

## 基于模糊支持向量回归的机场噪声预测

陈海燕<sup>1,2</sup> 杨冰欣<sup>1</sup> 徐涛<sup>2</sup> 王建东<sup>1</sup>

(1. 南京航空航天大学计算机科学与技术学院, 南京, 210016;

2. 中国民航信息技术科研基地, 天津, 300300)

**摘要:**在现行通用噪声计算模型基础上建立了一种基于模糊支持向量回归的机场噪声预测模型,通过计算样本的局部孤立因子来确定各个样本的模糊隶属度,以融入模糊支持向量回归算法中。最后,在某机场历史飞行数据的基础上,从对所提出模型的预测精度、抗干扰性、泛化能力进行了验证。结果表明,这种基于局部孤立因子的模糊支持向量回归算法能有效地预测机场周围的噪声等级,且该方法比标准支持向量回归具有更高的预测精度和更好的抗噪声能力。

**关键词:**机场噪声预测;支持向量机;模糊支持向量回归;模糊隶属度;局部孤立因子

**中图分类号:** TP18      **文献标志码:** A      **文章编号:** 1005-2615(2013)05-0722-05

## Airport Noise Prediction Based on Fuzzy Support Vector Regression

Chen Haiyan<sup>1,2</sup>, Yang Bingxin<sup>1</sup>, Xu Tao<sup>2</sup>, Wang Jiandong<sup>1</sup>

(1. College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, China;

2. Information Technology Research Base of Civil Aviation Administration of China, Tianjin, 300300, China)

**Abstract:** To predict and prevent the noise around the airport becomes an urgent problem. A new airport noise prediction model based on fuzzy support vector regression is established on the existing generic software for noise calculation. To integrate into the fuzzy support vector regression algorithm, the fuzzy membership of each sample is determined by its local outlier factor. Finally, the prediction accuracy, noise immunity, generalization ability of the proposed model are validated on the historic flight data of an airport. Experiments show that the fuzzy support vector regression algorithm based on local outlier factor can effectively predict the noise levels around airports, and is more accurate and better noise immunity than the standard support vector regression.

**Key words:** airport noise prediction; support vector machine; fuzzy support vector regression; fuzzy membership; local outlier factor

中国民航的迅猛发展令世界瞩目,不断新建、扩建的机场和持续增长的航空运输量,也使得机场噪声问题愈发严重,并成为影响民航业持续健康发展的障碍之一。因此,迫切需要利用先进的信息技术对机场噪声进行科学预测,进而为民航相关部门提供决策支持。目前,国内外主要采用噪声预测模型结合噪声预测软件对机场噪声作出预测,如集成

噪声模型(Integrated noise model, INM)<sup>[1]</sup>, Noisemap<sup>[2]</sup>等。随着 INM 的普及,INM 的问题不断涌现,如 Asensio 等人<sup>[3]</sup>指出 INM 无法考虑飞机滑行的问题,存在预测偏差,提出了滑行阶段的噪声强弱等级及传播方向预测方法。Yang 等人<sup>[4]</sup>指出 INM 中的模型有时不能给出接近真实环境的飞机噪声预测结果。因此,有必要针对中国机场噪

**基金项目:**国家自然科学基金重大(61139002)资助项目;中国民航信息技术科研基地开放基金(CAAC-ITRB-201203)资助项目。

**收稿日期:** 2012-11-02; **修订日期:** 2013-06-26

**通信作者:** 陈海燕,女,讲师, E-mail: chenhaiyan@nuaa.edu.cn。

声数据特点,建立适合中国的机场噪声预测模型。

机器学习方法也是对噪声进行预测的方法之一。Van<sup>[6]</sup>的研究表明,利用从实测数据中挖掘出的规律、模式与声学理论计算模型进行结合,可以有效提高噪声的预测准确率;Hameta 等人<sup>[6]</sup>指出噪声预测模型应当根据噪声源等各种因素的变化进行修正;尹志宇<sup>[7]</sup>利用 L-M 优化算法的多层神经网络预测模型对道路交通噪声进行了预测;Basner 等人<sup>[8]</sup>则通过 Markov 过程对飞机噪声的影响进行了预测。

支持向量机(Support vector machine, SVM)作为一种机器学习新方法,已表现出很多优于其他算法的性能,被广泛地应用于统计分类以及回归分析中。然而,由于机场噪声数据来自现实世界中实际监测,训练集中会含有大量不确定因素及不利于数据挖掘的孤立点,在这样的数据集上进行模型训练,结果往往差强人意。因此,为了提高 SVM 的抗噪能力和预测精度,本文将采用模糊支持向量机(Fuzzy support vector machine, FSVM)<sup>[9-10]</sup>建立机场噪声预测模型。根据中国机场噪声数据的特点,设计更高效的基于局部孤立因子的模糊隶属度计算方法,并验证模型的有效性和先进性。

## 1 模糊支持向量回归原理

### 1.1 支持向量回归

对于训练集  $T = \{(x_i, y_i)\}$ ,  $x_i \in \mathbf{R}^n$ ,  $y_i \in \mathbf{R}$ ,  $i = 1, 2, \dots, l$ , 支持向量回归(Support vector regression, SVR)的过程是通过一个非线性的映射  $\varphi$ , 将输入空间映射到一个高维空间中,并在这个空间内进行线性回归。这个问题可以转化为一个带约束的最优化问题

$$\min Q(\omega, \xi) = \frac{1}{2}\omega^2 + C \sum_i (\xi_i^- + \xi_i^+) \quad (1)$$

$$\text{s. t.} \begin{cases} y_i - (\omega \cdot \varphi(x)) - b \leq \epsilon + \xi_i^- \\ (\omega \cdot \varphi(x)) + b - y_i \leq \epsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \quad i = 1, 2, \dots, l \end{cases}$$

式中:  $\epsilon$  为不敏感损失函数;  $\xi_i^-$ ,  $\xi_i^+$  为松弛因子,其描述了样本点偏离回归间隔所带来的损失;  $C$  为各样本松弛项的惩罚因子,它表示对支持向量回归估计函数的误差大于  $\epsilon$  的样本点的惩罚。 $C$  的值越大,表示对该样本点带来的损失越重视,以至于对于目标函数的损失也越大。因此,对于偏离回归间隔的点,它们回归产生的误差较大,固定的惩罚因子  $C$  会使回归函数对于这类点非常敏感,进而去拟合这些点,如图 1 所示。从图中可以看出,由

于孤立点的加入,回归间隔会向靠近孤立点的方向移动,从而影响最后回归的精度。

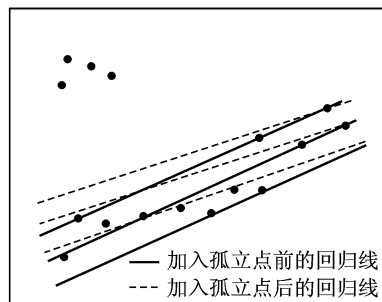


图 1 在  $\epsilon$ -SVR 中同等对待孤立点而带来的过拟合

### 1.2 模糊支持向量回归

为了提高 SVM 的抗噪声能力和分类精度, Lin 等人<sup>[9]</sup>将模糊隶属度的概念引入到 SVM 中,提出了模糊支持向量机。而模糊支持向量回归(Fuzzy support vector regression, FSVR)将模糊隶属度的概念引入到 SVR 中,为每个样本引入模糊隶属度  $\mu_i$ , 表示该样本对回归结果的重要程度。模糊化的输入样本集可表示为  $\{(x_1, y_1, \mu_1), (x_2, y_2, \mu_2), \dots, (x_i, y_i, \mu_i)\}$ , 其中  $0 < \mu_i < 1$ , 则上述的最优化问题可转化为

$$\min \frac{1}{2}\omega^2 + C \frac{1}{l} \sum_i \mu_i (\xi_i^- + \xi_i^+) \quad (2)$$

$$\text{s. t.} \begin{cases} y_i - (\omega \cdot \varphi(x)) - b \leq \epsilon + \xi_i^- \\ (\omega \cdot \varphi(x)) + b - y_i \leq \epsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \quad i = 1, 2, \dots, l \end{cases}$$

其 Wolfe 对偶可表示为

$$\max W = \sum_{i=1}^l (\alpha_i + \alpha_i^*) y_i - \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) (x_i \cdot x_j) \quad (3)$$

$$\text{s. t.} \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq \mu_i C \quad i = 1, \dots, l \end{cases}$$

可见,实施 FSVR 的关键问题在于模糊隶属度函数的确定。

## 2 基于局部孤立因子的模糊隶属度函数

针对机场噪声实测数据存在大量孤立点的问题,可通过设计一种能反应样本孤立程度的模糊隶属度函数来消除或减少孤立点对回归结果的影响,亦即,通过计算样本的孤立程度来确定其模糊隶属度取值。Markus 等人<sup>[11]</sup>在 2000 年提出的基于密度的孤立点识别方法,适合于比较复杂的数据分布情况,且计算复杂性不高,是公认的孤立性评价的

好方法;然而,研究表明,该方法只对高密度规则模式的孤立点有效,却不能处理低密度规则模式的孤立性评价,这样,对于一些处于边缘的密度较低的孤立点就无法得到正确的评价。在基于密度方法的基础上,通过对局部孤立因子(Local outlier factor, LOF)的定义进行修改,可以使 LOF 方法适用于评价各种密度规则模式样本。LOF 表示了一个对象相对于它的邻居的孤立程度。本文提出了通过计算 LOF 来确定每个样本的模糊隶属度,以此来刻画样本偏离回归间隔的程度。通过  $k$  距离、 $k$  距离邻居、可达距离、局部可达密度等概念来描述样本的邻居及样本与邻居之间的紧密关系。

**定义 1**  $k$  距离( $k$ -distance)。距离样本最近的  $k$  个邻居中距离最远的那个邻居与样本的距离,计算  $k$ -距离的目的是确定样本附近的邻居构成,以及判断样本周围的稀疏度。

**定义 2**  $k$ -距离邻居。样本  $p$  的  $k$ -距离邻居是指  $p$  的邻居中那些与  $p$  的距离不大于  $p$  的  $k$ -距离的所有邻居集合,定义为

$$N_k(p) = \{q \mid q \in D \setminus \{p\}, \text{dist}(p, q) \leq k\text{-distance}(p)\} \quad (4)$$

**定义 3** 可达距离。样本  $p$  关于  $o$  的可达距离为  $o$  的  $k$  距离和  $p$  到  $o$  距离中的较大值,即

$$\text{reach-dist}_k(p, o) = \max\{k\text{-distance}(o), d(p, o)\} \quad (5)$$

样本  $p$  距离  $o$  较近取  $o$  的  $k$  距离,反之,将  $p$  和  $o$  之间的距离作为  $p$  相对  $o$  的可达距离。

**定义 4** 局部可达密度。样本  $p$  局部可达密度的定义为

$$\text{lrd}_k(p) = \frac{|N_k(p)|}{\sum_{o \in N_k(p)} \text{reach-dist}_k(p, o)} \quad (6)$$

即  $p$  的  $k$  近邻集合中元素个数除以  $p$  到集合中所有点的相对可达距离之和。

**定义 5** 局部孤立因子。通过计算样本点周围  $k$  个点的平均局部可达密度和该点的局部可达密度的比来衡量该样本点的孤立度,比值接近 1,说明该点和邻居样本有着良好的关系,在数据分布上非常接近,并能很好地合并在一个类里面。比值大于 1 代表该点的相对于周围邻居的孤立度更高,计算公式如下

$$\text{lof}_k(p) = \frac{\sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)}}{|N_k(p)|} \quad (7)$$

基于局部孤立因子的隶属度函数定义如下

$$\mu_i = \begin{cases} (1 - \theta)^m + \sigma \overline{\text{lof}} < \text{lof}_k(p) \leq \text{lof}_{\max} \\ 1 - \theta & \text{lof}_{\min} \leq \text{lof}_k(p) \leq \overline{\text{lof}} \end{cases} \quad (8)$$

式中:  $\theta = \frac{\text{lof}_k(p) - \text{lof}_{\min}}{\text{lof}_{\max} - \text{lof}_{\min}}$ ,  $\text{lof}_k(p)$  为样本点的局部孤立因子,  $\text{lof}_{\min}$ ,  $\text{lof}_{\max}$  和  $\overline{\text{lof}}$  分别为样本局部孤立因子中的最小值、最大值和均值;  $\sigma < 1$ , 为足够小的正实数;  $m \geq 2$ , 基于局部孤立因子的模糊隶属度函数的含义如图 2 所示。从图中可以看出,样本的局部孤立因子越大,隶属度下降的越快,当  $\text{lof}_k(p)$  接近  $\text{lof}_{\max}$  时,隶属度接近一个非常小的正实数  $\sigma$ ,即越孤立的样本它对回归模型的影响越小。

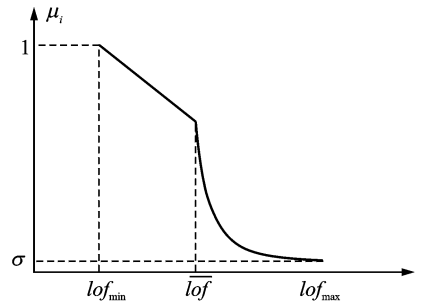


图 2 模糊隶属度函数示意图

### 3 基于 FSVR 的机场噪声预测模型

#### 3.1 数据集的准备

由于中国对机场噪声研究的起步较晚,目前还没有积累起相关的数据集。本文采用的数据集包含了监测点属性,如点坐标;飞过监测点上方单个飞行事件的相关属性,如飞行类型(起、降、盘旋)、飞机型号、发动机推力、航速、飞行高度、斜距、早中晚各时段的飞行次数;气象属性,如温度、湿度、压强。其中,一些属性值可从已获得的航班运行记录和局部的噪声监测数据中提取,另一些需根据其统计特性由程序模拟生成,如真实的航迹需在标准的航迹上作随机处理。

#### 3.2 孤立因子的计算

根据第 2 节提出的局部孤立因子的算法,计算上述数据集中每个样本的局部孤立因子,其中  $k$  值取 50。将数据集中的点通过主成分分析(Principal component analysis, PCA)降成三维并投影,如图 3 所示,其中较大的描点为数据集中局部孤立因子值较高的 100 个点。

将基于密度的孤立因子算法与 LOF 算法在上述数据集上进行计算效果的比较,用算得的孤立因

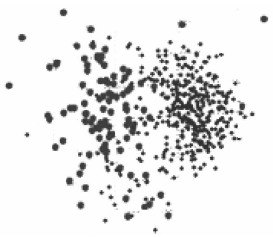


图 3 基于局部孤立因子的孤立点投影

子较大的前 100 个样本中包含的噪声样本个数来表示算法的孤立点识别能力,比较结果如表 1 所示。

表 1 不同孤立因子算法对比

算法名称	样本数	噪声点数	识别数	命中率/%
LOF	500	100	81	81
密度法	500	100	62	62

从表 1 可以看出,局部孤立因子算法由于引入了近邻的概念,能更好地识别位于样本类边缘的孤立点,因此比基于阈值的密度孤立因子算法有更好的识别效果,也更适用于机场噪声实测数据的样本分布规则。

### 3.3 FSVR 模型训练

在完成了孤立点清理和标准化的数据集上,设置隶属度函数参数  $m=10$ ,计算每个数据样本的隶属度,得到模糊数据集  $\{(x_1, y_1, \mu_1), (x_2, y_2, \mu_2), \dots, (x_i, y_i, \mu_i)\}$ 。在该模糊数据集上,采用 FSVR 算法,训练得到机场噪声预测模型,其中核函数采用径向基核(Radial basis function, RBF),并通过十折交叉验证(Cross validation)方法搜索到最佳的核函数参数  $g=0.5$ ,惩罚因子  $C=2$ 。

## 4 实验及结果分析

### 4.1 预测结果分析

为了验证基于局部孤立因子的 FSVR 的预测效果,利用训练出的模型在测试样本集上作预测,预测效果如图 4 所示。

从图中可以看出,单个点的预测误差基本控制在  $6\sim 7$  dB。考虑到训练集样本数量有限,如适当增加样本数量,将会进一步提高预测结果的精度。实践证明了通过 FSVR 对机场噪声进行预测的可行性。

### 4.2 抗干扰性实验结果分析

为了验证改进的模糊支持向量回归对噪声点的拟合效果,分别训练了  $\epsilon$ -SVR 和 FSVR 两个预测模型,拟合效果如图 5(a,b)所示。

从图中可以看出,随着噪声点的加入,  $\epsilon$ -SVR

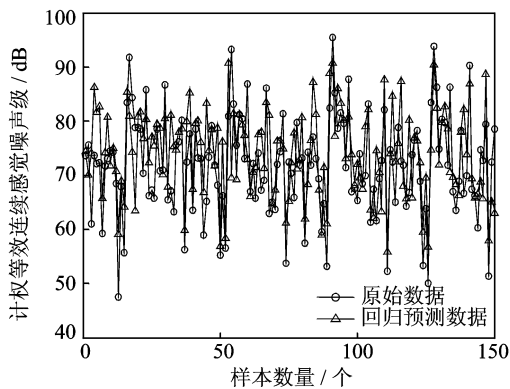
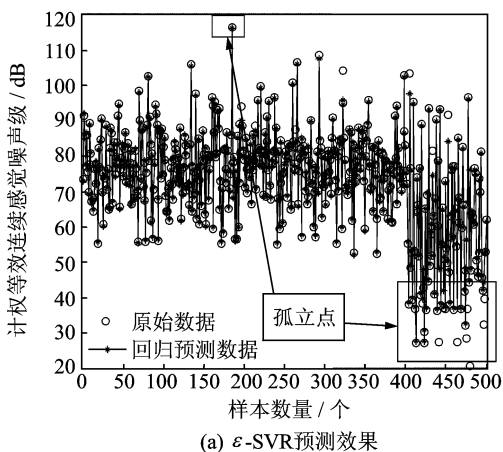
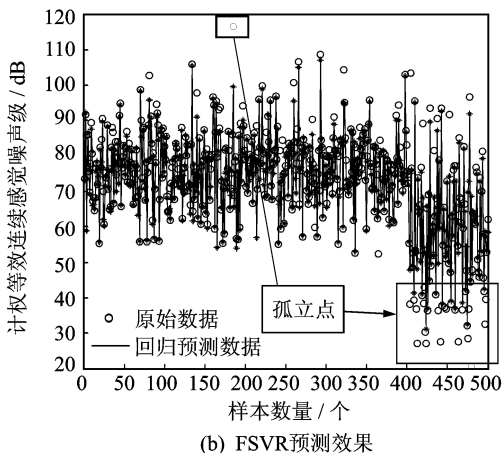


图 4 基于 LOF 的 FSVR 在测试集上的预测效果



(a)  $\epsilon$ -SVR 预测效果



(b) FSVR 预测效果

图 5  $\epsilon$ -SVR 和 FSVR 预测效果对比

对噪声点更为敏感,回归间隔向着噪声点的方向移动,在训练集上出现了过拟合现象,泛化能力大大降低;而 FSVR 能在一定程度上抑制噪声点的影响,表现稳定,具有更好的抗干扰能力。

### 4.3 FSVR 泛化能力验证

为了证明基于局部孤立因子的 FSVR 有更好的泛化能力,将其预测结果与加入噪声样本前的  $\epsilon$ -SVR、加入噪声样本后的  $\epsilon$ -SVR、基于密度法的 FSVR 的预测结果和基于紧密度的 FSVR<sup>[12]</sup>作对比,其预测均方误差(Mean squared error, MSE)结

果如表2所示。

表2 不同SVR预测结果对比

算法	加入噪声前的 $\epsilon$ -SVR	加入噪声后的 $\epsilon$ -SVR	基于密度法的 FSVR	基于紧密度的 FSVR	基于LOF的 FSVR
MSE	65.240 1	70.813 8	61.243 3	58.967 6	55.618

从表2的结果可以看出,随着噪声样本的加入,传统的支持向量回归方法  $\epsilon$ -SVR 的预测精度有所降低;而在包含噪声数据的数据集上,FSVR 的预测效果要明显好于  $\epsilon$ -SVR,说明随着模糊隶属度的加入使算法的抗干扰性增强;此外,本文提出的基于LOF的FSVR较基于密度法的FSVR能更好地描述样本偏离回归间隔的程度,从而给予样本更合理的隶属度,使得模型具有更高的预测精度和泛化能力。

另外,FSVR的算法复杂度主要取决于模糊隶属度的计算和FSVR模型的训练。表2中的3种FSVR在模糊隶属度计算时具有相同的复杂度;而在FSVR训练时与标准  $\epsilon$ -SVR 具有相同的复杂度。可见,本文所提出的基于LOF的FSVR能在不增加时间复杂度的情况下提高模型的预测精度和抗噪能力。

## 5 结束语

本文提出了一种基于局部孤立因子的模糊支持向量回归方法,并将其运用于机场噪声预测中。针对机场噪声实测数据含孤立点多的问题,提出基于局部孤立因子来计算样本的模糊隶属度;利用该值对传统支持向量回归方法中的惩罚因子进行加权,调节不同样本点偏离回归间隔带来的损失程度,使得模型具有更高的容错能力和更高的泛化能力。最后通过对比实验证明了该方法相比其他回归方法有着更高的预测精度,较Noisemap等软件有更好的可操作性和可扩展性。如何结合现实数据,并从中挖掘出更一般的关联关系,从而构建更精确的机场噪声预测模型是要进一步研究的方向。

## 参考文献:

[1] David W F, Gulding J, Dipardo J. Review of integrated noise model (INM) equations and processes [R]. NASA/CR-2003-212414,2003.

[2] Consulting W. The data entry component for the noisemap suite of aircraft noise models [EB/OL]. [2012-01-21]. <http://wasmerconsulting.com/baseops.htm>.

[3] Asensio C, Pav I, Ruiz M, et al. Estimation of directivity and sound power levels emitted by aircrafts during taxiing, for outdoor noise prediction purpose [J]. Applied Acoustics, 2007, 68(10): 1263-1279.

[4] Yang Yingjie, Chris H. Airport noise simulation using neural networks [C] // 2008 International Joint Conference on Neural Networks. Piscataway, NJ: IEEE, 2008: 1917-1923.

[5] Van F. Criteria for wind farm noise: Lmax and Lden [C] // Acoustics 08 Paris Press Conference. EAA, Paris: [s. n.], 2008: 4043-4048.

[6] Hameta J, Besnardb F, Doisy S, et al. New vehicle noise emission for French traffic noise prediction [J]. Applied Acoustics, 2010, 71(9): 861-869.

[7] 尹志宇. 基于L-M神经网络的道路交通噪声预测研究 [J]. 中国环境监测, 2009, 25(4): 84-87.  
Yin Zhiyu. Study on traffic noise prediction based on L-M neural network [J]. Environmental Monitoring in China, 2009, 25(4): 84-87.

[8] Basner M, Siebert U. Markov processes for the prediction of aircraft noise effects on sleep [J]. Medical Decision Making, 2010, 30(2): 275-289.

[9] Lin C F, Wan S D. Fuzzy support vector machines [J]. IEEE Trans on Neural Networks, 2002, 13(2): 464-471.

[10] 陈海燕, 王建东, 顾彬. 基于融合先验知识SVM的航班延误预警模型 [J]. 南京航空航天大学, 2009, 41(2): 243-247.  
Chen Haiyan, Wang Jiandong, Gu Bin. Early warning model of flight delay based on SVM with incorporated prior knowledge [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2009, 41(2): 243-247.

[11] Markus B, Kriegel H, Raymond T, et al. Identifying density-based local outliers [C] // Proc of the ACM SIGMOD Conference. New York, USA: ACM Press, 2000: 427-438.

[12] 张翔, 肖小玲, 徐光祐. 基于样本之间紧密度的模糊支持向量机方法 [J]. 软件学报, 2006, 17(5): 951-958.  
Zhang Xiang, Xiao Xiaoling, Xu Guangyou. Fuzzy support vector machine based on affinity among samples [J]. Journal of Software, 2006, 17(5): 951-958.