

自适应遗传算子优化BP神经网络的气温预测

窦柯宁, 孙小淇

青岛大学数学与统计学院, 山东 青岛

Email: 2019020136@qdu.edu.cn, sunxiaopi@live.com

收稿日期: 2020年12月25日; 录用日期: 2021年1月19日; 发布日期: 2021年1月28日

摘要

气温变化是由多种非线性因素引起的, 传统的数值气温预测模式由于其繁琐的步骤和较低的预测精度已无法满足人们的需求。鉴于BP神经网络对非线性的拟合有很强的映射能力, 因此考虑用BP神经网络对气温进行预测, 而滨海地区影响气温变化的因素更加复杂, 因此在BP神经网络中加入自适应遗传算子对其阈值、权值以及收敛速度进行优化以此提高气温预测的精度。为进一步提高模型预测精度以及数据处理效率, 文章初始对输入数据做了降维筛选主成分, 归一化等一系列优化处理, 在此基础上建立了BP神经网络以及加入自适应算子的BP神经网络两种气温预测模型。之后用MATLAB软件对山东省日照市的气温相关数据进行仿真, 对比BP和GA-BP模型的预测结果, 改进后模型预测的准确度更高, 以此证明了优化算法的有效性。

关键词

BP神经网络, 自适应遗传算子, 气温预测, 优化

Adaptive Genetic Operator Optimizes Temperature Prediction of BP Neural Network

Kening Dou, Xiaoqi Sun

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Email: 2019020136@qdu.edu.cn, sunxiaopi@live.com

Received: Dec. 25th, 2020; accepted: Jan. 19th, 2021; published: Jan. 28th, 2021

Abstract

The temperature change is caused by many nonlinear factors, and the traditional numerical tem-

perature prediction model cannot meet people's demands due to its complicated steps and low prediction accuracy. Given the BP neural network in nonlinear fitting has a strong ability to map, so consider using the BP neural network to forecast the temperature, and the influencing factors of temperature change are more complicated in the coastal area, thus to join in the BP neural network adaptive genetic operators, weights and threshold value to its convergence speed is optimized to improve the accuracy of the temperature prediction. To further improve the prediction accuracy and data processing efficiency of the model, a series of optimization processes, such as normalization of input data, dimension reduction, and principal component screening, was initially carried out in this paper. On this basis, two temperature prediction models, BP neural network, and BP neural network with the adaptive operator were established. Then MATLAB was used to simulate the air temperature-related data of Rizhao city and the prediction results of BP and GA-BP models were compared. The improved model had a higher prediction accuracy, thus proving the effectiveness of the optimization algorithm.

Keywords

BP Neural Network, Adaptive Genetic Operator, Temperature Prediction, Optimize

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人工神经网络(Artificial Neural Networks, ANNs)由大量相互连接的神经元构成,通过控制神经元间的连接强度对信息进行处理。将BP算法引入ANNs得到的BP神经网络在气象方面应用较晚,与传统数值预报模式相比,BP神经网络应用于天气预测,降低了实施过程的复杂程度,减少了数据的限制条件,提高了预测的准确率,因此更加符合人们的需求。曹兴锋[1]建立了一个前馈性BP神经网络,采用变梯度的方法对长江流域导致小麦赤霉病的气象进行了相关分析,得到了有效准确的预测结果,对控制该地区小麦赤霉病发病提供了理论指导。又有部分学者在BP模型的基础上,通过改进其输入层输入因子的方式提高BP模型的预测精度。例如:作者[2]在BP模型输入层中加入划分算法将降雨径流分为汛期和非汛期,改进后的模型对降雨径流量的预测准确率明显提高。

神经网络的训练是一个复杂的过程,并且随着网络权值以及训练样本集的变化,会导致网络训练结果随机性较大且容易陷入局部极值,因此部分学者开始尝试针对BP神经网络的缺点对其进行改进并运用于各个领域的预测。例如:艾洪福[3]在BP神经网络模型中加入了遗传算法建立了GA-BP预测模型,并对长春市PM2.5含量进行仿真;陈笑[4]建立了遗传算法优化的BP神经网络模型对地下水埋深进行模拟和预测,并使用RMSE、NSE、MAPE三种评价指标进行模型评估,最后以蒙城县为例进行仿真;袁俊杰[5]在其文章中建立了BP、GA-BP、PSO-BP三种模型对平流层风场风速进行预测,结果表明优化后的BP神经网络预测效果优于单一BP模型预测,并且加入粒子群算法的优化效果略好于加入遗传算法;谢梦蝶[6]运用遗传算法优化BP神经网络建立了股指预测模型,经实证研究表明优化后的GA-BP模型预测误差小于BP模型并且其收敛速度也有所提高。赵智[7]运用GA-BP神经网络对某地PM2.5的值进行预测仿真;杨云[8]运用GA-BP神经网络对供暖期的空气指数进行预测分析;张鹏达[9]运用GA-BP、BP神经网络分别对城市的空气质量进行预测仿真,以上文献均说明加入遗传算法或者其他改进算法降低BP神经网络模型预测误差是可行且有效的。

本文建立了一个自适应遗传算子优化的BP神经网络的气温预测模型,主要针对BP神经网络对阈值、权值敏感的缺点,与此同时提高其收敛速度。然后分别用BP和GA-BP种模型对日照市近年来的气温进行模拟训练及预测,分析对比其结果,显然加入自适应算子优化BP神经网络降低了单一BP神经网络模型的预测误差,具有实际应用价值,可以用于长期天气预报的气象需求,对于预防某些气候灾害提供参考。

2. 基本理论

2.1. BP 神经网络

BP (back propagation)神经网络[10]是ANNs中运用较广泛的一种网络模型,其数学模型如下图1:

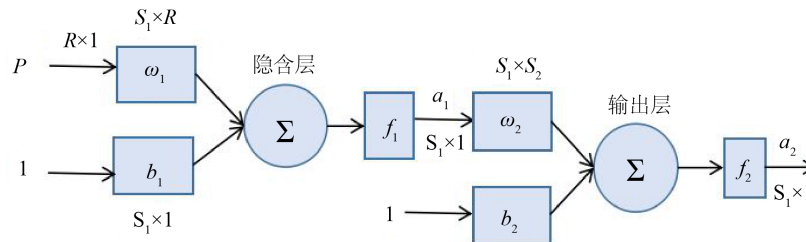


Figure 1. BP neural network mathematical model
图 1. BP 神经网络数学模型

具体算法过程如下:

1) 样本数据正向传输,经过输入层后加权输送给隐含层,然后隐含层各神经元汇总所有的输入,再通过转移函数产生某种输出,并通过下一层连接权重加权由输出层输出。

BP神经网络隐含层中第*i*个神经元的输出为:

$$a_{1i} = f_1 \left(\sum_{j=1}^n \omega_{1ij} p_j + b_{1i} \right), i = 1, 2, \dots, s_1$$

输出层中第*k*个神经元的输出为:

$$a_{2k} = f_2 \left(\sum_{i=1}^{s_1} \omega_{2ki} a_{1i} + b_{2k} \right), k = 1, 2, \dots, s_2$$

式中,*n*为输入层节点数,式中,*n*为输入层节点数,*p_j*为输入层第*j*个神经元的输入,*a_{1i}*为隐含层含层节点*i*的输出,*a_{2i}*为输出层节点*k*的输出,*w_{1ij}*为节点*i*和节点*j*的连接权值,*w_{2kj}*为节点*i*和节点*k*的连接权值,*b_{1i}*为隐含层节点*i*的阈值,*b_{2k}*为输出层节点*k*的阈值,*f₁(.)*和*f₂(.)*分别为隐含层和输出层的传递函数。

2) 比较期望输出与实际输出之间的误差。

误差函数为:

$$e = \frac{1}{2} \sum_{p=1}^q [d_o(k) - y_{o_o}(k)]^2$$

3) 误差反传阶段,需将误差送回,反向依次调整各连接层之间的权重并反复学习训练,一直循环直到输出层输出相对满意的结果。

2.2. 遗传算法理论

遗传算法(Genetic Algorithm, GA) [11]是建立在自然选择和基因遗传学原理的基础上发展而形成的

一种进化算法。其基本思路为根据待优化问题的目标函数设计一个适应度函数 $E(w)$ 然后对优化问题的参数按照一定的规则进行编码得到初始种群, 接着对种群进行评价、遗传操作。这样经过多代进化, 得到适应度最好的个体作为问题的最优解, 其中适应度值的好坏评估使用神经网络的实际输出值与期望输出值之间的均方差(MSE)。函数公式如下:

$$E(w) = N^{-1} \sum_{i=1}^N (t_i - a_i)^2$$

$$MSE = N^{-1} \sum_{i=1}^N (e_i)^2 = N^{-1} \sum_{i=1}^N (t_i - a_i)^2$$

式中 t_i 为第 i 个训练样本的网络输出值, a_i 为相应的期望输出值, N 为训练集的规模。

3. 预测模型算法设计及气温预测模型建立

3.1. 预测模型算法设计

BP 神经网络训练学习的过程实际是一个通过不断调整其参数而降低输出误差的过程, 加入自适应遗传算子优化意在进一步降低输出误差。故遗传算法优化神经网络的主要步骤分设计 BP 神经网络的结构, 设计 GA 的参数, BP 模型预测三步。其算法[12]流程如下图 2。

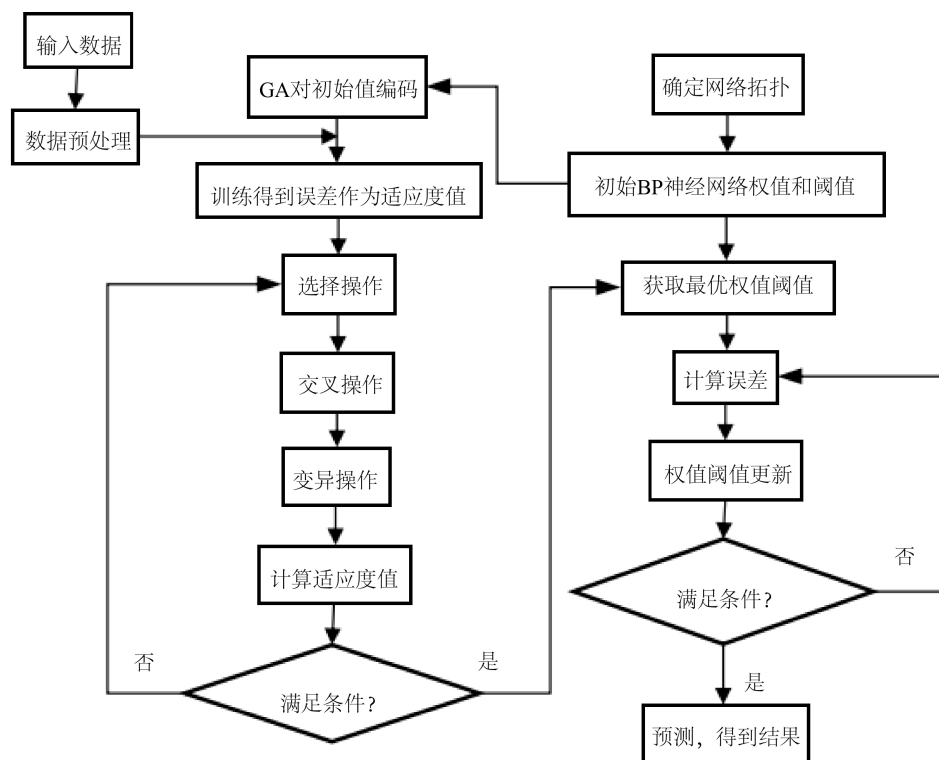


Figure 2. Genetic algorithm optimization of BP neural network algorithm program diagram
图 2. 优化 GA-BP 神经网络算法程序图

- 1) 种群初始化。个体编码中运用实数编码, 设置训练进化代数以及迭代次数等参数, 随机产生初始种群。
- 2) 获取个体评价函数。其中评价标准由 MSE 确定。
- 3) 选择运算。按照个体适应度值, 采用轮盘赌法。

4) 交叉运算。

自适应算子位置交叉:

$$\begin{cases} X_i(k+1) = \alpha_1 \times X_i(k) + (1-\alpha_1)X_j(k) \\ X_j(k+1) = \alpha_1 \times X_j(k) + (1-\alpha_1)X_i(k) \end{cases}$$

自适应算子速度交叉:

$$\begin{cases} V_i(k+1) = \alpha_2 \times V_i(k) + (1-\alpha_2)V_j(k) \\ V_j(k+1) = \alpha_2 \times V_j(k) + (1-\alpha_2)V_i(k) \end{cases}$$

5) 变异运算。

自适应算子位置变异:

$$X_i(k+1) = \begin{cases} X_i(k) + (X_i(k) - X_{\max}) * f(g) & r_1 \geq 0.5 \\ X_i(k) + (X_{\min} - X_i(k)) * f(g) & r_1 < 0.5 \end{cases}$$

自适应算子速度变异:

$$V_i(k+1) = \begin{cases} V_i(k) + (V_i(k) - V_{\max}) * f(g) & r_2 \geq 0.5 \\ V_i(k) + (V_{\min} - V_i(k)) * f(g) & r_2 < 0.5 \end{cases}$$

3.2. 气温预测模型设计

3.2.1. 数据选择及预处理

日照市位于东经 118°25'~119°39', 北纬 35°04'~36°04'之间, 东临黄海, 西接临沂市, 南与江苏省连云港市毗邻, 北与青岛市、潍坊市接壤, 隔海与日本、韩国相望。地处温带, 年平均气温全 12.7℃, 属典型的暖温带湿润季风区大陆性气候, 四季分明, 冷热季和干湿季的区别明显。本文选取日照市 2000 年 1 月到 2018 年 12 月共 19 年的逐月气象相关数据(来自中国气象网), 在对大量影响因素做主成分分析降维筛选后选择月平均风速、平均海平面气压、降水量、降水天数、平均能见度、最大持续风速、最低温度作为影响因子。MATLAB 仿真过程中, 2000~2015 年气温相关数据用于模型仿真训练, 参与仿真训练的数据充足且特征简单, 故仿真结果不易出现过拟合及欠拟合现象, 2015~2018 年的数据用于仿真预测。文中对输入数据做归一化处理, 最后再对输出数据做反归一化处理以提高神经网络模型的数据处理效率。具体方式如下:

$$x'_i = 2 * (x_i - x_{i\min}) * (x_{i\max} - x_{i\min})^{-1} + (-1)$$

式中, x_i 为输入向量 P 的第 i 维分量值, x'_i 为 x_i 归一化后的分量值, $x_{i\min}$ 为训练样本空间中输入向量第 i 维分量的最小值, $x_{i\max}$ 为最大值; y_i 为输出向量 T 的第 i 维分量值, $y_{i\min}$ 和 $y_{i\max}$ 分别为训练样本中输出空间第 i 维分量的最小值和最大值。

部分归一化数据见表 1。

3.2.2. 隐含层选择

对在神经网络预测模型的构建中, 包括 7 个输入参数(月平均风速、平均海平面气压、月降水量、月降水天数、月平均能见度、最大持续风速、最低温度), 1 个输出参数为月平均气温, 样本数为 228, 种群规模为 60, 进化代数为 60, 训练迭代次数 100, 学习率为 0.1, 动量系数为 0.8, 训练目标误差设置为 0.001。

Table 1. Normalized data

表 1. 归一化数据

	平均风速	海平面气压	降水量	最低温度
2015.1	0.7232	0.5878	0.8568	0.1617
2015.2	0.8221	0.5807	0.7420	0.1936
...
2018.12	0.4533	0.7238	0.9453	0.1694

不同隐含层节点数对应的网络输出误差性能不同, 隐含层的节点数[13]根据经验公式 $n = \sqrt{n_i + n_o} + c$ 确定。式中 n_i 表示输入层节点数; n_o 表示输出层节点数, 取 1 到 10 之间的任意常数, 误差性能由平均方差 MSE 衡量。表 2 为模型在输入层节点数为 7, 输出层节点数为 1 时, 隐含层节点数 n 与 MSE 的关系, 易知, 当隐含层节点数为 10 时输出误差性能最强, 故隐含层节点数取 10。此时 BP 神经网络拓扑结构图如图 3。

Table 2. The number n of hidden layer nodes and the relationship between the MSE

表 2. 隐含层节点数 n 与 MSE 的关系

隐含层节点数 n 与 MSE 的关系									
n	4	5	6	7	8	9	10	11	12
MSE	0.0021	0.0018	0.0016	0.0016	0.0017	0.0016	0.0012	0.0018	0.0017

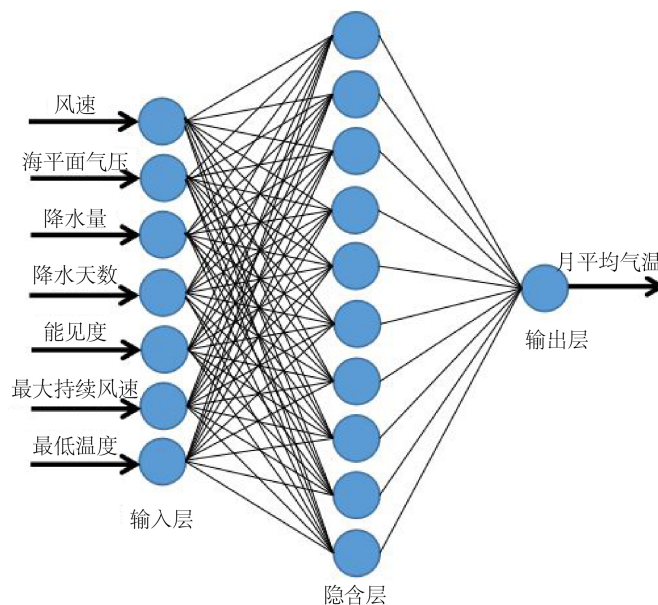


Figure 3. BP network topology

图 3. BP 网络拓扑结构

4. 仿真结果

在 MATLAB 对山东省日照市的气温相关数据仿真过程中, 得到 BP、GA-BP 神经网络训练误差变化曲线(图 4、图 5), 它们分别在迭代次数为 3、2 时输出性能达到最佳, 此时, BP 模型的输出误差为 0.00131, 并未达到期望输出误差, 但加入了自适应遗传算子的 BP 模型的输出误差为 0.00048722, 达到并超出了

期望输出误差。明显自适应遗传算子的加入在提高了神经网络收敛速度的同时改善了模型的预测性能。

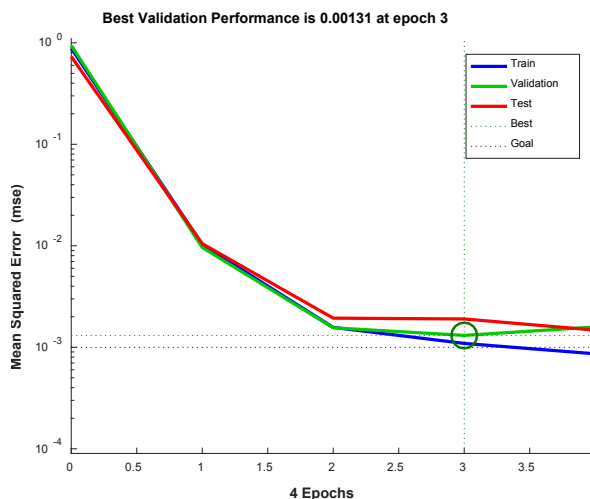


Figure 4. Prediction error performance of BP
图 4. BP 预测误差性能曲线图

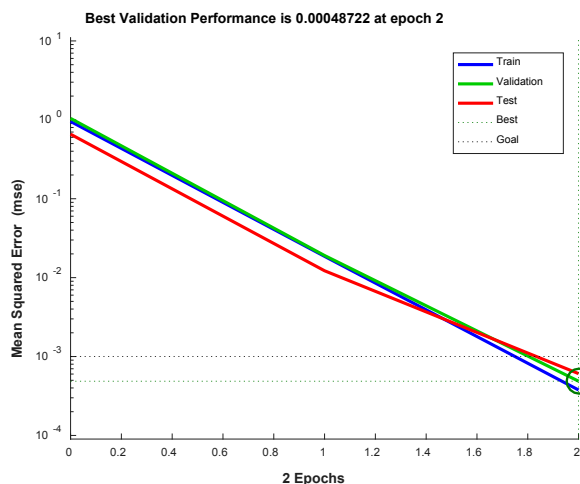


Figure 5. Prediction error performance of GA-BP
图 5. GA-BP 预测误差性能曲线图

图 6、图 7 分别为 BP、GA-BP 神经网络的气温仿真预测，显然，在 GA-BP 模型预测中大多数测试样本的真实值比 BP 模型更接近预测值，也可以说 GA-BP 预测模型的准确度要高于 BP 模型。

图 8，图 9 分别给出了 BP、GA-BP 神经网络的仿真预测误差曲线。由这两个曲线图我们不难看出 BP 模型的预测误差区间为 $[-3.5, 5)$ ，GA-BP 的气温预测误差区间为 $(-1.5, 4)$ ，从而不难分析出 GA-BP 预测误差范围小于 BP 模型，也可以说加入自适应算子的 GA-BP 模型预测效果优于 BP 模型。继续对两种模型做进一步精确评估得到表 3。

由表 3 容易看出 BP、GA-BP 两种模型对日照市 2015~2018 年月平均气温的预测精确程度，这说明加入遗传算法的 BP 神经网络的初始权值和阈值更加合理，并且对神经网络收敛速度慢的缺点有所改善，因此加入自适应算子优化遗传算法使得 BP 模型预测精度逐渐增加，这有力的说明了文中的优化算法是有效且合理的。

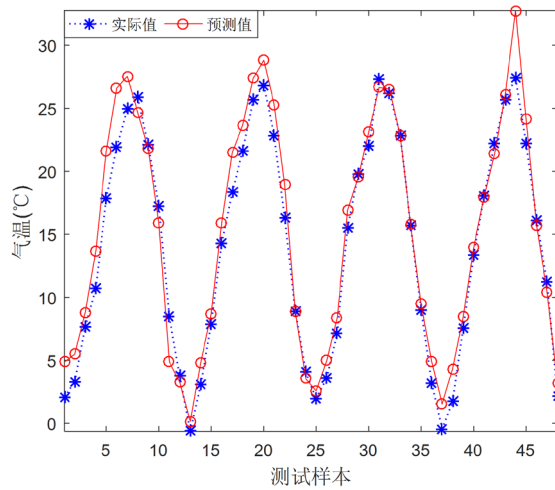


Figure 6. BP model prediction
图 6. BP 模型预测

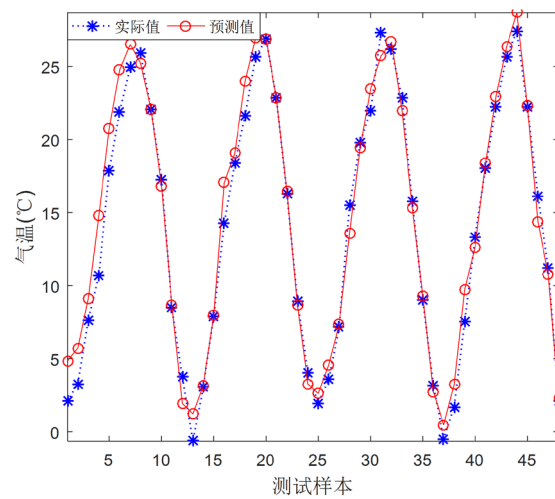


Figure 7. GA-BP model prediction
图 7. GA-BP 模型预测

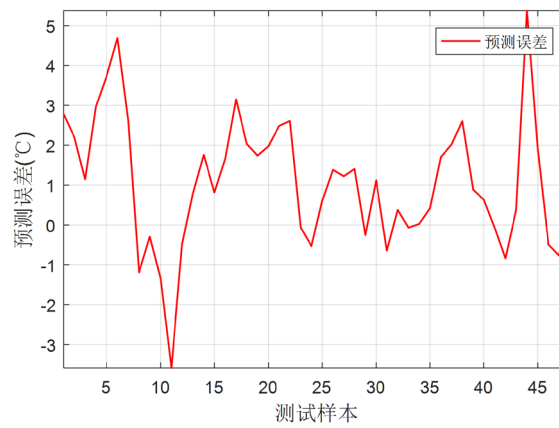


Figure 8. BP model prediction error
图 8. BP 预测误差曲线

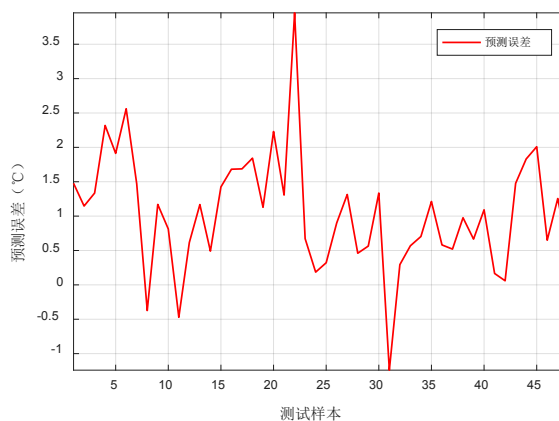


Figure 9. GA-BP model prediction error
图 9. GA-BP 预测误差曲线

Table 3. Evaluation of the prediction results of both models
表 3. 两种模型的预测结果评估

	SSE	RMSE	R ²	MAE	MAPE
BP	118.8880	1.6076	0.9686	1.5177	27.6342
GA-BP	32.8901	0.8456	0.9913	1.1274	50.1075

致 谢

本论文从选题到完成, 从内容到格式, 从标题到标点, 每一步都是在孙小淇导师的悉心指导下完成的。在写作过程中遇到了很多问题, 正是老师的耐心指导, 使得问题都得以解决。在此, 谨向导师表示崇高的敬意和衷心的感谢! 其次, 感谢我的同窗好友刘倩, 正是她的帮助和鼓励, 论文得以顺利完成。感谢我的家人, 他们给予我支持和鼓励, 使我克服困难。最后, 我还要感谢在我学习期间给我极大关心和帮助的同学, 同时也感谢杂志社编辑对我的直接和间接的帮助。

基金项目

山东省自然科学基金(ZR2019PA007)。

参考文献

- [1] 曹兴锋. 一种 BP 神经网络气象预测中的应用研究[J]. 科技通报, 2012, 28(8): 55-57.
- [2] Ju, Q., Yu, Z.B., Hao, Z.C., Ou, G.X., Zhao, J. and Liu, D.D. (2008) Division-Based Rainfall-Runoff Simulations with BP Neural Networks and Xinanjiang Model. *Neurocomputing*, **72**, 2873-2883. <https://doi.org/10.1016/j.neucom.2008.12.032>
- [3] 艾洪福. 基于 GA-BP 神经网络的“雾霾”天气预测研究[J]. 绿色科技, 2019(22): 95-96, 99.
- [4] 陈笑, 王发信, 戚王月, 周婷. 基于遗传算法的 BP 神经网络模型在地下水埋深预测中的应用——以蒙城县为例[J]. 水利水电技术, 2018, 49(4): 1-7.
- [5] 袁俊杰, 罗汝斌, 廖俊, 杨泽川, 王宁, 李珺. 3 种神经网络模型在平流层风场预测中的应用[J]. 控制与信息技术, 2019(5): 12-16.
- [6] 谢梦蝶, 秦江涛. 遗传算法优化 BP 神经网络预测股指研究[J]. 软件导刊, 2019, 18(4): 41-45.
- [7] 赵智. 基于遗传算法优化 BP 神经网络雾霾预测模型的研究[J]. 科技展望, 2015, 25(27): 226+228.
- [8] 杨云, 杨毅. 基于 GA-BP 神经网络的供暖期空气质量指数预测分析[J]. 陕西科技大学学报(自然科学版), 2016, 34(4): 171-176+186.

- [9] 张鹏达. 基于 BP 神经网络的城市环境空气质量预测模型[J]. 自动化技术与应用, 2014, 33(1): 9-11+19.
- [10] 袁冰清, 程功, 郑柳刚. BP 神经网络基本原理[J]. 数字通信世界, 2018(8): 28-29.
- [11] 高文. 基于遗传算法优化的 BP 神经网络对房价预测的研究[D]: [硕士学位论文]. 延安: 延安大学, 2019.
- [12] 朱云霞, 郭唐仪. 基于 GA-BP 神经网络的交叉口短时交通流预测[J]. 交通运输研究, 2019, 5(2): 45-51.
- [13] 成华义. GA-PSO-BP 神经网络在大气污染物浓度预测中的应用研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2014.