

基于 PSO-DE-BP 的光伏发电功率短期预测

刘春芳¹, 王攀攀², 曹菲¹

(1. 南瑞集团(国网电力科学研究院)有限公司, 南京 211106;

2. 江苏林洋新能源科技有限公司, 南京 210019)

摘要: 提高光伏发电功率预测精度对保障智能电网安全稳定运行有重要意义; 针对传统 BP 神经网络存在预测精度不高且收敛速度慢的弊端, 提出一种基于粒子群(PSO)差分进化(DE)并行计算优化 BP 神经网络的光伏发电短期预测方法; 首先分析影响因素重要程度, 通过带权重的欧式距离筛选相似的训练样本集; 其次, 对粒子群分组, 通过粒子群和差分进化混合算法对粒子组内和组间优化, 以保证种群多样性、提高预测稳定性和精度、避免局部最优; 然后, 建立预测模型, 通过基于 spark 的内存计算平台, 将 PSO-DE-BP 算法并行优化以提高算法运行效率; 最后, 根据不同天气类型的预测结果对模型进行分析验证, 此方法比 PSO-BP、BP 算法模型具有更高的稳定性和预测精度。

关键词: 光伏发电预测; BP 神经网络; 差分进化; 粒子群分组; Spark 并行计算

Short Term Prediction of Photovoltaic Power Generation Based on PSO-DE-BP

LIU Chunfang¹, WANG Panpan², CAO Fei¹

(1. NARI Group Limited Company (State Grid Electric Power Research Institute Company), Nanjing 211106, China;

2. Jiangsu Linyang Renewable Energy Technology Co., Ltd., Nanjing 210019, China)

Abstract: Improving photovoltaic power prediction accuracy is of great significance to ensure the safe and stable operation of smart power grid. Aimed at the disadvantages of low prediction accuracy and slow convergence speed of traditional BP neural network, a short-term photovoltaic power prediction method based on particle swarm optimization (PSO) differential evolution (DE) parallel computing optimization BP neural network is proposed. Firstly, the influencing factor is analyzed, and the selects similar training sample sets are selected by the weighted Euclidean distance. Secondly, the particle swarm is grouped by the hybrid algorithm based on particle swarm and differential evolution to optimize within and between particle groups, so as to ensure the PSO diversity, improve the prediction accuracy, and avoid the local optimization. Then, the prediction model is established, and the PSO-DE-BP algorithm is parallelized through the Spark-based memory computing platform. Finally, the model is analyzed and verified by the prediction results of different types of weather. This method has higher stability and prediction accuracy than PSO-BP and BP algorithms.

Keywords: prediction for the photovoltaic power generation; BP neural network; differential evolution; particle group; spark parallel computing

0 引言

光伏发电技术是当前利用太阳能的主要技术。然而, 受天气情况、太阳辐射强度、温度、湿度、云量等影响因素, 光伏发电技术有着间歇性、波动性、随机性特点^[1], 对智能电网的稳定运行带来极大挑战^[2]。因此, 需要对光伏发电功率预测进行研究, 预先得知日内光伏发电功率曲线, 使电力部门能以经济最优化调度机组^[3], 获取最大经济效益。

光伏发电预测技术一直是国内外专家学者研究的热门课题。目前光伏发电预测方式主要分为直接预测和间接预测。间接预测方式是先对地表辐射强度进行预测, 然后根

据光伏发电模型得到光伏发电功率^[4]。文献 [5] 采用了间接预测, 根据历史的光照强度、温度等因素, 利用支持向量机的方法来预测光照强度及温度, 再通过数学模型得出光伏发电功率, 但是未对天气类型分类, 精度不高。间接预测过度依赖于复杂的太阳辐射强度模型, 对含有不同类型光伏发电单元的系统来说, 转换效率、安装角度等参数存在差异。直接预测即直接对光伏电站的输出功率进行预测, 常用方法有支持向量机(SVM)^[6]、时间序列法^[7]、小波分析预测法^[8-9]以及人工神经网络法^[10-11]。文献 [10-11] 综合考虑了影响因素, 基于神经网络对光伏发电功率提前一天进行预测, 误差较大, 容易陷入局部最优。文献 [12] 基于差分进化优化 BP 神经网络对光伏发电功率预测, 有较

收稿日期: 2022-08-11; 修回日期: 2022-10-19。

基金项目: 国家电网公司总部科技项目资助(5100-202113396A)。

作者简介: 刘春芳(1990-), 女, 江苏丰县人, 硕士, 主要从事分布式发电、电能质量、电力计量方向的研究。

引用格式: 刘春芳, 王攀攀, 曹菲. 基于 PSO-DE-BP 的光伏发电功率短期预测[J]. 计算机测量与控制, 2023, 31(5): 180-186.

好的实用性和可行性, 但 BP 神经网络算法具有参数随机初始化、预测系统不稳定以及收敛速度慢等特点, 因此会出现早熟现象。

综合以上分析, 文章分析光伏发电的影响因素, 通过带权重的欧式距离公式筛选出相似样本集; 设计 PSO-DE-BP 混合算法对神经网络进行训练; 基于 SPARK 的平台对 PSO-DE-BP 并行计算; 通过实测数据仿真验证表明, PSO-DE-BP 算法比 PSO-BP、BP 算法精度更好且更稳定, 同时, 该方法具有良好的并行性能。

1 相似样本集筛选

1.1 影响因子分析

单位面积的光伏发电系统输出功率的数学模型^[13]

$$P_{pv_out} = \eta SI [1 - 0.005 (t_0 + 25)] \quad (1)$$

其中: η 是光伏阵列转换效率; S 是光伏阵列的面积, m^2 ; I 太阳辐射强度, kW/m^2 ; t_0 环境温度, $^{\circ}C$ 。

在固定的光伏发电系统中, 光伏阵列的面积是不变的, 在一个短期的输出功率预测模型中, 光伏阵列转换效率也基本上不变, 根据公式 (1) 的功率输出模型可知, 光伏发电功率主要的影响因素就是温度和太阳的辐射强度, 根据目前气象局提供的天气信息, 尽管缺乏太阳辐射强度的信息, 但是气象局提供的天气、云量、湿度以及温度能反应光照辐射强度^[14]。

图 1 为同一光伏电站不同天气类型下光伏发电功率对比曲线。天气类型对光伏发电系统输出功率影响非常明显, 且雨天和晴天输出功率值差距很大, 表明太阳辐射强度对光伏发电功率影响巨大。

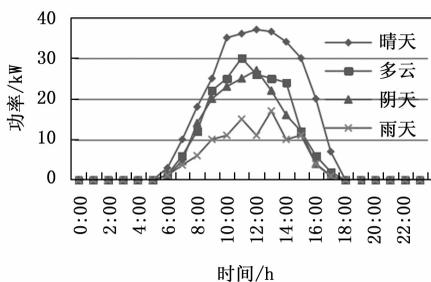


图 1 不同天气类型光伏功率输出曲线

表 1 选取 2021 年中江苏省某个光伏电站相同天气类型, 不同温度的历史数据, 分析温度与日光伏发电量的关系。

表 1 温度气象信息/ $^{\circ}C$

| 日期 | 1-2 | 3-22 | 5-1 | 7-31 | 9-8 | 11-2 |
|------|-----|------|-----|------|-----|------|
| 天气类型 | 晴 | 晴 | 晴 | 晴 | 晴 | 晴 |
| 最高气温 | 7 | 16 | 26 | 35 | 29 | 19 |
| 最低气温 | 0 | 4 | 14 | 25 | 23 | 12 |

图 2 表明, 温度与光伏发电先呈现正相关, 随着温度的升高, 光伏发电量增高, 因此, 光伏发电功率呈现出随

季节的周期性变化。

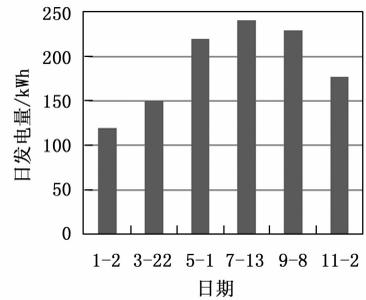


图 2 不同的温度下光伏日发电量

表 2 选择比较相邻几天且来自于同一个光伏电站的历史数据, 天气类型相同, 但相对湿度不同。

表 2 湿度气象信息

| 日期 | 天气类型 | 相对湿度/% |
|-----------|------|--------|
| 2021.9.6 | 多云 | 32.7 |
| 2021.9.10 | 多云 | 54.5 |
| 2021.9.14 | 多云 | 78.3 |

根据图 3 可知空气湿度明显影响着光伏发电系统的输出功率, 湿度越大, 光伏输出功率越小, 呈现负相关状态, 这是由于空气中的液态水能大量吸收太阳辐射能。

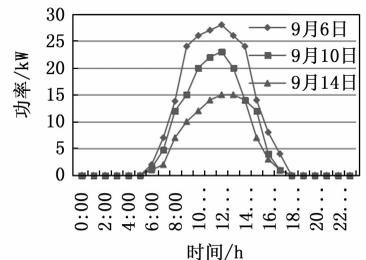


图 3 不同湿度下光伏发电功率变化曲线

表 3 中选取了 2021 年 8 月份的气象数据, 样本具有相近的温度以及不同的云量。结合图 4 表明, 云量和湿度相似, 与光伏发电呈现负相关。由于云将太阳辐射进行反射及散射, 大量削弱了太阳辐射强度。因此, 云量对预测每个小时的发电功率预测有着重要意义。

表 3 云量气象信息

| 日期 | 8-1 | 8-2 | 8-14 | 8-17 | 8-19 | 8-23 |
|----|-----|-----|------|------|------|------|
| 天气 | 晴 | 晴 | 多云 | 多云 | 阴 | 雨 |
| 云量 | 0.2 | 0.1 | 0.45 | 0.55 | 0.7 | 0.95 |

上述分析可知, 光伏发电系统与天气类型、温度、湿度有着密切的关系, 但每种因素与光伏发电功率关联度有所不同, 因此需要进一步分析影响因子的重要度。

1.2 相似样本选择算法

根据 1.1 可知, 影响因子的差距越大, 光伏发电系统的输出功率大小值和变化趋势差距越大。图 1 可知, 在晴天时的输出功率基本呈现上升、保持、下降的变化趋势。

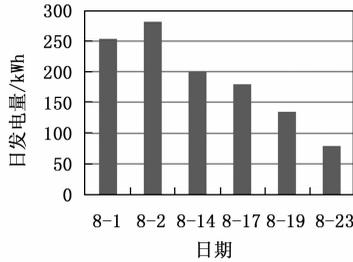


图 4 云量与光伏发电量

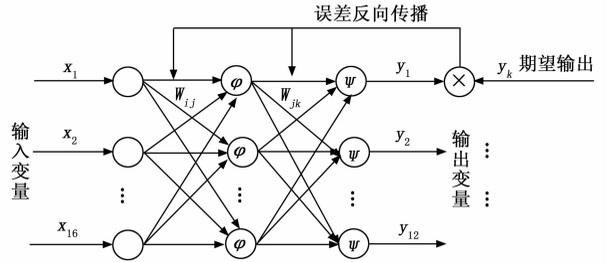


图 5 神经网络模型

然而在阴天、雨天，这种变化过程被破坏^[10]，输出功率曲线不稳定。文献 [15-16] 没有筛选相似的训练样本，在阴天或雨天时，功率预测的误差较大。文献 [10] 基于相似样本，用相似样本集进行训练和预测，减少一定的预测误差，但是只考虑了温度，缺少天气情况作为相似参考量。

然而，影响因子对光伏发电功率的影响程度不同。为了从光伏发电系统历史记录中选择出相似样本集，有效减少预测范围，提高预测精度，通过公式 (2) 得出各个因子的相关系数^[17]。

$$r_j = \frac{\sum_{i=1}^{NP} (x_i(j) - \bar{x}(j)) (p_i - \bar{p})}{\sqrt{\sum_{i=1}^{NP} (x_i(j) - \bar{x}(j))^2} \sqrt{\sum_{i=1}^{NP} (p_i - \bar{p})^2}} \quad (2)$$

其中： r_j 表示 j 影响因子相关系数； NP 任意样本集数目； $\bar{x}(j)$ 表示 j 影响因子的在样本集平均值； \bar{p} 对应日平均光伏发电功率； $x_i(j)$ ， p_i 分别对应 i 样本的影响因子和功率。

其中， w_i 表示 i 影响因子的权重。

$$w = \frac{|r_i|}{\sum_{k=1}^n |r_k|} \quad i=1, 2, \dots, n \quad (3)$$

若待预测日的气象特征向量记作 x_0 ， $x_0 = [x_{01}, x_{02}, x_{03}, x_{04}]$ ，第 j 个历史日的气象特征向量为 x_j ， $x_j = [x_{j1}, x_{j2}, x_{j3}, x_{j4}]$ ，其中 x_{01} ， x_{02} ， x_{03} ， x_{04} 分别代表待预测日的归一化后最高温度，最低温度，最高湿度，最低湿度，同理 x_j 中元素也代表 4 个影响因素。

选择相似样本集步骤如下：

- 1) 选择出和预测日相同天气情况 m 个历史样本，记作 Y 样本集。
- 2) 由 Y 样本集用计算公式 (2)，(3) 得出各个影响因子的权重值。
- 3) 计算预测日和样本集 Y 中历史记录的影响因素的欧式距离 d_i ， d_i 的计算公式如下：

$$d_i = \left[\sum_{k=1}^n w_k * (x_{0k} - x_{jk})^2 \right]^{1/2} \quad (4)$$

4) 将影响因素欧式距离集 $[d_1, d_2, \dots, d_m]$ 按距离值的大小升序排序，排列前 N 个历史样本形成训练样本集 Y_N 。

2 BP 神经网络预测模型设计

2.1 BP 神经网络

BP 神经网络是目前使用广泛的神经网络算法之一。如图 5 所示，典型的神经网络一般有输入层、隐含层、输出层。

2.2 预测模型建立

2.2.1 BP 神经网络输入层确定

光伏发电系统每天主要出力时间段为 6:00~18:00。由 1 节可知，天气类型严重影响光伏发电，即使选择了相同天气的样本集，比如，有着相同的晴转多云的天气类型，但是 12 个时间点有着不同天气类型。由表 4 可知，云量能体现各个时间段的天气类型^[18]，因此，以预测日 6:00~18:00 的每小时云量、最高温度、最低温度、最高湿度、最低湿度作为输入变量。

表 4 云量与天气类型关系表

| 天气类型 | 云量 |
|------|---------|
| 晴天 | 0~25% |
| 多云 | 25%~60% |
| 阴 | 60%~90% |
| 雨 | 90%以上 |

2.2.2 输出层节点数

输出层节点为预测日 6:00~18:00 的每个小时共 12 个时间段的光伏发电功率。

2.2.3 隐含层节点设计

隐含层节点过多、过少或训练时间过长以及精度不高都对网络的训练不利。充分考虑时间和精度，根据经验公式 (5) 进行设计^[19]：

$$J = \sqrt{I+O} + a \quad (5)$$

式中， I 是输入层节点数； O 为输出层节点数； J 是隐含层节点数； a 为 1~10 的常数。

通过反复调试神经网络模型，充分考虑时间和精度，隐含层选择为 10 个。

表 5 BP 神经网络输入输出变量

| 变量 | 变量名称 |
|----------------|-------------------------------|
| $X_1 - X_4$ | 预测日的最高气温,最低气温,最高湿度,最低湿度。 |
| $X_5 - X_{16}$ | 预测日的 6:00~18:00 每小时的云量 |
| $Y_1 - Y_{12}$ | 预测日的 6:00~18:00 每小时的平均光伏发电功率。 |

2.3 训练样本预处理

直接用原始的数据作为网络的输入与输出，会引起神经元过饱和。因此，神经网络中输入节点和输出节点的数据都必须归一化处理。通常 BP 神经网络使用 Sigmoid 函

数, 输出范围限定在 $[0, 1]$ 。用公式 (6) 对 1.2 筛选出待预测日训练样本集 Y_N 做归一化处理, 所有数据包括功率数据, 影响因素数据都转换到 $[0, 1]$ 区间。

$$X_j^* = (x_j - x_{j,min}) / (x_{j,max} - x_{j,min}) \quad j=1, 2 \cdots 21 \quad (6)$$

其中: X_j^* 为归一化后的数据; x_j 为原始数据; $x_{j,min}$, $x_{j,max}$ 某个时间段功率或影响因素的最小值和最大值; j 代表影响因子或功率。

输出样本反归一化:

$$y_k = (y_k^* * (y_{k,max} - y_{k,min})) + y_{k,min} \quad k=1, 2 \cdots 12 \quad (7)$$

式中, y_k 是实际输出功率; y_k^* 神经网络输出归一化后的值; $y_{k,max}$ 、 $y_{k,min}$ 是第 K 时间段功率最大值和最小值。

3 PSO-DE-BP 混合算法优化神经网络

3.1 基本的粒子群算法和差分进化算法

1) 粒子群算法^[20]通过粒子的相互作用, 对空间任何区域进行智能搜索。主要思想通过群体之间相互协作和信息带有记忆功能及共享信息, 寻找最优解。粒子的位置向量代表空间的一组解向量, 初始化时, 每个粒子都有随机的飞行方向, 以一定的速度寻找最优解。由于粒子共享信息功能, 粒子的位置、飞行速度大小和方向都随经验动态调整。

粒子速度更新公式:

$$v_{i,D}^{g+1} = \omega * v_{i,D}^g + c_1 * rand(1) * (p_{best}^G - x_{i,D}^G) + c_2 * rand(1) * (g_{best}^G - x_{i,D}^G) \quad (8)$$

粒子位置更新公式:

$$x_{i,D}^{g+1} = x_{i,D}^G + v_{i,D}^{g+1} \quad (9)$$

其中: G 为迭代次数; D 是空间维数, 对应优化参数的个数; $v_{i,D}^G$ 第 G 次 i 粒子的速度; p_{best}^G 是个体最优值; g_{best}^G 是全局最优值; ω 是惯性权重, c_1 , c_2 是学习因子; $x_{i,D}^{G+1}$ 第 $G+1$ 次的位置。

2) 差分进化算法是基于群体智能的随机并行优化算法^[21], DE 拥有记忆功能使其可以动态跟踪当前的搜索情况, 以调整策略, 寻找最优解。它的优化机制是根据不同个体之间的距离和方向信息来生成新的候选个体, 实现群体进化。主要采用变异、交叉、选择 3 种方式, 以这种顺序动态改变个体的位置, 寻找最优个体。

变异:

$$V_{i,D,G+1} = X_{h,D,G} + F \times (X_{r,D,G} - X_{k,D,G}) \quad (10)$$

其中: $X_{h,D,G}$, $X_{r,D,G}$, $X_{k,D,G}$ 是随机选择的个体, 且 $i \neq h \neq r \neq k$; $V_{i,G+1}$ 是第 i 个个体的新个体; F 是变异算子, 属于 $[0, 2]$ 的一个常数; D 是维数, 优化参数个数; G 表示代数。

交叉:

$$U_{i,j,G+1} = \begin{cases} V_{i,D,G+1} & R \leq CR \text{ or } J = Ir \\ X_{i,D,G} & R > CR \text{ and } j \neq Ir \end{cases} \quad (11)$$

式中, $V_{i,D,G+1}$ 变异后的 i 个体; $X_{i,D,G}$ 原个体; CR 是交叉率; Ir 是 $[0, D]$ 的随机整数。

选择: 贪婪策略选择后代。

$$X_{i,j,G+1} = \begin{cases} U_{i,D,G+1} & f(U_{i,D,G+1}) \leq f(X_{i,D,G}) \\ X_{i,D,G} & f(U_{i,D,G+1}) > f(X_{i,D,G}) \end{cases} \quad (12)$$

式中, f 是适应度函数。

3.2 PSO-DE-BP 混合算法优化神经网络

根据 2 节设计的神经网络模型, 用基于 PSO-DE-BP 的算法优化神经网络训练。BP 算法本质是梯度下降法, 由前面可知, 输入输出变量较多, 目标函数比较复杂, 容易陷入局部最优, 不稳定, 且波动性较大。本章节采用 PSO-DE-BP 混合算法优化 RBF 神经网络, 通过粒子分群组, 引入公式 (13) 保证群组内多样性, 运用改进 DE 的变异操作及选择操作, 使组与组之间保持关联, 寻找全局最优, 避免局部最优。

具体步骤如下。

步骤 1: 随机初始化粒子 $x_{i,D}$, 把神经网络需要优化的权值和阈值使其对应, 进行编码, 确定目标适应度值 E , 适应度函数 f_1 , 群组之间变异次数 G , 随机初始化粒子 v 。

步骤 2: 将粒子随机平均分成 m 群, 对群体进行编号 ID;

步骤 3: 每组个体根据粒子群迭代的算法, 寻找群内最优, 达到函数 f_2 目标值, 返回各组的群组最优值 $X_{best,D,g}$ (ID), 根据 $X_{best,D,g}$ (ID) 适应度的值大小, $m/2$ 组为较好群体, 其他的 $m/2$ 为较差群体。

粒子群迭代的目标函数:

$$f_2 = \frac{1}{|N| |R|} \sum_{i=1}^N \sqrt{\sum_{j=1}^D (x_{i,j} - \bar{x}_j)^2} \quad (13)$$

式中, $|R|$ 是搜索半径; $|N|$ 种每个群中粒子个数; $x_{i,j}$ 是第 i 个粒子第 j 维分量; \bar{x}_j 为每个种群中所有粒子的个体最优值在 j 维的均值。

f_2 函数的目标值的设定:

$$F = \lambda_0 e^{(\lambda_{max} - \lambda_{min}) * (g/G)} \quad (14)$$

其中: g 是当前变异次数; λ_0 是常数。

F 的值随着 g 次数增大, 自适应地改变, 粒子群开始, F 较大, 注重粒子群组之间交叉变异, 在保重粒子多样性同时, 向较好群体的方向进化, 避免局部最优。

步骤 4: 较好群体之间变异, 较差群体往较好群体进化, 用 DE 和 PSO 混合算法对粒子变异。

改进的变异操作:

$$v_{i,D,g-1} (ID) = W * v_{i,D,g} (ID) + F_1 \times (rand(X_{best}) - X_{i,D,G} (ID)) + F_2 \times (X_{best,D,g} (ID) - X_{i,D,G} (ID)) \quad (15)$$

$$V_{i,D,g+1} (ID) = X_{i,D,G} + v_{i,D,g+1} (ID) \quad (16)$$

$$F_1 = F_{max} = \frac{f_{best}}{f(ID)} * (F_{max} - F_{min}) \quad (17)$$

式中, ID 是群组号, 范围 $[1, m]$; i 代表 ID 群 i 个体; $rand(X_{best})$ 表示在较好的 $m/2$ 个群组当中某个组最优值, 随机选取, 共 $m/2$ 个; $X_{best,D,g} (ID)$ 当前群组最优个体;

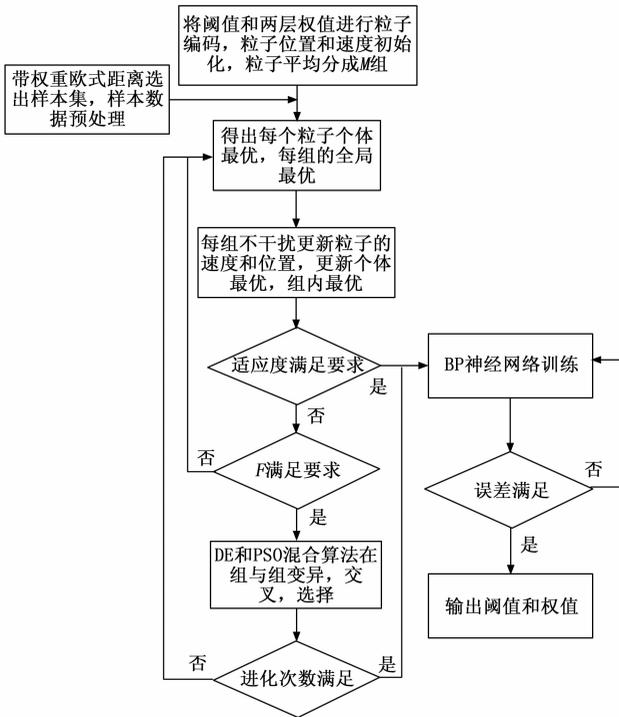


图 6 PSO-DE-BP 算法优化神经网络流程图

$f(ID)$ 当前群组最优适应度; f_{best} 最优群组的适应度; F_2 是 0 到 1 的随机数。

步骤 5: 每个群体中个体用公式 (11) 进行交叉, 然后在改进的选择策略上选择下一代个体, 避免贪婪策略出现早熟现象。

改进的选择操作:

$$X_{i,j,G+1} \begin{cases} u_{i,D,G+1} f(U_{i,D,G+1}) \leq f(X_{i,D,G}) \text{ or } R < \lambda \\ X_{i,D,G} f(U_{i,D,G+1}) > f(X_{i,D,G}) \end{cases} \quad (18)$$

$$\lambda = \min \left(\frac{f(X_{i,D,G})}{f(U_{i,D,G+1})}, 1 \right) \quad (19)$$

式中, f 是适应度函数, R 是 $[0, 1]$ 上的随机数。

步骤 6: 判断是否达到进化次数或适应度目标值, 是, 停止进化, 进入 BP 神经算法优化, 否则转到步骤 3。

适应度函数用训练样本的平均均方误差, 如公式 (20) 所示:

$$f_1 = \frac{1}{2 * NP} \sum_{n=1}^{NP} \sum_{i=1}^{i=12} (A_{n,i} - P_{n,i})^2 \quad (20)$$

式中, A 是实际值, P 是预测值, 由两层的权值和阈值以及 sigmoid 函数得出。

4 Spark 平台下 PSO-DE-BP 算法并行优化

随着光伏发电历史数据量快速增长, 尽管大量的数据能保证光伏发电功率预测的精度, 但是基于单节点难以处理大数据的需要^[22-23]。本章节基于 Spark 平台下, 研究 PSO-DE-BP 算法并行化。

在分布式内存计算框架 Spark 上进行运算, RDD 的数据集格式表示为^[24]:

$$(ZID, ID, x, v, F(p_h, f_h), (p_b, f_b), pf_{m/2,D+1})$$

其中: ZID 是群编号; ID 是群组编号; x, v 分别表示粒子当前位置向量和速度向量, F 为当前目标函数 f_2 的值, p_h, f_h 分别表示为个体经历的最佳位置和最佳适应度; p_b, f_b 分别表示子群组中粒子经历的最佳位置和群组最佳适应度, $pf_{m/2,j}$ 是存放较好 $m/2$ 群组的最佳位置和最佳适应度的 $m/2 \times (D+1)$ 矩阵, 表示需要优化参数个数和一个适应度。

PSO-DE-BP 算法并行化的具体步骤如下:

1) 初始化粒子群, 包括初始位置和初始速度, 差分进化次数, 随机对粒子群平均分成 M 组, 并且对群组进行 ID 编号。

2) 以 ID 编号为 key 值, 其他数据为 $value$, 进行 Map 操作, 通过映射函数 (20) 获得每一个粒子的个体适应度 f_h , 更新 (p_h, f_h) , 构造键值对 $(key, value)$ 。

3) Reduce 操作, 获取每个群组最佳适应度和最优解, 用公式 (13) 得出 F , 以 ID 为 $key, value$ 的 $(F, (p_b, f_b))$ 键值对, 再 Join 操作, 与步骤 2) 的键值对接, 更新 $(F, (p_b, f_b))$, 若差分进化迭代结束, 输出 p_b 。

4) p_b 满足目标适应度 E , 停止差分进化迭代, 输出最优 p_b , 否, 判断 F 是否满足, 否, Map 操作, 用粒子群算法公式 (8) ~ (9) 获得 x, v 更新 (x, v) , 同时以 ID 编号为 key 值, $value(x, v)$ 转至步骤 2), F 满足, 转至 e。

5) Map 操作, 以 ZID 为 key , 其他数据为 $value$, 重新构造键值对 $(key, value)$ 。

6) 再一次 Reduce 操作, 以 AID 为 $key, (pf_{m/2,D+1})$ 为 $value$, 选择较好 $m/2$ 群组 $(ph_{m/2,D+1})$ 的最佳位置和最佳适应度, Join 操作, 与步骤 5) 的键值对连接。

7) 进行一次 Map 操作, 以 ID 为 key , 其他数据为 $value$, 用改进的差分进化算法, 得出 x 和 v , 更新 (x, v) , 转到步骤 2)。

表 6 Spark 的基本操作

| Spark 基本操作方法 | 方法具体描述 |
|--------------|---|
| Map 操作 | 按照用户编写的映射函数, 形成新的 (key, value); |
| Reduce 操作 | 按照 key 分组, 调用用户编写函数进行组内处理, 每个组返回一个值; |
| Join 操作 | 相同 key, 不同 value 的 RDD 数据集对接, 更新 value; |

5 实验与分析

5.1 预测误差评估指标

为了对提出的算法有效地评估, 选用均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE) 作为光伏发电预测误差的指标。

MAPE 评估系统的预测能力, RMSE 评估预测值的离散程度^[25-26]。公式分别如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (A_{n,i} - P_{n,i})^2}{N}} \quad (21)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{A_{n,i} - P_{n,i}}{P_{n,i}} \right| \quad (22)$$

式中, N 为光伏发电功率输出个数; $P_{n,i}$ 为第 i 时间段实际的光伏输出功率; $A_{n,i}$ 是第 i 时间段预测的光伏输出功率。

5.2 预测模型验证

实验数据来源于本地 30 kW 分布式光伏电站 2020 年 3 月 24 到 2021 年 12 月 31 的光伏发电功率数据和气象信息, 每组数据样本包含 6:00~18:00 的 12 个时间点云量信息、日最高温度、日最低温度、最高湿度以及最低湿度共 16 个环境输入变量, 和 6:00~18:00 的 12 个时间点光伏输出功率共 12 个功率输出变量。

为了验证预测模型的有效性, 选取晴转多云、阴转多云以及雨天的天气类型的待测日, 采用本模型方法进行相似样本集筛选、训练、预测, 用 RMSE、MAPE 进行分析, 再通过 BP、PSO-BP、PSO-DE-BP 三种优化算法进行对比分析, 包括预测精度以及稳定性分析。

表 7~9 分别选取 2021 年 2 月 24 日, 天气类型为晴转多云; 2021 年 6 月 20 日, 天气类型为阴转多云; 以及 2021 年 9 月 17 日, 天气类型为雷阵雨转阴作为待测日, 采用本文提出的算法, 以各个时间点云量作为输入变量, 跟踪预

表 7 预测日天气类型为晴转多云误差分析表

| 时间 | 天气类型 | 实际值/kW | 预测值/kW | 相对误差 |
|-------|------|--------|--------|---------|
| 6:00 | 多云 | 1.31 | 1.09 | 0.167 9 |
| 7:00 | 晴 | 3.54 | 3.73 | 0.053 7 |
| 8:00 | 晴 | 8.37 | 8.09 | 0.033 5 |
| 9:00 | 晴 | 14.221 | 3.97 | 0.017 6 |
| 10:00 | 晴 | 20.382 | 0.44 | 0.002 9 |
| 11:00 | 多云 | 19.812 | 0.22 | 0.020 7 |
| 12:00 | 晴 | 22.312 | 1.68 | 0.028 2 |
| 13:00 | 多云 | 12.741 | 3.24 | 0.039 2 |
| 14:00 | 多云 | 6.27 | 6.88 | 0.097 3 |
| 15:00 | 多云 | 3.59 | 2.99 | 0.167 1 |
| 16:00 | 多云 | 1.23 | 1.03 | 0.162 6 |
| 17:00 | 多云 | 0.77 | 0.59 | 0.233 8 |
| 误差 | RMSE | 0.3925 | MAPE | 0.085 4 |

表 8 预测日天气类型为阴转多云误差分析表

| 时间 | 天气 | 实际值/kW | 预测值/kW | 相对误差 |
|-------|------|--------|--------|---------|
| 6:00 | 阴 | 0.33 | 0.42 | 0.272 7 |
| 7:00 | 阴 | 1.27 | 1.63 | 0.283 5 |
| 8:00 | 阴 | 6.55 | 6.86 | 0.051 9 |
| 9:00 | 阴 | 5.22 | 6.07 | 0.162 8 |
| 10:00 | 阴 | 6.38 | 6.72 | 0.053 3 |
| 11:00 | 多云 | 9.81 | 8.97 | 0.085 6 |
| 12:00 | 阴 | 7.91 | 7.27 | 0.080 9 |
| 13:00 | 多云 | 13.431 | 3.24 | 0.014 1 |
| 14:00 | 多云 | 10.271 | 0.88 | 0.059 4 |
| 15:00 | 多云 | 10.691 | 1.23 | 0.050 5 |
| 16:00 | 多云 | 4.55 | 5.13 | 0.127 5 |
| 17:00 | 多云 | 1.97 | 1.21 | 0.385 8 |
| 误差 | RMSE | 0.4513 | MAPE | 0.112 8 |

测一天当中的天气类型变化。由表 7~9 可以得出, 晴转多云天气预测精度高于阴转多云以及降雨天气。

5.3 PSO-DE-BP 算法性能分析

为了进一步明确算法在不同天气下功率预测的通用性, 分别使用 PSO-DE-BP、PSO-BP、BP 三种算法对模型进行训练。对晴天、阴天以及雨天进行相似天气分组, 求得每种天气状况下预测的平均误差, 分别见表 10~12。

表 9 预测日天气类型为雨天误差分析表

| 时间 | 天气 | 实际值/kW | 预测值/kW | 相对误差 |
|-------|------|--------|--------|----------|
| 6:00 | 多云 | 0.17 | 0.22 | 0.294 1 |
| 7:00 | 多云 | 2.35 | 2.03 | 0.136 2 |
| 8:00 | 阴 | 7.43 | 7.84 | 0.055 2 |
| 9:00 | 阴 | 9.36 | 9.49 | 0.013 9 |
| 10:00 | 阴 | 9.13 | 8.72 | 0.044 9 |
| 11:00 | 中雨 | 3.03 | 2.68 | 0.115 5 |
| 12:00 | 中雨 | 2.53 | 2.07 | 0.181 8 |
| 13:00 | 小雨 | 2.63 | 2.25 | 0.144 49 |
| 14:00 | 小雨 | 2.71 | 2.24 | 0.173 4 |
| 15:00 | 阴 | 6.19 | 5.03 | 0.187 4 |
| 16:00 | 多云 | 4.38 | 3.42 | 0.219 2 |
| 17:00 | 多云 | 2.48 | 2.97 | 0.197 6 |
| 误差 | RMSE | 0.5526 | MAPE | 0.147 0 |

表 10 晴天预测误差分析表

| 模型类别 | BP | PSO-BP | PSO-DE-BP |
|------|---------|---------|-----------|
| MAPE | 0.095 2 | 0.063 2 | 0.044 4 |
| RMSE | 1.733 1 | 1.200 8 | 0.216 7 |

表 11 阴天预测误差分析表

| 模型类别 | BP | PSO-BP | PSO-DE-BP |
|------|---------|---------|-----------|
| MAPE | 0.324 9 | 0.201 3 | 0.123 2 |
| RMSE | 2.568 5 | 1.564 9 | 0.546 8 |

表 12 雨天预测误差分析表

| 模型类别 | BP | PSO-BP | PSO-DE-BP |
|------|---------|---------|-----------|
| MAPE | 0.389 3 | 0.241 6 | 0.152 8 |
| RMSE | 3.625 6 | 2.514 9 | 0.584 3 |

由表 10~12 三种方法精度对比分析得出, 晴天预测误差明显优于阴天和雨天, 这是由于云量变化的随机性导致预测准确性降低。且 PSO-DE-BP 方法的误差要小于 BP 以及 PSO-BP 两种算法, 因此采用该算法可获得更稳定的网络。

5.4 算法并行性能分析

为了测试 PSO-DE-BP 算法的并行化性能, 采用加速比来测试算法的并行性能, 加速比是衡量并行系统或程序并行化的性能和效果的指标^[27], 如公式 (23) 所示, 其中 T_{single} 单机下运行的时间, T_{parallel} 表示该算法并行化下运行的

时间, S_{speed} 是衡量并行性能的指标, 该值越大, 算法的并行性能的效果越好。

$$S_{speed} = T_{single} / T_{parallel} \quad (23)$$

通过增加集群的节点以及样本的数据量, 来测试 PSO-DE-BP 算法的并行性能, 表格 13 所示, 单节点相当于单机运算。

表 13 可知, 随着节点数的增长, 加速比的增长率变缓。因为节点数越大, 集群之间的通信和任务调度越多, 消耗时间也增多。随着样本数据量的增加, 加速比与节点数更接近同比例增长。因此, PSO-DE-BP 算法能满足海量高维电力数据的负荷预测的性能要求。

表 13 加速比与群集节点及样本数量关系

| 数据量/倍 | 加速比(S_{speed}) | | |
|-------|--------------------|-------|-------|
| | 1 个节点 | 2 个节点 | 4 个节点 |
| 1 | 1 | 1.74 | 2.96 |
| 10 | 1 | 1.79 | 3.11 |
| 100 | 1 | 1.86 | 3.34 |
| 1 000 | 1 | 1.91 | 3.52 |

6 结束语

本文首先建立了一个以各个时间段的云量作为输入变量的神经网络结构, 用云量跟踪天气的变化情况, 有效提高了多云、阴雨天的预测精度。然后基于 PSO-DE-BP 混合算法对 BP 神经网络初始值进行优化, 通过粒子群分组优化, 分别用 PSO 和改进的 DE 对群内、群间迭代进化, 引进了种群多样性算法, 避免早熟现象, 寻找全局最优, 与 BP 算法和 PSO-BP 相比, 具有更高的稳定性和预测精度。

通过基于 Spark 的内存分布式计算思想, 引进到 PSO-DE-BP 混合算法中, 实行并行计算, 通过加速比分析, PSO-DE-BP 算法具有良好的并行性能和处理大规模数据的能力。

参考文献:

[1] 蒋小波, 徐小艳, 刘乐平, 等. 天气分类和 BP 神经网络的光伏短期出力预测 [J]. 电源技术, 2020, 44 (12): 1809-1813.

[2] 吴 硕. 光伏发电系统功率预测方法研究综述 [J]. 热能动力工程, 2021 (8): 1-7.

[3] 赖昌伟, 黎静华, 陈 博, 等. 光伏发电出力预测技术研究综述 [J]. 电工技术报, 2019 (6): 1201-1217.

[4] 邵堃侠, 郭卫民, 杨 宁, 等. 基于 K-means 算法的 RBF 神经网络预测光伏电站短期出力 [J]. 上海电机学院学报, 2017, 20 (1): 27-33.

[5] SAHA A K, CHOWDHURY S, CROSSLEY P A. Modeling and Performance Analysis of a Microturbine as a Distributed Energy Resource [J]. IEEE Transactions on Energy Conversion, 2009, 24 (2): 529-538.

[6] 张扬科, 李 刚, 李秀峰. 基于典型代表电站和改进 SVM 的区域光伏功率短期预测方法 [J]. 电力自动化设备, 2021

(11): 205-210

[7] 周 文, 孟 良, 杨正富, 等. 基于时间序列的大型光伏电站发电短期预测 [J]. 电源技术, 2021, 45 (11): 1490-1494.

[8] 陈 涛, 王 艳, 纪志成. 基于经验小波变换的光伏功率组合预测模型 [J]. 系统仿真学报. 2021 (11): 2627-26

[9] MANDAL P, MADHIRA S T S, HAQUE A U, et al. Forecasting Power Output of Solar Photovoltaic System Using Wavelet Transform and Artificial Intelligence Techniques [J]. Procedia Computer Science, 2012, 12: 332-337.

[10] 丁 明, 王 磊, 毕 锐. 基于改进 BP 神经网络的光伏发电系统输出功率短期预测模型 [J]. 电力系统保护与控制, 2012, 11: 93-99.

[11] 袁晓玲, 施俊华, 徐杰彦. 基于 BP 神经网络的光伏发电短期出力预测 [J]. 可再生能源, 2013 (7): 11-16.

[12] 李 倩, 周彬倩, 张建成, 等. 基于自适应差分进化和 BP 神经网络的光伏功率预测 [J]. 陕西电力, 2014 (2): 23-27.

[13] 王章权, 张 超, 何湘宁. 基于 Pspice 模拟行为模型的光伏阵列建模 [J]. 计算机仿真, 2007 (8): 225-228.

[14] 张雪莉, 刘其辉, 马会萌, 等. 光伏电站输出功率影响因素分析 [J]. 电网与清洁能源, 2012 (5): 75-81.

[15] 胡克用, 李云龙, 江 霞, 等. 改进神经网络模型在光伏发电预测中的应用 [J]. 计算机系统应用, 2019, 28 (12): 37-46.

[16] 左远龙, 黄玉水, 杨晓辉, 等. 基于 PFA-MBAS-BP 神经网络模型的光伏发电短期预测 [J]. 电力系统保护与控制, 2020 (15): 84-91.

[17] 徐维超. 相关系数研究综述 [J]. 广东工业大学学报, 2012 (3): 12-17.

[18] 徐 华, 盛立芳, 柴艺淳. 青岛地面太阳总辐射变化特征及影响因子分析 [J]. 中国海洋大学学报 (自然科学版), 2014 (5): 25-31.

[19] 李鹏梅, 臧传治, 王侃侃. 基于相似日和神经网络的光伏发电预测 [J]. 可再生能源, 2013 (10): 1-4.

[20] 高海兵, 高 亮, 周 驰, 等. 基于粒子群优化的神经网络训练算法研究 [J]. 电子学报, 2004 (9): 1572-1574.

[21] 杨启文, 蔡 亮, 薛云灿. 差分进化算法综述 [J]. 模式识别与人工智能, 2008 (4): 506-513.

[22] 骆剑彬, 姜绍飞, 任 晖, 等. Spark 云计算平台下的结构物理参数辨识 [J]. 振动与冲击, 2018, 37 (14): 67-73.

[23] 王诏远, 王宏杰, 邢焕来, 等. 基于 Spark 的蚁群优化算法 [J]. 计算机应用, 2015 (10): 2777-2780.

[24] 王保义, 王冬阳, 张少敏. 基于 Spark 和 IPPSO_LSSVM 的短期分布式电力负荷预测算法 [J]. 电力自动化设备, 2016 (1): 117-122.

[25] 靳 果, 朱清智, 闫 奇. 基于 PCA 和 ML-ELM-AE 的短期光伏功率预测 [J]. 控制工程, 2021, 28 (9): 1787-1796.

[26] 张雪松, 李 鹏, 周亦尧, 等. 基于贝叶斯概率的光伏出力组合预测方法 [J]. 太阳能学报, 2021 (10): 80-86.

[27] 曾 飞, 苗世洪, 尚亚男, 等. 分解协调式节点阻抗矩阵生成算法的并行实现 [J]. 电网技术, 2011 (9): 99-104.