

ECG Signal Analysis Using Temporary Dynamic Sequence Alignment

Análisis de Señales ECG empleando Alineación Secuencial Dinámica Temporal

Valentín Molina^{1*}, Gerardo Ceballos², Hermann Davila³

¹ Escuela Colombiana de Carreras Industriales, Bogotá, Colombia, amolinamecci@eccci.edu.co

² Universidad de los Andes, Mérida, Venezuela, ceballos@ula.ve

³ Escuela Colombiana de Carreras Industriales, Bogotá, Colombia, coordinación.biomédica@eccci.edu.co

Received: 09 May 2013

Accepted: 28 May 2013

Published: 30 July 2013

Abstract

This paper shows a feature extraction method for electrocardiographic signals (ECG) based on dynamic programming algorithms. Specifically, we applied local alignment technique for recognition of template in continuous ECG signals. First, we encoded the signal to characters based on the sign and magnitude of first derivative, then we applied local alignment algorithm to search for a complex PQRST template in target continuous ECG signal. Finally, we arrange the data for direct measurement of morphological features in all PQRST segment detected. To validate these algorithms, we contrasted them with conventional analysis by measuring QT segments in the Massachusetts Institute of Technology (MIT) data base. We obtained processing time at least 100 times lower than those obtained via conventional manual analysis and error rates in QT measurement below 5%. The automated massive analysis of ECG presented in this work is suitable for post-processing methods like data mining, classification, and assisted diagnosis of cardiac pathologies.

Keywords: ECG, Denoising, Dynamic Programming, Local Alignment, Template Classification.

Resumen

Este trabajo muestra un método de extracción de características de las señales electrocardíacas (ECG) basado en algoritmos de programación dinámica. Específicamente, se aplica la técnica de alineamiento local para el reconocimiento de una plantilla en la señal de ECG en tiempo continuo. En primer lugar, se codifica la señal a caracteres en base al signo y a la primera derivada, luego se aplica el algoritmo de alineamiento local para buscar una plantilla con el complejo PQRST en la señal ECG. Finalmente, se dispone de los datos para mediciones directas de las características morfológicas de todos los segmentos PQRST detectados. Para validar los algoritmos, se hace una comparación con el análisis convencional en mediciones del segmento QT sobre la base de datos del Massachusetts Institute of Technology (MIT). En los resultados se obtienen tiempos de procesamiento de al menos un centenar de veces más bajos que los obtenidos por el análisis convencional y las tasas de error en la medición del intervalo QT se encuentran por debajo del 5%. El análisis automático masivo del ECG que se presenta en este trabajo es adecuado para métodos de procesamiento orientados a la minería de datos, clasificación y diagnóstico asistido de patologías cardíacas.

Palabras Clave: Alineamiento Local, ECG, Filtrado, Programación dinámica, Clasificación de patrones

1. Introduction

Dynamic programming (DP) [1] is a technique for solving complex optimization problems that can be decomposed into relatively simpler sub-problems of the same nature. The solutions are stored in a table or matrix of dynamic programming, which allows

reusing sub-results to find the total solution, thus, avoiding repetitive and redundant calculations. The Temporal Dynamic Alignment *PD* uses algorithms to expand or contract a time series to fit another reference's series in order to minimize a distance function [2] [3] [4].

How to cite: Molina V., Ceballos G., Davila H., *ECG Signal Analysis Using Temporary Dynamic Sequence Alignment*, TECCIENCIA, Vol. 7 No. 14., 9-13, 2013, DOI: <http://dx.doi.org/10.18180/tecciencia.2013.14.1>

The work by Volkan Tuzcu and Selman Nas [5] presents a method to classify and identify waves in the ECG, using multi-resolution analysis by wavelet transform through the approximation of the signal's slope; dynamic programming is used only to compare the waves, while in the proposed methodology in this work it is also used to segment the ECG signal. Based on other studies [6]; wherein, the alignment is performed on data from capillary electrophoresis, conducted a series of innovations proposed algorithms to achieve a time alignment (local) of the ECG signal, indexing into an array data belonging to each PQRST waveform signal. This matrix contains the characteristic points for the wave's segmentation for analysis. Achieving this alignment of all ECG waves; allows faster quantitative analysis than the signal and indicates the result of a data set suitable for further processing of automatic classification of diseases. In order to eliminate the noise in the ECG signal, wavelet transform was used [7] [8] to threshold detail coefficients and use the second level of decomposition to lower the resolution of the data.

2. Problem

Electrocardiography signal analysis is significantly important to provide better diagnostics. Due to the large amount of data acquired, it becomes a rather tedious task for the specialists, leading to possible analysis errors. For this reason, it is important to implement robust and efficient data analysis methods toward possible criteria for a specific diagnosis [9] [10].

3. Theoretical framework

Alignment is a way to make an exact or approximate correspondence between character subsequence. It is mostly used to look for a pattern of characters, which are in the same order in the sequences involved [11]. The local alignment algorithm by Smith & Waterman (Gotoh optimization) [12], shows a variant for sequence aligning:

Let two sequences $\mathbf{x} = [x_1, x_2, \dots, x_n]$ and $\mathbf{y} = [y_1, y_2, \dots, y_m]$ a matrix is constructed to store the maximum of each alignment $M_{n+1, m+1}$ likewise, another array is created [12], which is indexed the direction from which comes each peak called $Ptr_{n+1, m+1}$. The local alignment algorithm, is based on the development of three stages [10]:

3.1 Initialization

The algorithm should be initiated under conditions equal to zero in each of the matrices.

$$P(i, j) = 0, \quad Q(i, j) = 0, \quad R(i, j) = 0 \quad (3.1)$$

$$M(i, j) = 0, \quad Ptr(i, j) = 0 \quad (3.2)$$

3.2 Filling the matrix M and Ptr

Each element of the dynamic programming matrix M is allocated from all neighboring maxima previously generated (3.3) (3.4).

$$M(i, j) = \max \begin{cases} P(i, j) \\ Q(i, j) \\ R(i, j) \end{cases} \quad (3.3)$$

$$P(i, j) = \max \begin{cases} P_{i-1, j-1} + Score(x_i, y_j) \\ Q_{i-1, j-1} + Score(x_i, y_j) \\ R_{i-1, j-1} + Score(x_i, y_j) \\ 0 \end{cases} \quad (3.4)$$

$$Q(i, j) = \max \begin{cases} P_{i, j-1} + d \\ Q_{i, j-1} + e \\ 0 \end{cases} \quad (3.5)$$

$$R(i, j) = \max \begin{cases} P_{i-1, j} + d \\ Q_{i-1, j} + e \\ 0 \end{cases} \quad (3.6)$$

In (3.4), (3.5), and (3.6) $Score(x_i, y_j)$, e and d refer to the scores get from character align x_i of the signal with the y_i from the other signal and penalties provided by the user in order to achieve better alignment. Matrices P , Q , and R , represent the maximum indexing analysis obtained from the *diagonal*, *left*, and *up* neighbors. Once each peak is found and indexed in M , Ptr must receive the provenance of each data (either *diagonal*, *left*, or *up*).

$$Ptr(i, j) = \begin{cases} diagonal & \text{if case} = 1 \\ left & \text{if case} = 2 \\ above & \text{if case} = 3 \\ none & \text{if } M(i, j) = 0 \end{cases} \quad (3.7)$$

3.3 Reconstruction of the optimal path

The best alignments will be reconstructed from positions (a,b) of M , being those in which the highest scores are generated for each alignment. Then by tracking Ptr , it will be easy find the optimum paths for each alignment.

4. Proposed Method

To implement local alignment algorithm for ECG, data analysis raises the following considerations:

A. Coding

It is proposed to encode each sample as the first ECG's derivative. Assuming that $Y(n)$ is the ECG signal:

$$Y'(n) = Y(n) - Y(n - 1) \quad (4.1)$$

Where $Y'(n)$ is the derivative of the ECG trace.

Then,

$$Encoding = \begin{cases} p & \text{if } Y'(n) \geq 0 \\ n & \text{if } Y'(n) < 0 \end{cases} \quad (4.2)$$

B. Alignment:

By using the algorithm developed by *Gotoh* [12], time alignment is done from a template or pattern with the ECG signal. The template is an ECG signal segment (QRST complex) properly selected from a normal or typical signal.

The robustness of the alignment is reflected in the ability to possess the algorithm to generate high or low scores in every moment of analysis, (4.3) shows the proposed similarity index, taking into account the similarity of characters (signs of slope) and the value of the slope.

$$Mscore(i, j) = score(x_i, y_i) + k * m_1(i) * m_2(j) \quad (4.3)$$

Where, $score(x_i, y_i)$ represents the score given to the similarity between the sign of the derivative of the sample x_i of the pattern and the sign of the derivative of the sample y_i of the encoded signal. $m_1(i)$ and $m_2(j)$ represent the value of the slope in the sample i of the pattern and in the sample j of the signal under analysis, respectively. k is a constant that controls the relationship between the two components of the index of similarity, by comparing the signs of slopes or their magnitudes.

C. Rearrangement of Alignments:

Upon reconstructing alignments that generate higher scores on M , it is necessary to reorganize the waves found to build the correspondence between all ECG waves.

This process is done by assigning each template sample a column number in the matrix and placing in each row a PQRST wave found where each sample is placed in the column corresponding to its pair in the template.

5. Results

The essential basis of the algorithm is the alignment of a pattern or template versus a full ECG signal; Figure 1 shows the dynamic programming matrix M representing its values as gray levels, the lighter gray is the magnitude of the matrix element represented.

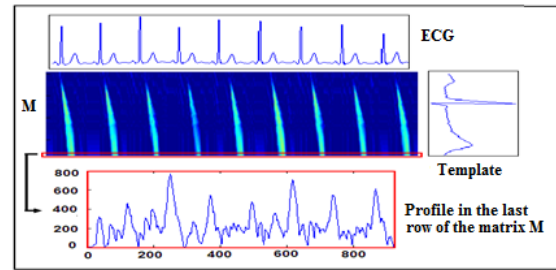


Figure 1. ECG Alignment versus Template

After filling the matrix M , it is necessary to seek the maximum get from the profile in the last row of the matrix. This consideration is essential, because its forced to make a full alignment between the pattern and each wave of the ECG signal.

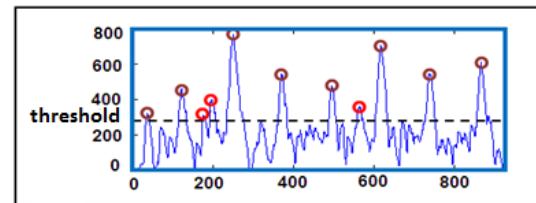


Figure 2. Location of the maximum

For the process of finding the maximum was chosen to take a threshold that allowed a total reconstruction of waves present in the signal, see Figure 2. In Figure 3 it is possible see the dynamic programming matrix M using the conventional similarity index and it is compared with the proposed method, adding the similarity among the magnitude of the slopes. The proposed methodology provides a better quality performance of the alignments.

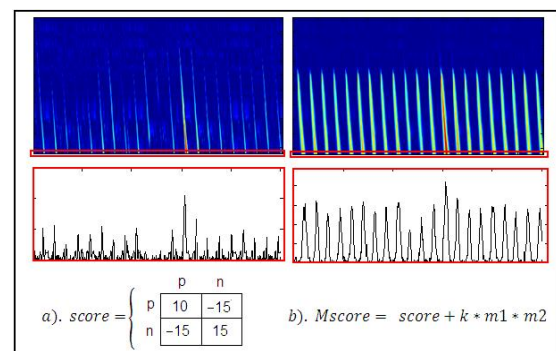


Figure 3. Profile in the last row of the matrix M

Once the maximum coordinates (a,b) have been located, the waves are reconstructed through Ptr , this is followed by rearrangement of the data considering the recesses inserted into each signal wave, because the waves in the reconstruction process do not show a direct correspondence among all the waves found Figure 4.

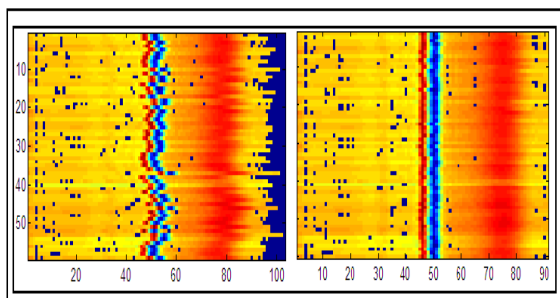


Figure 4. Alignment and rearrangement

After performing the alignment of all waves, points were chosen that characterize the beginning and end of the QT interval on the template, allowing placement of the columns corresponding to these critical points and recognizing them in all lined waves (Figure 5). To compare the results of each measurement method, the proposed algorithm was assessed by comparing with manual analysis make by a specialist of the same ECG signals.

Table 2. Manual method results

Signal	QT_v	QT_d	$Ef\%$	μQT	$\sigma^2 QT$	μQT_c	$\sigma^2 QT_c$	μFc	$\sigma^2 Fc$
e0611	79	79	100	0.4472	0.0119	0.4087	0.0117	45.7364	0.9389
e0119	99	99	100	0.4154	0.0134	0.4177	0.0154	61.115	4.7103
e0207	90	90	100	0.4353	0.012	0.4263	0.0131	56.3912	2.9593
sele0106	109	109	100	0.4638	0.0182	0.4707	0.0165	62.6475	5.2609
sel116	127	127	100	0.3572	0.015	0.3908	0.0166	78.5193	1.5644

Table 3. Proposed method results

Signal	QT_v	QT_d	$Ef\%$	μQT	$\sigma^2 QT$	μQT_c	$\sigma^2 QT_c$	μFc	$\sigma^2 Fc$
e0611	79	79	100	0.4405	0.005	0.4023	0.0052	45.7306	0.8812
e0119	99	97	97.9797	0.421	0.0131	0.4233	0.0162	61.1076	4.6935
e0207	90	89	98.89	0.4254	0.0129	0.4161	0.0148	56.3286	2.9286
sele0106	109	109	100	0.4427	0.0121	0.4486	0.0122	62.639	5.2369
sel116	127	127	100	0.3463	0.093	0.3784	0.0103	78.51	1.3866

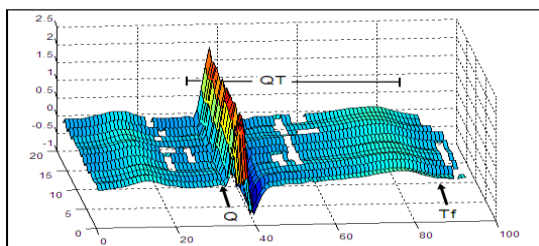


Figure 5. QT interval quantification

For Tables 2 and 3, QT_v represents the true QT intervals, QT_d are the QT intervals detected, $Ef\%$ is the efficiency provided; this is, defined as the ability to detect the same waves visually recognized by human expert ECG analysis. QT_c is the corrected QT interval [13], a sort of normalization with the instantaneous heart rate. Fc is the heart rate.

From the comparison of both tables, for instance, it is clear that automatic measurement of the QT segment differs by less than 5%, for each signal, from the manual measurement by an expert. The measurement of this parameter is very important for the quantification of heart disease, and in this case, demonstrates the potential use of the method proposed.

6. Conclusions

The proposed approach shows promising results to analyze the ECG signal. This methodology allows quickly and automatically segmenting all the PQRST waves of an electrocardiogram, it also allows any processing quantitative measurement as timeslots or wavelengths. Additionally, it is possible to get an ordered selection of data suitable to apply algorithms for training and classification tasks.

Processing times were obtained at least 100 times lower than those obtained by conventional manual analysis and QT measurement error rates were under 5%.

Table 1 demonstrates the large difference in execution time used by the two methodologies. This rapid analysis of massive ECGs obtained an ordering of the data suitable for automatic classification tasks, which represents the strength of our work.

Table 1. Computational analysis

Signal	N. Samples	N. Waves	Algorithm (min)	Manual (min)
e0611	12971	79	0.0436	11.61
e0119	12251	99	0.0434	10.54
e0207	12031	90	0.0436	10.36
sele0106	13134	109	0.04619	13.72
sel116	12125	127	0.0458	10.52

Tables 2 and 3 show the results obtained when measuring the QT interval of several signals from the MIT database, with the conventional manual method and the automatic method proposed.

There are several ideas to develop from this work, statistical evaluation of the method on a larger amount of data is necessary and involve the possibility of training to detect and classify diseases, although the similarity observed in an alignment of a template with standard features of an ECG signal can provide a first approximation to a possible diagnosis.

References

- [1] R. Bellman, *Dynamic Programming*, Princeton: Princeton University Press, 1957.
- [2] B. Huang, «ECG frame classification using dynamic time wrapping,» de *Electric and Computer Engineering, IEEE CCECE 2002*, 2002.
- [3] S. Chu, E. Keogh, D. Hart y M. Pazzani, «Interactive deepening dynamic time warping for time series,» de *The Second SIAM International Conference on Data Mining (SDM-02)*, 2002.
- [4] M. Kotas, «Application of dynamic time warping to ECG processing,» de *XI Conference "Medical Informatics & Technologies"*, 2006.
- [5] V. Tuzcu y S. Nas, «Dynamic time warping as a novel tool in pattern recognition of ECG changes in heart rhythm disturbances,» de *IEEE International Conference on Systems, Man and Cybernetics*, 2005.
- [6] G. A. Ceballos, J. L. Paredes y L. F. Hernandez, «Pattern recognition in capillary electrophoresis data using dynamic programming in the wavelet domain,» *Electrophoresis*, vol. 29, n° 13, pp. 2828-2840, 2008.
- [7] D. Cuesta Frau, D. Novák, J. Pérez Cortés, G. Andréu García, V. Eck, C. Sastre Mengual y M. Llorca Alcón, «Reducción del ruido en señales electrocardiográficas mediante la transformada Wavelet,» 2002.
- [8] M. Kaina, M. Fereniec y R. Maniewski, «Wavelet Denoising for Multi-lead High Resolution ECG Signals,» *Measurement Science Review*, vol. 7, n° 2, pp. 30-33, 2007.
- [9] C. San Martín y R. Carrillo, *Implementation of an isolated word recognizer speaker dependent*, Santiago de Chile: U.T.A Chile, 2004.
- [10] A. Manzo y J. Camarena, «A new and Efficient Alignment Technique for Evaluating the Similarity of Time Series,» de *Proceedings of the 17th International Congress on Computer*, Morelia, Mexico, 2001.
- [11] V. Ceballos, G. Grieken y Q. Paredes, «Pattern recognition in capillary electrophoresis using multiresolutional analysis and dynamic programming,» 2005.
- [12] O. Gotoh, «An improved algorithm for matching biological sequences,» *Journal of Molecular Biology*, vol. 162, n° 3, pp. 705-708, 1982.
- [13] S. Wong, G. Kervio, M. Altuve, F. Carré y G. Carrault, «Comparing Six QT Correction Methods in an Athlete Population,» 2012.