

基于广义回归神经网络的日总辐射曝辐量预估

庄述鹏¹, 宫 响^{1,2}, 林 婵³, 张淑华¹

(1. 青岛科技大学数理学院, 青岛 266061; 2. 中国海洋大学环境科学与工程学院, 青岛 266100;

3. 山东电力工程咨询院有限公司, 济南 250100)

摘 要: 采用广义回归神经网络(generalized regression neural network, GRNN)模型对山东烟台市福山气象站2000~2003年日总辐射曝辐量进行预估。模型通过交叉验证方法确定其关键参数(光滑因子),以日照时数、平均气压、平均气温、日最高气温、相对湿度、气溶胶光学厚度6个变量作为输入量。结果显示:GRNN网络的平均百分比误差为15.9%,均方根误差为2.32 MJ/m²,拟合优度为0.892,且模型的预估精度和拟合优度均明显优于LM-BP网络。气溶胶光学厚度对GRNN网络预估结果的准确度及拟合优度影响不大。对于无太阳辐射观测的地区,GRNN网络模型是预估当地日总辐射曝辐量的一种有效方法。

关键词: 人工神经网络; LM-BP网络; 太阳辐射; 预估

中图分类号: TK51

文献标识码: A

0 引 言

太阳能资源的有效开发利用是缓解大气污染的重要途径。一般以到达地面的日总辐射曝辐量反映太阳能资源丰富程度。中国太阳辐射观测站点较少且分布不均,很多地区的太阳能资源只能通过临近观测站点的辐射资料进行衡量评价,严重制约太阳能资源的开发利用。

影响太阳辐射的因素众多,主要包括:经纬度、海拔高度、日照时数、气温、气压、大气污染程度等,这使得日总辐射曝辐量序列既有确定性又包含随机性^[1]。人工神经网络模型具有非线性映射能力强等优点,近年来已成为预估太阳辐射资源的重要方法。周 晋等^[2]建立了北京日太阳总辐射月均值的BP神经网络预估模型;朱良山等^[1]和李 净等^[3]分别采用LM-BP神经网络模拟中国澳门的日辐射和西北地区的月平均辐射,模型预估误差较小,取得较好的预估结果。

径向基函数神经网络采用局部非线性映射函数,具有唯一最佳逼近的特性,且无局部极小。广义回归神经网络(GRNN)是径向基网络的一种,在

逼近能力和学习速度上较一般径向基网络更有优势,在交通、时间序列、互联网等众多领域显示出其优越性^[4-9]。本文采用GRNN网络模型对2000年~2003年山东省福山辐射站点的日总辐射曝辐量进行训练与验证,最后对比GRNN网络与LM-BP网络的预估效果。

1 数据及方法

1.1 研究区域概况

山东省太阳能资源较为丰富,年总辐射为4480~5800 MJ/m²,为Ⅱ类区(丰富区)。山东省位于34°15'N~38°13'N、114°21'E~122°25'E,南北最大长度约420 km,东西最大宽度约700 km,境内有沿海、平原、丘陵、山地等多种地形,使之太阳辐射的差异较大。山东省现在有3个太阳辐射观测站,分别为福山、莒县和济南。本文重点分析预估烟台市福山站日总辐射曝辐量数据。

1.2 数据处理

本文所用地面气象站的气象观测资料及太阳辐射资料均来自中国气象数据网(<http://data.cma.cn/>)。

收稿日期: 2016-08-29

基金项目: 国家自然科学基金(41406010)

通信作者: 宫 响(1977—),女,博士、讲师,主要从事海洋生态环境动力学及神经网络应用研究。gongxiang@qust.edu.cn

所用气象和辐射资料选用山东福山站 2000 年~2003 年观测数据,共 1461 组数据,每组数据包括日总辐射曝辐量、平均风速、最大风速、日照时数、平均气压、日最高气压、日最低气压、平均气温、日最高气温、平均相对湿度、最小相对湿度和降水量。

此外,城市大气污染程度指标选用气溶胶光学厚度(aerosol optical depth,AOD),采用 MODIS 遥感数据 (<https://ladsweb.nascom.nasa.gov/data/search.html>)。由于 MODIS 的 AOD 数据始于 2000 年 3 月,故将 2001 年~2003 年 1 月、2 月数据取平均值代替缺失的 2000 年 1 月、2 月数据。2000 年~2003 年其他缺失的 AOD 数据采用局部线性插值法补充,主要调用了 Matlab 中 Piecewise linear 和 Spline 函数。

1.3 研究方法

本文采用 GRNN 神经网络模型,预估福山站的日总辐射曝辐量,并比较测试样本预估值与真值,采用平均百分比误差 *MPE*(mean percentage error)和均方根误差 *RMSE*(root mean square error)这 2 个评价指标对模型结果进行评价分析。

1.3.1 模型介绍

GRNN 是由 The Lockheed Palo Alto 研究实验室的 Donald F. Specht 在 1991 年提出的一种新型的基于非线性回归理论的神经网络模型,具体的公式推导和理论可参考文献[10~12]。GRNN 由输入层、模式层、求和层和输出层等 4 层构成,对应网络输入量 $X=[x_1, x_2, \dots, x_n]^T$,其输出量为 $Y=[y_1, y_2, \dots, y_m]^T$,如图 1 所示。其中,输入层神经元数目等于输入向量的维数,各神经元是简单的分布单元,直接将输入量传递给模式层。模式层为径向基层,神经元个数等于学习样本数目 n , P_i (图 1)表示神经元传递函数,即径向基函数,通常采用高斯函数:

$$P_i = \exp[-(X - x_i)^T(X - x_i)/2\sigma^2], i = 1, 2, \dots, n \quad (1)$$

式中, X ——输入量; T ——矩阵转置; σ ——光滑因子,光滑因子越小,则函数对样本的逼近能力越强。

求和层中使用 2 类神经元计算公式进行求和。一类是对所有模式层神经元输出进行算术和,模式层与各神经元链接权值为 1,传递函数为:

$$S_D = \sum P_i \quad (2)$$

另一类计算公式是对所有模式层神经元的输出进行加权求和,模式层中第 i 个神经元与求和层中第 j 个分子求和神经元之间的链接权值为第 i 个输出样本 Y_i 中的第 j 个元素,传递函数为:

$$S_{Nj} = \sum y_{ij} P_i, j = 1, 2, \dots, m \quad (3)$$

输出层中的神经元数目等于学习样本中输出向量的维数 n ,各神经元将求和层的输出相除,神经元 j 的输出对应估计结果 $Y(X)$ 的第 j 个元素,即:

$$y_j = S_{Nj}/S_D, j = 1, 2, \dots, m \quad (4)$$

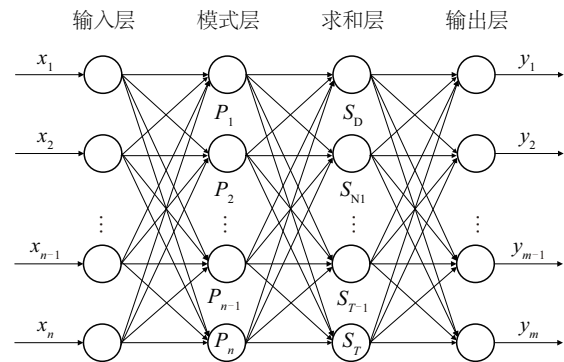


图1 广义回归神经网络 (GRNN) 模型结构图

Fig.1 Structure of generalized regression neural network (GRNN) model

1.3.2 光滑因子 (SPREAD) 确定方法

通过对 GRNN 网络的模型介绍可以看出,训练 GRNN 网络的关键是确定最优化光滑因子 σ ,本文采用了交叉验证算法来确定最优光滑因子。关于交叉验证的具体细节、相似度和稳定值的计算方法可参考文献[13]。本文采用四折交叉验证算法对数据进行训练,并记录每次训练得到的均方根误差,得到均方根误差趋于最小时所对应的最优光滑因子。具体方法为:

1) 选定光滑因子取值范围。*SPREAD* 值从 0.01 到 1 之间循环选择,每次循环间隔为 0.05。

2) 四折交叉验证。假设总训练集为 $A(X, Y)$, X 、 Y 分别为输入量和对应的输出量,将 $A(X, Y)$ 随机划分为 4 个相等大小的子集 $B(X, Y)$ 、 $C(X, Y)$ 、 $D(X, Y)$ 、 $E(X, Y)$ 。从这 4 个子集中任选 3 个,余下的一个作为测试集,利用选定的光滑因子,进行 GRNN 网络训练验证。

3) 计算均方根误差。重复进行 1) 和 2), 计算得

到每次 GRNN 网络预估值与真值间的均方根误差。

4) 确定最优光滑因子的取值。

1.3.3 模型建立

在输入量的选择上可参考文献[1~3],并结合相关性分析结果确定。模型输出量为日总辐射曝辐量。由于只针对一个观测站数据进行实验分析,所以拟输入量排除了经度、纬度、海拔高度3个

变量。根据拟输入量与输出量的相关性分析结果(表1),可知日照时数、气压、气温、相对湿度对日总辐射影响相对较大,而风速及降水量对日总辐射影响较小,因此选取日照时数、平均气压、平均气温、日最高气温、平均相对湿度作为输入量。此外,考虑到近年来城市大气污染程度的加剧,将AOD作为输入量之一。

表1 日总辐射曝辐量与影响因子之间的相关系数表

Table 1 Correlation coefficient between daily irradiation exposure of global radiation and impact factor

| 影响因子 | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | X_9 | X_{10} | X_{11} | X_{12} |
|------|--------|--------|---------|----------|----------|----------|---------|---------|-------|----------|----------|----------|
| 相关系数 | -0.004 | 0.053* | 0.874** | -0.473** | -0.500** | -0.468** | 0.530** | 0.579** | 0.007 | -0.276** | -0.342** | -0.131** |

注: X_1 —平均风速; X_2 —最大风速; X_3 —日照时数; X_4 —平均气压; X_5 —日最高气温; X_6 —日最低气温; X_7 —平均气温; X_8 —日最高气温; X_9 —AOD; X_{10} —平均相对湿度; X_{11} —最小相对湿度; X_{12} —降水量;**表示 $P<0.01$;*表示 $P<0.05$ 。

通过交叉验证方法对2000年~2003年福山气象站的1461组数据进行训练,确定最优光滑因子,建立效果最稳定的GRNN网络。利用该网络进行了120d的预估,即训练数据采用前1341组数据得出模型参数,验证数据采用后120组数据中已有的地面因子和AOD因子数据作为输入量,预估出输出量日总辐射曝辐量。

2 结果与分析

通过交叉验证法,求得不同光滑因子取值下GRNN网络预估值与真值的均方根误差,结果如图2所示。均方根误差能够反映出估值灵敏度及模型短期拟合信息,误差越小,模型灵敏度越高。当光滑因子值小于0.11,均方根误差会变大(图2);光滑因子值大于0.11时,虽然均方根误差不再改变(图2),但光滑因子值增大,径向基函数对样本的逼近能力变弱,且会导致数值计算上的困难。因此,本文GRNN网络取到的最优光滑因子为0.11,此时的GRNN网络模型最佳。

本文采用平均百分比误差和均方根误差2个指标对GRNN神经网络模型的预估结果进行误差评价与分析。模型验证结果如图3和图4所示。福山站GRNN网络模型预估效果较好,拟合优度 $R^2=0.892$,网络预估值与真值间的差值较小,范围在 $-8\sim 8\text{ MJ/m}^2$ (图3)。百分比误差多在13%以下(图4),平均百分比误差为15.9%,均方根误差为

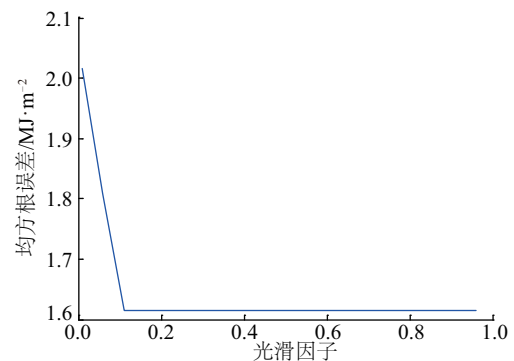


图2 光滑因子对均方根误差的影响

Fig. 2 Influences of smooth factor on the root mean square error of model

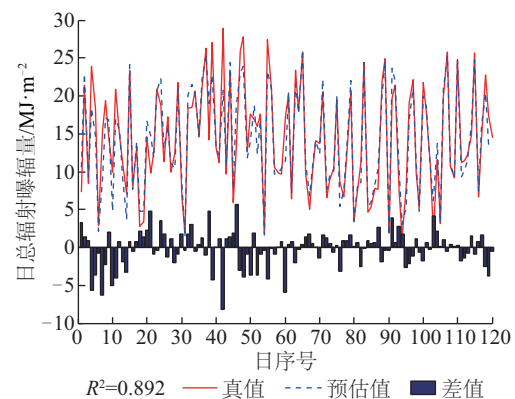


图3 福山站GRNN网络预估值与真值的对比
(差值为预估值减真值)

Fig. 3 Comparison of estimate values and real values of GRNN at Fushan meteorological station (Difference is that estimate value minus real value)

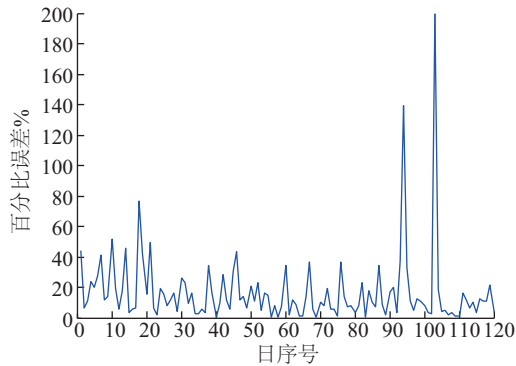


图4 福山站 GRNN 网络预估百分比误差

Fig. 4 Percentage error of the GRNN's estimate at Fushan meteorological station

2.32 MJ/m²(表 2)。个别日期百分比误差超过 100%(图 4),分析原数据发现,误差最大日期的平均气温在这一天变成零下温度,前后几天均为正值,表明气温对 GRNN 网络预估值影响较大。

表2 GRNN 模型与 LM-BP 模型误差统计指标

Table2 Error statistical indicators between GRNN model and LM-BP model

| 模型 | MPE/% | RMSE/MJ·m ⁻² |
|-------|-------|-------------------------|
| GRNN | 15.9 | 2.32 |
| LM-BP | 26.2 | 6.75 |

3 讨 论

3.1 LM-BP 网络预估结果对比

为检测 GRNN 神经网络的预估性能,与应用比较广泛的 LM-BP 网络的预估结果做了对比。LM 优化算法的网络训练参数采用 Matlab 工具箱中 LM 训练函数,隐含层神经元个数为 35,隐层数取 2 层。结果显示,LM-BP 网络训练时间为 21 s,与其相比,GRNN 网络训练时间较短(10 s)。

LM-BP 网络的预估结果如图 5 和图 6 所示。拟合优度 $R^2=0.487$,网络预估值与真值间的差值较小,范围在 -20 ~ 17 MJ/m²(图 5)。百分比误差多在 25% 以下(图 6),较 GRNN 网络预估结果,LM-BP 网络的预估值与真值间差值变大,最大达 20 MJ/m²,拟合优度变小($R^2=0.487$),拟合性能降低 40%。

福山站 LM-BP 网络的各项统计误差及拟合程度与 GRNN 网络的对比结果见表 2。福山站 LM-

BP 网络的平均百分比误差为 26.2%,均方根误差为 6.75 MJ/m²(表 2),均较 GRNN 网络预估结果误差显著偏大。综合以上可见,GRNN 网络比 LM-BP 网络的训练时间短,预测精度高,且表现出更好的拟合性。

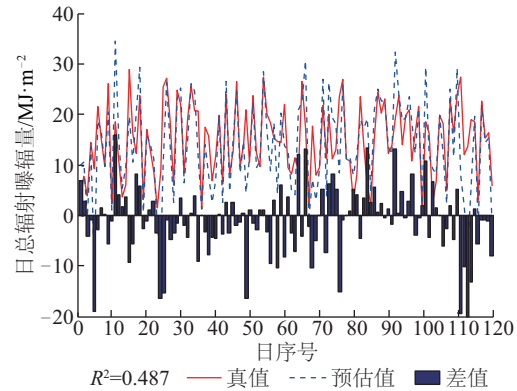
图5 福山站 LM-BP 网络预估值与真值的对比
(差值为预估值减真值)

Fig. 5 Comparison of estimate values and real values of the LM-BP Network at Fushan meteorological station (Difference is that estimate value minus real value)

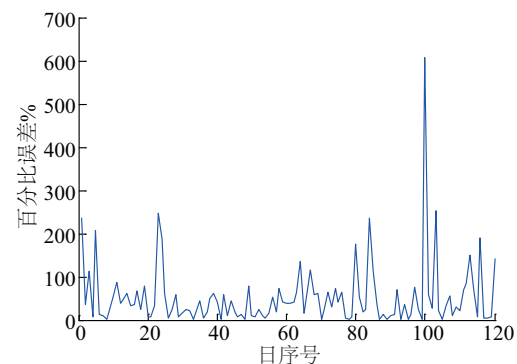


图6 福山站 LM-BP 网络预估百分比误差

Fig. 6 Percentage error of the LM-BP Network's estimate at Fushan meteorological station

3.2 大气污染程度对太阳辐射的影响

为分析城市大气污染程度对太阳辐射的影响,利用 GRNN 网络模型进行敏感试验。GRNN 网络模型中,不考虑输入量 AOD,首先经四折交叉验证,得到最优光滑因子为 0.16,这表明,相对原 GRNN 网络模型(包含 AOD 输入量),该网络模型对样本的逼近能力变弱。

不考虑 AOD 的影响,GRNN 网络预估结果的准确度及均方根误差略有降低($MPE=17.2\%$, $RMSE=$

2.41 MJ/m²),但拟合优度基本不变($R^2=0.885$),表明烟台市大气污染程度对太阳辐射影响不大,这可能与烟台市地处沿海,空气状况良好有关。

4 结 论

通过 GRNN 网络对烟台市福山气象站日总辐射曝辐量的预估研究表明,GRNN 网络的预估精度和拟合度明显优于 LM-BP 网络,适应于山东福山站的日总辐射曝辐量的模拟。GRNN 网络调整的参数较少,只有一个光滑因子参数,而 LM-BP 网络的隐含层神经元个数、隐层层数都需要人为确定;且 GRNN 网络训练时间比 LM-BP 网络短。因此 GRNN 网络是一种很有发展前景的日总辐射曝辐量模拟方法,可将其应用在无辐射观测地区的日总辐射曝辐量模拟。城市大气污染程度对烟台市日总辐射曝辐量有一定影响,其对重污染地区日总辐射曝辐量的影响有待进一步研究。

[参考文献]

- [1] 朱良山,林 婵,王起峰. LM_BP神经网络在总辐射日曝辐量计算中的应用研究[J]. 太阳能学报, 2013, 34(7): 1202—1205.
- [1] Zhu Liangshan, Lin Chan, Wang Qifeng. Applications of LM_BP network in daily irradiation exposure of global radiation calculate[J]. Acta Energaie Solaris Sinica, 2013, 34(7): 1202—1205.
- [2] 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, 24(4): 509—512.
- [2] Zhou Jin, Wu Yezheng, Yan Gang, et al. Solar radiation estimation using artificial neural networks[J]. Acta Energaie Solaris Sinica, 2005, 24(4): 509—512.
- [3] 李 净,王卫东,张福存. 利用LM-BP神经网络估算西北地区太阳辐射[J]. 干旱区地理, 2015, 38(3): 439—445.
- [3] Li Jing, Wang Weidong, Zhang Fucun. Simulation of solar radiation in northwest China based on LM-BP neural network[J]. Arid Land Geography, 2015, 38(3): 439—445.
- [4] 伊良忠,章 超,裴 峥. 广义回归神经网络的改进及在交通预测中的应用[J]. 山东大学学报: 工学版, 2013, 43(1): 9—14.
- [4] Yi Liangzhong, Zhang Chao, Pei Zheng. A modified general regression neural network with its application in traffic prediction[J]. Journal of Shandong University: Engineering Science, 2013, 43(1): 9—14.
- [5] 郑含博. 电力变压器状态评估及故障诊断方法研究[D]. 重庆: 重庆大学电气工程学院, 2012.
- [5] Zheng Hanbo. Study on condition assessment and fault diagnosis approaches for power transformers [D]. Chongqing: College of Electrical Engineering of Chongqing University, 2012.
- [6] 周建萍,闫澍旺. 广义回归神经网络预测加筋土支挡结构高度[J]. 岩土力学, 2002, 23(4): 486—490.
- [6] Zhou Jianping, Yan Shuwang. Generalized regression neural networks for predicting design height of GRW [J]. Rock and Soil Mechanics, 2002, 23(4): 486—490.
- [7] 冯志鹏,宋希庚,薛冬新. 基于广义回归神经网络的时间序列预测研究[J]. 振动、测试与诊断, 2003, 23(2): 105—109.
- [7] Feng Zhipeng, Song Xigeng, Xue Dongxin. General regression neural network based prediction of time series [J]. Journal of Vibration, Measurement & Diagnosis, 2003, 23(2): 105—109.
- [8] 卓 莹,张 强,龚正虎. 网络态势预测的广义回归神经网络模型[J]. 解放军理工大学学报: 自然科学版, 2012, 13(2): 148—151.
- [8] Zhuo Ying, Zhang Qiang, Gong Zhenghu. GRNN model of network situation forecast[J]. Journal of PLA University of Science and Technology: Natural Science Edition, 2012, 13(2): 148—151.
- [9] 王菲露,宋 杨. 基于广义回归神经网络的蛋白质二级结构预测[J]. 计算机仿真, 2012, 29(2): 184—187.
- [9] Wang Feilu, Song Yang. Protein secondary structure prediction based on generalized regression neural network[J]. Computer Simulation, 2012, 29(2): 184—187.
- [10] 张德丰. MATLAB神经网络应用设计[M]. 北京: 机械工业出版社, 2009.
- [10] Zhang Defeng. Application design of MATLAB neural network [M]. Beijing: Mechanical Industry Press, 2009.
- [11] 董长虹. MATLAB神经网络与应用[M]. 北京: 国防工业出版社, 2005.
- [11] Dong Changhong. Neural network and application of MATLAB [M]. Beijing: National Defense Industey Press, 2005.
- [12] MATLAB中文论坛. MATLAB神经网络30个案例分析[M]. 北京: 北京航空航天大学出版社, 2010.

- [12] Matlab Chinese BBS. 30 cases analysis of MATLAB neural network [M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2010.
- [13] 邱龙金, 贺昌政. 神经网络稳定性的交叉验证模型 [J]. 计算机工程与应用, 2010, 46(34): 43—45.
- [13] Qiu Longjin, He Changzheng. Cross validation model for neural network stability [J]. Computer Engineering and Applications, 2010, 46(34): 43—45.

ESTIMATE OF DAILY IRRADIATION EXPOSURE OF GLOBAL RADIATION USING GENERALIZED REGRESSION NEURAL NETWORK

Zhuang Shupeng¹, Gong Xiang^{1,2}, Lin Chan³, Zhang Shuhua¹

(1. School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China;

2. College of Environmental Science and Technology, Ocean University of China, Qingdao 266100, China;

3. Shandong Electric Power Engineering Consulting Institute Corp., Ltd., Ji'nan 250100, China)

Abstract: The generalized regression neural network (GRNN) model is used to estimate the daily irradiation exposure of global radiation from 2000 to 2003 at Fushan Meteorological Station, Yantai city, Shandong province. The cross validation method is adopted to determine the key parameter of GRNN model (smooth factor). The input parameters of GRNN model included sunshine duration, average pressure, average air temperature, daily maximum air temperature, relative humidity, and aerosol optical thickness at Fushan meteorological station. Results are promising with *MPE* is 15.9%, *RMSE* is 2.32 MJ/m², and the correlation coefficient is 0.892. The optimized GRNN presents the estimate better than LM-BP network. Aerosol optical thickness in the model has almost no influence on both *MPE* and *RMSE*. Therefore, using the GRNN model with the meteorological observation data is a very effective method to estimate the daily irradiation exposure of global radiation of some region which has not radiation observation site.

Keywords: neural networks; LM-BP networks; solar radiation; estimation