基于广义合成分析和深度神经网络的自回归系数估计方法

崔子豪,鲍长春
(北京工业大学信息学部,北京 100124)

摘 要: 自回归(AR)模型是一类描述时序序列相关性的有效方法,经典的 AR 系数估计方法对残差信号做了简单的假设,在噪声干扰等复杂场景中难以准确估计 AR 系数,而基于深度神经网络(DNN)的 AR(DNN-AR)系数估计方法在训练中容易受到莱文逊-杜宾迭代(LDR)解法的数值稳定性的影响. 为改善 DNN-AR 系数训练的稳定性和整体性能,在保证系统稳定性的前提下,本文利用精度转化提高系统运算速度的思路,提出了基于广义合成分析(GABS)模型的深度网络结构改善方法,提高了 AR 系数在含噪环境下估计的准确性和网络训练的稳定性. 组合 DNN 的 GABS(GABS-DNN)的模型由三个主要部分组成:修正器的谱增强网络、编码器的 DNN 预处理及 LDR 参数估计和解码器的 AR 系数到功率谱的转换. 在优化目标函数的过程中,引入了增强谱和观测谱的误差,减少了反向传播时 LDR 的梯度对增强网络的影响,实现了稳定估计含噪语音的 AR 系数.

关键词: AR 系数; 广义合成分析; 深度神经网络; 莱文逊-杜宾迭代解
中图分类号: TN912.35 **文献标识码:** A **文章编号:** 0372-2112 (2021)01-0029-11
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.12263/DZXB.20200644

Auto-Regressive Coefficient Estimation Based on the GABS and DNN

CUI Zi-hao, BAO Chang-chun
(*Faculty of Information Technology, Beijing University of Technology, Beijing* 100124, *China*)

Abstract: The auto-regressive (AR) model is an effective method to describe the correlation of time series. The classic AR coefficient estimation method utilizes a simple assumption about residual signal. It is a challenge to accurately estimate the auto-regressive coefficients in a complex environment such as noise or interference. Even though Deep Neural Networks (DNN) based AR (DNN-AR) coefficient estimation method can estimate the AR coefficients in a complex environment, the DNN-AR method is easily affected by the numerical stability of Levinson-Durbin recursion (LDR) approach during the training stage. The main target is to improve the stability and overall performance of the DNN-AR based method. In this paper, the precision transform method is utilized to improve computational efficiency while keeping system stability, and the generalized analysis-by-synthesis combing DNN (GABS-DNN) model is proposed for improving the accuracy of AR coefficient estimation and stability of the DNN training in the noisy environment. The GABS-DNN model consists of three main parts: spectrum enhancement network in the modifier, DNN preprocessing and LDR parameter estimation at the encoder, and the conversion from autoregressive coefficient to power spectrum at the decoder. In the process of optimizing the objective function, the error between the enhanced spectrum and the observed spectrum is added for reducing the influence of the gradient of the LDR on the enhanced network during back-propagation, which results in a stable estimation of the AR coefficients of noisy speech.

Key words: auto-regressive (AR) coefficients; generalized analysis-by-synthesis (GABS); deep neural networks (DNN); Levinson-Durbin recursion (LDR)

收稿日期: 2020-06-30; 修回日期: 2020-08-18; 责任编辑: 孙瑶
基金项目: 国家自然科学基金(No. 61831019, No. 61471014)

1 引言

自回归(Auto-Regressive, AR)模型能够描述工程[1~3]和经济[4]等领域内稳定的时变信号. 在语音信号处理中,AR 系数被广泛应用于语音编码[5~7]、语音识别[8,9]和语音增强[10~13]等.

如何从观测的信号片段提取出 AR 系数是一类经典的信号处理问题. 传统的 AR 系数提取方法首先假设高斯白噪声信号为激励信号,建立 Yule-Walker 方程,最后求解 Yule-Walker 方程以得到 AR 系数[14,15]. 莱文逊-杜宾迭代算法[16,17](Levinson-Durbin Recursion, LDR)是一类有效求解 Yule-Walker 方程的方法,该方法被广泛应用于 AR 系数的估计中. 在语音信号处理里,清音语音符合高斯白噪声激励的自回归模型,由此 LDR 也广泛应用于清音语音的 AR 系数估计. 浊音的激励模型类似于冲击串,与 LDR 估计 AR 系数的假设不符[18]. 因此,清浊语音的 AR 系数估计在以往的方法中并不一致,本文希望采用一个模型能够有效估计清音或者浊音的 AR 系数,进而能对任意语音片段进行估计.

在语音信号处理中,常采用功率谱密度(Power Spectral Density, PSD)和激励信号的先验知识能建立 AR 系数估计模型[6,18~20]. 其中,语谱扭曲(Spectrum Distortion, SD)、对数谱扭曲(Log-Spectrum Distortion, LSD)、Itakura-Saito 散度(Itakura-Saito Divergence, ISD)以及 Kullback-Leibler 散度(Kullback-Leibler Divergence, KLD)等常用在功率谱[21]上表达语音激励的先验假设. 根据在功率谱上最小 IS 散度是高斯激励最大似然解的特点,El-Jaroudi 和 Makhoul[19]提出了离散全极点(Discreate All-Pole, DAP)模型,解决了 LDR 由于傅里叶变换点不足而不能达到理想效果的问题. Murthi 和 Rao[20]提出一种基于最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)AR 系数的估计方法,尝试估计浊音信号的 AR 系数. Mads 等人[7]提出了稀疏线性预测编码(Sparse Linear Prediction Coding, SLPC)方法,其稀疏的激励假设更符合浊音的激励信号类似于周期冲击串模型,建立了一种稀疏的线性预测模型. 然而这些 AR 系数估计模型仅能够对单一目标进行估计.

深度学习是一类学习非线性映射关系的方法. 由于 AR 系数的估计方法,即 Yule-Walker 方程的求解过程可以视为一种非线性映射的过程,因此可利用深度神经网络(Deep Neural Network, DNN)模拟从语音信号特征中估计 AR 系数这一非线性过程. 并且根据研究目标的不同,常常选用包含了全连接层(Fully Connected, FC)、卷积层[22]或者循环网络[23,24]等不同结构的网络层来构建信号处理模型. 作者前期的基于 DNN 的半自定编码器(Part-defined Auto-Encoder, PAE)[11]将 AR 系数的估计问题转化为包络谱的估计问题,并构建基于 AR 系数的维纳滤波器(AR-wiener filter)实现语音增强. 在估计 AR 系数时,DNN 和 LDR 各有优劣,为此,作者前期研究的基于 DNN 预处理的 DNN-AR 的 LDR 模型能更为有效地估计 AR 系数[25]. 该方法与广义合成分析模型[28]类似,将 LDR 视为编码,将 AR 系数映射到包络谱的函数关系视为解码,则 DNN 预处理可以视为谱的修正模型,以放宽信号的波形匹配,扩展 AR 系数的应用范围.

然而对于 DNN-AR 模型,在作者前期的工作中[25]仅探讨了低阶 AR 系数在无噪声干扰下的估计. 面对更复杂的情况,例如高阶 AR 系数或者噪声干扰下的 AR 估计,LDR 的引入会带来数值稳定性的影响,从而提高对系统的要求[26,27],其主要原因是,在 Yule-Walker 方程中,矩阵的条件数随着信号反射系数接近 1 而急剧增加,在精度不足时的数值计算中,会导致 LDR 的不稳定. 同时,对于随机初始化的深度神经网络训练而言,其反射系数的搜索范围可能会覆盖整个取值范围,即[0, 1). 根据 LDR 数值稳定性的特征,一种有效避免训练不稳定的方法是提高计算精度,然而这意味着额外的运算时间和内存需求. 本文在 DNN-AR 模型的基础上,通过分析讨论模型 LDR 非线性映射所带来的稳定性问题,分析精度对 Toeplitz 矩阵的非负定条件和系统稳定性的影响,采用一种精度转换的方法在较好地在保证系统稳定性的前提下,提高了系统的运算效率. 此外,本文扩展了 DNN-AR 模型,引入广义合成分析策略(Generalised Analysis-By-Synthesis, GABS)[28],提出一种基于 GABS 和深度神经网络的 AR 系数估计方法(GABS-DNN),通过在深度学习中增加额外的反馈路径,一定程度上也能改善 DNN-AR 模型在含噪语音中可能出现的不稳定现象.

2 基于 GABS-DNN 的 AR 系数估计

2.1 经典 AR 估计模型与方法

p 阶 AR 模型处理平稳随机信号 $s(n)$ 可表示为

$$s(n) = -\sum_{i=1}^{p} a_i s(n-i) + e(n) \tag{1}$$

其中 a_i 指第 i 阶的 AR 系数,$e(n)$ 指高斯白噪声残差激励信号,其方差为 σ^2. AR 模型的 PSD 表示为

$$\hat{\phi}_S(k) = \frac{\sigma^2}{\left|1 + \sum_{i=1}^{p} a_i \mathrm{e}^{-\mathrm{j}\omega_k i}\right|^2}, k = 0, \cdots, N-1 \tag{2}$$

其中 $\omega_k = 2\pi k/N$,ϕ 代表功率谱,符号^表示估计的含义. 由此可以通过估计的 AR 参数替代真实的参数得到其功率谱密度. 经典的 AR 估计通过使激励信号目标函数的功率最小,在任意信号片段 $\boldsymbol{x}$ 中估计其 AR 系

数,$\boldsymbol{x}$ 和目标函数分别表示为

$$\boldsymbol{x}=[x(0)\quad x(1)\quad\cdots\quad x(N-1)]^{\mathrm{T}} \tag{3}$$

和

$$(\hat{\boldsymbol{a}},\hat{\sigma}^2)=\operatorname{argmin}_{\boldsymbol{a},\sigma^2}\sum_{n=0}^{N-1}e^2(n) \tag{4}$$

其中 AR 系数为

$$\boldsymbol{a}=[a_1\quad a_2\quad\cdots\quad a_p]^{\mathrm{T}} \tag{5}$$

假设观察信号在窗外值为 0,即对于 $n<0\vee n\geq N$ 有 $x(n)=0$,由式(4)可以导出

$$\hat{\boldsymbol{a}}=-\hat{\boldsymbol{R}}_X^{-1}\hat{\boldsymbol{r}}_X \tag{6}$$

其中 $\hat{\boldsymbol{R}}_X$ 和 $\hat{\boldsymbol{r}}_X$ 分别为估计的自相关矩阵和自相关矢量,如式(7)~(9)所示

$$\hat{\boldsymbol{r}}_X=[\hat{r}_X(1)\quad\cdots\quad\hat{r}_X(p)]^{\mathrm{T}} \tag{7}$$

$$\hat{\boldsymbol{R}}_X=\begin{bmatrix}\hat{r}_X(0) & \cdots & \hat{r}_X(p-1)\\ \vdots & \ddots & \vdots\\ \hat{r}_X(p-1) & \cdots & \hat{r}_X(0)\end{bmatrix} \tag{8}$$

$$\hat{r}_X(m)=\frac{1}{N}\sum_{n=0}^{N-1-m}x(n+m)x(n) \tag{9}$$

激励的方差 σ^2 为

$$\hat{\sigma}^2=\hat{r}_X(0)+\hat{\boldsymbol{r}}_X^{\mathrm{T}}\hat{\boldsymbol{a}} \tag{10}$$

由于 $\hat{\boldsymbol{R}}_X$ 是一个 Toeplitz 矩阵,LDR 可以非常有效地对参数 $\hat{\boldsymbol{a}}$ 进行估计[16,17]. 并且,由于式(1)为全极点模型,参数 $\hat{\boldsymbol{a}}$ 是稳定的.

2.2　基于 DNN 预处理的 AR 参数估计

相较于直接使用 DNN 估计其反射系数,基于 DNN 预处理的 AR 估计模型结合 LDR 算法使得其非线性映射模型更容易进行学习训练[25]. 为使其能够有效对含噪语音的 AR 系数进行估计,通过增加延迟器组使 DNN 网络可以更好利用相邻帧估计功率谱,本文将其扩展为如图 1 所示模型. 该模型与 GABS 有相似之处[28],DNN 增强含噪的语谱或者修正纯净语谱以扩展经典的 LDR 算法,使其能够使表 1 所示的测度达到最小. DNN 预处理用于修正功率谱,并将其转化为 Yule-Walker 方程所需的自相关序列,即

$$\boldsymbol{W}_X=\log_{10}(\boldsymbol{\Phi}_X) \tag{11}$$

$$\boldsymbol{w}_Y=\boldsymbol{f}_\theta(\boldsymbol{W}_X) \tag{12}$$

$$\hat{r}_X(m)=\frac{1}{N}\sum_{k=0}^{N-1}10^{w_Y(k)}\mathrm{e}^{\mathrm{j}\omega_k m} \tag{13}$$

其中 $m=0,1,\cdots,p$,$\boldsymbol{\theta}$ 表示 DNN 的参数,其多帧的含噪语谱可以表示为

$$\boldsymbol{\Phi}_X=[\phi_X^{-m};\quad\cdots;\quad\phi_X^{m}] \tag{14}$$

$$\phi_X=\left[\phi_X(0)\quad\cdots\quad\phi_X\left(\frac{N}{2}\right)\right]^{\mathrm{T}} \tag{15}$$

$$\boldsymbol{w}_Y=\left[w_Y(0)\quad\cdots\quad w_Y\left(\frac{N}{2}\right)\right]^{\mathrm{T}} \tag{16}$$

$$\phi_Y(k)=\phi_Y(N-k) \tag{17}$$

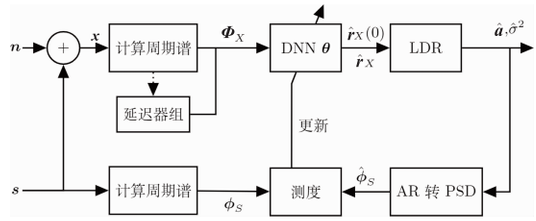


图1　DNN-AR估计模型的训练阶段原理框图

表 1　文中功率谱上的常用测度

测度	公式　$D(\phi_X,\tilde{\phi}_X)$
IS 散度 (IS-divergence)	$\frac{1}{K}\sum_{k=1}^{K}\left[\frac{\phi_X(k)}{\tilde{\phi}_X(k)}-\ln\frac{\phi_X(k)}{\tilde{\phi}_X(k)}-1\right]$
KL 散度 (KL-divergence)	$\frac{1}{K}\sum_{k=1}^{K}\left[\phi_X(k)\ln\frac{\phi_X(k)}{\tilde{\phi}_X(k)}-\phi_X(k)+\tilde{\phi}_X(k)\right]$
β 散度 (beta-divergence,β)	$\frac{1}{K\beta(\beta-1)}\sum_{k=1}^{K}[\phi_X^{\beta}(k)+(\beta-1)\tilde{\phi}_X^{\beta}(k)-\beta\phi_X^{\beta}(k)\tilde{\phi}_X^{\beta-1}(k)]$
语谱扭曲 (Speech Distortion,SD)	$\frac{1}{K}\sum_{k=1}^{K}[\phi_X(k)-\tilde{\phi}_X(k)]^2$
平均绝对误差 (Mean Absolute Error, MAE)	$\frac{1}{K}\sum_{k=1}^{K}\lvert\phi_X(k)-\tilde{\phi}_X(k)\rvert$
对数谱扭曲 (Log-Spectrum Distortion,LSD)	$\frac{1}{K}\sum_{k=1}^{K}\left[\log_{10}\frac{\phi_X(k)}{\tilde{\phi}_X(k)}\right]^2$
对数谱绝对误差 (1-norm of the Log-Spectrum Distortion,LSD-L1)	$\frac{1}{K}\sum_{k=1}^{K}\left\lvert\log_{10}\frac{\phi_X(k)}{\tilde{\phi}_X(k)}\right\rvert$

其中 ϕ_X^{-m} 指延迟 m 帧的 PSD.

这里,DNN 的输入和输出参数是固定且相似的,这使得 DNN 的映射 $\boldsymbol{f}_\theta(\cdot)$ 相对简单. 在纯净语音测试下,采用单位矩阵参数设计使 $\boldsymbol{\varphi}_Y=\boldsymbol{\varphi}_X$,以接近最小 IS 散度,从而使 DNN-AR 估计模型等效为经典的 LDR 算法. 由于基于 DNN 的语音增强常采用对数谱参数作为网络输入和输出,在此 $\boldsymbol{f}_\theta(\cdot)$ 也采用这一形式[29]. 基于上述网络模型,可以构建以功率谱为基础的目标函数,

$$L=\frac{1}{N}\sum_{i=1}^{N}D(\phi_S,\hat{\phi}_S) \tag{18}$$

其中,D 代表测度,常用测度如表 1 所示.

2.3　基于 GABS 和 DNN 的 AR 参数估计

DNN-AR 的非线性会导致在复杂输入环境下如噪声干扰下数值不稳定[27],而 DNN-AR 的非线性是由于 LDR 的非线性引起的. 针对数值不稳定问题,本文提出

如图 2 所示的基于 GABS 策略[28]和 DNN 的 AR 系数估计模型(GABS-DNN)的解决方法(如图 3 ~ 图 4 所示). 对于含噪语音的输入,语音增强网络 DNN1 对应于 GABS 的修正器,DNN 预处理和 LDR 为 GABS 的编码器,AR-PSD 为 GABS 的解码器,并引入观测谱 ϕ_S 和增强谱 $\bar{\phi}_X$ 的比较,以减弱非线性激活函数 LDR 对增强网络的影响,其增强网络 DNN1 处理流程为

$$\boldsymbol{W}_X = \log_{10}(\boldsymbol{\Phi}_X) \tag{19}$$

$$\bar{\boldsymbol{w}}_Y = \boldsymbol{f}_{\theta 1}(\boldsymbol{W}_X) \tag{20}$$

$$\bar{\phi}_X = 10^{\bar{w}_Y} \tag{21}$$

其修正网络 DNN 流程为

$$\bar{\boldsymbol{w}}_X = \log_{10}(\bar{\phi}_X) \tag{22}$$

$$\boldsymbol{w}_Y = \boldsymbol{f}_{\theta}(\bar{w}_X) \tag{23}$$

$$\hat{r}_X(m) = \frac{1}{N}\sum_{k=0}^{N-1} 10^{\boldsymbol{w}_Y(k)} \mathrm{e}^{\mathrm{j}\omega_k m} \tag{24}$$

采用的目标函数为

$$L = \frac{1}{N}\sum_{i=1}^{N} \lambda D(\phi_S, \bar{\phi}_X) + D(\phi_S, \hat{\phi}_S) \tag{25}$$

其中 λ 为权重参数,$D(\phi_S, \bar{\phi}_X)$ 为引入 ϕ_S 和 $\bar{\phi}_X$ 的比较. 理想情况下,增强网络 DNN1 的输出为纯净语音的功率谱参数,此时可以认为 DNN1 之后的处理过程为纯净语音的 DNN-AR 模型. 当网络输入为纯净语音时,可以忽略增强网络 DNN1 的作用,以纯净语谱输入网络 DNN 中,此时 GABS-DNN 网络与图 1 所示的 DNN-AR 估计模型相同.

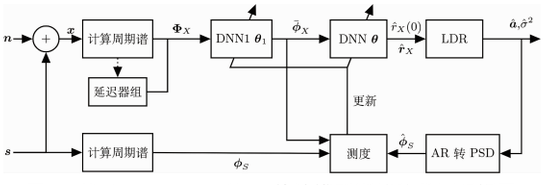


图2 基于GABS-DNN的AR估计模型训练阶段原理框图

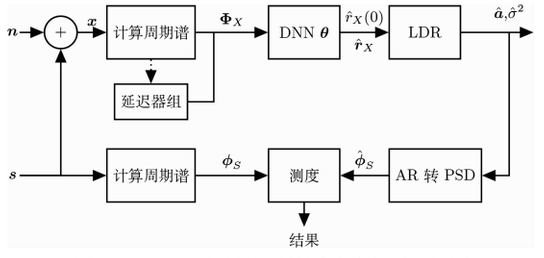


图3 DNN-AR估计模型的测试阶段原理框图

3 GABS-DNN 稳定性与精度变换方法

一般情况下,估计语音信号的 AR 系数基本是稳定的,但是, LDR 的数值稳定性使 AR 系数估计在深度学习的训练过程中仍有可能会遇到一些病态的、甚至负定的 Toeplitz 矩阵问题. 本部分研究深度学习的参数和超参数对 DNN-AR 网络数值稳定性的影响,并在保证系统稳定性的前提下提升运算性能.

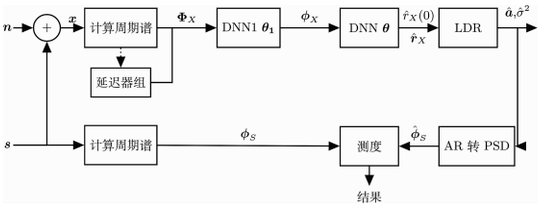


图4 基于GABS-DNN的AR估计模型测试原理框图

3.1 LDR 的数值稳定性

AR 系数与残差信号的关系可以表示为

$$\frac{\boldsymbol{a}-\hat{\boldsymbol{a}}}{\boldsymbol{a}} = \kappa(\boldsymbol{R}_X)\frac{\hat{\boldsymbol{e}}}{\hat{\boldsymbol{r}}_X} \tag{26}$$

其中,$\hat{\boldsymbol{e}}$ 表示残差信号,$\boldsymbol{a}$ 表示观察的 AR 参数,$\kappa(\hat{\boldsymbol{R}}_X)$ 是矩阵 $\hat{\boldsymbol{R}}_X$ 的条件数[30],即

$$\kappa(\hat{\boldsymbol{R}}_X) = \hat{\boldsymbol{R}}_X\hat{\boldsymbol{R}}_X^{-1} \tag{27}$$

Toeplitz 矩阵的条件数范围可以表示为[27]

$$\max\left\{\frac{1}{E_{n-1}}, \frac{1}{\prod_{j=1}^{n-1}(1-\hat{K}_j)}\right\} \leqslant \kappa(\hat{\boldsymbol{R}}_X) \leqslant n\prod_{j=1}^{n-1}\frac{1+|\hat{K}_j|}{1-|\hat{K}_j|} \tag{28}$$

其中 K_i 指第 i 个反射系数,E 为能量

$$\hat{K}_j = -\frac{c_1 + \sum_{j=1}^{i-1}\hat{a}_{i-1}, c_{i-j}}{E_{i-1}} \tag{29}$$

并且残差信号的幅值小于 $o(n)\prod_{j=1}^{n-1}(1+|\hat{K}_j|)$,其中 $o(n)$ 是与时序索引 n 相关的常数[27].

由于反射系数的范围是 $0 \leqslant |K_j| < 1$,当反射系数趋近于 1 时,$\prod_{j=1}^{n-1}\frac{1}{1-|K_j|}$ 的数值趋近于无穷,Toeplitz 矩阵成为病态矩阵,将导致 LDR 的数值稳定性问题.

3.2 LDR 的反向传播

深度学习的优化算法离不开计算梯度. 在 DNN-AR 模型中,LDR 在深度网络中可视为非线性激活函数,其 Yule-Walker 方程本身的条件数在深度学习中不仅影响当前批次的学习,也需要考虑下一批次的稳定性. 根据式(10)可知,其散度为

$$\nabla_{r_X}\sigma^2 = [1;\boldsymbol{a}] + \boldsymbol{r}_X^{\mathrm{T}}\nabla_{r_X}\boldsymbol{a} \tag{30}$$

由 AR 系数与残差信号间的关系式(26),可以近似构建 AR 系数与自相关系列的散度关系,

$$\hat{\boldsymbol{r}}_{X,1}\frac{\boldsymbol{a}-\hat{\boldsymbol{a}}_1-(\boldsymbol{a}-\hat{\boldsymbol{a}}_2)}{\hat{\boldsymbol{r}}_{X,1}-\hat{\boldsymbol{r}}_{X,2}} \approx \boldsymbol{a}C(n)\prod_{j=1}^{n-1}\frac{(1+|K_j|)^2}{1-|K_j|}$$

$$\hat{\boldsymbol{r}}_X\nabla_{r_X}(\boldsymbol{a}-\hat{\boldsymbol{a}})|_{r_X=\hat{r}_X} \propto \prod_{j=1}^{n-1}\frac{1}{1-|K_j|} \tag{31}$$

其中,$\boldsymbol{a}$ 是 AR 系数的期望值. 对于一个观测到的时域序

列,$\boldsymbol{a}$ 唯一. 将其带入式(30),增益的散度可以近似由其梯度关系代替. 忽略次要成分,则散度关系可以近似表示为

$$\nabla_{r_X}\sigma^2 \approx \nabla_{r_X}\boldsymbol{a} \approx C\prod_{j=1}^{n-1}\frac{1}{1-|K_j|} \tag{32}$$

其中常数 C 近似替代被忽略的次要成分. 由此,深度学习中 LDR 激活函数的散度受到 Yule-Walker 方程中的 Toeplitz 矩阵条件数影响,具体由反射系数 $\prod_{j=1}^{n-1}\frac{1}{1-|K_j|}$ 反应.

由于 IS 散度与残差信号的最小能量均可表示为 AR 模型高斯激励的最大对数似然解,即

$$(\hat{\boldsymbol{a}},\hat{\sigma}^2) = \arg\min_{\boldsymbol{a},\sigma^2}\sum_{n=0}^{N-1}e^2(n) \tag{33}$$

$$(\hat{\boldsymbol{a}},\hat{\sigma}^2) = \arg\min_{\boldsymbol{a},\sigma^2}D_{IS}(\phi_S,\hat{\phi}_S) \tag{34}$$

其中 D_{IS} 指 IS 散度. 假设式(33)和式(34)的散度具有相似关系

$$\nabla_{r_X}D_{IS}(\phi_S,\hat{\phi}_S) \approx \nabla_{r_X}\sum_{n=0}^{N-1}e^2(n) = \nabla_{r_X}\sigma^2 \tag{35}$$

与式(32)相似,它表示了功率谱密度关于自相关序列的散度. 考虑其主要的影响部分,有

$$\nabla_{r_X}D_{IS}(\phi_S,\hat{\phi}_S) \approx C\prod_{j=1}^{n-1}\frac{1}{1-|K_j|} \tag{36}$$

其 $k+1$ 次迭代可由第 k 次迭代表示

$$L^{k+1} = \frac{1}{N}\sum_{i=1}^{N}D(\phi_S,\mathrm{F}(\hat{\boldsymbol{r}}_x^k + l_r\nabla_{r_x}D(\phi_S,\hat{\phi}_S^k)\mid_{r_x=\hat{r}_x})) \tag{37}$$

其中,F 指自相关序列构建 Yule-Walker 方程后,将 LDR 估计的 AR 系数 $\hat{\boldsymbol{a}}$ 和增益 $\hat{\sigma}^2$ 转化为功率谱密度 $\hat{\phi}_S^{k+1}$ 的过程. 从式(37)可知,影响深度学习训练过程中的稳定性因素不仅包括观测信号本身的反射系数,也包括优化算法的超参数 l_r 等.

在 GABS-DNN 模型中,由于目标函数引入了增强谱和目标谱的比较,则对于增强网络 DNN1 中参数 θ_1 的更新,根据式(25)及其梯度关系,有

$$\nabla_{\theta_1}D_{IS}(\phi_S,\hat{\phi}_S) \approx \nabla_{\theta_1}\bar{\phi}_X\left(\nabla_{\phi_X}\boldsymbol{r}_X C\prod_{j=1}^{n-1}\frac{1}{1-|K_j|}+\lambda\right) \tag{38}$$

分析式(38),通过调整比例系数 λ,一定程度上可减弱 LDR 算法对增强网络产生的影响.

3.3　GABS-DNN 的数值精度

为保证系统稳定,还需要使构建的 Toeplitz 矩阵非负定,在数值计算中,采用双精度计算能在一定程度上能避免 Toeplitz 矩阵负定. 在 MATLAB 中由于采用了双精度计算,LDR 能稳定估计 AR 系数. 然而双精度的运算会带来额外的计算开销,为此本文采用精度转化的策略来保证计算稳定的同时提高运算效率.

对式(13)和式(36)所示全连接层输出的功率谱 $\bar{\phi}_X$,加上一理想高斯白噪声扰动 $e(k)\sim N(0,\sigma_b^2)$,且噪声与功率谱无关,根据帕斯瓦尔定理[31]可估计扰动的平均功率,

$$\sum_{k=0}^{N-1}\hat{r}_e(m) = \frac{1}{N}\sum_{k=0}^{N-1}|e(k)|^2 = \frac{1}{2\pi}\sigma_b^2 \tag{39}$$

$\hat{r}_e(m)$ 的期望为 $\frac{1}{2\pi N}\sigma_b^2$,理想情况下其构成的自相关矩阵 $\hat{\boldsymbol{R}}_e$ 为半正定矩阵,扰动后的 Toeplitz 矩阵期望任即为非负定矩阵,即

$$\hat{\boldsymbol{R}}_{X1} = \hat{\boldsymbol{R}}_X + \hat{\boldsymbol{R}}_e \tag{40}$$

在理想情况下,扰动并不会导致自相关矩阵负定,仅产生与 σ_b^2 相关的偏差. 由于 $\hat{r}_e(m)>0$,则实际中单点最大偏移值为 $\frac{1}{2\pi}\sigma_b^2$. 假设在自相关序列上加一相同分布的白噪声扰动,且高斯分布的绝大数绝对误差小于 $3\sigma_b$,则变换前最大误差 $3\sigma_b$ 与变换后最大误差 $\frac{1}{2\pi}\sigma_b^2$ 比与 σ 成反比. 假设 $\bar{\phi}_{e2}$ 代表较高精度的功率谱密度与真实功率谱密度的偏差,$\bar{\phi}_{X1}$ 代表较低精度的功率谱密度,假设在低精度转高精度时产生误差与 $e(k)$ 相近,即可表达为

$$\hat{\boldsymbol{R}}_{X1} = \hat{\boldsymbol{R}}_X + \hat{\boldsymbol{R}}_{e2} + \hat{\boldsymbol{R}}_e \tag{41}$$

其中,$\hat{\boldsymbol{R}}_X$ 为理想非负定的 Toeplitz 矩阵,$\hat{\boldsymbol{R}}_{e2}$ 为能保证系统稳定运行采用精度所引起的误差,则高精度在自相关函数时的精度误差 σ_{b2}^2 约等于在功率谱时的误差 σ_{b2}. 对于浮点数运算而言,只要保证低精度转高精度的误差小于 σ_{b2},且数值可由低精度表示,就能较好地保证系统稳定运行. 由此,在实际的后续 LDR 计算中,为保证 LDR 的稳定运行,神经网络的计算精度应该大于等于 32 位. 本文采用 32 位浮点数运算,并在估计修正的对数谱 $\bar{\phi}_Y$ 后转换为 64 位浮点数运算.

综上所述,影响 GABS-DNN 稳定性的因素包括观测序列的 Toeplitz 矩阵条件数相关的权重 η、测度 D、网络精度 b_θ、优化方法 opt 及学习率 l_r 等,可表达为

$$G_{\mathrm{opt}}(D,l_r,b_\theta,\eta) \tag{42}$$

其中 $\eta=\prod_{j=1}^{n-1}(1-|K_j|)$. 测度、网络精度、学习率和优化方法是训练前设定好的超参数外,在实验中应该关注权重 w.

4　实验设置与结果

对经典的 LDR、DAP 方法以及基于神经网络的 PAE、DNN-AR、GABS-DNN 模型,本文分别采用纯净和含噪语音进行测试,以验证精度变化方法、GABS-DNN 模型的有效性以及 DNN-AR 的适应能力.

4.1 训练和测试数据

本文采用的语音数据库和噪声数据库分别来自TIMIT[32]、LibriSpeech[33]和NoiseX-92[34]、DEMAND[35]. 语音数据的采样频率为16kHz,噪声重采样到16kHz以便与语音相对应. 纯净语音和含噪语音的傅里叶变换采用512点的正弦窗[36]. 训练含噪语音时,首先采用NoiseX-92噪声数据库中的多人讲话噪声(Babble)、F16环境噪声(F16)和工厂噪声(Factory1)中的片段与TIMIT数据库中每句语音按照信噪比(Signal-to-Noise Ratio,SNR)服从$U[-5,15]$dB进行混合,帧延迟集合为$[-5,-4,-3,-2,-1,1,2,3,4,5]$帧. 训练纯净语音时,SNR设为无穷大,且忽略延迟器组的输出. 本文随机选取2000句TIMIT数据库中的语音与上述3种噪声混合,得到6000句纯净语音和6000句含噪语音进行训练. 在训练GABS-DNN时,式(38)中的比例系数λ为1.

测试仍然采用Babble、F16和Factory1作为已见类别(seen)环境噪声,采用Demand数据库中的自助餐厅噪声(Cafeteria)作为未知类别(unseen)环境噪声,并采用TIMIT数据库和SpeechLibri数据库中各100条非训练样本语句分别作为已知类别和未知类别测试语音,与4种不同信噪比SNR$\in[-5,0,5,10]$dB的噪声数据混合作为观测语音的含噪语音,其他参数与训练时一致.

4.2 比较对象和参数设置

在纯净语音测试时,直接以经典的LDR、DAP以及简化的基于神经网络的PAE[11]作为比较对象,用于测试精度转换的可行性. 简化的PAE网络结构与图1类似,不同之处在于DNN的输出特征为反射系数与增益,再转为AR系数[37]. 采对于含噪语音,语音增强网络[38]和专门用于处理纯净语音的DNN-AR模型的组合(SE+DNN)作为比较对象,并比较GABS-DNN、SE+DNN以及处理含噪语音的DNN-AR的性能差异,各个模型的简称和含义如表2所示. 对于测试中所有的深度学习网络均采用三层全连接结构,深度网络对对数谱输入进行0均值归一化处理,对对数谱的输出去归一化,对反射系数输出则用双曲正切函数(tanh)作为网络输出层激励. 网络的隐藏层包含2048个神经元,后接带泄露修正线性单元(Leaky Rectified Linear Unit, Leaky-Relu)[39]和dropout层[40],Leaky-Relu负数的斜率为0.2,drop-out率为0.2. AR阶数为16,一般地,采用2e-4的学习率和Radam[41]算法进行训练,训练不稳定时采用更保守的5e-5的学习率和Radam算法进行训练,如果当前周期的误差大于两个周期前的误差,则学习率减半. 训练共有130个周期,每个周期有200个子集,每个子集约有6700个训练样本. 图1~4所示DNN的训练和测试在Pytorch[42]下完成,LDR和最终评测由MATLAB计算完成.

表2 测试模型的简称表

简称	纯净语音	含噪语音
LDR	LDR算法	纯净语音的LDR
DAP	DAP算法	–
DNN	DNN-AR模型	DNN-AR模型
SE+DNN	–	增强网络和纯净语音训练的DNN-AR模型
GABS	–	所提基于GABS和DNN的AR估计模型

为了便于比较不同精度的训练速度,训练和测试环境为:32GB DDR3 1867MHz内存、i7 4790 CPU以及VIDIA GeForce RTX 2060 SUPER.

4.3 精度测试

在测试GABS-DNN模型之前,先通过实验分析精度转换方法是否适用. 由表3可知,训练耗时明显缩短,采用精度转换方法的耗时约为64位浮点数运算耗时的1/6. 图5是结合了精度转换的DNN-AR与LDR、DAP、PAE测试结果的对比,其中,以DNN为基础的方法在测试和训练中的目标相同. 由图5可知,在IS散度下,LDR和DAP较优,这主要是由于该类方法是求解最小IS散度的精确解. PAE网络的测试结果较差,这可能是由于三层网络结构并不能准确估计反射系数. 图中四种模型的Beta散度和SD相似,LSD,AE和LSD-L1测度下DNN-AR方法较优,且在这几种目标函数的训练和测试中,除SD测度的64位浮点数运算的误差稍优于精度转换方法外,其余测试结果非常接近. 由此验证,可以用精度转换的方法保证系统稳定,并实现加速训练的目的.

表3 两种不同精度方法的训练耗时

	32位转64位	64位浮点数
平均耗时/s	6730	40540

图6比较了KL散度、IS散度、β散度、LSD,SD和MAE六种不同测度训练时的最小对数非线性权重值,即$\min(\log_{10}(\eta))$,该权重值越小,DNN-AR网络的非线性程度越高. 可以看出,IS散度和β散度在训练一定周期后网络的非线性程度降低,而MAE、SD和KL散度的非线性程度随着周期的增加逐渐增加,最后趋于平稳. 在IS散度、KL散度、β散度和LSD下,精度转换和64位浮点数方法网络的非线性程度相近,但是在MAE和SD测度下则存在一定差距,这可能是精度误差对以MAE为目标训练的Toeplitz矩阵产生了影响.

4.4 含噪语音的测试

由于精度转换方法能有效训练DNN-AR,且性能与64位浮点数的方法接近,因此在含噪语音训练和测试中,不再采用64位浮点数进行训练和测试. 在傅里叶频点足够的情况下,LDR与DAP模型结果相近,因此,含噪语音的测试比较对象以LDR为主.

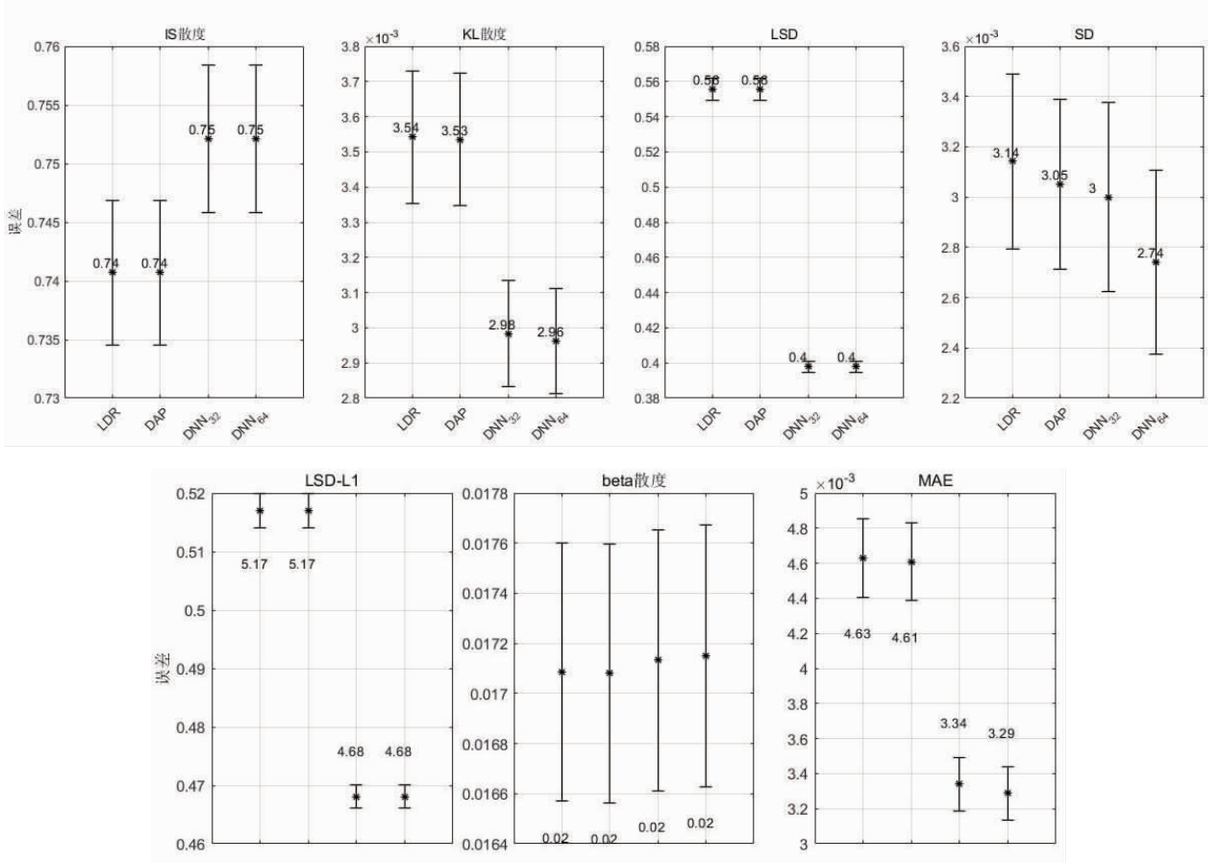
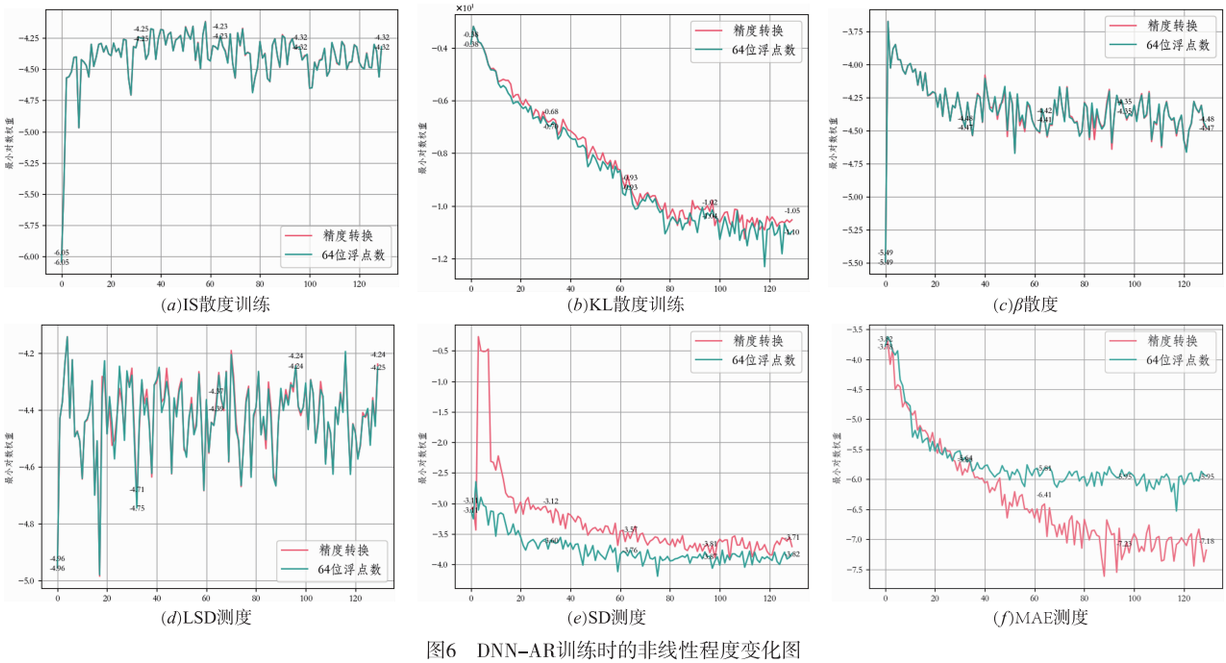


图5　DNN-AR，LDR和DAP在多种测度下的均值和95%置信区间结果图



(*a*)IS散度训练　(*b*)KL散度训练　(*c*)*β*散度

(*d*)LSD测度　(*e*)SD测度　(*f*)MAE测度

图6　DNN-AR训练时的非线性程度变化图

如图7所示,本文GABS-DNN模型的比较对象包括:估计纯净语音AR系数的LDR和DNN-AR方法以及估计含噪语音AR系数的DNN-AR模型(DNN)和SE + DNN模型. 其中,在MAE测度训练时,由于LDR的非线性和

所用优化方法 Radam 的影响,采用 2e-4 的学习率训练 DNN-AR 模型存在不稳定现象,由此采用 5e-5 的学习率进行训练.

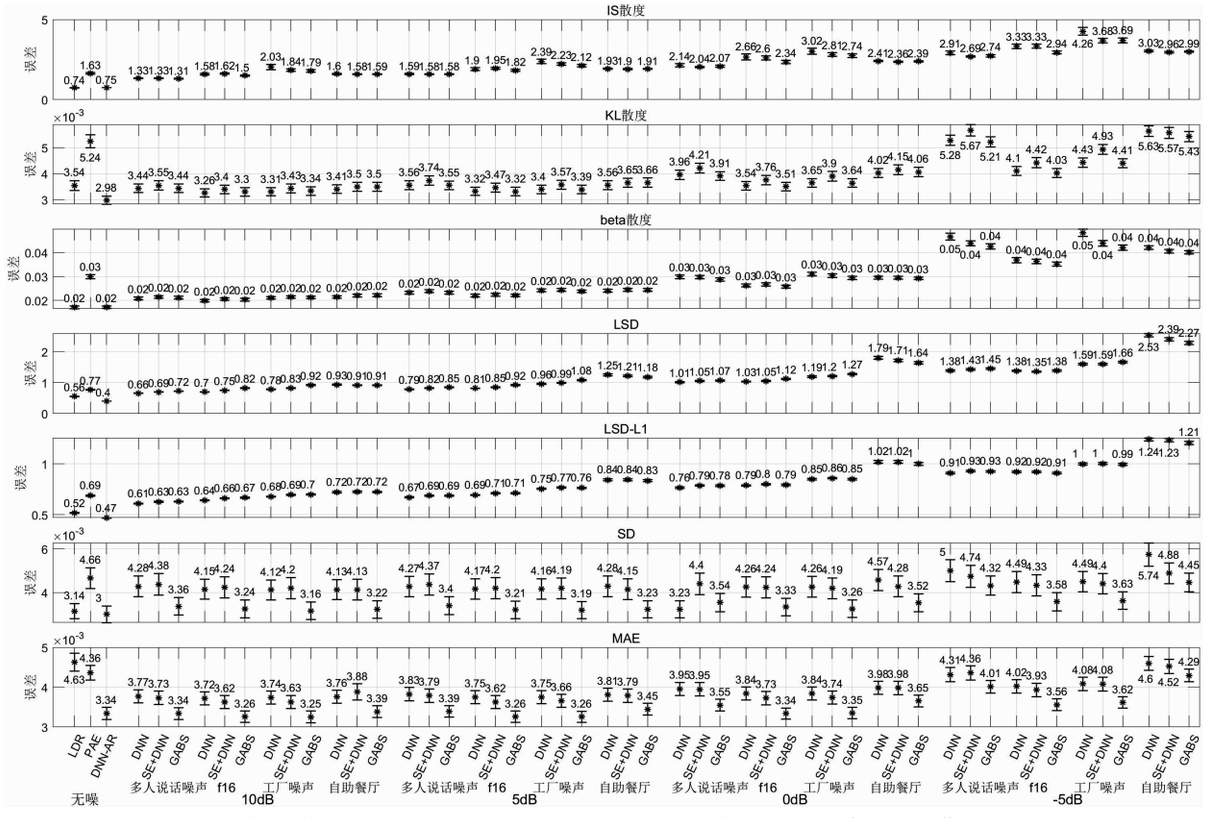


图7 四种噪声环境下DNN-AR, PAE和LDR在多种测度下对已知类别语音的均值和95%置信区间结果图

Babble、F16、Factory1、Cafeteria 干扰下的 AR 系数估计结果如图 7 所示. 在以 MAE 和 SD 测度为目标训练和测试时,GABS-DNN 的 AR 估计方法改进明显,尤其在 MAE 测度的情况下,5dB 和 10dB 的部分表现比纯净语音的 AR 估计方法更优,这可能是 GABS 策略一种语音参数估计的最优方法[28]. 在 IS 散度、beta 散度和 KL 散度测评中,三种方法较为接近,而在以 LSD 和 LSD-L1 为目标的训练测试中效果稍差于其他方法. DNN-AR 和 SE + DNN 模型在大部分测试中结果相近,然而在 IS 散度下 DNN-AR 的表现比 SE + DNN 差. 这可能是由于 DNN-AR 模型的参数量约为 SE + DNN 方法的一半.

未知类别的纯净语音及其在 Babble、F16、Factory1 和 Cafeteria 噪声干扰下的含噪语音的 AR 系数估计结果如图 8 所示. 在无噪环境下,对比图 8 与图 7 可知,除 KL 散度和 SD 测度外,DNN-AR 估计未知类别语音与已知类别语音的 AR 系数效果相近. 如图 7 所示,在有噪环境下的已知类别语音测试中,DNN、SE + DNN 与 GABS-DNN 对 AR 系数估计效果相近.

如图 8 所示,相比于估计已知类别语音的 AR 系数的估计,DNN、SE + DNN 与 GABS-DNN 估计未知类别语音的 AR 系数效果稍差. 并且在 IS 散度和 beta 散度测试中,随着信噪比逐渐降低,上述三种网络估计 AR 系数的性能变化趋势对已知类别和未知类别语音不一致,这可能需要采用更复杂的语音增强结构和优化策略来解决. 在 10dB 的 Cafeteria 噪声对 SpeechLibri 语音干扰下,SE + DNN 在 MAE 下表现不稳定,其原因从图 6 可知,在以最小 MAE 为目标函数时,精度转换网络模型稳定性弱于 64 位浮点数模型和以其他测度为目标的网络模型,这可能使单独训练增强网络数据容易受到非线性的影响.

5 总结

本文在 DNN-AR 模型的基础上,研究 LDR 的稳定性条件,给出一种精度转换的方法,该方法在较好保证系统稳定性的前提下提高训练效率. 同时,针对复杂场景即在噪声干扰下的 AR 系数估计中,引入 GABS 理论,提出一种基于 GABS 和 DNN 的 AR 系数估计方法. 实验对比表明,精度转换方法能够有效提高系统训练的效率,而且基于 GABS 和 DNN 的 AR 系数估计方法在一些测度下的估计准确性能有明显改善,且以纯净语音信号训练的 DNN-AR 模型能有良好的泛化性.

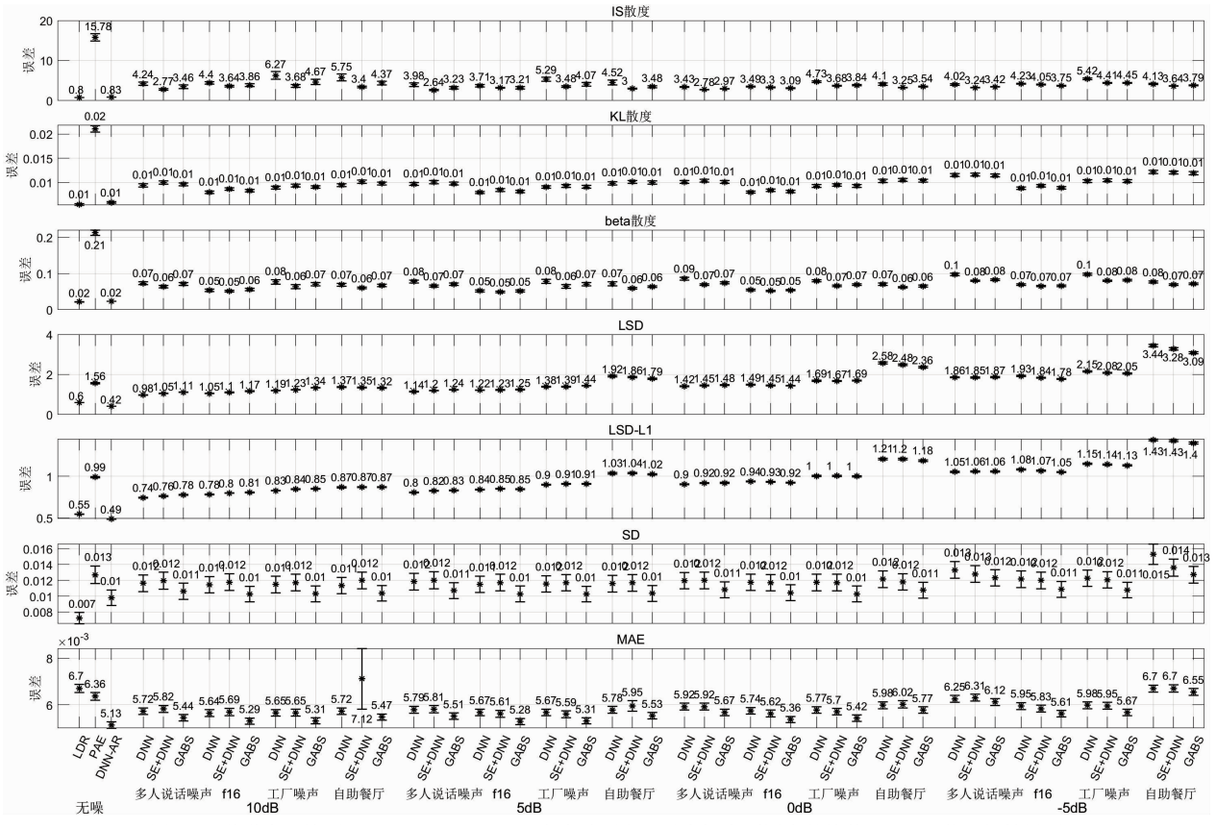


图8　四种噪声环境下DNN-AR, PAE和LDR在多种测度下对未知类别语音的均值和95%置信区间结果图

参考文献

[1] 舒新玲,周岱. 风速时程 AR 模型及其快速实现[J]. 空间结构,2003,(04):27 – 32 +46.
SHU X-L,ZHOU D. AR model of wind speed time series and its rapid implementation [J]. Spatial Structures,2003,(04): 27 – 32 +46. (in Chinese)

[2] 马秉伟,刘会金,周莉,崔福鑫. 一种基于自回归模型的间谐波谱估计的改进算法[J]. 中国电机工程学报,2005,(15):79 – 83.
MA B-W,LIU H-J,ZHOU L,CUI F-X. An improved algorithm of interharmonic spectral estimation based on AR model[J]. Proceedings of the CSEE,2005,(15):79 – 83. (in Chinese)

[3] 李星秀,韦志辉. 基于局部自回归模型的压缩感知视频图像递归重建算法[J]. 电子学报,2012,40(9):1795 – 1800.
LI Xing-xiu,WEI Zhi-hu. Compressed sensing video images recursive reconstruction algorithm based on local autoregressive model[J]. Acta Electronica Sinica,2012,40(9):1795 – 1800. (in Chinese)

[4] 吴桐雨,王健. 中国物流业、经济增长与技术创新——基于 2002 ~ 2017 年向量自回归模型的实证研究[J]. 工业技术经济,2019,38(03):116 – 122.

[5] 陈辉,张博霞. 自回归预测多级矢量量化线谱频率编码技术[J]. 西安科技大学学报,2017,37(05):736 – 741.
CHEN H,ZHANG B-X. Technology of multi-stage vector quantization with autoregressive prediction for linear spectrum frequency[J]. Journal of Xi'an University of Science and Technology,2017,37(05):736 – 741. (in Chinese)

[6] SCHROEDER M,ATAL B S. Code-excited linear prediction (CELP): High-quality speech at very low bit rates [A]. IEEE International Conference on Acoustics,Speech, and Signal Processing[C]. USA:IEEE,1985. 937 – 940.

[7] GIACOBELLO D,CHRISTENSEN M G,MURTHI M N, et al. Sparse linear prediction and its applications to speech processing[J]. IEEE Transactions on Audio, Speech, and Language Processing,2012,20(5):1644 – 1657.

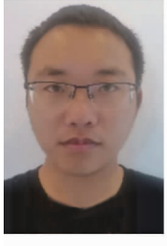
[8] 刘敬伟,王作英,肖熙. 基于自回归模型的加性噪声环境稳健语音识别[J]. 清华大学学报(自然科学版),2006,(01):50 – 53.
LIU J-W, WANG Z-Y, XIAO X. Autoregressive model-based robust speech recognition in additive noise environment [J]. Journal of Tsinghua University (Science and Technology),2006,(01):50 – 53. (in Chinese)

[9] ROE D. Speech recognition with a noise-adapting codebook [A]. IEEE International Conference on Acoustics,Speech,

and Signal Processing[C]. USA:IEEE,1987. 1139 – 1142.
[10] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing,1985,33(2):443 – 445.
[11] CUI Z H,BAO C C. Linear prediction-based part-defined auto-encoder used for speech enhancement[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA:IEEE,2019. 6880 – 6884.
[12] 何玉文,鲍长春,夏丙寅. 基于 AR-HMM 在线能量调整的语音增强方法[J]. 电子学报,2014,42(10):1991 – 1997.
HE Yu-wen, BAO Chang-chun, XIA Bing-yin. Online energy adjustment using AR-HMM for speech enhancement [J]. Acta Electronica Sinica,2014,42(10):1991 – 1997. (in Chinese)
[13] 孟宪波,鲍长春. 基于最小控制 GARCH 模型的噪声估计算法[J]. 电子学报,2016,44(3):747 – 752.
MENG Xian-bo, BAO Cang-chun. Noise estimate algorithm based on minima controlled GARCH model[J]. Acta Electronica Sinica, 2016, 44(3): 747 – 752. (in Chinese)
[14] WALKER G T. On periodicity in series of related terms [J]. Proceedings of the Royal Society of London (Series A, Containing Papers of a Mathematical and Physical Character),1931,131(818):518 – 532.
[15] YULE G U. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers[J]. Philosophical Transactions of the Royal Society of London (Series A, Containing Papers of a Mathematical or Physical Character), 1927, 226(636 – 646):267 – 298.
[16] LEVINSON N. The Wiener (root mean square) error criterion in filter design and prediction[J]. Journal of Mathematics and Physics,1946,25(1 – 4):261 – 278.
[17] DURBIN J. The fitting of time-series models[J]. Revue de l'Institut International de Statistique,1960,28(3):233 – 244.
[18] SHI L M, JENSEN J R, CHRISTENSEN M G. Least 1-norm pole-zero modeling with sparse deconvolution for speech analysis[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA:IEEE,2017. 731 – 735.
[19] EL-JAROUDI A, MAKHOUL J. Discrete all-pole modeling[J]. IEEE Transactions on Signal Processing,1991,39(2):411 – 423.
[20] MURTHI M N, RAO B D. All-pole modeling of speech based on the minimum variance distortionless response spectrum[J]. IEEE Transactions on Speech and Audio Processing,2000,8(3):221 – 239.
[21] GRAY A, MARKEL J. Distance measures for speech processing[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing,1976,24(5):380 – 391.
[22] KRIZHEVSKY A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[A]. Advances in Neural Information Processing Systems[C]. USA:Curran Associates, Inc,2012. 1097 – 1105.
[23] JI Y, ZHU W P, CHAMPAGNE B. Recurrent neural network-based dictionary learning for compressive speech sensing[J]. Circuits, Systems, and Signal Processing, 2019,38(8):3616 – 3643.
[24] 袁文浩,胡少东,时云龙,等. 一种用于语音增强的卷积门控循环网络[J]. 电子学报,2020,48(7):1276 – 1283.
YUAN Wen-hao, HU Shao-dong, SHI Yun-long, et al. A convolutional gated recurrent network for speech enhancement[J]. Acta Electronica Sinica, 2020, 48(7): 1276 – 1283. (in Chinese)
[25] CUI Z H, BAO C C, Nielsen J K, et al. Autoregressive parameter estimation with DNN-based pre-processing[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA: IEEE, 2020. 6759 – 6763.
[26] KRISHNA H, WANG Y. The split Levinson algorithm is weakly stable[J]. SIAM Journal on Numerical Analysis, 1993,30(5):1498 – 1508.
[27] BUNCH J R. The weak and strong stability of algorithms in numerical linear algebra[J]. Linear Algebra and Its Applications,1987,88:49 – 66.
[28] KLEIJN W B, RAMACHANDRAN R P, KROON P. Generalized analysis-by-synthesis coding and its application to pitch prediction[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA:IEEE,1992. 337.
[29] MICHELSANTI D, TAN Z H, SIGURDSSON S, et al. On training targets and objective functions for deep-learning-based audio-visual speech enhancement[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA:IEEE,2019. 8077 – 8081.
[30] CYBENKO G. The numerical stability of the Levinson-Durbin algorithm for Toeplitz systems of equations[J]. SIAM Journal on Scientific and Statistical Computing, 1980,1(3):303 – 319.
[31] PLANCHEREL M, LEFFLER M. Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies[J]. Rendicontidel Circolo Matematico di Palermo (1884 – 1940),1910,30(1):289 – 335.
[32] GAROFOLO J S, LAMEL L F, FISHER W M, et al. DARPA TIMIT acoustic-phonetic continous speech cor-

pus CD-ROM NIST speech disc 1-1. 1[J]. STIN, 1993, 93:27403.
[33] PANAYOTOV V, GUOGUO C, DANIEL P, et al. Librispeech:An ASR corpus based on public domain audio books[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. USA:IEEE, 2015. 5206 - 5210.
[34] VARGA A, STEENEKEN H J M. Assessment for automatic speech recognition:II. NOISEX-92:A database and an experiment to study the effect of additive noise on speech recognition systems[J]. Speech Communication, 1993, 12(3):247 - 251.
[35] THIEMANN J, NOBUTAKA I, EMMANUEL V. The diverse environments multi-channel acoustic noise database (DEMAND):A database of multichannel environmental noise recordings[J]. The Journal of the Acoustical Society of America, 2013, 133(5):3591.
[36] BARKER J, MARXER R, VINCENT E, et al. The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines[A]. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)[C]. USA:IEEE, 2015. 504 - 511.
[37] KAY S M. Modern Spectral Estimation:Theory and Application[M]. India:Pearson Education India, 1988.
[38] XU Y, DU J, DAI L R, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 23(1):7 - 19.
[39] XU B, WANG N, CHEN T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv Preprint, 2015, arXiv:1505. 00853.
[40] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout:A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1):1929 - 1958.
[41] LIU L, JIANG H, HE P, et al. On the variance of the adaptive learning rate and beyond[J]. arXiv Preprint, 2019, arXiv:1908. 03265.
[42] PASZKE A, GROSS S, MASSA F, et al. Pytorch:An imperative style, high-performance deep learning library [A]. Conference on Neural Information Processing Systems (NIPS)[C]. Vancouver, Canada:NIPS, 2019. 8026 - 8037.

作者简介



崔子豪　男,1991 年生于云南昆明. 现为北京工业大学博士研究生. 主要研究方向为语音增强.
E-mail:cuizihao@ emails. bjut. edu. cn

鲍长春(通信作者)　男,1965 年生于内蒙古赤峰. 现为北京工业大学教授、博士生导师. 主要研究方向为语音与音频信号处理.
E-mail:chchbao@ bjut. edu. cn