

# 基于模糊时序和支持向量机的高速公路 SO<sub>2</sub> 浓度预测算法<sup>①</sup>

岳鹏程, 张林梁, 马阅军

(山西省交通科学研究院, 太原 030006)

**摘要:** 针对现有 SO<sub>2</sub> 浓度预测方法中存在的污染物来源和影响因素认识不统一、小样本数据敏感、易于陷入局部最优等问题, 文中提出了基于模糊时序和支持向量机的高速公路 SO<sub>2</sub> 浓度预测算法, 为搭建高速公路环境健康监测系统提供了可靠的理论支持. 该方法依据 SO<sub>2</sub> 浓度的季节变动规律, 以季节作为时间序列, 以 24h 为粒化窗宽, 通过高斯核函数提取原始样本数据的特征值, 输入支持向量机训练模型, 并利用 k 重交叉验证法结合网格划分优化模型参数. 文中应用该方法建立了 SO<sub>2</sub> 浓度预测模型, 并以 2014 年 4 月至 2015 年 3 月山西省太旧高速公路某监测点 SO<sub>2</sub> 小时浓度监测值为样本数据, 在 MATLAB 平台下应用 LIBSVM 工具实现了计算过程. 结果表明, 基于模糊时序和支持向量机的高速公路 SO<sub>2</sub> 浓度预测算法不受机理性理论研究的限制, 支持小样本学习, 非线性拟合效果好, 泛化能力强.

**关键词:** 二氧化硫浓度; 支持向量机; 模糊时序; 高速公路; 预测算法

## Prediction for SO<sub>2</sub> Concentration Based on the Fuzzy Time Series and Support Vector Machine (SVM) on Expressway

YUE Peng-Cheng, ZHANG Lin-Liang, MA Yue-Jun

(Shanxi Transportation Research Institute, Taiyuan 030006, China)

**Abstract:** The present prediction methods for SO<sub>2</sub> concentration suffer from the disadvantages that there is no uniform understanding of pollutant sources and influencing factors, small sample data is sensitive, and prediction methods are easy to fall into local optimum etc. In order to solve these problems, a method for the prediction of SO<sub>2</sub> concentrations on expressway is proposed which is based on fuzzy time series and support vector machine (SVM), and provides a reliable theoretical support for building the highway environmental health monitoring system. Based on the seasonal variation of SO<sub>2</sub> concentrations, the method takes the season as time series, 24h for graining window width. Through the Gaussian kernel function to extract the eigenvalues of the original sample data, which are input support vector machine (SVM) model for training, and k-fold cross validation method combined with the grid division is used to optimize model parameters. Finally, a SO<sub>2</sub> concentrations prediction model is established with the method in this paper. By using 1h average SO<sub>2</sub> concentrations as sample data which are obtained by Shanxi taijiu expressway monitoring station from April 2014 to March 2015, the LIBSVM tool is used to realize the calculation process on the MATLAB platform. The results show that based on fuzzy time series and support vector machine (SVM), the forecasting methods of SO<sub>2</sub> concentration is not restricted by the research of machine rational theory, and supports small-sample learning, otherwise, the nonlinear fitting effect is perfect, and the ability of generalization is well.

**Key words:** SO<sub>2</sub> concentrations; support vector machine; fuzzy time series; expressway; prediction

研究表明, 大气中 SO<sub>2</sub> 气体浓度高于 0.5mg/m<sup>3</sup> 时开始对人体健康产生影响; 当 SO<sub>2</sub> 气体浓度高于 1mg/m<sup>3</sup> 时多数人感到上呼吸道和眼睛刺激; 当 SO<sub>2</sub> 气体浓度达到 400mg/m<sup>3</sup> 时会造成肺水肿及呼吸中

<sup>①</sup> 基金项目: 山西省交通厅科技项目(2015-1-25); 山西省基础研究项目(2015021126)

收稿时间: 2016-09-29; 收到修改稿时间: 2016-10-31 [doi:10.15888/j.cnki.csa.005795]

枢麻痹,直接危及人的生命<sup>[1]</sup>.大气中烟尘含有多种重金属及其氧化物,它们能催化SO<sub>2</sub>与之发生协同作用,形成毒性更强、破坏力更大的硫酸雾,日本四日市哮喘事件、伦敦烟雾事件等表明这种协同作用的严重危害性<sup>[2]</sup>.由此可见,大气中的SO<sub>2</sub>污染气体直接威胁着人类健康,如果能够有效预测大气中SO<sub>2</sub>浓度的变化和发展趋势,将有助于科学探索控制环境污染、改善环境质量的方法.国内外学者在这方面进行了许多尝试,提出了一些预测模型<sup>[3-5]</sup>.多因素气体浓度预测法,依赖于气体产生、扩散等机理性研究,以污染物质量浓度受某些影响因素影响为前提,把影响因素作为自变量,污染物浓度作为因变量,对污染物浓度与影响因素进行相关性分析,是一种应用最广泛的大气污染物浓度预测方法.但是,对于SO<sub>2</sub>浓度预测来说,这种方法存在诸多不确定性.首先,经研究发现的大气污染物SO<sub>2</sub>浓度变化的影响因素有污染物的排放量、污染源的布局、类型和气象条件以及燃料结构等,而实际上影响SO<sub>2</sub>浓度变化的因素还有很多,他们有的尚未被发现,有的则还不能确定;其次,即使假设已发现全部的影响因素,影响因素之间尚存在着因果关系、模糊性和耦合性,造成多因素回归分析法难以准确分析因素与因素之间以及因素与SO<sub>2</sub>浓度之间的相关性.对此,可充分利用机器学习算法解决复杂的多元非线性回归问题.

近年来迅速发展的人工神经网络,是典型的基于机器学习算法的建模方法,被广泛应用于大气污染物浓度预测领域<sup>[6,7]</sup>.已有的应用神经网络模型对SO<sub>2</sub>浓度进行预测的研究有:王丽梅等人利用B-P神经网络和西南某市近20年的样本数据,建立了城市大气二氧化硫浓度预测模型,对该市二氧化硫浓度作出了合理预测<sup>[8]</sup>;陈柳等人利用小波分解和重构的方法把一年的二氧化硫浓度变化划分为1月到9月和10月到12月两段,并分段建立神经网络预测模型,提高了模型预测精度<sup>[9]</sup>;姚宁等人考虑了气温、湿度、降水量等气象因素对大气污染物浓度的影响,建立起大气污染物浓度受气象因素影响的B-P神经网络模型,模型能够更加真实地反映了大气污染物浓度变化<sup>[10]</sup>.这些研究均取得了较好的效果.神经网络具有人脑思维的特点和自学习、自适应及自组织的功能,对处理任意复杂的非线性关系,具有适应性强、客观性好的优点,但也往往存在网络训练过拟合和局部极小等问题.支持向

量机(Support Vector Machine, SVM)是将统计学习理论与结构风险最小化原则相结合的一种机器学习技术,当样本数较少、信息量有限时,可以折中取舍训练精度和学习能力,提高泛化能力,在解决小样本、非线性等问题中具有其他模型不具备的优势,可有效避免神经网络中经常出现的过拟合及局部极小等问题,成为目前大气污染物浓度预测领域的研究热点<sup>[11-13]</sup>.鉴于对目前SO<sub>2</sub>浓度预测方法的对比分析,本文提出基于模糊时序和支持向量机的预测二氧化硫浓度变化趋势的新方法,利用模糊时序提取样本数据的特征值,作为支持向量机的输入,提高模型精度,解除了气体机理性研究对预测模型的限制,输入支持向量机后,再利用高斯核函数映射到高维空间解决非线性拟合问题,新方法能有效解决气体浓度预测中的多元非线性拟合问题,是一种对SO<sub>2</sub>气体污染物浓度预测的新尝试.

## 1 支持向量机

近年来,在环境污染和空气质量预测等领域已经开始有通过支持向量机进行回归分析或拟合预测的研究.本文主要是预测大气污染物SO<sub>2</sub>的浓度变化,基于应用了高斯核函数的支持向量机完成浓度样本数据从低维空间向高维空间的映射,实现非线性关系向线性关系的过渡,从而很好的解决了目前大气污染物SO<sub>2</sub>的浓度预测时存在的非线性内在依赖问题,即:

$$\Phi: X \rightarrow H \quad x \rightarrow \varphi(x) \quad (1)$$

其中,H代表高维空间, $\varphi(x)$ 代表映射函数,若存在

$$K(x, x') = (\varphi(x), \varphi(x')) \quad (2)$$

其中, $K(x, x')$ 代表核函数.根据Cover定理可知,低维空间的数据集通常线性不可分,将其映射到高维空间,线性可分相对容易.通过核函数 $K(x, x')$ 间接达到 $x \rightarrow \varphi(x)$ 的映射,以此将低维空间数据映射至高维空间,实现线性变换.

基于上述方法,利用核函数进行高维映射后线性计算,得到决策函数:

$$f(x) = \text{sign}((\omega, x) + b) \\ = \text{sign}(\sum_j y_j \alpha_j K(x, x_j) + b) \quad (3)$$

$$\omega = \sum_j \alpha_j K(x, x_j) \quad (4)$$

其中, $\omega$ 代表权值向量, $b$ 为阈值, $\alpha_i$ 代表拉格朗日乘子,经过高维映射之后,线性SVM的主要问题变为:

$$\begin{cases} \min J(\omega, b) = \frac{1}{2} \omega^T \omega \\ \text{s.t. } y_j (\omega^T \varphi(x_j) + b) \geq 1 \end{cases} \quad (5)$$

公式(5)对应的对偶形式为:

$$\begin{cases} \max W(\mathbf{a}) = \mathbf{e}^T \mathbf{a} - \frac{1}{2} \mathbf{a}^T \mathbf{Q} \mathbf{a} \\ \text{s.t. } \mathbf{y}^T \mathbf{a} = 0 \quad \alpha_j \geq 0, \forall j \end{cases} \quad (6)$$

式(6)中  $\mathbf{e}$  为单位向量,  $\mathbf{Q}$  为  $l \times l$  矩阵, 且  $Q_{ij} = y_i y_j K(x_i, x_j)$ .

## 2 时间序列的模糊粒化

### 2.1 模糊粒化高斯模型的建立

信息粒化是 1979 年由 L.AZadeh 教授提出的, 以相似性对象(以组的形式)作为整体或对整体对象进行有效分割的一种研究方法. 信息粒化的粒化方式包括非模糊的粒化方式(c-粒化)和模糊的粒化方式(f-粒化)两种<sup>[14-15]</sup>.

本文采用模糊信息粒化的粒化方式, 将样本数据根据时间序列分割成若干子序列(又称操作窗口), 并根据粒化产生的信息窗口进行模糊化, 生成多个模糊集(即模糊粒子), 在模糊信息粒化过程中, 模糊化的好坏度影响最后粒化的最终结果. 常用的模糊粒子主要有梯型、三角型、抛物型、高斯型等, 文中建模时采用非对称式高斯模糊模型. 其隶属函数和图像如下:

$$A(x, m, \sigma, \mu) = \begin{cases} \exp[-(x-m)/\sigma^2], & x \leq m \\ \exp[-(x-m)/\mu^2], & x > m \end{cases} \quad (7)$$

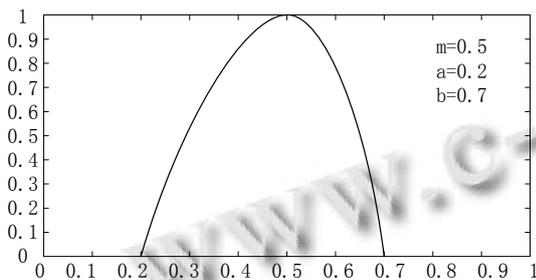


图1 非对称式高斯模糊集

### 2.2 SO<sub>2</sub> 浓度变化的时间序列分析

时间序列分析是指利用一组按时间顺序的数字序列(时间序列), 分析随机数据序列所反映出的统计规律, 用来解决实际问题. 通常一个时间序列会有趋势、季节变动、循环波动和不规则波动 4 种要素. 研究发现, 二氧化硫浓度变化具备季节变动和日循环波动两个时间序列要素<sup>[16]</sup>.

考虑山西省 2014 年到 2015 年的气象条件, 依据

传统的天文季节划分法将 1-3 月作为春季, 4-6 月作为夏季、7-9 月作为秋季、10-12 月作为冬季, 分季节研究 2014 年 4 月到 2015 年 3 月太旧高速公路阳泉服务区北区监测点的 SO<sub>2</sub> 浓度小时监测值的统计规律. 原始监测数据全部来自于山西省交通环境监测中心已建成的山西省交通运输环境监测信息平台, 统计结果如图 2 所示. 图 2 中直线表示 SO<sub>2</sub> 浓度监测值的变化范围, 直线的上端点表示最大值, 下端点表示最小值; 矩形表示 SO<sub>2</sub> 浓度变化的离散程度, 矩形上边缘表示 75%分位值, 下边缘表示 25%分位值, 矩形内方点表示平均值, 矩形内横线表示中值.

由图 2 可以看出 SO<sub>2</sub> 浓度监测值季节性差别明显:

(1)从直线表示的变化范围可以看出 SO<sub>2</sub> 浓度最大值春季最大, 夏季最小, 春季浓度最大值约为夏季浓度最大值的 20 倍, 而四个季节最小值基本相等, 所以春季浓度变化范围最大, 夏季最小; (2)从矩形表示的离散程度可以看出夏季和秋季 SO<sub>2</sub> 浓度变化幅度较小, 25%和 75%两个浓度分位值的差值在 3-6 $\mu\text{g}\cdot\text{m}^{-3}$  之间波动, 而春季和冬季的浓度变化幅度远大于夏季和秋季, 在 50-180 $\mu\text{g}\cdot\text{m}^{-3}$  之间波动; (3)由矩形内方点表示的平均值和横线表示的中值的相对位置来看, 夏季和秋季的中值与平均基本相等, SO<sub>2</sub> 浓度监测值两端分布均匀, 而春季和冬季的中值低于平均值, 监测数据整体呈葫芦样分布, 上端小下端大, 两端分布不均匀; (4)由矩形内方点表示的平均值可以看出, SO<sub>2</sub> 浓度每日上午(7:00-9:00)出现波峰、下午(14:00-16:00)出现波谷, 整体呈现类正弦曲线的循环波动. 这些变化都是由季节变动、气象条件变化、高速公路汽车通行量、车辆尾气排放情况以及其他不可控因素共同决定的.

在进行大气环境污染物浓度预测时, 预测多长时间段的变化趋势至少得需要相当时间长度的样本数据, 预测才能可靠. 研究以 2014 年 4 月至 2015 年 3 月山西省太旧高速公路阳泉服务区北区监测点一年的监测样本为数据基础, 分析 SO<sub>2</sub> 浓度的季节性周期变化规律. 在划分浓度分析的时间序列和决定粒化窗宽时, 考虑监测样本的最小采集周期为 1h 浓度值, 时间跨度为 1 年, 研究目的是 SO<sub>2</sub> 气体浓度的季节性周期变化规律, 以四个季节作为时间序列, 以天(24h)为时间子序列和粒化窗宽, 建立基于模糊时序的 SO<sub>2</sub> 浓度预测模型.

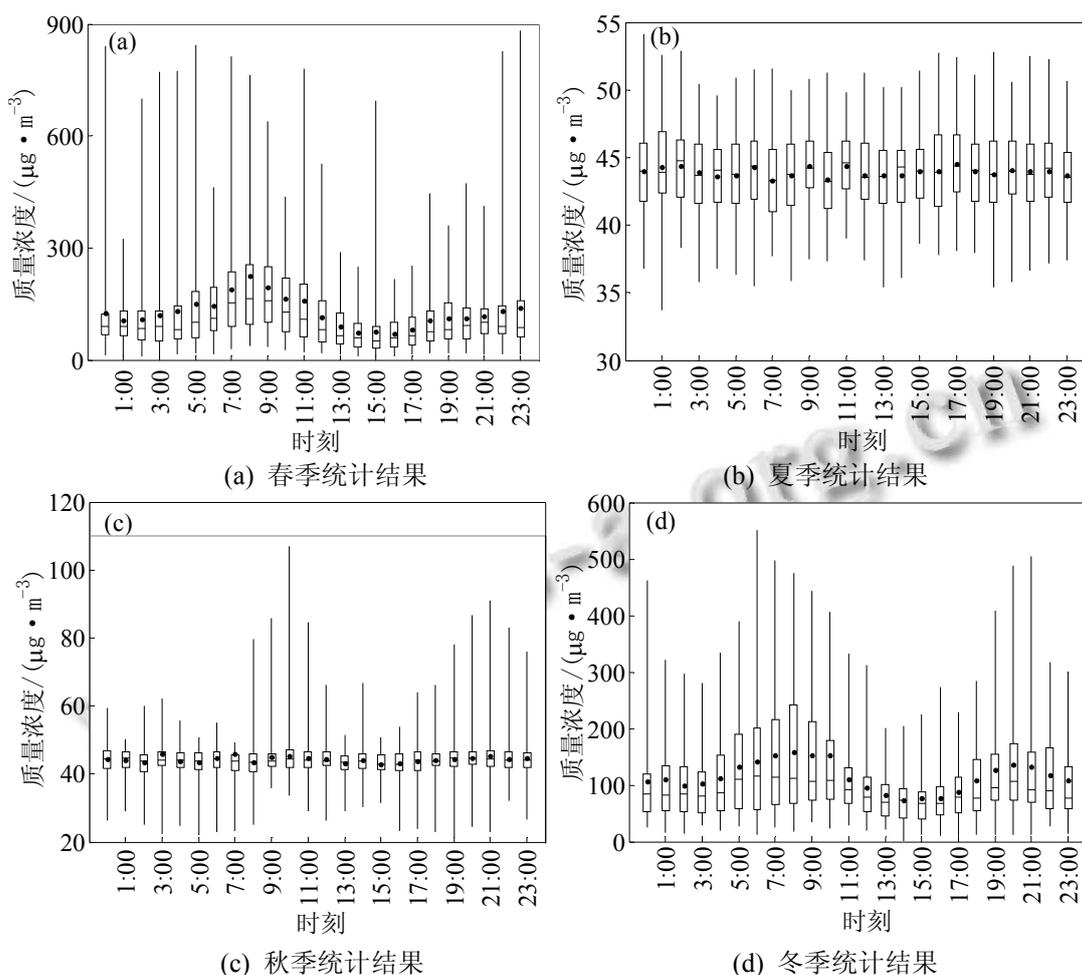


图 2 2014 年 4 月-2015 年 3 月山西省太旧高速公路阳泉服务区北区监测点 SO<sub>2</sub> 小时质量浓度统计结果

### 3 SO<sub>2</sub>浓度时序预测模型的建立

#### 3.1 模糊粒化高斯模型的建立

将山西省太旧高速公路阳泉服务区北区监测点 2014 年 4 月至 2015 年 3 月一年的样本数据按时间序列——季节划分为春季、夏季、秋季和冬季 4 个研究对象,并分别预测季节最后 1 天的浓度变化趋势和范围,用以验证预测结果的准确性.即将 2015 年 1-3 月的小时浓度监测值作为研究 SO<sub>2</sub> 气体浓度春季变化规律的研究对象,并预测 2015 年 3 月 31 日的 24h 浓度变化趋势,实际取得样本有效数据 1848 个,变化范围为  $y_1 \in (8.6, 883) \mu\text{g}\cdot\text{m}^{-3}$ ; 将 2014 年 4-6 月小时浓度监测值作为研究 SO<sub>2</sub> 气体浓度夏季变化规律的研究对象,并预测 2014 年 6 月 30 日的 24h 浓度变化趋势,实际取得样本有效数据 2150 个,变化范围为  $y_2 \in (33.7, 54.1) \mu\text{g}\cdot\text{m}^{-3}$ ; 2014 年 7-9 月小时浓度监测值作为研究 SO<sub>2</sub> 气体浓度秋季变化规律的研究对象,并预

测 2014 年 9 月 30 日的 24h 浓度变化趋势,实际取得样本有效数据 2208 个,变化范围为  $y_3 \in (21.3, 107.1) \mu\text{g}\cdot\text{m}^{-3}$ ; 2014 年 10-12 月小时浓度监测值作为研究 SO<sub>2</sub> 气体浓度冬季变化规律的研究对象,并预测 2014 年 12 月 31 日的 24h 浓度变化趋势,实际取得样本有效数据 2059 个,变化范围为  $y_4 \in (1.1, 550.2) \mu\text{g}\cdot\text{m}^{-3}$ .

分别将四个季节的 SO<sub>2</sub> 原始浓度监测样本数据,利用选定的 24h 粒化窗宽,划分为众多子窗口,每个子窗口又通过式(14)所示的高斯函数,筛选出一个模糊粒子,所有子窗口的模糊粒子组成了训练预测模型的输入样本集.以春季为例,原始浓度随时间的变化如图 3 所示,有效数据 1848 个,粒化子窗口数等于有效数据个数 1848 除以粒化窗宽 24 取整,为 77 个,模糊粒化处理后的样本数据如图 4 所示,图中绘制了三种模糊粒化参数 U、R 和 L 的浓度曲线,其中 U 表示

SO<sub>2</sub> 浓度变化的最大值、R 表示 SO<sub>2</sub> 浓度变化的平均值、L 表示 SO<sub>2</sub> 浓度变化的最小值。

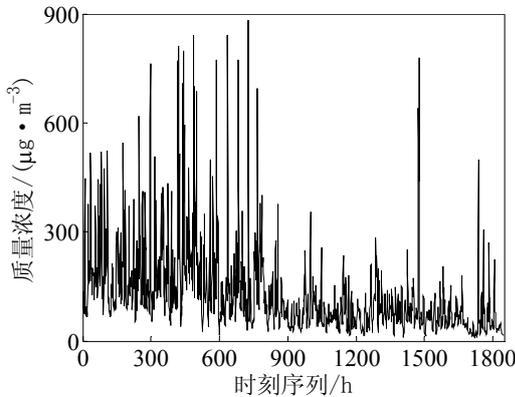


图 3 春季 SO<sub>2</sub> 质量浓度

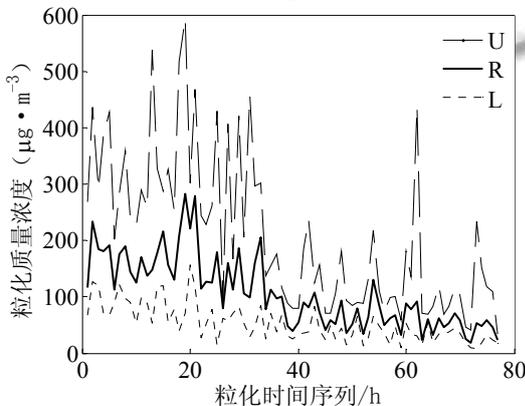


图 4 模糊粒化后的 SO<sub>2</sub> 质量浓度

### 3.2 优化支持向量机的性能参数

核函数的类型、核函数的参数 G 以及惩罚参数 C 都会对支持向量机的性能造成影响，目前的研究中，对于这些因素的选择尚无统一的模式规定，只做了参数优化的研究<sup>[17]</sup>。

通常的高斯核函数的表达式为：

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\delta^2}\right) \quad (7)$$

当  $\delta > 0$  时，仅对较小领域内的样本数据具有影响，当样本距远远大于  $\delta$  时，基本不会对样本数据产生影响，大气污染物 SO<sub>2</sub> 的浓度值日变化、季节变化明显，因此应用高斯核函数会较好的拟合效果。

通过对 SVM 进行距离 L1 优化后为：

$$\begin{cases} \min J(\omega, b, \xi) = \frac{1}{2} \mathbf{W}^T \mathbf{W} + C \sum_{j=1}^l \xi_j \\ \text{s.t. } y_j (\omega^T \varphi(x_j) + b) \geq 1 - \xi, \xi_j \geq 0 \end{cases} \quad (8)$$

式(8)对应的对偶问题为：

$$\begin{cases} \max \mathbf{W}(\alpha) = \mathbf{e}^T \alpha - \frac{1}{2} \alpha^T \mathbf{Q} \alpha \\ \text{s.t. } 0 \leq \alpha_j \leq C, \mathbf{y}^T \alpha = 0 \end{cases} \quad (9)$$

其中， $\mathbf{e}$  代表单位向量， $\mathbf{Q}$  为  $l \times l$  矩阵，且  $Q_{ij} = y_i y_j K(x_i, x_j)$ 。

根据核函数的线性组合原理，组合出复合核函数，为了提高高斯核函数的优势，文献[18-20]分别对高斯核函数进行了改进，如下所示：

$$k(x, y) = \exp\left(-\frac{1}{2} \sum_p \frac{(x_p - y_p)^2}{\delta_p^2}\right), \forall p \quad (10)$$

$$k(x, y) = \frac{1}{m} \sum_{t=1}^m \exp\left(-\frac{\|x_t - y_t\|^2}{2\delta_t^2}\right) \quad (11)$$

$$k(x, y) = \exp(-\lambda \|x_t - y_t\|^2) + 1 \quad (12)$$

$$k(x, y) = \left(\exp(-\lambda \|x_t - y_t\|^2) + \gamma\right) \quad (13)$$

$\lambda > 0, \gamma \geq 0, t \in N$

其中，前两个核函数易于陷入局部最优化，而后面的两个核函数通过与多项式函数组合形成新的核函数，在实际应用中达到了较好的效果。本文采用了式(13)改进的高斯核函数，目的是通过多项式核函数消除样本距极小值对回归拟合的影响。

### 3.3 基于 LIBSVM 的回归预测实现

支持向量机算法是通过台湾大学林智仁教授开发设计的 LIBSVM 工具箱实现的，较 MATLAB 原有 SVM 工具包，LIBSVM 具有使用方便、不受使用者基础知识的限制、兼容性强、能提供交互检验功能、拟合精度高等优势。应用 LIBSVM 工具箱对模糊粒化处理过的、输入 SVM 的样本集再进行回归预测的步骤是：1) 样本集数据归一化处理；2) 尝试在  $[2^{-10}, 2^{10}]$  范围内，充分利用 k 折交互检验法搜索惩罚参数 C 和核函数参数 G 的最佳回归参数；3) 得到最佳回归参数值后，分别对样本集的最小值、平均值和最大值进行回归预测，预测结果如图 5 所示，可以看出三者的拟合结果基本与原始数据保持一致，模型具有较好的预测能力。同理，可得到夏季、秋季和冬季的预测模型。

表 1 列出了春、夏、秋、冬四个时序模型的三种模糊粒化对象(最小值 L、平均值 R 和最大值 U)建模时所选取的惩罚参数 C 和核函数参数 G 的最优值比较结果，四季模型的三种模糊粒化对象实际值和预测值的对比结果以及模型误差分析结果，曲线图如图 6 所示。

从表 1 列出的模型参数最优值可以看出，搜索到

的模型参数最优值,保证了标准误差  $\delta_{MRE}$  足够小,能有效降低模型输入样本集的离散程度,提高模型的泛化性能.从表 1 列出的模型输出预测值与实际值的绝对误差范围来看:夏季模型三种模糊对象的绝对误差均小于其他三个季节,其中最小值 L 的绝对误差最小,仅为  $0.432\mu\text{g}\cdot\text{m}^{-3}$ ,最大值 U 的绝对误差最大,也只有  $5.803\mu\text{g}\cdot\text{m}^{-3}$  为冬季模型最大值 U 绝对误差的 11.6%,可见,夏季模型的预测结果最好.分析造成冬季模型最大值 U 绝对误差最大的原因是冬季  $\text{SO}_2$  污染气体排放受供暖期影响,浓度较其他季节明显偏高,最大值 U 为  $178\mu\text{g}\cdot\text{m}^{-3}$ ,计算其相对误差为 24.3%,能够满足

污染物浓度预测的误差要求.另外,根据环境空气质量标准 GB3095-2012 规定  $\text{SO}_2$  浓度限值执行 24 小时平均,冬季模型 U 预测值的较大误差对了解未来 24h  $\text{SO}_2$  浓度趋势并无直接影响,而准确的 R 预测值对预测  $\text{SO}_2$  浓度则起关键作用,对比四个模型三种模糊粒化对象的绝对误差,可以看出平均值 R 预测值与实际值的绝对误差较其他两个模糊粒化对象(最大值 U 和最小值 L)最小,分别为  $7.786\mu\text{g}\cdot\text{m}^{-3}$ (春季)、 $1.529\mu\text{g}\cdot\text{m}^{-3}$ (夏季)、 $4.93\mu\text{g}\cdot\text{m}^{-3}$ (秋季)和  $2.228\mu\text{g}\cdot\text{m}^{-3}$ (冬季).分析结果表明,所建模型预测结果较好,能够满足实际应用需求.

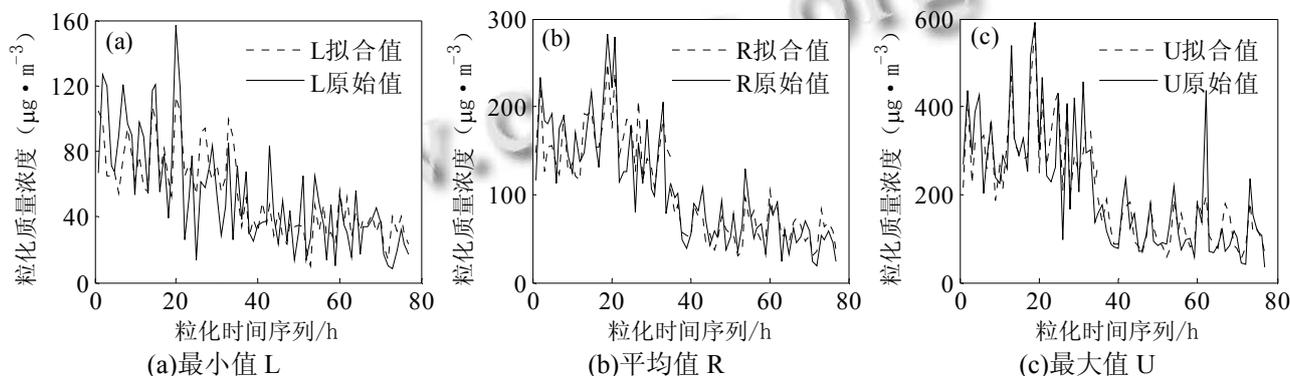


图 5 拟合值与原始值的对比结果

表 1 模型的性能参数和误差分析

时间序列 模型	模糊粒化 对象	最佳回归参数		标准误差 ( $\delta_{MRE}$ )	相关系数平 方( $r^2$ )	显著性水平		预测值 ( $\mu\text{g}\cdot\text{m}^{-3}$ )	实际值 ( $\mu\text{g}\cdot\text{m}^{-3}$ )	绝对误差 ( $\mu\text{g}\cdot\text{m}^{-3}$ )
		核函数参 数(G)	惩罚参数 (C)			$R_{0.05}$	$R_{0.01}$			
春季	L	1	0.5	0.014	0.580			23.029	15	8.029
	R	1	1	0.018	0.736	0.227	0.296	37.136	29.35	7.786
	U	1	0.25	0.053	0.639			66.287	48.2	18.087
夏季	L	0.5	0.25	0.002	0.494			39.432	39	0.432
	R	0.5	0.25	0.001	0.511	0.211	0.275	42.596	44.125	1.529
	U	0.036	0.25	0.002	0.478			45.797	51.6	5.803
秋季	L	0.016	0.25	0.005	0.417			33.402	22	11.402
	R	1	1	0.007	0.435	0.207	0.27	39.474	44.404	4.93
	U	0.25	0.25	0.007	0.556			54.607	79.8	25.193
冬季	L	1	0.25	0.019	0.703			29.154	12.9	16.254
	R	1	0.35	0.033	0.542	0.216	0.281	63.351	65.579	2.228
	U	1	0.25	0.057	0.562			134.737	178	43.263

监测点位于冠山药林寺自然保护区(省级),按照国家标准 GB3095-2012 规定执行一级标准<sup>[21]</sup>,24 小时平均值限值  $50\mu\text{g}\cdot\text{m}^{-3}$ .如图 6(a)所示,春季 R 预测值

$37.136\mu\text{g}\cdot\text{m}^{-3}$ ,实际平均值  $29.35\mu\text{g}\cdot\text{m}^{-3}$ ,实际值和预测值均小于  $50\mu\text{g}\cdot\text{m}^{-3}$ ,满足环境空气质量  $\text{SO}_2$  污染气体排放的一级标准;如图 6(b)所示,夏季 R 预测值

42.596 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际平均值 44.125 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际值和预测值均小于 50 $\mu\text{g}\cdot\text{m}^{-3}$ , 满足环境空气质量 SO<sub>2</sub> 污染气体排放的一级标准; 如图 6(c)所示, 秋季 R 预测值 39.474 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际平均值 44.404 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际值和预测值均小于 50 $\mu\text{g}\cdot\text{m}^{-3}$ , 满足环境空气质量 SO<sub>2</sub> 污染气体排放的一级标准; 如图 6(d)所示, 冬季 R 预测值 63.351 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际平均值 65.579 $\mu\text{g}\cdot\text{m}^{-3}$ , 实际值和预

测值均大于 50 $\mu\text{g}\cdot\text{m}^{-3}$  小于 150 $\mu\text{g}\cdot\text{m}^{-3}$ , 略高于环境空气质量 SO<sub>2</sub> 污染气体排放的一级标准低于二级标准, 冬季 SO<sub>2</sub> 污染气体排放受供暖期影响较其他三个季节浓度较高. 可见, 春季、夏季、秋季和冬季四个模型的预测值都能够执行与实际值同一级别的 SO<sub>2</sub> 污染气体排放标准, 模型预测结果能够表征气体浓度实际监测值的属性.

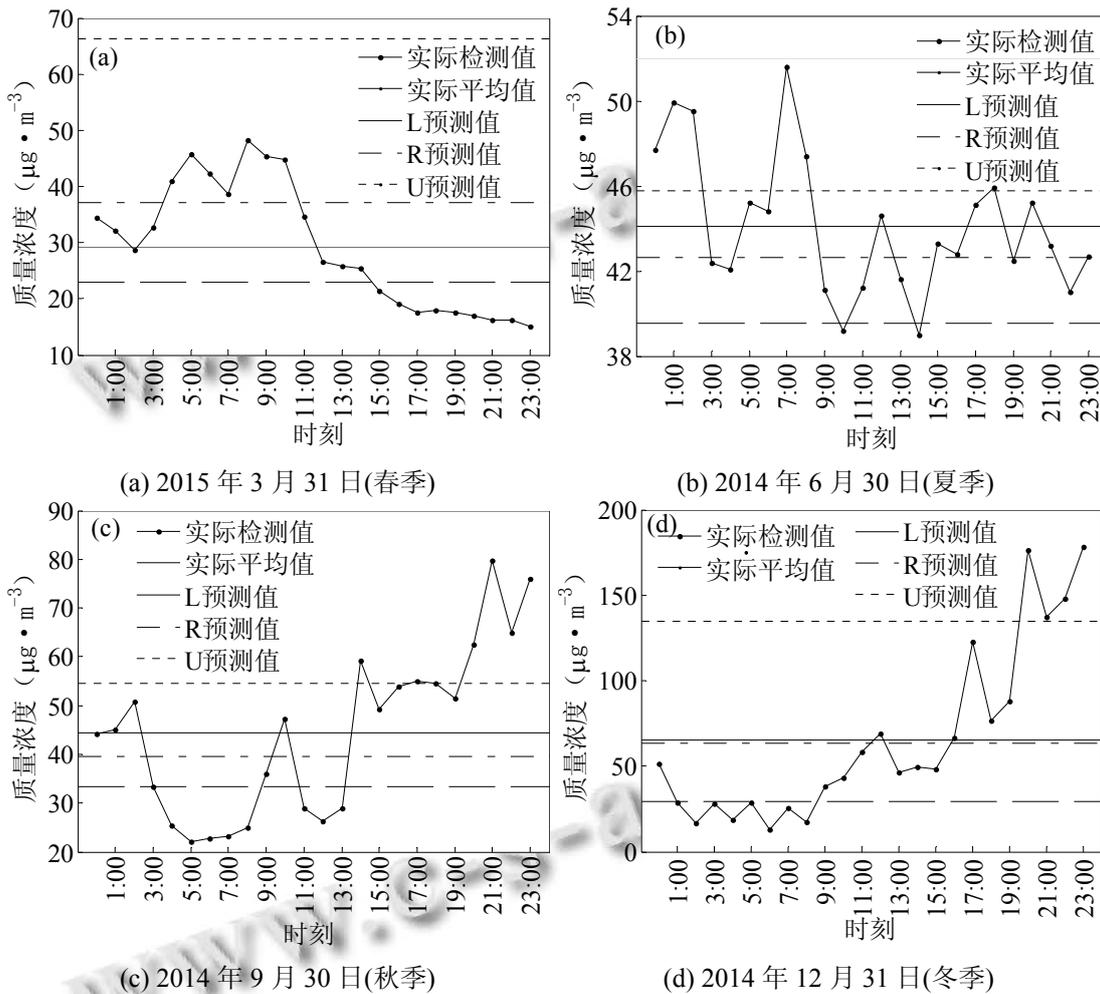


图 6 各季节最后一天的二氧化硫质量浓度预测值与实际值的比较结果.

#### 4 结论

1) 以山西省太旧高速公路阳泉服务区北区监测点 2014 年 4 月至 2015 年 3 月一年的样本数据作为训练模型的基础数据, 利用模糊时序处理后作为支持向量机的输入, 训练模型分别得到 SO<sub>2</sub> 浓度春、夏、秋、冬四个季节的预测模型. 基于模糊时序和支持向量机的 SO<sub>2</sub> 浓度预测方法研究的是 SO<sub>2</sub> 浓度的时间分布规律, 能有效克服污染物来源不确定、影响因素复杂、

非线性回归拟合易陷入局部最优等现有 SO<sub>2</sub> 浓度预测方法存在的问题.

2) 三种模糊粒化对象的预测值和实际值执行同一级别的 SO<sub>2</sub> 污染气体排放标准, 能够有效描述 SO<sub>2</sub> 浓度变化范围和变化趋势, 分析结果表明, 模糊时序数据处理中以季节作为时间序列, 以 24h 作为模糊粒化窗宽, 并利用高斯函数提取出各时间序列的三种模糊粒化对象最大值 U、最小值 L 和平均值 R 的特征值

的方法,能提高模型精度、改善拟合效果。另外,当具有足够长时间的样本数据时,同理可利用基于模糊时序和支持向量机的  $\text{SO}_2$  浓度预测方法预测更长时间范围的  $\text{SO}_2$  浓度时间分布规律。

3) 由于目前对于影响支持向量机性能影响因素的研究,尚无完善的理论依据,而核函数的类型、核函数参数  $G$  和惩罚参数  $C$  等性能参数的优选对模型预测精度影响较大,所以,接下来还可以考虑继续研究基于模糊时序和支持向量机的  $\text{SO}_2$  浓度预测方法中提高支持向量机性能的途径,进一步提高模型精度。

### 参考文献

- 1 姜少蓉,薛志钢,李薇,杜谨宏,张沿成,王文双.我国环境空气质量状况及大气污染对健康的影响.华北电力技术,2015,(8):7-13.
- 2 徐晨雨,王晶,陈跃龙.北京市城市道路交通与大气污染关系探究.商业经济,2015,(9):110-111,149.
- 3 刘永,郭怀成.城市大气污染物浓度预测方法研究.安全与环境学报,2004,4(4):60-62.
- 4 Pai TY, Lo HM, Wan TJ, et al. Predicting air pollutant emissions from a medical incinerator using grey model and neural network. Applied Mathematical Modelling, 2014, 39(5-6): 1513-1525.
- 5 任永建,高庆先,周锁铨.山西省阳泉市大气环境数值模拟.大气环境科学研究暨颗粒物污染防治与监测技术研讨会,2010.
- 6 荆涛,李霖,于文柱,王玉娟,郑永杰,田景芝.t 分布受控遗传算法优化BP神经网络的PM2.5质量浓度预测.中国环境监测,2015,31(4):100-105.
- 7 赵宏,刘爱霞,王恺,白志鹏.环境空气  $\text{SO}_2$  和  $\text{NO}_2$  浓度的 GA\_ANN 预测模型研究.计算机工程与应用,2010,46(8):199-201.
- 8 王丽梅,袁野,姚建.基于 B-P 神经网络的城市大气  $\text{SO}_2$  浓度预测.安徽农业科学,2011,39(7):4278-4280.
- 9 陈柳,马广大.大气中  $\text{SO}_2$  浓度的小波分析及神经网络预测.环境科学学报,2006,26(9):1553-1558.
- 10 姚宁,马青兰,张晶,文印.基于 AGNES 算法优化 BP 神经网络和 GIS 系统的大气污染物浓度预测.中国环境监测,2015,31(3):113-117.
- 11 Liu C B, Wang X F, Pan F. Parameters selection and stimulation of support vector machines based on ant colony optimization algorithm. Journal of Central South University, 2008,39(6):1309-1313.
- 12 付学良,杨洋,李纯子.基于混沌—支持向量机的大气污染物浓度预测模型.电子世界,2013,(4):78-79.
- 13 马博,董海鹰,任伟.基于模糊综合支持向量机的特高压变电站二次设备状态评估.计算机系统应用,2014,23(9):191-197.
- 14 Wang Q, Tian XM. Soft sensing based on fuzzy information granulation. Trans. of Beijing Institute of Technology, 2012, 32(9): 955-959.
- 15 刘杰,杨鹏,吕文生,刘阿古达木,刘俊秀.模糊时序与支持向量机建模相结合的PM2.5质量浓度预测.北京科技大学学报,2014,36(12):1694-1702.
- 16 郭龙.时间序列数据的周期性研究[硕士学位论文].成都:电子科技大学,2010.
- 17 陈俏,曹根牛,陈柳.支持向量机应用于大气污染物时间序列预测.计算机时代,2009,(9):1-3.
- 18 于进,钱锋.基于粒子群优化的高斯核函数聚类算法.计算机工程,2010,36(14):22-23.
- 19 徐咏梅,柳桂国,柳贺.高斯径向基核函数参数的 GA 优化方法.电力自动化设备,2008,28(6):52-55.
- 20 王行甫,陈家伟.基于高斯核的 SVM 的参数选择.计算机系统应用,2014,23(7):242-245.
- 21 国家环境保护部,国家质量监督检验检疫总局. GB3095-2012 环境空气质量标准.北京:中国环境科学出版社,2012.