QUATORZIEME COLLOQUE GRETSI - JUAN-LES-PINS - DU 13 AU 16 SEPTEMBRE 1993

EEG SIGNAL COMPRESSION BASED ON ADPCM AND NEURAL NETWORK PREDICTORS

S.BARGIOTTI(1), F.BARTOLINI(1), V.CAPPELLINI(1), A.MECOCCI(2)

(1) Dipartimento di Ingegneria Elettronica, Università di Firenze, via S.Marta 3, 50139, Firenze, Italy, E-Mail: barto@cosimo.ing.unifi.it

(2) Dipartimento di Elettronica, Università di Pavia, via Abbiategrasso 209, 27100, Pavia, Italy

RÉSUMÉ

Dans l'analyse des maladies encéphaliques le monitorage avec l'électro-encéphalogramme de longue dureé est devenue très important. Ce monitorage demande une grande capacitée de memorization. Dans cette presentation est faite une comparaison parmi des méthodes de compression du signal EEG sans perte d'information. Tous les méthodes présentés sont fondés sur le DPCM, la vrai comparaison est pourtant faite entre differentes predicteurs sois classiques que neurales.

ABSTRACT

In the analysis of encephalic diseases the long term electroencephalographical monitoring has become very inportant. Long time monitoring requires a large amount of data storage. In this work a comparison among some schemes for lossless EEG signal compression is presented. All the schemes are based on a DPCM method, the actual comparison is then performed among different predictors both classical and neural.

1. INTRODUCTION

Clinical data record is one of the most important problems of medical archiving, particularly in electroencephalography (EEG). Just a few years ago, the accumulation of EEG data required a large amount of paper, that was easily deteriorable. An alternative to paper records was digital storage on magnetic media. Moreover this new type of medium permits to maintain the data un-altered in the long time. With computerized EEG evaluation, the problem of a large amount of digited EEG data storage is reproposed. It becomes necessary to reduce the data-rate produced by a PCM coder by means of suitable compression techniques. The solution proposed in this paper are based on a differential coding scheme: this choice permits to satisfy the lossless constraint imposed by neurophisiologists. The system for lossless EEG data compression is composed by a predictor and a quantizer. The predictor provides an estimate of the actual samples of the EEG signal by means of a suitable combination of the previous samples. The quantizer is chosen in a way to obtain a lossless compression technique. The new source of data to be stored is composed by the quantized prediction error signal samples (whose variance is less than that of the original signal) that are codified with an entropy coding scheme. In this way, both the types of redundancy present in EEG signal are removed. Differential pulse code modulation is in fact used to remove the redundancy due to the temporal correlation of the signal, and the entropy coder removes the second type of redundancy due to a un-uniform probability distribution of the signal. In this paper a predictor based on Artificial Neural Network (ANN) is also proposed; by means of a suitable training, the weights of the network are determined once for all; a fixed nonlinear predictor is thus obtained, that shows performance comparable to that of the adaptive predictors.

2. EEG SIGNAL

EEG is a graphical representation of the brain electrical activity, it can be detected by means of electrodes placed over the human scalp. Studies have shown that the scalp EEG derives from graded synaptic potentials generated by cells in the cerebral cortex, named neurons. Such electrical activity presents a sort of synchronization, pointed out by some correllation in the waveform recorded by EEG analysis. The EEG potential normally varies from 10 to 200 μV but may rise to 1 mV during epileptic seizures. Its energy is mainly concentrated in the band from 0 to 30 Hz. Furthemore, EEG waveforms are quasistationary, i.e., can be treated as stationary waveforms over a period of less than ten seconds. EEG is very important because permits to detect any pathological process of the central nervous system. It makes expecially easy the diagnosis of encephalic tumour and epilepsy.

3. LINEAR PREDICTOR

The signal estimate can be obtained by means of a fixed predictor, whose coefficients are calculated once for all applying the Yule-Walker algorithm [1] over a stationary tract of the signal. This procedure permits to identify the AR model parameters in order to describe the EEG signal. As the EEG signal is quasistationary (it can be considered stationary over a period of less than ten seconds), the fixed predictor provide good results only over intervals that are statistically similar to that one employed for the Yule-Walker algorithm. The predictors used in our experiments have order 0 and 5. Obviously, the best results are provided by the predictor having higher order (its prediction error have the smallest variance). The worst estimate is obtained on the epileptic spikes. Better performance is obtained by means of adaptive predictors; in this case coefficients are modified in a way to follow the fast variations in the signal statistic. The adaptive predictors implemented are based on the LMS algorithm and are usually employed in speech coding. A second-order predictor is updated with the Transversal/Signal-Driven (TS) [2] algorithm. The coefficients are updated as

$$a_i(n+1) = a_i(n) + \frac{\mu\, \hat{e}_n\, \hat{x}_{n-i}}{\hat{\sigma}_x^2(n)} \qquad i = 1,2$$

$$\hat{\sigma}_x^2(n) = (1-\lambda)\sum_{k=0}^{\infty} \lambda^k \hat{x}_{n-k}^2 = \lambda\, \hat{\sigma}_x^2(n-1) + (1-\lambda)\hat{x}_n^2 \qquad 0 < \lambda < 1 \tag{1}$$

where $\hat{\sigma}_x^2(n)$ is an unbiased estimate of the reconstructed signal power. It is obtained by low pass filtering the istantaneous square value of the signal (the more λ approaches to 1 the more increases the attenuation of rapid variations). This type of filtering produces usually great benefits on the algorithm convergence. Parameter μ is the step size of the updates. The restrictions to a second-order predictor is due to easily grant the stability of the reconstruction filter. To ensure this stability of the coefficients are constrained to lie in the region

$$|a_2| \le 1 - \varepsilon \qquad |a_1| \le 1 - a_2 - \varepsilon \tag{2}$$

where ϵ is a small number constant used to prevent the coefficients from reaching the stable/unstable boundary. To determine the adaptivity parameters, a procedure "try and repeat" has been employed. After this optimization procedure, the following parameters were chosen:

- $\mu = 0.01$;
- $\lambda = 0.5$;
- $\epsilon = 0.1$.

A suitable model for the EEG signal (in a stationary tract) is an AR model of order 7: a predictor of order 2 can not then provide an satisfactory estimate; the corresponding prediction error signal has not a negligible correlation.

In a variant of the previous algorithm, a certain number of the previous prediction error samples also contribute to prediction:

$$\hat{x}_n = \sum_{k=1}^{p} a_k(n)\, \hat{x}_{n-k} + \sum_{k=1}^{q} b_k(n)\, \hat{e}_{n-k} \tag{3}$$

This solution permits to exploit the residual correlation present in the prediction error signal. Updating for $a_k(n)$ is the same as in equation (1) and for $b_k(n)$ occur according to

$$b_i(n+1) = b_i(n) + \frac{\mu_b\, \hat{e}_n\, \hat{e}_{n-i}}{\hat{\sigma}_e^2(n)} \qquad i = 1,..6 \tag{4}$$

where

$$\hat{\sigma}_e^2(n) = (1-\lambda_b)\sum_{k=0}^{\infty} \lambda_b^2\, \hat{e}_{n-k}^2 = \lambda_b\, \hat{\sigma}_e^2(n-1) + (1-\lambda_b)\hat{e}_n^2 \qquad 0 < \lambda_b < 1 \tag{5}$$

The same stability constraint is imposed to $a_k(n)$; no check is needed for $b_k(n)$. The parameters are:

- $\mu = 0.011$; $\mu_b = 0.01$;
- $\lambda = 0.99$; $\lambda_b = 0.99$;
- $\epsilon = 0.1$.

Also the simplified sign version has been considered:

$$a_i(n+1) = a_i(n) + \mu_a\, sign\{\hat{e}_n\}\, sign\{\hat{x}_{n-i}\} \qquad i = 1,2$$
$$b_i(n+1) = b_i(n) + \mu_b\, sign\{\hat{e}_n\}\, sign\{\hat{e}_{n-i}\} \qquad i = 1,2,..,6 \tag{6}$$

where:

- $\mu_a = 0.004$; $\mu_b = 0.004$;
- $\epsilon = 0.1$.

An alternative to the previous algorithms is the Transversal/Residual-Driven (TR): to update the transversal coefficients only autocorrelations of the quantized residual are used. Coeffients updates for a second-order AR filter can be approximated by updates of an equivalent higher order MA filter. The resulting updates are:

$$a_1(n+1) = a_1(n) + \frac{\mu\, \hat{e}_n\, \hat{e}_{n-1}}{\hat{\sigma}_e^2(n)}$$
$$a_2(n+1) = a_2(n) + \frac{\mu}{\hat{\sigma}_e^2(n)}\left(\hat{e}_n\hat{e}_{n-2} - 2a_1(n)\hat{e}_n\hat{e}_{n-1}\right) \tag{7}$$
$$\hat{\sigma}_e^2(n) = (1-\lambda)\sum_{k=0}^{\infty} \lambda^k \hat{e}_{n-k}^2 = \lambda\, \hat{\sigma}_e^2(n-1) + (1-\lambda)\, \hat{e}_n^2 \qquad 0 < \lambda < 1$$

After optimization the parameters are set to:

- $\mu = 0.0625$;
- $\lambda = 0.9$;
- $\epsilon = 0.1$.

4. NEURAL PREDICTOR

Another proposal made in this paper is the use of a predictor based on an Artificial Neural Network (ANN). The ANN employed is the well-known back-propagation network [3]. Its advantages is the facility of use and the speed in convergence. To train the ANN characteristic segments or graphoelements of the EEG signal were considered (epileptic spikes, artifacts, alpha rhythms, etc.). The ANN used is a feedforward network with a suitable number of inputs, one hidden layer and one output. The experiments have shown that a network with more than one hidden layer do not improve the performance of the associated predictor. The number of inputs selected is a tradeoff between simplicity of the network and suitable description of the signal. The squashing function is the hyperbolic tangent. The addition of a trained bias to the hidden layer produced a predictor with good performance. The ANN weights are trained once for all. If the training is mainly focused over epileptic spikes, the resulting predictor performs better than other neural predictors obtained by means of a generalized training and than the previously described adaptive predictors on spikes, but worse on other graphoelements. It has been observed that such predictors give a lower estimate of the signal in correspondence of spikes. Such behavior has an opposite effect in other situations with waveforms that are easily to predict but characterized by a large value of the amplitude (for example an alpha rhythm). An other strategy is to generalize the training presenting to the network all the different graphoelements. The resulting predictor have so good performance on all the signal.

5. QUANTIZATION AND CODING

The second important component in the predictive coding scheme is the quantizer. By knowing the prediction error p.d.f. (it was approximated by a gaussian), we could choose a quantizer that match the signal statistics [4]. Doing so, the MSE is minimized, but the reversibility of the compression/decompression process is not granted. This quantizer do not permit to control the peak error. In our experiments the source data are integer numbers (already uniformly quantized), while the prediction errors results to be real numbers; in this case a uniform quantizer approximating real number to the nearest integer, permits to perform a lossless compression/decompression process. In fact, if Q indicates such operation, results

$$Q(e(n)) = Q(x(n)-\tilde{x}(n)) = x(n)-Q(\tilde{x}(n)) \qquad (8)$$

where x(n) is the integer input sample and $\tilde{x}(n)$ its estimate. This relation permits to pass to the equivalent coding scheme that is lossless.

The new data source is composed by the quantized prediction error samples having and approximately gaussian distribution. An entropy coding scheme can be used to remove the redundancy due to a nonuniform probability distribution. Entropy coding is a source coding technique that assigns shorter code words to highly probable symbols and longer code words to less probable symbols. When the symbols are independent (in our case the error samples are quite uncorrelated), it is possible to generate codes having average word length approximately equal to the entropy of the source. The procedure for generating such a code is known as Huffman coding [5]. As the EEG signal is quasistationary, the prediction error variance is time varying; by monitoring the source entropy it is possible, when his variations are greater than a prefixed value, to generate again the code. On average, such an adaptive code is better than fixed but needs many operations. Let us now describe the algoritm employed; the original signal samples assume values from -128 to 127: so, it is reasonable to suppose that the quantized prediction error signal samples assume values from -255 to 255. In order to obtain a lossless compression/decompression technique, an entropy coding composed by 511 codeword must be used, with evident computation complexity. A better solution is to consider only the symbols which are statistically predominant, i.e., those that fall in the intervall $(\mu-3\sigma,\mu+3\sigma)$, where μ is the mean value and σ is the standard deviation. This choice reduce the number of codeword to about 30 (being the standard deviation on average equal to 5). Symbols that fall outer that intervall are described by means of a unique codeword, followed by the value of the original signal samples (i.e. a byte). Experimental trial have shown that the deterioration of performance due to this coding scheme with respect to a pure Huffman scheme is very small. Entropy is monitored every 128 samples.

6. RESULTS

The various algorithm were compared by using 18 traces of two EEG recording sessions of 200 s each (the first is referring to a normal person and presents the typical alpha rythm behavior, the second is referring to a persons having epileptic seizures); the signal was sampled 128 samples/s. A comparison is done among the predictors according to the variance of the resulting prediction error signal; the tested predictors are:

the simple zero order predictor (each sample is predicted as equal to the previous one) (ZOP);

a 5 order fixed predictor (the coefficients are determined by the Yule-Walker algorithm over an alpha trace) (AR5);

an adaptive tranversal **signal driven** algorithm with 2 poles (TS);

an adaptive algorithm with 2 poles and 6 zeroes using the **gradient** algorithm to update coefficients (ZG);

an adaptive algorithm with 2 poles and 6 zeroes using the simple **sign** algorithm to update coefficients (ZS);

an adaptive tranversal **residual driven** algorithm with 2 poles (TS);

a neural network with 10 inputs and 5 hidden nodes trained over a selection of alpha traces (N5A);

a neural network with 10 inputs and 5 hidden nodes trained over a selection of epileptic traces (N5E);

a neural network with 10 inputs and 20 hidden nodes trained over a selection of alpha traces (N20A);

a neural network with 10 inputs and 20 hidden nodes trained over a selection of epileptic traces (N20E);

a neural network with 10 inputs and 20 hidden nodes trained over a selection of alpha and epileptic traces (N20);

Table I

Predictor	Variance	
	alpha	epylectic
ZOP	23.6	20.8
AR5	14.7	14.8
TS	12.9	12.6
ZG	11.4	13.4
ZS	11.9	16.3
TR	13.0	12.2
N5A	15.8	17.2
N5E	19.9	17.1
N20A	14.6	18.1
N20E	19.5	16.9
N20	11.8	13.4

In Table I for each of these ten predictors the variance of the prediction error is given. The value in the table is the mean of the variance obtained in each of the 18 traces, alpha and epylectic rithms are distinguished. The best performance is obtained by the adaptive predictors, expecially by the classical transversal signal driven (TS). The neural network trained with care over a suitable selection of graphoelements gives (NN20) results which are very near to those of the adaptive predictors; this is maybe due to its intrinsic generalization capabilities; the advantage of the neural network predictors is that it can be trained once for all and its coefficients don't need to be updated with additional computation.

By using the entropy coding scheme described in section 5 it is possible to compress the EEG signal without loss of information to about the 45%. By accepting a small peak error (say 1 or 2) not perceivable with the usual EEG display systems even better compression ratios can be achieved (40% and 36% respectively).

REFERENCES

[1] J.G.Proakis, Digital communications, McGraw Hill, 1989

[2] P.Yatrou, P.Mermelstein, Ensuring predictor tracking in ADPCM speech coders under noisy transmission conditions, JSAC, vol. 6, February 1988, pp. 249-261

[3] J.L.McClelland, D.E.Rumelhart, Parallel distributed processing, MIT Press, Cambridge (MA), 1986

[4] J.Max, Quantizing for minimum distortion, IRE Transactions of Information Theory, vol. 6, March 1960, pp. 7-12

[5] D.A.Huffman, A method for the construction of minimum redundancy codes, Proceedings of IRE, vol. 40, (September 1952), pp. 1098-1101