

基于灰色 BP-NN 优化组合的 PM_{2.5} 预测

黄 鹰¹, 史爱武², 陈占龙¹, 张 威¹

(1. 中国地质大学(武汉) 地理与信息工程学院, 湖北 武汉 430070; 2. 武汉纺织大学 计算机学院, 湖北 武汉 430070)

摘 要: 针对传统的 BP 神经网络模型无法有效表达时间序列数据中存在的历史特征的缺陷, 提出利用灰色预测原理具备发现事物历史变化规律性的优势来解决 BP 神经网络预测模型的这一弱点, 最后得到的灰色 BP-NN 优化组合模型具备了更高的预测精度。实验采用中国气象站 2018 年 1 月至 2 月北京市 10 个监测点的 PM_{2.5} 质量浓度及其对应的每小时的空气污染物浓度、气象因子建立神经网络预测模型, 并采用灰色预测算法对神经网络模型进行改进, 改进后的结果为: 在系统误差上有了较大的降低, 同时预测结果与实测结果之间的拟合程度更好。

关键词: 空气污染; PM_{2.5} 浓度预测; 气象因子; 神经网络; 灰色预测算法; 时间序列数据; 拟合

中图分类号: TN711; TP183

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.191101

中文引用格式: 黄鹰, 史爱武, 陈占龙, 等. 基于灰色 BP-NN 优化组合的 PM_{2.5} 预测[J]. 电子技术应用, 2020, 46(6): 82-85, 92.

英文引用格式: Huang Ying, Shi Aiwu, Chen Zhanlong, et al. PM_{2.5} prediction based on the optimal combination of grey BP neural network[J]. Application of Electronic Technique, 2020, 46(6): 82-85, 92.

PM_{2.5} prediction based on the optimal combination of grey BP neural network

Huang Ying¹, Shi Aiwu², Chen Zhanlong¹, Zhang Wei¹

(1. School of Geography Information Engineering, China University of Geosciences, Wuhan 430070, China;

2. School of Mathematics and Computer, Wuhan Textile University, Wuhan 430070, China)

Abstract: Aiming at the defect that the traditional BP neural network model cannot effectively express the historical features existing in time series data, a method with the combination of BP neural network and grey forecast principle was proposed. Furthermore, grey forecast principle has the advantage of discovering the laws of historical changes, which can overcome the weakness of BP neural network prediction model and this method have higher prediction accuracy. The neural network prediction model was established by using the PM_{2.5} mass concentration of ten monitoring stations in Beijing in January and February 2018, as well as the corresponding hourly air pollutant concentration and meteorological factors. Meanwhile, the grey forecast algorithm was used to improve the neural network model. The results indicate that the improved method has the features of lower system error, and better fitting degree between the predicted result and the measured result.

Key words: air pollution; PM_{2.5} concentration prediction; meteorological factors; neural network; grey forecast algorithm; time series data; fitting

0 引言

近年来, 基于神经网络算法预测 PM_{2.5} 成为 PM_{2.5} 监测^[1]研究的热点。

人工神经网络具有很好的自适应性、自组织性和很强的自主学习能力^[2-4]。采用人工神经网络的方式去预测 PM_{2.5} 浓度值具备很高的适用性^[5]。但是直接采用典型的神经网络方法在收敛速度和泛化能力上并不理想, 所以一些学者通过用相关性分析的方法降低输入样本的维度, 可以在一定程度上解决收敛速度的影响。张怡文和李凤英等人分别采用了逐步回归和 Pearson 相关系数的方法分析影响 PM_{2.5} 相关因素的相关性, 降低输入样本的维数, 以此来提高算法的收敛速度^[6-7]。针对常用

的几种神经网络用于预测 PM_{2.5} 浓度值的方法存在的局部极值问题, 马天成等人则是将粒子群算法(Particle Swarm Optimization, PSO)与模糊神经网络进行融合, 发挥 PSO 算法全局寻优的特点, 预测 PM_{2.5} 颗粒物浓度的变化规律^[8]; 荆涛、李霖等人通过遗传算法与 BP(Back Propagation)神经网络算法相结合的方式也解决了 BP 神经网络算法训练过程中的局部极值问题^[9]。PM_{2.5} 数据及其相关影响因素数据都是具有一定时间相关性的时间序列数据, 其具备一定的历史特性, 而 BP 神经网络模型对这种特性无法进行有效的表达。裴雨潇等人通过把 PM_{2.5} 的数据构成时间序列, 并进行小波变换, 将低频部分和高频部分分别用不同的模型进行预测, 再将预测

值进行叠加,最后得到的结果比单纯用BP神经网络模型预测的效果要理想^[10]。目前国内外研究PM2.5预测模型多是基于神经网络结合其他算法进行模型的改进,这种方式相对比单纯用神经网络的方式在预测精度、泛化能力上都有较大的提升,但是PM2.5受其他因素的影响较大,考虑到从周围环境的角度来建立PM2.5模型的方式,预测精度还可以有所提升。

传统的灰色预测模型(Grey Model, GM(1, 1))将时间序列数据看成一个随时间变化的函数,但经过大量的实验表明经典灰色模型缺乏一定的预测稳定性,即使时间序列为纯指数序列,在做长期预测时仍存在较大的偏差^[11-12]。基于这个原因,谢乃明等人提出离散灰色预测模型(Discrete Grey Forecasting Model, DGM(1, 1)),并经过实验表明,DGM(1, 1)相较于传统的GM(1, 1)模型即使在时间序列数据大致符合指数增长规律也具有较好的预测精度,且对于长期预测有一定的优势^[13]。

气象因素和空气质量因素均为典型的时间序列数据,但是分析影响PM2.5相关因素时,发现无论是气象因素还是空气质量因素都是震荡序列,GM(1, 1)和DGM(1, 1)两种模型对于单调递增的时间序列数据都具备一定的预测精度,而对于震荡序列的预测精度则不能够确定。王岩、黄张裕等人提出一种基于震荡序列的灰色预测模型(Stochastic Discrete Grey Forecasting Model, SDGM(1, 1)),在进行震荡序列预测时,能够达到比较好的精度^[14]。

通过对PM2.5浓度影响因子中较为重要的一项影响因子——相对湿度值采用经典灰色预测模型(GM(1, 1))进行预测,将得到的相对湿度值曲线与采用改进后的灰色预测模型(DGM(1, 1), SDGM(1, 1))预测的预测值曲线进行对比,如图1所示。

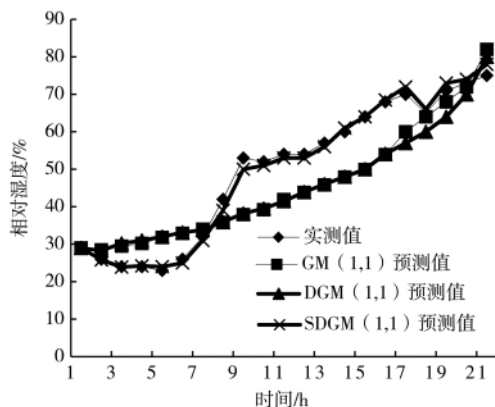


图1 模型预测曲线和实测曲线

由图1可知,基于震荡序列的改进之后的预测模型预测得到的相对湿度值与实测值曲线的走势相符,且相对偏差较小,拟合度非常好。因此,SDGM(1, 1)模型对于具有震荡时序特征的PM2.5相关影响因素的预测要更好。

本文提出的灰色BP-NN优化组合模型结合灰色预测方法与BP神经网络模型,利用灰色预测方法具备表

征事物历史变化规律的特性来改进原有的神经网络预测模型,选取相关的气象因素和相关空气质量因素,采用多元统计分析的方法取得与PM2.5相关性较大的影响因子作为模型的输入参数。

1 灰色预测算法优化BP神经网络模型

1.1 改进灰色预测模型

基于震荡序列的灰色预测模型建立过程大致如下:

对于序列 $x^0=(x_1^0, x_2^0, \dots, x_n^0)$,若满足: $\exists k^*, k^* \in (2, 3, \dots, n)$,使得 $x_k^0 - x_{k-1}^0 < 0, x_{k^*}^0 - x_{k^*-1}^0 > 0$,序列 x^0 即为震荡序列。 x^0 中元素最大值和最小值之差称为振幅,记为 $A(x^0)$ 。对于 $x^0=(x_1^0, x_2^0, \dots, x_n^0)$,将 $x^0 d=(x_1^0 d, x_2^0 d, \dots, x_{n-1}^0 d)$ 称之为平滑序列,其中平滑系数 d 由式(1)得出:

$$x_k^0 d = [(x_k^0 + A(x^0)) + (x_{k+1}^0 + A(x^0))] / 4$$

$$k=1, 2, 3, \dots, n-1 \quad (1)$$

由上可求出序列 $x^0=(x_1^0, x_2^0, \dots, x_n^0)$ 的平滑序列 $y^0=(y_1^0, y_2^0, \dots, y_n^0)$ 。由此建立离散灰色模型 $y_{k+1}^1 = \lambda_1 y_k^1 + \lambda_2 (\lambda_1$

和 λ_2 为待估系数),其中 $y_k^1 = \sum_{i=1}^k y_i^0$ (y_k^1 为累加算子序列的

第 k 个元素, y_i^0 为平滑序列的第 i 个元素),后面的推导过程与GM(1, 1)类似,这里就不做赘述。

1.2 神经网络与灰色预测相结合的优化组合模型

优化组合模型首先通过对影响PM2.5浓度变化因素的探究,进行灰色预测模型输入变量的选择;然后对单个选定的影响因素建立灰色预测模型,预测神经网络预测样本的输入参数值;然后基于神经网络较强的适用性和容错能力,对历史数据建立模型。通过灰色预测方法得到当前时间点的PM2.5相关影响因素数据,这部分数据具有一定的历史特征,将得到的数据代入到已经训练好的神经网络模型中,即得到需要进行预测的时间段的PM2.5的值。

(1)将取得的空气污染相关因素和气象相关因素原始数据通过缺失值补全、异常值处理、数据标准化等一系列数据预处理措施,作为系统的输入数据;

(2)采用多元统计分析的方式对输入数据进行一系列相关性分析、回归分析和通径分析后,取得符合条件的输入变量;

(3)在步骤(2)的基础上建立单变量的灰色预测模型和人工神经网络预测模型;

(4)利用步骤(3)中灰色预测模型得到的预测值代入训练好的人工神经网络模型中,得到需要预测的PM2.5数据。模型预测流程图如图2所示。

2 实验与结果分析

2.1 实验参数选择及实验数据

据统计研究发现,影响PM2.5浓度值的相关因素大致可以分为以下两类:

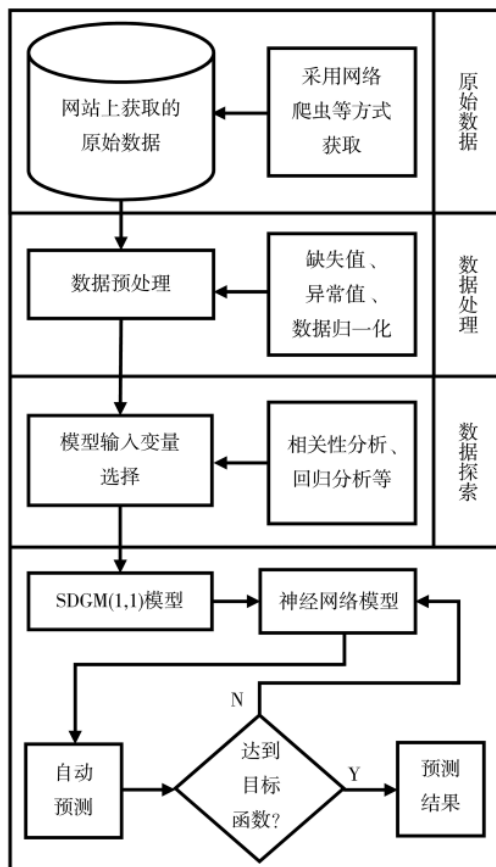


图2 基于灰色预测模型和神经网络的PM2.5预测模型流程图

(1) 空气污染相关因素: PM2.5 的来源主要有两种方式, 一种是由直接污染排放生成的, 另一种则是由前体物二次转化而来的, 其中最为主要的前体物包括 PM10、二氧化硫、氮氧化物和一氧化碳等空气污染物, 通过调研和统计分析的方法得到影响 PM2.5 浓度变化的主要空气质量因子有 CO、NO₂ 和 SO₂ 3 个影响因子。

(2) 气象相关因素: PM2.5 的值一方面受到产生 PM2.5 的污染物的影响, 另外一方面则与大气扩散有着很大的关系, 良好的大气扩散条件可以降低 PM2.5 的浓度值, 所以对影响 PM2.5 浓度变化的气象因子进行统计分析, 最终选择了相对湿度、日照时间、温度和风速作为神经网络输入参数中的气象因子。

参数选择过程如下:

(1) 数据预处理

由于模型的输入数据中的各个参数量纲不统一, 因此在进行数据分析时, 要先进行量纲统一的处理, 即归一化处理。

(2) 神经网络参数选择

本实验中 BP 神经网络结构采用 3 层神经网络, 隐藏层节点数采用试错的方法进行设计, Hinch-Nielsen 曾给出过单隐藏层神经元个数上界公式:

$$N_{\text{hid}} = N_{\text{input}} \times 2 + 1 \quad (2)$$

其中: N_{hid} 为隐藏层所应具有节点数最大值, N_{input} 为输

入层节点数。实验中确定的输入参数包括 4 个空气污染物因子和 4 个气象相关因子, 所以模型的输入层节点数为 8, 因此最大试错节点数为 17, 取 2~17 个隐藏层节点进行网络训练, 得到结果如表 1 所示。从表 1 中发现, 当隐藏层节点数为 14 时, 神经网络的训练误差最小, 所以最终确定的模型隐藏层节点数为 14 个。

表1 不同隐藏层节点数下神经网络的系统误差

神经元个数	网络误差	神经元个数	网络误差
2	49.4	10	30.98
3	48.3	11	29.43
4	46.17	12	25.61
5	40.63	13	23.47
6	39.12	14	18.45
7	38.55	15	27.23
8	36.89	16	34.11
9	33.99	17	32.13

(3) 传递函数的选择

选用 ReLU 函数作为预测模型的传递函数。

(4) 训练函数的设计

采用的训练函数是 Python 提供的 Adam 优化算法。

(5) 数据源

实验数据来源于 2018 年 1~2 月北京市 10 个监测点按小时统计数据, 数据类型包括气象因素(温度、湿度、气压、降水量、日照时间、风速)和空气污染物因素(SO₂、NO₂、CO、O₃)浓度。

2.2 实验结果与评价分析

实验结果采用计算平均相对误差(Mean Relative Error, MRE)、平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root-Mean-Square Error, RMSE)进行评价分析, 其计算公式如下:

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \bar{y}_i|}{y_i} \quad (3)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}_i| \quad (4)$$

$$\text{RMSE} = \sqrt{\left(\sum_{i=1}^N (y_i - \bar{y}_i)^2 \right) / N} \quad (5)$$

其中, y_i 为样本值, $\bar{y}_i$ 为样本预测值, N 为样本个数。

2.3 改进 BP 神经网络预测模型结果分析

根据前文介绍的数据预处理和参数选择以及灰色模型和神经网络模型原理和推导公式, 将采用以下步骤进行模型建立、学习和预测。

(1) 根据神经网络预测模型需采取 80% 的数据作为训练数据, 20% 的数据作为测试数据, 本文取得北京市 10 个检测点的 2017 年 12 月 11 日~2018 年 2 月 11 日的按小时统计数据, 包括气压、相对湿度、温度等气象数据和 PM2.5、PM10、SO₂、CO、NO₂ 等空气污染物数据共 1 488 条作为实验数据, 其中 1 190 条作为训练数据, 298 条作

计算机技术与应用

Computer Technology and Its Applications

为测试数据;

(2) 将这些原始数据采用前文提到的方法进行数据预处理,包括缺失值处理和异常数据处理;

(3) 采用多元统计方法进行输入变量选择,最终选择了相对湿度、风速、温度和气压 4 个气象因素以及 SO_2 、 NO_2 和 CO 3 个空气质量因素作为模型输入参数变量,按序分别设为 $X_i, i=1, 2, \dots, 7$;

(4) 通过 1 190 条训练数据中的各个输入参数变量,采用 Python 编程的方式建立单变量灰色模型得到 298 条预测数据所对应的输入参数变量的值。预测结果的精度等级采用后验差的方式进行检验得到如表 2 所示的结果。

表 2 灰色预测模型 $\text{PM}_{2.5}$ 相关因素参数精度表

时间/h	X_1	X_2	X_3	X_4	X_5	X_6	X_7
01-19:15	1 020.6	0.1	-3.1	49	8	16	0.6
01-19:16	1 020.9	0.9	-4	49	10	20	0.6
01-19:17	1 020.5	1.3	-2.7	44	11	29	0.7
01-19:18	1 020.4	1.4	-2	37	11	41	0.8
...	...	...	...	...	...	...	...
02-11:23	1 019.7	2.7	2.6	42	4	14	0.5
预测精度等级(均值)	好	好	合格	好	好	合格	好

(5) 将数据零均值标准化处理后代入到训练好的 3 层神经网络预测模型(输入层 7 节点,隐藏层 12 节点,

输出层 1 节点),得到预测值与实际值的对比,图 3 为改进神经网络预测的 $\text{PM}_{2.5}$ 值与实际值的对比图。作为对比,采用传统 BP 神经网络模型预测结果,如图 4 所示。

根据图 3 和图 4 可知,经过灰色模型进行改进的神经网络预测模型相比传统的基于 BP 神经网络的预测模型在系统误差上有了极大的降低,同时其拟合程度更好。

改进后的模型的主要区别在于对测试数据通过灰色预测方法进行预测,由于灰色模型具有发现事物潜在变化特征的性质,使得采用经过灰色预测进行处理过后的数据包含了历史因素的影响,从而对 BP 神经网络不能很好地利用时间序列中存在的规律的缺点进行了改善。经过改进后的模型计算系统误差与传统的基于 BP 神经网络模型的系统误差进行对比,如表 3 所示。

表 3 传统 BP 神经网络模型
与改进模型误差值对比

统计量	MRE	MAE	RMSE
传统 BP 神经网络模型	0.13	6.19	12.63
改进神经网络模型	0.08	4.77	9.16

3 结论

本文采用灰色预测与人工神经网络相结合的方式建立神经网络预测模型,有效改进了传统的 BP 神经网络模型无法有效地表达时间序列数据中存在的历史特

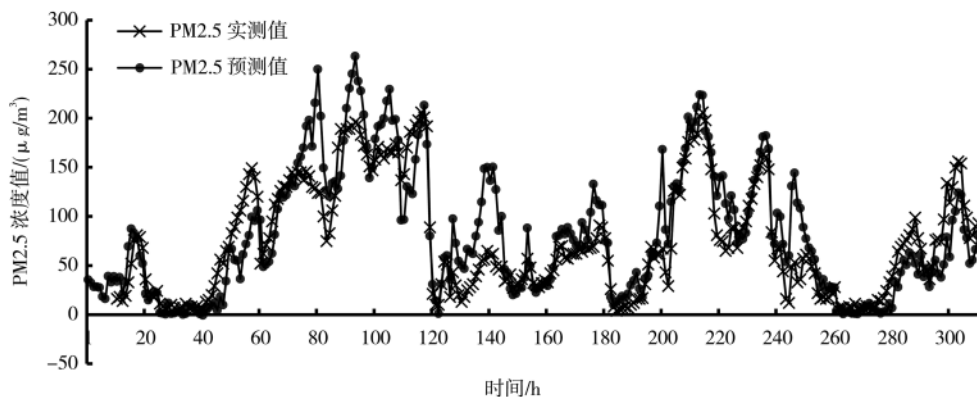


图 3 改进模型预测值与实际观察值对比图

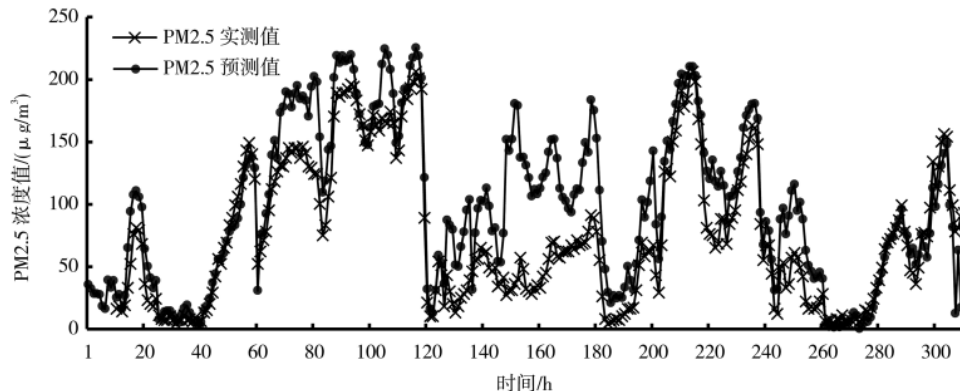


图 4 传统 BP 神经网络预测模型的预测值与实测值对比图

(下转第 92 页)

- [11] LU H, ZHU F, LIU Q, et al. Suppression of cable over-voltage in a high-speed electric multiple units system[J]. IEEE Transactions on Electromagnetic Compatibility, 2018, 61(2): 361-371.
- [12] RODRIGO M A, CASTRO H L C, MUNOZ F A. New clustering techniques based on current peak value, charge and energy calculations for separation of partial discharge sources[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2017, 24(1): 340-348.
- [13] LIU K, YI S, WANG G, et al. Passive target classification based on mode energy difference characteristic of the wavenumber spectrum[C]. 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2017.
- [14] ZHAO C, LI C, CAI W, et al. Research on AP clustering of weighted load index for demand response potential analysis[C]. 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), 2019.
- [15] 王宏, 赵菊敏, 李灯熬. 基于多特征融合的 J 波分类模型[J]. 电子技术应用, 2018, 44(11): 111-115.
- [16] MITICHE I, MORISON G, NESBITT A, et al. Classification of EMI discharge sources using time-frequency features and multi-class support vector machine[J]. Electric Power Systems Research, 2018, 163(PARTA): 261-269.

(下转第 100 页)

(上接第 85 页)

征的缺陷。实验结果证实: 3 种误差均有显著减小, 说明利用灰色预测原理具备发现事物历史变化规律性的特性来改进 BP 神经网络预测模型的性能, 提高了 PM_{2.5} 浓度预测精度, 并且得到的预测值曲线与实测值曲线拟合度更好。

本文采用按小时数据替代日均值数据, 提高了数据精度, 同时扩大了测试数据范围, 提高了神经网络模型泛化能力。由于传统采用日均值的数据量较少, 如果取得更多的数据进行分析, 则有可能会跨越季节, 而根据本文的分析季节对于 PM_{2.5} 浓度值的变化影响较大。因此选择按小时数据, 将数据范围控制在同一个季节环境背景下, 预测结果更好。

参考文献

- [1] 毛亚青, 胡展鹏, 俞啸, 等. 基于车载 GPS 终端的城市环境健康监测平台设计[J]. 电子技术应用, 2016, 42(5): 74-76, 79.
- [2] OPREA M, MIHALACHE S F, POPESCU M. Applying artificial neural networks to short-term PM_{2.5} forecasting modeling[C]. 12th IFIP International Conference on Artificial Intelligence Applications and Innovations(AIAI). Thessaloniki, Greece, 2016: 204-211.
- [3] 林承勇. 人工神经网络在预测 PM_{2.5} 浓度中的研究[D]. 成都: 电子科技大学, 2017.
- [4] FENG X, LI Q, ZHU Y J, et al. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation[J]. Atmospheric Environment, 2015, 107: 118-128.
- [5] GAO Y, WANG Z Y, LU Q C, et al. Prediction of vertical PM_{2.5} concentrations alongside an elevated expressway by using the neural network Hybrid model and generalized additive model[J]. Frontiers of Earth Science, 2017, 11(2): 347-360.
- [6] 张怡文, 敖希琴, 时培俊, 等. 基于 Pearson 相关指标的 BP 神经网络 PM_{2.5} 预测模型[J]. 青岛大学学报(自然科学版), 2017, 30(2): 83-87.
- [7] 李凤英, 白伟. 基于逐步回归分析的 T-S 模糊神经网络预测 PM_{2.5} 浓度研究[J]. 宁夏师范学院学报, 2016, 37(6): 5-13.
- [8] 马天成, 刘大铭, 李雪洁, 等. 基于改进型 PSO 的模糊神经网络 PM_{2.5} 浓度预测[J]. 计算机工程与设计, 2014, 35(9): 3258-3262.
- [9] 荆涛, 李霖, 于文柱, 等. t 分布受控遗传算法优化 BP 神经网络的 PM_{2.5} 质量浓度预测[J]. 中国环境监测, 2015(4): 105-110.
- [10] 裴雨潇, 刘奇洪, 丛小飞. 基于小波和过程神经网络的 PM_{2.5} 预测模型[J]. 环境科技, 2016, 148(2): 47-50.
- [11] 彭森, 连晓峰, 王昭洋, 等. 基于灰色预测和多目标决策的雾霾治理研究[C]. 全国仿真技术学术会议, 2015: 324-328.
- [12] 崔立志. 灰色预测技术及其应用研究[D]. 南京: 南京航空航天大学, 2010.
- [13] 谢乃明, 刘思峰. 离散 GM(1, 1) 模型与灰色预测模型建模机理[J]. 系统工程理论与实践, 2005, 25(1): 93-99.
- [14] 王岩, 黄张裕, 张玉爽, 等. 基于震荡序列的灰色预测模型在边坡沉降监测中的应用[J]. 测绘工程, 2017, 26(3): 69-71, 76.

(收稿日期: 2019-10-20)

作者简介:

黄鹰(1962-), 女, 硕士, 教授, 主要研究方向: 物联网与数据分析等。

史爱武(1972-), 通信作者, 男, 博士, 副教授, 主要研究方向: 大数据、云计算。

陈占龙(1980-), 男, 博士, 副教授, 主要研究方向: 地理空间信息。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所