Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2603–2606
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Rapid and brief communication

Wavelet feature domain adaptive noise reduction using learning algorithm for text-independent speaker recognition

Shung-Yung Lung*

Department of Information and Telecommunications Engineering, Ming Chuan University, Taoyuan County, Taiwan

Received 22 February 2006; received in revised form 8 January 2007; accepted 22 January 2007

Abstract

In this paper, a type of thresholding method is developed for adaptive noise reduction. Here, we propose a new type thresholding method. Unlike the standard thresholding functions, the new thresholding functions are infinitely differentiable. Gradient-based adaptive learning algorithms are presented to seek the optimal solution for noise reduction. Furthermore, the learning algorithm can be used for any speaker data derived from discrete wavelet transform. It is demonstrated that 94% correct classification rates can be achieved by the use of the first 32 variation features in TALUNG database.

Keywords: Speaker recognition; Discrete wavelet transform; Adaptive threshold; Neural network

1. Introduction

Noise reduction is a traditional problem in signal processing as well as many applications in the real world. Conventional linear system adaptive filtering techniques have been widely used in adaptive noise reduction problems. However, because of the linearity of the operation, the filter cannot change the intrinsic property of the original noised signal, such as regularity, etc. Indeed, the linear filter is a kind of linear manipulation of the spectrum of a signal because the complex exponential function is the eigenfunction of a linear system. Therefore, it is difficult to suppress the noise and keep the speech signal using linear filters when the spectrum of a signal is somewhat wideband and nonstationary, which is the usual case.

Recently, wavelet thresholding methods proved to be powerful tools for denoising problems [1–4]. The main purpose of these methods is to estimate a wide class of functions in some smoothness spaces, such as Karhunen Loeve transform (KLT), from their corrupted versions. The main wavelet thresholding scheme are the hard-thresholding and soft-thresholding. This technique is effective because the energy of a function with some smoothness is often concentrated on few coefficients while the energy of noise is still spread in all coefficients in the wavelet domain. In this paper, a new type of thresholding for noise reduction in speaker recognition application is developed. The optimal solution of different method is investigated.

2. Discrete wavelet transform (DWT)

The wavelet transform (WT) has been used [5,6] with limited success because it has significantly reduced with a wavelet basis of interpolating scaling functions. The DWT [7,8] performs the recursive decomposition of the lower frequency band obtained by the previous decomposition in dyadic fashion, thus giving a left recursive binary tree. The two wavelet orthogonal bases generated from a parent node are defined as

$$\psi_{j+1}^{2p} = \sum_{n=-\infty}^{\infty} h[n]\psi_j^p(k - 2^j n), \tag{1}$$

$$\psi_{j+1}^{2p+1} = \sum_{n=-\infty}^{\infty} g[n]\psi_j^p(k - 2^j n), \tag{2}$$

* Tel.: +886 3 3507001x3952; fax: +886 3 3593879.
E-mail address: sylung@mcu.edu.tw.

0031-3203/$30.00
doi:10.1016/j.patcog.2007.01.028

where $h[n]$ is the lowpass filter, $g[n]$ is the highpass filter and $\psi[n]$ is the wavelet function [7]. Due to the rich literature on the WT [1,2,4–8], we can skip its introduction and leave the reader to the references for the background information, if necessary.

3. Optimal solution of adaptive thresholding

In this paper, a new type of thresholding method for noise reduction in speaker recognition application is developed. The optimal solution of thresholding method in a mean square error (MSE) sense is discussed. The results show that the presented thresholding method is very effective in noise reduction application.

Suppose we have a speaker data set Ω that is partitioned into a variable of interest, $y \in R$ and a group of covariates $x \in R^d$. Suppose x_i are the signal samples, n_i are Gaussian white noise samples with i.i.d. distribution $N(0, \sigma^2)$. We assume that the following relationship exists:

$$y_i = m(x_i) + n_i, \tag{3}$$

where the regression curve $m(\,)$ is taken as the conditional mean function. Then we can approximate $m(\,)$ by

$$\hat{m}(x_i) = \sum_{j=1}^{K} w_j \psi^j(x_i), \tag{4}$$

where $\psi(x)$ represents a wavelet function in the frame.

Let $\hat{x}_i$ be the noise reduction output using the standard nonlinear soft-thresholding method [2], i.e.,

$$\hat{x}_i = \eta(y_i, t) \quad \forall t \geqslant 0. \tag{5}$$

Define the risk function to be the MSE, i.e.,

$$J(t) = E\left\{\sum_i \Delta_i^2\right\}, \tag{6}$$

with $\Delta_i = \hat{x}_i - x_i$.

Now, rewrite Eq. (5) as

$$\hat{x}_i = \begin{cases} y_i + t, & y_i < -t, \\ 0, & |y_i| \leqslant t, \\ y_i - t, & y_i > t. \end{cases}$$

The conditional risk function is

$$\begin{aligned} J(t|x_i) &= E\left\{\sum_i \Delta_i^2 | x_i\right\} \\ &= \int_{t-x_i}^{\infty} (\xi - t)^2 p_n(\xi)\,\mathrm{d}\xi \\ &\quad + x_i^2 \int_{-t-x_i}^{t-x_i} (\xi - t)^2 p_n(\xi)\,\mathrm{d}\xi \\ &\quad + \int_{-\infty}^{-t-x_i} (\xi + t)^2 p_n(\xi)\,\mathrm{d}\xi, \end{aligned} \tag{7}$$

where $p_n(\xi)$ is the probability density function (pdf) of the Gaussian distribution n_i. Taking derivative with respect to t, we obtain

$$\begin{aligned} h(t|x_i) &= \frac{\partial J(t|x_i)}{\partial t} \\ &= \int_{-\infty}^{-t-x_i} (\xi + t) p_n(\xi)\,\mathrm{d}\xi \\ &\quad + \int_{t-x_i}^{\infty} (\xi - t) p_n(\xi)\,\mathrm{d}\xi. \end{aligned} \tag{8}$$

Denote the optimal solution $t^* = \arg\min J(t)$ then $h(t^*|x_i) = 0$ must be hold. Since n_i is Gaussian distribution then

$$\begin{aligned} \frac{\partial h(t|x_i)}{\partial t} &= \frac{1}{\sigma^2} \int_{-\infty}^{-x} \xi(\xi - t) p_n(\xi - t)\,\mathrm{d}\xi \\ &\quad + \frac{1}{\sigma^2} \int_{-x}^{\infty} \xi(\xi + t) p_n(\xi + t)\,\mathrm{d}\xi \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{-x} \xi^2 p_n(\xi - t)\,\mathrm{d}\xi \\ &\quad + \frac{1}{\sigma^2} \int_{-x}^{\infty} \xi^2 p_n(\xi + t)\,\mathrm{d}\xi - \frac{t}{\sigma^2} h(t|x_i). \end{aligned}$$

Therefore, if we let t_z denote any zero of $\sum_i h(t|x_i)$, i.e., $\sum_i h(t_z|x_i) = 0$, then

$$\left.\frac{\partial \sum_i h(t|x_i)}{\partial t}\right|_{t=t_z} = \frac{1}{\sigma^2} \sum_i \left[\int_{-\infty}^{-x_i} \xi^2 p_n(\xi - t_z)\,\mathrm{d}\xi + \int_{-x_i}^{\infty} \xi^2 p_n(\xi + t_z)\,\mathrm{d}\xi\right] > 0.$$

This means that all zeros of $\sum_i h(t^*|x_i)$ must be the minimum points of $t^* = t_z$ and function $\sum_i h(t|x_i)$ increases monotonically in the neighborhood of t_z. There is at most one minimum solution t^* for $J(t)$.

4. Adaptive learning wavelet threshold algorithm (ALWT)

Recently, hard-thresholding and soft-thresholding in the WT domain have been studied in statistical estimation problems and proved to have many good mathematical properties [1–4]. The basic idea of the wavelet thresholding method is that the energy of a signal is often concentrated on few coefficients while the energy of noise is spread among all coefficients in the wavelet domain.

In many thresholding problems, such as training of an artificial neural network (NN) [8–10], a very troublesome issue is that there may be more than one local optimum. This often makes it difficult to find the global optimal solution of the problem. However, in this section, we discuss the supervised learning algorithms of the NN. The ALWT has only global optimal solution.

We construct a type of thresholding NN to perform the thresholding in the transform domain to achieve noise reduction. The