

DETECCION DE BORDES EN SECUENCIAS DE RANGO CONTINUO

EDGE DETECTION IN CONTINUOUS-RANGE

G.G. Aguirre Varela ^a, M.A. Ré ^{a,b}, N.M. López ^c

^a Facultad de Matemática de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Argentina

^b Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina.

^c Gabinete de Tecnología Médica, Facultad de Ingeniería. U.N. de San Juan.

Recibido: 28/12/17; aceptado: 24/07/18

Se presenta un método para la detección de bordes de dominio o cambios de estacionariedad en secuencias de valores de rango continuo como los obtenidos en Electromiografía (EMG) o en registros de Electroencefalograma (EEG). La detección del cambio de estacionariedad en una secuencia temporal como las mencionadas presenta interés para el reconocimiento del comienzo de una contracción muscular en EMG o del comienzo y propagación de una crisis epiléptica en el análisis del registro del EEG.

El punto de segmentación en una serie temporal se corresponde con la posición en la serie a partir de la cual cambian las propiedades estadísticas de los valores que la conforman. El método aquí propuesto se basa en el cálculo de la divergencia de Jensen-Shannon (DJS) entre los segmentos que forman la secuencia. La DJS es una medida de distancia entre distribuciones de probabilidad y para su evaluación aproximamos las distribuciones que corresponden a cada segmento por el método del *kernel* de densidad.

Para la aplicación del método se elige una posición en la secuencia como punto de segmentación y se calcula la DJS entre las distribuciones asociadas a las sub-secuencias que quedan así definidas. Se repite este proceso para cada posición en la secuencia y se identifica el punto de segmentación con la posición que arroja el valor máximo para la DJS.

Evaluamos método propuesto analizando señales sintéticas similares a señales electromiográficas.

Palabras clave: divergencia de Jensen Shannon, análisis de secuencias, segmentación.

It is presented a method for border detection or changes in stationarity in sequences of continuous range values such as those obtained in electromyography (EMG) or in electroencephalograms (EEG) records. The detection of a change of stationarity in a temporal sequence as those mentioned is of interest in the recognition of the start of muscle contractions in EMG or the start of an epileptic crisis in EEG records.

The segmentation point in a temporal series is the position in the sequence where the statistical properties change. The method that we propose is based on the Jensen-Shannon Divergence (JSD) calculation between the segments that makes the sequence. JSD is a measure of distance between probability distributions. For its calculation we approximate the probability densities in each segment by the kernel density approximation method.

To make use of this method a position in the sequence is chosen a segmentation point and the JSD is calculated between the probability densities of each segment so defined. The procedure is repeated for each position in the sequence. The segmentation point is identified with the position that gives the maximum value for the JSD.

We evaluate the method here proposed by analyzing synthetic signals similar to EMG signals.

Keywords: Jensen-Shannon Divergence, sequence analysis, segmentation.

I. INTRODUCCIÓN

En el análisis de series temporales electrofisiológicas tienen interés clínico el reconocimiento y detección de cambios de estacionariedad, i.e. el cambio de las propiedades estadísticas en intervalos de tiempo. En muchas ocasiones el análisis de este comportamiento se basa en una inspección visual del registro obtenido y en el reconocimiento de patrones. En general este análisis es efectuado por personas expertas pero introduce un elemento subjetivo y una difícil sistematización

En esta comunicación consideramos el uso de la divergencia de Jensen-Shannon (DJS), una medida de distancia entre distribuciones de probabilidad, para la detección de bordes en una secuencia de valores de rango continuo equivalente a una serie temporal. Estos bordes corresponden a cambios en la estacionariedad de la secuencia.^{1, 2,3,4}

La DJS ha sido utilizada con éxito para el análisis de segmentación de secuencias simbólicas⁵ (i.e. secuencias de rango discreto), donde la estimación de las distribuciones de probabilidad por las frecuencias de aparición relativas de cada símbolo para cada segmento es relativamente simple. En el caso de secuencias de

rango continuo^{1,2}, como en una señal electrofisiológica, la estimación de las densidades de probabilidad correspondientes presenta una dificultad mayor. Para superar esta dificultad utilizamos aquí la aproximación del *kernel* de densidad para la estimación de las densidades de probabilidad de cada segmento⁶. En particular este método de aproximación presenta la ventaja sobre el más usual de *binning* al requerir un menor número de datos.

II. MÉTODOS

DJS

La DJS es una medida de distancia entre distribuciones de probabilidad originada en teoría de la información (identificable con el contenido de información mutua). No requiere de la continuidad absoluta de las distribuciones de probabilidad consideradas^{4,7}, está siempre bien definida y acotada y su raíz cuadrada es una métrica en el espacio de las distribuciones de probabilidad.

Lin⁴ propone una versión generalizada como una medida de distancia entre varias distribuciones de probabilidad en la expresión:

$$D(\pi_1, \dots, \pi_N)[P_1 \dots P_N] = H[\sum_{i=1}^N \pi_i P_i] - \sum_{i=1}^N \pi_i H[P_i] \quad (1)$$

En esta, $\{P_i\}$ es un conjunto de distribuciones de probabilidad para la variable X y $\{\pi_i\}$ un conjunto de factores de peso asociados a las distribuciones de probabilidad que deben cumplir

$$\sum_{i=1}^N \pi_i = 1$$

A su vez $H[P]$ es la entropía de Gibbs- Shannon:

$$H[P] = - \sum_{k=1}^m p_k \ln p_k$$

En este trabajo utilizamos la versión continua de la DJS para dos densidades de probabilidad, que asociamos a las dos sub-secuencias que forman la secuencia a analizar; i.e. debemos considerar la expresión para la entropía de una densidad de probabilidad para la variable Y

$$H[\mu] = - \int_{-\infty}^{\infty} dy \mu(y) \ln[\mu(y)]$$

Segmentación

Consideremos una secuencia de n de valores reales que suponemos constituida por dos sub-secuencias de largo n_1 y n_2 respectivamente ($n_1 + n_2 = n$) con distribuciones de probabilidad $\mu_1(y)$ y $\mu_2(y)$. Ilustramos esquemáticamente la situación en la Figura 1.

Para detectar el punto de segmentación definimos un cursor móvil, p_c , que señala un elemento particular en

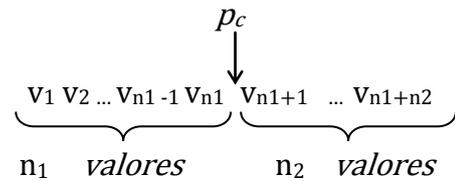


Figura 1. Esquema del problema de detección de segmentación.

la secuencia, dividiendo la misma en dos sub-secuencias con v_1 y v_2 elementos respectivamente ($v_1 + v_2 = n$). Este cursor se desplaza a lo largo de la secuencia y para cada posición se calcula la DJS (1) con $N = 2$ en la versión continua de la entropía

$$D(\pi_1, \pi_2)[\mu_1, \mu_2] = \sum_{i=1}^2 \pi_i \int_{-\infty}^{\infty} dy \mu_i(y) \ln[\mu_i(y)] - \int_{-\infty}^{\infty} dy \phi(y) \ln[\phi(y)] \quad (2)$$

En la ecuación (2) $\pi_i = v_i/n$, en tanto que, ϕ es la densidad marginal

$$\phi(y) = \pi_1 \mu_1(y) + \pi_2 \mu_2(y)$$

La DJS alcanza un valor máximo cuando $v_1 = n_1$, identificándose de este modo el punto de segmentación.

Método de aproximación

Para poder efectuar el cálculo de la DJS es necesario estimar las densidades de probabilidad presentes en la ecuación (2). Para esto usamos la aproximación del *kernel* de densidad⁶. Supongamos una realización de v_i repeticiones de una variable aleatoria Y con densidad de probabilidad $f(y)$. A partir de los valores registrados podemos aproximar

$$\hat{f}(y) = \frac{1}{v_i h} \sum_{i=1}^{v_i} K\left(\frac{y - y_i}{h}\right)$$

donde K es el *kernel* de densidad (función normalizada) y h un parámetro de suavizado. Aquí hemos usado un *kernel* gaussiano y el valor h informado como óptimo en la literatura⁷

$$h_0 = 1.06 \sigma (v_i)^{-\frac{1}{5}}$$

donde σ^2 es la varianza de la muestra calculada a partir del registro de valores.

Para realizar el cálculo de las integrales, primero, reconocemos en cada integral de la ecuación (2) el valor de expectación

$$\langle \ln[f(y)] \rangle = \int_{-\infty}^{\infty} dy f(y) \ln[f(y)]$$

Que aproximamos por

$$\langle \ln[f(y)] \rangle \cong \frac{1}{v_i} \sum_{j=1}^{v_i} \ln[\hat{f}(y_j)]$$

Así, la expresión para DJS queda

$$D^{(\pi_1, \pi_2)}[\mu_1, \mu_2] \cong \frac{1}{n} \left\{ \sum_{i=1}^2 \sum_{j=1}^{v_i} \ln[\hat{\mu}_i(y_j)] - \sum_{j=1}^n \ln[\hat{\phi}(y_j)] \right\} \quad (3)$$

Verificación

Verificamos el funcionamiento del método propuesto mediante 200 secuencias pseudo-aleatorias de $n = 2000$ valores. Cada secuencia está formada por dos sub-secuencias, con $n_1 = 1500$ y $n_2 = n - n_1 = 500$ valores respectivamente, generadas a partir de las densidades de probabilidad

$$\mu_i(y) = \frac{1}{\eta_i} \exp(-y/\eta_i)$$

con

$$\eta_1 = 1, \eta_2 = 2$$

Ilustramos en la Figura 2 el resultado obtenido al promediar la DJS sobre las 200 secuencias para cada posición del cursor. El método detectó correctamente el punto de segmentación en la posición $v_1 = 1500$.

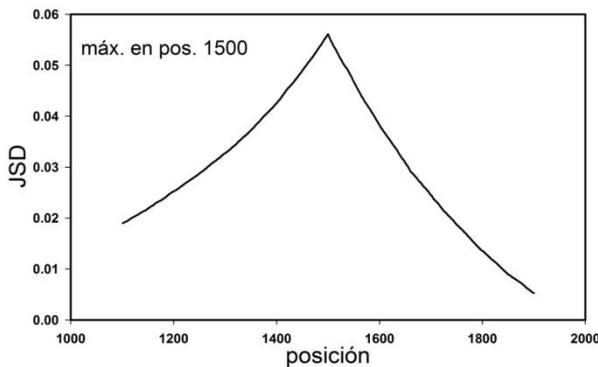


Figura 2. Promedio de 200 secuencias

III. RESULTADOS

Utilizamos el método propuesto para analizar 100 secuencias artificiales generadas siguiendo a Staude⁸. Las secuencias tienen una longitud equivalente a un registro de 10 segundos con una resolución de 1ms. En este lapso la señal pasa de un régimen estacionario a otro por un transitorio que se extiende desde los 4s hasta los 6.245s. En la Figura 3 se ilustra una de las secuencias generadas con la zona de transición resaltada.

En la Figura 4 ilustramos el resultado obtenido al promediar la DJS calculada, para cada posición del cursor, sobre las 100 secuencias. También se indica la posición del punto de segmentación detectado, correspondiente a la posición del máximo de la DJS. Como se puede ver, el punto de segmentación se encuentra aproximadamente en el centro de la zona de

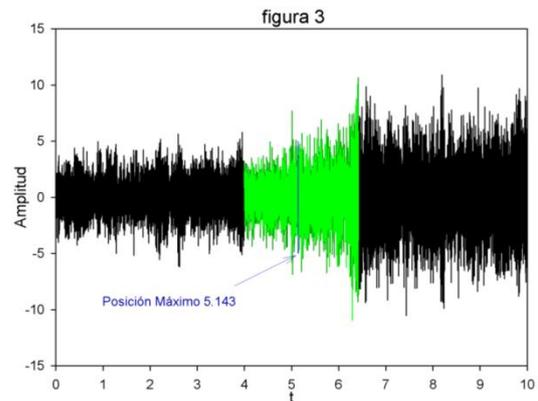


Figura 3. Secuencia generada con la zona de transición resaltada

transición. En la Figuras 3 se indica esta posición dentro de la secuencia.

En el inset de la Figura 4 se incluye el histograma correspondiente a la posición de segmentación detectada en cada secuencia. Puede verse que los puntos de segmentación están dentro de la zona de transición para todos los casos.

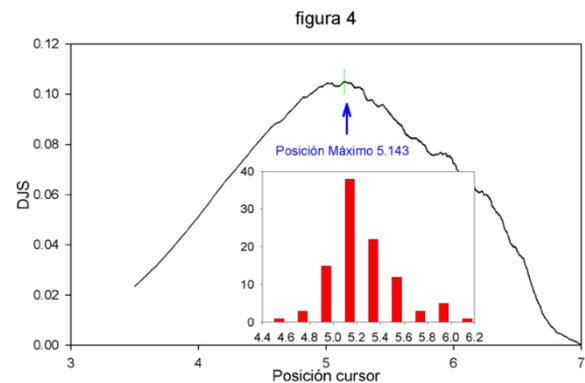


Figura 4. Resultado obtenido de promediar sobre las 100 secuencias la DJS calculada. Histograma correspondiente a la posición de segmentación detectada

Finalmente se consideró la unión de dos secuencias, a los fines de evaluar la capacidad de discriminación del método. En una primera corrida del programa se detectaron dos puntos de segmentación, correspondientes a las posiciones dentro de las zonas de transición. El valor de la DJS se ilustra con la línea roja en la Figura 5. Detectados estos dos puntos de segmentación se extrajo la zona central delimitada por estos puntos y se volvió a correr el programa, detectándose el punto de unión de las dos secuencias como un nuevo punto de segmentación, ilustrada con la línea negra. De esta manera se verifica que el uso iterativo del método permite distinguir nuevas regiones. En todos los casos la escala de gráfico fue elegida de tal forma de hacer apreciable el valor de la DJS.

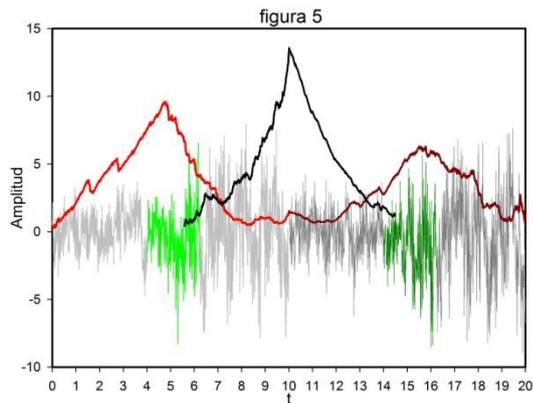


Figura 5. Uso iterativo del método

IV. CONCLUSIONES

En la presente comunicación se ha presentado un método para la detección de bordes o cambios de estacionariedad en secuencias de valores de rango continuo. El mismo está basado en el cálculo de la Divergencia de Jensen-Shannon (DJS) a partir de la aproximación de las distribuciones de probabilidad asociadas a los valores en los segmentos homogéneos. Esta aproximación es una alternativa al método de *binning*, en particular requiere un menor número de datos para conseguir buenos resultados. El método mostró un desempeño satisfactorio, detectando los puntos de segmentación en secuencias generadas a partir de distribuciones de probabilidad conocidas. Aún en los casos en los cuales había un transitorio en la señal. Además, mostró ser una herramienta que puede ser aplicada a la detección de puntos de segmentación en forma iterativa.

V. REFERENCIAS

- 1 - López N., Orosco E., di Sciacio F. Surface Electromyographic Onset Detection Based on Statistics and Information Content. *J. Phys.: Conference Series*, 332:012043, 2011.
- 2 - Pereyra M., Lamberti P., Rosso O. Wavelet Jensen-Shannon divergence as a tool for studying the dynamics of frequency band components in EEG epileptic seizures. *Phys. A*, 379:122–132, 2007.
- 3 - Grosse I., Bernaola-Galván P., Carpena P., Román-Roldán R., Oliver J. y Stanley H. Analysis of symbolic sequences using Jensen-Shannon divergence. *Phys. Rev. E*, 65:041905-1/16, 2002.
- 4 - Lin J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inform. Theory*, 37:145–151, 1991
- 5 - Ré M. y Azad R. Generalization of Entropy Based Divergence Measures for Symbolic Sequence Analysis. *PLoS ONE*, 9:e93532, 2014.
- 6 - Silverman B. W. *Monographs on Statistics and Applied Probability 26: Density Estimation for Statistics and Data Analysis* Chapman and Hall/CRC 1986.
- 7 - Steuer R., Kurths J., Daub C., Weise J., Selbig J. The Mutual Information: Detecting and Evaluating dependencies between variables. *Bioinformatics*, 18 S2:S231–S240, 2011.
- 8 - Staude G.H. Precise Onset Detection of Human Motor Responses Using a Whitening Filter and the Log-Likelihood-Ratio Test. *IEEE Transactions on Biomedical Engineering*, vol 48, NO. 11, 2001.