

Diseño de un algoritmo genético como método de selección de características en un sistema BCI basado en ritmos sensoriomotores

R. Corralejo Palacios¹, R. Hornero Sánchez¹, D. Álvarez González¹, L. M. San José Revuelta²

¹ Grupo de Ingeniería Biomédica, E. T. S. Ingenieros de Telecomunicación, Universidad de Valladolid, Valladolid, España, rebecca.corralejo@uva.es, robhor@tel.uva.es, dalvgon@ribera.tel.uva.es

² Laboratorio de Procesado de Imagen, E. T. S. Ingenieros de Telecomunicación, Universidad de Valladolid, Valladolid, España, lsanjose@tel.uva.es

Resumen

En este estudio se han seleccionado características extraídas de la señal EEG para ser utilizadas en un sistema Brain Computer Interface (BCI). A partir de la señal EEG pueden extraerse múltiples características para ser empleadas en la clasificación. Sin embargo, la utilización de características irrelevantes deteriora el rendimiento del clasificador. Por ello, es necesario seleccionar un grupo reducido de características que aporten información relevante para la clasificación. El presente trabajo propone el uso de un algoritmo genético (AG) como método de selección de características. Dicho método ha sido aplicado sobre el conjunto de datos III de la Competición BCI 2003 consiguiéndose un valor máximo de información mutua (MI) de 0.79 bit, mejorando así los resultados obtenidos por el método ganador de la competición (hasta en un 30%) y por otros métodos propuestos posteriormente (hasta en un 20%).

1. Introducción

Un sistema *Brain Computer Interface* (BCI) es aquel que monitoriza la actividad cerebral y traduce determinadas características, correspondientes a las intenciones del usuario, en comandos de control de un dispositivo [1]. Para registrar dicha actividad cerebral, el método más empleado es el electroencefalograma (EEG), ya que es una técnica portátil, de bajo coste y no invasiva [1].

Los sistemas BCI pueden emplear diferentes tipos de señales de control obtenidas a partir del EEG: potenciales evocados visuales (VEP), potenciales P300, ritmos sensoriomotores μ y β y potenciales corticales lentos (SCP). Es bastante habitual el uso de sistemas BCI basados en ritmos sensoriomotores μ y β . Estos ritmos (banda μ : 8-12 Hz, banda β : 16-24 Hz) presentan variaciones sobre la zona motora del córtex cerebral cuando se realiza o se imagina un movimiento [2]. En concreto, cuando el movimiento está relacionado con la mano derecha se produce una disminución de los ritmos en el hemisferio izquierdo y viceversa. Esta disminución de los ritmos se conoce como *Event Related Desynchronization* (ERD) y se observa en el hemisferio contralateral hacia el que se realiza el movimiento [3]. Debido a esto, detectando estas variaciones podría interpretarse la intención del usuario a partir de su señal EEG. Para ello, es necesario llevar a cabo varias etapas de procesamiento de señal. Los métodos de extracción de características realizan distintas combinaciones y

transformaciones sobre la señal EEG, de forma que las características obtenidas proporcionen la mayor capacidad discriminatoria posible. Por otro lado, los métodos de clasificación asignan, mediante diferentes métodos estadísticos, cada conjunto de características con la clase a la que más probabilidad tiene de pertenecer. Además, resulta también útil la selección de un grupo reducido de características que aporten información relevante para la clasificación, mejorando así el rendimiento del sistema.

Los algoritmos genéticos (AGs) son modelos computacionales inspirados en la evolución [4]. Como tales, codifican una solución potencial mediante una estructura similar a la de los cromosomas sobre los que aplican diferentes operadores de recombinación y mutación [5]. Los AGs mantienen una población de soluciones candidatas para un problema determinado y la hacen evolucionar iterativamente.

En el presente trabajo se emplea el conjunto de datos III de la competición BCI 2003, sobre el que se aplican distintos métodos de extracción de características. El objetivo consiste en el empleo de un AG como método de selección de características, de forma que se identifique qué características son más útiles a la hora de discriminar entre dos tipos de imágenes motoras. Las soluciones candidatas representarán diferentes combinaciones de distinto número de características. Así no será necesario probar todas las posibles combinaciones de características y, por tanto, se podrá ahorrar tiempo y recursos.

2. Señales EEG

En el presente estudio se emplea el conjunto de datos III de la competición BCI 2003 [6]. Dicho conjunto está formado por dos clases de intentos en los que un usuario imagina movimientos de la mano derecha o izquierda. Se proporciona el registro de los canales C3, Cz y C4, según el sistema 10-20. Durante cada intento se mostraba al usuario una indicación en forma de flecha cuyo sentido indicaba el movimiento que tenía que imaginar (mano derecha o izquierda). Inmediatamente después se mostraba una barra de realimentación que indicaba al usuario si estaba realizando correctamente la tarea.

Las señales están muestreadas a 128 Hz y filtradas entre 0.5 y 30 Hz. Se proporcionan 140 intentos etiquetados (se indica a qué clase pertenece cada intento) para

entrenamiento y 140 intentos sin etiquetar para validación. Cada grupo contiene 70 intentos de cada clase (derecha e izquierda). El objetivo de la competición consiste en proporcionar un valor de salida para cada instante de tiempo, de forma que el signo indique la clase (positivo para derecha, negativo para izquierda) y la magnitud refleje la probabilidad de que la estimación sea correcta en un rango entre 0 y 1.

3. Metodología

3.1. Extracción de características

A) Características espectrales a partir de la FFT

Consiste en la obtención de características mediante diferentes parámetros espectrales obtenidos a partir de la densidad espectral de potencia (PSD). Se calculan los momentos espectrales de primer y segundo orden de la PSD. También se calcula la potencia en las bandas μ y β como el área bajo la PSD. Por último, se incluyen dos características relacionadas con la diferencia de potencia entre canales en cada una de las dos bandas de frecuencia.

B) Transformada Wavelet Continua (CWT)

La CWT se define como la convolución de la señal EEG con una *wavelet* conjugada, desplazada y escalada. Las características extraídas son las amplitudes instantáneas de la CWT empleando dos *wavelets* Morlet, una centrada en la banda μ (10 Hz) y otra en la banda β (22 Hz) [3].

C) Transformada Wavelet Discreta (DWT)

La DWT descompone una señal en dos bandas de frecuencia, una entre 0 y $f_s/4$ y otra entre $f_s/4$ y $f_s/2$. Este proceso se puede repetir en la primera banda, dividiéndose esta entre 0 y $f_s/8$ y $f_s/8$ y $f_s/4$. En el presente estudio $f_s = 128$ Hz por lo que con un nivel de descomposición igual a 3 se obtienen cuatro bandas de frecuencia: (0-8 Hz), (8-16 Hz), (16-32 Hz) y (32-64 Hz). El ritmo μ se encontraría dentro de la segunda banda (coeficientes detalle nivel 3: CD_3) y el ritmo β estaría dentro de la tercera (coeficientes detalle nivel 2: CD_2).

D) Modelos autorregresivos (AR)

Los modelos AR son modelos paramétricos que describen la muestra actual como una combinación de las p muestras anteriores más un término de error. En este caso, se ha empleado el algoritmo de Burg para estimar los coeficientes y el orden del modelo empleado es $p = 4$.

E) Filtro adaptado al ritmo μ

Este método se basa en la obtención de un modelo parametrizado del ritmo μ presente en las señales. Se localiza la frecuencia fundamental del ritmo μ y se extraen los segmentos de 1 s de duración que obtienen la máxima correlación con un seno a dicha frecuencia. Estos segmentos se promedian y se analiza su espectro, a partir del cual se crea una plantilla del ritmo μ . Como característica se emplea la raíz cuadrada del máximo de la convolución circular de la plantilla con la señal EEG.

3.2. Selección de características

Se ha diseñado un AG para realizar la selección de

características a partir de las 28 características extraídas del conjunto de datos III de la Competición BCI 2003. Cada individuo de la población es un vector de 28 elementos que representan una característica concreta según se indica en la Tabla 1. Cada elemento se codifica mediante un 1 ó un 0, indicando si la característica correspondiente forma parte o no del subconjunto de características que representa el individuo.

Tanto al generar aleatoriamente la población inicial como al crear cada nueva generación, se establecen ciertas restricciones para que un individuo sea válido. En primer lugar, ha de tener un número de 1's comprendido entre 1 y 7, ya que el número máximo de características que se debería emplear en la clasificación es aproximadamente el número de intentos de cada clase dividido por 10 [7]. Además, los individuos deben cumplir que las características 1, 2 y 9 (potencias en la banda μ para los canales C3 y C4 y la diferencia entre ambas) no pueden incluirse a la vez dentro del mismo subconjunto de características. Debido a que la característica 9 es la diferencia entre las características 1 y 2, al estar relacionadas mediante una combinación lineal provocarían errores al aplicar el clasificador. De forma análoga para la banda β , no se pueden incluir simultáneamente las características 3, 4 y 10.

Se ha establecido un tamaño de 30 individuos para la población y se han realizado 200 iteraciones, ya que se ha observado experimentalmente que un número superior de iteraciones no mejora los resultados. Para medir la aptitud de cada individuo, se realiza la clasificación de los datos empleando el subconjunto de características que representa cada individuo aplicando el clasificador de Bayes y después se calcula el valor de MI máximo. Así, la función de aptitud coincide con el parámetro que se pretende optimizar en el estudio, que es la MI máxima.

Nº	Característica
1	Potencia en la banda μ , canal C3
2	Potencia en la banda μ , canal C4
3	Potencia en la banda β , canal C3
4	Potencia en la banda β , canal C4
5	Momento espectral de primer orden, canal C3
6	Momento espectral de primer orden, canal C4
7	Momento espectral de segundo orden, canal C3
8	Momento espectral de segundo orden, canal C4
9	Potencia en la banda μ , canal C3 - Potencia en la banda μ , canal C4
10	Potencia en la banda β , canal C3 - Potencia en la banda β , canal C4
11	Amplitud instantánea CWT en la banda μ , canal C3
12	Amplitud instantánea CWT en la banda μ , canal C4
13	Amplitud instantánea CWT en la banda β , canal C3
14	Amplitud instantánea CWT en la banda β , canal C4
15	Coefficientes detalle DWT nivel 3, canal C3
16	Coefficientes detalle DWT nivel 3, canal C4
17	Coefficientes detalle DWT nivel 2, canal C3
18	Coefficientes detalle DWT nivel 2, canal C4
19	Coefficiente AR 1 estimados para el canal C3
20	Coefficiente AR 2 estimados para el canal C3
21	Coefficiente AR 3 estimados para el canal C3
22	Coefficiente AR 4 estimados para el canal C3
23	Coefficiente AR 1 estimados para el canal C4
24	Coefficiente AR 2 estimados para el canal C4
25	Coefficiente AR 3 estimados para el canal C4
26	Coefficiente AR 4 estimados para el canal C4
27	Raíz cuadrada del máximo de la convolución circular de la plantilla del filtro adaptado al ritmo μ , con la señal del canal C3
28	Raíz cuadrada del máximo de la convolución circular de la plantilla del filtro adaptado al ritmo μ , con la señal del canal C4

Tabla 1. Relación de características extraídas a partir de las señales del Data Set III de la competición BCI 2003

En el AG diseñado se definen los siguientes operadores:

- Elitismo: permite que los individuos con mejor aptitud pasen intactos a la siguiente generación.
- Selección: se realiza empleando el método de la ruleta, así tienen mayor probabilidad de ser seleccionados los individuos con mejor aptitud.
- Cruce: dos soluciones se descomponen y se mezclan para formar nuevas soluciones. Viene determinado por P_{cru} , la probabilidad de que haya un cruce, y por N_{cru} , el número de puntos sobre los que se realiza.
- Mutación: alteración aleatoria de una solución candidata. Se ha empleado mutación monopunto determinada por el valor de probabilidad P_{mut} .

3.3. Método de clasificación

La clasificación se lleva a cabo mediante el clasificador de Bayes, siguiendo los pasos del método ganador de la competición [3]. Se asume una distribución gaussiana de cuatro dimensiones en cada instante de tiempo y la clasificación en un instante t se obtiene mediante el clasificador de Bayes. Sin embargo, la clasificación final en un instante t_0 incorpora información de los instantes anteriores ($t \leq t_0$), empleándose un valor de potencia discriminativa, ω_t , obtenida a partir del error de Bayes mediante el límite de Chernoff. Finalmente, se adapta el modelo a los requisitos de la competición para tener un valor de salida entre -1 y 1.

4. Resultados

4.1. Variación de las propiedades de elitismo

En primer lugar, se ha ejecutado el AG variando el parámetro N_{elit} , dejando pasar intactos a la generación siguiente un número de individuos variable: 0, 1 ó 2. El resto de parámetros se han mantenido constantes: $P_{cru} = 0.5$, $N_{cru} = 1$ y $P_{mut} = 0.125$.

En la Figura 1 se muestra la evolución de la función de aptitud, de la población global y del mejor individuo de cada generación. Se observa que en el primer caso, $N_{elit} = 0$, el algoritmo no converge a la solución óptima. Los valores de aptitud no presentan una tendencia monótona creciente sino que en ocasiones incluso empeora. Esto no ocurre cuando $N_{elit} = 1$ ó 2, ya que la aptitud del mejor individuo de cada generación es siempre igual o superior a la de la generación previa. La mayor aptitud se obtiene con $N_{elit} = 2$, a partir de la generación 61, alcanzándose un valor de MI máxima de 0.7728 bit. El subconjunto de características que representa este individuo está formado por siete características: 4, 5, 9, 12, 21, 22 y 26.

4.2. Variación de la probabilidad de cruce

Se ha analizado el efecto de la variación del parámetro P_{cru} variando la probabilidad de que dos individuos se combinen entre sí para dar lugar a nuevos individuos. Los valores empleados han sido $P_{cru} = 0.5$, 0.6 y 0.7. El resto de parámetros se han mantenido constantes: $N_{elit} = 2$, $N_{cru} = 1$ y $P_{mut} = 0.125$.

Los resultados se muestran en la Figura 2. En los tres

casos el algoritmo converge a una solución óptima, obteniéndose el mejor resultado con $P_{cru} = 0.7$. La MI máxima obtenida es de 0.7946 bit y se alcanza en la generación 102. El subconjunto de características que representa este individuo está formado por siete características: 4, 5, 9, 12, 14, 15 y 26, cinco de las cuales coinciden con las seleccionadas en el apartado anterior.

4.3. Variación del número de puntos de cruce

Se analiza también la variación del número de puntos de cruce, tomando valores de $N_{cru} = 1$ y 2. El resto de parámetros se han mantenido constantes: $N_{elit} = 2$, $P_{cru} = 0.6$ y P_{mut} disminuyendo de 0.1 a 0.005.

Los resultados se muestran en la Figura 3. En este caso se alcanza la misma solución con ambas configuraciones. La solución a la que convergen consigue un valor de MI de 0.7946 bit, y está formada por siete características (4, 5, 9, 12, 14, 15 y 26) que coinciden exactamente con las de la solución óptima obtenida en el apartado anterior.

4.4. Variación de la probabilidad de mutación

Por último, se ha analizado el efecto de la variación de la probabilidad de mutación tomando valores de P_{mut} constantes: 0.005, 0.05, 0.1 y 0.125 y variables: desde 0.1 hasta 0.005 y desde 0.125 hasta 0.005. El resto de parámetros se han mantenido constantes: $N_{elit} = 2$, $P_{cru} = 0.6$ y $N_{cru} = 1$.

Los resultados se muestran en la Figura 4. Se observa que si la probabilidad de mutación es demasiado pequeña el AG puede caer en un mínimo local y la solución a la que converge presenta una buena aptitud pero puede no ser la más óptima. Sin embargo, cuanto mayor es la probabilidad de mutación se observa que la aptitud global presenta más variaciones ya que se explora más el

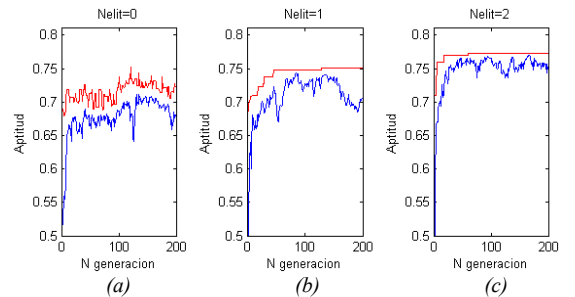


Figura 1. Evolución de la función de aptitud global (línea azul) y del mejor individuo (línea roja). Variación de las propiedades de elitismo: (a) $N_{elit}=0$, (b) $N_{elit}=1$, (c) $N_{elit}=2$

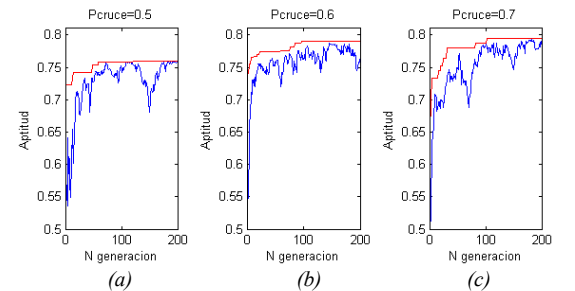


Figura 2. Evolución de la función de aptitud global (línea azul) y del mejor individuo (línea roja). Variación de la probabilidad de cruce: (a) $P_{cru} = 0.5$, (b) $P_{cru} = 0.6$, (c) $P_{cru} = 0.7$

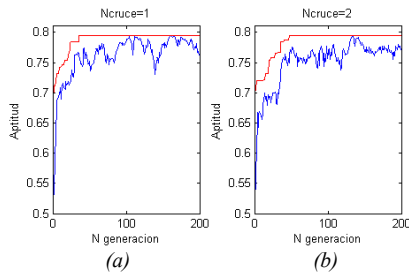


Figura 3. Evolución de la función de aptitud global (línea azul) y del mejor individuo (línea roja). Variación del número de puntos de cruce: (a) $N_{cru} = 1$, (b) $N_{cru} = 2$

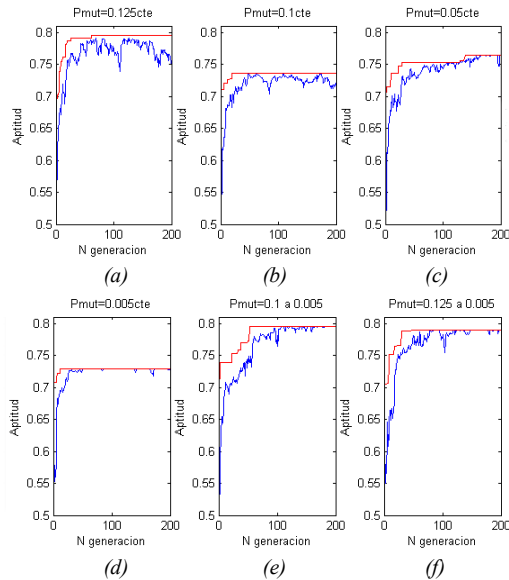


Figura 4. Evolución de la función de aptitud global (línea azul) y del mejor individuo (línea roja). Variación de la probabilidad de mutación: (a) $P_{mut} = 0.125$, (b) $P_{mut} = 0.1$, (c) $P_{mut} = 0.05$, (d) $P_{mut} = 0.005$, (e) $P_{mut} = 0.1$ a 0.005 , (f) $P_{mut} = 0.125$ a 0.005

subespacio de posibles soluciones. En el caso de emplear control de la diversidad, empleando P_{mut} variable, la aptitud global varía cada vez menos. Así se realiza una búsqueda más explorativa al comienzo y más explotativa a medida que se van creando nuevas generaciones.

Los mejores resultados se han obtenido con $P_{mut} = 0.125$ y con P_{mut} variable de 0.1 a 0.005, a partir de las generaciones 62 y 53 respectivamente. El valor de MI máxima obtenido es de 0.7946 bit. El subconjunto de características seleccionado está formado por las características 4, 5, 9, 12, 14, 15 y 26. En el caso de P_{mut} variable el conjunto seleccionado es el mismo pero intercambiando la característica 26 por la 24.

5. Discusión y conclusiones

A la vista de los resultados puede decirse que el AG diseñado es capaz de seleccionar un subconjunto óptimo de características que mejora hasta en un 30% (0.79 bit) el resultado obtenido por el método ganador de la competición (0.61 bit) y en un 20% los resultados obtenidos en estudios posteriores (0.66 bit) [8]. Se observa que todas las configuraciones óptimas convergen hacia subconjuntos de características bastante similares formados siempre por siete características obtenidas a partir de la FFT, la CWT, la DWT y los modelos AR. Sin

embargo, no se selecciona ninguna característica obtenida a partir del filtro adaptado lo que puede deberse a que la información que aporta este método es similar a la de la FFT o CWT y no contribuye con nueva información que pueda mejorar la clasificación. Por otro lado, características obtenidas a partir de la DWT, que no obtenían buenos resultados por separado [8], al ser combinadas con características extraídas a partir de otros métodos consiguen mejorar la clasificación.

A partir del presente trabajo se plantean varias líneas futuras. En primer lugar, una ampliación del número de características extraídas. Resultaría interesante obtener más parámetros a partir de los métodos de extracción de características estudiados: emplear nuevas bandas de frecuencia, más canales EEG, así como aplicar nuevos métodos de extracción de características. Así, al aplicar posteriormente un AG como método de selección se escogerán las más relevantes. Por otro lado, también podría aplicarse un AG con diferentes métodos de selección, cruce y mutación para estudiar si pueden mejorarse los resultados obtenidos en el presente estudio.

En resumen, a lo largo de este trabajo se ha justificado la necesidad de emplear algoritmos de selección de características que proporcionan información sobre qué conjunto de características puede discriminar mejor entre dos clases de imágenes motoras. Además, se ha realizado la selección mediante un AG consiguiéndose una mejora de hasta el 30% respecto al resultado ganador de la competición. Por lo tanto, el empleo de AGs puede resultar de utilidad para el control de aplicaciones BCI orientadas a mejorar la calidad de vida de las personas con grave discapacidad.

Referencias

- [1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, vol 113, 2002, pp 767-791.
- [2] Pineda JA, Allison BZ, Vankov A. The effects of Self-Movement, Observation, and Imagination on Rhythms and Readiness Potentials (RP's): Toward a Brain-Computer Interface (BCI). *IEEE Trans Rehab Eng*, vol 8, no 2, 2000, pp 219-222.
- [3] Lemm S, Schafer C, Curio G. BCI competition 2003-data set III: Probabilistic Modeling of Sensorimotor μ -Rhythms for Classification of Imaginary Hand Movements. *IEEE Trans Biomed Eng*, vol 51, no 6, 2004, pp 1077-1080.
- [4] Whitley D. A genetic algorithm tutorial. *Statistics and Computing*, vol 4, no 2, 1994, pp 65-68.
- [5] Yom-Tov E, Inbar GF. Feature Selection for the Classification of Movements From Single Movement-Related Potentials. *IEEE Trans Neural Systems and Rehab Eng*, vol 10, no 3, 2002, pp 170-177.
- [6] Pfurtscheller G, Schlögl A. Descripción del conjunto de datos III de la competición BCI 2003. Instituto de Ingeniería Biomédica de la Universidad Tecnológica de Graz, 2002.
- [7] Jain AK, Duin RPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE Trans Pat Anal and Mach Intel*, vol 22, no 1, 2000, pp 4-37.
- [8] Corralejo R, Hornero R, Abásolo D, Temprano J. Comparación de Métodos de Extracción de Características para su aplicación a Brain Computer Interface. *Libro de Actas del XXVI CASEIB*, Valladolid, 2008, pp 174-177.