第 40 卷 第 1 期
2021 年 1 月

应用声学
Journal of Applied Acoustics

Vol. 40, No. 1
January, 2021

◇ 研究报告 ◇

基于多域特征提取和深度学习的声源被动测距*

肖 旭[1,2,3] 王 同[1,2,3] 王文博[1,2,3] 苏 林[1,2] 马 力[1,2,3] 任群言[1,2,3†]

(1 中国科学院水声环境特性重点实验室 北京 100190)

(2 中国科学院声学研究所 北京 100190)

(3 中国科学院大学 北京 100049)

摘要: 采用一种基于多域特征提取的深度学习方法来实现声源被动测距。首先从声信号中提取多域特征,包含时域波形结构特征、时域包络特征、频域谱特征和基于短时傅里叶变换的时频联合域特征;然后基于不同谱表达计算出一组声学参数构成特征空间,在此基础上采用最大相关-最小冗余准则选出特征空间中与声源位置相关性高的关键特征作为模型输入;最后通过一种改进的深度神经网络实现声源距离的估计,引入自适应矩估计优化算法进行模型训练,利用L2和Dropout正则化策略实现网络参数稀疏化。通过声速正梯度浅海环境仿真实例对方法进行验证,对比分析了波形参数对测距性能和模型收敛速度的影响。结果表明,此方法在模型训练过程中收敛速度较快,预测性能较稳定,在所定条件下测试集上声源信号的综合测距精确率达到95%以上。

关键词: 多域特征提取;深度学习;声源被动测距

中图法分类号: TB566 **文献标识码:** A **文章编号:** 1000-310X(2021)01-0121-10

DOI: 10.11684/j.issn.1000-310X.2021.01.015

Source ranging based on deep learning and multi-domain feature extraction: synthetic results

XIAO Xu[1,2,3] WANG Tong[1,2,3] WANG Wenbo[1,2,3] SU Lin[1,2] MA Li[1,2,3] REN Qunyan[1,2,3]

(1 *Key Laboratory of Underwater Acoustics Environment, Chinese Academy of Sciences, Beijing* 100190, *China*)

(2 *Institute of Acoustics, Chinese Academy of Sciences, Beijing* 100190, *China*)

(3 *University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: Ranging of underwater acoustic sources is an important task in many practical applications. Deep neural networks (DNNs) have shown outstanding performance on illustrating complex nonlinear relationships. In terms of localization, DNNs were demonstrated to have low range estimation error under low SNR conditions. In this study, source localization is achieved by a deep learning method based on a set of multi-domain features extracted from sound signals. In this study, a comprehensive set of acoustic parameters are extracted from sound signals. The waveform, energy envelope, short-term Fourier transform are first estimated from the sound signals, based on which, the acoustic parameters are computed and used as the training data set for the estimation of source in a DNN. Training the deep architecture is achieved by Adam optimizer and dropout for parameter regularization with mean squared error (MSE) loss functions.

Keywords: Multi-domain extraction; Deep learning; Source ranging

2020-04-21 收稿; 2020-07-15 定稿

*国家自然科学基金项目 (11704396)

作者简介: 肖旭 (1996–), 男, 湖南长沙人, 博士研究生, 研究方向: 水声物理。

†通信作者 E-mail: renqunyan@mail.ioa.ac.cn

0 引言

声源被动测距作为声呐系统的重要功能之一，一直是水声工作者密切关注的问题[1]。由于水声观测信号受复杂的时、空、频变及强多途、高噪声和多普勒效应等因素影响，传统的匹配场处理方法往往面临计算量大和环境失配等问题。近年来，深度学习作为基于数据驱动方式的新兴分支，以其强大的特征提取能力和高效处理复杂、高维、非线性系统问题的独特优势[2]，为水声被动测距提供了一种新思路。

深度神经网络通过大量数据样本建立高维参数之间的复杂非线性映射，适用于物理建模困难的问题，引发了水声研究者的关注。Lefort 等[3] 通过水箱实验研究了机器学习算法在水声目标测距中的测距性能。Niu 等[4] 利用单水听器和残差卷积神经网络对声源进行定位。Wang 等[5] 提出了一种用于水下声源测距的深度迁移学习方法，将从仿真环境获得的预测能力迁移到实验海域。

水声接收信号包含声源和声信道的大量信息，其特征提取和构造是深度学习方法的关键环节。早期特征提取通常利用信号的自相关函数和功率谱估计，或采用时-频分析方法来提取一些时频联合域特征，如短时傅里叶变换 (Short-time Fourier transform, STFT)、Wingner-Ville 分布等。然而，无论是功率谱分析还是时频分析[6]，它们包含大量与声源位置不相关的信息或冗余信息，在形式上维数较大，一般无法直接应用于测距任务。而且，单一地采用某类特征通常会丢失掉部分特征，缺乏全面性。

针对以上问题，本文设计了一种基于多域特征提取和深度学习的方法来实现声源被动测距。首先从声信号中提取多域特征，包含时域波形结构特征、时域包络特征、频域谱特征和基于 STFT 的时频联合域特征。然后基于不同谱表达计算出一组声学参数构成特征空间，在此基础上采用最大相关-最小冗余准则 (Maximum relevance and minimum redundancy, mRMR) 选择特征空间中重要度高的关键特征 (与声源位置相关性大) 作为模型输入，最后通过一种改进的深度神经网络实现声源距离的估计。神经网络训练时采用自适应矩估计 (Adaptive moment estimation, Adam) 优化算法和均方误差 (Mean squared error, MSE) 代价函数进行更新模型参数，用 L2 和 Dropout 正则化策略实现网络参数正则化。通过浅海环境仿真实例验证了该方法的有效性，对比分析了波形参数对测距性能和模型收敛速度的影响。

1 理论模型

1.1 水声信号多域特征提取

声信号的多域特征可归纳为 6 类：时域波形结构特征、时域包络特征、频域谱特征、基于 STFT 的时频联合域特征、基于等效矩形带宽的听觉谱特征和基于正弦谐波模型的谐波谱特征[7]。每一类特征均包含了对应谱的多种声音特性，这些特征属性无法用单一尺度进行描述，只有在多维特征空间下才能表示。Peeters 等[8] 对这些声学特征进行总结，并通过多维标度法提取了合适的声学参数，使之与声信号在每个维度上的坐标呈现较大的相关性。

这些声学参数是基于不同谱表达计算的，其表征的物理含义各有不同，而水下声源物理特征和声场信息主要包含在信号的时域波形、频域能量分布中。因此，本文综合了以上更符合水声信号特点的前 4 类特征，并将所对应的特征归纳为时域特征和时频联合域特征。

1.1.1 时域特征

水声信号时域波形反映了信道对声信号传播的畸变作用，是获取声源总体特征最直接的来源。为了直接从时域提取特征，自相关系数是一种广泛使用的分类特征。首先是对原始信号 $s(t_n)$ 求自相关系数，t_n 代表信号时刻，保留前 N 维的自相关系数 ($c \in \{1, \cdots, N\}$) 表示为[9]

$$\rho(c) = \frac{1}{\rho(0)} \sum_{n=0}^{L_n-c-1} s(t_n)s(t_{n+c}), \tag{1}$$

其中，L_n 是分帧时的窗长，c 代表时间的滞后量。当声源信号为瞬态声信号时，其随时间变化经历 ADSR 过程，即激励阶段、衰减阶段、稳态持续阶段、释音残响阶段，如图 1 所示，其中激励阶段到衰减阶段以信号振幅峰值处为分界点，后三阶段统称为下降阶段。

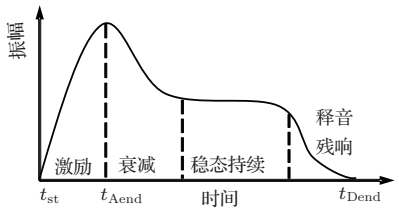


图 1 声信号 ADSR 过程

Fig. 1 ASDR process of sound signals

声源时域信号的幅度强弱变化和接收距离的相关性较强，为提取描述声源信号幅度变化的特征，对原始信号 $s(t_n)$ 进行 Hilbert 变换，然后使用截止频率为 5 Hz 的三阶 Butterworth 滤波器对振幅信号进行低通滤波，得到信号振幅包络 $e(t_n)$。估计信号的起始时间 $t_{\rm st}$、激励阶段和下降阶段的终止时间 $t_{\rm Aend}$、$t_{\rm Dend}$，在此基础上定义时域的 7 个声学参数，如表 1 所示。其中，为了估计激励过程的时间长度，定义对数激励时间为 [10]

$$\mathrm{LAT}=\lg(t_{\rm Aend}-t_{\rm st}). \tag{2}$$

相对应地，将激励阶段和下降阶段能量的平均时间斜率分别定义为激励斜率和下降斜率。估计振幅包络的最大值 $e_{\max}$。时间质心给出了信号能量质心所在的时刻，其定义式为 [11]

$$t_c=\frac{\sum\limits_{n=n_1}^{n_2}t_n e(t_n)}{\sum\limits_{n}e(t_n)}, \tag{3}$$

其中，n_1 和 n_2 为声信号的起始和终止时间对应的索引号，以滤除前后的空白时间。当信号声能量大于一定阈值 $\gamma e_{\max}$ 时，其持续时间定义为信号的有效时间，根据许多经验性的测试，γ 取 0.4 性能较稳定，见文献 [8]。除以上参数外，其余参数的公式定义可在文献 [8] 中一一得到。这些声学参数反映了声源的各种物理属性，例如对数激励时间反映了能量在上升过程中的时间长度，其与声源距离呈正相关，声源距离越远，声信号需达到稳态振动的时间越长。与波形包络相关的特征 (激励时间、时域质心等) 仅适用于瞬态信号，分析连续声信号时，应提取自相关系数、载波信号调制和后文定义的频域特征。

表 1　时域特征

Table 1　Temporal features

声学参数	符号	定义
自相关系数	AC	时域信号自相关系数
激励时间	ATT	激励过程的时间长度
衰减时间	REL	衰减过程的时间长度
对数激励时间	LAT	激励过程的时间长度的对数
激励斜率	ATK	激励阶段能量的平均时间斜率
下降斜率	DS	衰减阶段能量的平均时间斜率
时域质心	TC	信号能量包络的质心所在的时刻
有效持续时间	ED	信号能量超过设定阈值的时间
频率调制	FM	载波信号的频率
振幅调制	AM	载波信号的幅度

1.1.2　时频联合域特征

对原始信号进行 STFT，把每一帧进行快速傅里叶变换后的频域信号在时间上堆叠起来得到时频谱 $a_k(t_m)$，其中 m 代表帧数，k 代表频点数。对 $a_k(t_m)$ 进行标准化处理：

$$p_k(t_m)=\frac{a_k(t_m)}{\sum\limits_k a_k(t_m)}. \tag{4}$$

令第 k 个频点的频率为 f_k，得到谱的前四阶统计矩，分别定义为谱质心、谱延展、谱斜度和谱峰度 [12]，对于第 m 帧的频谱，表示为

$$\begin{cases}\mathrm{SC}(t_m)=\sum\limits_k f_k p_k(t_m),\\ \mathrm{SSP}(t_m)=\left(\sum\limits_k (f_k-\mu_1(t_m))^2\cdot p_k(t_m)\right)^{1/2},\\ \mathrm{SSK}(t_m)=\left(\sum\limits_k (f_k-\mu_1(t_m))^3\cdot p_k(t_m)\right)\Big/\mu_2^3,\\ \mathrm{SK}(t_m)=\left(\sum\limits_k (f_k-\mu_1(t_m))^4\cdot p_k(t_m)\right)\Big/\mu_2^4.\end{cases} \tag{5}$$

除了频谱的统计特征，根据频谱的斜率特征可以从 STFT 能量谱中提取谱斜率、谱衰减 [8]、谱滚降 [13]、谱通量 [14]。最后根据线谱特征提取信号的平坦度及谱峰度 [8]，其定义和物理含义如表 2 所示。对于第 m 帧，上述物理量计算公式分别为

$$\begin{cases}\mathrm{SSL}(t_m)=\dfrac{K\sum\limits_k f_k a_k(t_m)-\sum\limits_k f_k\cdot\sum\limits_k a_k(t_m)}{\sum\limits_k a_k(t_m)\left(K\sum\limits_k f_k^2-\left(\sum\limits_k f_k^2\right)^2\right)},\\ \mathrm{SD}(t_m)=\dfrac{1}{\sum\limits_k a_k(t_m)}\sum\limits_{k=2}^{K}\dfrac{a_k(t_m)-a_1(t_m)}{k-1},\\ \sum\limits_{f=0}^{\mathrm{SR}(t_m)}a_f^2(t_m)=0.95\sum\limits_{f=0}^{f_{\max}}a_f^2(t_m),\\ \mathrm{SV}(t_m)=\Delta f\cdot\sum\limits_k p_k(t_m),\\ \mathrm{SF}(t_m)=\dfrac{K\left(\prod\limits_k a_k(t_m)\right)^{1/K}}{\sum\limits_k a_k(t_m)},\\ \mathrm{SCR}(t_m)=\dfrac{K\max\limits_k^a k(t_m)}{\sum\limits_k a_k(t_m)},\end{cases} \tag{6}$$

其中，Δf 为信号 STFT 后两点的频率差，$f_{\max}$ 为奈奎斯特采样决定的最高频率，$\max\limits_k^a k(t_m)$ 是令

$a_k(t_m)$最大的$k(t_m)$的值。以上参数被证实对声信号的特征识别具有重要作用[15]，被试信号对于它们的变化具有较高的敏感度。

表 2　时频联合域特征

Table 2　Temporal-frequency features

声学参数	符号	定义
谱质心	SC	频谱一阶统计特性，反映声音明亮度
谱延展	SSP	频谱二阶统计特性，反映频谱围绕均值的变化程度
谱偏度	SSK	频谱三阶统计特性，反映频谱在其均值附近分布的不对称程度
谱峰度	SK	频谱四阶统计特性，反映频谱在平均频率附近的平坦程度
谱斜率	SSL	由线性回归方法描述的频谱幅度的下降斜率
谱衰减	SD	频谱下降时一组斜率的平均值
谱滚降	SR	下降至频谱总能量 95% 时对应的截止频率
谱通量	SV	描述声信号频谱包络面积的物理量
谱平坦度	SF	频谱几何平均值和算术平均值的比值
谱波峰	SCR	频谱的最大值和算术平均值的比值

1.2　基于 mRMR 准则的特征选择

以上声学参数作为水声信号的输入特征并不具有鲁棒性，不同任务 (如测距、识别) 的训练集拥有不同的最佳声学参数。在 1.1 节中给出的特征中，有些特征可能是冗余的甚至是不相关的，导致机器学习算法的效率降低、性能损失。

最大相关-最小冗余准则 (mRMR) 是一种综合考虑特征相关度和冗余度的特征重要性评价准则[16]。定义互信息 $I(A,B)$：

$$I(A,B)=\int p(A,B)\lg\left(\frac{p(A,B)}{p(A)p(B)}\right)\mathrm{d}A\mathrm{d}B, \quad (7)$$

其中，变量 A 和 B 的概率密度分别是 $p(A)$ 和 $p(B)$，其联合概率密度是 $p(A,B)$。设样本数量为 m，特征向量数量为 n，特征向量 $\boldsymbol{f}_i=[f_{(i,1)},f_{(i,2)},\cdots,f_{(i,m)}]^{\mathrm{T}}$，$I(\boldsymbol{f}_i,\boldsymbol{f}_j)$ 为样本中第 i 个和第 j 个特征的相关性，其中 $i,\ j=1,2,3,\cdots,n$。设 O_m 为类别标签，$I(\boldsymbol{f}_i,\boldsymbol{O})$ 为特征与输出类别 $\boldsymbol{O}$ 的相关性，其中向量 $\boldsymbol{O}=[O_1,O_2,O_3,\cdots,O_m]^{\mathrm{T}}$。利用最大相关标准式选择出与类别 $\boldsymbol{O}$ 相关性大的特征集合 $\mathbf{D}$：

$$\max\mathbf{D},\mathbf{D}(\mathbf{S},\boldsymbol{O})=\frac{1}{|\mathbf{S}|}\sum_{\boldsymbol{f}_i\in\mathbf{S}}I(\boldsymbol{f}_i,\boldsymbol{O}), \quad (8)$$

式 (8) 中，$|\mathbf{S}|$ 为集合 $\mathbf{S}$ 中所选特征的数量。利用最小冗余标准式剔除特征子集 $\mathbf{S}$ 中的冗余特征的集合 $\mathbf{R}$：

$$\min\mathbf{R},\mathbf{R}(\mathbf{S})=\frac{1}{|\mathbf{S}|^2}\sum_{\boldsymbol{f_i},\boldsymbol{f_j}\in\mathbf{S}}I\left(\boldsymbol{f}_i,\boldsymbol{f}_j\right). \quad (9)$$

综合以上条件，mRMR 方法计算式为

$$\max\boldsymbol{\theta}(\mathbf{D},\mathbf{R})=\mathbf{D}-\mathbf{R}. \quad (10)$$

给定具有 $N-1$ 个特征的集合 $\mathbf{S}_{N-1}$，总特征集合为 $\mathbf{F}$，计算集合 $\{\mathbf{F}-\mathbf{S}_{N-1}\}$ 中选择第 N 个特征使得式 (10) 中的集合 $\boldsymbol{\theta}(\boldsymbol{D},\boldsymbol{R})$ 最大：

$$\mathrm{mRMR}=\max_{\boldsymbol{f}_j\in\mathbf{F}-\mathbf{S}_{N-1}}\left(I(\boldsymbol{f}_j,\boldsymbol{O})-\sum_{\boldsymbol{f}_i\in\mathbf{S}_{N-1}}\frac{I(\boldsymbol{f}_i,\boldsymbol{f}_j)}{N-1}\right). \quad (11)$$

利用 mRMR 准则对特征空间进行预处理，可以剔除冗余特征，降低计算代价，产生紧凑性和泛化能力更强的模型。算法流程如下：

(1) 选择令相关性最大的特征 $\boldsymbol{f}_n$，即 $\max_{\boldsymbol{f}_n}I(\boldsymbol{f}_n,\boldsymbol{O})$，将所选特性特征添加到空集合 $\mathbf{S}$ 中。

(2) 在集合 $\mathbf{S}$ 的补集中找出具有非零相关性和零冗余的特征，如不包含，则转步骤 (4)；否则，选出相关性最大的特征 $\boldsymbol{f}_k$，即 $\max\limits_{\boldsymbol{f}_k\in\mathbf{S}^C,\boldsymbol{R}(\boldsymbol{f}_k)=0}I(\boldsymbol{f}_k,\boldsymbol{O})$，将选中的特征添加到集合 $\mathbf{S}$ 中。

(3) 重复步骤 (2)，直到 $\mathbf{S}$ 的补集中所有特征的冗余不为零为止。

(4) 选择 $\mathbf{S}$ 的补集中互信息熵最大且具有非零相关性和非零冗余的特征 $\boldsymbol{f}_l$，即 $\max\limits_{\boldsymbol{f}_l\in\mathbf{S}^C}I(\boldsymbol{f}_l,\boldsymbol{O})/\boldsymbol{R}(\boldsymbol{f}_l)$，将选择的特征加入集合 $\mathbf{S}$ 中。

(5) 重复步骤 (4)，直到 $\mathbf{S}$ 的补集中所有特征的相关性为零。

(6) 最后以随机顺序添加与 $\mathbf{S}$ 无关的特征。

1.3　改进的深度神经网络模型

1.3.1　传统前馈深度神经网络

传统的前馈深度神经网络 (Feedforward deep neural network, FF-DNN) 根据内部的神经网络层可以分为输入层 (输入声信号特征的层)、隐含层 (所有中间层) 和输出层 (输出目标距离估计值的层)。单层网络直接相互级联，某一层的任意一个神经元与其上一层的每一个神经元相连。其局部模型可描述为是一个线性运算加上一个非线性转移函数。

设深度神经网络层数为 L，l 是每一层的索引号，$\boldsymbol{x}^{(l)}$、$\boldsymbol{y}^{(l)}$ 分别是第 l 层的输入序列和输出序列，网络的输入 $\boldsymbol{y}^{(0)}=\boldsymbol{x}$，$\boldsymbol{w}^{(l)}$ 和 $\boldsymbol{b}^{(l)}$ 为第 l 层的权重矩阵和偏置向量，$f(z)$ 表示非线性转移函数。标准的 FF-DNN 网络描述如下 (对于 $l\in\{0,1,\cdots,L-1$ 层的第 i 个神经元$\}$) [17]：

$$\begin{cases}\boldsymbol{x}_i^{(l+1)}=\boldsymbol{w}_i^{(l+1)}\boldsymbol{y}^l+\boldsymbol{b}_i^{(l+1)},\\ \boldsymbol{y}_i^{(l+1)}=\boldsymbol{f}(\boldsymbol{x}_i^{(l+1)}).\end{cases}\tag{12}$$

训练模型时，利用寻优算法对距离估计的代价函数进行迭代优化求极小值，找到合适的线性系数矩阵 $\boldsymbol{w}$ 和偏置向量 $\boldsymbol{b}$：

$$(\boldsymbol{w}^*,\boldsymbol{b}^*)=\underset{\boldsymbol{w},\boldsymbol{b}\in\mathbb{R}^N}{\arg\min}\,J(\boldsymbol{w},\boldsymbol{b}),\tag{13}$$

其中，J 为代价函数，通常采用输出层输出的目标距离估计值与真实距离之间的均方误差：

$$J(\boldsymbol{w},\boldsymbol{b})=\frac{1}{2N}\sum_{j=1}^{N}||\boldsymbol{y}_j^L-\boldsymbol{z}_j||^2,\tag{14}$$

其中，N 是声信号训练样本数，j 是其样本索引号，$\boldsymbol{z}_j$ 是对应样本的真实距离。关于上述求解优化问题，常使用梯度下降法、共轭梯度法、拟牛顿法等数值优化方法。由于求 Hessian 矩阵及其逆计算量十分巨大，最常用的优化算法仍然是梯度下降算法。

在声源测距中，传统 DNN 存在以下缺陷：

(1) 测距误差较大。声源距离的代价函数可能高度非凸，迭代过程中容易陷入局部次优解或鞍点。

(2) 算法收敛速度慢。梯度下降法的初始学习率和调整策略需人工调节，相同的学习率被应用于各个参数，效率低下。

(3) 模型泛化性和鲁棒性弱。全连接网络的模型复杂度过高，参数稀疏度过低，易发生过拟合，以至于对环境变化和信号畸变过于敏感。

要解决以上问题，提高测距性能，重点在于如何改进寻优策略、加快收敛速度、防止过度学习。为此，本文引入 1 种自适应动态调整学习率的优化方法和 2 种网络参数稀疏化技术来改进网络模型。

1.3.2　Adam 优化算法

由于水下噪声、混响和水声信道的多途干扰，目标距离的代价函数通常为非凸函数且局部次优解，从而需要较大的学习率来跳出局部最优。然而，当在全局最优值附近搜索时，学习率太大会导致过度学习，降低声源测距精度。

Adam 算法是一种动态调整参数学习率的自适应优化方法 [18]。该方法通过梯度的一阶和二阶矩估计动态调整各网络参数的学习率，在迭代过程中通过偏差纠正使学习率维持在一定范围，从而获得平稳的参数更新，这是解决声源测距问题的理想方法。

设 t 为迭代次数，w 为待估参数，J 为代价函数，首先计算梯度的指数移动平均数 m_t。m_0 初值为 0。综合考虑之前时间步的梯度动量，设系数 β_1 为指数衰减率，有

$$m_t=\beta_1 m_{t-1}+(1-\beta_1)\nabla_w J(w_{t-1}).\tag{15}$$

计算梯度平方的指数移动平均数，v_0 初始化为 0。设系数 β_2 为指数衰减率，有

$$v_t=\beta_{t-1}+(1-\beta_2)\nabla_w J(w_{t-1})^2.\tag{16}$$

m、v 初始化为 0 会导致 m_t 偏向于 0，因此先进行偏差纠正再更新参数：

$$\begin{cases}\hat{m}_t=m_t/(1-\beta_1^t),\\ \hat{v}_t=v_t/(1-\beta_1^t),\\ w_t=w_{t-1}-\eta\cdot\hat{m}_t/\sqrt{\hat{v}_t},\end{cases}\tag{17}$$

式 (17) 中，η 为初始学习率。算法对更新的步长计算从梯度均值及梯度平方两个角度进行自适应的调节 [19]，起到提高迭代效率和测距精度的作用。

1.3.3　网络参数稀疏化

传统的 DNN 往往受限于特定的水声环境，对环境变化和信号畸变过于敏感，出现过拟合现象。具体表现在迭代过程中训练误差下降到一定程度时，测试误差反而开始增大。为了生成泛用性强的模型，将数据映射到网络特征后，网络特征之间的重叠信息应尽可能少，相关性尽可能低，从而近似于标准正交基。其主要方法是使特征产生稀疏性：稀疏特征有更大可能线性可分，或者对非线性映射机制有更小的依赖 [20]。

L2 正则化是一种简单且有效的网络参数稀疏化方法。在式 (14) 加入惩罚项，通过惩罚因子 λ 控制网络参数稀疏度：

$$\begin{aligned}&J(\boldsymbol{w},\boldsymbol{b})\\&=\frac{1}{N}\left(\sum_{j=1}^{N}\left\|\boldsymbol{y}_j^L-\boldsymbol{z}_j\right\|^2+\lambda\|\{\boldsymbol{w},\boldsymbol{b}\}\|^2\right).\end{aligned}\tag{18}$$

Dropout 正则化策略是另一种神经网络稀疏化手段，其核心在于每次权重更新迭代中，对网络的每

一层，随机将部分节点对应的权重值置零，使得线性系数矩阵和偏置向量达到稀疏化的效果，在一定程度上避免过拟合的问题。引入Dropout的神经网络描述由式(12)变为[17]：

$$\begin{cases} \boldsymbol{r}_j^{(l)} \sim \text{Bernoulli}(p), \\ \tilde{\boldsymbol{x}}^{(l)} = \boldsymbol{r}^{(l)} \cdot \boldsymbol{y}^l, \\ \boldsymbol{x}_i^{(l+1)} = \boldsymbol{w}_i^{(l+1)} \boldsymbol{y}^l + \boldsymbol{b}_i^{(l+1)}, \\ \boldsymbol{y}_i^{(l+1)} = f(\boldsymbol{x}_i^{(l+1)}), \end{cases} \tag{19}$$

式(19)中，符号$\cdot$表示向量点乘，$\boldsymbol{r}^{(l)}$是一个向量，其元素为服从伯努利随机分布的随机变量，分别以概率P和$1-P$取1和0为值，参数P是每个神经元的激活概率，通常P取[0.5,1.0]。使用该向量对上一层网络的输出$\boldsymbol{y}^{(l)}$进行采样，产生一个约减的输出$\tilde{\boldsymbol{y}}^{(l)}$用于下一次网络的输入。这个操作依次进行，从而可以生成一个稀疏的网络结构[17]。这种方法简单易行、节省运算资源，且不会提升优化过程的复杂度。改进后的DNN能够从有效的数据维度上学习到相对稀疏的特征，达到自动提取水声信号关键特征的效果。

2 数值仿真

通过KRAKEN声场计算工具，在声速正梯度浅海环境参数下生成仿真数据。图2描述了本文所使用的环境参数。

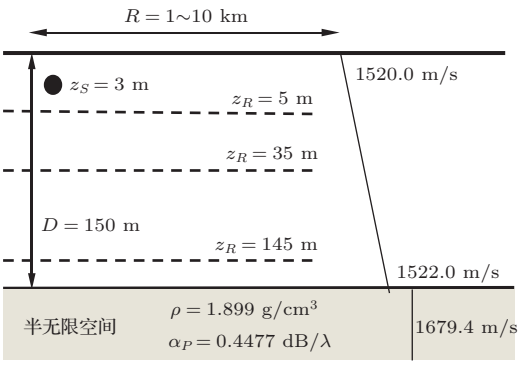


图2 环境参数

Fig. 2 Enviromental parameters

仿真数据包括连续波(Continuous wave, CW)在50 Hz、150 Hz和300 Hz的信号，线性调频(Linear frequency modulation, LFM)信号中心频率为500 Hz、1000 Hz和2000 Hz，频带宽度范围100 ~ 1000 Hz，信号长度0.2 ~ 1.0 s。将模拟接收信号拷贝100份并分成10组，每一组模拟接收信号分别添加信噪比(Signal-to-noise ratio, SNR)为1 dB、2 dB、3 dB、$\cdots$、10 dB的高斯噪声。接收点距离分布在1 ~ 10 km，深度是分布在5 ~ 145 m，网络输入训练集占总样本集80%，由16080个样本组成，剩余20%的数据作为测试集，由4020个样本组成。对生成样本进行多域特征提取，对每一帧得到的特征序列求统计特征，得到均值和方差，对所有帧的自相关系数和频域特征序列求统计特征，得到所有帧的时间均值和时间方差。最终提取到20100个样本的36维特征，作为DNN网络的输入特征，特征空间如图3所示。

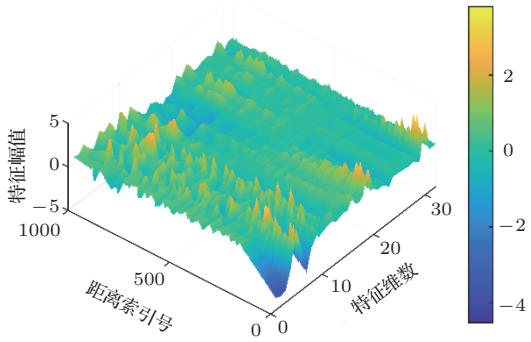


图3 特征空间

Fig. 3 Feature space

每个特征之间并不完全独立，有些特征与声源距离显著相关，这里根据1.2节中的mRMR算法进行特征选择。基于互信息的mRMR最高效和常用的[21]，在步骤(4)中，输出所选特征的互信息熵作为特征重要性的评价指标，对特征空间上所有特征进行重要度排序，结果如图4所示。

图4中的符号和表1、表2中一一对应，例如DS是下降斜率，SV_Mean是对信号每一帧的谱通量求的平均值；AC_Std是对每一帧自相关系数求的标准差。由排序结果可见与声源距离相关性最强的前3项特征是激励时间、下降斜率和谱通量均值，分别代表激励阶段的时间长度、衰减阶段能量的平均时间斜率和频谱包络面积的均值，这些物理量恰是反映声能量在传播过程中衰减的基本物理量。通常，为了兼顾特征集合的多样性和紧凑性，指标的阈值不宜过大或过小，经测试这里取0.03时，特征子集的维度为29，此时模型收敛性较好。最终得到与声源距离相关性最高的10维时域特征与19维频域特征。

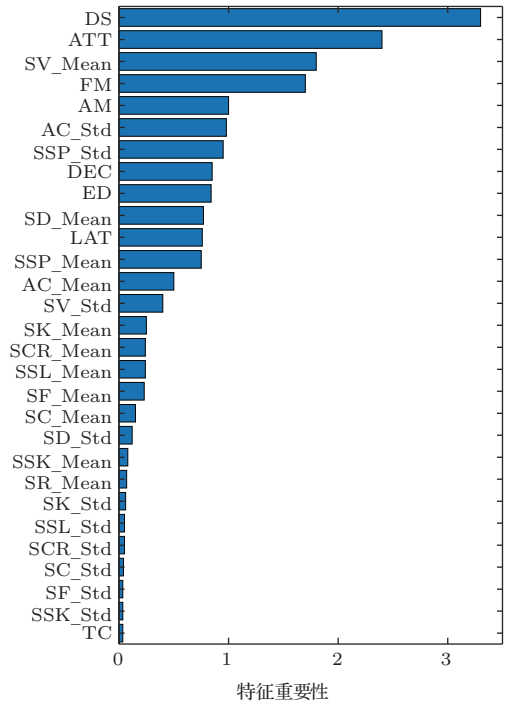


图 4　声学参数 mRMR 重要性排序

Fig. 4　The mRMR importance ordering of acoustic parameters

神经网络引入 Adam 优化算法进行网络训练，初始学习率采用 0.03，代价函数采用 MSE 函数，引入 Dropout 正则化处理，每次迭代神经元激活概率取 85%，初始权重由截断高斯分布模型产生，标准差为 0.1。经测试双曲正切函数、Sigmoid 函数和 ReLU 函数作为隐含层激活函数均未出现梯度消失现象，其中双曲正切函数在本问题中表现的收敛速度最快，且训练过程中未出现死神经元。隐含层激活函数采用双曲正切函数，输出层采用 200 个 Softmax 节点，对应不同的距离的概率，输出值最大的节点对应的距离为距离估计值。网络迭代次数设置为 20000 次。不同波形参数的单频信号和线性调频信号作为发射信号训练的 DNN 经 20000 次迭代后在测试集上的综合测距精确率达到 95% 以上，最高达到 98% 以上。以其中一组波形参数 (单频信号，$f_0 = 150$ Hz，$z_r = 35$ m，SNR $= 1 \sim 10$ dB) 的训练和测试结果为例，图 5 为该组信号的模型训练和测试结果。其中图 5(a) 给出了训练完成后模型最终在测试集上的距离估计结果，红线代表 KRAKEN 声场模型中给定的声源距离 (即真实距离)，蓝圈代表网络输出的估计距离；图 5(b) 给出了模型在训练集和测试集上的测量精度随迭代次数的变化曲线。

以单频信号作为发射信号，改变发射信号频率 ($f = 50$ Hz, 100 Hz, 150 Hz) 和声源深度 ($z_S = 0 \sim 140$ m, $\Delta z = 20$ m)，经 10000 次迭代，对比不同发射条件下的估计精确率，如图 6 所示，由图 6 可见波形参数和声源深度对模型性能的影响小，鲁棒性较好。

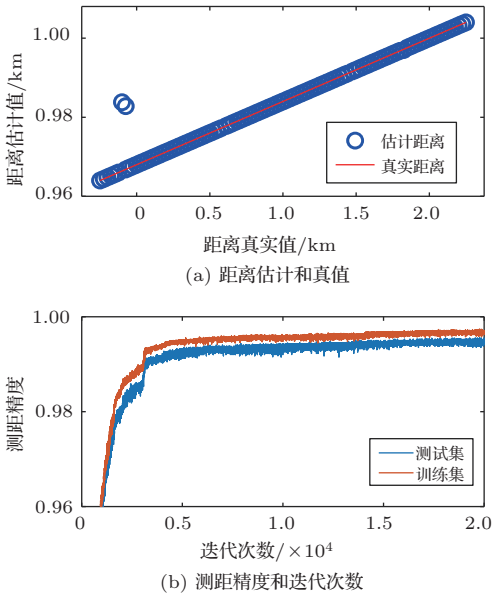


图 5　训练和测试结果

Fig. 5　The ranging accuracy by iteration times on validation and training sets

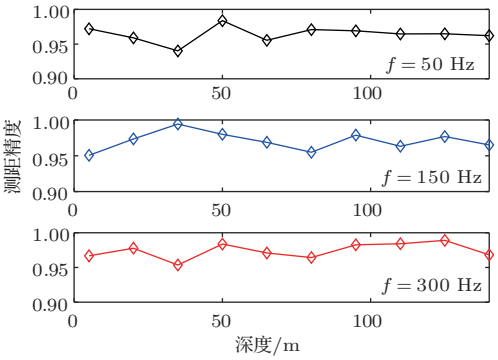
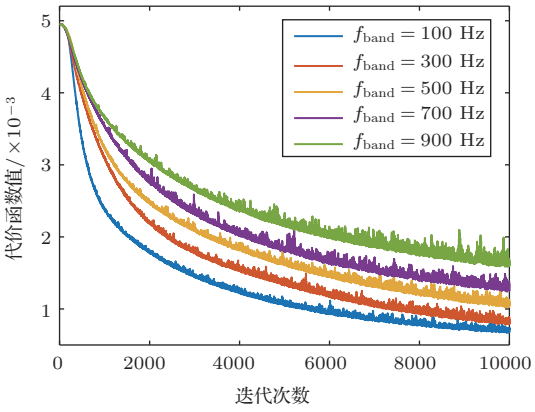


图 6　不同发射频率下各个深度的测距精确度

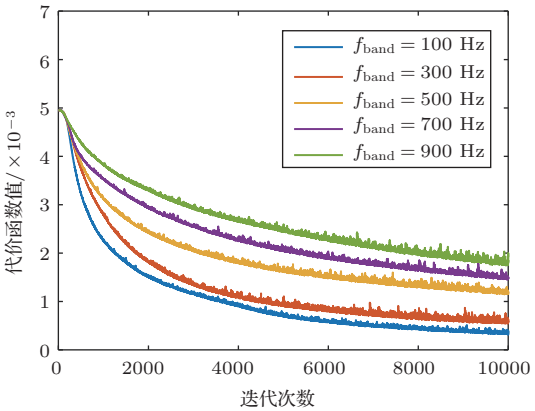
Fig. 6　Ranging accuracy at different depths with different transmission frequencies

以线性调频信号作为发射信号，分析模型的收敛速度和测量精度，经 10000 次迭代，对不同中心频率 ($f_c = 500$ Hz, 1000 Hz, 2000 Hz)、不同频带宽度 ($f_{\text{band}} = 100$ Hz, 300 Hz, 500 Hz, 700 Hz, 900 Hz) 和不同时间长度 ($T = 200$ ms, 400 ms, 600 ms, 800 ms, 1000 ms) 的信号源测距结果进行

对比。图7~图9给出了不同波形参数的信号的网络训练损失曲线，即训练过程中随着迭代次数增大，测试集上代价函数值的变化曲线。这里的代价函数值为测量的均方误差，下降越快，说明模型收敛速度越快。由此，图7表明模型的收敛速度和测量精度与频带宽度呈负相关，说明频带越宽所含信息越丰富，模型训练需要的时间越长；图8表明模型的收敛速度和测量精度与信号持续时间呈正相关，可见信号持续时间越长，呈现出的特征越明显；图9表明模型的收敛速度与中心频率和测量精度呈负相关，可见浅海中低频的信号特征更加集中，高频信号的特征更加分散，表明模型更适用于远程探测声呐。



(a) 中心频率为500 Hz



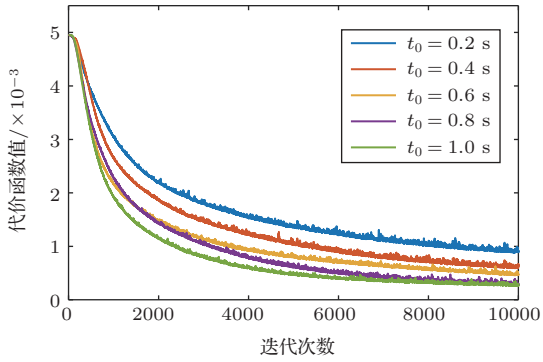
(b) 中心频率为1000 Hz

图7 信号时长为1.0 s时各频带宽度对应的网络训练损失曲线

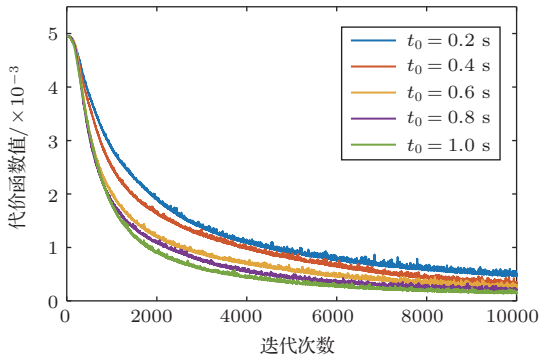
Fig. 7 Network training loss curve with different frequency bandwidth when the signal duration is 1.0 s

3 讨论和展望

相对已有的深度学习声源测距方法，本文提出的方法可提取信号的多域特征及对特征空间进行筛选，有利于产生紧凑性和泛化能力更强的模型；其次，改进了神经网络模型的寻优算法和参数稀疏化策略，可加快收敛速度并抑制模型过拟合。然而，机器学习是一种监督学习方法，其测距精度以建立功能全、信息丰富的数据库为代价。对于未知的海洋环境、时变的环境参数，需建立大量拷贝数据进行训练并加大网络的复杂度，同时需要已知的声源波形参数，对先验知识和计算资源有较大依赖性。



(a) 中心频率为500 Hz



(b) 中心频率为1000 Hz

图8 信号频带宽度为500 Hz时不同时间长度对应的网络训练损失曲线

Fig. 8 Network training loss curves with different time lengths when the frequency bandwidth is 500 Hz

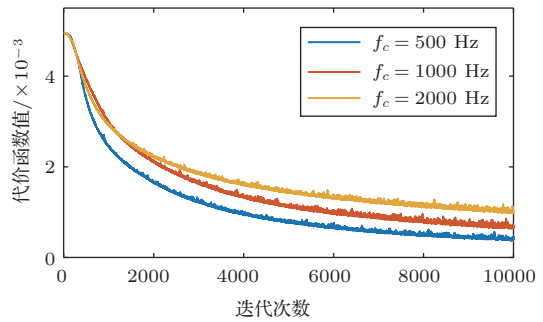


图9 信号频带宽度为500 Hz、时长为1.0 s时不同中心频率对应的网络训练损失曲线

Fig. 9 Network training loss curves with different center frequencies when the frequency bandwidth is 500 Hz and the duration is 1.0 s

目前已有相关研究[5]通过在真实海洋测试数据的基础上扩充仿真数据，以及采用迁移学习方法来减小深度学习方法对先验知识的依赖性。本文下一步工作将针对海洋测试中远距离测距的实际应用场景，通过建立数据库、优化神经网络结构、提高网络复杂度、增加输出节点数或改进输出层标签形式、增强自学习能力，以提升测距范围和测距分辨率。此外，可通过对数据库进一步扩充与声源目标识别相关的特征和标签，从而在声源被动测距的基础上同时执行目标识别相关的任务，如估计目标材料、形状、运动姿态等，以上需在后续工作中进一步探究。

4 结论

采用一种基于多域特征提取的深度学习方法来实现声源测距，通过浅海环境仿真实例验证了该方法的有效性，并分析了波形参数对测距性能和模型收敛速度的影响。本方法构建了声信号在时、频域的多维感知特征量，采用最大相关-最小冗余准则 (mRMR) 提取了声源和水下声场的关键信息，在传统深度神经网络的基础上引入自适应矩估计 (Adam) 优化、L2 正则化和 Dropout 正则化处理，提升模型的收敛速度和泛用性。结果表明：此方法在模型训练过程中收敛速度较快，预测性能较稳定，在所定条件下测试集上声源的综合测距精确率可达到 95% 以上，能够实现对声源距离的有效估计。此外，对不同发射信号训练效率的对比表明，算法性能对波形参数和声源深度具有良好的鲁棒性，模型收敛速度和测距精度对于带宽较小、持续时间长的瞬态发射信号较高。训练后的模型在单次测距任务中仅需执行毫秒级运算，可实现数据的实时处理。

本文提出采用的声源测距的深度学习方法，紧密结合了和海洋环境、传输距离相关的声源信号时频域多维感知特征，测试集上声源测距精度优良、可信。未来，建立并不断更新充实功能齐全、信息丰富的各种典型海洋环境参数、传输距离及声源信号多域特征大数据库，进一步优化算法和自学习能力后，本方法可望实现实时准确的水下目标被动定位、跟踪和分类识别，是今后深入研究的目标和方向。

参 考 文 献

[1] 张宗航, 孙超, 郭国强. 不确实海洋环境中的旁瓣结构声源测距方法 [J]. 电声技术, 2009, 33(7): 76–80.
Zhang Zonghang, Sun Chao, Guo Guoqiang. A method of scaling source range using matched field processor sidelobes in uncertain ocean environment[J]. Communication Electroacoustics, 2009, 33(7): 76–80.

[2] 胡志新. 基于深度学习的化工故障诊断方法研究 [D]. 杭州: 杭州电子科技大学, 2018.

[3] Lefort R, Real G, Drémeau A. Direct regressions for underwater acoustic source localization in fluctuating oceans[J]. Applied Acoustics, 2017, 116: 303–310.

[4] Niu H, Gong Z, Ozanich E, et al. Deep-learning source localization using multi-frequency magnitude-only data[J]. The Journal of the Acoustical Society of America, 2019, 146(1): 211.

[5] Wang W, Ni H, Su L, et al. Deep transfer learning for source ranging: deep-sea experiment results[J]. The Journal of the Acoustical Society of America, 2019, 146(4): EL317–EL322.

[6] 李思纯. 基于矢量水听器的目标特征提取与识别技术研究 [D]. 哈尔滨: 哈尔滨工程大学, 2008.

[7] 李晗, 陈克安, 田旭华. 基于平板冲击声的声源特性表征及自动识别 [J]. 应用声学, 2016, 35(4): 294–301.
Li Han, Chen Ke'an, Tian Xuhua. Characterization of sound source physical properties and automatic identi?cation based on impact sounds of plates[J]. Journal of Applied Acoustics, 2016, 35(4): 294–301.

[8] Peeters G, Giordano B L, Susini P, et al. The Timbre Toolbox: extracting audio descriptors from musical signals[J]. The Journal of the Acoustical Society of America, 2011, 130(5): 2902–2916.

[9] Brown J C, Houix O, Mcadams S. Feature dependence in the automatic identification of musical woodwind instruments[J]. The Journal of the Acoustical Society of America, 2001, 109(3): 1064–1072.

[10] Krimphoff J, McAdams S, Winsberg S. Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification[J]. Journal de Physique, 1994, 4(5): 625–628.

[11] Peeters G, Mcadams S, Herrera P. Instrument description in the context of MPEG-7[C]. Proc Icmc, 2000, 1(3): S125.

[12] 杨立学, 陈克安, 李双, 等. 飞机舱内声品质的音色参数表达 [J]. 西北工业大学学报, 2015, 33(3): 444–450.
Yang Lixue, Chen Ke'an, Li Shuang, et al. Timbre parameter representations of aircraft cabin sound quality[J]. Journal of Northwestern Polytechnical University, 2015, 33(3): 444–450.

[13] Scheirer E, Slaney M. Construction and evaluation of a robust multifeature speech/music discriminator[C]. Proc Icassp, 1997, 2.

[14] 朱知萌. 水下蛙人呼吸声信号特征提取研究 [D]. 哈尔滨: 哈尔滨工程大学, 2016.

[15] 旷玮, 姬培锋, 杨军. 笙的簧片物理参数与音色相关性的初步研究 [J]. 应用声学, 2016, 35(6): 494–504.
Kuang Wei, Ji Peifeng, Yang Jun. A study of the relationship between the physical parameters of Sheng reed and the timbre[J]. Journal of Applied Acoustics, 2016, 35(6):

494–504.

[16] 李梓瑞, 王慧琴, 胡燕, 等. 基于深度学习和最大相关最小冗余的火焰图像检测方法 [J]. 激光与光电子学进展, 2020, 57(10): 160–170.
Li Zirui, Wang Huiqin, Hu Yan, et al. Flame image detection method based on deep learning with maximal relevance and minimal redundancy[J]. Laser & Optoelectronics Progress, 2020, 57(10): 160–170.

[17] 杨阳, 张文生, 杨雪冰. 基于 Dropout 深度网络的两步图像标注算法 [J]. 计算机科学与探索, 2015, 9(12): 1494–1505.
Yang Yang, Zhang Wensheng, Yang Xuebing. Two steps image annotation algorithm based on deep network with Dropout[J]. Journal of Frontiers of Computer Science and Technology, 2015, 9(12): 1494–1505.

[18] 黄宇. 基于深度学习的跨物种 M6A 修饰位点预测研究 [D]. 青岛: 青岛大学, 2019.

[19] 肖琳. 基于深度学习多模型融合的路面裂缝识别算法研究 [D]. 西安: 长安大学, 2019.

[20] 周安众, 罗可. 一种卷积神经网络的稀疏性 Dropout 正则化方法 [J]. 小型微型计算机系统, 2018, 39(8): 1674–1679.
Zhou Anzhong, Luo Ke. Sparse Dropout regularization method for convolutional neural networks[J]. Journal of Chinese Computer Systems, 2018, 39(8): 1674–1679.

[21] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226–1238.

◇ 特讯 ◇

中国共产党优秀党员、著名声学家、教育家、我国声波测井理论的主要奠基人之一、吉林大学物理学院王克协教授因病医治无效，于 2020 年 11 月 18 日 9 时 05 分在北京逝世，享年 83 岁。

王克协教授 1938 年 1 月 7 日生于内蒙古通辽市，原籍辽宁大连市，1957 年考入东北人民大学 (今吉林大学) 物理系，1961 年加入中国共产党，1962 年毕业于吉林大学物理系理论物理专业，并留校任教。1990-1994 年，王克协教授担任吉林大学物理系主任，1991 年获国务院政府特殊津贴。

王克协教授是一位兢兢业业追求科学真理的科学家，他长期从事地声学与声测井科学研究工作，在国内最早之一开展了声波测井理论研究，1976 年成功地为石油系统举办了当时国内最早、最系统的声波测井物理基础培训班，为国内早期声波测井技术普及和发展做出了重大贡献。王克协教授学术生涯中始终紧跟声波测井专业发展前沿，从声波测井基础理论，到孔隙介质声学与井孔震电效应研究，再到多极声测井与地层各向异性研究、声波测井信号处理与地层参数反演研究、套管井固井质量声学检测方法研究、非线性声学与声弹性 - 地应力预测研究等，在众多方面均有建树。他在科研上取得的突出成绩得到了国内外同行的广泛关注和赞许，比如渗透率的直接求取、套管井水泥胶结评价技术、地层井孔附近应力集中分析等，在全球也只有屈指可数的几所大学和服务公司开展如此综合性的研究。王克协教授的科研工作大大缩短了国内声波测井研究与国外的差距，取得了令人瞩目的研究成果，出版了《声波测井理论、方法与应用——王克协教授科研论文选集》，极大地推动了我国声波测井技术的创新、发展与应用。1997 年，他主持的项目获吉林省教委科学技术进步一等奖，他本人在 2019 年度全国检测声学暨第十届全国储层声学与深部钻测技术前沿联合会议上被授予“元勋贡献奖”。

王克协教授将毕生奉献给高等教育和声波测井事业。他领导创建吉林大学声学专业，同时积极进行专业课程建设，先后独立开设 6 门硕士、博士专业课，奠定了有吉林大学特色的声学专业基础理论主干课程体系。他授课深入浅出，理论结合实际，先后主讲《理论声学基础》、《声测井理论与方法》、《孔隙介质声学》、《连续介质力学》、《非线性声学基础》、《数字信号处理》等课程。他指导并培养博士、硕士研究生近 50 名，为国家声波测井事业培养了一大批优秀人才，他的弟子们已在国内外相关领域中迅速成长，成为了技术专家和科研骨干，并传承着他的敬业精神。王克协教授不仅是一位成功的学者和导师，而且也为学生们树立了做人的榜样，他尊重学生，处处关心学生，对于家庭困难的学生，给予极大的鼓励和经济帮助，解除学生的后顾之忧，让学生全身心地投入到科研中去。王克协教授育人不倦的理念、敏锐的物理洞察、对科学探索的激情、严谨的治学态度、豁达的为人处世风格，都给学生们留下了深刻的印象和深远的影响。

王克协教授是一位德才兼备的优秀人民教师，深受学生、同事、同行的爱戴与尊敬。他是一位和蔼可亲、忠厚正派的长者，是新中国声波测井工作者们终身的老师，他的逝世是声学界的一大损失。王克协教授虽然走了，但他的精神永存，音容笑貌宛在，我们永远怀念他！他的高尚情操，教书育人、坚持不懈、勇于创新的精神，我们将铭记于心，并化悲痛为力量，为我国声波测井事业贡献力量，为实现中华民族的伟大复兴努力奋斗！

深切缅怀王克协教授，王克协教授千古！

吉林大学　崔志文

2021 年 1 月 1 日