The Use of Hybrid LVQ3 Neural Networks for Speaker Recognition Problems

Nabeel H.Kaghed **Eman S.Al_Shamery**
Computer Dept. Babylon University

Abstract

The main objective of this work is to develop the learning vector quantization (LVQ) neural network for the speaker recognition problems. The proposed methods consist of two main stages: the analysis stage and the recognition stage. The analysis stage was achieved by using wavelet transform coefficients, The neural networks was used in the recognition stage. The Four speaker recognition systems were differently constructed to appropriate the coefficients results in the stage of analysis which it was achieved by discrete wavelet packet transform (DWPT) coefficients. The first system was implemented by using the third development of LVQ (LVQ3). In other words it is pure neural network. The rest of the proposed systems were implemented by using different proposed activation functions of this network. The second system was implemented by combining the three sigmoidal functions in one function .The third system was implemented by a tansh function, and the last system was implemented using a wavenet network under which the second derivative of Haar wavelet function will be used. The work was applied on 10 persons (5 males and 5 females) which have converge ages. The results of the test showed that recognition rate averages for the five-recorded speaker statement which not previously trained.

Key words: discrete wavelet packet transform, sigmoid function, tansh function ,Haar function

الخلاصة

يهدف البحث إلى تطوير شبكة تعليم المتجه الكمي العصبية لحل مشاكل تمييز المتكلم أوتوماتيكيا .يتكون النظام المقترح من مرحلتين أساسية هما مرحلتي التحليل والتمييز .نفذت مرحلة التحليل باستخدام معاملات تحويل الموجية.استخدمت الشبكات العصبية في مرحلة التمييز حيث تم بناء أربعة أنظمة مختلفة في مرحلة التمييز لتلائم نتائج المعاملات في مرحلة التحليل التي نفذت باستخدام محول الموجية الجيبي المتقطع .نفذ النظام الأول باستخدام التطور الثالث لشبكة تعليم المتجه الكمي ،أي بتعبير آخر تطبيق الشبكة العصبية الصرفة .انحزت بقية الأنظمة المقترحة باستخدام دوال تنشيط مختلفة.نفذ النظام الثاني بتجميع الدوال السيجماوية الثلاث في دالة واحدة.نفذ النظام الثالث باستخدام دالة التانش أما النظام الرابع فقد نفذ باستخدام شبكة تحويل الموجية العصبية ووفقا للمشتقة الثانية لدالة هار .تم تطبيق النظام على عشرة أشخاص (٥ ذكور ٥ إناث) وبأعمار متقاربة. وضحت نتائج الاختبار معدلات نسب التمييز لخمسة تسجيلات للشخص لم يتم تدريبها مسبقا.

1.Introduction

One of the main problems in the speaker recognition systems is the non_stationary nature of the signal(Rabinar and et.al. ,1993), (Nejat ,1992).The wavelet transform is successfully applied to non_stationary signal for analysis, processing and provide alternative to the short time Fourier transform (Riol and et.al ,1991) ,(Graps ,1995) .The basis of a wavelet system is generated from simple scaling and translation (Walker ,1997),(Mersa ,1999) . The generating wavelet or mother wavelet, represented by $\psi(x)$ as shown in the following equation

$$\psi^{x}_{(a,b)} = 1/\sqrt{a}\psi((x-b)/a) \ldots\ldots(1)$$

Where

$\psi(x)$: the mother wavelet,

a: the scaling operator,

b: the translation operator.

Wavelets can approximate the time _frequency analysis using a mother wavelet which has square window in time space .The size of window can be almost freely

variable by two parameters (scale and translation). Thus wavelets can identify the localization of unknown signals at any level(Kobayashi and *et al.*,1994) .The wavelet coefficients are used as a feature extraction to the neural network, which is used as a pattern recognizer. It has been shown that neural network can realize any mapping (Nielsen, 1987) .The LVQ belongs to the competitive type of neural network which depend on the winner take all learning rule.The training depending on the competitive among neurons, one neuron is win in some time (Karayionnis, and *et al*, 1993), (Pandya , and *et al* ,1996)

Recently many of researches combine the wavelet and neural network in one neural network(Echauz and *et al*,1996), (Mathuar ,2001), (Echauz ,2002).the main objective of this work is to obtain wavelet neural network (WNN)as quickly and exactly as possible.

2.Wavelet transform

In order to work directly with the wavelet transform coefficients, we should present the relationship between expansion coefficients in terms of those at one scale higher .The relationship is especially practical by noting the fact that the original signal is usually unknown and only sampled version of the signal at a given resolution is available. For well_behaved scaling or wavelet functions, the samples of a discrete signal can approximate the highest achievable scaling coefficients .The scaling and wavelets coefficients, the following two relations relate coefficients at scale j to scaling coefficients at scale j+1

$$c_j(k) = \sum_m h(m-2k)c_{j+1}(m) \ldots.(2)$$

$$d_j(k) = \sum_m h'(m-2k)c_{j+1}(m) \ldots\ldots(3)$$

Where h(k)and h'(k) represent the wavelet and scaling filters respectively. Mathematically by assuming that cj(k)=x(k),(x(k) represent the signal),the discrete wavelet transform (DWT) coefficients are obtained by using the following set of equations

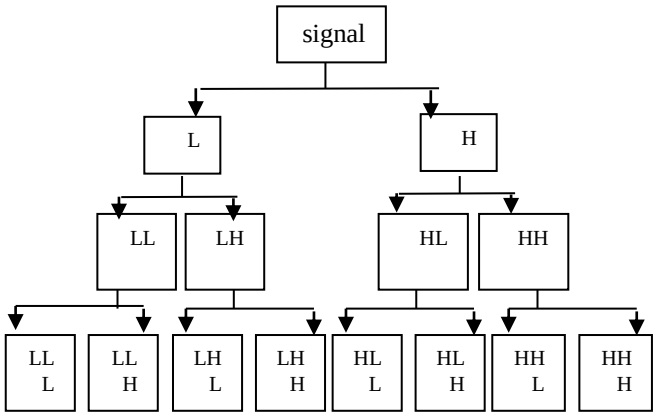
$$c_{j-1}(k) = \sum_m h(m-2k)c_j(m) \ldots.(4)$$

$$d_{j-1}(k) = \sum_m h'(m-2k)c_j(m) \ldots\ldots(5)$$

Where:
m:filter order,
h(m):low pass filter,
h'(m):high pass filter,
cj(m):input signal,
k:index of signal.

The calculations are continued until cj(k) and dj(k)are calculated. These coefficients called DWT of the original signal x(k).The pair of filters used in calculation of DWT are complementing low_pass and high_pass filters. The DWT divides the original signal bandwidth in logarithmic fashion. The first sequence dj-1(k) are the signal components from the upper half of the bandwidth. The lower half of the bandwidth is divided between the rest of DWT sequences. The remaining bandwidth is similarly between the rest of DWT sequences. These filters are refereed to as analysis filters in filter bank as well wavelet literatures. The wavelet analysis

work best in signal which composed of high frequency components of short duration ,in addition to low frequency component of long duration. The concept of changing resolution at different frequencies can be obtained by introducing what referred to as "wavelet packets" (Burrns , and et.al,1998),(Drygajlo ,2002). This system is suggested by Ronald Coifman to obtain the details coefficients in the high frequencies locations in addition to low frequencies locations (Burrns , and et.al,1998). The mechanism of this system is similar to DWT with single exception which is the splitting operation is taken by both the high and low frequencies locations. In the other word the filters bank is represented as a binary tree structure as shown in the figure (1) :



Fig(1)-three-stage wavelet decomposition,DWPTanalysis,tree.

Where the approximation bands describe the low frequencies while detail bands describe the high frequencies.
The number of coefficients are calculated according to the following equation:

$n = 2^{L+1} - 2$ ….(6)

n:the number of coefficients set,
L:the deep of tree

T=L*d…….(7)
T:the number of coefficients,
L:the deep of tree,
d:the length of mother wavelet.
The speech signal has been analyze to the third level by DWPT to obtain 14 vectors of coefficients, then we are taken the norm for each vector (according the equation (6)). The final result of analysis stage has fourteen coefficients for each speaker which they are used as input to neural network.

3.Neural networks

The learning vector quantization (LVQ) is a supervised learning extension the kohonen network method which used as pattern recognizer each neuron in output layer represent a class or category (several output maybe assigned to each class), the weight vectors are some times referred to as a reference or code book vector. The neuron with closest (Euclidean norm) weight vector is declared to be the winner. Kohonen [1990] is proposed several improvements to LVQ , named LVQ2, LVQ2.1,

LVQ3.All of these devolve around the manner in which the network is trained. Particularly, training will no longer be exclusively to the winning neuron. The training method rewards a winning neuron if it belong to the correct category by moving it towards the input vector. Conversely, if the wining neuron does not belong to the correct category, it is punished in that it is forced to move a way from the input. The principal idea behind these improvements is two neurons are modified rather than only a winning neuron in LVQ. In the other words the winner and runner-up are modified in the same time.

The general improvements have been summarized in the following algorithm (LVQ3).

Step0: Initialize reference vector (x1, x2…xn), initialize learning rate $\alpha(t)$.

Step1: While stopping condition is false do
Step 2-6.

Step2: For each training input vector **x** do
Step 3-4.

Step3: Find j so that ||**x-wj**|| is minimum.

Step4: Update **wj** as follow
if T=Cj then
wj(new)=**wj**(old)+ α [**x-wj**(old)].
else if T $\neq$ Cj then
if (min((dc1/dc2) and (dc2/dc1))>1- ε) and (max((dc1/dc2) (dc2/dc1))< 1- ε) then
wc1(new)=**wc1**(old)+ β (**x-wc1**(old))
wc2(new)=**wc2**(old)- β (**x-wc2**(old)).

Step5: Reduce learning rate

Step6: Test stopping condition:
The condition may specify a fixed number of iterations or the learning rate reaching is sufficiently small.

x: the input vector(training pattern vector),
T: the desired class (or category) for training vector,
C: the actual class of the output neuron,
||x-wj||: the Euclidean norm between weight vector and input vector,
dc: the distance from the input vector to the winner,
dr: the distance from the input vector to the runner-up,
j:index of winner neuron,
wc1:weight of the winner,
wc2:weight of the runner-up.

which impose further conditions when both winner and runner-up neurons belong to different classes, the winner is punished to move a way from the input and the runner-up reward to move a towards it. When both winner and a runner-up belong to the same class, both are trained.

The first implemented system the above algorithm applied in the recognition stage with the following parameter:

The length vector x=14(according to the number of the DWPT coefficients),
the number of classes=10
C,T=0,1,2,……..,9 (according to the number of speakers),
$\beta = 0.2\ \alpha$,

$\beta(t) = m1.\alpha(t)$,

$\varepsilon = 0.35$,

$\alpha(t) = \alpha(t) + \Delta\ \alpha(t)$,

the initial value of $\alpha(t)$ =0.01

$\Delta\ \alpha(t)$ =0.002,

the minimum of $\alpha(t)$ =0.0002.

The second proposed system the three sigmoidal functions are combined in one function as shown in equation (8), which it used as activation function to the LVQ3:

$$f(x) = s(x+2) - 2s(x) + s(x-2) \ldots\ldots(8)$$

The third proposed system the tansh function used as activation function according to the following equation

$$f(x) = e^{x} - e^{-x} / e^{x} + e^{-x} \ldots\ldots(9)$$

In the last system the second derivative of Haar wavelet function used as activation function according to following equation

$$f(x) = -xe^{-1/2x^2} \qquad .....(10)$$

In the last stage after the training algorithm were implemented in these systems, the test stage will be implemented. In this stage new untrained speech signals were used for the person under study, the system identify the speaker identity for the person with high success.

4.Conclution

1-The use of DWPT has strengthen the digital signal processing approach. So, in spite of the work has been done on a group of persons with converge ages and for the same speech statement under different records in different times ,but the recognition is still high.

2-The four methods have similar recognition rate (93%-95%).

3-The difference among these methods is the training period, under the same parameters and the initial weights. The number of epochs was as follows:

A-The pure LVQ3 is 35000 epoch.

B-The function which result from combining three sigmoidal functions is 6000 epoch.

C-The tansh function is 12000 epoch.

D-The second derivative of Haar wavelet function is 1800 epoch.

References

Burrns C., Gopinath R. and Guo H. ,1998, ”Introduction to wavelet and wavelet transforms”.

Drygajlo A., ”New fast wavelet packet transform algorithms for frame synchronized speech processing” ,andrzej.drygajlo@lst.de.epfl.ch.

Echauz J. and Vachtsevanos G. ,1996 ,”Elliptic and radial wavelet neural network" ,in proc(WAC'96),may 27-30 ,5.

Echauz J.,”strategies for fast training of wavelet neural networks”,, jechauz@exodo.upr.clu.edu.

Graps A. ,1995,”An introduction to wavelet”, IEEE computational science and engineering , 2 ,2.

Karayionnis, N. and Venetsanopoulos A., 1993,”Artificial NeuralNetwork, Learning Algorithm, Performance, Evaluation and Application”, Kluwer Academic publishers, London.

Kobayashi K. and Torioka T.,1994 ,”A wavelet neural network for function approximation and network optimization” ,proceeding of ANNIE ,94 ,4.

Mathuar A. ,2001,"Wavelet Self Organization Maps and Wavelet Neural Networks: A study,4,3,2394.

Mersa A. ,1999,”from Fourier transform to wavelet transform” ,spire ,lab UWN ,27 , October.

Nejat A. ,1992,”digital speech processing ,speech coding, synthesis and recognition”, Kuwer, Academic Publishers, London.

Nielsen H., 1987,”Kolomogorr’s mapping neural network existencetheorm”, proc. ICNN,11,3.

Pandya A., and Macy R. ,1996 ,”Pattern Recognition with Neural Network in C++”, CRC Press., Florida.

Rabinar L. and Schafer R. ,1993 ,”Digital processing of speech, signals”, prentice_ Hall, inc., Englewood, Cliffs, New Jersey.

Riol O. and Vetterli M. ,1991 ,”Wavelets and signal processing, IEEE signal processing magazine ,8 ,4 , October.

Walker J. ,1997,”Fourrier analysis and wavelet analysis ,44 ,6, ,june/july.