

TP1: Microestados en un registro electroencefalográfico

Juan E Kamienkowski, Luz Bavassi
Data Mining en Ciencia y Tecnología

29 de septiembre de 2021

La actividad del cerebro en durante el estado de reposo permite detectar características en los animales que no se pueden observar durante una tarea [3]. A primera vista, la actividad espontánea registrada con EEG da la impresión de ser bastante desorganizada. Sin embargo, cuando se analizan segmentos de tiempo corto, entre 80-150ms, se puede observar que dominan algunas configuraciones topográficas o mapas [4]. Estos períodos de cuasi-estabilidad se nombraron microestados funcionales. A partir del argumento que las diferentes topografías de EEG son generadas por configuraciones neuronales, se desprende que los microestados reflejan estados funcionales del cerebro [6].

Se sabe que los períodos offline, luego de un aprendizaje, favorecen el almacenamiento (consolidación) de la memoria [2]. En su trabajo, Poskanzer y colaboradores [5] mostraron existen 5 microestados que caracterizan el estado de reposo previo y posterior al aprendizaje, y que el incremento del microestado que presenta una activación occipital central correlaciona con la cantidad de palabras recordadas.

En este trabajo práctico tiene como objetivo principal el estudio y exploración de datos de EEG registrados durante períodos de reposo previo y posterior a un aprendizaje.

Objetivos específicos

- Calcular los microestados para un conjunto de registros de EEG. Analizar el número óptimo de conjuntos que se necesitan para segmentar los datos.
- Validación interna y externa de los clusters hallados. Comparar los conjuntos encontrados con diferentes métodos.
- Estudiar la dinámica de los microestados para las condiciones previa/posterior al aprendizaje.

Estructura de los datos:

Se van a utilizar los datos utilizados en el preTP. En la carpeta <https://drive.google.com/drive/folders/1NEXTebl0DIMUI7ShhLvEFmezd7uH0rQ2?usp=sharing> se encuentran los archivos separados por cada sujeto y cada período de reposo (bloque 1: reposo previo al

aprendizaje; bloque 2: reposo posterior al aprendizaje). La notación es la siguiente *suj-suj-bl-bloque.txt*.

Los archivos tienen datos de registros de 30 canales a una frecuencia de muestreo de 128 Hz. Los datos fueron previamente filtrados entre 2 y 20 Hz y re-referenciados al promedio de los electrodos.

Tarea 1: Inspección de los mapas topográficos para los máximos de GFP

Calcular el Campo de Potencia Global (GFP,[1]) a partir de la concatenación en el tiempo de los registros. El GFP está dado por la ecuación 1:

$$GFP_u = \sqrt{\frac{\sum_{i=1}^C (u_i - \langle u_i \rangle)^2}{C}} \quad (1)$$

Encontrar los picos máximos del GFP (P_{GFP}), guardar los mapas para cada máximo (\vec{x}_n).

Estudiar la presencia de mapas *outliers*. Para ello puede usar PCA y/o clustering jerárquico y/o matriz de distancias.

Tarea 2: Cálculo de microestados utilizando k-means modificados

Aplicar el algoritmo de K-means modificado sobre los P_{GFP} . Utilizar $k = 2 - 10$. ¿Es necesario repetir el procedimiento más de una vez para cada valor de k ?

Decidir el mejor valor de k . Calcular el promedio de Silhouette, el parámetro de Validación Cruzada y la Variancia global explicada para cada k . Realizar gráfico de las tres medidas para el k correspondiente.

Para el valor de k óptimo realizar el perfil de Silhouette completo. Inspeccionar, a partir de esta última figura, la presencia de *outliers*, si es necesario descartarlos del análisis y volver a calcular los clusters para este valor de k .

Tarea 3: Validación externa de los microestados

Realizar al menos dos métodos de clustering diferentes para los \vec{x}_n . Determinar la distancia que va a utilizar. Justificar brevemente porque se eligen esos métodos y nociones de distancia. Realice una inspección de las características de los clusters hallados y de los prototipos de cada cluster en caso que los hubiese (o de ejemplos en el caso que no), comparar con los encontrados por el método de k-means modificado.

Calcular la matriz de confusión entre algoritmos. Determinar el índice Rand.

Tarea 4: Visualización

Para los distintos algoritmos, visualizar los datos y las etiquetas de *cluster* asignadas en baja dimensión con alguna técnica de reducción (PCA, TSNE, MDS, etc). Esto se puede repetir para alguna variación que encuentren interesante, como para distintos valores de k o distintas nociones de distancia.

Tarea 5: Comparación de medidas de los microestados

Calcule el ajuste de los datos de EEG a los prototipos de microestados hallados en la **Tarea 2** para cada registro. Realice un suavizado de las series temporales (puede utilizar una ventana promedio móvil) y asigne a cada punto temporal el mapa prototipo que presente mayor correlación.

A partir de aca debe elegir al menos uno de los siguientes medidas para comparar entre ambas condiciones experimentales (antes y después del aprendizaje): (*ver medidas en d’Croz-Baron y colaboradores [3]*)

- Duración promedio de cada microestado.
- Frecuencia de ocurrencia.
- Fracción de tiempo cubierto.
- Global Explained Variance.

Formato

Les proponemos seguir el formato de publicación en una revista científica, pueden encontrar muchos formatos directamente en *Overleaf*, por ejemplo *IEEE Conference Template for ANCS 2019*. No es obligatorio seguir ese formato, pero si elegir el formato de alguna revista (también pueden encontrarlos disponibles para Word).

Las revistas suelen tener además instrucciones respecto al formato online, desde restricciones en el tamaño de cada sección, en el número de figuras/tablas, las secciones que debe contener, hasta formato de los números, referencias, etc.

Aquí ponemos nuestras restricciones, pero si quieren adaptarlo a alguna revista en particular también vale.

Secciones.

1. Título (máx. 100 caracteres), tiene que ser expresivo (no vale TP1).
2. Resumen (máx. 200 palabras), tiene que contener una descripción de todo el trabajo: Motivación, Antecedentes, Objetivos, Métodos, Resultados y alguna Conclusión.
3. Introducción Comienza con la motivación, sigue con los antecedentes, y termina siempre con un párrafo de objetivos (no es necesario que este dividido en sub-secciones). Típicamente, una vez que motivaron el trabajo y mostraron lo que hay hecho, viene una frase del estilo "Por ende, nos proponemos....".^o "Aquí nos proponemos...".

4. Métodos Detalle de los métodos a utilizar, en este caso no es necesario profundizar mucho pero pueden enumerarlos y sobre todo es el lugar para incluir cualquier método fuera de lo común que hayan utilizado.
5. Resultados y discusión Aquí se enumeran y discuten los resultados. Es muy importante que no sea una seguidilla de figuras y tablas. Como regla pueden considerar: *"Si una figura no se describe/comenta en el texto es que: O bien está de más y no hace a la historia, o bien se olvidaron de incluirla."*
6. Conclusiones Comienza generalmente con un resumen muy breve de los principales resultados obtenidos (uniendo distintas secciones), y luego se pasa a conclusiones generales, detallando problemas detectados, posibles explicaciones y trabajo a futuro.
7. Referencias Citas bibliográficas utilizadas durante el reporte. Si son sitios web o repositorios se incluyen generalmente al pie de la página que corresponde y no como cita bibliográfica.

Considerando Introducción, Métodos, Resultados y Conclusiones no deben superar las 5000 palabras (aprox.). Finalmente, considerando el formato de este trabajo pueden dividirlo en

1. Título
2. Resumen
3. Introducción
4. Experiencia 1: KMeans por ejemplo,
 Métodos
 Resultados y discusión
5. Experiencia 2: Otro algoritmo,
 Métodos
 Resultados y discusión
6. Experiencia N:
7. Conclusiones
8. Referencias

Nota final

El TP se realizará en grupos de a tres o cuatro personas. Se puede usar cualquier herramienta de análisis o combinación de herramientas, debiendo indicarla en el informe. El lenguaje (Python o R) en el que se desarrolle el TP no es excluyente.

La fecha límite de entrega es el día Domingo 17/10/2021 a las 23.55hs a través del campus.

Tabla de puntos: Cantidad máxima de puntos que se pueden obtener por ...

- ... la tarea 1: 1.5
- ... la tarea 2: 3
- ... la tarea 3: 2.5
- ... la tarea 4: 1.5
- ... la tarea 5: 1.5

Referencias

- [1] BRODBECK, V., KUHN, A., VON WEGNER, F., MORZELEWSKI, A., TAGLIAZUCCHI, E., BORISOV, S., MICHEL, C. M., AND LAUFS, H. Eeg microstates of wakefulness and nrem sleep. *Neuroimage* 62, 3 (2012), 2129–2139.
- [2] BROKAW, K., TISHLER, W., MANCEOR, S., HAMILTON, K., GAULDEN, A., PARR, E., AND WAMSLEY, E. J. Resting state eeg correlates of memory consolidation. *Neurobiology of learning and memory* 130 (2016), 17–25.
- [3] D’CROZ-BARON, D. F., BAKER, M., MICHEL, C. M., AND KARP, T. Eeg microstates analysis in young adults with autism spectrum disorder during resting-state. *Frontiers in human neuroscience* 13 (2019), 173.
- [4] MICHEL, C. M., AND KOENIG, T. Eeg microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage* 180 (2018), 577–593.
- [5] POSKANZER, C., DENIS, D., HERRICK, A., AND STICKGOLD, R. Using eeg microstates to examine post-encoding quiet rest and subsequent word-pair memory. *Neurobiology of Learning and Memory* 181 (2021), 107424.
- [6] POULSEN, A. T., PEDRONI, A., LANGER, N., AND HANSEN, L. K. Microstate eeglab toolbox: An introductory guide. *BioRxiv*, 289850 (2018).