

基于 GM-LSSVR 机场噪声时间序列预测模型

温冬琴 王建东 张霞

(南京航空航天大学计算机科学与技术学院 南京 210016)

摘要 随着我国民航事业的不断发展,机场噪声问题日益严重。针对机场噪声时间序列预测问题,在分析了噪声数据特点的基础上,提出了一种基于 GM-LSSVR 的噪声序列预测模型,即将机场噪声时间序列进行分解,并分别建立趋势项的 GM(1,1)模型和剩余项的 LSSVR 模型,用于机场噪声预测值的计算。在某机场实测数据中的应用表明,该方法的准确度明显优于其他预测方法。

关键词 时间序列分解,机场噪声预测模型,GM(1,1),最小二乘支持向量机,Spearman

中图分类号 TP399 **文献标识码** A

Prediction Model for Airport-noise Time Series Based on GM-LSSVR

WEN Dong-qin WANG Jian-dong ZHANG Xia

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract Along with the development of our country civil aviation, the airport noise is more and more serious. Aimed at the airport-noise time sequence prediction problem, this paper presented the prediction model based on GM-LSSVR, which is based on the analyses of noise's characteristic. Namely, the model decomposes the noise time sequence into tendency and the residual, then establishes the GM(1,1) model of trend and the LSSVR model of residual, which are used for airport-noise prediction model. The application in the measured data of an airport shows that the accuracy of this model is better than other prediction models.

Keywords Time series decomposition, Airport-noise prediction model, GM(1,1), LSSVR, Spearman

随着社会的进步和我国民航事业的不断发展,机场噪声日渐严重。为了更加有效地规划机场布局,合理地构建航班班次,需要对每天的机场噪声数据值进行研究。从时域上看,每天的机场噪声数据值构成了一种时间序列,进而将时间序列分析方法运用到噪声序列预测中。时间序列分析是概率统计学科的一个分支,其运用概率统计的理论和方法分析随机数据序列,并对其建立数学模型,在此基础上进一步分析随机数据的统计特性^[1]。该方法日趋成熟,被广泛地用到预报预测领域。

由 Vapnik(1995 年)提出的支持向量机^[2]是近年来受到广泛关注的一种新的数据挖掘方法,它是在小样本统计学习理论和结构风险最小化理论的基础上发展起来的一种新型的机器学习方法。支持向量机根据有限的样本信息,在模型的复杂度和学习能力之间寻求最佳折衷,以此获得最好的泛化能力。与经典支持向量机相比,最小二乘支持向量机将价值函数改为最小二乘价值函数以及用等式约束代替不等式约束^[3],避免了求解耗时的 QP 问题,提高了求解速度^[4]。其克服了传统学习器学习算法过拟合、局部优化的缺陷^[5],同时具有泛化能力强、抗干扰能力强等特点^[7]。目前,该方法已被广泛地应用到时间序列预测中。

文献[5]提出了一种基于 GM-LSSVM 的数据融合方法,

并将其用在无线传感器网络节点通信量的时间序列中。本文利用少量监测数据对模型进行训练,以灰色回归模型预测数据作为最小二乘支持向量机的输入数据,并对下一步未知信息进行预测。该融合方法对非线性、随机性和突发性的数据拟合度不佳的问题有了很大的改进。但是以该方式的融合算法在时间序列数据具有趋势时,预测效果不是很理想。

本文结合传统的时间序列分解方法^[6],针对文献[5]中的问题,基于 GM-LSSVR 噪声时间序列预测模型,将时间序列分解为趋势项和剩余项,然后利用 GM(1,1)对趋势项进行提取建模,LSSVR 对剩余项进行建模,两者的预测结果之和作为最终的预测值。本文提出的融合方式还未曾在文献中出现过,具有一定的新意。

1 噪声时间序列的分解模型

设机场噪声时间序列为 $Y = \{y_1, y_2, \dots, y_l\}$, 在趋势性显著 的情况下,将噪声时间序列分解为趋势项和剩余项的和,即 $y_i = T_i + M_i, (i = 1, 2, \dots, l)$, 其中 l 为序列长度, T_i 为趋势项,反映时间序列要素随时间的发展趋势; M_i 为剩余项,即原序列剔除了趋势项之后的数据项,反映时间序列要素的其它变化。进而对趋势项和剩余项分别建立灰色预测模型和最小二乘支持向量机回归模型。

到稿日期:2012-11-29 返修日期:2013-01-26 本文受基金项目:面向机场感知的噪声监测及其环境影响评估(61139002)资助。

温冬琴(1988—),女,硕士生,主要研究方向为数据挖掘和机场噪声时间序列分析,E-mail:wendongqinnuaa@163.com。

1.1 基于 spearman 法的趋势性检验

趋势性检验一般采用 spearman 秩次相关检验法^[9]。即对于给定的时间序列 $\{y_t\}$,在分析其与时序 $t(t=1,2,\dots,l)$ 的相关系数时,用其秩次(序列从大到小排列时相应的序号)代表,进而可求出序列与秩次的相关系数 R_s :

$$R_s = 1 - \frac{6 \sum_{t=1}^l (R_t - t)^2}{l^3 - l} \quad (1)$$

式中, l 为序列长度, R_t 为序列的秩次。

同时,构造统计量:

$$T = R_s \left(\frac{l-4}{1-R_s^2} \right)^{\frac{1}{2}} \quad (2)$$

T 服从自由度为 $(l-2)$ 的 t 分布^[10]。在显著水平 α 下,若有 $|T| > t_{1-\alpha/2}$ 成立,则说明序列趋势显著;反之,趋势不显著。若具有显著趋势,可用GM(1,1)进行提取。

1.2 基于 GM(1,1)模型的趋势项提取

灰色预测模型通过灰色系统理论中的序列算子弱化数据随机性,挖掘潜在规律,并利用离散的数据序列建立连续动态微分方程,进而进行预测^[11]。常用的灰色预测模型由包含单一变量的一阶微分方程构成^[12],称作GM(1,1)模型,其原理如下,设原始序列:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(l)) \quad (3)$$

通过一次累加,生成具有一定规律的新序列:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(l)) \quad (4)$$

式中, $x^{(1)}(k) = \sum_{m=1}^k x^{(0)}(m), k=1,2,\dots,l$

然后通过新序列建立白化方程:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (5)$$

式中, a 和 u 为待定参数。

进而构造背景序列:

$$Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(l)) \quad (6)$$

式中

$$z^{(1)}(k) = 0.5x^{(1)}(k-1) + 0.5x^{(1)}(k), k=2,3,\dots,l$$

将式(5)离散化,得到GM(1,1)的灰微分方程:

$$X^{(0)}(k) + aZ^{(1)}(k) = u \quad (7)$$

利用最小二乘法求解出参数 a 和 u ,进而得到新序列 $X^{(1)}$ 的预测方程:

$$X^{(1)}(k+1) = (X^{(0)}(1) - \frac{u}{a}) e^{-\hat{a}k} + \frac{u}{a}, k=1,2,\dots,l \quad (8)$$

在此基础上,通过一次累减生成得到原始序列的预测值,即

$$\begin{aligned} X^{(0)}(1) &= X^{(0)}(1) \\ X^{(0)}(k) &= X^{(1)}(k) - X^{(1)}(k-1) \\ &= (1 - e^{-a})(X^{(0)}(1) - \frac{u}{a})e^{-a(k-1)}, k=1,2,\dots,l \end{aligned} \quad (9)$$

1.3 基于最小二乘支持向量机(LSSVM)的余项建模

设训练样本 $\{(x_1, y_1), \dots, (x_l, y_l)\} \in R^n \times R$,其中 x_i 为输入向量, y_i 为输出量, l 为样本总数, n 为样本空间的维数。通过非线性映射 $z_i = \phi(x_i) \in R^m$,将原来低维空间映射到高维空间中($n \gg m$)。然后在高维空间采用线性函数 $g(x) =$

$w^T \phi(x) + b$ 进行拟合,并容许出现拟合误差,以求结构风险最小^[13]。

在最小二乘支持向量机中构造最小化目标函数:

$$\min J(w, e, b) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l e_i^2 \quad (10)$$

$$\text{s. t. } y_i = w^T \phi(x_i) + b + e_i, i=1,2,\dots,l$$

式中, $J(w, e, b)$ 为结构风险; C 为经验风险调整因子; e_i 为容许误差。

利用 Lagrange 法求解式(10)中的优化问题并根据 KKT 优化条件整理,得到 LSSVM 回归模型:

$$\hat{y} = g(x) = \sum_{i=1}^l \alpha_i k(x, x_i) + b \quad (11)$$

式中, $\alpha_i (i=1, \dots, l)$ 是拉格朗日乘子, b 为偏置量, $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 是根据 Mercer 条件定义的核函数。

2 算法流程图与建模步骤

2.1 算法流程图

根据算法思想,绘制流程图,如图1所示。

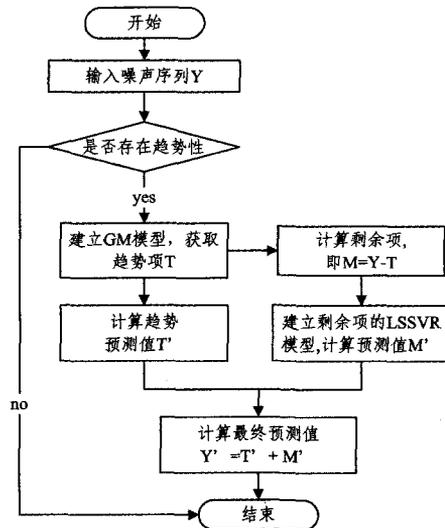


图1 算法流程图

2.2 建模步骤

步骤1 趋势性检验,采用 spearman 秩次相关检验法对噪声时间序列数据进行趋势判断,检测其是否具有一定的趋势。

步骤2 在趋势显著的基础上,对噪声时间序列进行分解,利用GM(1,1)对趋势进行提取,建立预测模型,获取趋势项序列 $\{T_i\}, i=1,2,\dots,l$,并计算趋势项预测值 \hat{T} 。

步骤3 根据公式 $M_i = y_i - T_i$,其中 $\{y_i\}$ 为原序列, $\{T_i\}$ 为趋势项序列, $i=1,2,\dots,l$,计算余项 $\{M_i\}$,对其分析并建立LSSVR回归预测模型,计算余项预测值 \hat{M} 。

步骤4 联合两模型作为最终的预测模型,即最终预测值 $\hat{y} = \hat{T} + \hat{M}$ 。

步骤5 评价,主要从平均绝对误差(MAE)、希尔不等系数(Theil IC)这两个评价标准来比较几个模型的预测精度。一般认为平均绝对误差越小,则预测精度越高;希尔不等系数一般介于0到1之间,数值越小表明拟合值与真实值的差异越小,即预测精度越高^[7]。

3 模型在机场噪声预测中的应用

3.1 数据选取

已有数据为国内某机场监测点 2010 年 2 月至 8 月的实测数据,由 15 个观测点每秒采集一次获得(单位 dB)。面对大量数据,我们选取 2 号和 12 号观测点 2 月 2 日至 8 月 20 日共 200 天的数据。由于机场噪声具有声级高、间断性等特点,我们取每天的平均值作为实验数据。在此基础上,我们选取前 190 天的数据作为训练集,后 10 天的数据作为测试集。

3.2 实例建模

1. 趋势项建模

利用 spearman 秩次相关检验法对数据进行趋势检验,当显著水平 $\alpha=0.01$ 时,查表得出其 t 分布值为 2.576,实验结果如表 1 所列。

表 1 趋势检验结果

实验数据	统计量 T	临界值	趋势性
2 号点	6.078	2.576	显著
12 号点	5.667	2.576	显著

由表 1 知两观测点的趋势显著,利用 GM(1,1)对 2 号点、12 号点进行趋势提取并建模预测,得到 2 号点趋势项模型参数 $a = -0.00042, u = 47.5541$,12 号点模型参数为 $a = -0.00038, u = 49.3684$,其趋势线分别如图 2 和图 3 所示。

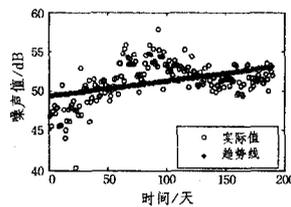
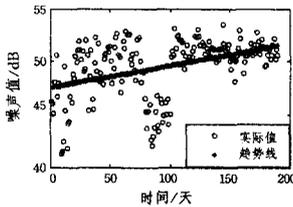


图 2 2 号点噪声数据的趋势线

图 3 12 号点噪声数据的趋势线

2. 剩余项建模

根据建模步骤 3 中的公式计算剩余项,然后用 LSSVR 对剩余项进行预测。将剩余项训练样本中 1~189 个数据作为输入向量,2~190 个数据作为决策向量。在 MatlabR2009a 平台上使用 LSSVMlabv1_8_R2006a_R2009a 工具箱,设置核函数为高斯径向基,利用训练集对模型参数寻优,得到回归模型。模型建立后,用第 190 天(即 8 月 10 号)的剩余项作为输入向量,预测出第 191 天的剩余项值。以此类推,预测出第 192~200 天的剩余项预测值,结果如表 2 所列。

表 2 趋势项和剩余项预测值

时间	2 号点		12 号点	
	趋势项	剩余项	趋势项	剩余项
08/11	51.515	-0.544	53.067	-0.823
08/12	51.537	0.310	53.088	-0.858
08/13	51.559	-0.126	53.108	-0.304
08/14	51.580	-0.347	53.128	-0.326
08/15	51.602	-0.681	53.148	-0.392
08/16	51.624	0.488	53.167	-1.023
08/17	51.645	0.969	53.189	-1.377
08/18	51.667	1.555	53.209	1.520
08/19	51.689	1.197	53.229	1.705
08/20	51.710	1.451	53.249	0.955

3. 预测模型

根据步骤 4 的公式计算最终预测值,并与其它模型进行比较,具体数值如表 3 和表 4 所列。

表 3 2 号点各模型的预测值

2 号点实际值	GM	LSSVM	文献[5]模型	本文模型
51.68	51.52	50.47	49.76	50.97
51.12	51.54	51.12	49.49	51.85
50.86	51.56	50.82	49.48	51.43
50.47	51.58	50.67	49.48	51.23
51.96	51.6	50.42	49.48	50.92
52.77	51.62	51.25	49.48	52.07
53.87	51.65	51.56	49.48	52.61
53.2	51.67	51.81	49.48	53.22
53.7	51.69	51.68	49.48	52.89
54.26	51.71	51.78	49.48	53.16

表 4 12 号点各模型的预测值

12 号点实际值	GM	LSSVM	文献[5]模型	本文模型
50.07	53.07	51.95	51.99	52.24
51.96	53.09	50.44	51.96	52.23
52.76	53.11	51.95	51.92	52.8
52.75	53.13	52.47	51.86	52.8
51.68	53.15	52.47	51.79	52.76
51.8	53.17	52.42	51.71	52.14
51.28	53.19	51.83	51.63	51.81
55.36	53.21	51.44	51.54	54.73
55.65	53.23	53.83	51.45	54.93
54.61	53.25	53.97	51.36	54.2

3.3 模型评价

本文中采用 MAE 和 Theil IC 作为评价标准,计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

$$Theil\ IC = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2}} \quad (13)$$

针对表 3 和表 4 所列的模型的预测值,各模型的 MAE, Theil IC 如表 5 所列。

表 5 两观测点各模型的 MAE, Theil IC 值

模型	2 号点		12 号点	
	MAE	Theil IC	MAE	Theil IC
GM	1.221	0.014	1.454	0.016
LSSVM	1.271	0.015	1.225	0.015
文献[5]模型	2.880	0.031	1.625	0.022
本文模型	0.770	0.008	0.525	0.008

从表 5 中可看出:对于 2 号观测点和 12 号观测点,文献 [5]中的 GM-LSSVM 融合模型的预测效果很不理想,说明了该方式的融合方法在序列具有趋势时具有一定的局限性;而本文所提出的模型在平均绝对误差、希尔不等系数方面与其他模型比较中,值均小于其他模型,说明该模型有很高的预测度,同时也说明该模型在机场噪声时间序列预测这方面优于上述其它模型。

结束语 本文针对机场噪声时间序列预测问题,在分析噪声数据特点的基础上,提出 GM_LSSVR 的分解模型。该模型将噪声时间序列分解为趋势项与剩余项的和,进而分别建立 GM(1,1)和 LSSVR 模型并进行预测,最后将两部分的和作为最终的预测值。用这一算法对国内某机场的实测数据进行预测,从常用的两种评价指标看出本文方法优于其它算法模型,能有效地提高预测精度。再则,本文提出的算法适用于时间序列的单步预测,且是具有显著趋势的时间序列的单

(下转第 220 页)

1) $\beta=1$ 时, $bnR_{\beta}A = \phi$, 所以 $bnR_{\beta}(bnR_{\beta}A) = bnR_{\beta}A = \phi$;

2) $B = \phi$ 时, $bnR_{\beta}B = \phi = B$;

3) 现 $\beta \in [0, 1)$, $B \neq \phi$, 若 $\exists [x]_R \subseteq bnR_{\beta}B$, 则 $|[x]_R \cap B| > \beta|[x]_R|$, $|[x]_R \cap B| < |[x]_R| - k$.

由 $|[x]_R \cap B| > \beta|[x]_R| \geq 0$ 有 $[x]_R \subseteq B$, 则 $|[x]_R \cap B| = |[x]_R| < |[x]_R| - k$, 这是一个矛盾式, 所以 $bnR_{\beta}B = \phi$.

推论 1 $bnR(bnRA) = \phi$.

命题 6 研究了边界 $bnR_{\beta}A$ 关于集合系统的一般性质, 命题 7 则得到了其幂作用性质. 基于双量化扩张性(命题 1)与命题 7 及推论 1 得到了 Pawlak 边界的幂作用结果.

结束语 本文利用笛卡尔积, 结合变精度上近似与程度下近似组建了新的双量化扩张粗糙集模型, 类似可以做其它平行推广. 其中提出的边界, 对经典 Pawlak 边界而言, 也具有双量化扩张性, 因此具有具体的双量化语义, 同时对局部不确定性进行了双量化的完备与精细刻画, 这对双量化的不确定性分析与应用具有意义. 本文研究了该边界的算法, 还需要对近似空间或粗糙集理论的双量化扩张与双量化不确定分析进行进一步深入与系统的研究.

参 考 文 献

- [1] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356
- [2] Yao Y Y. The superiority of three-way decision in probabilistic rough set models [J]. Information Sciences, 2011, 181: 1080-1096
- [3] Ziarko W. Variable precision rough set model [J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59
- [4] Yao Y Y, Wong S K M, Lingras P. A decision-theoretic rough set model [C]// The 5th International Symposium on Methodologies for Intelligent Systems. North-Holland, New York, 1990: 17-25
- [5] Yao Y Y, Lin T Y. Generalization of rough sets using modal logics [J]. Intelligent Automation and Soft Computing, 1996, 2(2): 103-120

- [6] Inuiguchi M, Yoshioka Y, Kusunoki Y. Variable-precision dominance-based rough set approach and attribute reduction [J]. International Journal of Approximate Reasoning, 2009, 50(8): 1199-1214
- [7] Wang J Y, Zhou J. Research of reduct features in the variable precision rough set model [J]. Neurocomputing, 2009, 72: 2643-2648
- [8] Mi J S, Wu W Z, Zhang W X. Approaches to knowledge reduction based on variable precision rough set model [J]. Information Sciences, 2004, 159(3/4): 255-272
- [9] Yanto I T R, Vitasari P, Herawan T, et al. Applying variable precision rough set model for clustering student suffering study's anxiety [J]. Expert Systems with Applications, 2012, 39(1): 452-459
- [10] Yao Y Y, Lin T Y. Graded rough set approximations based on nested neighborhood systems [C]// Proceedings of 5th European Congress on intelligent techniques and Soft computing, EUFIT'97. Verlag Mainz, Aachen, 1997: 196-200
- [11] Liu C H, Miao D Q, Zhang N, et al. Graded rough set model based on two universes and its properties [J]. Knowledge-Based Systems, 2012, 33: 65-72
- [12] Xu W H, Liu S H, Wang Q R, et al. The first type of graded rough set based on rough membership function [C]// 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Yantai, China, 2010: 1922-1926
- [13] Zhang X Y, Mo Z W, Xiong F, et al. Comparative study of variable precision rough set model and graded rough set model [J]. International Journal of Approximate Reasoning, 2012, 53(1): 104-116
- [14] Parthala N M, Shen Q, Jensen R. A distance measure approach to exploring the rough set boundary region for attribute reduction [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(3): 305-317
- [15] Parthala N M, Shen Q. Exploring the boundary region of tolerance rough sets for feature selection [J]. Pattern Recognition, 2009, 42(5): 655-667

(上接第 200 页)
步预测。

参 考 文 献

- [1] 张树京, 齐立心. 时间序列分析简明教程 [M]. 北京: 北方交通大学出版社, 2003
- [2] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995
- [3] 陈磊, 张土乔. 基于最小二乘支持向量机的时用水量预测模型 [J]. 哈尔滨工业大学学报, 2006, 38(9): 1528-1529
- [4] 王宇红, 黄德先, 高东杰, 等. 基于 LS-SVM 的非线性预测控制技术 [J]. 控制与决策, 2004, 19(4): 384
- [5] 王汝言, 唐季超, 吴大鹏, 等. WSN 中基于 GM-LSSVM 的数据融合方法 [J]. 计算机工程与设计, 2012, 33(9): 3372
- [6] 陈卫民, 陈志刚. 基于 PSR-LSSVM 的网络流量预测 [J]. 计算机

科学, 2012, 39(7): 92-95

- [7] 陈伟利, 范玉刚, 吴建德, 等. 基于 LSSVM/PID 复合逆系统的预测控制 [J]. 计算机科学, 2012, 39(8): 239-241
- [8] 易丹辉. 数据分析与 Eviews 应用 [M]. 北京: 中国统计出版社, 2002: 46-92
- [9] 黄显峰, 邵东国, 阳书敏. 降雨时间序列分解预测模型及应用 [J]. 中国农村水利水电, 2007(9): 6-7
- [10] 钱光兴, 崔东文. 盘龙河流域水文气象要素变化趋势分析研究 [J]. 广东水利水电, 2011, 4(4): 33
- [11] 刘思峰, 谢乃明. 灰色系统理论及其应用 (第 4 版) [M]. 北京: 科学出版社, 2008
- [12] 耿秋燕, 梁毅刚. 基于灰色自适应粒子群 LSSVM 的铁路货运量预测 [J]. 西南交通大学学报, 2012, 47(1): 145-146
- [13] 张淑宁, 王福利, 尤富强, 等. 基于鲁邦学习的最小二乘支持向量机及其应用 [J]. 控制与决策, 2010, 25(8): 1170