

CAPÍTULO II

Inteligencia artificial y cerebro computacional a través de la fusión de datos*

Humberto Bustince

La neurociencia computacional es un área de estudio que trata de entender el funcionamiento del cerebro, utilizando inteligencia artificial, matemáticas, ciencia de datos o física, entre otros. Dentro de la neurociencia computacional, el cerebro computacional investiga cómo actuar sobre objetos físicos con la actividad cerebral. En este trabajo vamos a presentar algunos de los recientes desarrollos en este último campo, centrándonos en cómo manejar los datos y la incertidumbre ligada a ellos. El objetivo es desarrollar técnicas que permitan, por ejemplo, recuperar movilidad a pacientes afectados de enfermedades neurológicas.

Palabras clave: neurociencia computacional, cerebro computacional, fusión de datos, incertidumbre.

* Este trabajo ha sido parcialmente financiado por el proyecto PID2022-136627NB-I00 financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

1. INTRODUCCIÓN

En los últimos años, es ya un tópico decir que la inteligencia artificial (IA) está adquiriendo un creciente protagonismo en prácticamente todos los campos de la actividad humana. En particular, los recientes desarrollos de Grandes Modelos de Lenguaje (*LLM*) como el ChatGPT han copado y siguen acaparando titulares en la prensa y dando lugar a numerosas discusiones sobre sus posibilidades y riesgos. Muchas de estas discusiones se mueven en terrenos más próximos a la ciencia ficción que a la realidad científica, contribuyendo en ocasiones a crear confusión, temores infundados o falsas esperanzas. Al mismo tiempo, no se pone el foco en problemas cruciales como la interpretabilidad, la explicabilidad o la necesidad de datos de calidad.

Por otra parte, hay muchos desarrollos que, a pesar de contar con una historia relativamente larga, han ido pasando más desapercibidos para la sociedad. Y ello, a pesar de que su impacto potencial en nuestras vidas resulta inmenso y de que están recibiendo fuertes inversiones económicas. Este es el caso, por ejemplo, de las investigaciones en el campo de la salud ligadas a la noción de medicina personalizada o de precisión (Smith, 2012), que vienen siendo impulsadas desde hace años y están ya transformando la forma en que se entiende y se practica la asistencia sanitaria.

Otro campo que puede ser clave para el futuro de la inteligencia artificial es el dedicado a la comprensión de la inteligencia humana. Dado que esta última depende, en gran medida, de que entendamos cómo funciona su elemento más representativo, el cerebro, no es de extrañar el gran esfuerzo que está atrayendo la neurociencia computacional (Barrenechea *et al.*, 2013; Lindsay, 2021); es decir, el estudio y comprensión de los mecanismos de funcionamiento del cerebro humano.

En este trabajo, vamos a considerar algunos recientes desarrollos en una de las áreas específicas de estudio dentro de la neurociencia computacional: el denominado cerebro computacional. En concreto, nos centraremos en cómo podemos actuar sobre dispositivos externos por medio de la actividad cerebral (Sejnowski y Churchland, 1992). Este desarrollo específico presenta el interés añadido de permitirnos discutir de forma simple algunos aspectos críticos para el desarrollo de la inteligencia artificial:

1. cómo podemos extraer información a partir de los datos disponibles, y
2. qué ocurre cuando existen problemas de incertidumbre o ruido ligados a los datos.

En lo fundamental, vamos a seguir los desarrollos en Ko *et al.* (2019), si bien también incluiremos resultados de trabajos posteriores. En todo caso, el esquema del trabajo es el siguiente. Comenzamos con una breve discusión sobre el concepto de inteligencia artificial. En la sección 3, comentamos las principales nociones relacionadas con la neurociencia computacional y el cerebro computacional. En la sección 4 tratamos los métodos no invasivos para el cerebro computacional, y en la sección 5, que se puede considerar el núcleo de esta propuesta, nos centramos en el problema de los datos. Terminamos con unas breves conclusiones y una lista de referencias.

2. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

Dado que vamos a utilizar el problema del cerebro computacional para entender algunos aspectos relevantes de la inteligencia artificial, es importante comenzar fijando el terreno sobre el que nos movemos. Así pues, empezamos tratando de responder la pregunta que sirve de título a esta sección.

El Reglamento de inteligencia artificial de la Unión Europea de 2024 (EU Artificial Intelligence Act, 2024), define la inteligencia artificial como:

“Sistemas automatizados diseñados para operar con diversos grados de autonomía y que, en función de objetivos explícitos o implícitos, generan resultados, predicciones, recomendaciones o decisiones que influyen en el entorno físico o virtual”.

Esta definición, aunque formalmente correcta, resulta bastante oscura y poco precisa. Por ello, en este trabajo vamos a considerar la que dio el investigador John McCarthy, uno de los padres de la inteligencia artificial, en la convocatoria de la Conferencia de Dartmouth de 1955 (McCarthy, 1955). Esta conferencia se considera el hito fundacional de la inteligencia artificial, y fue precisamente donde se acuñó el término, definido como sigue:

“inteligencia artificial es la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos requerirían inteligencia”.

El primer punto a resaltar en esta definición es que establece claramente el carácter híbrido de la inteligencia artificial. Por una parte, es una ciencia. Por tanto, requiere de sólidos fundamentos teóricos y del uso del método experimental para obtener y validar posibles resultados. Y, por otra parte, presenta un carácter fuertemente aplicado. Esta dualidad resulta especialmente clara en el caso del cerebro computacional, como veremos.

¿De qué forma puede desarrollarse la inteligencia artificial? O, dicho de otro modo, ¿cómo se pueden desarrollar esas máquinas inteligentes? Desde sus orígenes, se han utilizado dos posibles aproximaciones o paradigmas para responder a esta cuestión.

- El paradigma simbólico (Turing, 1950), que se centra en el desarrollo de sistemas que actúan de forma racional. Es decir, de algoritmos, sistemas expertos, que están, en general enfocados a la resolución de problemas concretos. Este paradigma fue el dominante desde los años 50 hasta la irrupción de las redes neuronales en los años 90.
- El paradigma conexionista (Medler, 1998), que pretende crear sistemas que imitan el cerebro humano. Es este paradigma el que se halla detrás de los desarrollos de redes neuronales actuales.

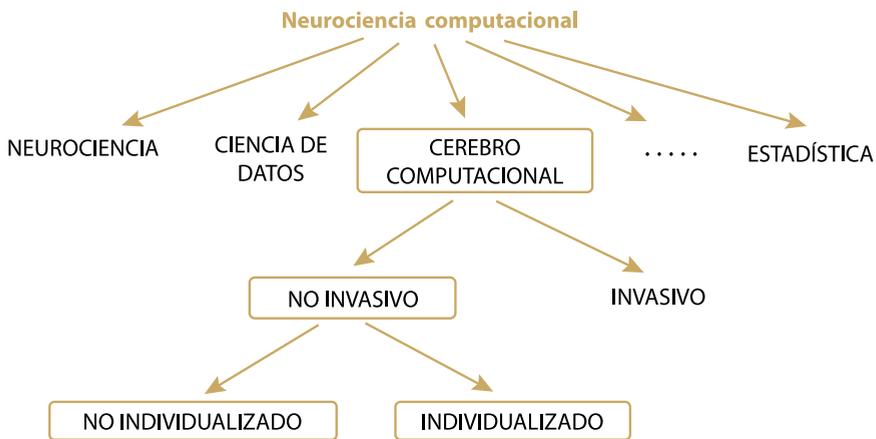
La neurociencia computacional se enmarca en el paradigma conexionista, ya que pretende profundizar en la comprensión del cerebro. Sin embargo, ambos paradigmas no son mutuamente excluyentes, ya que, como vamos a ver en el caso del cerebro computacional, también se va a hacer uso de técnicas y algoritmos propios del paradigma simbólico.

3. LA NEUROCIENCIA COMPUTACIONAL Y EL CEREBRO COMPUTACIONAL

La neurociencia computacional es un campo interdisciplinar que combina la neurociencia, la ciencia de datos, la informática, las matemáticas y la ingeniería para entender cómo funciona el cerebro y utilizar esta comprensión para desarrollar nuevas tecnologías (Trappenberg, 2010). Se trata, por tanto, de un campo muy amplio con numerosas ramificaciones y con un fuerte componente multidisciplinar. En la *figura 1* mostramos un esquema muy simplificado de su estructura, con algunos de sus elementos más relevantes.

Figura 1.

Esquema de la neurociencia computacional



Fuente: Elaboración propia.

Como ya hemos indicado, en este trabajo nos vamos a centrar específicamente en el problema del cerebro computacional. Es decir, en el estudio de interfaces cerebro-computador (*BCI*, por sus siglas en inglés), que permitan a los individuos controlar dispositivos externos con su actividad cerebral (Vidal, 1973).

Podemos decir, de forma simplificada, que existen dos métodos fundamentales para abordar los problemas del *BCI*: los invasivos (Polikov *et al.*, 2005) y los no invasivos (Bozinovski y Bozinovska, 2019). Empezamos proporcionando una breve visión de los métodos invasivos, considerando los desarrollos de la empresa Neuralink.

Los métodos invasivos consisten en utilizar dispositivos insertados dentro del propio cerebro para obtener información de la actividad de este. En la actualidad, el más conocido de estos métodos (aunque no el único existente) es Neuralink, desarrollado por la empresa del mismo nombre que fue fundada por Elon Musk en 2016 (<https://neuralink.com/>).

Según la información proporcionada por la propia empresa, Neuralink obtiene información de neuronas individuales, no de grupos de neuronas como sucede en los modelos no invasivos. En concreto, Neuralink utiliza un chip que incluye 64 cadenas flexibles de polímeros. Esto proporciona un total de 1.024 puntos que pueden utilizarse para registrar la actividad cerebral (Neuralink, 2019).

Aunque existen muchas dudas desde el punto de vista científico, la empresa ha informado de que el primer implante de uno de estos chips se llevó a cabo en enero de 2024 y de que, poco después, ya era posible detectar las señales de actividad neurológica del paciente. Más aún, la empresa añade que el dispositivo habría sido implantado en un segundo paciente en julio del mismo año, y que este paciente habría mejorado su capacidad para jugar a videojuegos y estaría entrenándose en el uso de *software* para el diseño de objetos 3D.

Con todo, es necesario mantener una actitud prudente ante estas afirmaciones. La comunidad científica ha expresado su preocupación por la falta de transparencia de estos desarrollos (Regalado, 2020). De hecho, el experimento fue inicialmente rechazado por la FDA (Food and Drug Administration, organismo responsable de autorizar ensayos clínicos y medicamentos en EE. UU.), aunque posteriormente sí dio su aprobación. En cualquier caso, es importante señalar que los pacientes no aparecen registrados en la base de datos de ensayos clínicos de Estados Unidos.

4. MÉTODOS NO INVASIVOS EN EL CEREBRO COMPUTACIONAL

Los métodos no invasivos tratan de obtener información por medio de dispositivos externos, como los electroencefalogramas. Entre estos métodos, son muy populares los basados en imagen motora. Es decir, métodos en los que el usuario imagina determinados movimientos que son transmitidos a los dispositivos externos y en los que las señales recibidas por estos deben utilizarse para recuperar el pensamiento del sujeto.

Dentro de los métodos no invasivos, y teniendo en cuenta el tipo de dispositivos a desarrollar, podemos distinguir dos grandes grupos:

- Los métodos que podemos denominar como no individualizados, en los que se trata de obtener información de diferentes sujetos para poder desarrollar dispositivos que sean, de alguna forma, universales. Esto es, aplicables, *a priori*, a cualquier individuo.
- Los métodos individualizados, que se centran en un sujeto concreto para desarrollar dispositivos específicos para dicho individuo.

Las técnicas utilizadas en los dos métodos son similares. De hecho, la principal diferencia entre ambos se halla en los datos que se utilizan para su desarrollo: datos provenientes de individuos diferentes en el primer caso, y de un solo individuo, en el segundo. Por ello, en este trabajo consideramos los métodos no individualizados.

En concreto, vamos a examinar el problema de identificar si un sujeto está pensando en mover su mano derecha o su mano izquierda, para el caso no individualizado, tal y como se discute en Ko *et al.* (2019). En primer lugar, describimos las principales etapas del tratamiento del problema.

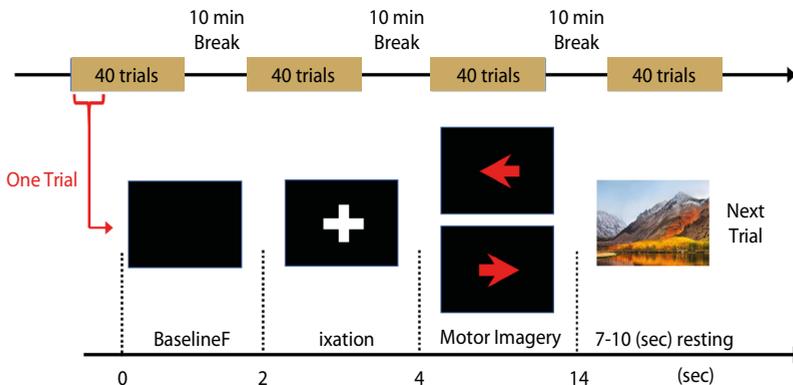
4.1. Captación de datos

El punto de partida, como en todos los problemas que hacen uso de técnicas de inteligencia artificial, son los datos. Tal y como se explica en Ko *et al.* (2019), en este caso los datos se obtienen en laboratorio. Concretamente, se consideran diez individuos con edades comprendidas entre los 18 y los 29 años, y que no presentan trastornos neurológicos. Cada participante se sitúa frente a un monitor y es conectado a un dispositivo individual que recoge las señales de su electroencefalograma. A cada uno de ellos se le pide pensar varias veces consecutivas en mover la mano derecha o la mano izquierda. Por medio de un electroencefalograma, se captan las señales eléctricas de la actividad neuronal cuando está pensando en el movimiento correspondiente.

Dado que estas señales son eléctricas, es necesario determinar qué frecuencias van a ser relevantes. Dependiendo del experimento, se utilizan varios canales distintos. La [figura 2](#) muestra en resumen el proceso de captación de datos.

Figura 2.

Procedimiento de captación de datos



Fuente: Ko *et al.* (2019).

En total, cada participante realizaba un total de 160 pruebas. Siguiendo la metodología estándar dentro del aprendizaje máquina, los datos asociados a cada una de estas pruebas se distribuyeron aleatoriamente en dos conjuntos disjuntos.

- 80 pruebas para entrenamiento, utilizadas para que el sistema aprenda a predecir si el sujeto está pensando en mover la mano derecha o la mano izquierda.
- 80 pruebas para test, que son utilizadas para comprobar la precisión del sistema una vez entrenado.

Es importante reseñar que, para que el problema esté equilibrado, cada uno de los conjuntos de 80 pruebas constaba de 40 ejemplos de pensar en mover la mano derecha y otros 40 de mover la mano izquierda.

4.2. Preprocesamiento de los datos

En general, en los problemas del cerebro computacional, como en la mayor parte de los problemas en inteligencia artificial, no es posible utilizar directamente los datos obtenidos. Es necesario someterlos a un proceso de "limpieza", conocido como preprocesamiento. En este problema en concreto, el preprocesamiento consta de dos etapas distintas:

- En primer lugar se aplica una transformada rápida de Fourier, para reducir la complejidad de los datos y descomponerlos en diferentes componentes de frecuencia.
- A continuación se usa *Common Spatial Pattern*, que es una técnica matemática habitualmente utilizada para separar las señales multivariadas en subcomponentes con máxima variación espacial (Pfurtscheller *et al.*, 1999).

De manera informal, el resultado de esta procesamiento es que las señales en cada banda vienen descritas por una serie de números que son la información (los datos) que va a utilizar nuestro sistema para predecir en qué está pensando el sujeto. Para cada sujeto y cada ensayo disponemos, por tanto, de cuatro conjuntos de números, correspondientes, respectivamente, a cada una de las bandas consideradas.

4.3. Clasificación

Este es el núcleo del método. Nuestro problema puede verse como un problema de clasificación: existen dos clases (mano derecha y mano izquierda) y para cada conjunto de datos (esto es, para cada individuo y cada prueba), debemos determinar a cuál de ellas pertenece. La pregunta es: ¿cómo podemos llevar a cabo esta clasificación?

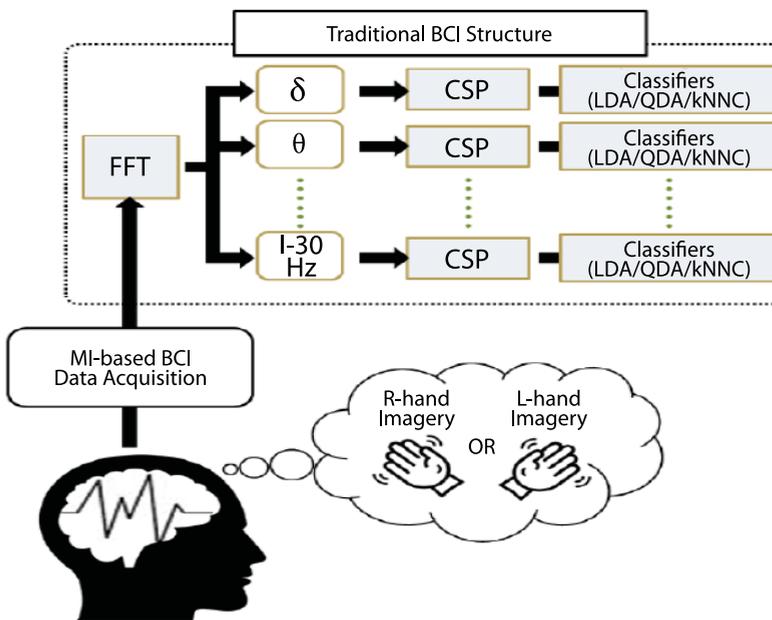
Existen muchos algoritmos de clasificación diferentes. Tres de los más conocidos son el análisis discriminante lineal (*LDA*), el análisis discriminante cuadrático (*QDA*) y la clasificación por los *k*-vecinos más próximos. La elección de estos métodos o de otros no es especialmente relevante para nuestra discusión, así que no vamos a entrar en muchos detalles acerca de los mismos.

Lo que sí es relevante es que entrenamos cada uno de los clasificadores con cada uno de los canales de cada uno de los ensayos. Esto es, para cada cada canal, enseñamos al clasificador a distinguir si estamos pensando en mover la mano derecha o la mano izquierda. De esta forma, una vez finalizado el entrenamiento, si introducimos una nueva señal en un canal dado, cada clasificador nos dirá la probabilidad de que el sujeto esté pensando en mover la mano derecha o la mano izquierda. Esto quiere decir que, para cada señal, podemos obtener varias respuestas posibles, una por cada clasificador. ¿Cómo podemos fusionarlas en la clasificación definitiva? Este es el paso clave en el estudio, que se realiza por medio de las denominadas funciones de fusión de datos.

La **figura 3** muestra un esquema del proceso completo descrito hasta aquí.

Figura 3.

Esquema del proceso tradicional de clasificación de los datos



Fuente: Ko et al. (2019).

5. LA FUSIÓN DE DATOS

5.1. Funciones de fusión de datos

Esta subsección es la más técnica desde un punto de vista matemático, pero es fundamental para poder entender plenamente los desarrollos posteriores del trabajo. Sin embargo,

aquellos lectores interesados en una visión más general del problema pueden omitirla y simplemente referirse a ella cuando se den definiciones concretas.

La fusión de datos es un proceso fundamental en casi todos los campos de la investigación. De una forma informal, podemos definirla como el proceso de obtener, a partir de un conjunto de datos homogéneos o heterogéneos, un nuevo valor que represente a todos ellos de la forma más precisa posible.

Uno de los métodos más utilizados para la fusión de información se basa en el uso de las llamadas funciones de agregación (Grabisch *et al.*, 2009). Una función de agregación es una función creciente que toma valores en el intervalo $[0,1]$ y devuelve otro valor también en el intervalo $[0,1]$, de modo, que, si todos los valores de entrada son 0, el valor devuelto es 0; y, si todos los valores de entrada son 1, el valor devuelto es 1.

Entre los ejemplos más conocidos de funciones de agregación se encuentran las medias (Beliakov *et al.*, 2016), como la media aritmética o la media geométrica, el máximo o el mínimo. Todos estos ejemplos consideran los datos de entrada individualmente. Esto es, no tienen en cuenta en general la posible relación que pueda existir entre datos diferentes, y que puede haber sido determinada de forma experimental. Sin embargo, en muchas ocasiones los datos no son independientes, sino que pueden reforzarse o inhibirse entre sí. Esto es particularmente cierto en el caso que nos ocupa del cerebro computacional, ya que parece razonable suponer que los diversos canales de la señal eléctrica tienen algún tipo de conexión entre ellos que va más allá de la mera correlación. Es decir, que la información de algunos canales es más significativa tomada junto a la de otros, que de forma individual, en especial a partir de determinados umbrales para los valores de la señal.

Para poder considerar la fusión de este tipo de datos que pueden estar relacionados entre sí, se considera toda una familia de funciones de agregación basadas en el concepto de medida difusa. Una medida difusa es simplemente una función que asigna un valor entre $[0,1]$ a cada posible subconjunto de los datos, con algunas restricciones. Si el subconjunto considerado es vacío, es decir, si no consideramos ningún dato, la medida toma el valor cero. Si, por el contrario, consideramos todos los datos disponibles a la vez, la medida asignada es 1; es decir, máxima. Finalmente, también se impone que si un subconjunto de datos A está incluido en otro subconjunto de datos B, la medida de A no puede ser mayor que la medida de B. De esta forma, podemos utilizar la medida para determinar hasta qué punto datos distintos se refuerzan entre sí.

¿Qué funciones de agregación podemos construir por medio de estas medidas? Aunque existen muchas posibilidades, las dos más utilizadas, y las que consideramos en este trabajo, son la integral Choquet (Choquet, 1953-54) y la integral Sugeno (Sugeno, 1974), que pasamos a describir brevemente.

Consideremos un conjunto de valores x_1, \dots, x_n en el intervalo $[0,1]$. Vamos a ordenar estos valores de forma creciente. Es decir los reordenamos por medio de una permutación $(\cdot): \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ de forma que:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} .$$

Por ejemplo, si tomamos $n=3$ y $x_1=0.6$, $x_2=0.2$, $x_3=0.4$, para reordenar debemos considerar la permutación $(1)=2$, $(2)=3$ y $(3)=1$, con lo que nos queda:

$$x_{(1)} = x_2 = 0.2 < x_{(2)} = x_3 = 0.4 < x_{(3)} = x_1 = 0.6.$$

tal y como queríamos. Este ordenamiento nos va a permitir ir considerando subconjuntos teniendo en cuenta qué datos son mayores o menores y la medida va a tener en cuenta este tamaño relativo de los datos, permitiendo dar importancia a los términos mayores. Para ello, vamos a definir, para cada $i=1, \dots, n$, $A_{(i)}$ como el conjunto de índices (originales) de los elementos mayores o iguales que $x_{(i)}$. Es decir, con el ejemplo anterior:

$$A_{(1)} = \{1, 2, 3\}, \quad A_{(2)} = \{1, 3\}, \quad A_{(3)} = \{1\}.$$

El objetivo de estas notaciones, que pueden parecer un poco complejas, es poder introducir las dos funciones siguientes. En primer lugar, definimos una medida m como una función que, dado un subconjunto de $\{1, \dots, n\}$, le asocia un valor entre 0 y 1 y que cumple las dos propiedades siguientes:

1. $m(\{1, \dots, n\})=1$ y $m(\emptyset)=0$.
2. Si $A \subseteq B$ entonces $m(A) \leq m(B)$.

Sin embargo, lo que no se exige en general es que la medida de una unión de dos conjuntos disjuntos sea la suma de sus medidas correspondientes. Este hecho nos permite tener en cuenta el posible reforzamiento de la información al tener en cuenta varios datos distintos. En efecto, si en un conjunto tenemos solo dos elementos a y b , esta definición nos permite decir que la medida del conjunto $\{a, b\}$ nunca va a ser más pequeña que las de a y b por separado, pero puede ser mayor. Usando un símil futbolístico, si tenemos dos delanteros de calidad, su efectividad es mayor si además resulta que son capaces de colaborar de forma adecuada.

Utilizando este concepto de medida, podemos considerar las dos definiciones siguientes:

- La integral Choquet de un conjunto de números x_1, \dots, x_n en $[0, 1]$ con respecto a la medida m se define como:

$$C_m(x_1, \dots, x_n) = \sum_{i=1}^n (x_{(i)} - x_{(i-1)}) m(A_{(i)})$$

donde establecemos que $x_{(0)}=0$.

- La integral Sugeno de un conjunto de números x_1, \dots, x_n en $[0, 1]$ con respecto a la medida m se define como:

$$S_m(x_1, \dots, x_n) = \max_{i=1}^n \min(x_{(i)}, m(A_{(i)}))$$

Es importante observar que en los últimos sumandos únicamente se tienen en cuenta para la medida, los mayores valores entre los datos de entrada, de modo que se puede valo-

rar la relación entre estos. Por ejemplo, si los valores más altos se obtienen para dos señales concretas, la medida permite resforzar su influencia en el resultado final. Experimentalmente, se ha comprobado que estas integrales proporcionan muy buenos resultados en problemas diversos, que van desde el procesamiento de imagen (Paternain, 2015) a la clasificación (Barrenechea *et al.*, 2013) o la toma de decisión (Lourenzutti *et al.*, 2017) cuando se comparan con las funciones de agregación más tradicionales (medias, productos...). De hecho, también es importante destacar que esto es debido, en parte, a que funciones como las medias ponderadas pueden recuperarse como ejemplos específicos de estas funciones, luego estamos considerando una familia más general que incluye las funciones usadas tradicionalmente en la literatura.

Sin embargo, en algunas ocasiones, estas funciones no captan adecuadamente la complejidad del problema a considerar. Este hecho ha llevado a investigar la forma de extenderlas o generalizarlas, de modo que, por una parte, se preserven algunas de sus propiedades más relevantes, y, por otra parte, se obtenga una mayor flexibilidad para tratar con problemas complejos. En particular, para el problema concreto del cerebro computacional, se han considerado generalizaciones de ambas integrales. En concreto, se trata de reemplazar el producto y el mínimo (y, por tanto, la linealidad) por funciones más generales que induzcan un comportamiento no lineal en la definición. Como muestra de estas generalizaciones, podemos dar los dos ejemplos siguientes:

1. Respecto a la integral Choquet, podemos hacer uso de la propiedad distributiva del producto y reescribir la integral Choquet como:

$$C_m(x_1, \dots, x_n) = \sum_{i=1}^n x_{(i)} m(A_{(i)}) - x_{(i-1)} m(A_{(i)})$$

Ahora, si reemplazamos el producto en cada uno de los términos de la derecha por sendas funciones F_1, F_2 con propiedades adecuadas, obtenemos las $CF1F2$ -integrales (Dimuro, 2020), definidas, para x_1, \dots, x_n en $[0,1]$, por:

$$C_m^{F_1, F_2}(x_1, \dots, x_n) = \min(1, \sum_{i=1}^n F_1(x_{(i)}, m(A_{(i)})) - F_2(x_{(i-1)}, m(A_{(i)})))$$

Obsérvese que introducimos el mínimo al comienzo simplemente para asegurar que el valor obtenido se encuentra entre 0 y 1. Las funciones F_1 y F_2 consideradas proporcionan mayor flexibilidad. Algunos ejemplos de funciones que podemos utilizar en esta construcción son $F_1(x, y) = \sqrt{xy}$ y $F_2(x, y) = \max(0, x+y-1)$, entre otras muchas.

2. De forma similar, podemos reemplazar el mínimo en la definición de integral de Sugeno por otra función M de dos variables y obtener la siguiente generalización (Bardozzo *et al.*, 2021):

$$S_m^M(x_1, \dots, x_n) = \max_{i=1}^n M(x_{(i)}, m(A_{(i)}))$$

Como función M en este caso podemos tomar, por ejemplo, el producto o las mismas funciones que en el caso anterior.

Es importante destacar que, al llevar a cabo estas generalizaciones, debemos pagar un cierto precio. En particular, perdemos la monotonía. Esto es, si tenemos un conjunto de datos

A y otro conjunto de datos B y todos los datos en A son mayores o iguales que los correspondientes datos en B , no tienen por qué cumplirse que el resultado de aplicar estas funciones a los datos de A sea mayor o igual que el resultado obtenido al aplicarlas a los datos de B . Sin embargo, esta propiedad de monotonía no se pierde completamente, ya que sigue siendo cierta si los datos crecen de una forma concreta, siguiendo una dirección o un conjunto de direcciones específicas. Esto es lo que se denomina monotonía direccional. Las funciones, como estas, que satisfacen esta forma menos restrictiva de monotonía y que toman el valor 0 si todos los valores de entrada son 0, y el valor 1 si todos los valores de entrada son 1, se denominan preagregaciones (Lucca *et al.*, 2016). Las preagregaciones están atrayendo un considerable interés en los últimos tiempos, ya que permiten mejorar los resultados obtenidos con muchos algoritmos que hacen uso de agregaciones. En este trabajo utilizaremos el término de función de fusión de datos o información para referirnos tanto a funciones de agregación como a preagregaciones.

5.2. La fusión difusa multimodal de datos en el cerebro computacional

El uso de las funciones de fusión de datos introducidas en la subsección anterior tiene como objetivo mejorar los resultados que puede proporcionar cada uno de los clasificadores individualmente. Visto de otra forma, se trata de aprovechar la información obtenida de todos los clasificadores considerados para lograr una mejor predicción en el problema. La pregunta es: ¿cómo podemos llevar a cabo esta fusión de datos?

Existe una amplia literatura sobre los denominados ensembles. Es decir, sobre cómo llevar a cabo esta fusión de los datos provenientes de diversos clasificadores (Bolón-Canedo y Alonso-Betanzos, 2019). En este trabajo, vamos a considerar la aproximación a este problema dada en Ko *et al.* (2019), y denominada Fusión Difusa Multimodal, y que lleva a cabo la fusión de la información en varias etapas, como se describe más abajo. La elección de este método es debida a los buenos resultados obtenidos. Se trata, por tanto, de una elección empírica, y, a día de hoy, se carece de una justificación completa para estos buenos resultados.

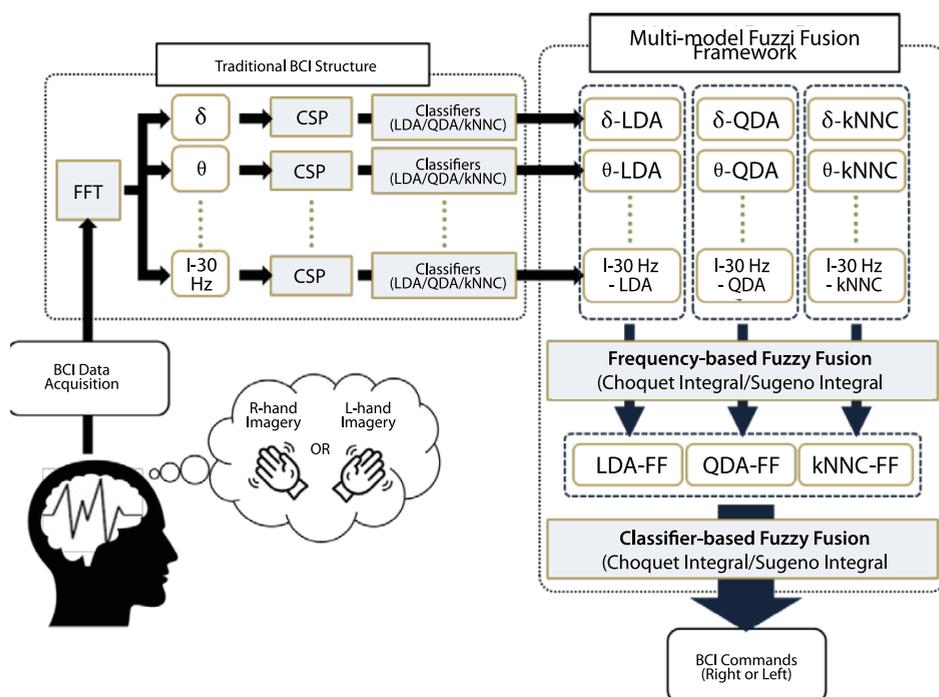
La Fusión Difusa Multimodal consta de tres etapas:

1. En primer lugar, para cada clasificador, obtenemos el resultado que dicho clasificador proporciona para cada uno de los canales por separado.
2. A continuación, también para cada clasificador, fusionamos los resultados obtenidos en el paso 1. Es decir, para cada clasificador obtenemos un único resultado (probabilidad de que el sujeto esté pensando en mover la mano derecha y probabilidad de que esté pensando en mover la mano izquierda). Al completar este paso, tenemos, por tanto, tantos resultados como clasificadores.
3. Finalmente, fusionamos los resultados de todos los clasificadores considerados, para obtener una respuesta definitiva.

En la **figura 4** mostramos el esquema completo de este algoritmo, añadiendo los nuevos pasos de función de información al esquema tradicional mostrado en la **figura 3**.

Figura 4.

Esquema del proceso completo de clasificación de los datos



Fuente: Ko *et al.* (2019).

Es importante destacar que en las dos últimas etapas, en que se lleva a cabo la fusión de datos, no tiene por qué utilizarse la misma función.

Desde el punto de vista teórico, el método propuesto tiene sentido y parece razonable. Sin embargo, ¿proporciona alguna mejora el hecho de considerar fusión de datos en dos etapas distintas? En Ko *et al.* (2019) se considera la respuesta a esta pregunta. Los resultados de esta comparación que se pueden encontrar en ese trabajo reflejan que, en general se obtienen una mejora de la precisión frente a la obtenida fusionando solo los resultados de todos los clasificadores o la precisión fusionando solo los resultados de las bandas para cada clasificador. Además, la tasa de acierto obtenida se aproxima a un 75 %.

Ahora bien, para poder valorar la calidad del método, es necesario tener en cuenta que este ha sido desarrollado con un conjunto de datos muy específico obtenido en laboratorio y que tanto las funciones como las medidas han sido escogidas *ad hoc*. ¿Qué sucede si se trata de extender el mismo método a otros conjuntos de datos, provenientes del mundo real?

En Fumanal-Idocin *et al.* (2024) se incluye una interesante discusión sobre los resultados de aplicar este mismo método a un problema más complejo: determinar si el individuo está pensando en mover la mano izquierda, la mano derecha, un pie o la lengua. Aunque formalmente el problema es muy similar al anterior, la precisión del método cae drásticamente a valores alrededor de un 60 o 65 % en el mejor de los casos. Esto claramente indica que existe un amplio margen de mejora en el algoritmo y plantea una cuestión: ¿a qué puede deberse esta pérdida de precisión?

Obviamente, esta pregunta no admite una respuesta simple ni única. A partir de los resultados en Fumanal-Idocin *et al.* (2024), vamos a discutir una de las posibles causas del problema y vamos a ver cómo podemos modificar el método de fusión de información para tenerla en cuenta.

5.3. El problema de la incertidumbre

En todo problema de fusión de información, uno de los principales factores que pueden afectar al carácter más o menos óptimo de la solución obtenida es el de la calidad de los datos disponibles. De hecho, no es exagerado afirmar que la base del éxito de cualquier aplicación de la inteligencia artificial se halla, precisamente, en los datos utilizados. No es casualidad que la explosión de desarrollos de inteligencia artificial que hoy vivimos haya venido de la mano del desarrollo de técnicas para poder tratar computacionalmente grandes cantidades de datos.

Cuando consideramos una aplicación concreta, como puede ser el cerebro computacional, hay dos preguntas fundamentales a responder relativas a los datos:

1. ¿De cuántos datos disponemos? Algunos métodos, como las redes neuronales profundas, solo proporcionan resultados óptimos si se dispone de cantidades suficiente mente grandes de datos para entrenarlas. Esto hace que su aplicación en algunos problemas concretos pueda resultar muy compleja o incluso ineficiente.
2. ¿Cómo de buenos son los datos de los que disponemos? En ocasiones, se considera que disponer de datos es equivalente a disponer de largas tablas de valores. Sin embargo, la realidad es más compleja. En primer lugar, existe la posibilidad de sesgos, es decir, que los valores de los que disponemos (si nos centramos en el caso numérico) no reflejen adecuadamente todos los casos de interés, o que sobre-representen algunos ejemplos o clases frente a otros. Pero, más aún, aunque se extreme el cuidado para evitar esta situación, puede ser que los datos no sean totalmente correctos. En particular, si los datos se obtienen a partir de sensores (utilizando esta palabra en su sentido más general) y si utilizamos ordenadores para su tratamiento, nunca vamos a poder garantizar la exactitud de los valores con los que trabajamos. Es decir, nuestros datos van a estar afectados de incertidumbre.

La incertidumbre, entendida como la imposibilidad de conocer el valor exacto de un dato, es inherente a la ciencia de datos, y por ende, a todos los procesos que utilizan datos. En el caso concreto del cerebro computacional, recordemos que los datos se obtienen utilizando electroencefalogramas fijados a la superficie de la cabeza. Esto implica que las señales eléctricas

cas captadas por estos dispositivos están inevitablemente afectadas de una importante cantidad de ruido. Es decir, que los valores que estamos midiendo, no son los reales, sino que son perturbaciones de esos valores reales, y no tenemos forma de conocer cuál es el valor exacto de los mismos. Dado que nuestros algoritmos hacen uso de estos valores, esta incertidumbre debe afectar al resultado final. Por tanto, una posible vía de mejora de los algoritmos es reducir la cantidad de ruido para tratar de obtener datos para nuestro uso los más próximos posibles a los datos reales. Esta es una aproximación seguida en muchos casos. Así, por ejemplo, es posible captar las señales dentro de una jaula de Faraday, para evitar distorsiones debidas a campos electromagnéticos externos, o intentar mejorar la sensibilidad de los sensores utilizados.

Pero en muchas ocasiones, no es posible eliminar completamente el ruido, por lo que debemos asumir que está presente. ¿Qué podemos hacer en este caso? Una salida es tenerlo en cuenta a la hora de fusionar la información. O, más concretamente, incluirlo dentro del proceso de fusión incorporándolo de alguna manera a la representación de los datos que consideramos.

¿Cómo es posible hacer esto? Una forma muy simple es por medio de intervalos. La idea no es en sí misma original ya que, de hecho, el uso de intervalos es habitual en campos como la Estadística. Se trata de una idea natural y simple de desarrollar. Supongamos que hay un dato real, que no conocemos, pero que podemos acotar. Para fijar las ideas, supongamos que el dato real es 0.3, pero que no lo conocemos, porque la medición que nosotros hacemos para obtener ese dato está afectada por ruido. Lo que podemos tratar de hacer es dar una cota inferior y una cota superior para ese valor real que no conocemos. Así, podemos decir que el dato está en el intervalo $[0.2, 0.5]$. El punto clave es darnos cuenta de que podemos ir un paso más allá y representar el dato, desconocido por ese intervalo. Es decir, el dato con el que vamos a trabajar es, directamente, el intervalo $[0.2, 0.5]$.

¿Cómo obtenemos este intervalo? Aunque existen métodos muy elaborados, es posible hacerlo de una manera muy simple. Podemos, por ejemplo, realizar varias mediciones y considerar como el intervalo resultante aquel dado por el menor y el mayor de los valores medidos. También es posible diseñar funciones específicas que dado un dato concreto, le asocien un intervalo, de forma similar a como se pueden construir intervalos de confianza en Estadística. No vamos a entrar en estos detalles, porque nos alejarían del objetivo de esta exposición.

Sin embargo, en la representación escogida del valor por medio del intervalo, hay un elemento esencial a tener en cuenta. Queremos representar, también, la incertidumbre asociada a ese mismo dato. Para entender este punto, imaginemos que todos los datos con los que trabajamos están en el intervalo $[0,1]$. Si un dato concreto x viene representado por el intervalo $[0.34, 0.35]$, estamos bastante seguros de cuál es el valor real. Sin embargo, si otro dato y viene representado por el intervalo $[0,1]$, estamos diciendo que de hecho no tenemos ni idea de cuál es el valor real. En otras palabras, podemos considerar que la amplitud del intervalo que utilizamos para representar un dato concreto (la diferencia entre el extremo superior y el extremo inferior del intervalo) es una medida de la incertidumbre que tenemos respecto al valor real de dicho dato. La construcción de intervalos para reemplazar a un dato teniendo en cuenta la incertidumbre ligada a él puede hacerse de numerosas maneras diferentes. Por ejemplo, es posible realizar diferentes mediciones de la misma señal y considerar el intervalo entre el menor

valor obtenido y el mayor valor obtenido. O puede asumirse que la incertidumbre depende del valor medido. por ejemplo, experimentalmente puede determinarse que, para valores muy próximos a cero podemos asumir, en función de las características de nuestros dispositivos que existe una incertidumbre muy pequeña, y que esta va creciendo a medida que aumentan los valores observados; o viceversa. Lo importante es que, al final, transformamos cada dato en un intervalo.

Desde un punto de vista matemático, la naturaleza de los intervalos es diferente a los de los números reales. En particular, si nos restringimos a subintervalos del intervalo unidad $[0,1]$, hay dos puntos clave que deben tenerse en cuenta a la hora de extender el método que hemos explicado.

- En primer lugar, dados dos números reales distintos cualesquiera, uno siempre es menor o igual que el otro. Sin embargo, esto no es cierto dados dos intervalos, ya que no existe un orden "natural" para compararlos. Esto representa un problema, ya que las integrales difusas que hemos utilizado necesitan ordenar primero los datos de entrada.
- Además, en el caso concreto de la integral Choquet, es necesario considerar diferencias entre dos datos. Pero la diferencia entre intervalos tampoco está definida, en general.

Afortunadamente, es posible superar ambas objeciones gracias a desarrollos teóricos recientes. Así, el concepto de orden admisible (Bustince *et al.*, 2013) define una relación de orden entre intervalos que es total. Es decir, tal que permite comparar dos intervalos cualquiera para decidir cuál es mayor y cuál es menor. Estos órdenes admisibles no son únicos, sin embargo, y diferentes órdenes pueden dar lugar a resultados distintos. Esta es una línea en la que se está investigando en la actualidad.

En cuanto al segundo punto, podemos considerar que la diferencia de dos números es una medida de cómo son de distintos. Por tanto, para llevar a cabo un proceso similar con intervalos, una posibilidad es introducir una función, llamada de dismilitud, que devuelva un valor numérico de cómo de diferentes son los intervalos considerados. Esta idea ha dado lugar a la introducción de la noción de d-integral (Bustince *et al.*, 2021).

En resumen, es posible reproducir el método presentado anteriormente trasladando las correspondientes funciones de fusión de datos al marco intervalar. Este estudio ha sido llevado a cabo en Fumanal-Idocin *et al.* (2021) y en Fumanal-Idocin *et al.* (2022), donde se comprueba que el uso de intervalos, en combinación con una adaptación intervalar de la integral de Sugeno, permite mejorar los resultados significativamente. En concreto, y para comprobar la validez del método, se utilizan datos del dataset CBCIC (Chowdhury y Andreu-Perez, 2021). Este contiene datos de diez pacientes afectados de ictus cerebral que les ha provocado problemas de movilidad en las manos. Para cada paciente se incluyen datos de 80 pruebas pensando en mover la mano derecha o la mano izquierda. Se trata de un dataset estándar dentro de la literatura del cerebro computacional y que ha sido utilizado, por ejemplo, en la competición del Congreso Internacional WCCI 2020 para evaluar los resultados de diferentes algoritmos.

En la **tabla 1** se muestran los resultados alcanzados, comparando los mejores resultados obtenidos por el método presentado con intervalos, y los mejores resultados obtenidos con otros métodos del estado del arte. En concreto, comparamos el método basado en el uso de la integral de Sugeno intervalovalorada con otros métodos del estado del arte. Se observa que el método intervalo valorado logra un mayor porcentaje de datos correctamente clasificados (segunda columna) con una mejor precisión (tercera columna).

Tabla 1.

Resultados con el dataset CBCIC

<i>Modelo</i>	<i>Precisión</i>	<i>F1-score</i>
IV Sugeno	0.8175 ± 0.1342	0.8149 ± 0.1366
EEG Net (Lawhern <i>et al.</i> , 2018)	0.6562 ± 0.1232	0.5933 ± 0.1712
Shallow Net (Tibor <i>et al.</i> , 2017)	0.7453 ± 0.13289	0.7342 ± 0.1489
Deep Net (Tibor <i>et al.</i> , 2017)	0.5331 ± 0.1356	0.4218 ± 0.1282
Multiscale CSP (Hersche <i>et al.</i> , 2018)	0.7956 ± 0.1144	0.7911 ± 0.1175
Gradient Boosting (Vijay <i>et al.</i> , 2020)	0.5956 ± 0.1203	0.5354 ± 0.1169

Fuente: Fumanal-Idocin *et al.* (2022).

Merece la pena destacar que algunos de los métodos del estado del arte considerado en esta tabla, como EEG Net, Shallow net y Deep Net, son, de hecho, redes neuronales profundas, a las que el método con intervalos considerado supera. Como ya hemos comentado, en el problema del cerebro computacional las redes neuronales profundas suelen carecer de datos suficientes para dar un rendimiento óptimo. Este es el caso, sobre todo, para los métodos individualizados.

5.4. El problema con cuatro clases: escogiendo la mejor función de fusión

Como ya hemos señalado anteriormente, los resultados empeoran significativamente si los sujetos, en lugar de pensar solo en mover la mano derecha o la mano izquierda, pueden pensar también en mover la lengua o uno de los pies. En este caso, incluso el uso de intervalos para considerar la incertidumbre no parece ser suficiente para obtener resultados suficientemente competitivos.

Si queremos considerar este problema, es necesario tener en cuenta un aspecto que, hasta ahora, hemos soslayado. ¿Es nuestro resultado indiferente a la función o funciones escogidas para llevar a cabo la fusión de los datos? Los experimentos muestran claramente que no es así, y que, de hecho, la función escogida puede influir significativamente. Pero entonces, ¿cómo podemos escoger la función más apropiada?

En general, no existe una respuesta a esta pregunta, ya que son todavía necesarios muchos estudios en profundidad que permitan clarificar la naturaleza del problema desde un punto de vista matemático para determinar las propiedades más deseables de la función de fusión. De hecho, el procedimiento más habitual actualmente es el ensayo y error. Utilizamos diversas funciones y nos quedamos con aquella o aquellas que mejor resultado dan.

Sin embargo, en la actualidad se está explorando una vía para intentar avanzar hacia una respuesta más satisfactoria. La idea es recordar que el objetivo de las funciones de agregación, en particular, y de las funciones de fusión, en general, es tratar de obtener un único valor (real o intervalar) que represente de la foma más adecuada posible los datos de entrada considerados. Esto puede interpretarse de la manera siguiente: queremos escoger como resultado de la fusión un valor que sea lo "más parecido" posible a los datos de entrada. La clave está en dar un significado apropiado a la expresión entrecomillada.

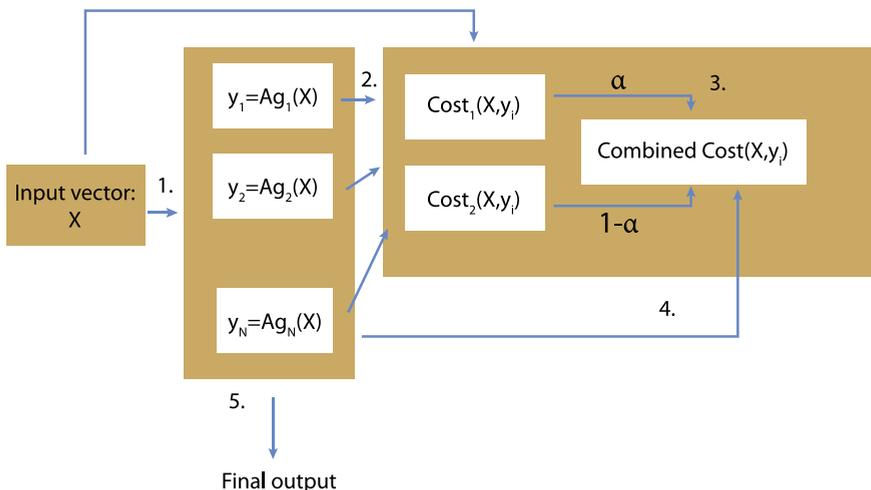
Esta idea está detrás del desarrollo de dos ideas en la literatura de inteligencia artificial: las funciones penalty (Calvo y Beliakov, 2010; Bustince *et al.*, 2017) y las desviaciones moderadas (Altalhi *et al.*, 2019). Sin entrar en detalles excesivamente técnicos, la idea detrás de ambas es similar y sería la siguiente. Supongamos que tenemos un conjunto de valores x_1, \dots, x_n a fusionar y un conjunto de funciones M_1, \dots, M_k que podemos utilizar. Cada una de estas funciones nos proporciona un valor y_1, \dots, y_k como resultado de la fusión. Pues bien, escogemos como valor final el y_j más parecido a los x_1, \dots, x_n . Por fijar ideas, este valor puede ser el que minimiza el error cuadrático medio:

$$\frac{1}{n} \sum_{i=1}^n (x_i - y_j)^2,$$

aunque en general las funciones penalty y de las desviaciones moderadas proporcionan aproximaciones más ricas a la idea de "mayor parecido". Esquemáticamente, el método se muestra en la [figura 5](#), donde las funciones de comparación aparecen como funciones de coste que se combinan entre sí por medio de un parámetro α para afinar los resultados (Fumanal-Idocin *et al.*, 2024).

Figura 5.

Esquema de uso de funciones penalty y desviaciones moderadas



Fuente: Fumanal-Idocin *et al.* (2024).

Un aspecto a destacar del uso de funciones penalty y desviaciones moderadas es que la selección del valor representativo se realiza cada vez que es necesario llevar a cabo un proceso de fusión. Es decir, a la hora de fusionar los datos de la señal 1 del paciente 1, el método puede escoger el resultado obtenido con una función M_1 , pero al fusionar los datos de la señal 2 del mismo paciente, elegir el resultado obtenido con una función distinta M_2 . De esta forma se pueden tener en cuenta posibles diferencias en la naturaleza de los datos, incluyendo diferentes niveles de incertidumbre. Tal y como se muestra en (Fumanal-Idocin *et al.*, 2024), esta metodología proporciona resultados cercanos al estado del arte cuando se considera el problema de las cuatro clases discutido anteriormente.

Hay que tener en cuenta que esta es una línea de investigación en desarrollo, y que para poder explotarla plenamente se necesita un mejor conocimiento de la naturaleza de las señales electroencefalográficas para poder determinar como debe medirse de la forma más adecuada el parecido. Además, desde el punto de vista computacional, este es un método muy costoso, ya que, para obtener un resultado óptimo, es deseable disponer de un amplio conjunto de funciones de fusión de información, que es necesario utilizar una y otra vez para evaluar los datos disponibles.

6. CONCLUSIONES

En este trabajo nos hemos centrado en el problema de identificar en qué está pensando un individuo a través de las señales de un encefalograma. En concreto, hemos considerado el problema de determinar si está pensando en mover la mano derecha o la mano izquierda, y el problema más complejo de si está pensando en mover una de las manos, la lengua o un pie. Este problema nos ha permitido presentar dos de los puntos clave en el desarrollo de la inteligencia artificial actual: la forma de fusionar datos para obtener información útil, y el problema de la calidad de esos mismos datos. En concreto, en el cerebro computacional, los datos están fuertemente afectados de incertidumbre, y es necesario tener en cuenta esta incertidumbre.

Por supuesto, el problema considerado admite otros abordajes, como pueden ser entender las señales del electroencefalograma como una serie temporal o diseñar redes neuronales profundas *ad hoc*. Así mismo, las técnicas que hemos discutido en estas páginas son, en general, aplicables a un amplio rango de problemas de la inteligencia. El cerebro computacional proporciona un marco muy apropiado para entender su relevancia sin necesidad de desarrollos técnicos muy complejos para el lector no experto. En cualquier caso, se trata de un área de investigación en plena ebullición en la actualidad y es seguro que los próximos años darán a lugar a desarrollos muy relevantes.

Referencias

- ALTALHI, A. H., FORCÉN, J. I., PAGOLA, M., BARRENECHEA, E., BUSTINCE, H., y TAKAC, Z. (2019). Moderate deviation and restricted equivalence functions for measuring similarity between data. *Information Sciences*, 501, 19–29.

- BARDOZZO, F., DE LA OSA, B., HORANSKÁ, L., FUMANAL-IDOCIN, J., DELLI PRISCOLI, M., TROIANO, L., TAGLIAFERRI, R., FERNANDEZ, J., y BUSTINCE, H. (2021). Sugeno integral generalization applied to improve adaptive image binarization. *Information Fusion*, 68, 37–45.
- BARRENECHEA, E., BUSTINCE, H., FERNÁNDEZ, J., PATERNAIN, D., y SANZ, J. A. (2013). Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems. *Axioms*, 2(2), 208–223.
- BELIAKOV, G., BUSTINCE, H., y CALVO, T. (2016). *A practical guide to averaging functions*. Springer.
- BOLÓN-CANEDO, V., y ALONSO-BETANZOS, A. (2019). Ensembles for feature selection: A review and future trends. *Information Fusion*, 52, 1–12.
- BOZINOVSKI, S., y BOZINOVSKA, L. (2019). Brain-computer interface in Europe: The thirtieth anniversary. *Automatika*, 60(1), 36–47.
- BUSTINCE, H., BELIAKOV, G., DIMURO, G.P., BEDREGAL, B., y MESIAR, R. (2017). On the definition of penalty functions in data aggregation. *Fuzzy Sets and Systems*, 323, 1–18.
- BUSTINCE, H., FERNANDEZ, J., KOLESÁROVÁ, A., y MESIAR, R. (2013). Generation of linear orders for intervals by means of aggregation functions. *Fuzzy Sets and Systems*, 220, 69–77.
- BUSTINCE, H., MESIAR, R., FERNÁNDEZ, J., GALAR, M., PATERNAIN, D., ALTALHI, A., DIMURO, G. P., BEDREGAL, B., y TAKAC, Z. (2021). d-Choquet integrals: Choquet integrals based on dissimilarities. *Fuzzy Sets and Systems*, 414, 1–27.
- CALVO, T., y BELIAKOV, G. (2010). Aggregation functions based on penalties. *Fuzzy Sets and Systems*, 161(10), 1420–1436.
- CHOQUET, G. (1953–54). Theory of capacities. *Annales de l'Institut Fourier*, 5, 131–295.
- CHOWDHURY, A., y ANDREU-PEREZ, J. (2021). Clinical brain-computer interface challenge 2020 (cbcic at wcci2020): Overview, methods and results. *IEEE Transactions on Medical Robotics and Bionics*, 3(3), 661–670.
- DAYAN, P., y ABBOTT, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, Mass: MIT Press.
- DIMURO, G. P., LUCCA, G., BEDREGAL, B., MESIAR, R., SANZ, J. A., LIN, C. T., y BUSTINCE, H. (2020). Generalized CF1F2-integrals: from Choquet-like aggregation to ordered directionally monotone functions. *Fuzzy Sets and Systems*, 378, 44–67.
- EU ARTIFICIAL INTELLIGENCE ACT. (2024). Ley de inteligencia artificial de la UE (2024). <https://artificialintelligenceact.eu/es/>
- FUMANAL-IDOCIN, J., TAKAC, Z., FERNÁNDEZ, J., SANZ, J. A., GOYENA, H., LIN, C. T., WANG, Y.-K., y BUSTINCE, H. (2021). Interval-valued aggregation functions based on moderate deviations applied to motor-imagery-based brain-computer interface. *IEEE Transactions on Fuzzy Systems*, 30(7), 2706–2720.
- FUMANAL-IDOCIN, J., TAKAC, Z., HORANSKÁ, L., ASMUS, T., DIMURO, G., VIDAURRE, C., FERNANDEZ, J., y BUSTINCE, H. (2022). A generalization of the Sugeno integral to aggregate interval-valued data: an application to brain computer interface and social network analysis. *Fuzzy Sets and Systems*, 451, 320–341.
- FUMANAL-IDOCIN, J., VIDAURRE, C., FERNÁNDEZ, J., GÓMEZ, M., ANDREU-PEREZ, J., PRASAD, M., y BUSTINCE, H. (2024). Supervised penalty-based aggregation applied to motor-imagery based brain-computer-interface. *Pattern Recognition*, 145, 109924.
- GRABISCH, M., MARICHAL, J. L., MESIAR, R., y PAP, E. (2009). *Aggregation Functions*. Cambridge, U.K.: Cambridge University Press.

- HERSCHE, M., RELLSTAB, T., SCHIAVONE, P. D., CAVIGELLI, L., BENINI, L., y RAHIMI, A. (2018). Fast and accurate multiclass inference for MI-, BCIs using large multiscale temporal and spectral features. En *2018 26th European Signal Processing Conference (EUSIPCO)*, 1690–1694.
- KO, L.-W., LU, Y. C., BUSTINCE, H., CHANG, Y.-C., CHANG, Y., FERNANDEZ, J., WANG, Y.-K., SANZ, J. A., DIMURO, G. P., y LIN, C.-T. (2019). Multimodal fuzzy fusion for enhancing the motor imagery-based brain computer interface. *IEEE Computational Intelligence Magazine*, 14(1), 96–106.
- LAWHERN, V. J., SOLON, A. J., WAYTOWICH, N. R., GORDON, S. M., HUNG, C. P., y LANCE, B. J. (2018). EEGnet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5), 056013.
- LINDSAY, G. (2021). *Models of the mind: How physics, engineering and mathematics have shaped our understanding of the brain*. Bloomsbury Sigma.
- LOURENZUTTI, R., KROHLING, R. A., y REFORMAT, M. Z. (2017). Choquet based TOPSIS and TODIM for dynamic and heterogeneous decision making with criteria interaction. *Information Sciences*, 408, 41–69.
- LUCCA, G., SANZ, J. A., DIMURO, G. P., BEDREGAL, B., MESIAR, R., KOLESÁROVÁ, A., y BUSTINCE, H. (2016). Preaggregation functions: construction and an application. *IEEE Transactions on Fuzzy Systems*, 24(2), 260–272.
- MCCARTHY, J., MINSKY, M. L., ROCHESTER, N., y SHANNON, C. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*. <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>
- MEDLER, D. A. (1998). A Brief History of Connectionism. *Neural Computing Surveys*, 1, 61–101.
- NEURALINK, MUSK, E. (2019). An integrated brain-machine interface platform with thousands of channels. *bioRxiv* 10.1101/703801.
- PATERNAIN, D., FERNÁNDEZ, J., BUSTINCE, H., MESIAR, R., y BELIAKOV, G. (2015). Construction of image reduction operators using averaging aggregation functions. *Fuzzy Sets and Systems*, 261, 8–111.
- PFURTSCHELLER, G., GUGER, C., y RAMOSER, H. (1999). EEG-based brain-computer interface using subject-specific spatial filters. *Engineering applications of bio-inspired artificial neural networks*, Lecture Notes in Computer Science, 1607/1999, 248–254.
- POLIKOV, V. S., TRESKO, P. A., y REICHERT W. M. (2005). Response of brain tissue to chronically implanted neural electrodes. *Journal of Neuroscience Methods*, 148(1) 1–18.
- REGALADO, A. (2020). Elon Musk's Neuralink is neuroscience theater. *MIT Technology Review*. www.technologyreview.com/2020/08/30/1007786/elon-musks-neuralink-demo-update-neuroscience-theater/
- SEJNOWSKI, T. J., y CHURCHLAND, P. S. (1992). *The computational brain*. Cambridge, Mass: MIT Press.
- SMITH, R. (2012). Stratified, personalised, or precision medicine. *British Medical Journal*. <https://blogs.bmj.com/bmj/2012/10/15/richard-smith-stratified-personalised-or-precision-medicine/>
- SUGENO, M. (1974). Theory of fuzzy integrals and its applications. *Doctoral Thesis*. Tokyo Institute of Technology.
- TIBOR, S. R., TOBIAS, S. J., JOSEF, F. L. D., MARTIN, G., KATHARINA, E., MICHAEL, T., FRANK, H., WOLFRAM, B., y TONIO, B. (2017) Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11), 5391–5420.
- TRAPPENBERG, T. P. (2010). *Fundamentals of Computational Neuroscience*. Oxford University Press Inc.
- TURING, A. M. (1950). I.—Computing Machinery and Intelligence. *Mind* LIX, 236, 433–460

- VIDAL J. J. (1973). Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1), 157–180.
- VIJAY, M., KASHYAP, A., NAGARKATTI, A., MOHANTY, S., MOHAN, R., y KRUPA, N. (2020). Extreme gradient boosting classification of motor imagery using common spatial patterns. En *2020 IEEE 17th India Council International Conference (INDICON)*, 1–5.