

文章编号: 2096-1618(2017)06-0584-06

# 基于 PCA 的太阳辐射观测算法研究

郑 丹<sup>1</sup>, 马尚昌<sup>1,2</sup>, 张素娟<sup>1,2</sup>

(1. 成都信息工程大学电子工程学院, 四川 成都 610225; 2. 中国气象局大气探测重点开放实验室, 四川 成都 610225)

**摘要:**研究太阳辐射对认识气候变化有重要影响,而中国地面台站的太阳辐射观测数据不足,质量不佳。针对提高太阳辐射观测的精确度,提出利用主成分分析法(PCA)对影响太阳辐射观测的多个气象要素进行降维处理,剔除冗余变量,结合遗传算法(GA)获取BP模型的最优权值阈值,并用实测数据加以验证该模型的可行性,充分提高模型性能以实现太阳辐射的观测研究。结果表明:基于PCA的GA-BP网络模型的观测精度高于传统BP模型,该模型有效地提高模型的泛化能力,具有一定的可行性及指导意义。

**关键词:**太阳辐射;主成分分析;遗传算法;BP神经网络;误差分析

**中图分类号:**TP183

**文献标志码:**A

**doi:**10.16836/j.cnki.jcui.2017.06.002

## 0 引言

太阳辐射是地球大气的主要能源,是促进地球上的水、大气、生物活动和变化的主要动力,在许多领域都具有重要作用<sup>[1]</sup>。太阳辐射的影响因子多,且与多种气象要素的耦合关系复杂性强,因此需要大量的广范围的太阳辐射台站,经长时间观测,综合各种气象要素进行分析。而中国太阳辐射观测站点少、分布不均、测量时间短、太阳辐射数据的不足和质量不高严重制约着气象、水文、能源、生态等多种模型的建立和发展,针对这些问题,有必要采取间接法获取太阳辐射数据,可见太阳辐射数据的研究具有基础性和重要意义。

现如今,如何对太阳辐照度进行有效、准确地预测已成为重要课题。目前对太阳辐射的间接计算法可简单归纳为3种,第1种方法是通过大气辐射传输模型,模拟其传输的物理过程,但该方法较为复杂、输入参数难以获取,局限性太大;第2种方法为经典模型,利用常规气象要素及台站地理位置信息,模拟太阳辐射<sup>[2]</sup>。而典型的晴天太阳辐射、半正弦、ARMA逐时等模型不能很好地表达与多种气象因素之间的多变量、强耦合的关系及应对太阳辐射强度的时变性、随机性,同时用某个确切模型对太阳辐射进行准确细致描述的难度系数太大,应用没有普适性的传统模型,会引起较大误差并难以移植,所以也不能满足需求;文献[3]主要是建立智能算法的预测模型,如采用神经网络(neural network, NN)模拟太阳辐射,具有较强学习

推广力,能非线性建模等特点。文献[4]利用典型的NN对太阳辐射进行估算,得出的精度优于经验公式;文献[5]提出基于蚁群算法改进BP神经网络模型,选取最优影响因子作为输入变量,对隐含层数进行优化处理,有效提高了辐照强度的预测精度;文献[6]则采用主成分分析法对输入变量进行降维处理、并进行网络结构参数的调整以及网络训练初始权值的选择,从这3方面出发,结合多种算法,实现对模型泛化能力的提高;文献[7],在晴天条件下利用LS-SVM方法模拟常规气象要素对太阳辐射的影响,针对非线性、小样本等问题建立逐时太阳辐射模型,降低维数、避免出现局部极小值。

提出一种基于主成分分析(PCA)的GA-BP神经网络的综合模型,采用PCA对输入变量(光照强度、温度、湿度、风速、气压等常规气象数据)进行降维处理,删除冗余信息,利用GA算法优化BP网络模型模拟太阳辐射,并与传统的BP模型进行对比,证明该模型的有效性。

## 1 直接辐射表数据可行性分析

研究所用到的太阳辐照度、光照强度及其他常规气象数据分别由同处于成都信息工程大学综合气象观测场(30°35'06"N, 103°59'23"E, 海拔:488 m)的直接辐射表、数字日照计和自动气象站获取。太阳辐射的实测值来源于热电型直接辐射表,对辐射表测量准确度的验证、数据可行性的分析则成为首要任务。

通过双轴跟踪辐射表观测到的总辐射 $E_g$ 、散射辐

射  $E_d$  及当时的太阳高度角  $H_A$  计算出水平面直接辐射  $S_L$ 、垂直面直接辐射  $S$ :

$$S_L = E_g - E_d$$
$$S = S_L / \sin(H_A)$$

太阳直接辐照度值随太阳高度角变化而变化,为此需对以下参数进行进一步的确定。

- (1) 随季节而变的太阳赤纬角为
- $$\delta = 23.45 \times \sin(360^\circ \times (284 + n) / 365)$$
- (2) 太阳时角用  $\omega$  表示,但时角受时差的影响,成都地区以北京时间计时,两地实际上存在  $(120^\circ - \text{当地经度})$  的时差<sup>[8]</sup>,所以以中国北京时间为依据的成都地区时角  $\omega$  为

$$\omega = (12 - t) \times 15^\circ + (120^\circ - \varphi)$$

- (3) 太阳高度角的表达式为

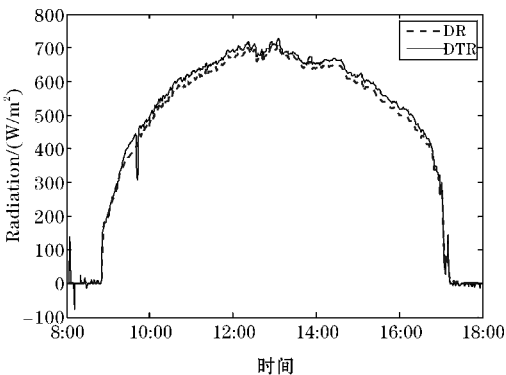
$$\sin(\alpha) = \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \cos(\omega)$$

式中,  $\delta$  为太阳赤纬;  $n$  为按天数顺序排列的积日;  $\omega$  为太阳时角;  $t$  为北京时间;  $H_A$  为太阳高度角;  $\varphi$  为当地纬度。

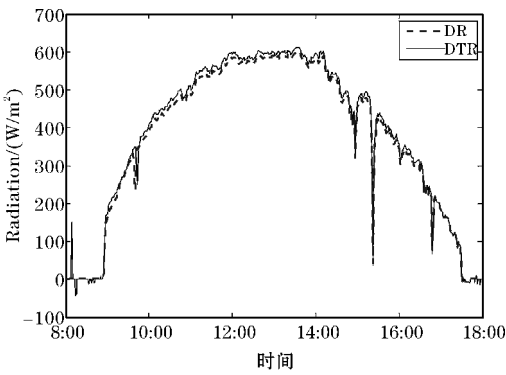
通过直接辐射表(DR)与双轴跟踪辐射表(DTR)所测得的直接辐射值对比分析,验证直接辐射表的准确度。抽取 2017 年 1 月 12 日、1 月 13 日、2 月 17 日和 4 月 14 日的 8:00–18:00 的辐射数据,统计 DR 和 DTR 两种辐射表所实测的直接辐射值,得出其相对误差如表 1 所示,曲线结果如图 1 所示。

表 1 直接辐射数据

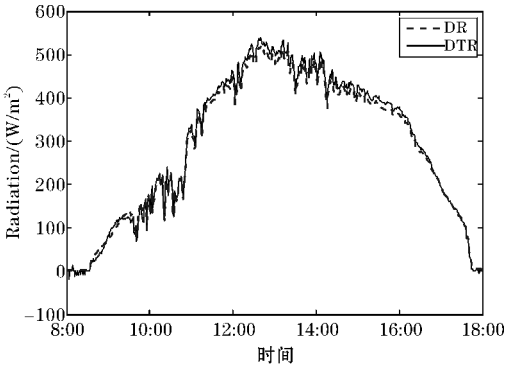
日期	1 月 12 日	1 月 13 日	2 月 17 日	4 月 14 日
相对误差	0.0245	0.0269	0.0285	0.0471



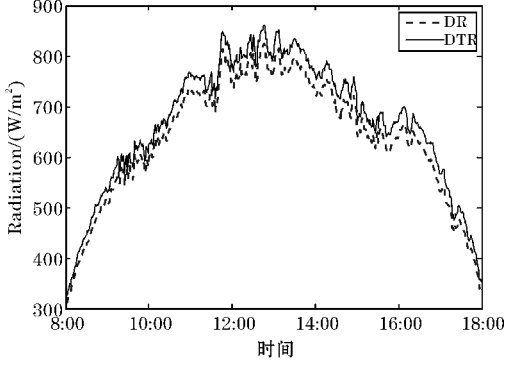
(a) 2017 年 1 月 12 日



(b) 2017 年 1 月 13 日



(c) 2017 年 2 月 17 日



(d) 2017 年 4 月 14 日

图 1 直接辐照值对比

通过两种不同辐射表所测直接辐射值的相互验证,由表 1 的相对误差统计及图 1 的曲线结果显示:相对误差均小于 5%,直接辐射表的测量准确度较好,可将其作为太阳辐射数据对比分析的参考标准。

## 2 太阳辐射观测的主成分分析

主成分分析法 (pinciple components analysis, PCA) 是经验正交函数分解系列方法之一<sup>[9–10]</sup>。运用

降维思想,将多个独立变量用少数综合的变量概括,简化输入变量,便于理解描述,促进训练效率的提升并缩短时间,有效改进太阳辐射的估算模型。以晴天天气类型为例,对太阳辐射观测的影响因素进行主成分分析,其主要分析步骤为:

### (1) 数据标准化

主成分分析的结果受量纲影响,应先把各变量数据标准化,设影响太阳辐照度的气象影响因素由  $X = [X_1, X_2, \dots, X_p]$  构成,对原始数据进行标准化:

$$x_i = (X_i - \bar{X}_i) / S_i$$

式中:  $\bar{X}_i$  为样本均值,  $S_i$  为样本标准差。

(2) 计算相关系数矩阵  $R$

根据  $[x_1, x_2, \dots, x_p]$ , 计算样本的相关系数  $R = [r_1, r_2, \dots, r_p]$ 。

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

式中:  $r_{ij}(i, j = 1, 2, \dots, p)$  为原变量  $x_i$  与  $x_j$  的相关系数。

(3) 计算特征值和特征向量

求出特征值, 并按次进行排列:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \dots \lambda_p \geq 0$$

其特征值  $\lambda_i$  所对应的特征向量为  $\mu_i (i = 1, 2, \dots, p)$ , 则可通过  $Z = [\mu_1, \mu_2, \dots, \mu_p]^T X$  的线性变换表示降维后的新变量, 其中:  $\|\mu_i\| = 1$ 。

(4) 求各个主成分的贡献率及累计贡献率主成分  $i$  的贡献率为

$$\alpha_i = \lambda_i / \sum_{k=1}^p \lambda_k \quad (i = 1, 2, \dots, p)$$

前  $n$  个主成分的累计贡献率为

$$\alpha_n = \sum_{k=1}^n \lambda_k / \sum_{k=1}^p \lambda_k \quad (i = 1, 2, \dots, p)$$

当  $\alpha_n \in [85\%, 95\%]$  时, 则选取前  $n$  个作为模型的输入变量。

3 基于 PCA 的 GA-BP 的太阳辐射观测模型

3.1 GA 优化 BP 神经网络权值和阈值

遗传算法 (genetic algorithm, GA) 是模拟生物在自然环境中的遗传和进化过程形成的具有较强容错性和适应性的随机搜索与优化算法<sup>[11-12]</sup>, 其本质上是一种并行、高效、简便、快捷的全局搜索方法。它是一种基于概率的搜索技术, 利用群体搜索策略和个体信息互换, 自动获取、积累有关搜索空间的信息, 跳出局部极值以获取最优解<sup>[13-14]</sup>, 其算法运算流程如图 2 所示。

对比于传统的算法, GA 具有以下特点及优势:

(1) GA 以问题解集为基础, 而非从单个解逐次迭代, 它能同时从不同解获得多个机制, 因此更利于全局择优;

(2) GA 针对参数的编码集进行计算, 而非参数本身, 对于处理只有代码而无数值的优化问题, 更具独特优势;

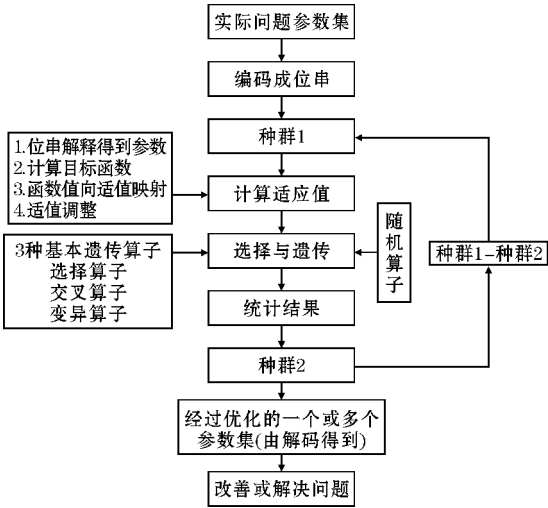


图 2 遗传算法流程图

(3) GA 的搜索信息为目标函数值, 而无须过多附加操作, 成功避免了函数求导障碍;

(4) GA 具有隐含的并行性, 使用的几种基本遗传算子都是随机操作, 增加了其灵活性;

(5) GA 具有自组织、自适应、自学习等特性, 且可扩展能力强, 易于同别的算法相结合, 生产综合双方优势的混合算法。

3.2 GA-BP 神经网络的建立

选取并收集相关晴天天气条件下的气象数据, 将其进行主成分分析, 提取主要影响因素, 作为 GA-BP 神经网络模型的输入变量 ( $Z_1, Z_2, Z_3$ ), 通过对初始种群数、遗传代数、遗传算法等的设定来获取 BP 网络最优连接权及阈值, 建立有 7 个隐层神经元的 BP 训练, 并对其余数据进行测试, 其 GA-BP 模型结构如图 3 所示。

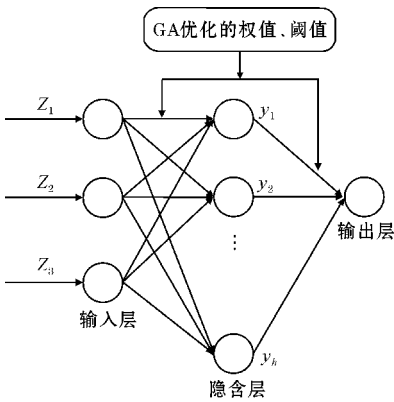


图 3 GA-BP 模型结构

3.3 太阳辐射观测模型性能评估

采用 MAPE (mean absolute percentage error)、MSPE (mean square percentage error) 以及 RE (relative error)

指标对太阳辐射观测模型结果进行评估<sup>[15]</sup>,这些指标能减少正负抵消,很好地衡量预测模型的无偏性,达到其评估能力。

平均绝对误差:

$$E_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{|Y_i - y_i|}{y_i} \right)$$

均方百分比误差:

$$E_{\text{MSPE}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_i - y_i}{y_i} \right)^2$$

式中: $Y_i$  表示太阳辐射预测值, $y_i$  代表太阳辐射实测值, $n$  为样本数。

4 算例分析

4.1 基于 PCA 的天气因素相关性分析

收集整理成都信息工程大学气象观测场的光照强度及自动站气象数据,以晴天数据为例,对影响太阳辐射观测的 5 个原始气象数据进行主成分分析,选取更适合的输入变量。原始数据为晴天天气条件下,8:00–18:00 光照强度、温度、湿度、风速、气压的分钟数据。主成分分析结果见表 2。

表 2 特征值及其累计贡献率

编号	初始特征值		
	特征值	贡献率/%	累计贡献率/%
1	3.0890	61.78	61.78
2	1.0169	20.34	82.12
3	0.5492	10.98	93.10
4	0.3193	6.39	99.49
5	0.0257	0.51	100.00

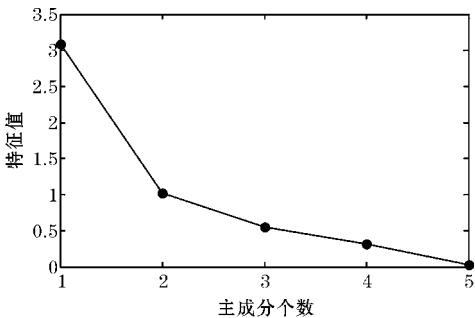


图 4 碎石分布

由表 2 可以确定,前 3 个特征值的累计贡献率已达到 93.10%,所反映的分析效果较好,选取前 3 个主成分为输入变量。同时在图 4 中,主成分个数为 2 时,虽然出现明显拐点,但此时包含的原始信息量仅为

82.12%,因此提取的主成分个数选取 3。

用  $X_1 \sim X_5$  表示原有的 5 个天气变量, $Z_1 \sim Z_3$  为降维后提取的主成分,其各主成分的表达式为

$$\begin{aligned} Z_1 &= 0.4238X_1 + 0.1025X_2 + 0.8843X_3; \\ Z_2 &= 0.5451X_1 + 0.0251X_2 - 0.2192X_3; \\ Z_3 &= -0.5283X_1 + 0.1158X_2 + 0.3735X_3; \end{aligned}$$

4.2 GA 优化 BP 模型输入变量的权值和阈值

设有  $N$  个个体的参数编码种群,其染色体长度为  $L=L_1 \cdot R + R \cdot L_2 + R + L_2$  ( $L_1$ 、 $L_2$  为输入、输出向量的维数, $R$  为常数),设定初始种群数为 30,遗传代数为 100,按照给定的适应值进行迭代,得到适应度曲线图(图 5),从图 5 可以看到,25 次迭代后,其最佳适应度处于稳定不变状态,100 代终止后获取一组最优的初始权值和阈值,如表 3 所示。

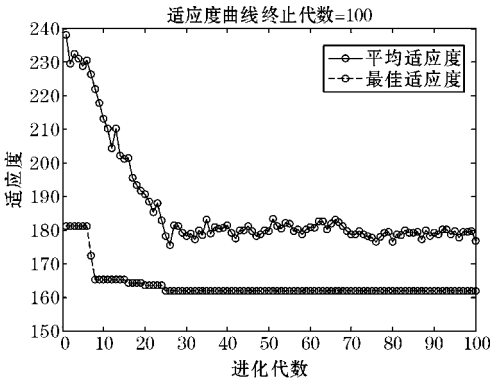


图 5 适应度曲线

表 3 GA 优化后的最优初始权值和阈值

序号	权值 $w_1$	权值 $w_2$	权值 $w_3$	阈值 $b_1$
1	0.1341	-0.0574	-2.4679	-2.0509
2	-1.4949	-0.3146	0.8278	-2.6271
3	1.4930	2.9557	2.5932	1.2111
4	-2.4466	2.7212	-2.0232	-1.1100
5	2.8227	0.5820	-1.5586	0.7007
6	-2.5782	-1.1998	1.8813	-1.9574
7	-2.5397	-0.8721	-2.2079	0.9084

4.3 GA-BP 神经网络模型

从模型的迭代情况来看,9 月 11 日 GA-BP 经过 64 次迭代达到训练误差值,BP 迭代 84 次;1 月 13 日 GA-BP 经过 136 次迭代而 BP 是 118 次;2 月 17 日的迭代次数分别为 188 次、104 次,GA-BP 在第 168 次迭代时训练误差达到 0.00303,如图 6 所示。



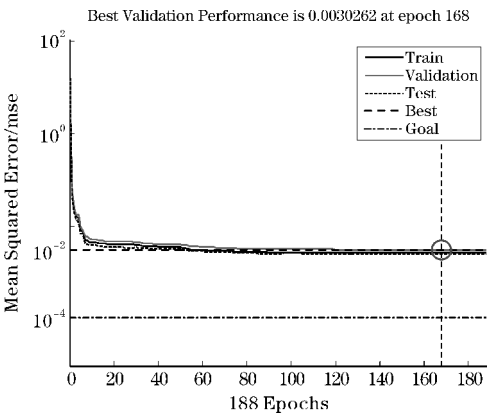
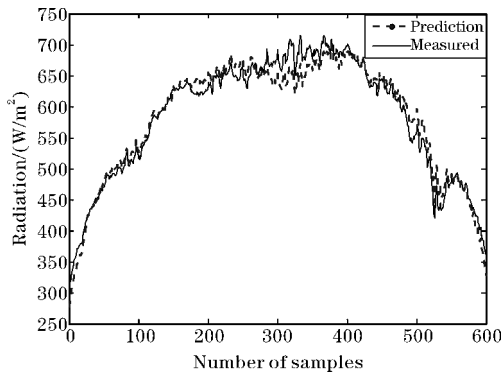
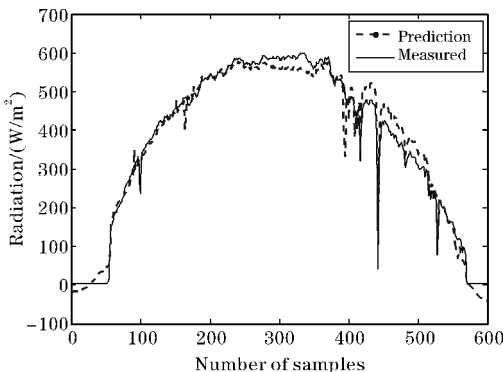


图 6 2017 年 2 月 17 日训练过程中误差变化曲线

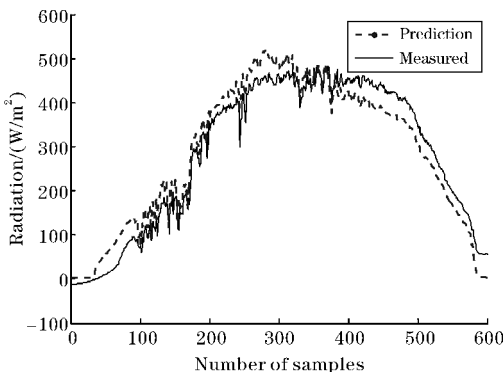
选取 6 个月的晴天数据进行训练,对 2016 年 9 月 11 日、2017 年 1 月 13 日以及 2 月 17 日做太阳辐照度的预测,其经过基于 PCA 的 GA-BP 模型优化处理后的结果对比图,如图 7 所示。



(a) 2016 年 9 月 11 日



(b) 2017 年 1 月 13 日



(c) 2017 年 2 月 17 日

图 7 预测值与实测值结果对比

表 4 不同模型的误差结果

日期	模型	评价指标		
		MSPE	MAPE	RE
2016-09-11	性回归	0.0046	0.0889	0.0713
	BP 模型	0.0045	0.0826	0.0639
	GA-BP	0.0043	0.0792	0.0681
2017-01-13	线性回归	0.1610	1.3041	-0.5323
	BP 模型	0.1834	1.5226	0.6169
	GA-BP	0.1115	1.5048	-1.5548
2017-02-17	线性回归	0.1633	0.8133	0.6958
	BP 模型	0.1707	0.9703	-0.3315
	GA-BP	0.0956	0.7087	0.0622

为更形象直观的比较出优化模型(GA-BP)对太阳辐射预测效果的提高,进行了相比于未优化模型(BP),GA-BP 模型的均方百分比误差(MSPE)和平均绝对误差(MAPE)的减少量统计,见图 8。

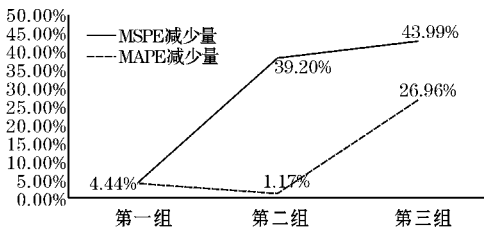


图 8 改进后的模型误差减少量统计图

由表 4 可看出:相比于线性回归和 BP 模型所测量的结果误差,经过 GA-BP 模型改进后的结果明显得到了改善。从图 8 可明显看出:GA-BP 模型相对于 BP 模型的 MSPE 最高减少了 43.99%,最低减少了 4.44%,而 MAPE 最高减少了 26.96%,最低减少了 1.17%,可以看出基于 PCA 的 GA-BP 模型的泛化能力得到了一定程度的提高。

## 5 结束语

利用常规气象数据,建立了将 PCA、GA 和 BP 神经网络混合的优化模型,应用于太阳辐照度的估算,与较单一的 BP 模型估算结果进行比较,结果表明:采用 PCA 简化输入变量的维数,减少了计算量,而遗传算法对模型权值和阈值的优化处理,提高了搜索的速度和效率,同时也增大了达到全局收敛的可能性,与单一的 BP 模型对比发现,优化的 BP 模型的测量精度更高,更适合太阳辐射的预测,所以基于 PCA 的 GA-BP 模型对太阳辐射的估算是可行有效的,具有一定的实用价值。

建立的仿真模型对晴天天气类型进行预测,其预测结果表现较好。但是,预测的结果很大程度上会受

天气类型变化的干扰,在后续研究中,如何高效地识别出不同天气类型,并针对突变天气的预测模型的可用性、有效性的研究将继续深入探讨。

## 参考文献:

- [1] 左大康,王懿贤,陈建绥. 中国地区太阳总辐射的空间分布特征[J]. 气象学报,1963,33(1):78-95.
- [2] 曹双华,曹家枏. 太阳逐时总辐射混沌优化神经网络预测模型研究[J]. 太阳能学报,2006,27(2):164-169.
- [3] 田景文,高美娟. 人工神经网络算法研究及应用[M]. 北京:北京理工大学出版社,2006.
- [4] 周晋,吴业正,晏刚,等. 利用神经网络估算太阳辐射[J]. 太阳能学报,2005,26(4):509-512.
- [5] 贺文,齐爽,陈厚合. 蚁群BP神经网络的光伏电站辐照强度预测[J]. 电力系统及其自动化学报,2016,28(7):26-31.
- [6] 夏玫. BP神经网络泛化能力改进研究[D]. 太原:太原科技大学,2009.
- [7] 苏高利,柳钦火. 基于LS-SVM方法的晴空逐时太阳辐射模型[J]. 北京师范大学学报(自然科学版)2007,43(3):274-278.
- [8] 王炳忠,张富国,李立贤. 我国太阳能资源极其计算[J]. 太阳能学报,1980,1(1):1-9.
- [9] 邢杰,萧德云. 基于PCA的概率神经网络结构优化[J]. 清华大学学报(自然科学版),2008,48(1):141-144.
- [10] 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.
- [11] 周文全. BP网络泛化能力的增强改进[D]. 广州:暨南大学,2012.
- [12] 廖东进,刘晓龙. 基于BP神经网络的光伏电站辐照度预测研究[J]. 太阳能,2016(10).
- [13] 米增强. 光伏电站辐照度ANN预测及其两维变尺度修正方法[J]. 太阳能学报,2013,34(2):251-258.
- [14] 孙妮妮. BP神经网络算法改进及其应用研究[D]. 重庆:重庆大学,2009.
- [15] 司杨. 基于模式匹配的地面接受太阳辐射功率预测方法研究[J]. 青海大学学报(自然科学版),2012,30(1):30-34.

## Research on Solar Radiation Observation Algorithm based on PCA

ZHENG Dan<sup>1</sup>, MA Shang-chang<sup>1,2</sup>, ZHANG Su-juan<sup>1,2</sup>

(1. College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China; 2. The Key Laboratory of China Meteorological Administration, Chengdu 610225, China)

**Abstract:** The study of solar radiation has an important impact on the understanding of climate change, while the solar radiation observation data of ground stations in China is insufficient and of poor quality. In order to improve the accuracy of solar radiation observation, this paper proposes a method of principal component analysis (PCA) to reduce the number of meteorological factors that influence the solar radiation, which eliminate redundant variables. Combining with the genetic algorithm (GA) to obtain the optimal weights and thresholds of BP model, and using the measured data to verify the feasibility of the model, it can fully improve the network performance to achieve the observation of solar radiation. The results show that the accuracy of GA-BP network model based on PCA is higher than traditional BP model, which can improve the generalization ability of the model and has certain feasibility and guidance significance.

**Keywords:** solar radiation; principal component analysis (PCA); genetic algorithm (GA); BP neural network (BPNN); error analysis