

# 基于多核最小二乘支持向量回归的 TDOA-DOA 映射方法

张峰 陈华伟 李妍文

(南京航空航天大学电子信息工程学院, 南京, 210016)

**摘要:** 基于到达时间差 (Time difference of arrival, TDOA) 估计的方法是声源波达方向 (Direction of arrival, DOA) 估计中的一类重要方法。其中由 TDOA 到 DOA 的映射是该类方法的关键步骤。本文提出了一种基于多核聚类最小二乘支持向量回归 (Least-squares support vector regression, LS-SVR) 的 TDOA-DOA 映射方法, 并且分析了其稀疏化处理后的性能。为了提高混响噪声环境下的 TDOA-DOA 映射性能, 本文还给出了一种基于归一化中值滤波的 TDOA 估计离群值消除方法。仿真结果表明, 本文提出的方法要优于现有的最小二乘方法以及单核 LS-SVR 方法。

**关键词:** 声源波达方向估计; 到达时间差估计; 最小二乘支持向量回归; 多核学习

**中图分类号:** TN911.7      **文献标志码:** A

## TDOA-DOA Mapping Using Multi-kernel Least-Squares Support Vector Regression

Zhang Feng, Chen Huawei, Li Yanwen

(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China)

**Abstract:** In sound source direction of arrival (DOA) estimation, one of the typical methods is based on the time difference of arrival (TDOA). For the TDOA-based sound source DOA estimation, the TDOA-DOA mapping is a crucial step. Here, we propose a TDOA-DOA mapping approach based on the multi-kernel least-squares support vector regression (LS-SVR), and also analyze its performance with sparsification. In addition, we present an outlier detection method based on the normalized median filtering to post-process the TDOA estimation for improving the performance of TDOA-DOA mapping in noisy reverberant environments. Simulation results show that the proposed method is superior to its counterparts, such as LS and single-kernel LS-SVR methods.

**Key words:** sound source DOA estimation; TDOA estimation; least-squares support vector regression (LS-SVR); multi-kernel learning

## 引 言

声源波达方向 (Direction of arrival, DOA) 估计是音频与语音信号处理领域的一个重要研究方向, 在视频会议系统<sup>[1]</sup>、机器人听觉<sup>[2]</sup>和说话人识别系统<sup>[3]</sup>等诸多领域具有广泛应用。目前人们已提出了

多种声源 DOA 估计方法<sup>[4-9]</sup>,其中典型的声源 DOA 估计方法包括基于到达时间差(Time difference of arrival, TDOA)的方法、基于可控响应功率(Steered response power, SRP)的方法<sup>[4-5]</sup>以及高分辨空间谱估计方法<sup>[6-7]</sup>等。相比较,SRP 方法和高分辨空间谱估计法需要对目标声源的方位进行全空间搜索,因此计算量较大。而基于 TDOA 估计的方法,其计算复杂度相对较小。因此,基于 TDOA 估计的方法得到了广泛研究,成为实际应用中的一类重要方法。基于 TDOA 估计的声源 DOA 估计方法利用互相关方法估计出声音信号到达各麦克风的时间差,从而估计出声源 DOA。该方法通过两步实现,第 1 步是 TDOA 估计(亦称为时延估计),第 2 步为 TDOA-DOA 映射。目前人们提出了很多时延估计的方法提高混响噪声环境下的声源 DOA 估计性能。相比较而言,对于 TDOA-DOA 映射的研究尚不够深入。典型的 TDOA-DOA 映射方法包括最大似然法和最小二乘法。与最小二乘法相比最大似然法的缺点是需要已知估计量的统计特征,而这些信息在实际应用中往往无法预知。因此最小二乘法得到了广泛应用<sup>[10]</sup>。最小二乘映射方法存在一些局限性:(1)对于非线性阵列不一定能得到闭式解,因而需要进行搜索求解,这使得计算量增大;(2)在声源位于阵列端射方向附近时其性能会急剧恶化。为克服这一问题,文献[11]提出了采用最小二乘支持向量回归(Least-squares support vector regression, LS-SVR)机构造非线性映射,取得了比最小二乘法更高的声源 DOA 估计精度。此外文献[12]还分析了核函数的选取对 LS-SVR 映射性能的影响。

现有的基于 LS-SVR 的 TDOA-DOA 映射方法仅仅局限于单核函数。单核函数方法虽然可以解决非线性模式分析问题,但在数据异构不规则、样本不平坦分布等复杂情形下,其效果不够理想。多核函数的方法可以克服这一缺点<sup>[13]</sup>。本文研究了基于多核函数 LS-SVR 的 TDOA-DOA 映射方法以提升声源 DOA 的估计性能。此外针对混响噪声环境 TDOA 估计精度较低的情况,给出了一种基于归一化中值滤波的 TDOA 估计值野点消除方法。考虑到多核 LS-SVR 方法的训练集均为支持向量,本文通过稀疏化处理降低了训练集中冗余支持向量的计算量。

## 1 基于 LS-SVR 的声源 TDOA-DOA 映射方法

考虑由  $M$  个全向麦克风组成的均匀线阵,相邻阵元之间的距离为  $d$ 。在远场情况下,定义  $\theta(-90^\circ \leq \theta \leq 90^\circ)$  为声源方向与阵列法向之间的夹角。输出信号可表示为

$$x_m(n) = \delta_m s[n - \tau_{1m}(\theta)] + v_m(n) \quad (1)$$

式中: $x_m(n)$ ,  $m=1, \dots, M$  为麦克风输出信号; $\delta_m$  为传播过程中的衰减因子; $s(n)$  为声源信号; $\tau_{1m}(\theta) = (m-1)d\sin\theta/c$  为第  $m$  个麦克风和参考麦克风之间的相对时延,其中  $c$  为声速, $v_m(n)$  为第  $m$  个麦克风的加性噪声。在运用 LS-SVR 估计声源 DOA 时,首先由模型参数训练出 LS-SVR。然后运用 LS-SVR 模型来估计声源 DOA<sup>[11]</sup>。其中的关键步骤是获取麦克风的观测特征,这里可以将 TDOA 作为特征来构造 LS-SVR。基于 LS-SVR 的映射方法属于离线工作模式。对于训练角度  $\theta_k$ ,其 TDOA 特征向量  $\mathbf{I}_k$  ( $k=1, \dots, N$ ) 可表示为

$$\mathbf{I}_k = [\tau_{12}, \dots, \tau_{1M}, \tau_{23}, \dots, \tau_{2M}, \dots, \tau_{(M-1)M}]^T \quad (2)$$

式中: $\tau_{ij} = (j-i)d\sin\theta/c$  为离线条件下第  $i$  和第  $j$  个麦克风之间的 TDOA 估计值。通过训练集 TDOA,可以由式(3)得到 LS-SVR 模型的系数  $\mathbf{a}, b$ , 则

$$\begin{bmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & \mathbf{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{\Theta} \end{bmatrix} \quad (3)$$

式中: $\mathbf{I}$  为全“1”列向量; $\mathbf{I}$  为  $N$  阶单位矩阵; $\mathbf{\Omega}$  中元素  $\Omega_{kl} = K(\mathbf{I}_k, \mathbf{I}_l)$ ,  $K(\cdot, \cdot)$  为核函数; $\mathbf{a} = [\alpha_1, \dots, \alpha_N]$ 。文献[12]的研究结果表明径向基核函数的性能较好,其定义为

$$K(\mathbf{I}_k, \mathbf{I}_l) = \exp\left\{-\frac{\|\mathbf{I}_k - \mathbf{I}_l\|}{(2\sigma^2)}\right\} \quad (4)$$

式中“ $\|\cdot\|$ ”为  $l_2$  范数。在估计声源 DOA 的过程中,首先通过观测麦克风估计出 TDOA 的特征向量,则

$$\mathbf{\Gamma} = [\hat{\tau}_{12}, \dots, \hat{\tau}_{1M}, \hat{\tau}_{23}, \dots, \hat{\tau}_{2M}, \dots, \hat{\tau}_{(M-1)M}]^T \quad (5)$$

式中:  $\hat{\tau}_{ij}$  为第  $i$  和第  $j$  个麦克风之间的 TDOA 估计值。本文选用相位变换加权的广义互相关法(Generalized cross correlation with PHase transform, GCC-PHAT)<sup>[14]</sup>来估计 TDOA,以提高算法的抗干扰能力<sup>[15]</sup>。运用先前得到的 LS-SVR 模型,则声源 DOA 估计为

$$\hat{\theta}_{\text{LS-SVR}} = \sum_{k=1}^N \alpha_k K(\mathbf{\Gamma}_k, \mathbf{\Gamma}) + b \quad (6)$$

## 2 基于多核 LS-SVR 的 TDOA-DOA 映射方法及其稀疏化分析

针对声源 DOA 估计中的 TDOA-DOA 映射问题,本文提出了基于中值滤波的时延离群值消除方法以及基于聚类方法的多核 LS-SVR 映射方法,并在此基础上对多核 LS-SVR 映射方法进行稀疏化分析。

### 2.1 基于中值滤波的 TDOA 估计离群值处理

随环境混响时间的增大和信噪比的降低,TDOA 估计值的误差会增大,从而导致特征向量  $\mathbf{\Gamma}$  的准确度降低。由于标准 LS-SVR 缺乏稀疏性,易受异常估计值的影响,这时若直接进行 TDOA-DOA 映射,最终的 DOA 估计精度往往较差。为有效改善少量离群值所造成的 TDOA 映射 DOA 时的不稳定性,本文采用中值滤波技术<sup>[16]</sup>消除 TDOA 特征向量  $\mathbf{\Gamma}$  中的孤立离群值。由式(3)可知,用于 LS-SVR 映射的 TDOA 分量之间存在一定的比例关系,因此可以根据此比例关系构造一个较为宽松的约束条件,即对归一化后的 TDOA 特征向量  $\mathbf{\Gamma}'$  提出一个门限。门限之外的 TDOA 分量,用归一化后的 TDOA 特征向量  $\mathbf{\Gamma}''$  的中值代替,以减小离群值的影响。根据这一思想,这里采用归一化后中值滤波消除离群值的方法。在一个  $M$  元线性阵列中,每一个训练角度  $\theta_n (n=1, \dots, N)$  的 TDOA 特征向量可以表示为  $\mathbf{\Gamma}_n = [\tau_{ij}]$ ,  $1 \leq i < j \leq M$ , 输入向量  $\mathbf{\Gamma}_n$  的维数  $\lambda = C_M^2$ 。各 TDOA 分量之间的比例关系为  $\boldsymbol{\varepsilon} = [j-i]$ , 则除以比例系数后的 TDOA 特征向量可表示为

$$\mathbf{\Gamma}'_n = \frac{\mathbf{\Gamma}'_n}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \tau_{ij} \\ j-i \end{bmatrix} = [\tau'_{ij}] \quad (7)$$

计算除以比例系数后 TDOA 分量的中值  $\text{median}\{\tau'_{ij}\}$ ,  $1 \leq i < j \leq M$ , 则归一化后 TDOA 特征向量可表示为

$$\mathbf{\Gamma}''_n = \frac{\mathbf{\Gamma}'_n}{\text{median}[\tau'_{ij}]} = \begin{bmatrix} \tau'_{ij} \\ \text{median}[\tau'_{ij}] \end{bmatrix} = [\tau''_{ij}] \quad (8)$$

将  $\tau''_{ij}$ ,  $1 \leq i < j \leq M$  与门限  $[0.75, 1.25]$  相比较,若在门限内,则不作处理;若在门限之外,则用 1 代替,得到  $\mathbf{\Gamma}'''_n$ 。则经过中值滤波后的 TDOA 向量为

$$\hat{\mathbf{\Gamma}}_n = \mathbf{\Gamma}'''_n \cdot \boldsymbol{\varepsilon} \quad n=1, \dots, N \quad (9)$$

通过仿真实验对该方法的性能进行了验证。这里选用尺寸为  $5 \text{ m} \times 4 \text{ m} \times 3 \text{ m}$  的矩形混响房间。麦克风阵列列为四元均匀直线阵列,阵元间距为  $4 \text{ cm}$ ,阵列离地高度为  $0.5 \text{ m}$ ,声源与麦克风阵列中心的距离为  $2 \text{ m}$ ,如图 1 所示。设定  $\text{SNR} = 20 \text{ dB}$ ,  $\text{RT}_{60} = 210 \text{ ms}$  时,  $\text{RT}_{60}$  为混响时间,运用 GCC-PHAT 方法估计 TDOA,在未采用中值滤波方式及采用中值滤波方式处理 TDOA 的两种情况下,分别采用最小二乘法、LS-SVR 这两种映射方法,以 50 次蒙特卡洛实验的均方根误

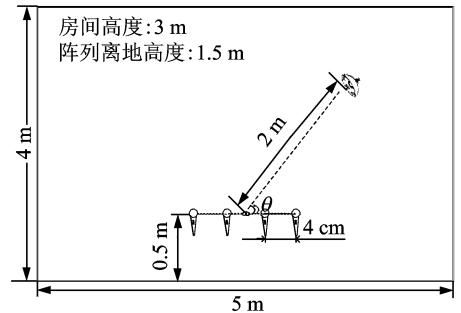


图 1 仿真实验麦克风阵列的配置图

Fig. 1 Configuration of microphone array in simulation room

差为评价指标,仿真结果如图 2 所示。

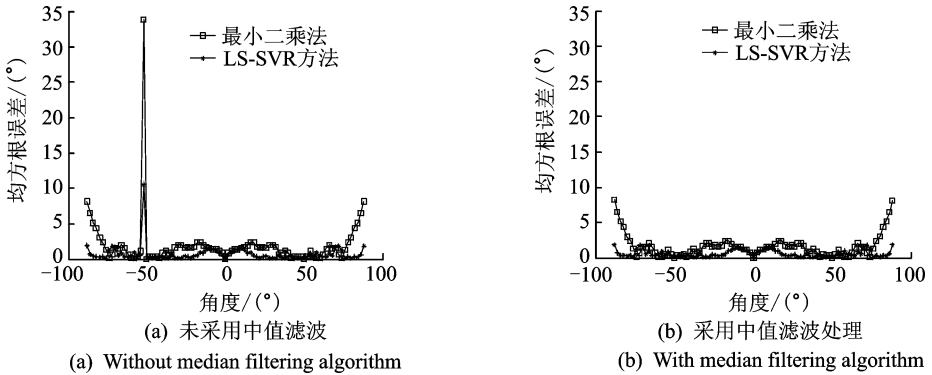


图 2 采用中值滤波处理前后,最小二乘法与 LS-SVR 方法的性能对比

Fig. 2 Performance comparison of least squares method and LS-SVR method with or without median filtering algorithm

由图 2(a)可知,受混响噪声影响,声源的入射角度在 $-50^\circ$ 附近出现了较大的估计误差。这是由于 TDOA 估计的精度较差,导致映射 DOA 时出现了较大误差,从而出现了估计结果不稳定的情况。尤其是采用最小二乘法时,最大均方根误差已超过 $30^\circ$ 。虽然此时 LS-SVR 方法要优于最小二乘法,但由于离群值的出现,该方法也出现了较大的估计误差。而在图 2(b)中,最小二乘法以及 LS-SVR 方法的映射性能都没有很大起伏。对比图 2(a)与图 2(b)可知,在相同的仿真条件下,采用中值滤波方式处理 TDOA 可有效消除 TDOA 中的离群值,提高 TDOA-DOA 映射性能的鲁棒性。

## 2.2 聚类多核 LS-SVR 映射方法

核方法虽然能在一定程度上解决非线性模式分析问题,但当样本特征含有异构信息、样本规模很大且数据在高维空间中分布不平坦时,采用单个核函数进行直接映射往往不能达到实际的应用需求,因此出现了很多核组合方法的研究。多核学习方法是一类性能更好的基于核的学习模型,采用多核学习方法可以提升映射性能。在声源 DOA 估计中,由于声源信号受到混响和噪声的影响,估计时延准确度因环境不同而存在一定差异;而且对于不同的声源 DOA,时延估计的准确度也有一定区别。这时采用单核 LS-SVR 不能很好地对 TDOA 进行拟合,因此可以将多核方法应用到声源 DOA 估计中,以提高估计精度。多核(Multiple kernel, MK)学习方法的核心思想就是将  $R$  个核函数进行凸组合。只要满足 Mercer 条件,组合之后的函数就可看作是一种核函数<sup>[17]</sup>。如

$$K(\mathbf{\Gamma}_i, \mathbf{\Gamma}_j) = \sum_{k=1}^R \beta_k K_k(\mathbf{\Gamma}_i, \mathbf{\Gamma}_j) \quad \beta_k \geq 0 \quad (10)$$

式中: $K_k$  为满足 Mercer 条件的核函数, $\beta_k$  为每个核函数的权重系数。由 Mercer 条件可知,凸组合之后的  $K(\mathbf{\Gamma}_i, \mathbf{\Gamma}_j)$  仍为核函数。对多个核函数进行凸组合的方法有很多,最简单的方法就是线性组合。虽然线性组合方式可能会丢失某些核函数信息<sup>[18]</sup>,但由于非线性组合的计算复杂度过高<sup>[19]</sup>,因此本文研究了线性组合核函数的性能。为将多核 LS-SVR 方法应用到声源 DOA 的 TDOA-DOA 映射中,需要对 TDOA 特征向量训练集分组,对每一个分组分别找出拟合效果最好的核函数,即对当前的第  $k$  组,令其权重系数  $\beta_k$  为 1,而其他权重系数为 0。聚类是数据挖掘领域的一种重要方法,它根据样本间关联度对样本进行分类。该方法要求同一类别内的样本相似度尽量高,而不同类别间样本相异度尽量低<sup>[20]</sup>。K-Means 聚类方法的基本思想是以追求误差函数最小值为目标将样本划分为预定的类数  $Q$ 。这里将 K-Means 聚类方法应用于声源 DOA 估计中 TDOA 的预分类。K-Means 算法的主要步骤是:先指定聚类

数目  $Q$ ,  $Q$  个初始聚类中心初值以及收敛条件; 然后根据一定的相似性度量准则, 将所有数据分配到最近的聚类中心, 形成  $Q$  类, 并将各类的平均矢量作为该类的聚类中心; 反复迭代直到达到终止条件<sup>[21]</sup>。将 K-Means 聚类算法应用到声源 DOA 估计中的 TDOA-DOA 映射问题的具体过程为: 首先将 TDOA 特征向量训练集  $\Phi = [\Gamma_1, \Gamma_2, \dots, \Gamma_n, \dots, \Gamma_N]^T$  中的样本分成  $Q$  类, 初始聚类中心集合表示为  $C = [c_1, c_2, \dots, c_k, \dots, c_Q]$ , 用  $d_{ik}(\Gamma_i, c_k)$  表示样本  $\Gamma_i$  与其所属聚类中心  $c_k$  之间的距离, 用目标函数  $J$  来表示训练集内所有样本与所属聚类中心距离的总和, 即

$$J = \sum_{k=1}^Q \sum_{i, i \in n_k} d_{ik}(\Gamma_i, c_k) \quad (11)$$

式中:  $n_k$  为第  $k$  组的样本数据个数, 目标函数  $J$  的值越小表示该聚类越紧凑, 聚类效果越好。因此可以通过追求目标函数  $J$  的极小值来选出最优聚类方案。选择欧氏距离作为样本数据  $\Gamma_i$  与所属聚类中心  $c_k$  间的相似性指标,  $\Gamma_i^{(k)}$  为属于第  $k$  组的数据样本, 则目标函数  $J$  可定义为

$$J = \sum_{k=1}^Q J_k = \sum_{k=1}^Q \left( \sum_{i, i \in n_k} |\Gamma_i^{(k)} - c_k|^2 \right) \quad (12)$$

聚类过程就是寻找最佳聚类中心  $c_k$ , 使目标函数  $J$  为最小值, 这时聚类中心为

$$c_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \Gamma_i^{(k)} \quad (13)$$

将 K-Means 聚类算法应用到声源 DOA 估计后, 得到  $Q$  个 TDOA 聚类中心。根据训练集的分类情况, 有对应类别的训练集通过经验选择法选出最优参数组合, 构造出核函数  $K^k(\Gamma_i, \Gamma_j)$ 。以欧式距离为度量标准, 计算当前 TDOA 特征向量  $\Gamma$  与每个聚类中心  $c_k$  的距离, 比较并归入距离最近的一组, 记为第  $k$  组。对于不同的 TDOA 特征向量  $\Gamma$ , 根据聚类结果, 选用对应的核函数  $K^k$ 。当  $l=k$  时,  $\beta_l=1$ ; 其他时候  $\beta_l=0, k, l=1, \dots, Q$ , 根据式(13)估计出声源 DOA, 则

$$\hat{\theta}_{\text{LS-SVR}} = \sum_{l=1}^Q \sum_{n=1}^N \alpha_n^l \beta^l K^l(\Gamma_n, \Gamma) + b^l \quad (14)$$

与最小二乘映射方法相比, 对于任意结构的阵列, LS-SVR 方法总存在闭式解, 因而无需进行搜索求解, 从而减小了计算量。仿真实验结果表明, 与单核 LS-SVR 方法相比, 多核 LS-SVR 方法进一步提升了声源 DOA 估计性能。

### 2.3 稀疏化处理

由于 LS-SVR 将标准 SVM 中的不等式约束变成等式约束。因而涉及到的所有向量都参与运算, 即都为支持向量, 这使得 LS-SVR 缺乏稀疏性。本文研究表明, 对于 TDOA-DOA 映射问题, TDOA 集合中存在一定的冗余性, 可以通过稀疏化处理剔除一部分向量, 在保证性能的前提下降低运算量。稀疏 LS-SVR 代表性的方法为 Suykens 提出的基于最小支持权重的剪枝稀疏方法<sup>[22]</sup>。该方法的核心思想为: 在 LS-SVR 中, 由公式  $\alpha_n = \gamma e_n$  可知,  $\alpha_n$  的数值与误差系数  $e_n$  成比例, 所以  $\alpha_n$  反映了训练数据的重要程度。据此, 可以通过消除一些具有较小  $|\alpha_n|$  的支持向量减少支持向量的个数, 从而得到第  $p$  次删除支持向量之后的参数以及  $b^{(p)}$ 。再由式(14)得到当前迭代次数下的声源 DOA, 则

$$\tilde{\theta}_{\text{LS-SVR}} = \sum_{l=1}^Q \sum_{n=1}^N \alpha_n^{lp} K^l(\Gamma_n, \Gamma) + b^{(p)} \quad (15)$$

综上所述, 基于稀疏多核 LS-SVR 的声源 DOA 估计算法为

#### 步骤 1 模型训练

- (1) 将声源 DOA 的分布情况离散化为  $\Theta = [\theta_1, \theta_2, \dots, \theta_N]^T$ 。
- (2) 对于每一个训练角度, 计算出加入理想语音后的 TDOA 特征向量  $\Gamma_k$ 。
- (3) 对纯净语音训练集使用 K-Means 聚类算法, 将其分成  $Q$  组并找出  $Q$  个聚类中心。

(4) 对每一个分组,通过求解式(5),得出 LS-SVR 模型的参数  $\alpha^k, b^k$ ; 构造出核函数  $K^k(\mathbf{F}_i, \mathbf{F}_j)$ 。

### 步骤2 剪枝法删减支持向量

(1) 由于  $\alpha_n = \gamma e_n$ , 因而较小的系数  $\alpha$  对应较小误差  $e$ ; 对  $|\alpha_n|$  进行排序并选取较小的比例(本文取 5%)。

(2) 剔除这些  $\alpha$  对应的训练数据,并运用原来的参数组合重新训练,得出新的  $\alpha^{k(\rho)}$ 。

(3) 查看 DOA 估计性能,并重复步骤 1(4)-步骤 2(2),直到性能满足要求,或达到预设的循环次数  $S$ 。

### 步骤3 DOA 估计

(1) 通过麦克风阵列观测数据,构造当前 DOA 的 TDOA 特征向量  $\mathbf{F}$ 。

(2) 采用中值滤波方法对 TDOA 进行预处理。

(3) 计算 TDOA 特征向量  $\mathbf{F}$  与每个聚类中心  $\mathbf{c}_k$  的距离  $d_k$ , 并归入距离最近的一组,记为第  $k$  组。

(4) 通过式(15)估计出声源 DOA。

## 3 仿真实验与分析

### 3.1 聚类多核 LS-SVR 映射方法仿真实验

首先对本文提出的聚类多核 LS-SVR 映射方法进行仿真。选用矩形混响房间尺寸为  $5\text{ m} \times 4\text{ m} \times 3\text{ m}$ 。麦克风阵列为四元均匀直线阵列,阵元间距为  $4\text{ cm}$ ,阵列离地高度为  $0.5\text{ m}$ ,声源与麦克风阵列中心的距离为  $2\text{ m}$ ,其示意图如图 1 所示。

#### 3.1.1 单核与多核映射方法的性能比较(SNR=20 dB, $RT_{60}=120\text{ ms}$ )

设定以测试角度为横坐标,50 次蒙特卡洛实验后每个角度所对应的均方根误差为纵坐标。运用 GCC-PHAT 方法获取 TDOA,以径向基核为核函数比较采用由 1 个和 4 个核函数构造的 LS-SVR 映射 DOA 时的性能,如图 3 所示。在此仿真实验条件下,由于混响和噪声对信号干扰较小,估计出的 TDOA 准确度较高。由图 3 可以看出,无论是最大均方根误差还是整体表现,由 4 个径向基核构造的多核 LS-SVR 的性能都明显优于单核 LS-SVR,尤其是在端射方向;并且由 4 个径向基核构造的 LS-SVR 映射得到的角度基本可以控制在  $1^\circ$  以内,性能较为稳定,而单核构造的 LS-SVR 得到的角度起伏较大。

#### 3.1.2 单核与多核映射方法的性能比较(SNR=10 dB, $RT_{60}=300\text{ ms}$ )

这里考虑降低信噪比,增大混响时间。以测试角度为横坐标,50 次蒙特卡洛实验后每个角度所对应的均方根误差为纵坐标。运用 GCC-PHAT 方法获取 TDOA,采用中值滤波的方法消除离群值,比较采用由 1 个和 4 个径向基核构造的 LS-SVR 映射 DOA 时的性能,如图 4 所示。在此仿真条件下,信号受混响影响较大,估计出的 TDOA 准确度下降。因而采用本文提出的中值滤波方法对 TDOA 进行处理,有效消除离群值,使得 TDOA 准确度得到一定提升。从图 4 可以看出,由 4 个径向基核构造 LS-SVR 时的性能在整体上优于由单核构造 LS-SVR 时的性能。由 4 个径向基核构造的 LS-SVR

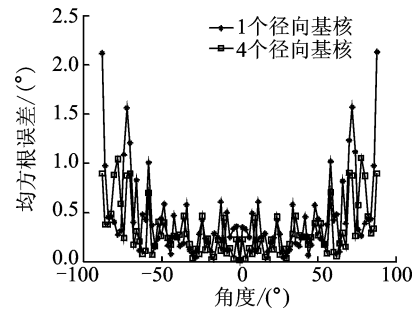


图3 单核与四核 LS-SVR 的性能比较  
Fig. 3 Performance comparison of single-kernel LS-SVR and quad-kernel LS-SVR

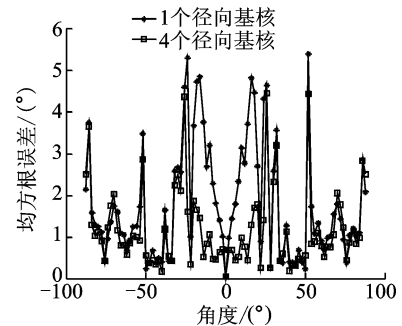


图4 消除离群值后,单核与四核 LS-SVR 的性能比较  
Fig. 4 Performance comparison of single-kernel LS-SVR and quad-kernel LS-SVR after outlier detection

的最大均方根误差也小于单核 LS-SVR。单核 LS-SVR 在优化参数过程中考虑的是最小化全局损失量, 此时的单核 LS-SVR 更加侧重边界性能, 即认为声源入射方向与阵列法向夹角较大时的误差小, 整体误差就小, 所以图 4 中单核 LS-SVR 的边界性能与四核 LS-SVR 的边界性能相当。四核 LS-SVR 的特点是可以分区段优化参数, 不同核函数针对的角度区域不同, 所以在声源入射方向与阵列法向夹角较小时, 四核 LS-SVR 的性能明显优于单核 LS-SVR, 改善程度可达  $2^\circ \sim 3^\circ$ 。

### 3.1.3 不同混响条件下单核与多核映射方法的性能比较

设定 SNR=20 dB, 无混响以及  $RT_{60}$  分别为 120, 160, 210, 250 和 300 ms 时, 以测试角度为横坐标, 50 次蒙特卡洛实验后每个角度所对应的均方根误差为纵坐标, 运用 GCC-PHAT 方法获取 TDOA, 比较采用由 1, 2, 3, 4 个径向基核构造的 LS-SVR 映射 DOA 时的性能, 如图 5 所示。由于 DOA 估计的精度受混响影响较大, 因而本文针对混响对多核 LS-SVR 的性能作了进一步的分析。对于图 5, 在整体上, 无论由单核还是多核构造的 LS-SVR, 其平均均方根误差总是随着  $RT_{60}$  的增大而增加; 在同一仿真条件下, 即具有相同 SNR 和  $RT_{60}$  时, 由多核构造的 LS-SVR 性能总是优于单核 LS-SVR, 且随着核函数数目的增多而进一步提升。

### 3.1.4 不同环境条件下各映射方法性能的综合比较

分别设定 SNR 为 5, 10 和 20 dB, 无混响以及  $RT_{60}$  分别为 120, 160, 210, 250 和 300 ms 时, 以 50 次蒙特卡洛实验后所有角度对应的平均均方根误差和最大均方根误差为评价指标, 运用 GCC-PHAT 方法获取 TDOA, 然后采用中值滤波的方法消除离群值, 采用最小二乘法 and 由 1, 2, 3, 4 个径向基核构造的 LS-SVR 进行 DOA 映射, 其性能比较如表 1 所示。

表 1 不同 SNR 以及不同  $RT_{60}$  时, 单核 LS-SVR、多核 LS-SVR 与最小二乘法的性能比较

Tab. 1 Performance comparison of single-kernel LS-SVR, multi-kernel LS-SVR and least squares method versus different SNR and reverberation time

SNR/dB	$RT_{60}/ms$	平均均方根误差/最大均方根误差				
		单核/ $(^\circ)$	双核/ $(^\circ)$	三核/ $(^\circ)$	四核/ $(^\circ)$	最小二乘法/ $(^\circ)$
5	0	0.36/1.82	0.35/1.80	0.34/1.79	0.33/1.81	0.63/4.07
	120	0.57/2.04	0.54/2.13	0.44/1.93	0.43/1.85	1.14/6.70
	160	0.69/1.82	0.63/2.02	0.53/2.11	0.52/2.10	1.44/7.31
	210	0.91/2.62	0.83/2.05	0.74/2.24	0.72/2.32	1.85/8.24
	250	1.39/3.77	1.25/3.70	1.19/3.67	1.02/3.44	2.28/8.97
	300	3.04/7.96	2.67/7.57	2.29/7.93	2.16/6.92	3.39/10.09
10	0	0.24/1.57	0.23/1.61	0.22/1.59	0.22/1.63	0.62/4.35
	120	0.48/1.95	0.43/1.61	0.36/1.55	0.35/1.46	1.07/6.49
	160	0.58/1.72	0.53/1.79	0.42/1.69	0.42/1.60	1.36/7.12
	210	0.75/2.35	0.67/1.71	0.58/2.06	0.57/2.06	1.75/8.06
	250	0.96/3.26	0.84/2.23	0.80/2.23	0.69/2.71	2.11/8.80
	300	1.79/5.39	1.63/5.50	1.29/4.97	1.17/4.45	2.77/9.71
20	0	0.23/1.92	0.23/1.99	0.22/1.98	0.22/1.96	0.72/5.02
	120	0.45/2.13	0.40/1.09	0.33/1.07	0.31/1.05	1.03/6.54
	160	0.51/1.82	0.46/1.27	0.34/1.24	0.33/1.25	1.30/7.20
	210	0.63/1.91	0.55/1.80	0.47/1.57	0.46/1.58	1.68/8.23
	250	0.80/3.19	0.71/2.11	0.66/2.15	0.56/2.88	2.03/9.02
	300	1.64/17.05	1.46/17.34	1.38/14.86	1.16/13.84	2.75/9.94

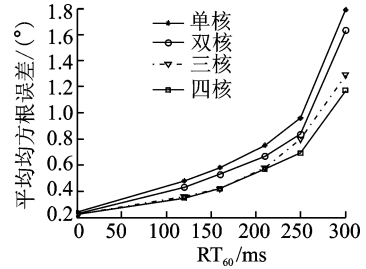


图 5 由不同个数核函数构造的 LS-SVR 性能随  $RT_{60}$  的变化

Fig. 5 Performance of LS-SVR constructed by different numbers of kernel function versus reverberation time

为全面展示各种混响噪声条件下不同映射方法的性能,表 1 给出了各种仿真条件下,多核构造的 LS-SVR、单核 LS-SVR 以及最小二乘法的映射性能。在总体上,随着 SNR 的下降以及  $RT_{60}$  的增大,各种映射方法的平均均方根误差和最大均方根误差都增大,即性能都变差。在 SNR 和  $RT_{60}$  相同时,多核 LS-SVR 的性能要优于单核 LS-SVR 以及最小二乘法;一般情况下,核的个数越多,多核 LS-SVR 的性能越好;且混响时间越大,多核 LS-SVR 映射方法的性能优势越明显。

### 3.2 稀疏化分析仿真实验

对本文提出的基于稀疏多核 LS-SVR 的声源 DOA 估计算法进行了仿真验证,在指定 SNR 和  $RT_{60}$  下,采用中值滤波方法处理 TDOA,给出了该方法随稀疏次数的性能变化情况;并与未稀疏的多核 LS-SVR、未稀疏的单核 LS-SVR 和稀疏单核 LS-SVR 等映射方法进行了比较。

#### 3.2.1 稀疏多核与未稀疏多核映射方法的性能比较

设定 SNR=20 dB,  $RT_{60}=250$  ms 时,运用 GCC-PHAT 方法获取 TDOA,使用中值滤波方法处理 TDOA,以测试角度为横坐标,50 次蒙特卡洛实验后每个角度所对应的均方根误差为纵坐标,采用 4 个径向基核构造 LS-SVR,然后比较未稀疏与稀疏 8 次时的多核 LS-SVR 性能,如图 6 所示。由图 6 可知,稀疏 8 次后的多核 LS-SVR 性能与未稀疏的多核 LS-SVR 性能十分接近,都能保持较高的准确度;尤其是在端射方向附近,这两种方法的性能几乎一致。虽然在 DOA 较小时,稀疏 8 次后的多核 LS-SVR 方法与未稀疏的多核 LS-SVR 的性能相比存在一定差异,但两者之间的差异几乎可以忽略。所以稀疏多核 LS-SVR 在降低运算量的情况下仍能保持较好的映射性能。

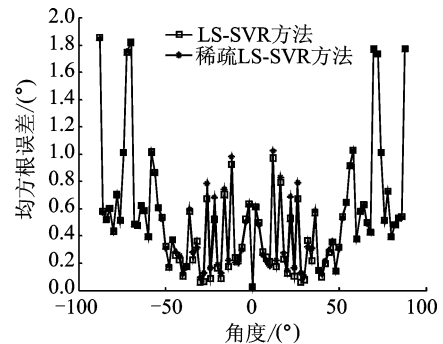


图 6 稀疏 8 次多核 LS-SVR 方法与未稀疏方法的性能比较

Fig. 6 Performance comparison of multi-kernel LS-SVR after eight-time sparsification and without sparsification

#### 3.2.2 稀疏次数对稀疏多核映射方法性能的影响

设定 SNR=20 dB,  $RT_{60}=250$  ms,运用 GCC-PHAT 方法获取 TDOA,使用中值滤波方法处理 TDOA,再采用四核 LS-SVR 映射 DOA,以 50 次蒙特卡洛实验的所有 DOA 的平均均方根误差为评价指标,稀疏 LS-SVR 方法的平均均方根误差随稀疏次数的变化情况如图 7 所示。由图 7 可知,随着稀疏次数的增多,平均均方根误差缓慢增大,尤其是稀疏次数在 5 次以内时,性能几乎保持一致。即便稀疏到 10 次时,平均均方根误差只增加了 0.05。当稀疏次数增大到 15 次时,平均均方根误差增加了 0.11。这是由于在一定范围内数据量减小时,只是剔除了一部分冗余的支持向量,对测试效果几乎无影响。而当数据量减小到一定程度时,由于训练阶段对映射曲线的学习不够全面,因而对于部分测试数据没有达到理想的映射效果,因此不能无限地进行稀疏。

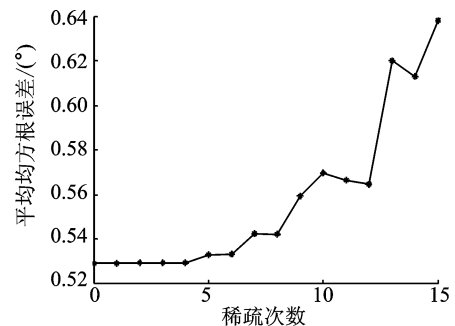


图 7 多核 LS-SVR 方法的性能随稀疏次数变化趋势

Fig. 7 Performance of multi-kernel LS-SVR method versus time of sparsification

#### 3.2.3 单核和多核映射方法随稀疏次数的性能变化

设定 SNR=20 dB,  $RT_{60}=250$  ms,运用 GCC-PHAT 方法获取 TDOA,使用中值滤波方法处理 TDOA,采用由单核、双核、三核以及四核构造的 LS-SVR 映射 DOA,并在此基础上分别进行稀疏化,以 50 次蒙特卡洛实验的所有 DOA 的平均均方



根误差和最大均方根误差为评价指标。表 2 呈现了不同个数的核函数构造的 LS-SVR 在不同稀疏次数下的性能。稀疏之前训练集的样本数目为 89, 由于每次剔除当前训练集中的 5% 的数据, 因而 1 次稀疏后支持向量比例为原支持向量集合的 95.5%。稀疏 5 次, 10 次和 15 次后支持向量比例分别为原支持向量集合的 79.8%, 64% 和 52.8%。由表 2 可知, 一般情况下, 在指定核函数个数时, 随着稀疏次数增多, 无论由多少个核构造 LS-SVR, 其性能总是随着稀疏次数缓慢变化。当稀疏次数增大到一定数目时, 其估计性能可能会急剧恶化, 如表 2 中单核和三核构造的 LS-SVR。但在稀疏 10 次以内时, 各种 LS-SVR 方法都能保持稳定的性能, 即至少可以剔除 36% 的冗余向量, 且在指定稀疏次数时, 随着核的个数增多, 多核 LS-SVR 的性能也得到提升, 例如与单核 LS-SVR 相比, 四核 LS-SVR 的平均均方根误差减少了  $0.28^\circ$  左右, 最大均方根误差减少了  $1.4^\circ$  左右。同时, 在稀疏 15 次时, 即剔除 47.2% 的冗余向量时, 四核 LS-SVR 仍能保持很好的 DOA 估计性能。总而言之, 稀疏多核 LS-SVR 方法在保持优越的 DOA 估计性能的同时, 有效减小了测试时的运算量。

表 2 不同数目核构造的 LS-SVR, 性能随稀疏次数的变化

Tab. 2 Performance of LS-SVR constructed by different numbers of kernel function versus time of sparsity ( $^\circ$ )

稀疏次数	平均均方根误差/最大均方根误差			
	单核	双核	三核	四核
0	0.81/3.24	0.71/2.18	0.66/2.22	0.53/1.85
1	0.81/3.23	0.71/2.18	0.66/2.23	0.53/1.85
5	0.82/3.20	0.71/2.13	0.67/2.24	0.53/1.85
10	0.85/3.23	0.70/2.06	0.67/2.09	0.57/1.84
11	0.93/3.24	0.70/2.05	0.67/2.05	0.57/1.84
12	1.70/6.89	0.70/2.02	0.67/2.09	0.56/1.85
13	1.98/7.70	0.71/2.03	0.71/2.06	0.62/1.85
14	4.57/19.78	0.73/2.09	1.32/11.09	0.61/1.88
15	4.42/19.27	0.75/2.12	1.40/12.41	0.64/1.89

## 4 结束语

本文针对声源 DOA 估计中的 TDOA-DOA 映射问题, 提出了基于聚类方法的多核 LS-SVR 映射方法, 在此基础上对多核 LS-SVR 映射方法进行稀疏化分析。仿真结果表明, 该映射方法的性能要优于现有的单核 LS-SVR 以及最小二乘法。一般情况下, 随着核个数增多, 多核 LS-SVR 的性能也更好。混响时间较大时, 多核 LS-SVR 映射法的性能优势体现得更明显。由于多核 LS-SVR 映射方法运用的 LS-SVR 中, 所有训练集都为支持向量, 缺乏稀疏性, 使得测试时运算量较大。因而对于基于 LS-SVR 的 TDOA-DOA 映射中的多核函数方法, 有必要通过稀疏化消除 TDOA 集合中存在的冗余。仿真结果表明, 与基本的多核 LS-SVR 相比, 稀疏多核 LS-SVR 映射方法不仅能保持良好的 DOA 估计性能, 而且有效减小了测试时的运算量。

## 参考文献:

- [1] Zhang C, Florencio D, Ba D E, et al. Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings[J]. *IEEE Transactions on Multimedia*, 2008, 10(3): 538-548.
- [2] 李晓飞, 刘宏. 机器人听觉声源定位研究综述[J]. *智能系统学报*, 2012, 7(1): 9-20.  
Li Xiaofei, Liu Hong. A survey of sound source localization for robot audition[J]. *CAAI Transactions on Intelligent Systems*, 2012, 7(1): 9-20.
- [3] 栗志意, 张卫强, 何亮, 等. 基于核函数的 IVEC-SVM 说话人识别系统研究[J]. *自动化学报*, 2014, 40(4): 780-784.  
Li Zhiyi, Zhang Weiqiang, He Liang, et al. Speaker recognition with kernel based IVEC-SVM[J]. *ACTA Automatica Sinica*, 2014, 40(4): 780-784.
- [4] Hahn W R, Tretter S A. Optimum processing for delay-vector estimation in passive signal arrays [J]. *IEEE Transactions on Information Theory*, 1973, 19(5): 608-614.

- [5] 谭颖,殷福亮,李细林.改进的 SRP-PHAT 声源定位方法[J].电子与信息学报,2006,28(7):1223-1227.  
Tan Ying, Yin Fuliang, Li Xilin. Sound localization method using modified SRP-PHAT algorithm[J]. Journal of Electronics & Information Technology, 2006, 28(7): 1223-1227.
- [6] Schmidt R O. Multiple emitter location and signal parameter estimation[J]. IEEE Transactions on Antennas & Propagation, 1986,34(3):276-280.
- [7] Roy R, Kailath T. ESPRIT-estimation of signal parameters via rotational invariance techniques[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1990, 37(7): 984-995.
- [8] 何赛娟,陈华伟,尹明婕,等.基于差分麦克风阵列和语音稀疏性的多源方位估计方法[J].数据采集与处理,2015,30(2):372-381.  
He Saijuan, Chen Huawei, Yin Mingjie, et al. Direction finding of multiple sound sources based on sparseness of speech signals and differential microphone array[J]. Journal of Data Acquisition and Processing, 2015,30(2):372-381.
- [9] 郭业才,宋官琨琨,吴礼福,等.基于圆形麦克风阵列的声源定位改进算法[J].数据采集与处理,2015,30(2):344-349.  
Guo Yecai, Song Gong Kunkun, Wu Lifu, et al. Improved algorithm of sound source localization using circular microphone array[J]. Journal of Data Acquisition and Processing, 2015,30(2):344-349.
- [10] Huang Y, Benesty J, Chen J. Acoustic MIMO signal processing [M]. Berlin: Springer, 2006.
- [11] Chen H, Ser W. Sound source DOA estimation and localization in noisy reverberant environments using least-squares support vector machines[J]. Journal of Signal Processing Systems, 2011, 63(3):287-300.
- [12] 张峰,陈华伟,王天南,等.声源 DOA 估计中 LS-SVR 核函数选取研究[J].电声技术,2014,38(5):57-61.  
Zhang Feng, Chen Huawei, Wang Tiannan, et al. Selection research of LS-SVR kernels for sound source DOA estimation [J]. Audio Engineering,2014,38(5):57-61.
- [13] 汪洪桥,孙富春,蔡艳宁,等.多核学习方法[J].自动化学报,2010,36(8):1037-1050.  
Wang Hongqiao, Sun Fuchun, Cai Yanning, et al. On multiple kernel learning methods[J]. ACTA Automatica Sinica, 2010,36(8):1037-1050.
- [14] Knapp C, Carter G. The generalized correlation method for estimation of time delay [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, 24(4): 320-327.
- [15] 梁宇,马良,纳霞等.基于广义互相关算法的时延估计[J].计算机科学,2011,38(B10):454-456.  
Liang Yu, Ma Liang, Na Xia, et al. Research of time delay estimation based on GCC algorithm[J]. Computer Science, 2011, 38(B10): 454-456.
- [16] Hong Y, Kwong S, Wang H. Decision-based median filter using k-nearest noise-free pixels[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Taiwan, China; ICASSP, 2009: 1193-1196.
- [17] 赵犁丰,李新,王栋.多核模糊聚类算法的研究[J].中国海洋大学学报:自然科学版,2009,39(5):1047-1050.  
Zhao Lifeng, Li Xin, Wang Dong. Clustering algorithm based on multiple kernel SVM[J]. Periodical of Ocean University of China,2009,39(5):1047-1050.
- [18] Cortes C, Mohri M, Rostamizadeh A. Learning non-linear combinations of kernels[J]. Advances in Neural Information Processing Systems,2009,22:396-404.
- [19] Wang Z, Jie W, Gao D. A novel multiple Nystrom-approximating kernel discriminant analysis[J]. Neurocomputing, 2013, 119: 385-398.
- [20] 贺玲,吴玲达,蔡益朝.数据挖掘中的聚类算法综述[J].计算机应用研究,2007,24(1):10-13.  
He Ling, Wu Lingda, Cai Yichao. Survey of clustering algorithms in data mining[J]. Application Research of Computers, 2007,24(1):10-13.
- [21] 梁焯炜.K-均值聚类算法的改进及其应用[D].长沙:湖南大学,2012:20-24.  
Liang Yewei. Improvent of K-means clustering algorithm and its application[D]. Changsha: Hunan University, 2012:20-24.
- [22] Suykens J A K, Lukas L, Vandewalle J. Sparse approximation using least squares support vector machines[C]//IEEE International Symposium on Circuits and Systems. Geneva: ISCAS, 2000:757-760.

## 作者简介:



张峰(1989-),男,硕士研究生,研究方向:音频与语音信号处理,E-mail: zhang-feng@nuaa.edu.cn.



陈华伟(1977-),男,教授,研究方向:音频与语音信号处理、阵列信号及统计信号处理。



李妍文(1992-),女,硕士研究生,研究方向:音频与语音信号处理。

