RESULTS ON ORIGINAL AND COMPRESSED HEARTBEATS CLASSIFICATION USING A MULTI-LAYER PERCEPTRON

Monica Fira, Liviu Goras
Faculty of Electronics and Telecommunications, ‘Gh. Asachi’ Technical University
mfira@scs.etc.tuiasi.ro

Abstract:
The aim of this communication is to compare several results on heartbeats classification using a Multi-Layer Perceptron (MLP) trained with databases consisting of original cardiac patterns as well as reconstructed ones from compressed signals obtained by means of the method presented in [1]. Moreover, the possibility of using Principal Components Analysis (PCA) applied both on original and compressed ECG signals have been investigated.

Introduction

Stocking and transmission of such large amounts of data are time and resource consuming operations and practice lead to conceiving solutions that enable handling a large amount of information with acceptable expenses (minimal). The solution to this problem, in the world of computers leads to developing methods of data compression.

Many algorithms for ECG compression have been proposed in the last thirty years; they have been mainly classified into three major categories: direct data compression, transformation methods, and parametric techniques.

One of the most difficult problems in ECG compression applications and reconstruction is defining the error criterion. In most ECG compression algorithms, the percentage root-mean-square difference (PRD) measure is employed. Other error measures such as the PRD with various normalizations, root mean square error (RMS) and signal to noise ratio (SNR) are used as well. However, the clinical acceptability of the reconstructed signal should always determine through visual inspection as well.

Materials and methods

The patterns used for classification were ECG signals from the MIT-BIH database [2] as well as reconstructed ones from compressed ECG by means of the method presented in [1]. The proposed method is a hybrid one i.e., a combination between signal processing (resembling the peak-picking compression techniques) and information transmission theory based techniques. It can be viewed as two cascaded blocks.

The first block extracts the essential information from the ECG signal. The signal reconstruction is based on the localization of the most significant part of the local minima and maxima (amplitude and location) together with some other points in order to decrease the error. The second block encodes the information from the first block, through delta coding followed by the LZW coding.

Compression - The preprocessing

The pre-processing stage consists of a filtering with a 6-degree Savitzky-Golay filter (SGF) using a 17 points constant window. SGF’s also called digital smoothing polynomial filters or least-squares smoothing filters are typically used to "smooth out" a noisy signal whose frequency span (without noise) is large. They perform much better than standard averaging FIR filters, which tend to filter out a significant portion of the signal's high frequency content along with the noise. Although SGF’s are more effective in pre-serving the high frequency components of the signal, they are less successful than standard averaging FIR filters in rejecting the noise. SGF’s are optimal in the sense that they minimize the least-squares error in fitting a polynomial to frames of noisy data.

The ECG skeleton

The main idea of this step is to extract the local minima and maxima from the filtered ECG signal and rounding the extracted values to the nearest integer. This is equivalent to a non-uniform sampling followed by a quantization. We will call the resulting discrete signal with non-uniformly spaced samples the signal skeleton. Knowing the location and the amplitude of the local extremes it is possible to reconstruct the ECG signal without loss of relevant information. This fact was confirmed by the error and the distortion values between the original ECG signal and the reconstructed ECG signal as well as by specialist physicians through visual inspection.

The hysteretic filtering

In order to improve the compression rate without significantly increasing the distortion between the original ECG signal and the reconstructed one from the skeleton samples having variation of the extreme values under a prescribed threshold have been discarded. This has been done with a hysteretic filtering in two steps.
In the first step the aim was improving the compression quality. The hysteretic filtering is characterized by a thresh-old denoted TH2. Samples for which the difference from previous ones is less than TH2 are discarded. The calculation of TH2 consists in the computation of a first threshold,

\[ TH1 = \sqrt{ST(A[n])} \]

where \( ST \) is the standard deviation of \( A[n] = \text{amp}[n] - \text{amp}[n-1] \) and \( \text{amp}[n] \) represent the amplitude of the \( n \)-th sample of the skeleton. The standard deviation of skeleton samples having amplitude variations \( A[n] = |\text{amp}[n] - \text{amp}[n-1]| \) less than \( TH1 \) is then calculated. The TH2 threshold is deter-mined using the formula

\[ TH2 = k \cdot ST(A[n]) \]

where \( k \) was initially chosen equal to 2.

The reconstruction errors based on the skeleton obtained in the above manner are rather high mainly for the QRS complex when adjacent skeleton samples are rather far one to the other.

The reconstruction error can be decreased by adding inter-mediate points to the skeleton resulted after the application of the TH2 threshold. The location of the intermediary points which will be added to the skeleton is determined through a third threshold, TH3 as follows: where the absolute value of the difference between two the successive amplitudes is higher than TH3, a sample of the original signal taken in the middle of the distance between the skeleton samples is added to the skeleton.

This TH3 threshold has been adopted according to the following formula:

\[ TH3 = \frac{1}{4N} \sum_{n=1}^{N} A[n] \]

The LZW coding

The LZW coding [8], [9] is a lossless “dictionary based” compression algorithm which looks for repetitive sequences of data and builds a dictionary based on them.

As a preliminary step, the obtained skeleton is delta coded both for amplitudes and distances. The LZW algorithm is then used to compress the results.

Patterns

The input patterns have been produced based on the R wave localization/detection presented in [3]. Following the R wave detection, segmentation has been realized as follows: a pattern begins from the middle of the RR interval between the previous and current heartbeat, and finishes at the middle of the next RR interval.

As the segmented output patterns had different dimensions, in order train of the MLP neural network each pattern was resampled to 100 samples. This figure has been chosen to decrease the dimension of the patterns while conserving the waveform. Even though the initial number of samples was about 300, through resampling to 100 using the above procedure there was practically no loss of information (visual inspection made by the specialist physician).

A MLP with tangent hyperbolical transfer function and a back-propagation algorithm with cross-validation have been used for training.

PCA (Principal component analysis)

PCA is also called the (discrete) Karhunen-Loève transform or the Hotelling transform.

Principal Components Analysis (PCA) is a multivariate procedure which rotates the data such that maximum variabilities are projected onto the axes. A set of correlated variables are transformed into a set of uncorrelated variables which are ordered by reducing variability. The uncorrelated variables are linear combinations of the original variables, and the last of these variables can be removed with minimum loss of real data.

The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as is possible. It computes a compact and optimal description of the data set.

The first principal component is the combination of variables that explains the greatest amount of variation. The second principal component defines the next largest amount of variation and is independent to the first principal component. There can be as many possible principal components as there are variables.[10], [11]

The Karhunen-Loève transform is therefore equivalent to finding the singular value decomposition of the data matrix \( X \),

\[ X = W \Sigma V^T \]

and then obtaining the reduced-space data matrix \( Y \) by projecting \( X \) down into the reduced space defined by only the first \( L \) singular vectors, \( W_L \):

\[ Y = W_L^T X = \sum_i V_i^T \]

The matrix \( W \) of singular vectors of \( X \) is equivalently also the matrix \( W \) of eigenvectors of the matrix of observed covariances \( C = X X^T \),

\[ XX^T = W \Sigma^2 W^T \]

By finding the eigenvalues and eigenvectors of the covariance matrix, the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset.[10]

Classification

The possibility of the heartbeats classification using the PCA applied on the patterns obtained from both original and compressed ECG has been investigated using a MLP trained and tested with the principal components corresponding to the most significant eigenvalues.

The MIT-BIH Arrhythmia database was used to assess the proposed classification algorithm and to compare it with other known classification methods. The 8 envisaged classification classes correspond to the most frequent types of heartbeats i.e.:

- atrial premature beat
- normal beat
- left bundle branch block beat
- right bundle branch block beat
- premature ventricular contraction
- fusion of ventricular and normal beat
• paced beat
• fusion of paced and normal beat.

Results

The MIT-BIH Arrhythmia database [2] was used to evaluate the proposed compression algorithm and compare it with other known compression methods. The ECG signals were digitized through sampling at 360 samples per second, quantized and encoded with 11 bits. For the tests the first 10000 samples out of 24 MIT-BIH records [2] have been used.

After the pre-processing stage (including the Savitzky-Golay filtering, extracting of all local maxima and minima, hysteretic filtering, followed by the delta coding) an average compression rate of 7.41:1 and an average PRD of 1.17% were obtained. After the LZW compression applied to the two previously obtained vectors, a final average compression rate of 18.27:1 is obtained (see Table 1). Since the LZW compression is a lossless compression, the PRD is conserved.

The mean value of the compression ratio (CR) for the 24 records analyzed is 18.2725, and the percentage root-mean-square difference (PRD) is 1.17% [1]. Using the training of a MLP (100-50-8 configuration) with original patterns and testing with patterns derived from the compressed signal a classification ratio of 82.5% has been obtained.

Table 1 - Average final results of the compression algorithm for 24 of records

<table>
<thead>
<tr>
<th>Average CR</th>
<th>Average PRDN</th>
<th>Average PRD</th>
<th>Average RMS</th>
<th>Average BPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.27:1</td>
<td>17.3742</td>
<td>1.1703</td>
<td>11.3507</td>
<td>260</td>
</tr>
</tbody>
</table>

As already mentioned, the performance evaluation of different coders should be doubled by visual inspection of the reconstructed signals by a cardiologist. Each compressed signal was scored by comparing its inspection of the reconstructed signal by a cardiologist. If the reference diagnosis was the same, the compressed signal was considered to be “correctly diagnosed”. If the diagnoses were different, or if the quality of the compressed ECG was too poor to make a diagnosis, the compressed signal was considered to be “incorrectly diagnosed”. In our cases the verdict of the cardiologist physician was “correctly diagnosed” for all records.

The average PRD or RMS and average CR values obtained using the proposed method of compression is compared to other methods in literature in Table 2.

Table 2 - PRD comparison of different algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average of errors (PRD or RMS)</th>
<th>Average of CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet</td>
<td>18.2 RMS</td>
<td>21.4:1</td>
</tr>
<tr>
<td>SPHIT</td>
<td>3.57 PRD</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>4.85 PRD</td>
<td>16:1</td>
</tr>
<tr>
<td></td>
<td>6.49 PRD</td>
<td>20:1</td>
</tr>
<tr>
<td>JPEG2000</td>
<td>2.19 PRD</td>
<td>12:1</td>
</tr>
<tr>
<td></td>
<td>2.74 PRD</td>
<td>16:1</td>
</tr>
<tr>
<td></td>
<td>3.26 PRD</td>
<td>20:1</td>
</tr>
<tr>
<td>Proposed</td>
<td>1.17 PRD</td>
<td>11.35 RMS</td>
</tr>
<tr>
<td></td>
<td>18:27:1</td>
<td></td>
</tr>
</tbody>
</table>

The results obtained from the classification with patterns obtained from original ECG depend on the MLP configuration and are presented in Table 3.

Table 3 - Results of pattern classification with various MLP configurations

<table>
<thead>
<tr>
<th>Pattern</th>
<th>MLP configuration</th>
<th>Accuracy of classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original pattern</td>
<td>100-50-8</td>
<td>91.54 %</td>
</tr>
<tr>
<td>Original pattern</td>
<td>100-30-8</td>
<td>85.70 %</td>
</tr>
<tr>
<td>Original pattern</td>
<td>100-10-8</td>
<td>85.31 %</td>
</tr>
<tr>
<td>Original pattern</td>
<td>100-50-30-8</td>
<td>92.25 %</td>
</tr>
</tbody>
</table>

Table 4 - Matrix of confusion for original pattern classification with MLP configuration 100-50-8 and classification=91.54%

Table 5 - Matrix of confusion for training of a MLP (100-50-8 configuration) with original patterns and testing with patterns derived from the compressed signal, classification ratio=82.5%

Table 6 - Results of the pattern classification with various MLP configurations
When PCA has been applied on patterns obtained from the compressed signal, it has been found that in order to reconstruct the patterns with good quality the first 30 principal components are necessary.

Discussions

The above results compare favourably to previously reported results. To the authors’ knowledge, the number of papers dealing with ECG classification into a higher number of classes is rather small, most of them treating the problem of simultaneous detection or of classification of few pathologies. Among them, De Chazal [4] reports a classification accuracy of 93.6% for 5 classes (with a total of 15 pathologies) while Prasad [5] and Osowski [6] using wavelet and SVM respectively, report 96%.

Conclusions

A skeleton-based adaptive hysteretic filtered ECG data compression technique has been proposed. It was tested and compared with different ECG compression algorithms.

The proposed method is rather fast and easy to implement and leads to high CR with a good reconstructive quality.

The mean value of the CR for the 24 records analyzed is 18.27:25 and the PRD=1.17%.

All clinical information is preserved after compression with the proposed algorithm. The algorithm was tested for the compression of normal, abnormal and affected by noise ECG signals extracted from the MIT-BIH database.

Visual inspection by the cardiologist physician of the original and reconstructed ECG signals, led to “correct diagnose” for all records.

Using patterns from compressed ECG a good classification (82%) and a good reconstruction quality have been obtained for a compression rate of 18.27:1.

It has been also found that in the case of PCA applied to patterns obtained from compressed ECG, the number of necessary principal components to be used in a reconstruction and, subsequently of a good classification should be not smaller than 30. The architecture of the MLP using the PCA method is smaller in size which suggests the application of this method in training a MLP for classification.

Acknowledgments

The Grant no. 124/2006 TD of CNCSIS has supported from the research for this paper.

References