文章编号：1000-8608(2018)03-0316-08

基于改进粒子群优化 BP_Adaboost 神经网络的 $PM_{2.5}$ 浓度预测

李晓理*[1,2,3]，梅建想[1]，张　山[1]

（1.北京工业大学 信息学部，北京　100124；
2.计算智能与智能系统北京市重点实验室，北京　100124；
3.数字社区教育部工程研究中心，北京　100124）

摘要： 为了提高大气污染物浓度预测精度，采用灰色关联分析选取影响大气中 $PM_{2.5}$ 浓度的主要因子，并以此作为神经网络输入变量，建立一种基于 BP_Adaboost 神经网络的 $PM_{2.5}$ 浓度预测模型.用改进粒子群算法来选择 BP_Adaboost 神经网络权重和阈值，可以有效避免神经网络在训练时陷入局部最优解.根据北京市海淀区万柳监测站和朝阳区北京工业大学监测点每小时监测的大气污染物浓度和气象条件，分别选择 2014-11-01～2014-11-25 和 2017-07-07～2017-08-06 数据作为实验研究对象.仿真结果表明，在 $PM_{2.5}$ 浓度预测中，相比于 BP_Adaboost、BP 和广义回归神经网络 3 种预测模型，改进粒子群优化 BP_Adaboost 神经网络预测性能更优.

关键词： 灰色关联分析；BP_Adaboost 神经网络；$PM_{2.5}$ 浓度预测模型；改进粒子群算法

中图分类号： X831　　**文献标识码：** A　　**doi：** 10.7511/dllgxb201803013

0　引　言

随着工业化程度不断加大，废气、废渣排放增加，大气污染愈发严重，尤其是颗粒物质（PM）污染，这一问题已引起国家相关部门高度关注.颗粒物质不仅会影响居民生活、出行，破坏生态系统，而且跟人类健康密切相关.在过去几十年间，研究表明，人口每日死亡率增长和各种疾病症状（如哮喘、慢性支气管炎、肺功能下降等）与大气中颗粒物质浓度有显著关系.因此，许多业界人士对如何预测大气污染物中 $PM_{2.5}$ 浓度开展了相应研究，为居民生活、出行提供决策.最初，统计学方法被用于预测空气污染物，Kumar 等用自回归整合移动平均模型预测印度德里郊区每日臭氧、一氧化碳、氮氧化物平均浓度[1]；Elbayoumi 等结合大气污染物变量和气象条件，利用多元线性回归模型预测当地室外 $PM_{2.5}$ 浓度随着时间和季节的变化趋势[2].Ishak 等利用多元线性回归、模糊系统和广义可加模型研究突尼斯当地历史时刻 $PM_{2.5}$ 浓度对未来空气质量预测的影响[3].同时，Nieto 等通过建立多元自适应回归样条预测模型，分析出了西班牙奥维耶多当地气象分布特征[4].虽然这些回归模型被应用于大气污染预测，但是复杂地形、气象条件、污染物等多种因素共同影响，导致预测模型呈现出高度非线性，增加了 $PM_{2.5}$ 浓度预测难度.因此，传统回归模型在用于大气预测中，往往难以满足允许精度，预测性能欠佳.

大量研究表明，相比于传统统计学方法，神经网络具有良好的非线性逼近、自组织和自学习能力，无须确定输入和输出之间函数关系，只需通过对数据样本进行训练，利用训练好的网络来对输入数据进行预测[5]，可以有效解决建模难题，因此其被广泛应用于大气污染预测.孙宝磊等采用 MIV 方法筛出主要因子，建立 BP 神经模型预测昆明地区日均污染物浓度[6]；针对芬兰赫尔辛基城区，Kukkonen 等建立多种神经网络模型预测 $PM_{2.5}$ 小时平均质量浓度[7]；同样地，石灵芝等建

收稿日期： 2017-11-19；　**修回日期：** 2018-03-28.
基金项目： 国家自然科学基金资助项目（61473034，61673053）；北京科技新星计划交叉学科合作项目（Z161100004916041）.
作者简介： 李晓理*（1971-），男，博士，教授，博士生导师，E-mail：lixiaolibjut@bjut.edu.cn；梅建想（1992-），男，硕士生，E-mail：844208164@qq.com.

立 BP 神经网络模型预测湖南长沙火车站 PM_{10} 小时平均质量浓度[8]. Elangasinghe 等用 BP 神经网络结合 k-means 聚类算法[9]，建立了一个基于监测站气象条件和 $PM_{2.5}$、PM_{10} 浓度预测模型，能够有效预测新西兰南侧岛屿空气主要污染指数. 由此看出，BP 神经网络模型具有较好预测性能，已引起人们广泛关注. 但是，BP 是一类弱预测器，存在过拟合和容易收敛于局部极小点问题. 因此，模型预测精度有待进一步提高.

为此本文将 BP 神经网络和 Adaboost 算法结合，提出一种基于 BP_Adaboost 神经网络[10-11]的 $PM_{2.5}$ 浓度预测模型，建模时考虑影响 $PM_{2.5}$ 浓度的多种因素，并用灰色关联分析选取影响 $PM_{2.5}$ 浓度的主要因子作为神经网络输入变量；此外，运用改进粒子群算法来优化神经网络权重和阈值，以有效避免陷入局部最优解；最后以北京市海淀区万柳监测站和朝阳区北京工业大学监测点的实时数据为例，验证模型预测性能.

1　改进粒子群优化 BP_Adaboost 神经网络模型建立

1.1　灰色关联分析

灰色系统理论认为，含有已知信息或不确定性信息的系统，从表面上看，数据可能是随机的、杂乱无章的，但是其仍然是有界和有序的，数据集内会呈现一定规律. 正如对天气预报系统而言，大气中 $PM_{2.5}$ 浓度受多种因素共同影响，各因素间相互关系无法定量分析，还会在一定范围内波动，属于动态变化的. 因此，可借助灰色关联分析方法鉴别各因素之间发展趋势的相互依赖程度，找出各因素对 $PM_{2.5}$ 浓度的影响程度[12-13]，其计算步骤如下：

步骤 1　建立原始数据矩阵 $\boldsymbol{x}_i$

$$\boldsymbol{x}_i=(x_i(1)\quad x_i(2)\quad\cdots\quad x_i(k)) \tag{1}$$

式中：$x_i(k)$ 为第 i 个因素在第 k 时刻的原始数据；$i=1,2,\cdots,7$，$k=1,2,\cdots,n$，n 为原始数据长度.

步骤 2　计算初始化变换矩阵 $\boldsymbol{x}_i'$

$$\boldsymbol{x}_i'=(x_i(1)/x_i(1)\quad x_i(2)/x_i(1)\quad\cdots\quad x_i(k)/x_i(1))=(x_i'(1)\quad x_i'(2)\quad\cdots\quad x_i'(k)) \tag{2}$$

步骤 3　计算差序列 $\Delta_{oi}(k)$

$$\Delta_{oi}(k)=|x_i(k)-x_i'(k)| \tag{3}$$

步骤 4　计算关联系数 $\xi_{oi}(k)$ 和灰色关联度 γ_{oi}

$$\xi_{oi}(k)=\frac{\min\limits_i\min\limits_k\Delta_{oi}(k)+\varphi\max\limits_i\max\limits_k\Delta_{oi}(k)}{\Delta_{oi}(k)+\varphi\max\limits_i\max\limits_k\Delta_{oi}(k)} \tag{4}$$

其中 φ 为分辨系数，其作用在于提高关联系数间的差异显著性，$\varphi\in(0,1)$，经过多次反复试验可得，本文中 φ 取为 0.6.

$$\gamma_{oi}=\frac{1}{n-1}\sum_{k=1}^{n}\xi_{oi}(k) \tag{5}$$

在灰色关联分析过程中，以 PM_{10}、NO_2、CO、O_3、SO_2 的浓度和温度、相对湿度为变量，分别得出北京市海淀区万柳监测站和朝阳区北京工业大学监测点数据关联系数依次为 $\boldsymbol{\xi}_{o1}=$(0.982 6　0.872 1　0.745 8　0.722 5　0.898 2　0.927 1　0.931 8)，$\boldsymbol{\xi}_{o2}=$(0.972 9　0.884 5　0.752 7　0.710 9　0.872 6　0.942 8　0.930 1).

从上述结果可以看出，PM_{10} 浓度、温度、相对湿度、NO_2 浓度和 SO_2 浓度关联系数较大，因此，提取出这 5 个变量作为影响 $PM_{2.5}$ 浓度的主要因子.

1.2　BP_Adaboost 神经网络

1.2.1　*核心思想*　在绝大多数集成学习算法通过构造越来越复杂的预测器来提高预测精度时，Adaboost 却追求将最简单的弱预测器组合得到强预测器. 在训练子预测器的方法上，Adaboost 提供了重要启示：打破已有样本分布，重新采样使预测器更多地关注难学习的样本. 在算法使用上，仅需要指定迭代次数，不需要任何先验知识，一切运行过程中的参数由算法自适应地调整，因此被评价为最接近“拿来即用”的算法[14].

Adaboost 算法是组合多个弱预测器输出生成强预测器[15]. 首先，从样本空间中抽取 m 组作为训练数据，每组训练数据的权重均初始化为 $1/m$. 然后，分别训练弱预测器，迭代运行 T 次后，每组训练数据权重依据预测结果进行动态调整，若预测误差未能达到允许值将增大其权重，进一步加强学习. 经过反复迭代后，弱预测器将得到一个预测函数序列 $f_1,f_2,\cdots,f_T$，若预测结果越好其权重越大，反复迭代 T 次，通过弱预测函数加权得出强预测函数 $h(x)$，实现数据预测.

1.2.2　*算法流程与步骤*　基于 BP_Adaboost 模型的 $PM_{2.5}$ 浓度预测算法流程图如图 1 所示，算法步骤如下.

步骤 1　数据获取和网络初始化. 选取 m 组训练数据，赋予训练数据权重分布为 $D_t(i)=1/m$，依据样本输入和输出维数确定网络结构，网络初始权重和阈值由改进粒子群算法优化获得.

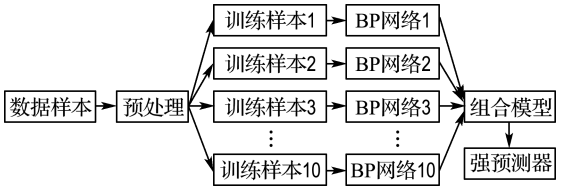


图 1 BP_Adaboost 算法结构

Fig. 1 The structure of BP_Adaboost algorithm

步骤 2 弱预测器. 训练第 t 个弱预测器时得到预测序列 $g(t)$的预测误差和

$$e_t=\sum_{i=1}^{m}D_t(i);\ g(t)\neq y \tag{6}$$

式中:y 为期望值.

步骤 3 计算预测序列权重. 依据预测误差 e_t 计算权重 a_t:

$$a_t=\frac{1}{2}\ln\left(\frac{1-e_t}{e_t}\right) \tag{7}$$

步骤 4 测试数据权重调整. 根据权重 a_t 调整下一轮训练样本权重,如下:

$$D_{t+1}(i)=\frac{D_t(i)}{B_t}\exp[-a_t y_i g_t(x_i)] \tag{8}$$

式中:B_t 是归一化因子,y 为期望值.

步骤 5 强预测函数. 经过 T 次迭代,由 T 组弱预测器函数 $f(g_t,a_t)$生成强预测器函数 $h(x)$.

$$h(x)=\mathrm{sgn}\left[\sum_{t=1}^{T}a_t f(g_t,a_t)\right] \tag{9}$$

1.3 改进粒子群优化 BP_Adaboost 神经网络

1.3.1 *改进粒子群算法* 粒子群(particle swarm optimization,PSO)算法是一种模拟生物机制的全局随机搜索算法,其通过搜索当前最优值来找到全局最优值[16]. 首先,对粒子进行初始化,接着经过数次迭代寻找最优解,在此过程中粒子速度和位置更新公式如下:

$$v_{is}=\omega v_{is}(t+1)+c_1 r_{1s}(t)(p_{is}-x_{is}(t))+ c_2 r_{2s}(t)(p_{gs}(t)-x_{gs}(t)) \tag{10}$$

$$x_{is}(t+1)=x_{is}(t)+v_{is}(t+1) \tag{11}$$

$$\omega=\omega_{\max}-\frac{t(\omega-\omega_{\min})}{t_{\max}} \tag{12}$$

式中:p_{is}是第 i 个粒子所经历过最好位置,p_{gs}是粒子群中最好位置;学习因子 c_1 和 c_2 分别为调节粒子飞向自身最好位置和全局最好位置方向的步长,取 $c_1=1.496\ 2$,$c_2=1.496\ 2$;r_{1s}和 r_{2s}为服从[0,1]均匀分布伪随机数;v_{is}是粒子速度,其值变化范围为$-10\sim10$;ω 为惯性权重,用来控制前一速度对当前速度的影响,ω 越大,越有利于全局搜索,而 ω 越小则越有利于精确局部搜索,寻找到最优解,因此,采用变化惯性权重可以有效避免陷入局部最优解. 为了更好地平衡算法全局和局部搜索能力,避免算法早熟和在全局最优解附近产生振荡现象,本文采用权重线性递减 PSO 算法. 其中,ω 随算法迭代次数变化公式为式(12),t 和 $t_{\max}$分别表示当前和最大迭代步数,设定 $t_{\max}=100$,$\omega_{\max}=0.9$,$\omega_{\min}=0.4$,ω 在不断迭代过程中线性递减,该算法能够寻找到最优解且具有很好收敛性. 由式(10)、(11)和(12)可知,惯性权重、学习因子和随机数共同决定粒子飞行速度,需要调整参数较少,避免只依赖经验对 ω 进行调节.

将粒子群算法全局搜索能力与神经网络学习算法相结合,既可解决盲目寻优问题,又能避免发生局部收敛情况. 在运用 PSO 算法时,将粒子群初始位置向量设为 BP_Adaboost 神经网络权重和阈值,然后用改进粒子群算法在整个粒子群中搜索最优位置,使均方误差最小化,直至满足算法停止条件. 此时,将最优位置向量赋给 BP_Adaboost神经网络权重和阈值,再训练网络,直至算法停止. 相比于随机初始化神经网络权重和阈值而言,经优化后的神经网络不易陷入局部极小点,从而能够提高算法性能.

1.3.2 *基于改进粒子群 BP_Adaboost 神经网络* 在搜索空间中,粒子群按事先设定飞行速度,在不断搜索和寻优过程中,其根据个体和群体飞行经验进行动态更新,其算法流程图如图 2 所示,其算法步骤如下:

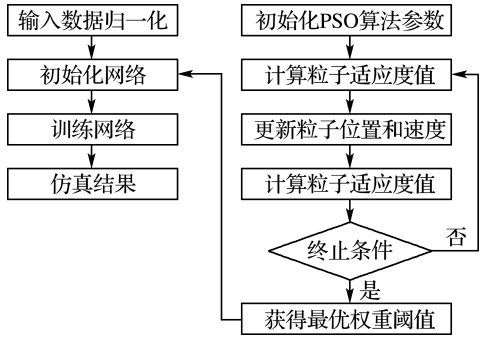


图 2 PSO 算法流程图

Fig. 2 The flow chart of PSO algorithm

步骤 1 初始化 PSO 算法相关参数.

步骤 2 计算每个粒子适应度值.

步骤 3 比较每个粒子当前和历史最好位置的适应度值,若较好,则把它视为当前最好位置.

步骤 4　比较每个粒子当前和全局最好位置的适应度值，若较好，则把它视为当前全局最好位置.

步骤 5　根据式(10)、(11)和(12)对速度和位置进行更新.

步骤 6　如果满足终止条件，则输出最优个体，并赋给 BP_Adaboost 神经网络，否则返回到步骤 2.

步骤 7　训练已构建好的 BP_Adaboost 神经网络，进行仿真实验.

2　$PM_{2.5}$ 实验数据分析和预测

北京作为中国首都，近些年来，由于人流量较大，工业化程度不断加大以及周边城市的影响，市内空气污染严重. 在北京市内，有 35 个监测站，包括 23 个环境评估点和 1 个郊区控制点，负责监测城市空气质量. 同时，6 个边界地区监测点用来反映周边城市对北京市环境的影响. 在二环、三环以及四环主干道路口设置 5 个交通监测点，用来反映交通流量信息. 本文选取了北京市海淀区万柳监测站和朝阳区北京工业大学监测点数据作为实验研究对象，该数据分别来自于城市空气项目[17]和空气质量传感网络监测仪，后者性能指标、外观和数据界面如图 3 所示. 大量研究表明，$PM_{2.5}$ 主要化学成分很复杂，含有碳物质(有机碳、元素碳)、硫酸盐、硝酸盐、铵盐等[18]，而且各物质在大气中还存在复杂物理和化学变化，因此，可以得出 $PM_{2.5}$ 浓度与大气污染物变量和气象条件相关.



(a) 性能指标　　(b) 外观　　(c) 数据界面

图 3　空气质量传感网络监测仪

Fig. 3　Air quality sensor network detector

由于影响 $PM_{2.5}$ 浓度因素很多，如气象条件、大气污染物浓度、交通流量、工业排放废气和废物等，考虑到监测站数据样本的重要性和有效性，选取北京市海淀区万柳监测站(2014-11-01～2014-11-25)和朝阳区北京工业大学监测点(2017-07-07～2017-08-06)每小时监测数据进行实验，如图 4 和 5 所示. 此外，为了研究两个监测站在 1 d 内 6 种大气污染物浓度变化所呈现规律，以万柳监测站(2014-11-08)和北京工业大学监测点(2017-07-20)数据为例进行分析，如图 6 和 7 所示.

从图 4 和 5 可以看出，$PM_{2.5}$ 浓度与 PM_{10} 浓度呈现出一致变化，具有很强相关性，而与温度和相对湿度有着负相关的变化趋势. 此外，在夏季期间，多种污染物浓度较低，尤其是 NO_2 和 SO_2 浓度较低，空气质量较好，而在冬季多种污染物浓度相对较高. 在不同季节，$PM_{2.5}$ 浓度差异性显著，一般地，冬季 $PM_{2.5}$ 浓度相对较高，主要来源于燃煤量急剧增加，导致大量颗粒性物质被排放到大气中，而且在冬季温度和相对湿度较低，大气结构稳定，导致污染物累积易形成污染事件.

从图 6 和 7 可以看出，1 d 当中各污染物浓度呈现两头高中间低的变化趋势，即凌晨和夜间污染物浓度较高，其他时间段相对较低，这与温度和相对湿度有着直接关系. 在凌晨和夜间，温度和相对湿度较低，不利于各种污染物扩散，导致污染物不断累积. 但是，随着温度升高和相对湿度变大，污染物消散加速，污染物浓度会降低.

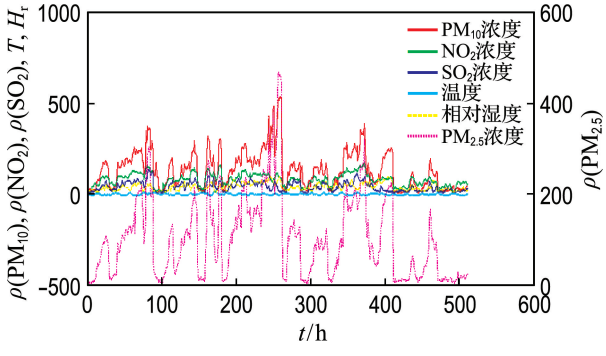


图 4　万柳监测站 2014-11-01～2014-11-25 数据

Fig. 4　The data from November 1 to November 25, 2014 at Wanliu station

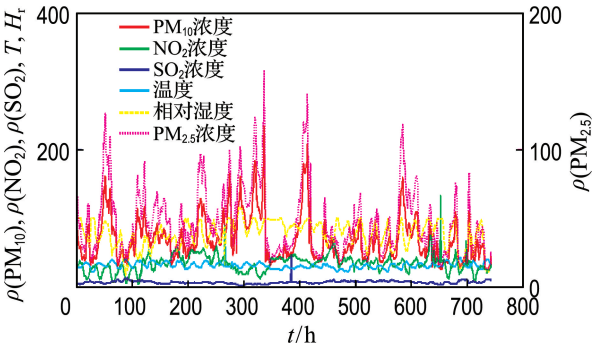


图 5 北京工业大学监测点 2017-07-07～2017-08-06 数据

Fig. 5 The data from July 7 to August 6, 2017 at Beijing University of Technology point

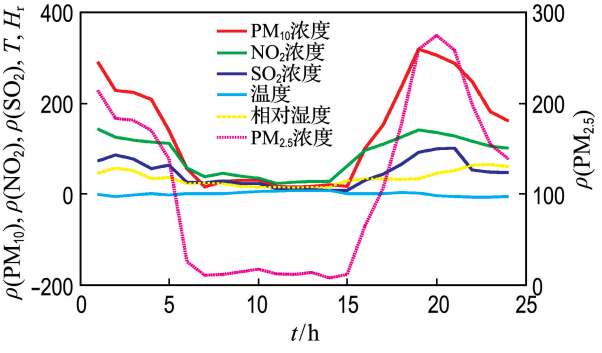


图 6 万柳监测站 2014-11-08 数据

Fig. 6 The data on November 8, 2014 at Wanliu station

此外，上述这些数据是每间隔 1 h 监测的，为了消除各数据之间量纲差异性，对于该神经网络的训练数据和测试数据，皆采用最大最小法进行归一化：

$$x_{\text{norm}}=\frac{x_k-x_{\min}}{x_{\max}-x_{\min}} \tag{13}$$

式中：$x_{\min}$ 为数据序列最小值，$x_{\max}$ 为数据序列最大值，$x_{\text{norm}}\in[0,1]$.

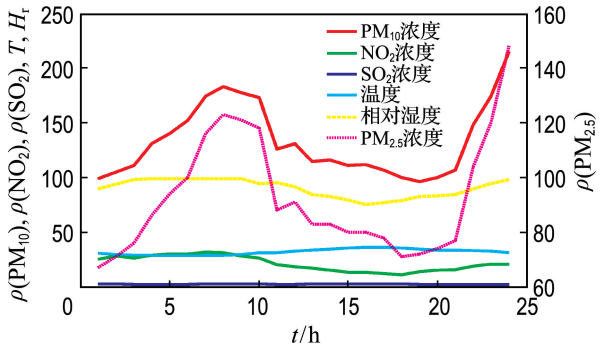
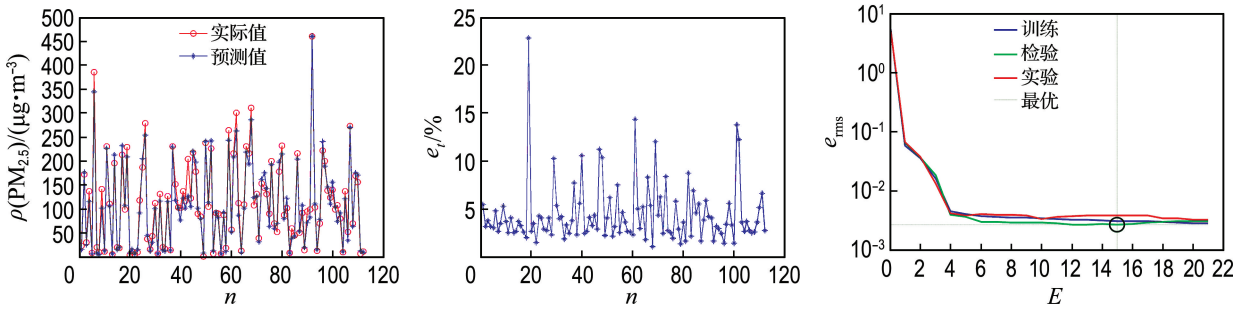


图 7 北京工业大学监测点 2017-07-20 数据

Fig. 7 The data on July 20, 2017 at Beijing University of Technology point

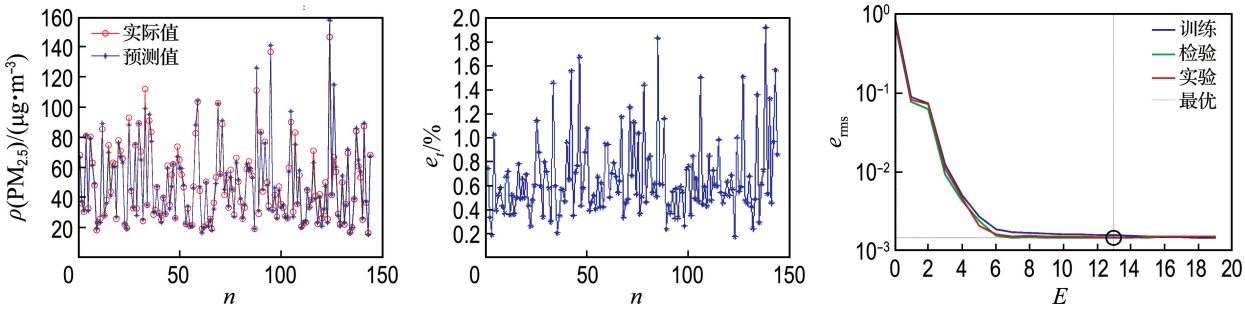
在本次实验中，由灰色关联分析结果知，下一个小时 $PM_{2.5}$ 浓度与当前时刻 PM_{10} 浓度、温度、相对湿度、NO_2 浓度和 SO_2 浓度关联系数较大，故选取当前时刻 PM_{10} 浓度、NO_2 浓度、SO_2 浓度、温度和相对湿度作为该神经网络输入变量，大气中下一个小时 $PM_{2.5}$ 浓度作为输出变量. 在实验过程中，选取万柳监测站(2014-11-01～2014-11-25)总共 512 个冬季数据样本和北京工业大学监测点(2017-07-07～2017-08-06)总共 744 个夏季数据样本，再分别从中随机选取 400 个和 600 个数据样本训练神经网络，其网络训练误差曲线如图 8(c)、9(c)所示，确定其权重和阈值，使其收敛，并能够满足在不同季节数据样本下预测大气中 $PM_{2.5}$ 浓度的要求. 经过多次重复实验，选取训练误差最小，最终确定 BP 神经网络结构为 5-6-1，BP_Adaboost 神经网络由 10 个 BP 神经网络模型训练生成. 待神经网络训练完成后，为了验证其预测性能，再分别用剩余 112 个和 144 个数据样本来测试网络，从而实现提前 1 h 预测大气中 $PM_{2.5}$ 浓度的功能.



(a) 模型预测结果 (b) 模型预测相对误差 (c) 网络训练误差曲线图

图 8 万柳监测站预测结果

Fig. 8 Forecast results for Wanliu station



（a）模型预测结果　　（b）模型预测相对误差　　（c）网络训练误差曲线图

图 9　北京工业大学监测点预测结果

Fig. 9　Forecast results for Beijing University of Technology point

3　结果和讨论

3.1　性能指标

在本文中，选取 3 种不同算法进行对比，为了衡量其预测性能，定义 3 种统计指标用来评估，依次为均方根误差（e_{rms}）、平均绝对百分比误差（e_{map}）、相关系数（R^2），其计算公式如下：

$$e_{rms}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i-Y_i)^2} \tag{14}$$

$$e_{map}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{O_i-Y_i}{O_i}\right|\times 100\% \tag{15}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(O_i-Y_i)^2}{\sum_{i=1}^{n}(\overline{Y}-Y_i)^2} \tag{16}$$

式中：O_i 为 $PM_{2.5}$ 浓度在第 i 时刻实际值，Y_i 为同一时刻模型预测值，$\overline{Y}$ 为模型预测输出平均值. e_{rms} 反映模型预测输出值稳定性，e_{map} 反映模型预测输出值偏离 $PM_{2.5}$ 浓度实际值程度，两者值皆越小，性能越好；R^2 反映 $PM_{2.5}$ 浓度实际值与模型预测输出值相关联程度，其值越接近 1，性能越好.

3.2　预测结果

针对万柳监测站和北京工业大学监测点每小时监测数据，运用改进粒子群优化 BP_Adaboost 神经网络（MPSO-BP_Adaboost）建模进行仿真实验，其预测结果如图 8 和 9 所示. 此外，将本文中所采用预测模型和 BP_Adaboost、BP、广义回归神经网络（GRNN）3 种预测模型进行对比，其性能指标见表 1 和 2.

从图 8 和 9 可以看出，针对万柳监测站和北京工业大学监测点，采用改进粒子群优化 BP_Adaboost 神经网络建模来对大气中未来 1 h $PM_{2.5}$ 浓度进行预测，预测输出值和期望输出值基本相吻合，能够达到高精度预测的效果. 在某些时间点预测时，可能由于外部环境突变、一些人为因素和神经网络自身性能等原因，存在预测误差，但是，网络输出相对误差只是在一个较小范围内波动，在大气预测应用方面可以接受，因此，BP_Adaboost 神经网络能够很好实现预测功能.

表 1　万柳监测站 4 种预测模型性能指标对比

Tab. 1　The performance index contrast of four kinds of prediction models at Wanliu station

模型	均方根误差	平均绝对百分比误差/%	R^2
MPSO-BP_Adaboost	6.097 1	4.982 0	0.927 1
BP_Adaboost	8.157 2	5.443 0	0.901 8
BP	12.006 2	7.986 0	0.872 7
GRNN	9.851 2	7.253 0	0.892 0

表 2　北京工业大学监测点 4 种预测模型性能指标对比

Tab. 2　The performance index contrast of four kinds of prediction models at Beijing University of Technology point

模型	均方根误差	平均绝对百分比误差/%	R^2
MPSO-BP_Adaboost	4.211 8	1.090 0	0.950 2
BP_Adaboost	7.036 4	3.410 0	0.910 1
BP	10.901 4	7.820 0	0.882 5
GRNN	9.681 1	7.630 0	0.902 4

从表 1 和 2 来看，针对两个监测站点数据样本，实验结果略微不同，主要由于万柳监测站数据样本数值变化范围较大，加大了预测难度. 为了保证模型有较高预测精度，在数据样本具有真实性和可靠性前提下，数据样本量应该尽可能大，包含不同条件下数据样本信息，使模型训练充分，提高

其泛化能力.此外,基于本文实验部分所述建立的MPSO-BP _ Adaboost、BP _ Adaboost、BP 和GRNN 4种预测模型,都具有有效性和可靠性.但是,MPSO-BP_Adaboost 预测模型相比于其他3种预测模型而言,均方根误差和平均绝对百分比误差均是最小,R^2 为最大,这表明改进粒子群优化 BP_Adaboost 在用于预测大气中 $PM_{2.5}$ 浓度时,模型性能要优于另外3种预测模型,从而能够更好地实现预测大气中 $PM_{2.5}$ 浓度的目标,为人们出行提供参考.

4 结 语

本文运用改进粒子群优化 BP_Adaboost 神经网络来预测北京市海淀区万柳监测站和朝阳区北京工业大学监测点下一个小时 $PM_{2.5}$ 浓度.根据站点监测数据和神经网络预测输出结果,得出如下结论:(1)影响大气中 $PM_{2.5}$ 浓度有多种因素,相互之间可能存在强耦合,利用灰色关联分析找出对 $PM_{2.5}$ 浓度影响较大的因子,并用于 BP_Adaboost 神经网络建模,可以有效缩短建模时间和更精确建模.(2)BP_Adaboost 神经网络是强分类预测器,相比于其他神经网络,在大气预测方面,更利于加强神经网络泛化能力和提高预测模型精度.(3)采用改进粒子群优化 BP_Adaboost 神经网络,可以有效避免神经网络在训练时陷入局部最优解,进一步改善预测模型性能.在本文中,未能从理论角度去证明改进粒子群算法可以避免陷入局部最优解,这将在以后研究中予以考虑.其次,只做了短期预测,在未来工作中,将尝试去做长期预测,使之更具有实用行和可靠性.最后,由于 BP 神经网络结构只能凭借试凑法来确定,将考虑用 PSO 算法去优化网络结构[19-20],解决这一难题并应用于大气中其他污染物浓度预测.

参考文献:

[1] KUMAR U, JAIN V K. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO) [J]. **Stochastic Environmental Research and Risk Assessment**, 2010, **24**(5):751-760.

[2] ELBAYOUMI M, RAMLI N A, YUSOFA N F F M, *et al*. Multivariate methods for indoor PM_{10} and $PM_{2.5}$ modelling in naturally ventilated schools buildings [J]. **Atmospheric Environment**, 2014, **94**(8):11-21.

[3] ISHAK A B, MOSLAH Z. Analysis and prediction of PM_{10} concentration levels in Tunisia using statistical learning approaches [J]. **Environmental and Ecological Statistics**, 2016, **23**(3):1-22.

[4] NIETO P, ANTÓN J, VILÁN J, *et al*. Air quality modeling in the Oviedo urban area (NW Spain) by using multivariate adaptive regression splines [J]. **Environmental Science and Pollution Research**, 2015, **22**(9):6642-6659.

[5] MOUSTRIS K P, ZIOMAS I C, PALIATSOS A G. 3-Day-ahead forecasting of regional pollution index for the pollutants NO_2, CO, SO_2, and O_3 using artificial neural networks in Athens, Greece [J]. **Water Air and Soil Pollution**, 2010, **209**(1/4):29-43.

[6] 孙宝磊,孙 鬲,张朝能,等. 基于BP神经网络的大气污染物浓度预测[J]. 环境科学学报, 2016, **37**(5):1864-1871.
SUN Baolei, SUN Hao, ZHANG Chaoneng, *et al*. Forecast of air pollutant concentration by BP neural network [J]. **Acta Scientiae Circumstantiae**, 2016, **37**(5):1864-1871. (in Chinese)

[7] KUKKONEN J, PARTANEN L, KARPPINEN A, *et al*. Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki [J]. **Atmosphere Environment**, 2003, **37**(4):4539-4550.

[8] 石灵芝,邓启红,路 婵,等. 基于BP人工神经网络大气颗粒物 PM_{10} 浓度预测[J]. 中南大学学报(自然科学版), 2012, **43**(5):1969-1974.
SHI Lingzhi, DENG Qihong, LU Chan, *et al*. Prediction of PM_{10} mass concentrations based on BP artificial neural network [J]. **Journal of Central South University (Science and Technology)**, 2012, **43**(5):1969-1974. (in Chinese)

[9] ELANGASINGHE M A, SINGHAL N, DIRKS K N, *et al*. Complex time series analysis of PM_{10} and $PM_{2.5}$ for a coastal site using artificial neural network modelling and *k*-means clustering [J]. **Atmospheric Environment**, 2014, **94**(1):106-116.

[10] 李 航. 统计学习方法[M]. 北京:清华大学出版社, 2012.
LI Hang. **Statistical Learning Methods** [M]. Beijing: Tsinghua University Press, 2012. (in Chinese)

[11] GAO Yun, WEI Xin, ZHUANG Wenqin, *et al*. QoE prediction for IPTV based on BP _ adaboost neural networks [C] // **2017 13th International Wireless Communications and Mobile Computing**

Conference（**IWCMC**）. Valencia：IEEE，2017：32-37.

[12] TSAI M S，HSU F Y. Application of grey correlation analysis in evolutionary programming for distribution system feeder reconfiguration [J]. **IEEE Transactions on Power Systems**，2010，**25**(2)：1126-1133.

[13] 罗庆成，徐国新. 灰色关联分析与应用[M]. 南京：江苏科学技术出版社，1989：110-125.
LUO Qingcheng，XU Guoxin. **The Analysis Application of Grey Correlation** [M]. Nanjing：Jiangsu Science and Technology Publishing House，1989：110-125.（in Chinese）

[14] HASTIE T，TIBSHIRANI R，FRIEDMAN J. **The Elements of Statistical Learning** [M]. New York：Springer，2001：337-384.

[15] 王小川，史　峰. MATLAB 神经网络 43 个案例分析[M]. 北京：北京航空航天大学出版社，2013：115-126.
WANG Xiaochuan，SHI Feng. **The Analysis of 43 Neural Network Cases Using MATLAB** [M]. Beijing：Beijing University of Aeronautics and Astronautics Press，2013：115-126.（in Chinese）

[16] 钱　锋. 粒子群算法及其工业应用[M]. 北京：科学出版社，2013：120-156.
QIAN Feng. **Particles Swarm Optimization and Its Industrial Application** [M]. Beijing：Science Press，2013：120-156.（in Chinese）

[17] ZHENG Yu，CAPRA L，WOLFSON O，*et al*. Urban computing：concepts，methodologies，and applications [J]. **ACM Transactions on Intelligent Systems and Technology**，2014，**5**(3)：112-118.

[18] 曹军骥. $PM_{2.5}$ 与环境[M]. 北京：科学出版社，2014.
CAO Junji. **$PM_{2.5}$ and the Environment in China** [M]. Beijing：Science Press，2014.（in Chinese）

[19] WU Jiansheng，JIN Long. Study on the meteorological factor prediction model using the learning algorithm of neural ensemble based on PSO algorithm [J]. **Journal of Tropical Meteorology**，2009，**15**(1)：83-88.

[20] SHEIKHAN M，SHA′BANI A A. PSO-optimized modular neural network trained by OWO-HWO algorithm for fault location in analog circuits [J]. **Neural Computing and Applications**，2013，**23**(2)：519-530.

$PM_{2.5}$ concentration prediction using BP_Adaboost neural network based on modified particle swarm optimization

LI　Xiaoli[*1,2,3]，　MEI　Jianxiang[1]，　ZHANG　Shan[1]

（1. Faculty of Information Technology，Beijing University of Technology，Beijing 100124，China；
2. Beijing Key Laboratory of Computational Intelligence and Intelligent Systems，Beijing 100124，China；
3. Engineering Research Center of Digital Community，Ministry of Education，Beijing 100124，China）

Abstract：In order to improve the prediction accuracy of the atmospheric pollutants concentration，the gray correlation analysis is used to select the main factors affecting the $PM_{2.5}$ concentration in the atmosphere. Regarding them as the input variables，a model based on BP_Adaboost neural network is proposed to predict the $PM_{2.5}$ concentration. The modified particle swarm optimization（MPSO）algorithm is applied to choose the weight and threshold of BP_Adaboost neural network，which can availably avoid falling into local optimal solution. According to the concentration of air pollutants and meteorological condition，the data between November 1，2014 to November 25，2014 and July 7，2017 to August 6，2017，which are monitored every hour by the Wanliu station of Haidian distinct and Beijing University of Technology point of Chaoyang distinct in Beijing，are used as the experiment object. The simulation results show that the $PM_{2.5}$ concentration prediction performance of MPSO-BP_Adaboost neural network is better than that of BP_Adaboost，BP and generalized regression neural network.

Key words：gray correlation analysis；BP_Adaboost neural network；$PM_{2.5}$ concentration prediction model；modified particle swarm optimization（MPSO）algorithm