

# 基于特征向量的最小二乘支持向量机 PM<sub>2.5</sub> 浓度预测模型

李 龙, 马 磊, 贺建峰\*, 邵党国, 易三莉, 相 艳, 刘立芳

(昆明理工大学 信息工程与自动化学院, 昆明 650500)

(\* 通信作者电子邮箱 jfenghe@kmust.edu.cn)

**摘 要:** 针对大气中细颗粒物(PM<sub>2.5</sub>)浓度预测的问题, 提出一种预测模型。首先, 通过引入综合气象指数综合考虑风力、湿度、温度等因素; 然后, 结合实际二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度、一氧化碳(CO)浓度和PM<sub>10</sub>浓度等构成特征向量; 最后, 利用特征向量和PM<sub>2.5</sub>浓度数据来建立最小二乘支持向量机(LS-SVM)预测模型。经2013年城市A和城市B环境监测中心的数据预测分析表明, 引入综合气象指数后预测的准确性提高, 误差降低近30%。说明该模型能够较为准确地预测PM<sub>2.5</sub>浓度, 并具有较高的泛化能力。此外还分析了PM<sub>2.5</sub>浓度与住院率、医院门诊量的关系, 发现了它们的高度相关性。

**关键词:** PM<sub>2.5</sub>浓度预测; 综合气象指数; 特征向量; 相关性分析; 最小二乘支持向量机

**中图分类号:** TP391.4; TP18 **文献标志码:** A

## PM<sub>2.5</sub> concentration prediction model of least squares support vector machine based on feature vector

LI Long, MA Lei, HE Jianfeng\*, SHAO Dangguo, YI Sanli, XIANG Yan, LIU Lifang

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming Yunnan 650500, China)

**Abstract:** To solve the problem of Fine Particulate Matter (PM<sub>2.5</sub>) concentration prediction, a PM<sub>2.5</sub> concentration prediction model was proposed. First, through introducing the comprehensive meteorological index, the factors of wind, humidity, temperature were comprehensively considered; then the feature vector was conducted by combining the actual concentration of SO<sub>2</sub>, NO<sub>2</sub>, CO and PM<sub>10</sub>; finally the Least Squares Support Vector Machine (LS-SVM) prediction model was built based on feature vector and PM<sub>2.5</sub> concentration data. The experimental results using the data from the city A and city B environmental monitoring centers in 2013 show that, the forecast accuracy is improved after the introduction of a comprehensive weather index, error is reduced by nearly 30%. The proposed model can more accurately predict the PM<sub>2.5</sub> concentration and it has a high generalization ability. Furthermore, the author analyzed the relationship between PM<sub>2.5</sub> concentration and the rate of hospitalization, hospital outpatient service amount, and found a high correlation between them.

**Key words:** Fine Particulate Matter (PM<sub>2.5</sub>) concentration prediction; comprehensive meteorological index; feature vector; correlation analysis; Least Squares Support Vector Machine (LS-SVM)

## 0 引言

细颗粒物(Fine Particulate Matter, PM<sub>2.5</sub>)是指大气中空气动力学当量直径小于等于2.5 μm的颗粒物,其数值越高,代表颗粒物浓度越高,意味着空气污染越严重。虽然PM<sub>2.5</sub>只是地球大气成分中含量很少的组成部分,但是它对空气质量和能见度等指标有重要影响。近期,我国多地出现雾霾天气,严重影响了人们的生活。二氧化硫、氮氧化物和PM<sub>2.5</sub>是雾霾的主要构成物质,PM<sub>2.5</sub>是其中加重雾霾污染天气的罪魁祸首,成为了影响人们正常生活的重要指标,因此准确预测PM<sub>2.5</sub>的浓度(浓度量纲为μg/m<sup>3</sup>,后文中提及的浓度均以此单位计)变得越来越重要。

影响PM<sub>2.5</sub>浓度的因素包括空气中二氧化硫(SO<sub>2</sub>)、二氧化氮(NO<sub>2</sub>)、一氧化碳(CO)、可吸入颗粒物(PM<sub>10</sub>)含量、臭氧(O<sub>3</sub>)和气象因素等。预测PM<sub>2.5</sub>浓度值变化规律对未

来空气质量监测有重要意义。

预测PM<sub>2.5</sub>的浓度,最重要的是分析各个影响因素与PM<sub>2.5</sub>之间的复杂关系。近年来,专家学者开展了一些相关的工作。秦侠等<sup>[1]</sup>提出了一种基于神经网络的方法预测大气污染物浓度;陈柳等<sup>[2]</sup>提出了一种基于支持向量机(Support Vector Machine, SVM)和时间序列的大气污染浓度预测模型;苏静芝等<sup>[3]</sup>提出了一种加入气象因素的人工神经网络的大气污染物浓度预测模型;陈俏等<sup>[4]</sup>提出了一种基于支持向量机和回归法的大气污染物浓度预测模型;魏振钢等<sup>[5]</sup>提出一种基于高斯模型的大气污染物浓度预测模型;王燕等<sup>[6]</sup>提出一种基于箱模型的大气污染物浓度预测模型。

这些文献通过研究个别因素对大气污染物浓度的影响提出了预测方法,但是气象因素对于PM<sub>2.5</sub>的影响是十分复杂的,实际情况中往往是不同气象因素相互影响的结果。如果分别考虑各个因素,则不能很好地体现多个因素相互作用对

收稿日期:2014-04-12;修回日期:2014-05-12。

基金项目:国家自然科学基金资助项目(11265007);教育部留学回国人员科研启动基金资助项目(2010-1561)。

作者简介:李龙(1988-),男,黑龙江大庆人,硕士研究生,主要研究方向:数据挖掘;马磊(1978-),男(回族),云南昆明人,讲师,硕士,主要研究方向:数据挖掘、生物信息学、医疗信息系统;贺建峰(1965-),男,云南昆明人,教授,主要研究方向:数据挖掘、医学图像处理。

PM<sub>2.5</sub> 浓度产生的耦合效应,也就不能准确建立预测 PM<sub>2.5</sub> 浓度的模型。

本文旨在用与 PM<sub>2.5</sub> 浓度相关性较强的因素组成气象特征向量,用对最小二乘支持向量机(Least Squares Support Vector Machine, LS-SVM)模型进行训练,得到准确度较高的预测 PM<sub>2.5</sub> 浓度模型。

通过对文献的分析,本文引入了综合气象指数这一指标,综合考虑了气象因素对 PM<sub>2.5</sub> 的影响<sup>[7]</sup>。同时从 SO<sub>2</sub> 浓度、NO<sub>2</sub> 浓度、PM<sub>10</sub> 浓度、CO 浓度、臭氧(O<sub>3</sub>)的 1 h 浓度、O<sub>3</sub> 的 8 h 浓度等因素中选取与 PM<sub>2.5</sub> 浓度相关性较强的因素,用这些因素与综合气象因素组成特征向量,用特征向量和历史 PM<sub>2.5</sub> 浓度数据训练 LS-SVM 模型,再用训练得到的模型来对 PM<sub>2.5</sub> 浓度进行预测。通过仿真研究探讨引入综合气象指数、LS-SVM 模型的预测效果,此外加入了 PM<sub>2.5</sub> 浓度与日门径量关系的讨论。

## 1 本文方法

### 1.1 数据选取

城市 A、B 和 C 均为我国重要的工业城市,雾霾天气时有发生。PM<sub>2.5</sub> 浓度长期持续较高,空气污染情况严重,属于我国 PM<sub>2.5</sub> 污染的典型城市。研究这些城市得出的结论具有一定的可推广性,故本文选取来自城市 A、B 和 C 的环境监测中心官方网站及该城市的医院的数据,进行相关问题的研究分析探讨。

### 1.2 综合气象指数

目前影响空气中 PM<sub>2.5</sub> 浓度的因素有很多,影响空气污染指数的主要因素可以归纳为 3 类:气象类、化学类和地理类。结合数据和实际情况分析,本文将空气湿度、风力、温度这 3 个因素与 PM<sub>2.5</sub> 的形成机制相结合,提出综合气象指数公式这一新的概念。综合气象指数的公式为:

$$D = 1.8 \times T + 0.55(1 - H) + 3.2W^{1/2} + 27 \quad (1)$$

其中:  $D$  为综合气象指数(无量纲);  $T$  为平均气温(单位为 °C);  $H$  为平均相对湿度(单位为 %);  $W$  为平均风速(单位为 m/s)。

本文用统计产品与服务解决方案(Statistical Product and Service Solutions, SPSS)统计软件分析最高温度、最低温度、平均温度、最大风速、平均风速、最大湿度、平均湿度、最小湿度、综合气象指数与 PM<sub>2.5</sub> 浓度的相关性。各因素与 PM<sub>2.5</sub> 浓度的相关性用相关系数表示,绝对值大的因素相关性大。可以看出本文提出的综合气象指数与 PM<sub>2.5</sub> 浓度的相关性最强,如表 1 所示。

表 1 各气象因素与 PM<sub>2.5</sub> 浓度的相关系数

气象因素	相关系数	气象因素	相关系数
平均湿度	-0.075 87	最大气温	-0.680 77
最大湿度	-0.066 63	平均气温	-0.720 73
最小湿度	-0.110 50	平均风速	-0.278 01
最小气温	-0.710 80	最大风速	-0.224 86
综合气象指数	-0.736 43		

### 1.3 特征向量选取

事实上,气象因素只是影响 PM<sub>2.5</sub> 浓度的一项重要因素,其他因素还应该包括二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度、一氧化碳(CO)浓度和 PM<sub>10</sub> 浓度等。因此,基

于对文献的分析,本文将各因素分别进行了量化计算并整理本文提出的综合气象指数、二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度等相关因素数据,最后选择相关性较强的因素构成特征向量。

为了能够准确地得到 PM<sub>2.5</sub> 浓度预测模型,必须要用可靠的相关性分析方法选准构成特征向量的因素,本文采用计算相关系数和绘制热度图两种方法来分析和验证各因素与 PM<sub>2.5</sub> 浓度相关性的大小。

相关系数又称皮(尔逊)氏积矩相关系数,该系数是说明两个因素之间相互关系密切程度的统计分析指标。相关系数用希腊字母  $\gamma$  表示,  $\gamma$  值的范围在 -1 和 +1 之间:  $\gamma > 0$  为正相关;  $\gamma < 0$  为负相关;  $\gamma = 0$  表示不相关,  $\gamma$  的绝对值越大,相关程度越高。

相关系数计算公式<sup>[8]</sup>为:

$$\gamma = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{\sum XY - \frac{\sum X \cdot \sum Y}{n}}{\sqrt{\left[ \sum X^2 - \frac{(\sum X)^2}{n} \right] \left[ \sum Y^2 - \frac{(\sum Y)^2}{n} \right]}} \quad (2)$$

本文运用 SPSS 统计软件对 PM<sub>2.5</sub> 浓度与二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度、一氧化碳(CO)浓度、PM<sub>10</sub> 浓度、O<sub>3</sub> 的 1 h 浓度、O<sub>3</sub> 的 8 h 浓度、综合气象指数等因素作相关性分析。如图 1 所示,可以看出:二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度、一氧化碳(CO)浓度、PM<sub>10</sub> 浓度与 PM<sub>2.5</sub> 浓度呈正相关;O<sub>3</sub> 的 1 h 浓度、O<sub>3</sub> 的 8 h 浓度、综合气象指数与 PM<sub>2.5</sub> 浓度呈负相关;其中综合气象指数、二氧化硫(SO<sub>2</sub>)浓度、一氧化碳(CO)浓度、一氧化碳(CO)浓度、PM<sub>10</sub> 浓度与 PM<sub>2.5</sub> 浓度相关性较强(相关系数绝对值大于 0.6)。

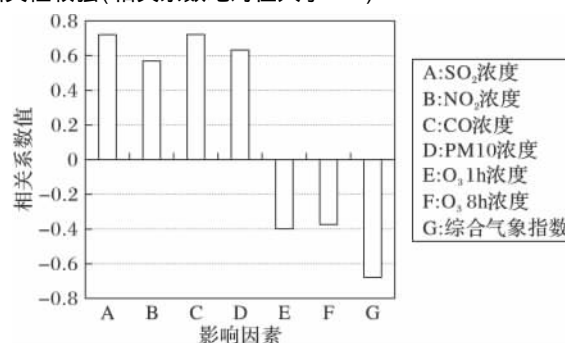


图 1 各影响因素与 PM<sub>2.5</sub> 浓度的相关性比较

热度图(Heat Map)是指将特征向量各个数据用不同颜色表示热度,从而体现个因素的相关性。热度图的好处在于直观地表现出了每一个特征与 PM<sub>2.5</sub> 的相关情况。不同颜色表示不同的相关性,其中浅色代表相关系数为正,即正相关;深色代表相关系数为负,即负相关;白色代表不相关;同时颜色越深表示相关性越强。

本文运用 R 软件对 PM<sub>2.5</sub> 浓度与二氧化硫(SO<sub>2</sub>)浓度、二氧化氮(NO<sub>2</sub>)浓度、一氧化碳(CO)浓度、PM<sub>10</sub> 浓度、O<sub>3</sub> 的 1 h 浓度、O<sub>3</sub> 的 8 h 浓度、综合气象指数等因素的相关性绘制了热度图(如图 2 所示)。在图 2(O 表示综合气象指数)中,

将样本中 PM2.5 浓度与其他因素相关系数的值按从高到低排序,以方便得到其与各因素之间的关系,用排序后的 PM2.5 浓度相关系数绘图得到颜色变化为从浅色到深色渐变的图像。可以看出综合气象指数的变化趋势较平滑,其热度图颜色呈现出从深色到浅色变化的趋势,即综合气象指数值高时,PM2.5 浓度低;综合气象指数值低时 PM2.5 浓度高;其与 PM2.5 成负相关关系。二氧化硫 (SO<sub>2</sub>) 浓度、二氧化氮 (NO<sub>2</sub>) 浓度、一氧化碳 (CO) 浓度、PM10 浓度变化趋势也较平滑,其热度图颜色呈现出从浅色到深色变化的趋势,即这几个指标值高时,PM2.5 浓度高;这几个指标值低时 PM2.5 浓度低,其与 PM2.5 成正相关关系。其他指标呈现出的相关性趋势不明显。

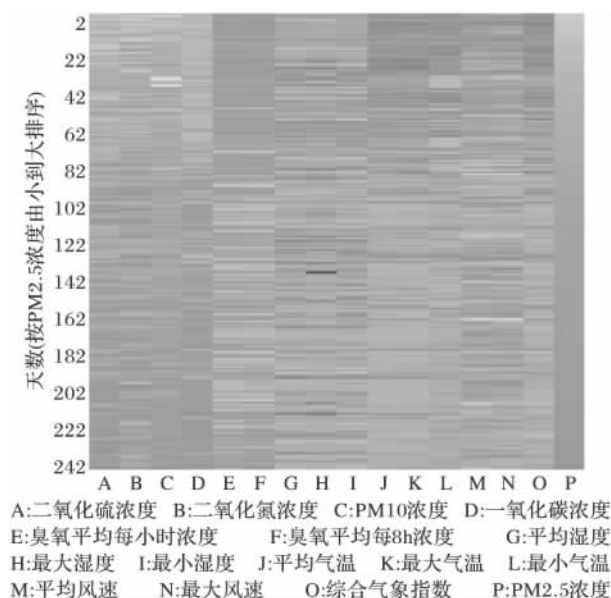


图2 各因素与 PM2.5 浓度的热度图

通过热度图分析得到结论与相关系数分析结论一致,相互验证结果显示了本文依据因素相关性选取出的特征向量的正确性。

综合以上分析,本文选择综合气象指数、二氧化硫 (SO<sub>2</sub>) 浓度、一氧化碳 (CO) 浓度、二氧化氮 (NO<sub>2</sub>) 浓度和 PM10 浓度构成特征向量。

#### 1.4 模型的建立

为了提高预测精度,选择合理的预测模型显得尤为重要。尽管研究表明用支持向量机 (Support Vector Machine, SVM) 进行大气污染物浓度的预测具有突出的优点,但是该方法仍然存在一些问题,特别是 SVM 中有若干不等式,计算比较复杂。通过筛选,本文采用 LS-SVM 模型,它将 SVM 中的不等式约束转换成等式约束,极大地简化了计算,提高了模型的训练效率<sup>[9]</sup>。LS-SVM 模型具有如下优势:1) 具有非线性和高维模式识别能力,可以识别影响 PM2.5 产生及演变的因素之间的复杂关系;2) 针对 PM2.5 浓度及其相关的监测数据较少的情况,可以发挥其在解决小样本问题上的优势;3) 泛化能力强,在模型训练完毕后,对于新的样本也能完成输入到输出的映射。

在 LS-SVM 模型算法中,输入空间的函数估计可转换为如下的最优问题:

$$\min \varphi = \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2$$

使得

$$y_i = w \cdot \varphi(x_i) + b + \xi_i \quad (3)$$

其中:  $w$  为权值向量,  $\xi_i$  为松弛变量,  $C$  为惩罚因子,  $b$  为偏差值,  $i$  为天数。

基于结构风险最小化原则确定决策函数参数  $w, b$ , 可等效为求解以下优化问题:

$$\min R = C \sum_{i=1}^n e_i^2 + \frac{1}{2} \|w\|^2 \quad (4)$$

其中:  $R$  为期望风险,  $e$  为模型对训练样本的预测误差。

在引入拉格朗日函数后,根据 Mercer 条件和最优化条件 KKT (Karush-Kuhn-Tucker) 可以得到矩阵方程:

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (5)$$

其中:  $y = [y_1, y_2, \dots, y_i]$ ,  $a = [a_1, a_2, \dots, a_i]$  为模型的参数,  $\Omega$  中的元素为本文数据样本中的特征向量的数据  $\Omega_{ij} = K(x_i, x_j)$  ( $i, j = 1, 2, \dots, l$ )。求解式 (5) 后得到 LS-SVM 的函数估计为:

$$f(x) = \sum_{i=1}^l a_i K(x_i, x_j) + b \quad (6)$$

其中  $K()$  为径向基函数 (Radial Basis Function, RBF)<sup>[10]</sup>。

最小二乘支持向量机求解线性方程,其满足极值条件,但不能保证是全局最优解,最小二乘支持向量机具有更快的求解速度,求解所需的计算资源较少,故本研究选择使用 LS-SVM 建立模型。

本文用特征向量和 PM2.5 浓度作为样本对 LS-SVM 模型进行训练,构造的训练样本形式为  $\{x_{i1}, x_{i2}, x_{i3}, x_{i4}, D_i, y_{i1}\}$ 。其中:  $x_{i1}, x_{i2}, x_{i3}, x_{i4}, D_i, y_{i1}$  分别代表二氧化硫 (SO<sub>2</sub>) 浓度、二氧化氮 (NO<sub>2</sub>) 浓度、一氧化碳 (CO) 浓度、PM10 浓度、综合气象指数和 PM2.5 浓度。

## 2 仿真结果与分析

### 2.1 模型预测结果对比

本研究通过查阅文献发现,秦霞等<sup>[1]</sup>的研究误差较低但是泛化能力较低,国外文献中大多数研究 PM2.5 浓度同各污染物浓度的定性关系和空间分布<sup>[11-13]</sup>,研究某一城市 PM2.5 浓度和其他污染物定量关系的文献较少,故本文选择在两城市两方法之间比较。本文选取城市 A<sup>[14]</sup> 环境监测站官方网站 2013-01-01—2013-09-19 的 PM2.5 浓度数据和中国天气网的气象数据。将所得数据分成训练组 (180 d) 和测试组 (32 d)。使用训练组数据对 LS-SVM 模型训练得到训练后的模型,再用训练后的模型计算测试数据,得到测试值 PM2.5 浓度,将得到的结果与测试组真实的 PM2.5 浓度对比,验证预测的准确度。

当构成特征向量的因素只选取二氧化硫 (SO<sub>2</sub>) 浓度、二氧化氮 (NO<sub>2</sub>) 浓度、一氧化碳 (CO) 浓度,不加入综合气象指数时,预测出的 PM2.5 浓度如图 3(a);加入综合气象指数因素后,预测出的 PM2.5 浓度如图 3(b)。将特征向量中有无综合气象指数的结果进行对比分析 (即改进前与改进后进行对比分析);改进前后的仿真效果对比如图 3 所示。

从图 3 来看,改进模型后的预测值走势更趋近于真实值,验证了使用改进后的特征向量 (加入了二氧化硫 (SO<sub>2</sub>) 浓度、二氧化氮 (NO<sub>2</sub>) 浓度、一氧化碳 (CO) 浓度等因素) 训练 LS-

SVM 模型能很好地反映 PM<sub>2.5</sub> 的产生及演变规律,说明该方式是有效的。

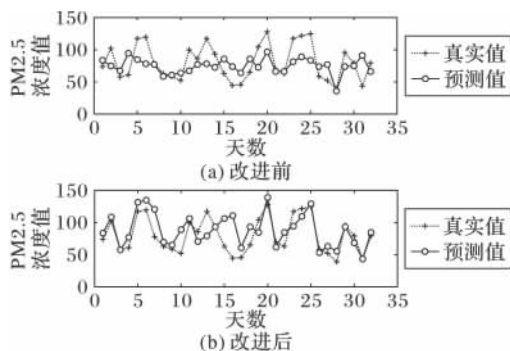


图 3 PM<sub>2.5</sub> 浓度预测模型效果对比(城市 A)

为了证明改进后的 LS-SVM 模型具有泛化能力,并进一步验证改进后的模型的有效性,本文选取与城市 A 地理位置、气候条件不同的城市 B<sup>[15]</sup> 再次进行 PM<sub>2.5</sub> 浓度预测,得出结果如图 4 所示。观察对比城市 B 预测曲线(图 4 中的实线)与实际情况(图 4 中的虚线)相符程度,发现对城市 B PM<sub>2.5</sub> 浓度的预测有效,说明基于特征向量的 PM<sub>2.5</sub> 浓度预测模型具有泛化能力。另外,图 4 将特征向量中有无综合气象指数的结果进行对比分析(即改进前与改进后进行对比分析)。

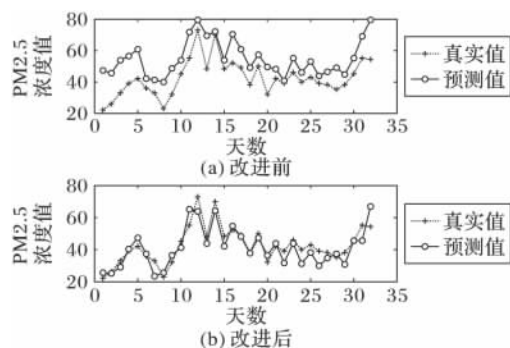


图 4 PM<sub>2.5</sub> 浓度预测模型效果对比(城市 B)

## 2.2 模型预测结果评价

本文使用均方误差(Mean Squared Error, MSE)和绝对误差(Absolute Error, ABS)对预测结果进行评价。

均方误差是衡量“平均误差”的一种较方便的方法,可以评价数据的变化程度。数理统计中均方误差是指参数估计值与参数真值之差平方的期望值,记为  $MSE$ 。 $MSE$  是衡量“平均误差”的一种较方便的方法, $MSE$  可以评价数据的变化程度, $MSE$  的值越小,说明预测模型描述实验数据具有更好的精确度。

$MSE$  的计算公式如下:

$$MSE = E[(y_i - \hat{y}_i)^2] \quad (7)$$

绝对误差为准确值  $x$  与其近似值  $x^*$  之差,记为  $e(x^*) = x^* - x$ ,简记为  $e^*$ 。但一般来说,不能准确知道  $e(x^*)$  的大小,可以通过测量或计算  $|e(x^*)| = |x^* - x| \leq \varepsilon(x^*)$  估计其绝对值的上界  $\varepsilon(x^*)$  叫作近似数  $x^*$  的绝对误差限,简称误差限,简记为  $\varepsilon^*$ 。

$ABS$  的计算公式如下:

$$ABS = |y_i - \hat{y}_i| \quad (8)$$

其中  $y_i$  和  $\hat{y}_i$  表示真实值和预测值。

选取城市 A 环境监测指标进行测试时,模型改进前后的  $MSE$  和  $ABS$  分别降低了 29% 和 27%。选取城市 B 环境监测指

标为测试数据时,模型改进前后的  $MSE$  和  $ABS$  分别降低了 68% 和 47%,如表 2 所示。

表 2 城市 A 和 B PM<sub>2.5</sub> 浓度预测模型改进前后误差对比

城市	参数	改进前	改进后	误差降低比例/%
A	$MSE$	624.4849	443.2745	29
	$ABS$	20.8579	15.2839	27
B	$MSE$	160.9934	50.9441	68
	$ABS$	10.9251	5.7826	47

从表 2 可看出,本文所采用的预测方法其预测精度高于原有的不采用特征向量的预测法。因此,采用本文方法可得到更好的预测结果,并且通过不同城市的测试说明该模型具有一定的泛化能力。

## 3 PM<sub>2.5</sub> 浓度与日门诊量、住院率的关系

由于 2013 年 11—12 月雾霾天气席卷我国部分地区,使得呼吸系统、心脑血管疾病的发病率明显上升。故本文收集城市 C<sup>[16]</sup> 某医院 2013-11-01—2013-11-30 医院每日门诊量、住院率数据和该地区该时间段的 PM<sub>2.5</sub> 浓度数据,使用 Excel 软件将其归一化处理后,与 PM<sub>2.5</sub> 浓度与日门诊量、住院率分别进行比对,发现 PM<sub>2.5</sub> 浓度与日门诊量、住院率曲线走势相趋近,如图 5~6 所示。其中:住院率是归一化后的住院率,其值超过 1 并不代表住院率超过 100%。经观察易知,PM<sub>2.5</sub> 与疾病发生的相关性极大<sup>[17-18]</sup>。

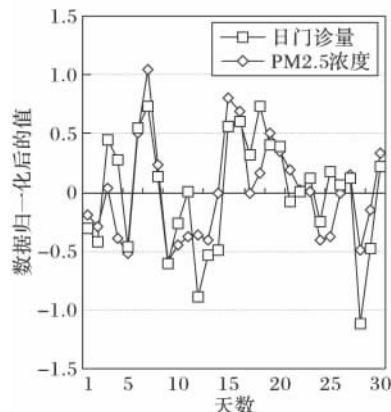


图 5 PM<sub>2.5</sub> 浓度与日门诊量(2013-11-01—11-30)的关系

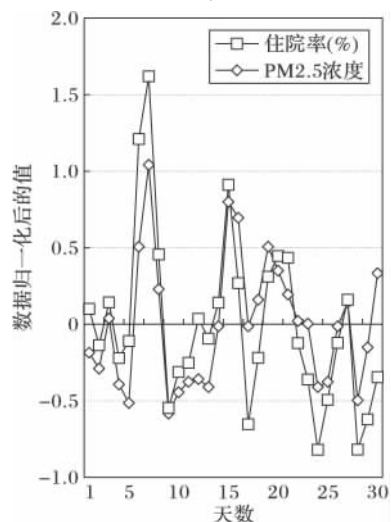


图 6 PM<sub>2.5</sub> 浓度与住院率的关系(2013-11-01—11-30)

虽然 LS-SVM 模型能够建立 PM<sub>2.5</sub> 浓度的以特征向量为调控因子的关系链——“特征向量→PM<sub>2.5</sub> 浓度”,但是该模型所构建的 PM<sub>2.5</sub> 关系链只能反映 PM<sub>2.5</sub> 形成的过程,没有涉及 PM<sub>2.5</sub> 与人类健康(如住院率、日门诊量)等后续问题的相关关系。事实上,PM<sub>2.5</sub> 会对人类健康产生严重影响,已经成为不争的事实。下一步工作将会在本文基础上,对 PM<sub>2.5</sub> 的浓度变化给医院日门诊量、病人住院率带来怎样的影响进行更进一步的定量研究,进而实现从海量气象数据、环境数据和医疗数据中挖掘出它们内在的联系和实用信息的目的。

#### 4 结语

污染物浓度和气象因素一直是影响 PM<sub>2.5</sub> 预测的主要因素,分别分析单一气象因素的影响,不能很好地体现出多个气象因素产生的耦合效果对气象因素的影响规律。本文引入综合气象因素这一指标,在综合考虑 SO<sub>2</sub> 浓度、NO<sub>2</sub> 浓度、CO 浓度和 PM<sub>10</sub> 浓度的基础上,将其与综合气象因素组合构成特征向量,使用特征向量和 PM<sub>2.5</sub> 浓度值,建立 LS-SVM 模型。通过城市 A 和城市 B 环境监测站的数据研究表明,使用本文方法较为合理,同时预测精度较高,根据测试,预测结果比较逼近真实数据,模型泛化能力较强。

另外,本文结合实际情况,对 PM<sub>2.5</sub> 与医院门诊量等相关指标进行了定性的分析,发现它们有高度的相关性,这为下一步的通过 PM<sub>2.5</sub> 浓度预测日门诊量等工作奠定了一些基础。

虽然本文设计的模型预测效果较好,但是由于 PM<sub>2.5</sub> 的形成机制十分复杂,影响 PM<sub>2.5</sub> 浓度的未知因素仍然存在,这使得本文模型的预测数据与真实数据还有一定差距,仍存在改进与提高的空间。

#### 参考文献:

- [1] QIN X, LEI L, YAO X. Methods to improve the generalization of BP neural network applied in air pollution forecasting [J]. Journal of Beijing University of Technology, 2007, 33(8): 849–852. (秦侠, 雷蕾, 姚小丽. 大气污染预测中提高 BP 网络泛化能力的方法 [J]. 北京工业大学学报, 2007, 33(8): 849–852.)
- [2] CHEN L, WU D, CHEN Q. Wavelet analysis and support vector machine used in predicting atmospheric pollution [J]. Xi'an University of Science and Technology, 2010, 30(6): 726–730. (陈柳, 吴冬梅, 陈俏. 小波分析及支持向量机应用于大气污染预测 [J]. 西安科技大学学报, 2010, 30(6): 726–730.)
- [3] SU J, QIN X, LEI L, et al. Study of neural networks in air pollution forecasting application [J]. Sichuan Environment, 2008, 27(2): 98–101. (苏静芝, 秦侠, 雷蕾, 等. 神经网络在空气污染预报中的应用研究 [J]. 四川环境, 2008, 27(2): 98–101.)
- [4] CHEN Q, CAO G, CHEN L. SVM applied to predict the concentration of atmospheric pollutants [J]. Computer Technology and Development, 2010, 20(1): 250–252. (陈俏, 曹根牛, 陈柳. 支持向量机应用于大气污染物浓度预测 [J]. 计算机技术与发展, 2010, 20(1): 250–252.)
- [5] WEI Z, GUO Z, ZHANG L, et al. The application of an air pollution model based on the Gauss mode [J]. China Ocean University: Natural Science, 2008, 38(2): 327–330. (魏振钢, 郭遵强, 张琳, 等. 基于高斯模式的大气污染模型的应用 [J]. 中国海洋大学学报: 自然科学版, 2008, 38(2): 327–330.)
- [6] WANG Y, DONG L, CHEN J. The study of three-dimensional multi-box model of prediction of atmospheric environment [C]// Proceedings of the 2007 China Environmental Science Society Annual Conference. Beijing: China Environmental Science Press, 2007: 621–625. (王燕, 董丽, 陈剑. 三维多箱模型预测大气环境的研究 [C]// 2007 中国环境科学学会学术年会优秀论文集. 北京: 中国环境科学出版社, 2007: 621–625.)
- [7] CHARRON A, HARRISON R. Fine (PM<sub>2.5</sub>) and coarse (PM<sub>2.5-10</sub>) particulate matter on a heavily trafficked London highway: sources and processes [J]. Environmental Science and Technology, 2005, 39(20): 7768–7776.
- [8] YAN L. Correlation coefficient and partial correlation coefficient in the relevant analysis [J]. Journal of Yunnan Institute of Finance, 2003, 19(3): 78–80. (严丽坤. 相关系数与偏相关系数在相关分析中的应用 [J]. 云南财贸学院学报, 2003, 19(3): 78–80.)
- [9] HE H, ZHOU X, ZENG J. Short-term load forecasting simplified LS-SVM model and implementation [J]. Computer Simulation, 2011, 28(1): 302–306. (贺红林, 周翔, 曾劲松. 短期负荷预测的简化 LS-SVM 模型及实现 [J]. 计算机仿真, 2011, 28(1): 302–306.)
- [10] YAN W, SHAO H. Support vector machines and comparison and application of least squares support vector machine [J]. Control and Decision, 2003, 18(3): 358–360. (阎威武, 邵惠鹤. 支持向量机和最小二乘支持向量机的比较及应用研究 [J]. 控制与决策, 2003, 18(3): 358–360.)
- [11] LIU Y, PACIOREK C, KOUTRAKIS P. Estimating regional spatial and temporal variability of PM<sub>2.5</sub> concentrations using satellite data, meteorology, and land use information [J]. Environmental Health Perspectives, 2009, 117(6): 886–892.
- [12] TURPIN B, LIM H. Species contributions to PM<sub>2.5</sub> mass concentrations: revisiting common assumptions for estimating organic mass [J]. Aerosol Science and Technology, 2001, 35(1): 602–610.
- [13] McKENDRY I. Evaluation of artificial neural networks for fine particulate pollution (PM<sub>10</sub> and PM<sub>2.5</sub>) forecasting [J]. Journal of the Air and Waste Management Association, 2002, 52(9): 1096–1101.
- [14] Xi'an Environmental Monitoring Station. Daily air quality [EB/OL]. (2013-09-20) [2014-01-05]. <http://www.xianemc.gov.cn/>. (西安市环境监测站. 空气质量日报 [EB/OL]. (2013-09-20) [2014-01-05]. <http://www.xianemc.gov.cn/>.)
- [15] Wuhan Environmental Monitoring Station. Daily air quality [EB/OL]. (2013-09-20) [2014-01-05]. <http://www.whemc.cn/>. (武汉市环境监测站. 空气质量日报 [EB/OL]. (2013-09-20) [2014-01-05]. <http://www.whemc.cn/>.)
- [16] Ningbo Environmental Monitoring Station. Daily air quality [EB/OL]. (2013-12-20) [2014-01-05]. <http://www.nbemc.gov.cn/>. (宁波市环境监测站. 空气质量日报 [EB/OL]. (2013-12-20) [2014-01-05]. <http://www.nbemc.gov.cn/>.)
- [17] SHANG Y, SUN Z, CAO J, et al. Systematic review of Chinese studies of short-term exposure to air pollution and daily mortality [J]. Environment International, 2013, 54: 100–111.
- [18] WANG D, WANG B, BAI Z. PM<sub>2.5</sub> pollution and daily mortality residents meta analysis of the relationship [J]. Journal of Environment and Health, 2012, 29(6): 529–532. (王德庆, 王宝庆, 白志鹏. PM<sub>2.5</sub> 污染与居民每日死亡率关系的 Meta 分析 [J]. 环境与健康杂志, 2012, 29(6): 529–532.)