Texto

Descripción generada automáticamente

**INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA**

**ESCUELA DE INGENIERÍA Y GESTIÓN**

*Análisis del rendimiento en el algoritmo de machine learning que emula ondas ERP P300 usado en un experimento con interfaces cerebro computador en pacientes con ELA (Esclerosis Lateral Amiotrófica).*

**AUTOR: Chavez Montaño, Alexander. (Leg. Nº 506218)**

**TUTOR: Ramele, Rodrigo**

**Trabajo final presentado para la obtención del título de especialista en ciencia de datos.**

**BUENOS AIRES**

**Segundo cuatrimestre, 2023**

*A mis dos emes: Macu y Marco.*



**Tabla de contenido**

1. Introducción … 5

2. Estado del arte … 6

3. Definición del problema … 5

4. Justificación del estudio … 5

5. Alcances del trabajo y limitaciones … 5

6. Hipótesis … 5

7. Objetivos … 5

8. Metodología … 5

Técnicas

Herramientas

9. Referencias-Bibliografía … 5

**Abstract**

Este trabajo complementa la investigación previa realizada en los experimentos descritos en el artículo EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica): parte del resultado de dicha investigación fue la de investigar algoritmos de machine learning que emulan ondas ERP P300: acá se realiza un análisis exploratorio tanto del objeto de estudio como de las herramientas computacionales disponibles para luego realizar un abanico de pruebas que arrojen distinta *performance* y nos permitan proponer mejoras en la preconfiguración de dicho algoritmo.

**1. Introducción**

La electroencefalografía es una de las herramientas clínicas que, a lo largo de las últimas décadas, se ha convertido en uno de los principales métodos para obtener imágenes en tiempo real del comportamiento cerebral de manera no invasiva, portátil y móvil más usado en el ambiente médico [1]. Dentro de la electroencefalografía tenemos un conjunto de ondas con distintas características que varían en sus propiedades físicas como amplitud o frecuencia, como también en el origen y la ubicación en las distintas zonas del cerebro. La onda P300 se obtiene de ubicar un canal en el lóbulo parietal y su comportamiento es reactivo debido a estímulos esperados pero infrecuentes relacionados con actos cognitivos.

La electroencefalografía , sin embargo, está expuesta a alteraciones no deseadas en sus resultados, ya que, por más controlado, preciso y consistente que sea el ambiente donde se realiza el experimento o la toma de muestra, estaremos sujetos a factores fuera de nuestro control. Esta problemática se suele afrontar generando ambientes de pruebas donde se pueda recrear la situación con la mayor fidelidad posible. Los experimentos pasados y éste trabajo integrador se cimentan en la base de datasets sintéticos, artificiales, con los que se simulan respuestas de ondas ERP P300 a partir de electroencefalogramas reales, con resultados previamente conocidos, a fin de trabajar en la performance de algoritmos que logren resultados con mejoras en el tiempo.

Los métodos y los procedimientos cuantitativos para automatizar la decodificación de ondas EEG como la P300 se basa en EEG no invasivo [2]. Sin embargo, los métodos de la decodificación de señales, basadas en detección de formas de onda, y además con algoritmos de machine learning son relativamente escasos.

Se pretende entonces, a través de este trabajo, darle continuidad a la investigación previa realizada en los experimentos descritos en el artículo EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica): en éste trabajo se realiza un análisis exploratorio de electroencefalogramas llamados pasivos; pacientes que participaron del experimento pero desconociendo las reglas de interacción con el *speller*, explicado más adelante en profundidad. En una etapa posterior se “inyectan” potenciales P300 en los lugares donde sabemos de antemano suceden los eventos, y realizamos modificaciones en las propiedades de las ondas que arrojan distinta *performance* para permitirnos obtener mejoras en la preconfiguración del algoritmo.

**2. Estado del arte y Marco conceptual**

El estado del arte es el estado de la cuestión, dónde está parada la investigación hasta ése momento.

La electroencefalografía es una de las herramientas clínicas que, a lo largo de las últimas décadas, se ha convertido en uno de los principales métodos para obtener imágenes en tiempo real del comportamiento cerebral de manera no invasiva, portátil y móvil más usado en el ambiente médico **[8]**. Sin embargo, está expuesta a alteraciones no deseadas en sus resultados, ya que, por más controlado, preciso y consistente que sea el ambiente donde se realiza el experimento o la toma de muestra, estaremos sujetos a que el objeto de estudio, en este caso es el ser humano, incurrirá en desconcentración o desenfoque al momento de hacer las pruebas y esto modificar la respuesta esperada.

Dentro de la electroencefalografía tenemos un conjunto de ondas con distintas características que varían en sus propiedades físicas como amplitud o frecuencia, como también en el origen y la ubicación en las distintas zonas del cerebro. La onda P300 se obtiene de ubicar un canal en el lóbulo parietal y su comportamiento es reactivo debido a estímulos esperados pero infrecuentes relacionados con actos cognitivos.

Los métodos y los procedimientos cuantitativos para automatizar la decodificación de ondas EEG como la P300 se basa en EEG no invasivo **[2]**. Sin embargo, los métodos de la decodificación de señales, basadas en detección de formas de onda, y además con algoritmos de machine learning, es relativamente escaso.

Marco conceptual

: define con precisión los conceptos centrales

del dominio del problema (Palabras Clave).

Marco teórico

: explica las teorías sobre el

Estado del Arte

: sintetiza el estado actual que se encuentre en

otras investigaciones o situaciones similares a la estudiada.

Antecedentes

: describe aquellos antecedentes que contemplen

diversas formas de resolución de ese problema , que se hayan

construido en forma previa a la investigación.

**3. Definición del problema**

La continuidad de los proyectos de investigación es vital para obtener avances y mejoras en los resultados de los experimentos. Particularmente, las investigaciones en el tratamiento de señales electroencefalográficas con modelos de machine learning son un campo de estudio relativamente nuevo con relativamente poca disponibilidad de datos, lo que genera obstáculos que impiden realizar experimentos comparativos a gran escala. En el caso de los experimentos descritos en el artículo EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica), su continuidad se encontraba pausada por razones ajenas a éste documento. Es posible que, una vez reanudada estas investigaciones, se puedan dar saltos posteriores con experimentos comparativos a gran escala.

1. Introducción 100%

2. Estado del arte 30%

3. Definición del problema 100%

4. Justificación del estudio

5. Alcances del trabajo y limitaciones

6. Hipótesis

7. Objetivos 30%

8. Metodología 30%

Técnicas

Herramientas

9. Referencias-Bibliografía 30%

**7. Objetivos**

**General.**

Darle continuidad a los experimentos e investigaciones previas realizadas en el Instituto Tecnológico de Buenos Aires ITBA, analizando y obteniendo resultados del rendimiento en el algoritmo de machine learning que emula ondas ERP P300 usado en un experimento con interfaces cerebro computador en pacientes con ELA (Esclerosis Lateral Amiotrófica).

Los traje del documento de la materia. Refinarlos y ya quedan listos.

**Específicos.**

Identificar las mejorías generadas al ingresar progresivamente mas elementos a la muestra.

Analizar y encontrar cuáles son las variables que, una vez alteramos sus dimensiones, impactan en mayor medida en la reconstrucción de la onda.

Ajustar y calibrar los hiperparámetros en los algoritmos de machine learning a medida que la muestra crece y las distintas variables de las ondas involucradas se vayan alterando.

Describir todos y cada uno de las etapas que se requieren hacer en la implementación de los algoritmos y la obtención de los resultados, de manera que puedan ser fácilmente identificables posibles elementos de mejora en futuras investigaciones.

**8. Metodologías a usar**

Los traje del documento de la materia. Refinarlos y ya quedan listos.

Explayarme en la librería mne principalmente.

Describir brevemente las otras librerías como numpy.

La metodología a usar será CRISP-DM con la aclaración que no tendrá etapa de despliegue, ya que es un trabajo de investigación. El lenguaje de programación elegido es phyton.

Las librerías especializadas serán scipy para todas las herramientas matemáticas necesarias como

álgebra lineal, integración, interpolación, FFT, procesamiento de señales e imágenes, etc. Y numpy para el manejo de las señales EEG, ya que requieren estructuras de datos especiales.

Preparación de los datos:

Los datos ya se encuentran disponibles. Los EEG que serán usados son de experimentos anteriores en el Departamento de ingeniería de computación del Instituto Tecnológico de Buenos Aires, y una base de datos en la web: [P300 speller with ALS patients (008-2014)](http://bnci-horizon-2020.eu/database/data-sets) para la etapa de comparación.

La exploración de métodos basados en formas de onda se realiza siguiendo las directrices PRISMA. Los criterios para identificar métodos que se basan en la forma de onda de la señal:

• Forma de onda

• Inspección visual.

• Coincidencia de patrones en el dominio del tiempo.

• Extracción de características.

• Creación de un diccionario de plantillas en base a las características.

Modelado:

En base a los criterios, los algoritmos que se evaluarán son los siguientes:

• Búsqueda de coincidencias.

• Entropía de permutación.

• Código de cadena de pendiente horizontal.

• Transformación de característica de escala invariante.

Siendo los dos últimos los de mayor rendimiento y en los cuales se enfocará el experimento. Es acá donde se pretende lograr mejores resultados.

El próximo paso es aplicar los algoritmos de clasificación. En todos los datasets se aplicará cross validation para luego usar Support Vector Machine SVM, la base de la comparación será una señal reconstruida y normalizada.

Evaluación:

Se revisarán los resultados obtenidos de la clasificación con las modificaciones realizadas en la señal y los hiperparámetros seleccionados, que en el caso de SVM es el hiperplano con el que separa el conjunto de datos.

En función de los resultados obtenidos se generará nuevamente el ciclo de pruebas hasta alcanzar los valores óptimos.

**9. Referencias / Bibliografía**

**(OJO QUE YA ESTÁN ORGANIZADAS Y CITADAS)**

1. Hartman, A.L. Atlas of EEG Patterns; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.

2. Guger, C.; Allison, B.Z.; Lebedev, M.A. Introduction. In Brain Computer Interface Research: A State of the Art Summary 6; Springer: Cham, Switzerland, 2017; pp. 1–8.

3. Ramele, R.; Villar, A.J.; Santos, J.M.; EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces: Computer Engineering Department, Instituto Tecnológico de Buenos Aires (ITBA), Published: 16 November 2018.

4. Harari, Y.N; Homo Deus: A Brief History of Tomorrow; Jerusalem University

5. Skoog, D.A.; West, D.M.; Holler, F.J.; Crouch, S.R. Analytical Chemistry: An Introduction; Saunders College, Publishing: Philadelphia, PA, USA, 2000.

6. Owens, T.J.; Zandt, G.; Taylor, S.R. Seismic evidence for an ancient rift beneath the Cumberland Plateau, Tennessee: A detailed analysis of broadband teleseismic P waveforms. J. Geophys. Res. Solid Earth 1984, 89, 7783–7795.

7. Stockman, G.; Kanal, L.; Kyle, M. Structural pattern recognition of carotid pulse waves using a general waveform parsing system. Commun. ACM 1976, 19, 688–695.

8. Hartman, A.L. Atlas of EEG Patterns; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.

|  |  |
| --- | --- |
| **Paper** | **Título** |
| brainsci-08-00199.pdf | EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces |
| P300 dataset of 8 healthy subjects.pdf | P300 dataset of 8 healthy subjects.pdf |
| fncom-13-00043.pdf | Histogram of Gradient Orientations of Signal Plots Applied to P300 Detection |
| UMA-BCI Speller.pdf | UMA-BCI SPELLER: PLATAFORMA DE COMUNICACIÓN DE FÁCIL CONFIGURACIÓN BASADA EN EL BCI2000 P300 SPELLER |
| P300 Speller with patients with ALS | P300 Speller with patients with ALS |
| Picton 1992 | The P300 wave of the human Event-Related- Potential. |
| vucic2020.pdf | P300 jitter latency, brain-computer interface and amyotrophic lateral sclerosis |
| tesis\_n3966\_Gambini.pdf | Modelos de segmentación basados en regiones y contornos activos aplicados a imágenes de radar de apertura sintética |
| fnins-07-00267.pdf | MEGandEEGdataanalysiswithMNE-Python |