文章编号:1004-9037(2014)02-0254-05

基于短时能量和最小相对均方误差准则的神经网络语音水印方法

郝 欢[1] 陈 亮[1] 张翼鹏[2]

(1.解放军理工大学通信工程学院,南京,210007; 2.南京炮兵学院作战实验中心,南京,211132)

摘要:针对传统最小均方误差(Least mean square error, LMS)和最小二乘准则(Recursive least squares, RLS)的神经网络语音水印的局限性,提出了基于短时能量和最小相对均方误差(Least relative mean square error, LRMS)准则的神经网络语音水印算法。首先在首帧语音中嵌入同步序列,然后求出每帧的短时能量并对大于设定阈值的语音帧进行小波变换,最后利用以 LRMS 准则构建的神经网络实现水印的嵌入和提取。通过合理设定短时能量阈值,实现了水印容量和鲁棒性的平衡,而采用 Levenberg-Marguardt(LM)算法迅速地让网络收敛。理论分析和实验结果表明,与文献[8]相比,本文提出的神经网络方案收敛速度更快,对于噪声、低通滤波、重采样和重量化等攻击有更强的鲁棒性,性能平均提高了 5%。

关键词:短时能量;最小相对均方误差;小波变换;Levenberg-Marguardt 算法

中图分类号:TN392 **文献标志码**:A

Neural Network Speech Watermarking Method Based on Short-Term Energy and Least Relative Mean Square Error Criterion

Hao Huan[1], *Chen Liang*[1], *Zhang Yipeng*[2]

(1. College of Communications Engineering, PLA University of Science & Technology, Nanjing, 210007, China; 2. Operational Experiment Center, Nanjing Artillery Academy, Nanjing, 211132, China)

Abstract: In order to overcome the weakness of least mean square error (LMS) and the recursive least squares(RLS), a new neural network speech watermarking method based on short-term energy and least relative mean square error(LRMS) is proposed. Firstly, a synchronization sequence is embedded into the first frame of the speech. In addition, the short-term energy of each frame is calculated and discrete wavelet transform(DWT) is performed for the speech frame larger than the threshold. Finally, the watermark is embedded and extracted via the trained LRMS based neural network. The balance of the watermarking capacity and robustness is achieved by setting a reasonable short-term energy threshold and the network converges fast by Levenberg-Marguardt(LM) algorithm. The theoretical analysis and the experimental results show that, compared with reference [8], the improved neural network scheme converges faster and gets better robustness against attacks such as additive noise, low-pass filtering, resampling, re-quantifying, et al. Moreover, the performance achieves 5% increase on average.

Key words: short-term energy; least relative mean square error; discrete wavelet transform; Levenberg-Marguardt(LM) algorithm

引 言

随着多媒体技术和互联网的飞速发展,数字音像制品在网络上广泛传播,版权保护问题越来越严峻。数字水印技术作为信息隐藏技术领域的重要分支[1-2],是在不影响原始载体感知质量的条件下,利用视觉或者听觉冗余向载体中嵌入具有特定信

基金项目:国家自然科学基金(61072042)资助项目。

收稿日期:2012-09-24;**修订日期**:2012-11-22

息的过程[3]。

目前,绝大多数的水印算法都是通过对宿主的时域或者变换域的数值进行修改来实现水印信息的嵌入[4-6]。BP 神经网络是目前应用较为广泛的一种前向型人工神经网络的结构形式,有很强的非线性拟合能力,具有大规模并行信息处理、自适应、自组织和实时学习的特点[7]。利用神经网络强大的泛化能力找出水印嵌入前后的潜在关系,能够有效对抗信道中的各种攻击。文献[8]提出了一种基于 LMS 准则的神经网络水印算法,以预测值与目标值之间的误差平方和最小为准则,神经网络训练时无法对较小值进行充分训练,容易受到信道中各种干扰而产生误码,因此应用范围受到一定的限制。

采用 LRMS 准则的神经网络以目标值和输出值之间的相对误差最小为收敛准则,输出值与目标值之间的相对偏差较小,十分适合于通过修改空域或者变换域系数实现水印嵌入的语音水印系统。然而,在清音或者能量较小的浊音信号中嵌入水印容易受到噪声或滤波攻击的干扰,攻击前后的相对变化超过神经网络的泛化能力产生误码。因而需要在宿主语音中找出那些能量较大的,适合水印嵌入的语音帧进行水印信息的嵌入。

1　短时能量

语音信号的能量随时间而变化,清音和浊音之间的能量差别相当显著[9]。通过计算语音的短时能量,可以方便地描述出语音的这种特征变化情况。短时能量常用于清浊音判决和语音识别,在信噪比较高的情况下,短时能量还可以作为区分有声和无声的依据[10]。定义短时能量为

$$E_n=\sum_{m=-\infty}^{\infty}[x(m)\times w(n-m)]^2=\sum_{m=n-N+1}^{n}[x(m)\times w(n-m)]^2 \tag{1}$$

式中:N 为窗长,即短时能量是一帧语音样点值的加权平方和。当采用矩形窗时,式(1)简化为

$$E_n=\sum_{m=n-N+1}^{n}x^2(m) \tag{2}$$

由式(2)可知,E_n 越大,该帧语音的平均幅度越大,受攻击特别是噪声攻击的影响越小。因而选择的语音帧能量阈值越大,水印的鲁棒性越强,但整个语音段中可以选择的语音帧越小,水印容量也越小,通过合理设置阈值可以获得水印容量和鲁棒性的平衡。对于一段测试语音,基于 LMS 准则的神经网络在不同短时能量阈值条件下的抗噪性能如图 1 所示。

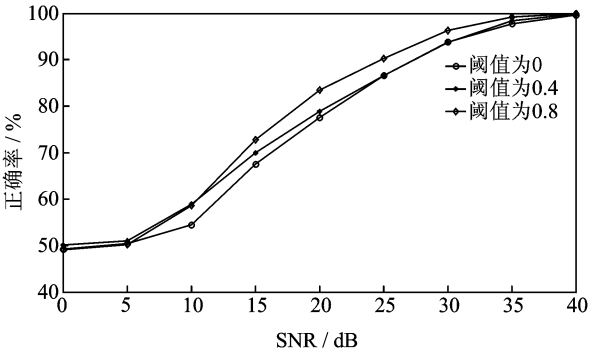

图 1　不同阈值条件下水印提取正确率

Fig. 1　BAR of different thresholds

由图 1 可知,随着阈值的增大水印的抗噪性能也越来越好,然而相应的水印容量也越来越小,分别为 327,227 和 212 b,因而需要综合考虑系统对水印容量和鲁棒性的要求来设定一个合适的短时能量阈值。

2　基于短时能量和最小相对均方误差准则的神经网络水印方法

2.1　系统原理

基于短时能量和 LRMS 准则的神经网络水印算法首先在第一帧语音中嵌入同步序列对水印信息进行定位,然后设定短时能量阈值找出适合水印嵌入的语音帧,利用语音的短时相关性和神经网络强大的非线性映射能力建立小波变换重要低频系数之间的隐含关系,通过修改重要低频系数实现水印的嵌入,利用数据之间的隐含关系完成水印信息的提取。整个水印嵌入和提取框图如图 2 所示。

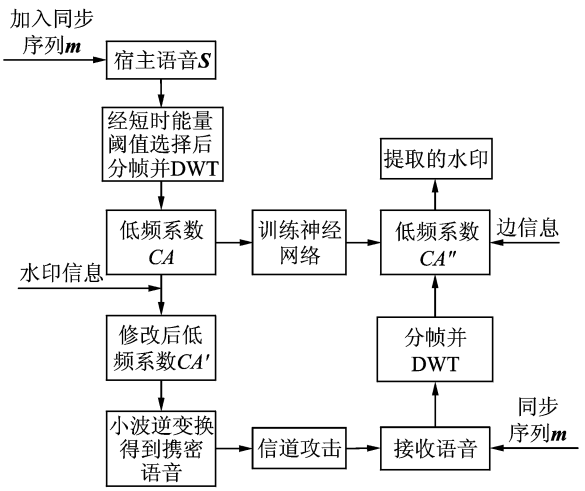

图 2　水印的嵌入和提取框图

Fig. 2　Block diagram of embedding and extracting watermark

2.2 神经网络结构和训练方法

人工神经网络有很多种模型，其中误差反向传播 BP 神经网络是目前应用最为广泛的网络之一，一般包括输入层、隐层和输出层。虽然隐层可以有多层，但是 Hornik 等的研究表明具有单隐层的神经网络就可以以任意精度逼近任一函数[11]。一个单隐层的 BP 神经网络结构模型如图 3 所示。

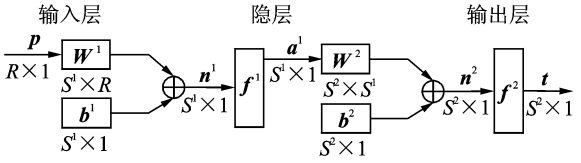

图 3 三层 BP 神经网络

Fig. 3 BP neural network with three layers

网络的训练采用 LRMS 收敛准则，对于每一个输入样本，计算网络输出 $\boldsymbol{t}$ 与目标输出 $\boldsymbol{T}$ 的相对偏差。通过 LM 算法[12]调整网络参数，使得网络输出与目标输出的相对均方误差最小化，即

$$\min F(\boldsymbol{w},\boldsymbol{b})=E\left\{\left|\frac{e_i(\boldsymbol{w},\boldsymbol{b})}{t_i}\right|^2\right\}=E\left\{\left|\frac{t_i-T_i}{T_i}\right|^2\right\} \tag{3}$$

相对于 LMS 准则，LRMS 准则的误差函数多除了一个目标输出项，计算复杂度只是多了一次除法操作。为了使网络对各种常见攻击有很好的鲁棒性，采用 LMS 准则的神经网络需要设置很小的目标收敛误差，而一个较大的目标相对收敛误差就可以使采用 LRMS 准则的神经网络有很好的性能。在没有明显增加计算复杂度的情况下，只需很少的迭代步数就达到设置的目标输出相对误差，训练时间更短。较小值在攻击特别是噪声攻击中容易产生误码，每个数据经过 LRMS 准则训练后，网络输出与目标输出的偏差分布较均匀，较小值得到充分训练，在各种常见攻击中鲁棒性也就更强。

2.3 水印嵌入

水印嵌入算法如下：

(1)对宿主语音 $\boldsymbol{S}$ 进行分帧，$\boldsymbol{S}=[\boldsymbol{S}_1,\boldsymbol{S}_2\cdots]$，为便于水印信息的定位，在第一帧语音中嵌入一段伪随机序列 $\boldsymbol{m}$ 作为同步序列，然后计算每帧的短时能量并找出大于设定阈值的语音，对相应位置 $\{M^i\}$ 的语音帧进行离散小波变换，$[CA^i,CD^i]=\mathrm{DWT}(\boldsymbol{S}_i)$，其中 CA^i 为每帧的低频分量。

(2)对每帧的 CA^i 分别按照绝对值的大小进行排序，找出绝对值较大的 10 个值 $\{CA_j^i\}=[CA_1^i\ CA_2^i\ CA_3^i\ CA_4^i\ CA_5^i\ CA_6^i\ CA_7^i\ CA_8^i\ CA_9^i\ CA_{10}^i]$ 及其相应的位置 $\{I_j^i\}$。构建神经网络，利用每帧中的 $[CA_2^i\ CA_3^i\ CA_4^i\ CA_5^i\ CA_7^i\ CA_8^i\ CA_9^i\ CA_{10}^i]$ 8 个值来预测另外一个值 CA_6^i 的大小。

(3)修改 CA_6^i 实现水印信息的嵌入，具体方法如式(4)所示

$$CA_6^{i\prime}=\begin{cases}CA_6^i\times(1+\alpha\times w_i) & w_i=1\\ CA_6^i & w_i=0\end{cases} \tag{4}$$

式中，α 为水印的嵌入强度，w_i 为第 i 比特的水印信息。

(4)利用修改后的低频分量 $CA_6^{i\prime}$ 和高频分量 CD^i 进行小波逆变换，$\boldsymbol{S}_i'=\mathrm{IDWT}(CA^{i\prime},CD^i)$，完成水印信息的嵌入。

2.4 水印提取

水印检测算法如下。

(1)对于接收端收到的语音，首先利用同步序列 $\boldsymbol{m}$ 定位出水印的嵌入位置，然后根据 $\{M^i\}$ 找出嵌入水印的语音帧。对每帧语音 $\boldsymbol{S}_i'$ 进行离散小波变换，$[CA^{i\prime\prime},CD^{i\prime\prime}]=\mathrm{DWT}(\boldsymbol{S}_i')$。

(2)利用边信息 $\{I_j^i\}$ 和每帧的低频分量 $CA^{i\prime\prime}$ 得到参与训练神经网络的绝对值较大的 9 个值 $\{CA_j^{i\prime\prime}\}=[CA_2^{i\prime\prime}\ CA_3^{i\prime\prime}\ CA_4^{i\prime\prime}\ CA_5^{i\prime\prime}\ CA_6^{i\prime\prime}\ CA_7^{i\prime\prime}\ CA_8^{i\prime\prime}\ CA_9^{i\prime\prime}\ CA_{10}^{i\prime\prime}]$，再通过训练好的神经网络用 $[CA_2^{i\prime\prime}\ CA_3^{i\prime\prime}\ CA_4^{i\prime\prime}\ CA_5^{i\prime\prime}\ CA_7^{i\prime\prime}\ CA_8^{i\prime\prime}\ CA_9^{i\prime\prime}\ CA_{10}^{i\prime\prime})]$8 个值来预测 $CA_6^{i\prime\prime}$。

(3)提取出水印，具体方法如式(5)所示

$$w_i'=\begin{cases}1 & \left|\dfrac{CA_6^{i\prime\prime}-CA_6^{i\prime\prime\prime}}{CA_6^{i\prime\prime\prime}}\right|>\alpha'\\ 0 & \text{其他}\end{cases} \tag{5}$$

式中：α' 为提取阈值，可根据信道条件进行相应的设置；$CA_6^{i\prime\prime\prime}$ 为预测值；w_i' 为提取得到的水印信息。

3 算法性能分析与测试

本文实验中采用标准语音库中 8 kHz，16 位量化，长度为 21 s 的国内外男女声各 5 段作为载体，每帧语音长度为 512，三级小波分解，小波基为“haar”，短时能量的阈值 $\theta=0.8$，单隐层神经网络，输入神经元个数为 8，隐层神经元个数为 12，输出神经元个数为 1，嵌入强度 $\alpha=0.3$，攻击噪声采用高斯白噪声，加噪实验为 50 次求平均值。

3.1 不可感知性测试

语音质量评价标准包括主观评价和客观评价两种，本文采用 ITU-TP. 862 语音质量评价标准(Perceptual evaluation of speech quality，PESQ)

计算出的 MOS 分和 SNR 来测试水印性能，SNR 由式(6)得到

$$\mathrm{SNR}=10\log\left[\frac{\sum_{n=1}^{Num}s(n)^2}{\sum_{n=1}^{Num}((s(n)-s'(n))^2}\right] \tag{6}$$

式中：Num 为语音长度，$s(n)$为原始语音，$s'(n)$为携密语音。一段汉语女声嵌入水印前后的波形图如图 4 所示。

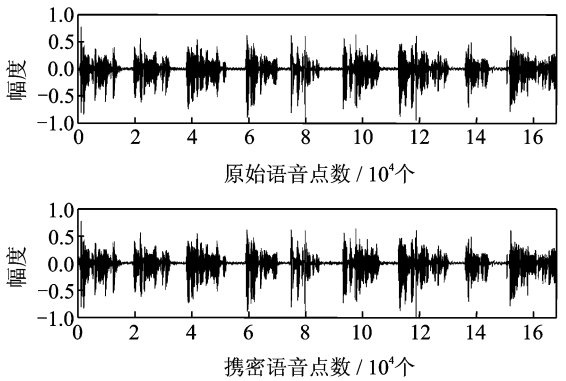

图 4　嵌入水印前后语音波形比较

Fig. 4　Waveform comparison before and after embedding watermark

由图 4 可以看出：水印嵌入前后几乎看不出波形失真。经过试听也察觉不到嵌入水印前后的语音变化，证明水印嵌入算法有良好的不可感知性，具体的测试结果如表 1 所示。

表 1　水印算法透明性测试

Table 1　Transparency test for watermark algorithm

语音类型	汉语女声	汉语男声	英语女声	英语男声
MOS 分	4.2	4.3	4.0	4.3
SNR	23.52	23.64	23.46	23.97

3.2　鲁棒性测试

为了测试算法对各种信道攻击的性能，对携密语音进行了加噪、低通滤波、重采样、重量化等操作。本文算法与文献[8]算法在引入短时能量阈值前后对不同噪声的平均性能如图 5 所示。

从图 5 可以看出采用 LRMS 准则的算法对强噪声有更好的鲁棒性，这是因为基于 LRMS 准则的神经网络以预测值和目标值的相对误差最小为收敛目标，较小目标值得到充分训练，从而有更强的拟合能力，对噪声的鲁棒性也就更强。短时能量阈值引入后，水印不会嵌入到能量较小的语音帧，减小了噪声的干扰，降低了误码率。文献[8]的神经网络水印算法在网络训练过程中偏好收敛于较大值而忽略较小值，即便引入短时能量阈值去掉能量较小的语音帧也只在高信噪比条件下才获得了较低的误码率。相比而言，本文算法对噪声有很好的鲁棒性。即便在 15 dB 情况下也获得了 93.6% 的正确率。在不同滤波器截止频率下算法平均性能如表 2 所示。

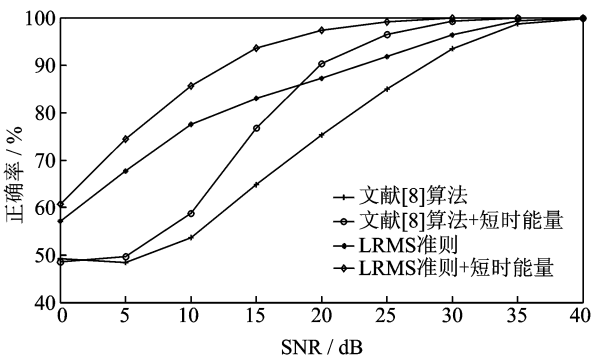

图 5　两种准则在引入短时能量阈值前后水印提取正确率

Fig. 5　BAR of two criterions before and after introducing short-time energy threshold

表 2　不同滤波器截止频率下算法性能比较

Table 2　Performance comparison for different filters cut off frequency

低通滤波器截止频率/kHz	1	1.5	2	2.5	3
文献[8]算法	0.915	0.974	0.983	0.995	1
LRMS 准则	0.993	0.996	0.996	1	1
文献[8]算法＋短时能量阈值	0.935	0.974	0.983	0.995	1
LRMS 准则＋短时能量阈值	0.995	0.995	0.995	1	1

从表 2 可以看出本文算法较文献[8]对滤波攻击有更好的性能。在信道条件较恶劣，低通滤波截止频率为 1 kHz 情况下仍然获得了很低的误码率。算法对重采样、重量化的平均性能如表 3 所示。

表 3　对重采样、重量化算法性能比较

Table 3　Performance comparison for resampling and requantization

攻击	8 k-16 k-8 k	16 bit-8 bit-16 bit
文献[8]算法	1	0.933
LRMS 准则	1	0.956
文献[8]算法＋短时能量阈值	1	0.985
LRMS 准则＋短时能量阈值	1	0.995

由表 3 可知,算法对于重采样都有很好的鲁棒性,对于重量化本文算法优于文献[8]算法。由于重量化过程中引入了量化噪声,那些较小值在重量化前后的相对变化较大,本文算法采用的是短时能量阈值和 LRMS 准则,确保了水印不会嵌入到能量较小的语音帧中,每一个目标值都得到了充分训练,较好的抵抗了这种干扰。

4 结束语

本文结合短时能量和神经网络,提出了一种基于短时能量和 LRMS 准则的神经网络水印方法。算法利用短时能量阈值找出适合嵌入水印的语音帧,以 LRMS 为收敛准则,采用 LM 训练算法的神经网络嵌入和提取水印。仿真结果表明,算法能够实现水印信息的盲提取,与文献[8]相比,本文算法在没有明显增加计算复杂度的情况下网络训练时间更短,对于噪声、低通滤波和重量化攻击,性能平均提高了 5%。

参考文献:

[1] 孙建国,门朝光,姚爱红,等. 基于量子纠错理论的数字水印技术[J]. 高技术通讯,2010,20(6):585-589.
Sun Jianguo, Men Chaoguang, Yao Aihong, et al. Digital watermarking based on quantum error correction coding[J]. Chinese High Technology Letters, 2010,20(6):585-589.

[2] 谢春辉,程义民,陈扬坤. 数字图像中扩频水印的盲提取方法[J]. 数据采集与处理,2011,26(1):26-30.
Xie Chunhui, Cheng Yimin, Chen Yangkun. Blind extraction method for spread-spectrum watermark on digital image[J]. Journal of Data Acquisition and Processing, 2011,26(1):26-30.

[3] Koz A, Alatan A A. Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2008, 18(3): 326-337.

[4] Zhang Yongmei, Ma Li, Xing Xiujuan. A multi-purpose video watermarking algorithm based on wavelet transform and image partition[C]// Second International Conference on Intelligent System Design and Engineering Application. [S. l.]: IEEE, 2012: 76-79.

[5] Mansouri A, Aznaveh A M, Torkamani-Azar F, et al. Low complexity video watermarking in H. 264 compressed domain[J]. Information Forensics and Security, IEEE Transactions on, 2010, 5(4): 649-657.

[6] ValizadehA, Wang Z J. An improved multiplicative spread spectrum embedding scheme for data hiding [J]. Information Forensics and Security, IEEE Transactions on, 2012,7(4):1127-1143.

[7] 余华,黄程韦,金赟,等. 基于粒子群优化神经网络的语音情感识别[J]. 数据采集与处理,2011,26(1):57-62.
Yu Hua, Huang Chengwei, Jin Bin, et al. Speech emotion recognition based on particle swarm optimizer neural network[J]. Journal of Data Acquisition and Processing, 2011,26(1):57-62.

[8] Chen Liang, Hao Huan, Zheng Guohong. An audio watermarking of wavelet domain based on BP neural network[C]// Proceeding of the WCSP, International Conference on Wireless Communication and Signal Processing. Nanjing, China: [s. n.], 2011:1-5.

[9] 张雄伟,陈亮,杨吉斌. 现代语音处理技术及应用[M]. 北京:机械工业出版社,2003:8-9.
Zhang Xiongwei, Chen Liang, Yang jibin. Modern audio processing technology and application[M]. Beijing: China Machine Press, 2003:8-9.

[10] 胡光锐,韦晓东. 基于倒谱特征的带噪语音端点检测[J]. 电子学报,2000,28(10):1-3.
Hu Guangrui, Wei Xiaodong. Endpoint detection of noisy speech based on cepstrum[J]. Acta Electronica Sinica, 2000,28(10):1-3.

[11] Cybenko G. Approximation by superposition of a sigmoidal function[J]. Mathematics of Control, Signals, Systems, 1989,2(4):303-314.

[12] Hagan M T, Menhaj M B. Training feedforward networks with the marquardt algorithm[J]. Neural Networks, IEEE Transactions on, 1994,5(6):989-993.

作者简介:郝欢(1988-),男,硕士研究生,研究方向:语音与图像处理、音频水印,E-mail:263624712@qq.com;陈亮(1974-),男,教授,研究方向:多媒体信息处理,网络信息处理,信息隐藏技术;张翼鹏(1988-),男,助教,研究方向:语音与图像处理,量子神经网络。