

Filtro supervisado para la selección multi-objetivo de características de electroencefalogramas para BCI

Pedro Martín-Smith¹, Julio Ortega¹, Javier Asensio-Cubero², Andrés Ortiz³

¹Depto. Arquitectura y Tecnología de Computadores, CITIC, Universidad de Granada

²Neuralcubes Ltd. (Reino Unido)

³Depto. de Ingeniería de Comunicaciones, University of Malaga

pmartin@ugr.es, jortega@ugr.es, javier@neuralcubes.co.uk, aortiz@ic.uma.es

Resumen. En este artículo se propone un procedimiento de tipo filtro para la selección de características mediante optimización multi-objetivo en problemas de clasificación de patrones con un número muy grande de características. Este tipo de problemas es relativamente frecuente en muchas aplicaciones bioinformáticas donde, además, el número de características es mucho mayor que el número de patrones disponible para construir el clasificador, y aparecen los denominados problemas de “maldición de dimensionalidad” (*curse of dimensionality*). La principal aportación del procedimiento que se propone es la definición de un conjunto de funciones de utilidad a partir de las etiquetas de los patrones del conjunto de entrenamiento, que permiten una búsqueda eficaz del mejor subconjunto de características a través de un procedimiento evolutivo de optimización multi-objetivo. La propuesta se ha evaluado a través de las tareas de una interfaz cerebro-computador (*Brain-Computer Interface*, BCI) que implica resolver un problema de clasificación de electroencefalogramas. Este problema de clasificación se ha abordado aquí mediante clasificadores LDA y técnicas de análisis de señal en los dominios temporal y espectral, basadas en multi-resolución (*Multi-Resolution Analysis*, MRA), para definir las características de los patrones entre las que se realiza la selección. Los resultados obtenidos ponen de manifiesto ciertos beneficios del filtro propuesto tales como requerir un menor tiempo de procesamiento y ofrecer mejores capacidades de generalización que otros procedimientos de selección de tipo *wrapper* con los que se ha comparado aquí.

Keywords: Interfaces cerebro-computador (*Brain-Computer Interfaces*, BCI), Filtros para selección de características, Selección de características, Optimización multi-objetivo, Análisis Multi-resolución (*Multi-Resolution Analysis*, MRA).

1 Introducción

En muchas aplicaciones de clasificación de patrones de dimensión elevada es preciso utilizar alguna técnica para seleccionar las componentes o características relevantes de los patrones. De hecho, reducir las dimensiones de los patrones a clasificar contribuye a mejorar la tasa de aciertos y la interpretabilidad de los clasificadores

cuando el número de características es muy grande en comparación con el número de patrones de entrenamiento disponibles, situación en la que aparecen los problemas relacionados con la denominada “maldición de la dimensionalidad” (*curse of dimensionality* en inglés). En bioinformática, son frecuentes los problemas de modelado con patrones de dimensión muy elevada, por lo que es imprescindible ser capaz de seleccionar eficientemente las componentes/características relevantes de los patrones, como se pone de manifiesto en [1], una revisión de las técnicas de selección de características utilizadas en bioinformática que además analiza el uso de dichas técnicas en aplicaciones bioinformáticas tales como el análisis de secuencias de aminoácidos, el análisis de *microarrays*, y la espectrografía de masas. En [2] se considera la selección de características para espacios de gran dimensión y pocos patrones en la búsqueda de conjuntos de genes cuyos niveles de expresión puedan servir para la diagnosis y la prognosis. La reducción de dimensiones de los patrones también se ha aplicado en la clasificación de electroencefalogramas (EEG) en el estudio de los ataques de epilepsia [3]. La caracterización de los EEG para los problemas de clasificación propios de las interfaces Cerebro-Computador (*Brain-Computer Interfaces*, BCI) [4] da lugar a patrones con un número muy elevado de características, debiendo aplicarse alguna técnica para seleccionar las esenciales: (1) la presencia de ruido o valores atípicos (*outliers*) en las características; (2) la necesidad de representar la información temporal de los EEG; (3) la no-estacionariedad de los EEG; y (4) en reducido número de patrones (EEGs) disponibles para el entrenamiento del clasificador. Esta circunstancia ocasiona problemas relacionados con la denominada “maldición de la dimensionalidad” (*curse of dimensionality*) debido a que el número de patrones necesario para definir adecuadamente una clase aumenta exponencialmente con la dimensión de los patrones [5].

Aquí se aborda la selección de características basada en optimización multi-objetivo, para problemas de clasificación supervisada. Para evaluar nuestra propuesta se ha utilizado un problema de imaginación de movimiento (*motor imagery*, MI) usual en BCI en el que se ha aplicado el análisis multi-resolución (*multi-resolution analysis*, MRA) [6] para definir las componentes de los patrones (EEGs). Mediante el MRA se aplica una secuencia de espacios sucesivos de aproximación verificando una serie de restricciones para conseguir una descripción lo más próxima posible a la señal que se pretende caracterizar [7], por lo que se trata de una técnica útil siempre que la señal muestre diferencias en los espacios de aproximación sucesivos. Por otro lado, en los problemas de imaginación de movimiento (MI) en una extremidad se trata de aprovechar las amplificaciones y atenuaciones de corta duración asociados a los eventos de sincronización (*event related synchronization* (ERS) y desincronización (*event related desynchronization*, ERD) que se ocasionan cuando los usuarios imaginan movimientos [8]. El análisis ERD/ERS es complejo porque los patrones ERD/ERS son débiles y ruidosos y se producen en distintas localizaciones del córtex, en instantes diferentes dentro de un mismo ensayo, y en bandas de frecuencia distintas. Además, existen diferencias entre usuarios diferentes, e incluso pueden producirse cambios en los patrones en una misma sesión de captura de EEG para el mismo usuario. Consecuentemente, se dispone de un número relativamente pequeño de patrones de dimensión muy elevada, constituyendo un buen escenario para evaluar

las aproximaciones multi-objetivo para la selección de características que se consideran aquí.

En el artículo, la Sección 2 plantea el problema de selección de características como un problema de optimización multi-objetivo, en tanto que la Sección 3 describe la aproximación de tipo filtro que se propone y constituye la principal aportación del trabajo. La Sección 4 considera la caracterización a través del MRA del conjunto de datos de la aplicación de BCI que se ha utilizado para el análisis experimental del artículo. En la misma Sección 4 también se describen los procedimientos de selección utilizados para la comparación experimental con el procedimiento propuesto. Finalmente, los resultados experimentales se proporcionan en la Sección 5, y las conclusiones en la Sección 6.

2 Selección de características multi-objetivo

En este artículo, la selección de características se aborda como un problema de optimización multi-objetivo. Las funciones de coste que definen los objetivos a optimizar evalúan la utilidad de las características seleccionadas para alcanzar una clasificación correcta de los patrones.

Un problema de optimización multi-objetivo busca la combinación de valores en las variables de decisión, $\mathbf{x}=[x_1, x_2, \dots, x_n] \in \mathbb{R}^n$, que satisfice un conjunto dado de restricciones, $g(\mathbf{x}) \leq 0$, $h(\mathbf{x})=0$, y optimiza el vector de funciones objetivo, $\mathbf{f}(\mathbf{x})=(f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$. Dado que, usualmente, los objetivos están en conflicto (reducen el valor en uno de ellos implica incrementar el valor de otros), los procedimientos de optimización multi-objetivo no proporcionan una única solución sino el conjunto de soluciones *no dominadas* que constituyen el denominado *frente de Pareto* entre las que el usuario seleccionaría la más adecuada para el caso. Las soluciones que constituyen el frente de Pareto son óptimas en el sentido de que ninguna solución es peor que las otras soluciones del frente cuando se consideran todos los objetivos.

La Figura 1 proporciona un esquema del procedimiento de selección de características basado en un procedimiento evolutivo de optimización multi-objetivo que proponemos en este trabajo. Cada individuo de la población codifica una selección de características de los patrones de entrada a clasificar. Para definir las funciones de utilidad usadas como funciones de coste u objetivos en el procedimiento evolutivo, se han propuesto diversas alternativas que se pueden clasificar como de tipo *wrapper*, filtro, embebidas, o híbridas. En [9] se revisan las características y las principales propuestas de cada una de dichas alternativas. En los procedimientos de tipo *wrapper* y embebidos, las funciones objetivo dependen del clasificador que se vaya a utilizar puesto que, para evaluar la utilidad de las características seleccionadas, se usan las medidas de las prestaciones que proporciona el clasificador. En cambio, en las alternativas basadas en filtros se tienen en cuenta propiedades de los patrones de entrenamiento como sus distancias euclídeas, funciones basadas en medidas de correlación o de consistencia, medidas de información, etc. Por tanto, en la selección mediante filtros, para calcular los valores de las funciones de utilidad no hay que dedicar tiempo a entrenar el clasificador que se va a usar en la aplicación.

Las etapas del algoritmo evolutivo de optimización multi-objetivo que se utiliza en este artículo corresponden al algoritmo NSGA-II [10], en la que los individuos se han codificado convenientemente y se han implementado operadores genéticos específicos para la aplicación considerada.

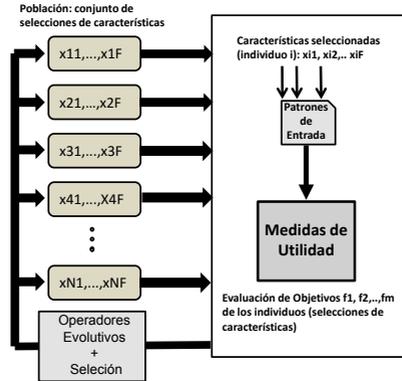


Figura 1. Esquema de una aproximación para la selección de características basada en un procedimiento evolutivo de optimización multi-objetivo

3 Un filtro para selección multi-objetivo con etiquetas.

Las funciones de utilidad utilizadas para la selección de características pueden variar según la clasificación sea supervisada o no supervisada [11]. Entre los trabajos publicados que consideran la selección de características en problemas de clasificación supervisada están [11-15]. En este caso de clasificación supervisada se dispone de las etiquetas de los patrones de entrenamiento y los procedimientos de selección multi-objetivo suelen utilizar el número de características seleccionadas y las prestaciones del clasificador como funciones de utilidad [12, 13]. En el caso de los problemas de BCI, los patrones de entrenamiento están etiquetados y se podrían utilizar medidas de acierto de clasificación para el conjunto de patrones de clasificación como medidas de utilidad. No obstante, el uso de funciones de utilidad relacionadas con la calidad de los agrupamientos (*clustering*) que se pueden alcanzar para una selección de características dada puede ser muy adecuado para modelar la estructura intrínseca de los patrones antes de entrenar el clasificador. Además, en ese tipo de aproximación se podría seguir la misma estrategia que en los algoritmos de *clustering* semi-supervisados, en los que la clave radica en aprovechar el conocimiento que aportan las etiquetas disponibles para los patrones para mejorar la calidad del agrupamiento [16,17].

```

Entrada
Patrones={Pi / i=1,...,n} // n es el número de patrones.
Etiquetas = {Li / i=1,...,n} // Li ∈ {1,...,NC}, Li etiqueta de Pi
// NC es el número de etiquetas diferentes (clases)
S(m)=0 (m=1,...,NC) // variables temporales para evaluar las
//coincidencias/diferencias entre parejas de patrones
SS(m,k)=0 (m=1,...,NC; k=1,...,Qt) // medidas de coincidencias y Q es el número de
// parejas posibles de patrones

Salida
{F(1),...,F(NC)} ; funciones de coste (tantas como etiquetas o clases)

01  Calcular Di,j=d(Pi,Pj) (i,j=1,...,n, i<j);
    // Distancias Euclídeas entre los patrones i y j

02  Δsort = {Δk (k=1,...,Qt=n(n-1)/2)}
    // con Δk = (Pσ1(k), Pσ2(k), d(Pσ1(k), Pσ2(k)))
    // tal que σ1(k) < σ2(k) y d(Pσ1(k), Pσ2(k)) ≤ d(Pσ1(k+1), Pσ2(k+1))
    // para k=1,...,Qt-1.

03  for k=1 to W (W < Qt)
    {
04      if (Lσ1(k) = Lσ2(k)) then
05          S(Lσ1(k)) = S(Lσ1(k)) + 1 //coincidencia (la misma clase)
06      else
07          S(Lσ1(k)) = S(Lσ1(k)) - 0.5; //clases diferentes
08          S(Lσ2(k)) = S(Lσ2(k)) - 0.5;
09      end
10      for m=1 to NC SS(m,k) = S(m) ;
    }
11  for m=1 to NC
12  {
    F(m) = -mediak (SS(m,k)) ;
    }
    
```

Figura 2. Cálculo de las funciones de utilidad en la aproximación de tipo filtro propuesta para la selección de características

La Figura 2 contiene el pseudo-código del procedimiento de tipo filtro que proponemos para evaluar la idoneidad (*fitness*) de los individuos de la población del algoritmo evolutivo multi-objetivo utilizado para la selección de características. Existen tantas funciones objetivo (o funciones de utilidad) como clases. En el procedimiento, dados los n patrones de entrenamiento, P_1, \dots, P_n , y sus etiquetas, L_1, \dots, L_n , correspondientes a las NC diferentes clases, el primer paso (paso 01) de la Figura 2 se dedica al cálculo de las distancias euclídeas entre cada par de patrones posible. A continuación, en el paso 02, los pares de patrones se ordenan según sus distancias (de menor a mayor valor), construyéndose el conjunto Δ_{sort} con los elementos Δ_k ($k=1, \dots, Qt=n(n-1)/2$) que, como se observa en la Figura 2, están definidos por dos patrones y su distancia mutua, donde los índices $\sigma_1(k)$ y $\sigma_2(k)$ corresponden, respectivamente, al primero y al segundo patrón en Δ_k . Los elementos del conjunto Δ_{sort} se utilizan, en los pasos 03 a 10 de la Figura 2, para calcular un conjunto de medidas de coincidencia $SS(m,k)$, donde $m=1, \dots, NC$ y $k=1, \dots, W$, siendo W un valor, que en este artículo hemos fijado a $W=Qt-1$. Si dos patrones consecutivos tienen la misma etiqueta (corresponden a la misma clase), la medida de coincidencia para dicha etiqueta se incrementa en 1 dado que se supone que patrones próximos que pertenecen al mismo *cluster* deberían estar marcados con la misma etiqueta (por

pertenecer a la misma clase). En cambio, si dos patrones próximos no pertenecen a la misma clase, la medida de coincidencia de ambas etiquetas se decrementan en 0.5 (dado que se distribuye la unidad que habría que decrementar entre las dos etiquetas consideradas). Para terminar, en los pasos 11 a 12, se determinan las funciones de coste o utilidad $F(1)$ a $F(NC)$ a partir de la expresión $F(m) = \sum_k SS(m,k)/W$, $m=1, \dots, NC$, $k=1, \dots, W$. Estas funciones nos permiten evaluar la consistencia entre pares de patrones próximos y probabilidades elevadas de coincidencia entre sus etiquetas.

4 Aplicación a la selección de características en BCI

Los datos utilizados pertenecen a la base de datos del laboratorio de BCI de la Universidad de Essex (RU) y contienen patrones que corresponden a tres tipos distintos de movimiento imaginado (mover la mano derecha, la mano izquierda, o un pie) por 10 sujetos de entre 24 y 50 años de edad (58% mujeres, 50% no expertos en uso de técnicas de BCI) a lo largo de cuatro tomas distintas, conteniendo 120 repeticiones por clase de movimiento imaginado y sujeto [6]. Las características de cada EEG (patrón) se obtienen a partir del análisis multi-resolución (*multi-resolution analysis*, MRA) que se describe en [6] y mediante el cual, para cada señal medida por un electrodo se generan varios segmentos para los cuales se obtienen dos conjuntos de coeficientes de *wavelets* (coeficientes de detalle y de aproximación) a distintos niveles de detalle. Así, si se tienen S segmentos, E electrodos, y L niveles de *wavelets*, cada patrón vendrá caracterizado por $2 \times S \times E \times L$ conjuntos de coeficientes, donde el número de coeficientes de cada conjunto depende del nivel. En el conjunto de datos de la Universidad de Essex utilizado, $S=20$, $E=15$, y $L=6$, por lo que tenemos 3600 conjuntos de coeficientes con entre 4 y 128 coeficientes por conjunto (según el nivel). En total, cada patrón viene descrito por 151200 coeficientes. Dado que el número de patrones que se tienen para los conjuntos de entrenamiento y de test es de 178, es evidente que es esencial disponer de un procedimiento eficiente de selección de características.

En [6] se aplica un procedimiento sencillo, que también utilizamos aquí, que asigna un único coeficiente por segmento, electrodo, nivel de *wavelet*, y tipo de coeficiente. Este coeficiente se obtiene calculando la varianza del conjunto de coeficientes correspondientes al segmento, electrodo, nivel y tipo y normalizando dicho valor entre 0 y 1. En cualquier caso, aunque ahora cada patrón está caracterizado por $2 \times S \times E \times L$ coeficientes, todavía sigue ocurriendo que el número de patrones es demasiado pequeño en comparación con el número de características.

Una aproximación para evitar el problema de la maldición de la dimensionalidad sin tener que seleccionar características, consiste en distribuir las características entre un conjunto de clasificadores y utilizar un módulo que elija por mayoría entre las salidas que proporcionan esos clasificadores. En [6] se utiliza esta estrategia, utilizando un conjunto de $2 \times S \times L$ clasificadores LDA (*linear discriminant analysis*) con tantas entradas como electrodos, E . En lo que sigue, esta alternativa se denominará OPT0 (de *option 0*). Además de utilizar esta alternativa, también se han considerado otras dos aproximaciones de selección de tipo *wrapper*, OPT1 y OPT2 [19], basadas en optimización multi-objetivo para comparar con el procedimiento que

se propone aquí. La alternativa OPT1 utiliza un único clasificador LDA y tiene como entradas las características encontradas por el procedimiento de selección multi-objetivo. En cambio, la alternativa OPT2 selecciona entre los $2 \times S \times L$ clasificadores LDA con E entradas de la alternativa OPT0.

Para evaluar las prestaciones de OPT1 y OPT2 una vez entrenados los clasificadores con el conjunto de características seleccionadas (un individuo de la población utilizada en el algoritmo evolutivo) se ha utilizado tanto la tasa de acierto en la clasificación (*accuracy*) como el comportamiento del clasificador para los patrones que no se han utilizado en el entrenamiento, es decir, su capacidad de generalización. Es por ello que se han utilizado dos funciones de costo para evaluar la selección de características en OPT1 y OPT2. La primera de las funciones de costo es el índice Kappa [18], que proporciona una descripción de las prestaciones del clasificador mejor que la razón de acierto dado que tiene en cuenta la distribución de error por clase. La otra función de costo utiliza la validación cruzada 10-*fold* aplicada con los patrones del conjunto de entrenamiento.

5 Resultados experimentales

En esta sección se comparan los resultados de los procedimientos de OPT0, OPT1, y OPT2 con los del procedimiento de selección de características de tipo filtro que se describe en la Sección 3, y que denominaremos FOPT1. Los procedimientos se han implementado en MATLAB[®], y se han ejecutado en un computador con procesador P8700 a 2.53 GHz y 4GB de RAM. Como se ha indicado más arriba, los experimentos se han realizado con datos del laboratorio de BCI de la Universidad de Essex para 10 individuos (designados como 101 a 110) organizados en dos ficheros por individuo. El fichero *x#* contiene los patrones de entrenamiento para el individuo # y el fichero *xe#* contiene los patrones de test para dicho individuo #. Las etiquetas de los patrones pueden tomar tres valores, uno para cada uno de los tres posibles movimientos a los que nos hemos referido más arriba. Igualmente, tal y como se ha dicho, la alternativa OPT0 no implica ninguna selección de características ya que es la propia estructura del sistema la que evita la maldición de la dimensionalidad a costa de una mayor complejidad en el clasificador. La alternativa OPT1 utiliza un procedimiento de selección de tipo *wrapper* basado en optimización multi-objetivo sobre un espacio de $S \times L \times E \times 2 = 20 \times 6 \times 15 \times 2 = 3600$ características, mientras que la alternativa OPT2 utiliza un espacio de búsqueda de dimensión $20 \times 6 \times 2 = 240$. Nuestro procedimiento FOPT1 también considera un espacio de dimensión de 3600 componentes como OPT1. La complejidad de los clasificadores en OPT1 y FOPT1 es menor que los que se consideran en OPT0 y OPT2.

El algoritmo evolutivo de optimización multi-objetivo que se ha utilizado es el bien conocido NSGA-II, incluyendo los operadores de cruce y mutación. Se utiliza cruce binario simulado con una probabilidad de 0.5 (una probabilidad de mutación de 0.5), con índices de distribución de 20 tanto para la mutación como para el cruce. También se han hecho algunos experimentos con diferentes tamaños de población y generaciones, para determinar los valores más bajos en estos parámetros que proporcionasen resultados competitivos con OPT0. Así, se han utilizado 50

iteraciones y 50 generaciones para OPT1 y OPT2, y 200 individuos y 300 generaciones para FOPT1. Al final de la sección se considera el efecto de los cambios en estos parámetros. En cualquier caso, hay que tener en cuenta que en este trabajo se pretende mostrar que la selección de características multi-objetivo proporciona ciertas mejoras de clasificación en tareas de BCI utilizando MRA para la extracción de características, y no se ha llevado a cabo un ajuste exhaustivo de los parámetros del procedimiento evolutivo de optimización.

Tabla 1. Comparación de los distintos métodos de selección y clasificación con los datos de la Universidad de Essex (índices Kappa evaluados con los patrones de entrenamiento)

Subject	OPT0	OPT1	OPT2	FOPT1
	Índice Kappa (x#)	Índice Kappa (x# media.std)	Índice Kappa (x# media.std)	Índice Kappa (x# media.std)
101	0.790	0.738±0.022*	0.828±0.011 [^]	0.744±0.014*
102	0.857	0.747±0.017*	0.855±0.011	0.767±0.023*
103	0.757	0.665±0.012*	0.747±0.013	0.659±0.020*
104	0.899	0.819±0.013*	0.902±0.010	0.844±0.018*
105	0.757	0.646±0.023*	0.751±0.013	0.628±0.023*
106	0.774	0.604±0.023*	0.776±0.014	0.604±0.015*
107	0.857	0.816±0.018*	0.880±0.011 [^]	0.820±0.020*
108	0.774	0.508±0.021*	0.745±0.023*	0.500±0.018*
109	0.790	0.597±0.021*	0.770±0.015*	0.595±0.023*
110	0.883	0.821±0.016*	0.877±0.026	0.831±0.018*

Tabla 2. Comparación de los distintos métodos de selección y clasificación con los datos de la Universidad de Essex (índices Kappa evaluados con los patrones de test)

Subject	OPT0	OPT1	OPT2	FOPT1
	Índice Kappa (xe#)	Índice Kappa (xe# media.std)	Índice Kappa (xe# media.std)	Índice Kappa (xe# media .std)
101	0.438	0.393±0.046*	0.437±0.033	0.468±0.028 [^]
102	0.455	0.302±0.074*	0.429±0.023	0.406±0.025*
103	0.279	0.249±0.046	0.325±0.017 [^]	0.354±0.029 [^]
104	0.564	0.510±0.056	0.545±0.035	0.639±0.045 [^]
105	0.287	0.191±0.040*	0.240±0.031	0.248±0.041
106	0.321	0.193±0.070*	0.319±0.028	0.271±0.028*
107	0.631	0.560±0.041*	0.634±0.019	0.636±0.018
108	0.254	0.088±0.036*	0.184±0.027*	0.190±0.028*
109	0.388	0.207±0.071*	0.333±0.026*	0.318±0.041*
110	0.648	0.450±0.036*	0.605±0.041*	0.559±0.022*

Las Tablas 1 y 2 comparan los índices Kappa que se obtienen a través de las cuatro aproximaciones consideradas (OPT0, OPT1, OPT2, y FOPT1). Las columnas "Índice Kappa index (x#)" de la Tabla 1 proporcionan los valores de dicho índice evaluando el clasificador correspondiente con el mismo conjunto de patrones utilizado para el entrenamiento. En la Tabla 2, las columnas "Índice Kappa (xe#)" indican que la evaluación se ha hecho con el conjunto de patrones de test, que no se utilizan en el entrenamiento de los clasificadores.

Los resultados de las Tablas 1 y 2 corresponden a los valores promedio de la tasa de aciertos y la desviación típica tras 15 repeticiones de cada experimento. Se han aplicado los test de Kolmogorov-Smirnov y de Kruskal-Wallis con un nivel de confianza del 95% para determinar si las diferencias observadas entre los distintos

procedimientos son estadísticamente significativas. En las Tablas 1 y 2, \wedge y $*$ significan que, respectivamente la mejora o el empeoramiento observados con respecto a OPT0, son estadísticamente significativos.

En cuanto al tiempo de ejecución requerido por las distintas alternativas, el valor medio es de $4.533\pm 45s$ para OPT1, $13.353\pm 1031s$ para OPT2, y $11.598\pm 67s$ para FOPT1. Teniendo en cuenta las diferencias observadas en los tiempos de ejecución se podría indicar que la comparación realizada no es del todo justa e intentar mejorar las prestaciones de las alternativas que requieren menos tiempo de ejecución, fundamentalmente OPT1, incrementando el número de individuos en la población o las generaciones hasta que se tengan tiempos similares a los de OPT2 y FOPT1. Los experimentos que se han realizado para OPT1, utilizando los datos de dos individuos (104 y 107) y poblaciones de 100 individuos y 60 generaciones, corresponden a un tiempo medio de ejecución del mismo orden que el de FOPT1 ($11.042\pm 162s$), y proporcionan valores medios de 0.515 ± 0.047 para $x104$ y de 0.580 ± 0.047 para $x107$. En ambos casos, los resultados obtenidos por FOPT1 siguen siendo mejores.

6 Conclusiones

Este artículo propone un nuevo procedimiento de tipo filtro para la selección de características mediante un algoritmo evolutivo de optimización multi-objetivo que utiliza un objetivo por cada una de las clases del problema. La propuesta ha sido comparada con el procedimiento descrito [6], que aborda el problema de la maldición de la dimensionalidad mediante la composición de varios clasificadores que tienen como entradas subconjuntos diferentes de las características de los patrones. De esta forma, el número de patrones es mayor que el número de características utilizadas como entrada en cada clasificador. No obstante, en este tipo de aproximaciones el número de clasificadores que hay que entrenar es elevado, igual que el tiempo necesario para entrenarlos y para calcular, a partir de la señal, el valor de todas las características. Se trata de una dificultad considerable desde el punto de vista de las aplicaciones con requisitos de tiempo real. Por otra parte, nuestro filtro multi-objetivo también ha sido comparado con otras dos alternativas multi-objetivo previamente propuestas [19] con dos funciones de costo relacionadas, respectivamente, con la tasa de acierto del clasificador y con la capacidad de generalización del clasificador.

Los resultados experimentales muestran que la selección de características multi-objetivo proporciona prestaciones similares a las que se consiguen con OPT0 utilizando todas las características de los patrones de entrada con lo que mediante OPT1, OPT2 y FOPT1 se puede reducir la complejidad de los clasificadores. Además, nuestra propuesta FOPT1 mejora la alternativa OPT1 que utiliza la misma estructura de clasificador. Además, FOPT1 también mejora OPT0 y OPT2 para algunos individuos y parece que la capacidad de generalización que proporciona FOPT1 es mejor que las de OPT1 y OPT2 puesto que las diferencias entre los valores de los índices Kappa en las Tablas 1 y 2 son menores para FOPT1.

Como trabajo futuro analizaremos los conjuntos de características seleccionados desde el punto de vista del conocimiento que proporcionen acerca de los electrodos y segmentos relevantes. Además continuaremos analizando distintos índices de validación de *clusters* tanto en procedimientos de tipo *wrapper* como de tipo filtro.

Agradecimientos. Este trabajo ha sido financiado por el proyecto TIN2015-67020-P (Ministerio de Economía y Competitividad y fondos FEDER). Agradecemos al profesor John Q. Gan de la Universidad de Essex el acceso a las bases de datos utilizadas y su ayuda en este trabajo.

Referencias

1. Y. Saecys; I. Inza; P. Larrañaga: "A review of feature selection techniques in bioinformatics". *Bioinformatics*, Vol. 23, N. 19, pp. 2507-2517, 2007.
2. C. Sima; E. Dougherty: "What should be expected from feature selection in small-sample settings". *Bioinformatics*, Vol. 22, pp. 2430-2436, 2006.
3. N. Acir; C. Güzelis: "An application of support vector machine in bioinformatics: Automated recognition of epileptiform patterns in EEG using SVM classifier designed by a perturbation method". *Advances in Information Systems. LNCS Vol. 3261*, pp. 462-471. 2005.
4. F. Lotte; M. Congedo; A. Lécuyer; F. Lamarche; B. Arnaldi: "A review of classification algorithms for EEG-based brain-computer interfaces". *Journal of Neural Engineering*, Vol. 4, 2007.
5. S.J. Raudys; A.K. Jain: "Small sample size effects in statistical pattern recognition: Recommendations for practitioners". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, pp. 252-264, 1991.
6. J. Asensio-Cubero; J.Q. Gan; R. Palaniappan: "Multiresolution analysis over simple graphs for brain computer interfaces". *Journal of Neural Engineering*, Vol. 10, No. 4, doi:10.1088/1741-2560/10/4/046014, 2013.
7. I. Daubechies: "Ten Lectures on Wavelets". SIAM, Philadelphia, 2006.
8. G. Pfurtscheller; F.H. Lopes da Silva: "Event-related EEG/MEG synchronization and desynchronization: Basic principles". *Clinical Neurophysiology*, Vol. 110, No. 11, pp. 1842-1857, 1999.
9. V. Bolón-Canedo; N. Sánchez-Marroño; A. Alonso-Betanzos: "A review of feature selection methods on synthetic data". *Knowl. Inf. Syst.*, Vol. 34, pp. 483-519, 2013.
10. K. Deb; S. Agrawal; A. Pratap; T. Meyarivan: "A fast elitist Non-dominated Sorting Genetic Algorithms for multi-objective optimisation: NSGA-II". In *Proc. of the 6th Int. Conf. on Parallel Problem Solving from Nature (PPSN VI)*, LNCS 1917, pp.849-858, Springer-Verlag, 2000.
11. J. Handl; J. Knowles: "Feature selection in unsupervised learning via multi-objective optimization". *Intl. Journal of Computational Intelligence Research*, Vol. 2, No. 3, pp. 217-238, 2006.
12. C. Emmanouilidis; A. Hunter; J. MacIntyre: "A multi-objective evolutionary setting for feature selection and a commonality-based crossover operator". In *Proc. of the 2000 Congress on Evolutionary Computation*, IEEE Press, New York, NY, pp. 309-316, 2000.
13. L.S. Oliveira; R. Sabourin; F. Bortolozzi; C.Y. Suen: "A methodology for feature selection using multi-objective genetic algorithms for handwritten digit string recognition". *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 17, No. 6, pp. 903-929, 2003.
14. Y. Kim; W.N. Street; F. Menezes: "Evolutionary model selection in unsupervised learning". *Intelligent Data Analysis*, Vol. 6, No. 6, pp. 531-556, 2002.
15. M. Morita; R. Sabourin; F. Bortolozzi; C.Y. Suen: "Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition". In *Proc. of the Seventh International Conference on Document Analysis and Recognition*, IEEE Press, New York, NY, pp. 666-671, 2003.
16. H. Gan; N. Sang; R. Huang; X. Tong; Z. Dan: "Using clustering analysis to improve semi-supervised classification". *Neurocomputing*, Vol. 101, pp.290-298, 2013.
17. S. Basu; A. Banerjee; R.J. Rooney: "Semi-supervised clustering by seeding". In *Proc. of the 19th International Conference on Machine Learning*, pp.11-18, 2003.
18. J. Cohen: "A coefficient of agreement for nominal scales". *Educ. Psychological Meas.*, Vol. 20, pp. 37-46, 1960.
19. J. Ortega; J. Asensio-Cubero; J.Q. Gan; A.Ortiz: "Evolutionary multi-objective feature selection in multiresolution analysis for BCI". In *Proc. of IWBBIO*, Granada, 2015.