基于相似日和GA－DBN神经网络的光伏发电短期功率预测

光伏电站输出功率受多种外界环境因素影响显著，存在非线性、波动大等缺点。针对这一问题，提出改进的深度置信网络（Deep Belief Network，DBN）的方法。首先利用遗传算法（Genetic Algorithm，GA）为DBN神经网络选取最优的初始权值；其次利用灰色关联度法选择与预测日气象特征相似度高的日期，将这些日期的天气数据和历史发电功率作为训练样本训练DBN神经网络，建立短期光伏预测模型；最后通过仿真算例分析验证了该方法对传统DBN模型预测准确度的提升，且具有一定的可行性。

梁彩霞　高赵亮／江苏大学电气信息工程学院

梁彩霞／硕士研究生

关键词／Keywords
短期功率预测·
相似日·
深度置信网络·
遗传算法·

0　引言

得益于经济社会快速发展，光伏产业在电力系统中得到了广泛应用。光伏发电具有低碳环保、安全可靠、维护简单和规模灵活等优点[1]。光伏发电系统受到天文、气象和地理环境等多种因素的影响，输出功率具有不确定性。若大规模光伏发电系统接入电网，会对电网造成冲击，影响整个电网的安全稳定运行。因此，光伏发电功率的预测对于电网调度、电力负荷的配合具有重大意义，如何提高短期功率预测的准确度是国内外学者致力于研究的方向之一。

目前，对于光伏发电短期功率预测的方法主要有传统预测法[2-4]和人工智能算法。传统预测方法对于变化规律明确、趋势单一的数据序列具有较好的预测结果，但这些方法采用的数学模型较为简单，对于非线性系统的预测准确度有限。人工智能算法突破了传统算法的局限性，近年来得到广泛应用，主要有支持向量机[5-6]（Support Vector Machine，SVM）、小波分析法[7]和人工神经网络法[8-9]等。

基于此，本文选择人工智能算法中的DBN神经网络来进行光伏输出的短期预测，但由于光伏输出受气象因素影响较大，在不同气象环境下表现出不同的发电特点。因此很多学者已经提出了按照天气类型分类预测的方式。文献[10-11]分别按照季节和天气类型来分类，但按照一种天气类型来分类太过笼统，没有考虑到各种环境因素的综合影响。因而本文利用灰色关联法来选择预测日的相似日，其中涉及了六种气象特征。

遗传（GA）算法[12]也常用于神经网络预测中，用来优化神经网络的初始权值和阈值，可有效避免算法陷入局部最优解和收敛速度慢的缺陷。文献[13]利用改进的GA对神经网络的权值进行优化，最终提高了负荷预测的准确度。

综上所述，本文提出了利用改进的深度置信网络（DBN）模型进行预测。利用GA为DBN神经网络选取最优的初始值，并且利用灰色关联度法选择与预测日气象特征相似度高的日期，将这些日期的气象数据和历史发电功率作为训练样本对DBN神经网络进行训练，最后利用训练好的模型对预测日的输出功率进行预测。仿真结

国家自然基金（51477070），江苏高校优势学科建设工程资助项目 PAPD（20116）。

果验证了所建立的模型可提高传统DBN模型预测的准确度，并具有一定的可行性。

1 选取相似日

1.1 天气类型编码

影响光伏发电输出功率的因素很多，如所处地理位置、气象环境以及所安装光伏阵列固有的系统信息等。对于确定的光伏电站来说，由于其历史发电功率序列来自同一发电系统，数据本身就包含了光伏阵列的系统信息，所以本文不考虑光伏阵列的转换效率、安装角度和位置等隐含在历史功率数据中的因素[14]。

光照强度会随着天气类型的变化而变化，不同天气类型下光伏阵列的输出差异很大。根据不断完善的光伏监控数据系统，对大量有效的光伏电站输出量依据天气类型进行统计分析，将不同天气类型映射为0～1之间的具体数值作为模型的一个输入变量[15]。

1.2 基于灰色关联法选择相似日

（1）构造气象特征向量

根据影响光伏发电输出的各气象因素，构造每天的气象特征向量，即

$$D=[w,\overline{E},T^{\mathrm{h}},T^{\mathrm{l}},H^{\mathrm{h}},H^{\mathrm{l}}] \quad (1)$$

式中，w为天气类型；$\overline{E}$为日平均辐照度；T^{h}和T^{l}分别为每日的最高和最低温度；H^{h}和H^{l}分别为每日的最高和最低湿度。

（2）计算相似度

1）对气象特征向量中各个分量进行归一化处理，即

$$x_j(i)=\frac{d_j(i)-m(i)}{M(i)-m(i)} \quad (2)$$

式中，$d_j(i)$为第j日的第i个分量；$M(i)$、$m(i)$分别为所有特征向量第i个分量的最大值和最小值，$i=1\sim6$。

归一化后预测日和第j日的特征向量分别为

$$x_0=[x_0(1),x_0(2),x_0(3),x_0(4),x_0(5),x_0(6)] \quad (3)$$

$$x_j=[x_j(1),x_j(2),x_j(3),x_j(4),x_j(5),x_j(6)] \quad (4)$$

2）计算x_0和x_j在第k个分量的关联系数为

$$x_j(k)=\frac{\min\limits_j\min\limits_k\left|D_j(k)\right|+r\max\limits_j\max\limits_k\left|D_j(k)\right|}{\left|D_j(k)\right|+r\max\limits_j\max\limits_k\left|D_j(k)\right|} \quad (5)$$

式中，$D_j(k)=x_0(k)-x_j(k)$；r为分辨系数，一般取$r=0.5$。

3）综合各个分量的关联系数，采用连乘方式定义相似度可有效解决各个分量权重的问题，x_0和x_j的相似度定义为

$$F_j=\prod_{k=1}^{6}x_j(k) \quad (6)$$

2 深度置信网络(DBN)原理

深度置信网络由若干个受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)层层堆叠而成，RBM的模型如图1所示。其结点包括可见结点和隐含结点两种，可见层和输入层的结点间双向对称连接，同层的结点之间没有连接。DBN的模型如图2所示，DBN由可见层、隐含层和输出层构成，第一个RBM的可见层即为DBN的可见层，前一个RBM的输出作为下一个RBM的输入。传统的RBM只能处理二值数据，而连续型受限玻尔兹曼机（Continuous Restricted Boltzmann Machine, CRBM）的神经元具有能够连续变化的状态值和可以处理实值数据的优点[16]，本文中的DBN采用CRBM。

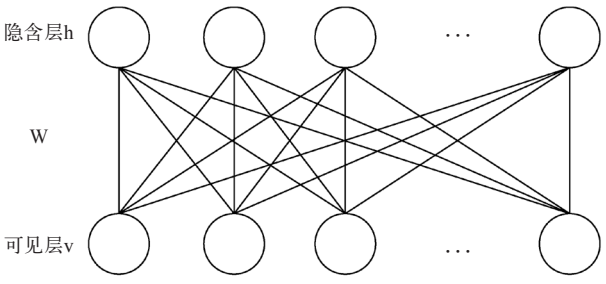

图1 RBM的结构

DBN模型的训练过程主要可分为预训练和微调整两个阶段。在预训练阶段，采用无监督式逐层贪心训练的方法，随机选择训练样本，输入至DBN网络，训练第一个RBM，使其能够捕获输入

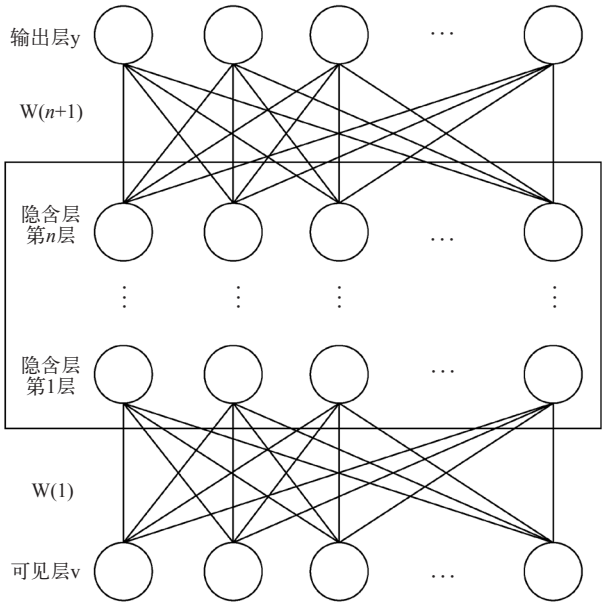

图2 DBN的结构

数据的重要特征，再将第一个RBM的输出作为下一个RBM的输入，依次类推，直至DBN模型完成训练。此时DBN的输出特性并未达到最优状态。在微调整阶段，采用有监督的训练方式，对DBN的权重矩阵和偏置量进行微调，提高模型预测的准确度。

3 GA－DBN预测模型设计

人工神经网络的收敛性能受其初始值影响较大。DBN神经网络的初始权值和阈值是随机选取的，如果这些初始值选取不当，则会导致收敛到全局最优的速度缓慢或者陷入局部最优解。GA是一种多点搜索全局性优化算法，有较大概率收敛到全局最优解。考虑到DBN的特点，将GA与DBN网络结合起来，利用GA为DBN选取一组优秀的初始权值和阈值，克服初始值选取的盲目性。这样不仅能够发挥神经网络的非线性泛化映射能力，还能使DBN神经网络克服收敛速度慢和陷入局部最优解的缺点[17]。

3.1 遗传算法的应用

GA是基于自然进化原理的优化算法。在由待优化参数组成的种群中，对个体按照选择的适应度函数进行筛选，并进行遗传学中的交叉变异等操作，个体产生子代的概率与其适应度成正比，因而在不断进化筛选的过程中，子代在继承上一代的基础上，适应度函数值不断提高。在满足结束条件时，选取个体适应度值最高的个体。GA训练DBN神经网络初始值的流程如图3所示，具体步骤如下。

1）种群初始化。个体编码采用实数编码的方式，这样可避免复杂的编码解码过程，每个个体均为一个实数串，由输入层和隐含层、隐含层和隐含层、隐含层和输出层的连接权值以及隐含层和输出层的阈值组成。

2）适应度函数的选择为

$$F_i = 1/E_i \tag{7}$$

式中，E_i为第i个个体学习之后的预测输出与期望输出之间的误差。

3）基因选择。将得到的各个体的适应值按照降序排列，采用轮盘赌选择法进行操作，得到其在子代中出现的概率，适应度越高的个体遗传到子代的概率越大，个体i被选择的概率为

$$P_i = F_i / \sum_{i=1}^{c} F_i \tag{8}$$

式中，c为个体总数目。

4）基因交叉。采用实数交叉法对配对的两个个体以概率p_c交换其中部分基因，生成两个新的个体。第m和n个个体在第j位交叉操作如下，即

$$\begin{cases} g'_{mj} = g_{mj}r + g_{nj}(1-r) \\ g'_{nj} = g_{nj}r + g_{mj}(1-r) \end{cases} \tag{9}$$

式中，g_{mj}、g_{nj}分别为交叉前第m和第n个个体在第j位的基因；$r \in [0,1]$。

5）基因变异。为增加种群中个体的多样性，以一个较低的概率p_m进行变异操作。随机选取第m个个体的第j位基因进行变异，操作方法如下，即

$$g'_{mj} = \begin{cases} g_{mj}r + (g_{mj} - g_{\max})r_1(1 - s/s_{\max})\ r_2 \geqslant 0.5 \\ g_{mj}r + (g_{\min} - g_{mj})r_1(1 - s/s_{\max})\ r_2 < 0.5 \end{cases} \tag{10}$$

式中，$g_{\min}$、$g_{\max}$分别为第m个个体的第j位基因上的最小值和最大值；s为当前的迭代次数；$s_{\max}$为

最高进化次数；r_1为随机数；$r_2 \in [0,1]$。

6）判断是否满足终止条件。变异操作之后，若达到设定的最高迭代次数或计算得到的适应度函数值达到设定的准确度，则将其中适应度最高的个体输出；否则进行下一轮的复制、交叉或变异操作，直至满足条件。

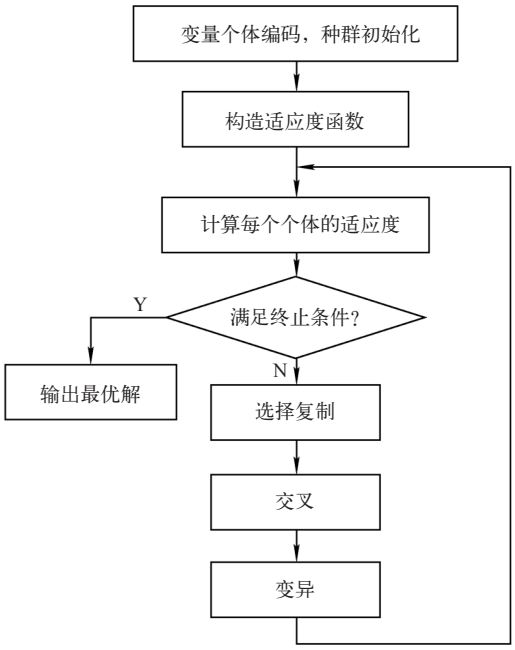

图3 遗传算法流程

3.2 GA－DBN模型的建立

利用GA所获得的初始权值矩阵和阈值构成初始的DBN网络。之后进行DBN的训练阶段，训练过程[19]如下。

1）设结点状态用$\{S_i\}$表示，与结点i相连的结点j的状态用S_j来表示，权重矩阵为$\boldsymbol{W}$。随机选取一个训练样本输入到第一个CRBM中（即DBN的可见层），更新第一层隐含层各结点的状态S_j，即

$$S_j = \varphi_j\left(\sum_i \omega_{ij} S_i + \sigma N_j(0,1)\right) \tag{11}$$

式中，$\varphi_j(x)$为一个S型函数；$N_j(0,1)$为标准正态分布。

2）根据步骤1）得到的隐含结点状态S_j，按照式（12）更新可见结点状态，记为S_i，即

$$S_i = \varphi_j\left(\sum_j \omega_{ij} S_j + \sigma N_i(0,1)\right) \tag{12}$$

3）根据步骤2）更新的可见结点状态S_i，按照式（13）再次更新隐含结点状态，记为S_j，即

$$S_j = \varphi_j\left(\sum_i \omega_{ij} S_i + \sigma N_j(0,1)\right) \tag{13}$$

4）将第一个CRBM的隐含层结点状态S_j作为第二个CRBM的输入，再按照步骤1）～步骤3）更新第二个CRBM各个结点的状态。按照这个步骤层层更新下去，直至将整个网络的结点状态更新。

5）随机选取下一个训练样本，跳转到步骤1）。当输入所有训练样本时，本轮训练结束，则根据式(14)计算每个权重值的变化量，以更新各个CRBM的权重矩阵，即

$$\Delta\omega_{ij} = \eta_\omega(< S_i S_j > - < S_i S_j >) \tag{14}$$

6）转到步骤1），进行下一轮训练，直至达到设定的最高训练次数或网络的权重矩阵的变化量足够小，即当$||DW(i)|| < e$时，结束DBN训练，DBN网络建模完成。

4 算例分析

本文所用数据来自浙江龙游光伏发电站(119.17E，29.03N)，是装机容量为10 MW的分布式屋顶光伏电站，共有17个逆变器，时间区间为2015年6月30日~2016年4月30日。功率实测值每隔1 h采集一次，即每日的7：00~12：00的12个整点。将历史发电量数据和气象数据分为训练样本和测试样本。利用上文所述的灰色关联法选取预测日的20个相似日构成训练样本。利用GA找到DBN网络初始最优权值和阈值，再利用训练样本训练DBN神经网络，得到光伏预测模型。随机选取2015年8月18日，当日天气类型为晴，气温为25～33 ℃，湿度范围为60%～92%。利用本文预测模型预测该日的12个光伏发电量数据，选择与预测日相似度最高的日期发电量作为预测模型的输入。预测结果如图4所示。除利用本文模型进行发电量的预测之外，还利用了传统的DBN网络预测进行对比验证。

对于预测结果的评判标准，本文采用平均绝对百分比误差MAPE和均方根误差RMSE。其中，MAPE值用作分析整体预测的准确度，RMSE用作

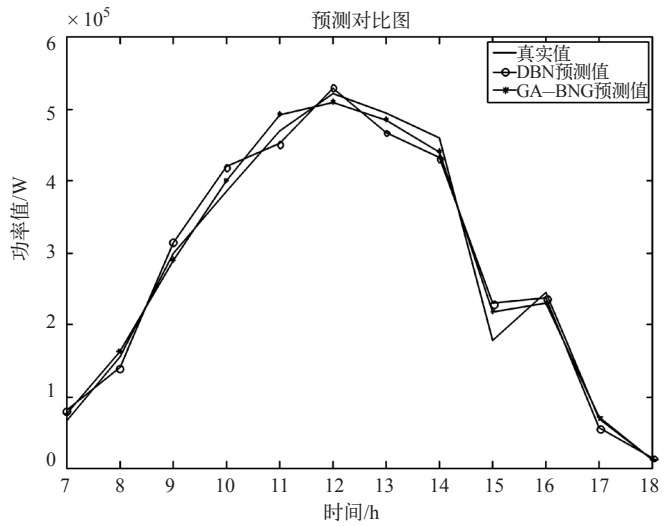

图4 光伏发电量预测结果

衡量预测值与真实值的偏离程度。具体计算的方法为

$$\mathrm{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|(y_i-y_i')/y_i\right|\times 100\% \tag{15}$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-y_i')^2}{n}} \tag{16}$$

式中，y_i为实际功率输出；y_i'为预测输出功率；n为序列数据总数。

对两种预测方法进行预测指标分析，其结果如下表所示。

表 预测误差指标分析

误差指标（%）	传统DBN模型	本文模型
MAPE	10.74	6.09
RMSE	13.42	8.47

由上表可知，本文预测方法的MAPE值和RMSE值分别为6.09%和8.47%，而传统的DBN模型MAPE值和RMSE值分别为10.47%和13.42%。基于本文设计模型预测结果的MAPE值和RMSE值明显低于传统的DBN模型，准确度更高，验证了所提出的光伏发电预测模型的有效性和准确度。

5 结束语

光伏发电系统的发电量受多种外界因素影响，具有不确定性。大规模的光伏并网对电网的影响不容忽视。基于此，本文将GA与DBN神经网络结合，即利用GA为DBN神经网络选择最优的初始值，并利用灰色关联度选取与预测日相似的日期历史数据作为DBN网络的训练样本。对该预测模型进行光伏输出功率短期预测，得出以下结论：

1）利用灰色关联度的方法，按照气象特征选取与预测日相似度高的日期进行模型的训练，更符合光伏功率预测的需要，建立的预测模型进行预测时更具有针对性，从而提高了整体预测的准确度。

2）本文将GA与DBN网络相结合，利用GA为DBN网络选取一组最优的初始权值和阈值，可有效避免传统算法容易局部最优解和收敛速度慢的缺点。

参考文献

[1] 赵玉文．太阳能利用的发展概况和未来趋势[J]．中国电力，2003，36(9)：63-69.

[2] 安德洪，柳湘月，刘嘉焜，等．基于季节ARIMA模型的电力负荷建模与预报[J]．天津大学学报：自然科学与工程技术版，2004，37(2)：184-187.

[3] 李钜，李敏，刘涤尘．基于改进回归法的电力负荷预测[J]．电网技术，2006，30(1)：99-104.

[4] 李栓，刘莉，刘阳．趋势外推法在电力负荷预测中的应用[J]．沈阳工程学院学报：自然科学版，2005，1(2)：64-65.

[5] 王继东，宋智林，冉冉．基于改进支持向量机算法的光伏发电短期功率滚动预测[J]．电力系统及其自动化学报，2016，28(11)：9-13.

[6] 吴坚，郑照红，薛家祥．深度置信网络光伏发电短时功率预测研究[J]．中国测试，2018(5)：6-11.

[7] Kawasaki N，Oozeki T，Otani K，et al．An evaluation method of the fluctuation characteristics of photovoltaic systems by using frequency analysis[J]．Solar Energy Materials and Solar Cells，2006，90(18–19)：3356-3363.

[8] 肖同录，赵增顺．基于深度信念网络的短期电力负荷预测[J]．电子世界，2014(10)：186-187．

[9] 杨晓帆，陈廷槐．人工神经网络固有的优点和缺点[J]．计算机科学，1994，21(2)：23-26．

[10] 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．

[11] Friedrich L，Afshari A．Short-term forecasting of the abu dhabi electricity load using multiple weather variables[J]．Energy Procedia，2015，75：3014-3026．

[12] 黄建国，罗航，王厚军，等．运用GA-BP神经网络研究时间序列的预测[J]．电子科技大学学报，2009，38(5)：687-692．

[13] 王武，张元敏，蔡子亮．基于遗传优化神经网络的电力系统短期负荷预测[J]．电力系统保护与控制，2008，36(9)：39-42．

[14] 张岚，张艳霞，郭嫦敏，等．基于神经网络的光伏系统发电功率预测[J]．中国电力，2010，39(9)：75-78．

[15] 倪春花，陈国恩，朱伟，等．基于相似日理论和BP神经网络的光伏发电功率预测[J]．电气应用，2016，35(1)：42-48．

[16] Chen H，Murray A F．Continuous restricted boltzmann machine with an implementable training algorithm[J]．IEE Proceedings - Vision，Image and Signal Processing，2003，150(3)：153-158．

[17] 陈艳．基于遗传神经网络的短期电力负荷预测研究[D]．大连：大连理工大学，2005．

[19] 晁静．基于DBN的汇率预测研究[D]．南京：南京大学，2012．

(收稿日期：2018－08－14)EA

典藏技术 革新应用

专业专注

时刻关注并推动中国电气技术的发展

《电气应用》是一本拥有光荣发展历程的全国优秀科技期刊。

自1982年创刊至今，已走过三十多年艰苦创业、改革发展的光辉历程。其所荣获的“全国优秀期刊、中国期刊方阵‘双奖期刊’和国家期刊奖‘百种重点期刊’、中文核心期刊”等诸多荣誉见证了编辑们的专业与勤奋，也博得了广大读者和业内专家的支持与厚爱，在众多电工类专业技术期刊中颇具影响力。

国内统一刊号：CN11-5249/TM 国际标准刊号：ISSN1672-9560 国内邮发代号：82-341 国外发行代号：M1409