

Aplicación de clasificadores en la detección de patologías en señales electrofisiológicas de potenciales evocados

- Juan Luis Barberia,^{1,2} Yanina Corsaro,^{1,2} Lorenzo Licata Caruso,^{1,2} Marco Gabriel Baldiviezo,^{1,2} Juan Facundo Fernandez Biancardi,^{1,2} Masia Rebeca Hernando,^{1,2} Camila Bontempo,^{1,2} Adrian Paglia,^{1,2} Matias Rodriguez,^{1,2} Walter Legnani^{1,2}

¹Centro de Procesamiento de Señales e Imágenes

²Facultad Regional Buenos Aires. Universidad Tecnológica Nacional

El objetivo de este trabajo fue evaluar la performance de tres algoritmos de clasificación automática de señales: Random Forest, Support Vector Machines y Nearest Neighbor [1] para distinguir señales de potenciales evocados de individuos normoyentes y patológicos.

Para esto se construyó un dataset compuesto por mediciones provenientes de señales de potenciales auditivos evocados de corta latencia extraídas del banco de datos de Physionet [2]. La señal estímulo del estudio consistió en tonos de diferente intensidad y diferente frecuencia. Se tomaron las señales en su formato original en bruto, de las cuales 500 corresponden a pacientes normoyentes y 500 a patológicos. Las señales fueron procesadas para incorporarlas al dataset mediante un denoising constituido por onditas en base a umbrales, tomando los primeros 5000 puntos de cada señal. Se subdividió el dataset por bandas de frecuencia del estímulo: 1 kHz y 4 kHz, y de intensidad sonora: 10 dB a 40 dB, 40 dB a 60 dB y 60 dB a 100 dB.

Con el objetivo de disminuir el sobreajuste y validar, se utilizó un 80% para el modelado con los diferentes algoritmos y un 20% para validación. A través de diferentes magnitudes correspondientes a sistemas dinámicos no lineales, como los exponentes de Lyapunov a través de una adaptación del algoritmo de Wolf [3], la Dimensión Fractal a través del algoritmo de Higuchi [4] y a través del cálculo de la Entropía de Permutación [5], Entropía de Permutación Ponderada [6] y Modificada [7] se calcularon las diferentes variables que constituyeron el dataset. Para el cálculo de la función de distribución de las entropías se utilizó una dimensión de embedding de 3 y 4, y un time delay de 15 y 20, siendo este el rango de valores habituales que se utiliza en la literatura de física médica, obteniéndose con estos los mejores resultados para este tipo de señales.

La performance del clasificador se ponderó mediante la elaboración de las curvas características operativas del receptor (en inglés ROC) [8]. A partir de esto se calculó el porcentaje de aciertos en la predicción del modelo, el área bajo la curva ROC, la sensibilidad, la especificidad y la precisión.

Los tres clasificadores empleados mostraron una elevada performance en el trabajo, resultando para algunas combinaciones del dataset mejores el Random Forest y el Support Vector Machines.

Referencias:

- [1] Aggarwal, CH, *Data Classification Algorithms and Applications*, (2015).
- [2] [Physionet](#) Consultado 17/07/2021.
- [3] Alan Wolf, Jack B. Swift, Harry L. Swinney, and John A. Vastano, *Determining Lyapunov exponents from a time series*, Department of Physics, University of Texas, Austin, USA, (1984).
- [4] T. Higuchi, *Approach to an irregular time series on the basis of the fractal theory*, Phys. D **31**, 277 (1987).
- [5] C. Bandt and B. Pompe, *Permutation Entropy: A Natural Complexity Measure for Time Series*, Phys. Rev. Lett. **88**, 174102 (2002).
- [6] B. Fadlallah, B. Chen, A. Keil, and J. Principe, *Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information*, Phys. Rev. E **87**, 022911 (2013).
- [7] Cuesta-Frau, D. *Using the Information Provided by Forbidden Ordinal Patterns in Permutation Entropy to Reinforce Time Series Discrimination Capabilities*, Entropy **22**, 494 (2020).
- [8] Wojtek J. Krzanowski, *ROC Curves for Continuous Data*, CRC Press (2009).