

基于宽度学习系统的声源区域定位方法

张晓萌, 孙昊, 王硕朋, 徐静

(河北工业大学人工智能与数据科学学院, 天津 300130)

摘要: 针对传统的声源定位技术在室内环境中定位精度不足和时效性差等问题, 提出了一种基于宽度学习系统 (Boarding Learning System, BLS) 的声源区域定位方法。方法首先将采集到的数据的特征作为网络的特征节点, 再将所有映射的特征和随机生成权重增强节点直接连接到输出端, 对应的输出系数通过伪逆求得。最后利用训练好的宽度学习系统网络模型对测试点进行预测, 判断测试点所属区域。实验结果表明, 基于宽度学习系统的声源区域定位方法在室内环境中的定位精度和实时性方面都取得了良好的效果。

关键词: 声源定位; 机器学习; 室内环境; 宽度学习系统

中图分类号: TN912 **文献标识码:** B

Regional Localization Method for Sound Source Based on BLS

ZHANG Xiao-meng, SUN Hao, WANG Shuo-peng, XU Jing

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300130, China)

ABSTRACT: To solve the problem of the lack of positioning accuracy and timeliness in traditional unstructured indoor environments, this paper proposed a regional localization method for sound source based on broad learning system (BLS). The method firstly took the characteristics of the collected data as the characteristic nodes of the network. Then, all mapped features and randomly generated weight enhancement nodes were directly connected to the output, and the corresponding output coefficients were obtained through pseudo-inverse. Finally, the trained BLS network model was used to predict the test points, determine the area. The experimental results show that the regional localization method for sound source based on BLS has achieved good results in the positioning accuracy and real-time performance in the indoor environment.

KEYWORDS: Sound source localization; Machine learning; Indoor environments; BLS

1 引言

随着语音处理技术和人工智能技术的发展, 自然语音以其友好的方式在人机交互领域备受关注。运用麦克风阵列进行声源定位研究已经成为信号处理领域的一个研究热点, 在移动机器人、语音识别、视频会议系统、语音增强等方面都有较多的应用^[1-4]。

传统的声源定位方法大体上可分为三类: 基于高分辨率谱估计技术、最大输出功率的可控波束形成技术、基于声达时间差的声源定位技术。文献[5]利用基于高分辨率谱估计的 music 算法, 实现了多目标定位技术, 但该技术仅适用于

远场模型, 反射作用产生的混响会严重干扰使定位的准确性。文献[6]利用一种基于波束形成正则化和数据缩放的逆问题的混合方法来检测声源, 但计算复杂度较高, 耗时较长。文献[7]利用广义互相关 (GCC) 算法进行时延估计来进行声源定位, 但对噪声和混响的抗干扰能力不足。这些传统的方法都存在一个共性: 都是从信号处理的角度来处理声源定位问题。而在实际情况下, 感兴趣的声源位置仅限于某些预定义的区域, 因此对于这些典型的场景, 从机器学习的角度来处理声源定位问题更为合理^[8]。文献[9]将 BP 神经网络应用于传感器阵列, 使其在一定程度上提高了在室内环境定位的准确性, 但还存在模型的训练时间长、定位精度不足、拟合能力不足等问题。文献[10]提出利用 RBF 核支持向量机来构建加权的最小方差无失真响应波束形成器, 能有效处理近场的单声源定位问题, 但计算量较大仍然是存在的问题。文献[11]提出了一种基于改进的支持向量回归的室内定位方法, 提出的算法可以减少室内复杂环境带来的噪声干扰, 与

基金项目: 国家自然科学基金 (61773151, 61703135, 61503118); 河北省自然科学基金 (F2017202119, F2016202327)

收稿日期: 2018-12-15 修回日期: 2019-01-31

传统的支持向量回归定位算法相比有更高的定位精度。文献[12]利用概率神经网络(PNN)算法进行声源定位,实现了更准确的时延估计并解决了声源定位在高混响和低信噪比下定位精度不足的缺点。

基于此,本文利用机器学习的算法通过分类的手段进行声源定位,针对在某些实际情况下感兴趣的声源位置仅局限于某些预定义的区域^[13],以及已有的定位技术在室内环境中定位精度不足和实时性差等问题,提出了一种基于宽度学习系统(BLS)的声源区域定位方法。宽度学习系统(BLS)的实质是随机向量函数链接神经网络(random vector functional link neural network, RVFLNN),是一种无需深度结构的高效增量学习系统^[14],其算法复杂度低,训练速度快提高了定位系统的实时性,最后通过实验证明了该方法的有效性。

2 宽度学习系统

2.1 宽度学习系统模型

宽度学习系统是基于传统的随机向量函数链接神经网络(random vector functional link neural network, RVFLNN)^[15]构建的,其与直接采用输入并建立增强节点的RVFLNN不同,宽度学习系统以映射作为输入并构建映射特征,BLS不采用深度的结构而基于单隐层神经网络构建,在精度不够准确时,深度神经网络会采用增加层数或调整参数个数的方式,而BLS则采用横向扩展的方式,利用输入映射的特征作为网络的特征节点,再增强为随机生成权重的增强节点,并将映射特征与增强节点直接连接到输出端,对应的输出系数可以通过伪逆^[16]计算得出,BLS无需从头开始学习,只需调节与新增节点有关的权重。

假设给定输入数据 X ,并利用函数 $\phi_i(XW_{e_i} + \beta_{e_i})$ 映射产生第 i 组映射特征 Z_i 。其中 W_{e_i} 是具有适当维度的随机权重系数。记 $Z^i = [Z_1, \dots, Z_i]$ 表示前 i 组所有映射特征。类似的,第 j 组增强节点 $\xi_j(Z^i W_{h_j} + \beta_{h_j})$ 被记为 H_j ,同时前 j 组所有的增强节点被记为 $H^j = [H_1, \dots, H_j]$,实际上,根据建模任务的复杂度可以选择不同的 i 和 j 。当 $i \neq k$ 时 ϕ_i 和 ϕ_k 可以是不同的函数,同样,当 $j \neq r$ 时 ξ_j 和 ξ_r 也可以是不同的函数。为了充分利用稀疏自编码器特性的优点,利用线性逆问题并对初始 W_{e_i} 进行微调以获得更好的特征。具体细节如下。

假设输入数据集 X 有 N 个样本,每个样本都有 M 个维度, Y 为 $R^{N \times C}$ 的输出矩阵。对于 n 个特征映射,每个映射生成 k 个节点,那么第 i 组映射特征 Z_i 为

$$Z_i = \phi_i(XW_{e_i} + \beta_{e_i}) \quad i = 1, \dots, n \quad (1)$$

其中 W_{e_i} 和 β_{e_i} 是随机产生的,将所有特征节点表示为 $Z^n = [Z_1, \dots, Z_n]$,并将第 m 组增强节点表示为

$$H_m = \xi(Z^n W_{h_m} + \beta_{h_m}) \quad (2)$$

因此,宽度学习的模型可以表示为

$$Y = [Z_1, \dots, Z_n | \xi(Z^n W_{h_1} + \beta_{h_1}), \dots, \xi(Z^n W_{h_m} + \beta_{h_m})]$$

$$= [Z_1, \dots, Z_n | H_1, \dots, H_m] W^m \\ = [Z_n | W^m] W^m \quad (3)$$

其中 $W^m = [Z^n | H^m]^+ Y$, W^m 是宽度结构的连接权重。图1为上述宽度学习系统结构图。

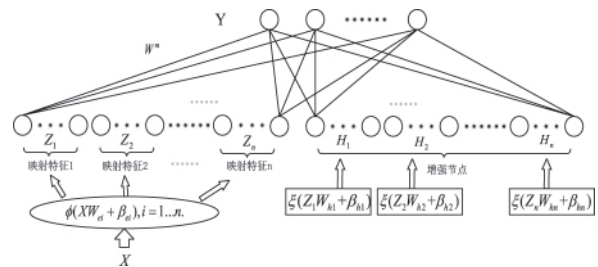


图1 宽度学习系统结构图

2.2 增量学习算法

增量学习算法是宽度学习算法的扩展方法,在流行的深层网络结构中,当现有模型无法满足训练精度时一般需要增加网络的深度或者隐含层节点数^[17],需要重新设置新结构的参数,相反在宽度学习系统中,可以利用增量学习算法高效的重新构建宽度学习系统模型,仅仅需要更新权重而无需重新训练整个网络结构,附加 $(n+1)$ 个特征映射及 p 个增加节点的增量网络结构如图2所示。假设初始结构由 n 组特征映射节点和 m 组增强节点组成,那么第 $(n+1)$ 个特征映射节点可以表示为

$$Z_{n+1} = \phi(XW_{e_{n+1}} + \beta_{e_{n+1}}) \quad (4)$$

相应的增强节点随机生成如下

$$H_{ex_m} = [\xi(Z_{n+1}W_{ex_1} + \beta_{ex_1}), \dots, \xi(Z_{n+1}W_{ex_m} + \beta_{ex_m})] \quad (5)$$

其中 W_{ex_j} 和 β_{ex_j} 是随机生成的,因此,该算法仅需计算附加映射特征的伪逆从而完成快速增量学习。

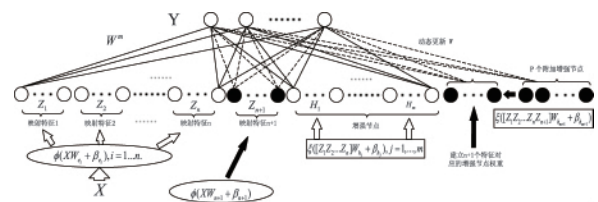


图2 宽度学习系统的增量网络结构

3 系统分析

3.1 建立信号模型

在非结构化的室内环境中,麦克风接收到的信号中除了存在各个麦克风之间由于与声源的距离不同造成的接收到的信号的相位差异外还存在声波在空气中传播造成的幅度衰减。若二维空间内有单一固定声源 $s(t)$,对于由 M 个麦克风组成的阵列,则第 i 个麦克风接收到的信号可以写成:

$$x_i(t) = a_i s(t - \tau_i) + n_i(t), \quad i = 1, 2, \dots, M \quad (6)$$

式中 i 表示第 i 路麦克风, a_i 和 τ_i 分别表示接收到声源信号

的幅度衰减因子以及相对时延 $n_i(t)$ 则是各种噪声的总和。假定声音信号与各麦克风接收到的噪声互不相关,各麦克风的噪声也不相关。

本系统采用 4 个传感器,声源位于四路麦克风阵列之中,二维空间下的系统模型如图 3 所示。

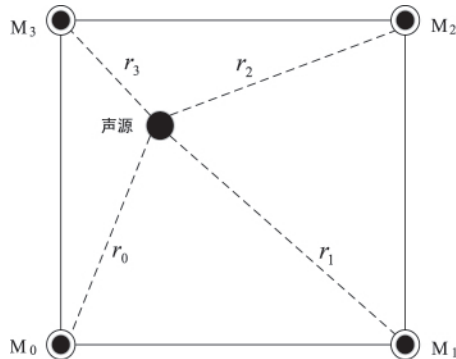


图 3 系统模型

3.2 构建定位数据集

基于 BLS 的声源区域定位方法需要构建定位数据库来获得训练样本集和测试样本集,在样本选取阶段,首先在待定位区域分成若干个子区域,然后按照一定的间隔布置定位参考点和测试点,麦克风 (M_0, M_1, M_2, M_3) 对声源在各个参考点所发出的声音信息进行采集,然后经声音放大模块及数据采集模块完成信号处理过程。每个参考点的位置坐标为 $L_i = (x_i, y_i)$,以每个参考点处的声达时间差为特征构建定位数据,用 T 表示定位数据库,则其可表示为

$$T = \begin{bmatrix} \Delta t_{11} & \Delta t_{12} & \Delta t_{13} \\ \vdots & \vdots & \vdots \\ \Delta t_{i1} & \Delta t_{i2} & \Delta t_{i3} \\ \vdots & \vdots & \vdots \\ \Delta t_{n1} & \Delta t_{n2} & \Delta t_{n3} \end{bmatrix} \quad (7)$$

式中: $\Delta t_{i1}, \Delta t_{i2}, \Delta t_{i3}$ 分别表示在第 i 个参考点处的基准麦克风相对于其它三个麦克风所接收信号的声达时间差; n 表示参考点的个数 $i=1, 2, 3, \dots, n$ 。本文利用短时平均能量和短时过零率的端点检测方法计算声源信号端点^[18],麦克风 (M_0, M_1, M_2, M_3) 接收信号的端点为 r_0, r_1, r_2, r_3 ,可算出第 i 个参考点处的声达时间差:

$$\Delta t_{ij} = \frac{\Delta r_j}{f} \times T, j=1, 2, 3, \quad (8)$$

其中, $\Delta r_j = r_j - r_0$; $f=100\text{kHz}$ 是声音采集卡的采样频率; $T=1\text{s}$ 为采样周期。

3.3 BLS 算法实现

基于 BLS 的声源区域定位方法总体框架如图 4 所示,包括训练过程及分类过程,其具体算法实现步骤如下:

3.3.1 BLS 算法实现

1) 首先将待定位区域分成若干个子区域,然后在待定位

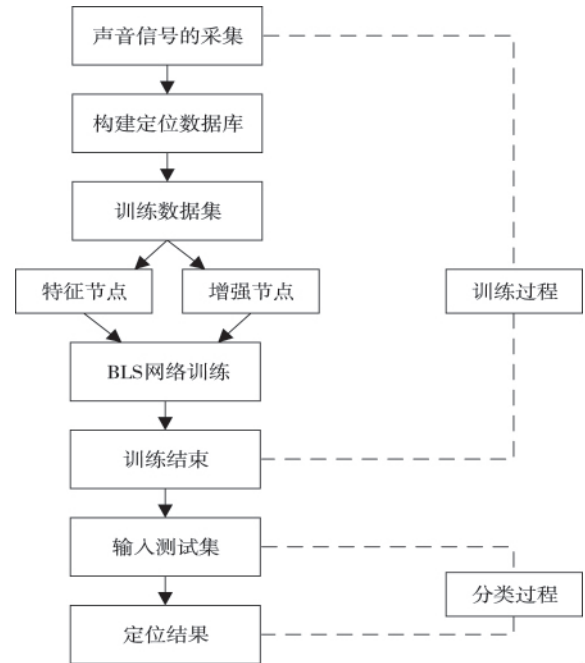


图 4 基于 BLS 的声源区域定位总体框架

的区域均匀布置参考点随机布置测试点以此构建定位数据库,选取参考点的声达时间差信息作为训练数据集,每个参考点所在的区域对应训练数据集的标签,随机选取的测试点的声达时间差信息作为测试数据集,每个测试点所在的区域对应测试数据集的标签。记 N_1 : 每个窗口特征节点的个数; N_2 : 特征节点的窗口个数; N_3 : 增强节点的个数; $H_{1(s \times f)}$: 训练集 s 表示样本个数 f 表示特征数。

2) 利用 BLS 网络的随机向量函数链接特性,对每个样本的特征进行一次权值随机的卷积和偏置得到新的特征,对新的特征进行稀疏处理及归一化。对于 N_2 个特征窗口,均生成 N_2 个特征节点,每个节点是 s 维特征向量。对特征节点矩阵进行标准化与增广,得到增强节点,增加网络的非线性因素。

3) 求特征节点和增强节点到目标值的伪逆^[19],完成整个 BLS 网络的输入和权重的训练。

3.3.2 分类过程

将测试集样本输入到已经训练好的网络模型中,最终得到待定位区域中的每一个测试点所属的定位区域,求得定位准确率。

4 实验

4.1 系统组成

基于 BLS 的声源区域定位方法由麦克风阵列,由数据采集及放大模块和计算机构成。麦克风阵列由体积小且灵敏度高的 MPA201 全指向性麦克风组成,数据采集及放大模块由数据采集卡 NI9215A 和集成运算放大器 LM324 对麦克风阵列所接收的信号进行采集和放大。

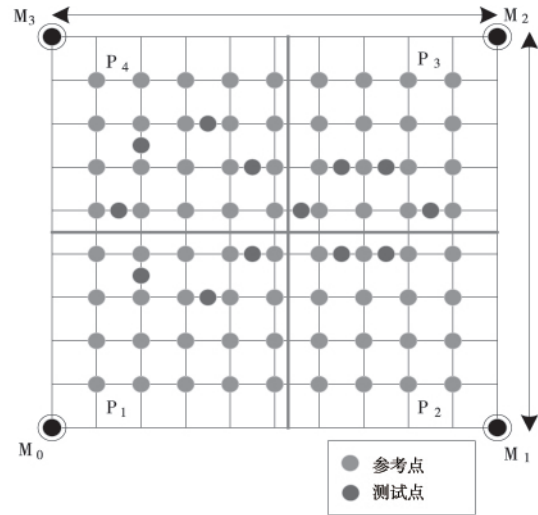
4.2 实验环境的建立

为验证本文提出的基于 BLS 的声源区域定位方法的有效性,在河北工业大学武清实验室进行了实验,实验室墙壁无隔音处理且噪音约为 40dB,实验场景如图 5 所示。图 5(a)为实验环境,图 4(b)为局部平面图。整个定位区域为长方形,长 5.93m、宽 5.43m,面积约为 32.2m²。在定位区域内的



(a) 实验环境

四个顶点布置高为 1.36m 的麦克风(M_0, M_1, M_2, M_3)。然后将定位区域分成 4 个子区域(P_1, P_2, P_3, P_4)。总共选取 72 个点作为定位参考点,相邻两个定位参考点间隔 0.593m。随机选取分界线附近的 13 个点为测试点。利用一小型蓝牙音箱作为声源,调整其高度与麦克风高度大致相同,在各个参考点及测试点发声供麦克风(M_0, M_1, M_2, M_3)接收。



(b) 局部平面图

图 5 实验场景

4.3 实验结果

实验在 Windows7 下 MATLAB R2016a 环境下进行测试,首先设置单一结构的 BLS 网络模型,初始值为 10 × 10 个特征节点和 500 个增强节点进行测试,测试结果如表 1 所示。然后将初始的增强节点改为 300,令其每次动态增加 50 个增强节点直到达到 500 训练结束,测试结果如表 2 所示。其次,本实验使用三个动态增量来动态增加:1) 特征节点;2) 相应的增强节点;3) 附加增强节点。设置网络初始值为 10 × 6

个特征节点和 300 个增强节点。特征节点每 10 步动态更新,从 60 增加到 100。相应的增强节点每次增加 15,并且附加增强节点每次增加 35。表 3 列出了在不同动态结构中使用增量学习的定位结果。

表 1 单一结构的宽度学习系统模型定位结果

特征节点数量	增强节点数量	测试准确率	训练时间(s)	测试时间(s)
100	500	92.31%	0.086	0.045

表 2 逐步增加 50 个增强节点的定位结果

特征节点数量	增强节点数量	测试准确率	每更新一次的训练时间(s)	累计训练时间(s)	每更新一次的测试时间(s)	累计测试时间(s)
100	300	92.31%	0.107	0.107	0.055	0.055
100	300→350(50)	92.31%	0.141	0.248	0.011	0.066
100	350→400(50)	92.31%	0.023	0.271	0.001	0.067
100	400→450(50)	92.31%	0.014	0.285	0.002	0.068
100	450→500(50)	92.31%	0.015	0.300	0.001	0.069

表 3 不同动态结构中的定位结果

特征节点数量	增强节点数量	测试准确率	每更新一次的训练时间(s)	累计训练时间(s)	每更新一次的测试时间(s)	累计测试时间(s)
60	300	92.31%	0.060	0.060	0.032	0.032
60→70	300→350(15 + 35)	92.31%	0.040	0.100	0.014	0.046
70→80	350→400(15 + 35)	92.31%	0.029	0.129	0.011	0.025
80→90	400→450(15 + 35)	92.31%	0.020	0.149	0.017	0.042
90→100	450→500(15 + 35)	92.31%	0.017	0.166	0.008	0.050

从实验结果可以看出,基于 BLS 的声源区域定位方法可以将随机选取的分界线附近的 13 个测试点准确无误的分

— 300 —

类,精准的确定声源所属的区域,而且训练速度极快。同样,在使用增量学习算法对模型进行动态更新可以提供一兼

容的结果,训练时间略大于单一结构的宽度学习系统模型,但训练精度相同。

4.4 对比实验

为了验证宽度学习算法在声源定位中的性能,本文将其与 k 近邻算法(KNN)、BP(Back Propagation)神经网络和支持向量机(SVM)进行比较,为了保证实验的一致性,选取相同的实验数据,同样在 Windows7 下 MATLAB R2016a 环境下进行测试,每组分别测试五次,求得平均测试结果,测试结果如表 4 所示:

表 4 不同模型下定位准确率

	模型			
	SVM	BP	KNN	BLS
平均测试准确率(%)	51.85	84.31	92.40	92.89
时间(s)	0.02	0.65	0.7	0.10

实验结果表明,相比于 k 近邻算法(KNN)、BP(Back Propagation)神经网络和支持向量机(SVM)三种模型,基于 BLS 的声源区域定位方法不仅准确率高而且耗时短,验证了所提出方法的有效性和可行性。

5 结论

针对在某些实际情况下感兴趣的声源位置仅局限于某些预定义的区域,以及已有的定位技术在室内环境中定位精度不足和实时性差等问题,提出了一种基于宽度学习系统(BLS)的声源区域定位方法,并通过实际实验验证了该系统的有效性。而且宽度学习系统以及它的扩展结构在实验中都表现出了快速且高精度的性能,通过逐步更新建模系统,无需从一开始重新训练整个系统。尤其当系统收集到新输入数据时,现有网络结构迫切需要更新以反映系统的真实性和完整性。在人工智能不断发展的今天,这一功能完全适用于大数据时代,具有一定的实用性。

参考文献:

[1] 孙昊,仲维灿,刘贺洋. 正四面体麦克风阵列声源定位模型研究[J]. 计算机仿真, 2015, 32(2): 378-382.

[2] D A Hambrook, M Ilievski, M Mosadeghzad, et al. A Bayesian computational basis for auditory selective attention using head rotation and the interaural time-difference cue[J]. Plos One, 2017, 12(10): e0186104.

[3] P Aarabi, S Zaky. Robust sound localization using multisource audio-visual information fusion[J]. Inform Fusion, 2006, 5(2): 209-223.

[4] 程宁,刘文学. 基于听觉感知特性的信号子空间麦克风阵列语音增强算法[J]. 模式识别国家重点实验室, 2009, 35(12): 1481-1487.

[5] L Kumar, A Tripathy, R M Hegde. Robust Multi-Source Localization Over Planar Arrays Using MUSIC-Group Delay Spectrum[J]. IEEE Transactions on Signal Processing, 2014, 62(17): 4627-4636.

[6] T Padois, P A Gauthier, A Berry. Inverse problem with beamform-

ing regularization matrix applied to sound source localization in closed wind-tunnel using microphone array[J]. Journal of Sound & Vibration, 2014, 333(25): 6858-6868.

[7] 赵圣,崔牧凡,尤磊,等. 基于小型麦克风阵列的声源定位技术[J]. 华中科技大学学报(自然科学版), 2013, 41(s1): 188-191.

[8] Li Y, Chen H. Reverberation Robust Feature Extraction for Sound Source Localization Using a Small-Sized Microphone Array[J]. IEEE Sensors Journal, 2017, 19(9): 1-1.

[9] Fu T, Zhang Z, Liu Y, et al. Development of an artificial neural network for source localization using a fiber optic acoustic emission sensor array[J]. Structural Health Monitoring, 2015, 14(2): 168-177.

[10] D Salvati, C Drioli, G L Foresti. A weighted MVDR beamformer based on SVM learning for sound source localization[J]. Pattern Recognition Letters, 2016, 84: 15-21.

[11] 姚英彪,毛伟勇,姚瑞丽,等. 基于改进支持向量回归的室内定位算法[J]. 仪器仪表学报, 2017, 38(9): 2112-2119.

[12] Sun Y, Chen J, Yuen C, et al. Indoor Sound Source Localization with Probabilistic Neural Network[J]. IEEE Transactions on Industrial Electronics, 2017, 64(9): 1-1.

[13] 崔林林,宋萍,石庚辰. 基于区域划分的声定位算法及精度分析[J]. 吉林大学学报(工), 2017, 47(1): 288-293.

[14] Chen C, Liu Z. Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture[J]. IEEE Transactions on Neural Networks & Learning Systems, 2017, 29(1): 10-24.

[15] C L P Chen and J Z Wan. A rapid learning and dynamic stepwise updating method for flat neural networks and the application to timeseries prediction[J]. IEEE Trans. Syst., Man, Cybern. B, Cybern, Feb. 1999, 29(1): 62-72.

[16] C Leonides. Control and dynamic systems V18[M]. in Advances in Theory and Applications (Control and dynamic systems). Amsterdam, The Netherlands: Elsevier, 2012.

[17] 段艳杰,吕宜生,张杰,等. 深度学习在控制领域的研究现状与展望[J]. 自动化学报, 2016, 42(5): 643-654.

[18] 王硕朋,杨鹏,孙昊. 基于声音位置指纹的室内声源定位方法[J]. 北京工业大学学报, 2017, 43(2): 224-229.

[19] Arthur E Hoerl, Robert W Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems[J]. Technometrics, 2000, 42(1): 55-66.



[作者简介]

张晓萌(1994-)女(满族),河北省保定市人,硕士研究生,主要研究领域为机器学习、声源定位。

孙昊(1979-)男(汉族),吉林省吉林市人,副教授,硕士研究生导师,主要研究领域为移动机器人、机器学习。

王硕朋(1987-)男(汉族),河北省邢台市人,博士生,主要研究方向为声源定位。

徐静(1992-)女(汉族),山西省大同市人,硕士研究生,主要研究领域为声源定位。