功率的波动性开始增大,3 种单一方法预测误差也越来越大,尤其是在小雨天气下,ARMA 模型的平均绝对误差达到 17.78%. 相比于晴天而言,预测精度大大下降. 总体而言,由于 BP 神经网络法的输入变量考虑了日照时长和温湿度,所以预测误差低于 ARMA,但是与 LSSVM 相比,其预测性能还不算理想。

表 3 单一模型预测误差统计

Table 3 Statistics of forecasting errors for single models

天气类型	评价指标	LSSVM	ARMA	BP	天气类型	评价指标	LSSVM	ARMA	BP
晴天	$e_{MAPE}$	8.77%	10.99%	10.59%	阴天	$e_{MAPE}$	13.09%	17.97%	17.28%
	$e_{TIC}$	0.029 4	0.037 5	0.030 8		$e_{TIC}$	0.055 3	0.075 6	0.071 0
多云	$e_{MAPE}$	10.50%	14.20%	13.91%	小雨	$e_{MAPE}$	13.99%	17.78%	17.29%
	$e_{TIC}$	0.034 4	0.046 5	0.035 5		$e_{TIC}$	0.057 8	0.096 6	0.062 9

4.5 组合模型的预测

在获得单一方法预测结果的基础上,分别采用基于误差平方和最小确定权重的组合预测模型和基于相关系数最大确定权重的组合预测模型<sup>[17]</sup>与本文所提基于交叉熵理论的组合预测模型进行对比分析.

表 4 给出了不同预测日下交叉熵组合预测模型的权重值, $w_1$ 、 $w_2$  和  $w_3$  分别为 LSSVM、ARMA 和 BP 法的权重值(考虑篇幅限制,这里给出了部分时刻下单一方法的权重值). 由表 4 可以看出,3 种方法在不同时刻的权重值均不相同. 在晴天 13:00,3 种方法的预测误差均较大,此时 ARMA 和 BP 方法与其他单一方法的相互支持程度较小,求解的权重值均近似为 0,避免了单一方法对整体预测精度的影响. 在阴天 17:00 和小雨 7:30 时,3 种方法预测误差近似,需提高与其他方法相互支持程度较大的 ARMA 的权重,求解的权重值分别为 0.505 和 0.636. 由此可见,采用交叉熵理论动态求得的权重值可以客观地反映不同预测方法间的信息融合.

表 4 交叉熵模型权重值

Table 4 Weights of cross entropy model

天气类型	时刻	$w_1$	$w_2$	$w_3$	天气类型	时刻	$w_1$	$w_2$	$w_3$
晴天	10:00	0.076	0.855	0.069	阴天	9:00	0.277	0.717	0.006
	13:00	0.996	0.004	0.000		15:00	0.000	0.001	0.999
	16:00	0.006	0.913	0.081		17:00	0.385	0.505	0.110
多云	8:00	0.532	0.002	0.466	小雨	7:30	0.273	0.636	0.091
	12:00	0.239	0.413	0.348		11:00	0.167	0.523	0.310
	15:30	0.074	0.633	0.293		13:30	0.643	0.001	0.356

图 6 为不同预测日下 3 种组合预测方法的光伏发电功率预测结果. 与图 5 相比,可以明显看出组合方法所得预测发电功率值更加接近实际的发电功率值.

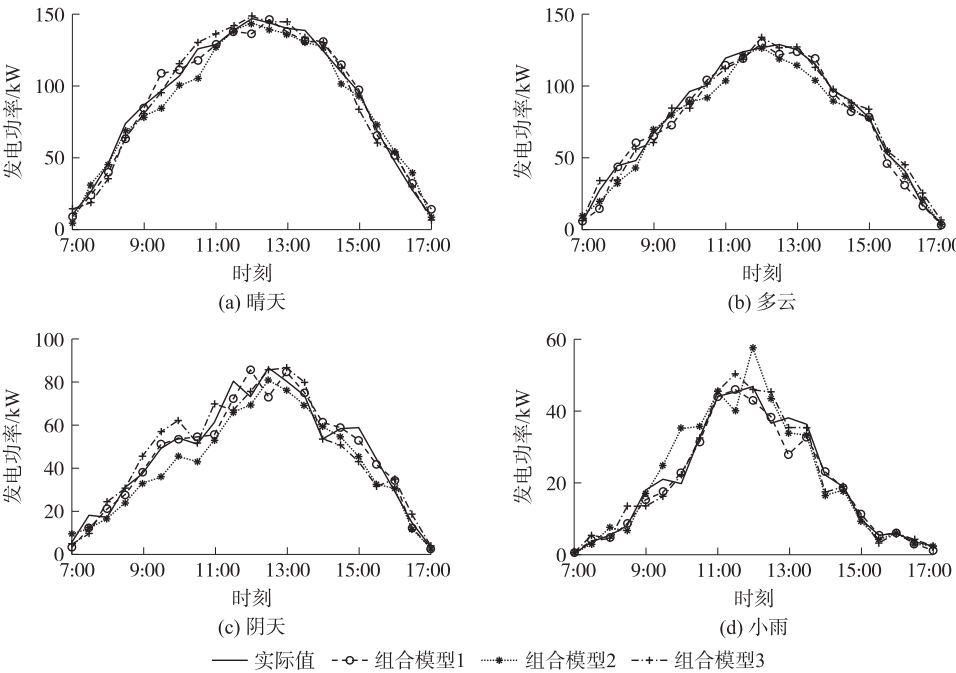


图 6 组合模型的预测结果

Fig. 6 Forecasting results of combination models



表 5 给出了不同预测日下组合预测方法的误差统计. 由表 5 见,组合预测方法在 4 种天气类型下的预测精度较单一方法更高. 晴天时,组合模型 3 的 MAPE 值比单一方法中最佳的 LSSVM 提高了 4.34%,且高于组合模型 1 和组合模型 2. 由评价指标可以看出,在不同预测日下组合模型 2 和组合模型 3 的预测误差相差无几,预测性能相当. 此外,在不同预测日下,组合模型 3 的 TIC 指标值在 3 种预测方法中均最低,充分表明了本文所提基于相似日和交叉熵理论的组合预测模型的优越性.

表 5 组合模型预测误差统计

Table 5 Statistics of forecasting errors for combination models

天气类型	评价指标	组合模型 1	组合模型 2	组合模型 3	天气类型	评价指标	组合模型 1	组合模型 2	组合模型 3
晴天	$e_{MAPE}$	5.35%	5.43%	4.43%	阴天	$e_{MAPE}$	10.47%	10.52%	9.62%
	$e_{TIC}$	0.021 8	0.021 9	0.019 9		$e_{TIC}$	0.049 4	0.051 7	0.046 1
多云	$e_{MAPE}$	8.24%	8.37%	7.32%	小雨	$e_{MAPE}$	11.35%	11.37%	10.20%
	$e_{TIC}$	0.026 7	0.027 7	0.024 9		$e_{TIC}$	0.054 9	0.060 8	0.046 7

5 结语

本文针对光伏发电功率的间歇性与波动性,提出了一种基于相似日和交叉熵理论的短期光伏发电功率组合预测模型,得到以下结论:

(1)LSSVM、ARMA 和 BP 神经网络法可对光伏发电短期功率进行预测,但单一预测方法在不同预测日下的误差均不相同. 本文算例中,LSSVM 预测精度高于 BP 神经网络法,BP 神经网络法预测精度高于 ARMA.

(2)组合预测模型可以有效地降低不同预测日下的预测误差,基于误差平方和的组合模型和基于相关系数的组合模型在预测误差上较为接近,但交叉熵组合模型预测误差更低,预测性能更佳.

(3)交叉熵组合预测方法是在每一个预测时刻,通过最小化组合预测方法与各单一预测方法之间的相互支持度来动态设置权重,在光伏发电随机性与波动性较大时也能保持较高的预测精度,能适应不同预测日下的光伏发电短期功率预测,具有一定实用价值.

[ 参考文献 ] (References)

[ 1 ] 丁明,王伟胜,王秀丽,等. 大规模光伏发电对电力系统影响综述[J]. 中国电机工程学报,2014,34(1):1-14.  
DING M,WANG W S,WANG X L,et al. A review on the effect of large-scale PV generation on power systems[J]. Proceedings of the CSEE,2014,34(1):1-14.(in Chinese)

[ 2 ] 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.  
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,2016,40(4):140-151.(in Chinese)

[ 3 ] 许童羽,马艺铭,曹英丽,等. 基于主成分分析和遗传优化 BP 神经网络的光伏输出功率短期预测[J]. 电力系统保护与控制,2016,44(22):90-95.  
XU T Y,MA Y M,CAO Y L,et al. Short term forecasting of photovoltaic output power based on principal component analysis and genetic optimization of BP neural network[J]. Power system protection and control,2016,44(22):90-95.(in Chinese)

[ 4 ] TENG X L,GAO Z H,ZHANG Y Y,et al. Key technologies and the implementation of wind,PV and storage co-generation monitoring system[J]. Journal of modern power systems & clean energy,2014,2(2):104-113.

[ 5 ] 高相铭,杨世凤,潘三博. 基于 EMD 和 ABC-SVM 的光伏并网系统输出功率预测研究[J]. 电力系统保护与控制,2015,43(21):86-92.  
GAO X M,YANG S F,PAN S B. A forecasting model for output power of grid-connected photovoltaic generation system based on EMD and ABC-SVM[J]. Power system protection and control,2015,43(21):86-92.(in Chinese)

[ 6 ] LORENZ E,HURKA J,HEINEMANN D,et al. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems[J]. IEEE journal of selected topics in applied earth observations and remote sensing,2009,2(1):2-10.

[ 7 ] 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. (in Chinese)
- [8] 兰华, 廖志民, 赵阳. 基于 ARMA 模型的光伏电站出力预测[J]. 电测与仪表, 2011, 48(2): 31-35.  
LAN H, LIAO Z M, ZHAO Y. ARMA model of the solar power station based on output prediction[J]. Electrical measurement & instrumentation, 2011, 48(2): 31-35. (in Chinese)
- [9] 叶林, 陈政, 赵永宁, 等. 基于遗传算法—模糊径向基神经网络的光伏发电功率预测模型[J]. 电力系统自动化, 2015, 39(16): 16-22.  
YE L, CHEN Z, ZHAO Y N, et al. Photovoltaic power forecasting model based on genetic algorithm and fuzzy radial basis function neural network[J]. Automation of electric power systems, 2015, 39(16): 16-22. (in Chinese)
- [10] 朱永强, 田军. 最小二乘支持向量机在光伏功率预测中的应用[J]. 电网技术, 2011, 35(7): 54-59.  
ZHU Y Q, TIAN J. Application of least square support vector machine in photovoltaic power forecasting[J]. Power system technology, 2011, 35(7): 54-59. (in Chinese)
- [11] 王新普, 周想凌, 邢杰, 等. 一种基于改进灰色 BP 神经网络组合的光伏出力预测方法[J]. 电力系统保护与控制, 2016, 44(18): 81-87.  
WANG X P, ZHOU X L, XING J, et al. A prediction method of PV output power based on the combination of improved grey back propagation neural network[J]. Power system protection and control, 2016, 44(18): 81-87. (in Chinese)
- [12] 杨锡运, 刘欢, 张彬, 等. 基于熵权法的光伏输出功率组合预测模型[J]. 太阳能学报, 2014, 35(5): 744-749.  
YANG X Y, LIU H, ZHANG B, et al. A combination method for photovoltaic power forecasting based on entropy weight method[J]. Acta energiae solaris sinica, 2014, 35(5): 744-749. (in Chinese)
- [13] 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): 66-71.  
CHEN T, SUN G Q, WEI Z N, et al. Photovoltaic power generation forecasting based on similar day and CAPSO-SNN[J]. Electric power automation equipment, 2017, 37(3): 66-71. (in Chinese)
- [14] 杨锡运, 刘玉奇, 张璜, 等. 含并网光伏电站的系统可靠性评估方法[J]. 高电压技术, 2016, 42(9): 2689-2696.  
YANG X Y, LIU Y Q, ZHANG H, et al. Reliability evaluation method of grid connected with photovoltaic power station[J]. High voltage engineering, 2016, 42(9): 2689-2696. (in Chinese)
- [15] 陈宁, 沙倩, 汤奕, 等. 基于交叉熵理论的风电功率组合预测方法[J]. 中国电机工程学报, 2012, 32(4): 29-34.  
CHEN N, SHA Q, TANG Y, et al. A combination method for wind power prediction based on cross entropy theory[J]. Proceedings of the CSEE, 2012, 32(4): 29-34. (in Chinese)
- [16] 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): 153-157.  
LI J H, CHEN G P, GE P J, et al. Output power forecasting of PV generation system based on similar day theory[J]. East China electric power, 2012, 40(1): 153-157. (in Chinese)
- [17] 陈华友. 基于相关系数的优性组合预测模型研究[J]. 系统工程学报, 2006, 21(4): 353-360.  
CHEN H Y. Research on properties of superior combined forecasting models based on correlation coefficients[J]. Journal of systems engineering, 2006, 21(4): 353-360. (in Chinese)

[责任编辑: 陈 庆]