

Detección del complejo QRS mediante búsqueda aproximada de cadenas

Miguel E. Rubio Rincón, Martín O. Méndez, Alfonso Alba
Facultad de Ciencias, Universidad Autónoma de San Luis Potosí

Resumen— Los algoritmos de Búsqueda Aproximada (BA) son aquellos que tienen como objetivo buscar en una cadena de caracteres, las instancias más similares a una cadena patrón. Un área interesante para la aplicación de estos algoritmos es el análisis de señales biomédicas como el Electrocardiograma (ECG). Dentro del análisis de ECG una de las tareas más importantes es la detección de complejos QRS ya que mediante este se puede evaluar la variabilidad de la frecuencia cardíaca (VFC). En este trabajo se propone el uso de un algoritmo de consenso para la BA con el fin de detectar los complejos QRS. Este método consiste en realizar una codificación diferencial a la señal de ECG para así poder aplicar el algoritmo de BA. El método propuesto resultó ser una herramienta computacionalmente eficiente y de simple implementación para la detección de los complejos QRS.

Palabras clave— Búsqueda Aproximada, Electrocardiograma, Variabilidad de la frecuencia cardíaca.

I. INTRODUCCIÓN

El estudio de Electrocardiograma (ECG) es una de las herramientas principales en el área de la cardiología y es muy utilizado en casi todas las áreas de la medicina debido a que el ECG es una señal de fácil medición. El ECG contiene información sobre el funcionamiento del corazón y el sistema cardiovascular mediante el estudio de los potenciales bioeléctricos del miocardio de una forma no invasiva. Debido a esto, el desarrollo de métodos para analizar y detectar las ondas que componen el ECG es un área importante de investigación.

Una de las ondas más importantes y características del ECG es el complejo QRS, el cual tiene una forma singular que facilita su búsqueda, y mediante el análisis de las posiciones de estos, se puede evaluar la variabilidad de la frecuencia cardíaca (VFC). La VFC ha sido estudiada en forma exhaustiva en diferentes circunstancias como en anestesia y durante el sueño, debido a su estrecha relación con el sistema autónomo [5]. En las últimas décadas se han propuesto diferentes metodologías para la detección de los complejos QRS, que van desde la aplicación de filtros [6] hasta técnicas muy complejas basados en los puntos que componen el complejo QRS [4]. Sin embargo, es de utilidad desarrollar técnicas de bajo costo computacional y flexibles que puedan ser útiles para encontrar posibles deformaciones del complejo QRS que ayuden a detectar diferentes patologías cardíacas.

Una de las alternativas más interesantes son los algoritmos de búsqueda aproximada que suelen tener aplicaciones como búsqueda en línea, corrección de ortografía, detectores de plagio etc. [7]. Estos algoritmos

pueden proporcionar otro enfoque para el estudio de señales cardiovasculares, ya que mediante la codificación a símbolos de una señal, se pueden encontrar de forma simple complejos que en la señal cruda son difíciles de ubicar, además de tener la posibilidad de realizar la clasificación de diferentes tipos de formas de un mismo complejo.

En este trabajo se propone un método novedoso para la detección de complejos QRS mediante el uso de un algoritmo de búsqueda aproximada de cadenas [1].

II. METODOLOGÍA

El método presentado en este trabajo consta de 4 pasos fundamentales. El primer paso es la codificación diferencial; esta consiste en transformar la señal de ECG a una cadena de símbolos que cuente con propiedades que favorezcan la localización de los complejos QRS. El segundo paso consta de la adaptación y aplicación del algoritmo descrito en [1]. Debido a que este método puede presentar una cantidad considerable de falsos positivos; como tercer paso se propone un post-proceso para la reducción de falsos positivos. El último paso es una adaptación de los positivos encontrados con el fin de aproximarlos a las posiciones de los picos R.

A. Codificación.

Debido a que en este trabajo se utiliza un algoritmo de búsqueda aproximada de cadenas para encontrar los complejos QRS dentro de registros de ECG; es necesario realizar una codificación de la señal biológica a símbolos de un alfabeto discreto.

1) *Derivación de la señal*: En una señal de ECG los complejos QRS pueden variar considerablemente lo cual dificulta la detección en base a la amplitud, por lo tanto es necesario simplificar el patrón y la cadena de búsqueda. Para esto, se propone utilizar un filtro derivativo FIR de 5 puntos con un kernel $G = [-1 \ -2 \ 0 \ 2 \ 1]/8$ propuesto en [6], para calcular la primera derivada de la señal; de esta forma el patrón a buscar es mucho más simple ya que los complejos QRS cuentan con patrón de derivadas muy característico; además, este filtro también puede reducir ciertos tipos de ruido y reducir la interferencia debida a otras oscilaciones presentes en el ECG.

2) *Normalización*: Debido a que los registros de ECG varían de manera considerable entre pacientes, equipo y protocolo utilizado para adquirir las señales, una manera de hacer que el método propuesto sea invariante a las escalas en amplitud y tiempo consiste en normalizar la derivada del ECG, la cual se propone realizar de la siguiente manera:

Fig. 1. Ejemplo de la codificación de un registro de ECG.

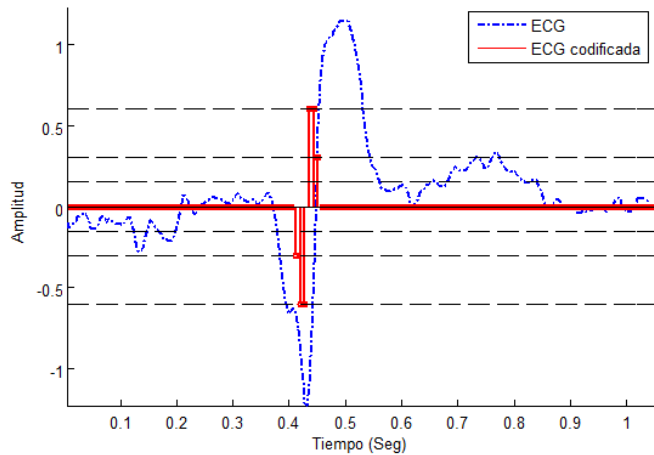


Fig. 1. Ejemplo de la codificación de un registro de ECG

Teniendo $D(n)$ como la derivada de la señal previamente calculada, la cual tiene una frecuencia de muestreo F_s , y una longitud M ; si se desea normalizar cada S segundos, la longitud de los intervalos a normalizar T quedara definida por:

$$T = (S)(F_s). \quad (1)$$

Los intervalos a normalizar I_n están definidos por:

$$I_n = T(N - 1), T(N - 1) + 1, \dots, T(N) \quad (2)$$

donde $N = 1, 2, \dots, \lfloor M/T \rfloor$.

Considerando que $\text{prom}_{10}(X)$ es una función que calcula el promedio de los 10 valores mayores (en valor absoluto) de X , la señal normalizada se calcula mediante:

$$Dn(I_n) = D(I_n) / \text{prom}_{10}(|D(I_n)|). \quad (3)$$

De esta forma cada intervalo de $D(n)$ se normaliza entre el promedio de los 10 números con mayor valor absoluto dentro del mismo intervalo.

3) *Discretización*: Una vez normalizada la derivada, se desea discretizar la señal resultante para representarla mediante un alfabeto finito. En este trabajo se optó por realizar una discretización a 7 niveles cuyos umbrales, seleccionados empíricamente, están dados por $[-0.6, -0.3, -0.15, 0, 0.15, 0.3, 0.6]$, los cuales definen rangos que se representan con distintos símbolos. De esta forma se obtiene la señal $Dc(n)$ la cual esta codificada en un alfabeto de tamaño igual a 7.

El número de niveles fue seleccionado debido a que el algoritmo no funciona muy bien con alfabetos pequeños (de 5 o menos símbolos), y al utilizar un alfabeto muy grande aumenta la complejidad de las instancias. Se optó por utilizar estos niveles ya que estos resultaron ser umbrales clave de discriminación en base a la comparación de varios complejos QRS de la base de datos [2] con las diferentes ondas de ECG.

B. Aplicación del Algoritmo de Búsqueda Aproximada.

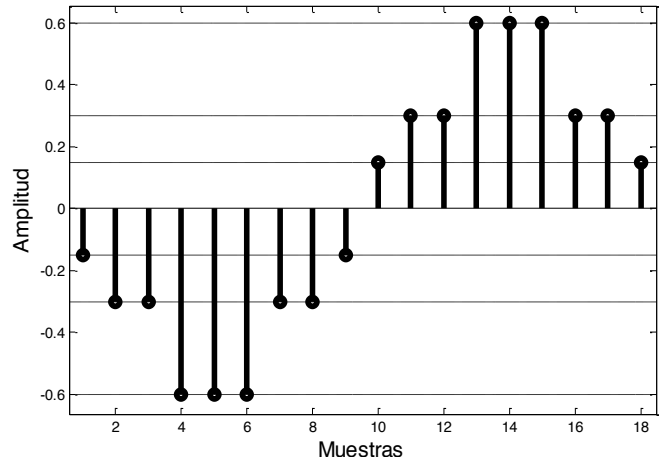


Fig. 2. Patrón genérico de la derivada codificada para búsqueda de picos-R.

1) *Diseño de la cadena patrón*: En la búsqueda aproximada de cadenas generalmente se conoce el patrón de búsqueda. En este caso esta tarea es un poco más complicada ya que al tratarse de señales biológicas no se conoce con exactitud este patrón, sin embargo se sabe que las proporciones de un complejo QRS son muy similares entre pacientes sanos y afortunadamente la codificación que se realiza a esta señal generaliza aún más el parecido entre estos complejos; por lo tanto, para realizar el modelado de este patrón se analizaron 3 registros de pacientes sanos obtenidos de la base de datos “Fantasia” de Physionet [2] con una frecuencia de muestreo de 250 Hz, y empíricamente se estableció el modelo mostrado en la figura 2 en base a la observación y los resultados obtenidos mediante pruebas con modelos de diferentes características. Una de las ventajas de utilizar búsqueda aproximada es que no solamente se buscan aquellas instancias que son idénticas al patrón sino que también busca aquellas que sean suficientemente parecidas a este, por lo tanto con un patrón genérico se pueden encontrar varios complejos QRS que sean diferentes entre sí, incluso de pacientes con alguna patología. De manera empírica hemos determinado que una distancia máxima de $k = 3$ entre la cadena patrón y la señal ECG produce los mejores resultados en la detección del complejo QRS.

2) *Ajuste de parámetros*: Una ventaja de la búsqueda aproximada de cadenas es que se pueden ajustar los parámetros de búsqueda para poder controlar el grado de similitud requerido para reportar una instancia aproximada de la cadena patrón. Una vez que el patrón fue seleccionado, el algoritmo [1] solamente utiliza 2 parámetros. El primer parámetro es la distancia de edición máxima k que determina el número máximo de alteraciones que puede sufrir una instancia encontrada con respecto a la cadena patrón. El algoritmo [1] es un algoritmo de consenso, debido a que cada símbolo de la cadena de búsqueda emite votos a las posiciones donde podría iniciar una instancia de la cadena patrón, por lo tanto es necesario utilizar también un umbral U

el cual es el número de votos mínimo para que una posición en la cadena de búsqueda sea considerada como el inicio de una instancia aproximada de la cadena patrón (i.e., un positivo).

Aparte del ajuste de parámetros, el algoritmo permite asignar un peso distinto a cada caracter de la cadena patrón, de manera que símbolos específicos puedan emitir más de un voto dependiendo de su importancia. En algunos registros debido a ruido o a patologías, se pueden presentar impulsos en la señal $Dc(n)$, sin embargo estos no suelen llegar a los valores máximos absolutos, por lo tanto una buena forma de darle prioridad a los picos R es asignando un peso de 3 para los caracteres extremos, aquellos que se refieren a los niveles -0.6 y 0.6, mientras que los demás tendrán un peso de 1.

El algoritmo regresa la lista de posiciones donde fueron detectados los positivos Lp y el vector $C(i)$ el cual contiene el número de votos para cada posición i en la cadena de búsqueda.

C. Reducción de falsos positivos.

El algoritmo mencionado anteriormente puede presentar muchos falsos positivos en algunas zonas del registro; sin embargo estos suelen estar cerca de un verdadero positivo. Gracias al ajuste realizado en el peso de la votación para los niveles extremos (-0.6 y 0.6), los verdaderos positivos tienden a presentar un valor mayor en $C(i)$, por lo cual se propone el siguiente método:

Considerando dm como la distancia mínima permitida entre cada pico R, se determina, para cada positivo Lp_j , si la distancia RR con respecto al positivo anterior es razonablemente grande mediante la condición:

$$|Lp_j - Lp_{j-1}| \geq dm. \quad (4)$$

En caso de no cumplirse esta condición, significa que uno de los dos positivos es un falso positivo, así que se elimina aquel con menor valor en $C(i)$.

D. Localización de picos R.

Debido a que el algoritmo reporta las posiciones donde podría empezar una cadena patrón, es necesario ajustar estas posiciones a los picos R correspondientes, ya que el error entre una posición Lp y el inicio del complejo QRS correspondiente puede tener una magnitud hasta de k . Además, se asume que el pico R se localiza a la mitad del patrón que representa el complejo QRS. Por lo tanto, se propone la siguiente ecuación como una forma sencilla de aproximar las posiciones de los picos R:

$$LP_j = Lp_j + \frac{M}{2} - 1 - k, \forall j \quad (5)$$

E. Complejidad computacional

Tanto el filtrado, como la normalización requieren un número fijo de operaciones por cada muestra de la señal ECG. Por otra parte, el algoritmo de búsqueda aproximada de cadenas tiene una complejidad de $O(nkw)$, donde n representa la longitud de la señal a analizar, k representa el número máximo de inserciones o supresiones entre la cadena patrón y las instancias aproximadas de la misma, y w es el número máximo de ocurrencias de un símbolo dado dentro de la cadena patrón. Para esta aplicación, tanto k como w toman valores fijos (3 y 4, respectivamente), por lo tanto, la complejidad del algoritmo propuesto es de orden $O(n)$.

III. RESULTADOS PRELIMINARES Y DISCUSIÓN

Se utilizó la base de datos [3] la cual consiste en 48 registros de ECG con una duración aproximada de 30 min, una frecuencia de muestreo de 360Hz y una resolución de 11 bits con un rango dinámico de 10mV. De los 48 registros se seleccionaron 20 para las pruebas preliminares, y se formaron dos grupos de 10 registros cada uno. El grupo A consta de registros que presentaron una forma normal de complejos QRS y muy poco ruido. El grupo B consta de registros que presentaron múltiples formas de arritmia ventricular y/o un exceso de ruido en la medición. Para cada registro se conoce el número total de latidos TB y la posición de cada complejo QRS.

Los parámetros utilizados en esta prueba fueron $k=3$, $U=17$ y $dm=100$ muestras (0.27 seg). Los valores k y U seleccionados fueron aquellos con los que se obtuvo un mejor resultado promedio mediante el análisis de todos los registros.

Con los parámetros anteriores se procesaron todos los registros de cada grupo y se calculó el número de verdaderos positivos TP , falsos positivos FP y falsos negativos FN . Posteriormente se calculó un porcentaje de error mediante la siguiente fórmula:

$$Error = \frac{FP+FN}{TB} \times 100\%. \quad (6)$$

En la Tabla 1 se observa una comparación de los porcentajes de error obtenidos mediante el método propuesto de Búsqueda Aproximada (BA) y los de Pan Tompkins (PT) y Wavelet (WT), cuyos resultados fueron obtenidos de [4].

Se puede observar que en los resultados generales el método propuesto BA tuvo un mayor porcentaje de error en todos los casos del grupo A y en la mayoría de los casos del grupo B; sin embargo, debido a que BA tiene un rango de búsqueda más amplio que PT, tuvo un mejor resultado promedio, con respecto a PT, en el Grupo B y en los resultados generales. Dado que el método de Pan Tompkins es considerado el estándar de oro, podemos decir que el método propuesto tiene un desempeño competitivo (similar a PT).

Por otro lado, el algoritmo WT es considerablemente mas eficiente; sin embargo, es también el que demanda mayores recursos computacionales. En [4] se menciona que el tiempo promedio requerido por el algoritmo WT para procesar 10

minutos de ECG es de 60 segundos, mientras que el algoritmo BA requiere únicamente de 5 segundos (en un CPU AMD Athlon a 1.8 GHz, similar al utilizado en [4]). En un CPU moderno, como el Intel Core i7 a 2.4 GHz utilizado en nuestras pruebas, el algoritmo BA requiere en promedio de 0.75 s para procesar 10 minutos de ECG.

Cabe mencionar que el BA utiliza un patrón de búsqueda para complejos QRS de pacientes sanos, y sin embargo da buenos resultados generales con el grupo B, esto es debido a que este método tiene la ventaja de poder aproximarse a varios tipos de instancias de búsqueda.

IV. CONCLUSIÓN

La búsqueda aproximada de cadenas resulta ser una herramienta eficiente y de simple implementación que muestra potencial para la detección de los complejos QRS. Resultados preliminares, basados en un algoritmo de consenso para la búsqueda aproximada de cadenas, muestran un desempeño razonablemente bueno, competitivo contra el estado del arte. Los resultados sugieren que el método propuesto tiene la capacidad de reconocer complejos QRS con morfologías atípicas a partir de un patrón simple, por lo que vale la pena continuar desarrollando y mejorando el

Grupo A			
Registro	Error BA	Error PT	Error WT
100	0.09%	0.00%	0.00%
101	0.16%	0.43%	0.00%
103	0.05%	0.00%	0.00%
111	0.19%	0.05%	0.09%
112	0.12%	0.04%	0.12%
114	0.53%	1.06%	0.16%
115	0.05%	0.00%	0.00%
116	0.91%	1.04%	0.04%
119	2.01%	0.05%	0.05%
121	0.54%	0.59%	0.16%
Total grupo A	0.46%	0.33%	0.06%
Grupo B			
Registro	Error BA	Error PT	Error WT
102	1.37%	0.00%	0.11%
105	1.56%	3.46%	1.09%
106	1.58%	0.35%	0.25%
107	0.70%	0.90%	0.00%
108	5.22%	12.54%	1.59%
109	0.24%	0.04%	0.00%
113	3.23%	0.00%	0.11%
118	0.04%	0.04%	0.04%
124	0.62%	0.00%	0.00%
200	0.92%	0.35%	0.04%
Total Grupo B	1.43%	1.77%	0.32%
TOTAL	0.95%	1.05%	0.19%

Tabla 1. En esta tabla se muestran los resultados obtenidos con los métodos de Búsqueda Aproximada (BA), Pan Tomkins (PT) y Wavelet (WT)

método propuesto. Como trabajo a futuro, algunas perspectivas de mejora consisten en utilizar un conjunto de patrones (en lugar de uno solo) para abarcar distintas morfologías y escalas, y adaptar de mejor manera el algoritmo de búsqueda para esta aplicación en particular.

Finalmente, cabe mencionar que el algoritmo propuesto es computacionalmente muy eficiente y puede adaptarse para su ejecución en tiempo real.

RECONOCIMIENTOS

Este trabajo se realizó con el apoyo de los fondos CONACyT 154623 y 180604.

BIBLIOGRAFÍA

- [1] A. Alba, M. Rubio-Rincon, M. Rodriguez-Kessler, E. R. Arce-Santana, M. O. Mendez, "Un algoritmo de consenso para la búsqueda aproximada de patrones en cadenas de proteínas," *Revista Mexicana de Ingeniería Biomédica*, vol. 33, pp. 87-99, 2012.
- [2] Iyengar N, Peng C-K, Morin R, Goldberger AL, Lipsitz LA. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *Am J Physiol* 1996;271:1078-1084.
- [3] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med and Biol* 20(3):45-50 (May-June 2001). (PMID: 11446209)
- [4] Yun-Chi Yeh, Wen-June Wang, "QRS complexes detection for ECG signal: The Difference Operation Method", *Computer Methods and Programs in Biomedicine*, Vol. 91, Issue 3, September 2008, Pages 245-254
- [5] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability, Standards of measurement, physiological interpretation, and clinical use", *European Heart Journal* (1996) 17, 354-381.
- [6] Valtino X. Afonso, "ECG QRS detection", *Biomedical digital signal processing*, Pages 236-264.
- [7] G. Navarro, "A guided tour to approximate string matching," *ACM Computing Surveys*, vol. 33, pp-31-88, 2001.
- [8] Yanjun Li, Hong Yan, Feng Hong, Jinzhong Song, "A new approach of QRS complex detection based on matched filtering and triangle character analysis", *Australasian Physical & Engineering Sciences in Medicine*, September 2012, Volume 35, Issue 3, pp 341-356
- [9] P. Kathirvel, M. Sabarimalai Manikandan, S. R. M. Prasanna, K. P. Soman, "An Efficient R-peak Detection Based on New Nonlinear Transformation and First-Order Gaussian Differentiator", *Cardiovascular Engineering and Technology*, December 2011, Volume 2, Issue 4, pp 408-425.