

# 基于卡尔曼滤波融合算法的空气质量指数预测\*

郭利进<sup>1,2</sup> 井海明<sup>1,2#</sup> 南亚翔<sup>1,2</sup> 修春波<sup>1,2</sup>

(1.天津工业大学电气工程与自动化学院,天津 300387;

2.天津工业大学电工电能新技术天津市重点实验室,天津 300387)

**摘要** 分析了卡尔曼滤波算法的基本原理及其对空气质量指数(AQI)的预测机制。利用自回归滑动平均模型(ARMA)为卡尔曼滤波建立数学模型,提出了将径向基函数(RBF)神经网络融合于卡尔曼滤波,实现了新的融合算法对AQI进行预测。根据AQI时间序列的特点,建立了自回归预测模型,进而建立卡尔曼滤波的状态方程和测量方程。采用随机梯度逼近训练算法训练RBF神经网络,用RBF神经网络的输出作为卡尔曼滤波测量方程的观测值。仿真结果表明,融合了RBF神经网络后的卡尔曼滤波预测算法改善了单一方法预测滞后的现象,减小了误差,提高了预测精度。

**关键词** 卡尔曼滤波 空气质量指数 自回归滑动平均模型 径向基函数

DOI:10.15985/j.cnki.1001-3865.2017.04.009

**Prediction of air quality index based on Kalman filtering fusion algorithm** GUO Lijin<sup>1,2</sup>, JING Haiming<sup>1,2</sup>, NAN Yaxiang<sup>1,2</sup>, XIU Chunbo<sup>1,2</sup>. (1.College of Electrical Engineering and Automation, Tianjin Polytechnic University, Tianjin 300387; 2.Laboratory of New Technology about Energy and Electrical Engineering, Tianjin Polytechnic University, Tianjin 300387)

**Abstract:** The basic principle of the Kalman filtering algorithm and its prediction mechanism of air quality index (AQI) were analyzed. This paper using auto-regressive and moving average model (ARMA) to establish a mathematical model for Kalman filtering and put forward radial basis function (RBF) neural network merging with Kalman filtering to achieve a new fusion algorithm for AQI forecast. According to the characteristics of AQI time series, auto-regressive prediction model was established, then Kalman filtering state equation and measurement equation were established. The stochastic gradient approximation algorithm was used to train the RBF neural network, and the output of RBF neural network was used as the observation value of the Kalman filtering measurement equation. Simulation results showed that compared with a single method, the Kalman filtering prediction algorithm combined with RBF neural network had improved the lag phenomenon, reduced errors and raised the prediction accuracy.

**Keywords:** Kalman filtering; air quality index; auto-regressive and moving average model; radial basis function

近年来我国经济快速发展,导致环境污染程度越来越大,城市空气污染问题尤为严重,空气质量问题已经影响到我国民众的正常生活。我国环保部门对此十分重视,中国空气质量在线监测分析平台对空气质量指数(AQI)、PM<sub>10</sub>、PM<sub>2.5</sub>、SO<sub>2</sub>、NO<sub>2</sub>、O<sub>3</sub>、CO等数据进行实时监测并播报。然而,大气空间是一个多因素干扰、强滞后的庞大系统,且空气污染物浓度还受污染物排放、复杂的地理情况及气象条件等因素的影响<sup>[1]</sup>,因此AQI具有较强的非线性特性<sup>[2]</sup>。在解决非线性系统中,人工神经网络是非常实用的工具,应用十分广泛。反向传播(BP)神经网络能够准确描述AQI与影响因子之间的非线性特性关系<sup>[3-4]</sup>,针对PM<sub>10</sub>、SO<sub>2</sub>、CO、NO<sub>x</sub>等污染物浓

度的预测,可以通过三维空气质量模型,形象地表示各种污染物对AQI的影响<sup>[5-7]</sup>。但BP神经网络也存在一定的缺陷,如预测结果误差较大、风速变化极点时易产生振幅、训练时易陷入局部最小值等缺点。本研究对卡尔曼滤波融合算法预测AQI及融合的机制进行研究,利用自回归滑动平均模型(ARMA)建立数学预测模型,吸收多种方法的优点,提出径向基函数(RBF)神经网络预测数据融合在卡尔曼滤波的方法,提高了AQI预测精度。

## 1 卡尔曼滤波融合算法的AQI预测原理

### 1.1 卡尔曼滤波的基本原理

卡尔曼滤波实际是通过一种递推算法,利用状

第一作者:郭利进,男,1971年生,博士,教授,主要从事智能控制系统与数据挖掘等方面的研究。# 通讯作者。

\* 国家自然科学基金资助项目(No.61203302);天津市应用基础与前沿技术研究计划项目(No.14JCYBJC18900)。

态方程递推性以及最小均方差最优估计<sup>[8]</sup>,得到离散型观测数据,再对系统的观测数据进行线性化处理。卡尔曼滤波算法的前提是要建立线性模型,线性离散动态系统可以由状态方程(见式(1))和观测方程(见式(2))表示:

$$X(t) = \Phi(t, t-1) \cdot X(t-1) + \Gamma(t, t-1)w(t) \quad (1)$$

$$Z(t) = H(t)X(t) + v(t) \quad (2)$$

式中:  $X(t)$  为  $t$  时刻的状态向量;  $t$  为离散时间;  $\Phi(t, t-1)$  为  $t-1$  时刻到  $t$  时刻的状态转移矩阵;  $\Gamma(t, t-1)$  为  $t-1$  时刻到  $t$  时刻的激励转移矩阵;  $w(t)$  为系统噪声向量;  $Z(t)$  为  $t$  时刻的观测向量;  $H(t)$  为  $t$  时刻的预测输出转移矩阵;  $v(t)$  为观测噪声向量。

在卡尔曼滤波算法基本方程中,假设  $w(t)$  和  $v(t)$  为高斯白噪声,通常情况,可根据经验给协方差矩阵赋值,通过卡尔曼滤波算法递推预测方程,可以得到  $t$  时刻状态的最优估计值<sup>[9]</sup>。

### 1.2 建立线性离散动态系统模型

建立状态空间模型是系统获得预测结果的重要条件。结合 AQI 时间序列平稳连续的特点,本研究采用时间序列模型建立卡尔曼滤波所需要的观测方程和状态方程<sup>[10]</sup>。自回归滑动平均 (ARMA) 模型是研究时间序列的重要方法,由自回归模型 (AR 模型) 与滑动平均模型 (称 MA 模型) 为基础“混合”构成,应用十分广泛。ARMA 模型中,假设某变量  $t$  时刻的数值为  $x_t$ ,  $t$  时刻前有  $p$  个过去时间点,通过对过去时间点数值  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  进行线性回归,并与随机干扰噪声组成线性组合,获取  $x_t$  与过去值之间的表达式。ARMA 模型表达式如下:

$$x_t - \varphi_1 x_{t-1} - \dots - \varphi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3)$$

式中:  $\varphi_i (i=1, 2, \dots, p)$  为 AR 模型参数;  $a_t, a_{t-1}, \dots, a_{t-q}$  为平均值等于 0, 方差等于  $\sigma_a^2$  的不相关随机序列,称其为白噪声;  $\theta_j (j=1, 2, \dots, q)$  为 MA 模型参数;  $p, q$  分别为 AR 模型和 MA 模型的阶数。

### 1.3 RBF 神经网络获取观测值

神经网络系统在控制领域应用比较广泛,它具有能处理复杂情况、满足高设计要求和控制不确定因素等优势<sup>[11]</sup>。RBF 神经网络是一种前向型神经网络,网络由 3 层组成,即输入层、隐含层和输出层,结构简单且逼近能力强。

输入层为第 1 层,是由信号源节点组成;隐含层为第 2 层,单元数根据所描述的具体需要而定,如单

层或多层的非线性处理神经元;第 3 层为输出层,根据输入模式做出响应。

RBF 神经网络的基本思想:将径向基函数作为单元的“基”构成隐含层空间,将输入矢量不通过权连接,而是直接映射到隐含层空间,当径向基函数中心点确定后,映射关系也随之确定。

与 BP 神经网络相比, RBF 神经网络隐含层节点函数为径向函数,克服了 BP 神经网络训练时间长和计算复杂的不足。同时, RBF 神经网络具有良好的泛化能力、实时性好、更好的逼近能力和自适应能力。RBF 神经网络输出层满足映射  $f(x)$ , 方程式如(4)所示。

$$y = f(x) = \sum_{k=1}^N \omega_{lk} \eta_k(x - c_k) \quad l=1, 2, \dots, m \quad (4)$$

$$\eta(x) = \exp(-x^2/\sigma^2) \quad (5)$$

式中:  $y$  为映射函数输出向量;  $x$  为映射函数输入向量;  $\omega_{lk}$  为第  $l$  个隐含层节点到第  $k$  个输出层神经元的加权值;  $c_k$  为 RBF 神经网络中心点值;  $m$  为隐含层单元的个数;  $N$  为输出层神经元个数;  $\eta_k(x)$  为第  $k$  个输出层的转换函数;  $\sigma$  为基函数宽度参数。

本研究中把  $t$  时刻的 AQI 数据作为 RBF 神经网络的输出,把  $t-1 \sim t-p$  时刻的 AQI 数据作为 RBF 神经网络的输入,选取输出层神经元个数  $N=1$  的 RBF 结构模型,如图 1 所示。

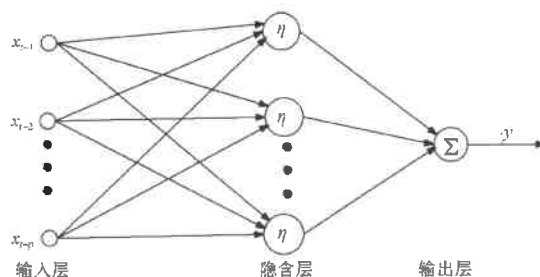


图 1 RBF 结构图

Fig.1 RBF structure diagram

本研究采用随机梯度训练算法,即用已知 AQI 数据作为训练样本集。随机梯度法在训练 RBF 神经网络时使输出层的加权值,中心点值和基函数宽度参数同步进行有监督学习。所谓监督学习是指分析训练样本数据,产生一个推断的功能,其可有效地获取神经网络输出层的加权值、中心点值和基函数宽度参数。

RBF 神经网络映射机制为:(1)非线性变换单元只对其中心点附近的输入敏感,距离中心点越远,非线性变换单元的输出减小,直至一个很小值,表现

为“局部敏感”;(2)非线性变换单元输出减小的速度由  $\sigma$  决定,  $\sigma$  越大,减小速度越慢,反之,则减小速度越快;(3)非线性变换单元经过输出层的幅值调整,来获得期望的输出。

单一算法满足不了对 AQI 预测结果的要求,本研究将 RBF 神经网络所获得的观测值代入到卡尔曼滤波方程中,形成了预测 AQI 的新方法,基于卡尔曼滤波的融合算法,实现高效预测。融合过程中,首先将 AQI 数据 1 代入到经过 ARMA 建模后的卡尔曼滤波  $\hat{X}(t-1|t-1)$  中, AQI 数据 2 经过 RBF 神经网络预测后赋值给观测方程  $Z(t)$ , 卡尔曼滤波将  $\hat{X}(t-1|t-1)$  和  $Z(t)$  进行融合,最后得出预测后的 AQI 数值  $\hat{X}(t|t)$ 。

1.4 AQI 预测误差指标

AQI 预测误差指标是衡量预测方法优劣的重要条件。在预测分析的过程中,可采用多种参数评价一个预测方法的精确度,如最大误差、均方根误差、平均绝对误差、平均相对误差等。在实际应用中,一个或两个参数很难保证预测的精确性,所以通常情况下,应该综合考虑多个参数,才能衡量预测的准确度。

2 AQI 预测的仿真和分析

2.1 数据来源

根据相关规定,地级及以上环境保护行政主管部门或其授权的环境监测站需向公众公开空气质量数据,包括 AQI、PM<sub>10</sub>、PM<sub>2.5</sub>、SO<sub>2</sub>、NO<sub>2</sub>、O<sub>3</sub>、CO 等数据,并进行时报和日报。本研究使用的数据来源于中国空气质量在线监测分析平台。随着纳入考核的污染物种类增多和计算方法复杂程度加大, AQI 指数受到更多因素的影响,具有较强的非线性特性。根据《环境空气质量指数(AQI)技术规定(试行)》(HJ 633—2012), AQI 等级划分如表 1 所示。

表 1 AQI 等级划分  
Table 1 The grading level of AQI

| 数值      | AQI 等级 | AQI 类别 |
|---------|--------|--------|
| 0~50    | I      | 优      |
| 51~100  | II     | 良      |
| 101~150 | III    | 轻度污染   |
| 151~200 | IV     | 中度污染   |
| 201~300 | V      | 重度污染   |
| >300    | VI     | 严重污染   |

本研究选取中国空气质量在线监测分析平台发布的天津市 2015 年 1—8 月份 AQI 数据作为 RBF 神经网络的训练样本,9 月份 AQI 数据作为 RBF 神经网络的预测性能测试样本,经测试, RBF 神经网络预测性能较好。由状态方程和观测方程可知,

AQI 预测精度为 1 h,根据  $t$  时刻之前的 AQI 实际值,获得  $t$  时刻的 AQI 预测值,形成连续预测。使用 2015 年 10 月 1 日起连续 1 000 h 的 AQI 作为实验样本,由于数据过多,影响仿真效果,因此取 10 月 1 日起前 300 h 的 AQI 数据进行仿真。

2.2 卡尔曼滤波算法预测 AQI

根据时间序列分析法得到了高阶自回归模型,确定系统状态方程参数。用卡尔曼滤波方法对 AQI 指数进行预测,预测结果如图 2 所示。

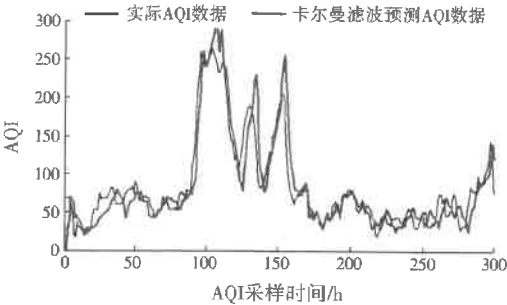


图 2 卡尔曼滤波的 AQI 预测结果  
Fig.2 The AQI prediction result of kalman filtering

由图 2 可见,采用 ARMA 建立模型的卡尔曼滤波算法对 AQI 进行预测,预测的结果基本反映出 AQI 的变化趋势,但存在较大的滞后延迟现象,致使预测结果的误差较大,并未达到预期的预测结果。

2.3 BP 神经网络预测 AQI

利用 BP 神经网络对 AQI 进行预测,得到的预测结果如图 3 所示。

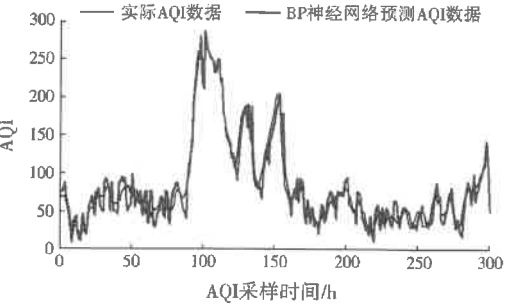


图 3 BP 神经网络的 AQI 预测结果  
Fig.3 The AQI prediction result of BP neural network

由图 3 可见, BP 神经网络算法在 AQI 预测中具有一定的预测能力和预测精度,在某些点处有较强的非线性逼近能力。但在 AQI 波动大的极点处影响了 BP 神经网络的泛化,产生振荡,致使预测效果和预测精度降低。

2.4 基于卡尔曼滤波的融合算法预测 AQI

将 RBF 神经网络的预测结果作为卡尔曼滤波方程的观测值,实现融合预测,如图 4 所示。

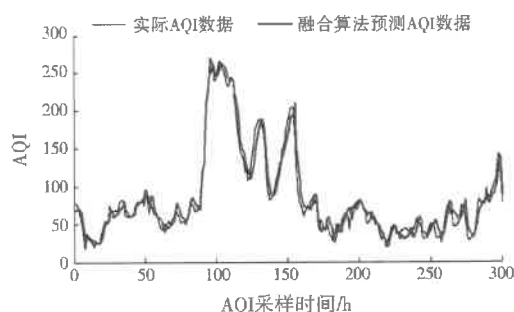


图4 融合算法的AQI预测结果

Fig.4 The AQI prediction result of fusion algorithm

由图4可见,利用卡尔曼滤波算法与ARMA建模和RBF神经网络算法进行融合,从而进一步优化预测结果,预测精度优于单一采用卡尔曼滤波算法和单一BP神经网络算法。融合算法与单一卡尔曼滤波算法预测结果相比,延迟现象得到极大地改善;与BP神经网络算法相比,在数据极值点的预测结果误差大大减小,预测效果有明显改善。不同方法预测误差指标如表2所示。

表2 不同方法预测性能指标

Table 2 The performance index of different predict methods

| 预测方法       | 最大误差 | 均方根误差 | 平均绝对误差 | 平均相对误差/% |
|------------|------|-------|--------|----------|
| BP神经网络     | 57.3 | 21.7  | 23.5   | 22.4     |
| ARMA 卡尔曼滤波 | 50.3 | 20.5  | 22.4   | 20.7     |
| 融合算法       | 31.3 | 11.2  | 13.1   | 12.9     |

通过对表2中各种AQI预测算法误差指标对比可知,BP神经网络算法和ARMA卡尔曼滤波算法对AQI的预测结果都有较大的误差,尤其是BP神经网络算法,误差较为明显。BP神经网络对历史信息的利用率和记忆能力有限,从而增加了神经元状态错误变化的机率。基于卡尔曼滤波的融合算法吸收了RBF神经网络良好的泛化能力、适应新数据能力强、学习速度快。该融合算法预测模型减少对真实数据的依赖,调整能力大大增强,提高了预测精度,减少了预测误差。

### 3 结论

通过对不同算法的预测机制进行分析,利用ARMA模型建立卡尔曼滤波的状态方程和观测方程,利用RBF神经网络的预测值作为卡尔曼滤波方程的观测值,用卡尔曼滤波算法将ARMA和RBF神经网络融合在一起,实现了AQI的融合预测。仿真实验结果验证了融合算法的有效性,减小单一预测方法造成的误差积累,并减少极值点较大波动数据点的误差。融合算法获取了两种算法的优点,预测

结果十分理想。

### 参考文献:

- [1] 白鹤鸣,沈润平,师华定,等.基于BP神经网络的空气污染指数预测模型研究[J].环境科学与技术,2013,36(3):186-189.
- [2] 李小飞,张明军,王圣杰,等.中国空气污染指数变化特征及影响因素分析[J].环境科学,2012,33(6):1936-1943.
- [3] 王俭,胡筱敏,郑龙熙,等.基于BP模型的大气污染预报方法的研究[J].环境科学研究,2002,15(5):62-64.
- [4] 白鹤鸣,沈润平,师华定,等.空气中PM<sub>10</sub>浓度的BP神经网络预报研究[J].环境研究与监测,2010,23(2):29-32.
- [5] YU Shaocai, MATHUR R, SARWAR G, et al. Eta-CMAQ air quality forecasts for O<sub>3</sub> and related species using three different photochemical mechanisms (CB4, CB05, SAPRC-99): comparisons with measurements during the 2004 ICARTT study[J]. Atmospheric Chemistry & Physics, 2010, 10(6):3001-3025.
- [6] EDER B, KANG Daiwen, RAO S T, et al. Using national air quality forecast guidance to develop local air quality index forecasts[J]. Bulletin of the American Meteorological Society, 2010, 91(3):313-326.
- [7] EDER B, KANG Daiwen, MATHUR R, et al. A performance evaluation of the national air quality forecast capability for the summer of 2007[J]. Atmospheric Environment, 2009, 43(14):2312-2320.
- [8] 张贤达.确定ARMA模型MA阶数的一种方法[J].自动化学报,1994,20(1):80-84.
- [9] 修春波,任晓,李艳晴,等.基于卡尔曼滤波的风速序列短期预测方法[J].电工技术学报,2014,29(2):253-259.
- [10] 王学林,周俊,姬长英,等.基于卡尔曼滤波的RBF神经网络和PD复合控制研究[J].计算机测量与控制,2009,17(8):1551-1553.
- [11] 刘进宝,丁涛.基于径向基函数神经网络的风速预测[J].太阳能报,2012,33(7):1131-1135.

编辑:丁怀 (收稿日期:2016-08-20)

(上接第387页)

- [19] HE Y T, CHEN C C, TRAINA S J. Inhibited Cr(VI) reduction by aqueous Fe(II) under hyperalkaline conditions[J]. Environmental Science & Technology, 2004, 38(21):5535-5539.
- [20] TZOU Y M, WANG Shangli, WANG M K. Fluorescent light induced Cr(VI) reduction by citrate in the presence of TiO<sub>2</sub> and ferric ions[J]. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2005, 253(1/2/3):15-22.
- [21] CRITTENDEN J C, ZHANG Yin, HAND D W, et al. Solar detoxification of fuel-contaminated groundwater using fixed-bed photocatalysts[J]. Water Environment Research, 1996, 68(3):270-277.
- [22] 孟凡生.中国铬渣污染场地土壤污染特征[J].环境污染与防治,2016,38(6):50-53.

编辑:丁怀 (收稿日期:2016-10-12)