Estudio sobre Ensembles de Rankings de Características

Borja Seijo-Pardo, Verónica Bolón-Canedo, Iago Porto-Díaz y Amparo Alonso-Betanzos

Departamento de Computación - Universidad de A Coruña
Campus de Elviña s/n 15071 - A Coruña, España

Resumen En los últimos años, el aprendizaje *ensemble* se ha convertido en un foco de gran atención, partiendo de la suposición de que la salida obtenida al combinar múltiples expertos es mejor que la que se puede alcanzar utilizando un único experto. Esta idea de aprendizaje *ensemble* se puede adaptar para ser usada en el proceso de selección de características, donde los diferentes métodos de selección actúan como expertos. En este trabajo se proponen varios métodos *ensemble* de selección de características basados en la combinación de diferentes *rankings* de características, intentando superar el problema de seleccionar de antemano el método más apropiado para cada escenario. Los resultados de los *rankings* individuales son combinados con diferentes métodos de unión y la adecuación de los *ensembles* se demuestra mediante un método de clasificación SVM. Los resultados obtenidos sobre cinco conjuntos de datos muestran que el uso de los métodos *ensemble* propuestos obtienen rendimientos mejores o similares a los que se consiguen con los métodos individuales de selección de características, consiguiendo liberar al usuario de un análisis previo para elegir el mejor método en cada escenario.

1. Introducción

En los últimos años el tamaño de los conjuntos de datos, utilizados en el área del aprendizaje automático, ha sufrido un notable incremento. Debido a ello, la *selección de características* tiene un papel importante en este área a la hora de extraer información y conocimiento útil. Para ello se identifican y eliminan las características que se consideran irrelevantes y redundantes, reduciendo el tamaño de los conjuntos de datos, lo que conlleva una disminución del coste computacional de los algoritmos de aprendizaje y un menor uso de tamaño de almacenamiento. Además, estudios previos han demostrado que el rendimiento de los modelos de clasificación mejora cuando las características irrelevantes y redundantes son eliminadas del conjunto de datos original [17].

Tradicionalmente, los métodos de aprendizaje automático han utilizado un único modelo de aprendizaje para resolver un problema concreto. Sin embargo, a lo largo de los últimos años, se ha observado que combinando varios modelos diferentes sobre el mismo problema se pueden obtener mejores resultados.

Esta idea está basada en la suposición de que combinar la salida de múltiples expertos en un problema particular produce mejores resultados que los que se pueden obtener utilizando un único experto [9], siendo los métodos más populares *Bagging* [3] y *Boosting* [18], que aportan garantías teóricas de rendimiento y resultados experimentales sólidos. A pesar de que la mayoría de los trabajos existentes aplican el aprendizaje *ensemble* sobre el proceso de clasificación, esta técnica también puede verse como un medio para mejorar otras disciplinas de aprendizaje automático, como el proceso de selección de características.

En el proceso de selección de características existen dos enfoques diferentes a la hora de evaluar las características de un conjunto de datos: (i) *evaluación de subconjunto* y (ii) *evaluación individual* [24]. En el primero de los casos, se generan sucesivos subconjuntos de características de acuerdo a una estrategia de búsqueda prefijada. Estos subconjuntos son evaluados iterativamente, siguiendo un criterio de bondad, hasta alcanzar el subconjunto final de características seleccionadas. En el segundo caso, se obtiene un *ranking* de todas las características del conjunto de datos, asignando un nivel de relevancia a cada una de ellas. Los métodos de evaluación de subconjunto tienen la ventaja de detectar redundancia y dependencia entre los atributos, pero pueden sufrir de un problema de ineficiencia computacional y por ello en este trabajo se han utilizado los métodos de evaluación individual.

Aparte de esta clasificación, se pueden distinguir tres aproximaciones diferentes en base a la relación entre el algoritmo de selección de características y el método de aprendizaje inductivo usado para inferir un modelo: (i) *métodos de filtrado*, (ii) *métodos envolventes* o *wrappers* y (iii) *métodos embebidos* [5]. Los métodos de filtrado utilizan las propiedades inherentes de los datos de entrenamiento para realizar el proceso de selección de características, independientemente del algoritmo de inducción. Estos métodos presentan un bajo coste computacional y una buena capacidad de generalización. Los métodos envolventes o *wrappers* utilizan el propio algoritmo de inducción como parte de la función de evaluación de los subconjuntos de características. Esta interacción con el clasificador suele aportar mejor precisión en los resultados, pero tiene un gran coste computacional y los resultados pueden sufrir de sobreajuste. Por último, los métodos embebidos buscan un subconjunto óptimo de características dentro de la construcción del clasificador, y se pueden ver como una búsqueda en el espacio combinado del subconjunto de funciones e hipótesis. Esta aproximación es capaz de obtener las dependencias de características a un menor coste computacional que los *wrappers*, pero también puede sufrir de sobreajuste. Por ello, en este artículo se han seleccionado los métodos de filtrado y embebidos, debido a que son capaces de reducir la dimensión de los datos sin comprometer el tiempo y la memoria requeridos por los algoritmos de aprendizaje automático.

En el ámbito de la selección de características, se han realizado varios estudios, aplicando el aprendizaje *ensemble*, con el fin de lograr una mejor precisión de clasificación. Uno de estos estudios [19] ha sido realizado sobre 21 conjuntos de datos *UCI* [2], comparando cinco medidas de diversidad con respecto a su posible uso en el *ensemble* de selección de características. Este estudio considera cuatro

estrategias de búsqueda para el *ensemble* de selección de características junto con el subespacio aleatorio simple. A partir de esta idea de múltiples criterios de evaluación de selección de características, surgen varios *ensembles* formados por diferentes métodos de selección de características. Con el objetivo de mejorar el rendimiento y la estabilidad de los resultados se desarrolla el algoritmo *Multicriterion Fusion-based Recursive Feature Elimination* [23] *(MCF-RFE)*. Además, existen otros estudios donde los métodos de selección que forman el *ensemble* son métodos de tipo *ranker*, combinados después para producir un único *ranking* final. Tres técnicas de *ranking* de características, basadas en métodos de filtrado de uso general para problemas de clasificación de texto, fueron utilizadas por Olsson y Oard [13], donde se utilizan tres métodos de unión diferentes. Wang et al. realizan algunos trabajos destacados en este área, como [21].

La idea de este trabajo es aplicar el aprendizaje *ensemble* al proceso de selección de características, utilizando métodos de filtrado y embebidos para aportar diversidad. El objetivo marcado es el de introducir diversidad y aumentar la regularidad del proceso de selección de características, de manera que se aprovechen las ventajas de los métodos individuales que forman el *ensemble* y se superen sus puntos débiles. Se denomina regularidad a la capacidad de los métodos *ensemble* para obtener resultados aceptables independientemente del conjunto de datos sobre el que es aplicado y sus propiedades particulares. También se debe mencionar que usando los métodos *ensemble* se va a intentar liberar al usuario de seleccionar el mejor método de selección de características para cada escenario, ya que esta aproximación obtiene resultados aceptables independientemente de las propiedades de los datos.

2. Ensemble de Selección de Características

En la Figura 1 se puede ver un esquema gráfico simplificado de la composición de un *ensemble* de métodos *ranker*, el cual se va a describir a continuación.

Existen numerosas y variadas formas de diseñar las aproximaciones *ensemble* de selección de características. De entre todas las opciones existentes, en este trabajo se ha optado por realizar un *ensemble* de métodos *ranker*. En este tipo de diseño *ensemble* se intenta maximizar la diversidad de métodos individuales de selección de características que lo componen, obteniendo así resultados de mayor regularidad, robustez y precisión [21]. Para ello se generan R modelos utilizando diferentes métodos *ranker* de selección de características, todos empleando el mismo conjunto de datos de entrenamiento. Finalmente, las r salidas obtenidas de los diferentes métodos son combinadas, usando algún método de unión para obtener un único *ranking* final con el que entrenar el clasificador. Debido a que el *ranking* final obtenido contiene todas las características del conjunto de datos, es necesario establecer un umbral T con la finalidad de obtener un subconjunto de características representativo. Una vez obtenido este *ranking* final se realiza el entrenamiento de un clasificador.

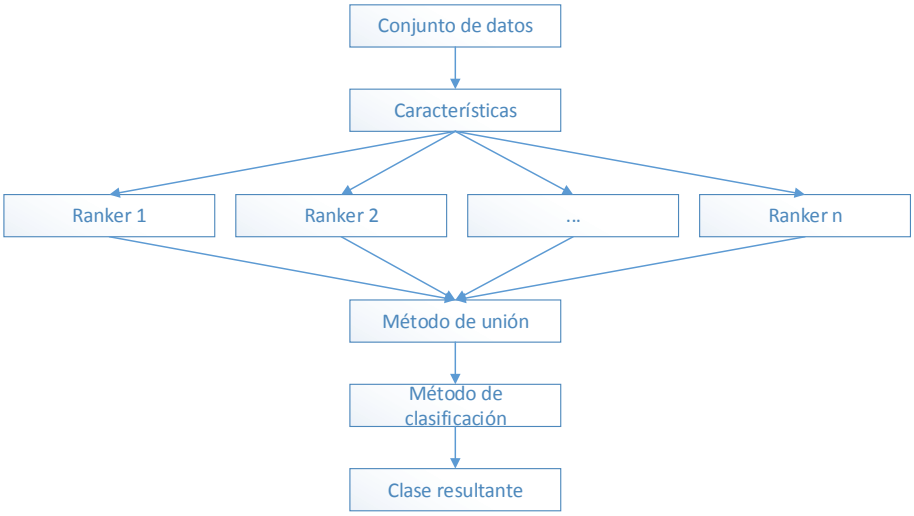

Figura 1. Esquema de composición de un *ensemble* de métodos *ranker*.

3. Estudio Experimental

El rendimiento de los *ensembles* propuestos se prueba sobre cinco conjuntos de datos diferentes, pertenecientes al repositorio de conjuntos de datos UCI [2]. Estos conjuntos de datos se muestran en el Cuadro 1, formando un interesante punto de referencia para comprobar la adecuación de los *ensembles* propuestos.

Conjunto de datos	N° muestras	N° Características	Clases
Yeast	1 484	8	10
Spambase	4 601	57	2
Madelon	2 400	500	2
Connect4	67 557	42	3
Isolet	7 797	617	26

Cuadro 1. Conjuntos de datos utilizados en el estudio experimental. Todos ellos están disponibles para descarga en [2]

Entre la amplia gama de métodos *ranker* de selección de características disponibles en la literatura, se han seleccionado cuatro métodos de filtrado (*Chi-Cuadrado* [10], *Information Gain* [16], *mRMR* [15] y *ReliefF* [8]) y dos métodos embebidos (*SVM-RFE* [4] y *FS-P* [12]) para realizar este estudio. En particular, el método *mRMR* realiza la búsqueda en el espacio de características de manera

similar a los métodos de subconjunto, pero con la diferencia de que el número de características a seleccionar es un parámetro que debe establecer el usuario. Por lo tanto, a efectos prácticos para este estudio se comporta como el resto de métodos de tipo ranking, donde necesitamos establecer un umbral para decidir con qué características vamos a entrenar el clasificador correspondiente. Este conjunto de métodos *ranker* ha sido elegido porque (i) utilizan diferentes métricas, lo que asegura una gran diversidad en el *ensemble*; y (ii) son ampliamente utilizados por los investigadores en la selección de características.

Los diferentes *rankings*, obtenidos como resultado de aplicar los métodos anteriores de manera individual sobre el conjunto de datos, son combinados mediante diversos métodos de unión para obtener un único *ranking* final. En este trabajo se han seleccionado siete métodos de unión diferentes: *SVM-Rank* [6] (basado en el algoritmo *SVM* para el aprendizaje de funciones *ranking*), los métodos *min*, *median*, *mean* y *geommean* [22] (basados en operaciones aritméticas simples) y los métodos *Stuart* [1] y *RRA* [7] (utilizan distribuciones estadísticas de ordenación),

Puesto que los métodos de selección utilizados son todos métodos *ranker*, es decir, no seleccionan un subconjunto de características sino que devuelven el conjunto total ordenado, es necesario establecer un valor de umbral para obtener un subconjunto de las características más relevantes. En este estudio se ha optado por dos valores de umbral diferentes: (i) el umbral $\log_2(n)$ –donde n es el número de características de un conjunto de datos determinado– ya que es el umbral recomendado en [21] y (ii) un valor de umbral correspondiente con el 10 % de las características del conjunto.

Para comprobar la adecuación de los metodos *ensemble* propuestos, en términos de error de clasificación, se elige un clasificador de tipo *SVM* con función de base radial *(SVM-RBF)* [14]. Para ello se lleva a cabo una validación cruzada *10-fold*, consistente en dividir el conjunto de datos en diez particiones y repetir el proceso de aprendizaje diez veces. Cada una de las veces, una partición es usada como datos de *test* y las nueve particiones restantes se usan juntas como datos de entrenamiento. Finalmente, se calcula el error medio obtenido en las diez repeticiones, y la desviación típica que presentan estos datos. Los Cuadros 2 y 3 muestran la media de los porcentajes de error sobre el conjunto de *test*. Teniendo diez errores diferentes como resultado de la validación cruzada *10-fold*, se realiza el *test* de Kruskal-Wallis, con la hipótesis nula de que el conjunto de métodos comparados tiene un comportamiento similar, para comprobar si existen diferencias significativas para un nivel de significación $\alpha = 0{,}05$. A continuación, en caso de que se rechace la hipótesis anterior, se realiza un *test* de comparación múltiple (Tukey's) [20]. Los resultados de estos *tests* se muestran en las tablas marcando con una cruz aquellos métodos cuyo error de *test* no sea significativamente peor que el mejor resultado individual obtenido para cada uno de los conjunto de datos.

El experimento realizado consiste en una comparación entre el uso de diferentes métodos de selección de manera individual y el uso de métodos *ensemble*. En primer lugar, se ha realizado un estudio donde se han probado todas las posibles

combinaciones de cuatro, cinco y seis métodos *ranker*, uniendo los *rankings* parciales de acuerdo a los métodos de unión indicados anteriormente para obtener un único *ranking* final. De este estudio preliminar, cuyos resultados totales no se presentan por limitaciones de espacio, se ha observado que los *ensembles* de cinco y seis *rankers* obtienen, en media, los mejores resultados. En este trabajo se muestran comparaciones entre los métodos individuales base, los diferentes *ensembles* formados por cinco *rankers* y los de seis *rankers*, donde la diferencia entre los dos tipos de *ensemble* es la inclusión del método *Chi-Cuadrado* para formar el modelo final.

En las siguientes secciones se responderá a las preguntas: "¿Alguno de los métodos individuales aporta mejores resultados que el resto?", "¿Los métodos *ensemble* mejoran a los métodos individuales?", "¿Cuál es el mejor método de unión?", y, finalmente, "¿Aporta alguna mejora añadir el método *Chi-Cuadrado* en la formación del *ensemble*?". Para responder a esta serie de preguntas, se van a proporcionar dos figuras comparativas diferentes (Figura 2 y 3).

Ranker (Método de unión)	Yeast	Spambase	Madelon	Connect4	Isolet
Ensemble5 (SVM-Rank)	$54{,}58^{\dagger} \pm 3{,}33$	$12{,}22^{\dagger} \pm 2{,}98$	$33{,}71^{\dagger} \pm 4{,}82$	$31{,}31^{\dagger} \pm 0{,}71$	$50{,}58^{\dagger} \pm 3{,}94$
Ensemble5 (min)	$53{,}91^{\dagger} \pm 3{,}52$	$20{,}04 \pm 1{,}78$	$33{,}46^{\dagger} \pm 4{,}40$	$32{,}28 \pm 0{,}60$	$46{,}80^{\dagger} \pm 3{,}14$
Ensemble5 (median)	$57{,}62^{\dagger} \pm 5{,}20$	$11{,}76^{\dagger} \pm 1{,}09$	$33{,}46^{\dagger} \pm 4{,}45$	$31{,}48^{\dagger} \pm 0{,}44$	$55{,}36 \pm 4{,}76$
Ensemble5 (mean)	$54{,}24^{\dagger} \pm 3{,}22$	$14{,}76^{\dagger} \pm 4{,}52$	$33{,}50^{\dagger} \pm 4{,}30$	$31{,}40^{\dagger} \pm 0{,}68$	$51{,}26 \pm 4{,}62$
Ensemble5 (geommean)	$53{,}90^{\dagger} \pm 3{,}52$	$12{,}20^{\dagger} \pm 2{,}92$	$33{,}33^{\dagger} \pm 4{,}37$	$31{,}25^{\dagger} \pm 0{,}73$	$51{,}72 \pm 3{,}26$
Ensemble5 (stuart)	$54{,}65^{\dagger} \pm 4{,}16$	$11{,}30^{\dagger} \pm 1{,}37$	$33{,}33^{\dagger} \pm 4{,}37$	$31{,}22^{\dagger} \pm 0{,}78$	$50{,}57^{\dagger} \pm 3{,}99$
Ensemble5 (RRA)	$56{,}20^{\dagger} \pm 3{,}94$	$13{,}67^{\dagger} \pm 3{,}15$	$33{,}50^{\dagger} \pm 4{,}44$	$30{,}96^{\dagger} \pm 0{,}78$	$50{,}97^{\dagger} \pm 4{,}16$
Ensemble6 (SVM-Rank)	$55{,}52^{\dagger} \pm 4{,}73$	$16{,}74^{\dagger} \pm 4{,}32$	$33{,}46^{\dagger} \pm 4{,}24$	$31{,}24^{\dagger} \pm 0{,}75$	$48{,}99^{\dagger} \pm 2{,}31$
Ensemble6 (min)	$53{,}91^{\dagger} \pm 3{,}52$	$20{,}04 \pm 1{,}78$	$33{,}50^{\dagger} \pm 4{,}45$	$32{,}28 \pm 0{,}60$	$47{,}62^{\dagger} \pm 6{,}16$
Ensemble6 (median)	$58{,}76^{\dagger} \pm 3{,}92$	$11{,}85^{\dagger} \pm 1{,}13$	$33{,}38^{\dagger} \pm 4{,}36$	$31{,}44^{\dagger} \pm 0{,}32$	$51{,}61 \pm 2{,}97$
Ensemble6 (mean)	$58{,}36^{\dagger} \pm 4{,}47$	$13{,}76^{\dagger} \pm 3{,}99$	$33{,}71^{\dagger} \pm 4{,}31$	$31{,}34^{\dagger} \pm 0{,}71$	$48{,}72^{\dagger} \pm 1{,}62$
Ensemble6 (geommean)	$54{,}24^{\dagger} \pm 3{,}22$	$13{,}13^{\dagger} \pm 3{,}69$	$33{,}33^{\dagger} \pm 4{,}37$	$31{,}06^{\dagger} \pm 0{,}50$	$49{,}53^{\dagger} \pm 1{,}52$
Ensemble6 (stuart)	$58{,}76^{\dagger} \pm 3{,}92$	$14{,}54^{\dagger} \pm 4{,}26$	$33{,}33^{\dagger} \pm 4{,}37$	$31{,}18^{\dagger} \pm 0{,}81$	$49{,}60^{\dagger} \pm 1{,}52$
Ensemble6 (RRA)	$58{,}76^{\dagger} \pm 3{,}92$	$18{,}45^{\dagger} \pm 3{,}78$	$33{,}42^{\dagger} \pm 4{,}36$	$30{,}73^{\dagger} \pm 0{,}82$	$49{,}94^{\dagger} \pm 4{,}17$
Chi-Cuadrado	$58{,}76^{\dagger} \pm 3{,}92$	$13{,}00^{\dagger} \pm 1{,}97$	$33{,}92^{\dagger} \pm 4{,}64$	$30{,}76^{\dagger} \pm 0{,}52$	$50{,}99^{\dagger} \pm 3{,}95$
InfoGain	$58{,}76^{\dagger} \pm 3{,}92$	$13{,}41^{\dagger} \pm 1{,}87$	$33{,}83^{\dagger} \pm 4{,}70$	$30{,}76^{\dagger} \pm 0{,}52$	$58{,}23 \pm 6{,}83$
mRMR	$53{,}90^{\dagger} \pm 3{,}52$	$22{,}82 \pm 1{,}99$	$42{,}17 \pm 2{,}77$	$32{,}22 \pm 0{,}35$	$43{,}50^{\dagger} \pm 1{,}92$
ReliefF	$58{,}76^{\dagger} \pm 3{,}92$	$21{,}28 \pm 3{,}10$	$33{,}21^{\dagger} \pm 4{,}24$	$30{,}74^{\dagger} \pm 0{,}49$	$59{,}47 \pm 1{,}71$
SVM-RFE	$55{,}86^{\dagger} \pm 4{,}44$	$12{,}54^{\dagger} \pm 2{,}58$	$33{,}50^{\dagger} \pm 4{,}35$	$33{,}90 \pm 0{,}44$	$48{,}12^{\dagger} \pm 5{,}98$
FS-P	$54{,}38^{\dagger} \pm 3{,}11$	$12{,}00^{\dagger} \pm 1{,}90$	$33{,}50^{\dagger} \pm 4{,}42$	$34{,}16 \pm 0{,}41$	$61{,}86 \pm 1{,}33$

Cuadro 2. Umbral $\log_2(n)$: porcentaje medio estimado de error de *test*. Las cruces muestran los resultados que no son significativamente diferentes respecto al mejor método individual para cada uno de los conjunto de datos.

Ranker (Método de unión)	Yeast	Spambase	Madelon	Connect4	Isolet
Ensemble5 (SVM-Rank)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}26^{\dagger} \pm 1{,}52$	$36{,}29^{\dagger} \pm 3{,}12$	$31{,}40^{\dagger} \pm 0{,}87$	$51{,}49^{\dagger} \pm 1{,}93$
Ensemble5 (min)	$50{,}27^{\dagger} \pm 3{,}97$	$19{,}63^{\dagger} \pm 2{,}46$	$35{,}21^{\dagger} \pm 2{,}79$	$32{,}09 \pm 0{,}64$	$58{,}48 \pm 2{,}27$
Ensemble5 (median)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}67^{\dagger} \pm 1{,}57$	$36{,}63^{\dagger} \pm 3{,}86$	$31{,}51^{\dagger} \pm 0{,}66$	$52{,}79^{\dagger} \pm 2{,}57$
Ensemble5 (mean)	$54{,}18^{\dagger} \pm 2{,}72$	$13{,}02^{\dagger} \pm 3{,}70$	$36{,}21^{\dagger} \pm 3{,}61$	$31{,}36^{\dagger} \pm 0{,}75$	$51{,}22^{\dagger} \pm 1{,}81$
Ensemble5 (geommean)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}28^{\dagger} \pm 1{,}54$	$37{,}25^{\dagger} \pm 3{,}82$	$31{,}20^{\dagger} \pm 0{,}78$	$54{,}89 \pm 2{,}10$
Ensemble5 (stuart)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}28^{\dagger} \pm 1{,}54$	$37{,}04^{\dagger} \pm 4{,}26$	$31{,}16^{\dagger} \pm 0{,}70$	$51{,}65^{\dagger} \pm 1{,}15$
Ensemble5 (RRA)	$54{,}18^{\dagger} \pm 2{,}72$	$14{,}50^{\dagger} \pm 4{,}12$	$36{,}54^{\dagger} \pm 3{,}50$	$31{,}01^{\dagger} \pm 0{,}44$	$50{,}29^{\dagger} \pm 1{,}36$
Ensemble6 (SVM-Rank)	$54{,}18^{\dagger} \pm 2{,}72$	$14{,}21^{\dagger} \pm 4{,}10$	$35{,}46^{\dagger} \pm 4{,}05$	$31{,}06^{\dagger} \pm 0{,}86$	$49{,}84^{\dagger} \pm 1{,}28$
Ensemble6 (min)	$50{,}27^{\dagger} \pm 3{,}97$	$19{,}63^{\dagger} \pm 2{,}46$	$35{,}17^{\dagger} \pm 2{,}84$	$32{,}09 \pm 0{,}64$	$57{,}93 \pm 2{,}50$
Ensemble6 (median)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}87^{\dagger} \pm 1{,}60$	$34{,}92^{\dagger} \pm 3{,}83$	$31{,}41^{\dagger} \pm 0{,}65$	$48{,}93^{\dagger} \pm 1{,}78$
Ensemble6 (mean)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}35^{\dagger} \pm 1{,}55$	$36{,}00^{\dagger} \pm 3{,}06$	$31{,}28^{\dagger} \pm 0{,}68$	$50{,}62^{\dagger} \pm 0{,}82$
Ensemble6 (geommean)	$54{,}18^{\dagger} \pm 2{,}72$	$11{,}56^{\dagger} \pm 1{,}85$	$36{,}04^{\dagger} \pm 3{,}86$	$30{,}94^{\dagger} \pm 0{,}63$	$54{,}76 \pm 1{,}48$
Ensemble6 (stuart)	$54{,}18^{\dagger} \pm 2{,}72$	$12{,}69^{\dagger} \pm 3{,}82$	$36{,}37^{\dagger} \pm 3{,}50$	$31{,}16^{\dagger} \pm 0{,}71$	$51{,}55^{\dagger} \pm 2{,}21$
Ensemble6 (RRA)	$54{,}18^{\dagger} \pm 2{,}72$	$12{,}95^{\dagger} \pm 3{,}72$	$36{,}00^{\dagger} \pm 4{,}03$	$30{,}73^{\dagger} \pm 0{,}63$	$48{,}75^{\dagger} \pm 1{,}46$
Chi-Cuadrado	$59{,}97^{\dagger} \pm 5{,}18$	$13{,}17^{\dagger} \pm 1{,}02$	$34{,}29^{\dagger} \pm 3{,}34$	$30{,}81^{\dagger} \pm 0{,}58$	$75{,}06 \pm 1{,}93$
InfoGain	$55{,}13^{\dagger} \pm 4{,}99$	$13{,}39^{\dagger} \pm 1{,}24$	$33{,}62^{\dagger} \pm 3{,}50$	$30{,}76^{\dagger} \pm 0{,}54$	$48{,}62^{\dagger} \pm 2{,}30$
mRMR	$55{,}13^{\dagger} \pm 4{,}99$	$22{,}78 \pm 2{,}24$	$46{,}42 \pm 3{,}46$	$32{,}29 \pm 0{,}49$	$47{,}15^{\dagger} \pm 1{,}71$
ReliefF	$55{,}13^{\dagger} \pm 4{,}99$	$20{,}08 \pm 3{,}14$	$33{,}17^{\dagger} \pm 3{,}13$	$30{,}70^{\dagger} \pm 0{,}50$	$58{,}38 \pm 2{,}23$
SVM-RFE	$54{,}31^{\dagger} \pm 6{,}50$	$12{,}50^{\dagger} \pm 1{,}41$	$31{,}71^{\dagger} \pm 2{,}56$	$33{,}92 \pm 0{,}59$	$51{,}58^{\dagger} \pm 3{,}33$
FS-P	$54{,}66^{\dagger} \pm 4{,}33$	$12{,}17^{\dagger} \pm 1{,}52$	$33{,}96^{\dagger} \pm 2{,}94$	$34{,}18 \pm 0{,}60$	$64{,}95 \pm 4{,}53$

Cuadro 3. Umbral del 10 %: porcentaje medio estimado de error de *test*. Las cruces muestran los resultados que no son significativamente diferentes respecto al mejor método individual para cada uno de los conjunto de datos.

3.1. ¿Alguno de los Métodos Individuales Aporta Mejores Resultados que el Resto?

Tomando como referencia los resultados aportados en los Cuadros 2 y 3 y las gráficas de barras mostradas en la Figura 2, donde se comparan los métodos *ensemble* con los métodos individuales, se puede observar que no existe un método individual que mejore los resultados medios obtenidos por los *ensemble* para los diferentes escenarios planteados en este estudio. El método individual que aporta una mejor precisión media es el método *Information Gain*, el cual obtiene resultados no significativamente diferentes respecto al mejor en 9 de los 10 escenarios diferentes estudiados.

En cuanto a los *ensembles*, el *Ensemble5 (SVM-Rank)* y el *Ensemble6 (mean)* obtienen resultados no significativamente diferentes respecto al mejor en la totalidad de los escenarios estudiados.

3.2. ¿Los Métodos *Ensemble* Mejoran a los Métodos Individuales?

En vista de los resultados aportados en los Cuadros 2 y 3 y las gráficas de barras mostradas en la Figura 2, donde se comparan los métodos *ensemble* con

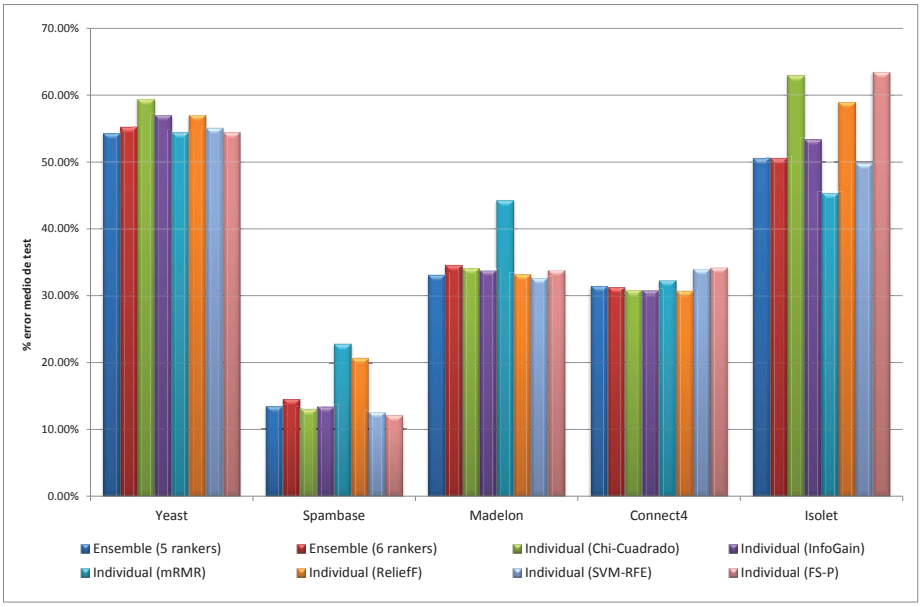

Figura 2. Comparación gráfica entre los métodos *ensemble* y los métodos individuales.

los métodos individuales, se puede observar que los modelos *ensemble* aportan los mejores resultados medios independientemente del dominio del problema estudiado. Así, los métodos *Ensemble5* obtienen resultados no significativamente diferentes respecto al mejor en 62 de las 70 configuraciones diferentes estudiadas. Por su parte, los métodos *Ensemble6* obtienen resultados no significativamente diferentes respecto al mejor en 64 de las 70 configuraciones diferentes estudiadas, por lo que, en general, la metodología *ensemble* aporta buenos resultados en el 90 % de los escenarios *ensemble* planteados en este trabajo.

Por otro lado, los métodos individuales obtienen resultados no significativamente diferentes respecto al mejor en 42 de las 60 configuraciones diferentes estudiadas, lo que indica que la metodología individual aporta resultados buenos en el 70 % de los escenarios individuales planteados en este trabajo. A pesar de que un método individual concreto puede obtener mejores resultados en un conjunto de datos específico, los métodos *ensemble* aportan mejores resultados en media e independientemente del dominio del problema estudiado, mejorando así la regularidad del proceso de selección de características.

3.3. ¿Cuál Es el Mejor Método de Unión?

Tomando como referencia los resultados aportados en los Cuadros 2 y 3 y las gráficas de barras mostradas en la Figura 3, donde se comparan los diferentes métodos de unión, se puede observar que no existe un método concreto que aporte mejores resultados que el resto para los diferentes escenarios. Si bien el

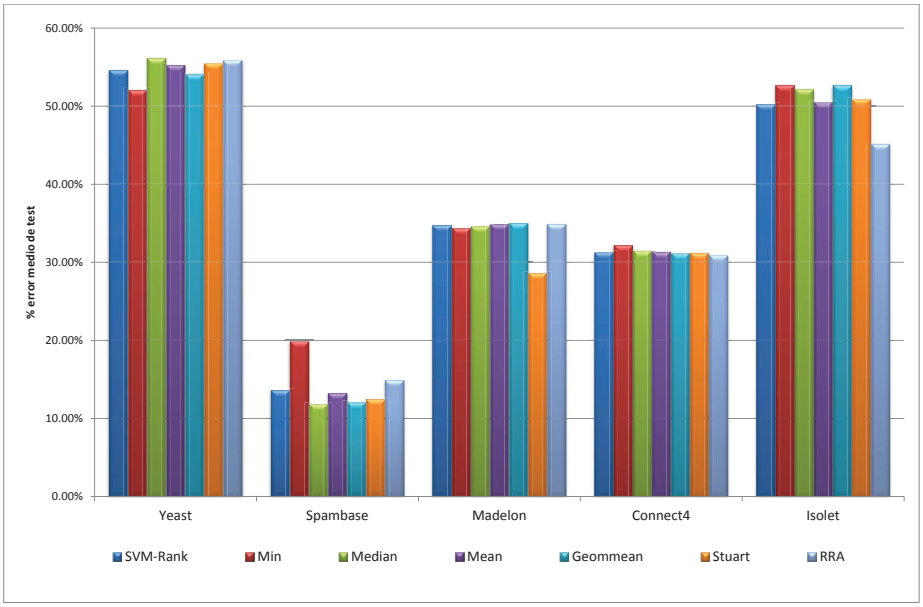

Figura 3. Comparación gráfica entre los diferentes métodos de unión.

método *Stuart* alcanza los mejores resultados sobre el conjunto de datos *Madelon*, y el método *RRA* en el conjunto de datos *Isolet*, ambos métodos no presentan resultados relevantes en los escenarios restantes. Además, el método *min* obtiene resultados dispares según el conjunto de datos analizado, siendo capaz de obtener el mejor resultado en el conjunto de datos *Yeast*, y el peor en *Spambase*. Esto se debe al tratamiento que realiza este método para resolver los empates en los *rankings* de características.

3.4. ¿Aporta alguna Mejora Añadir el Método *Chi-Cuadrado* en la Formación del *Ensemble*?

En vista de los resultados aportados en los Cuadros 2 y 3, donde se comparan los diferentes métodos *ensemble* y los métodos individuales entre si, se observa que los métodos *Ensemble5* obtienen resultados no significativamente diferentes respecto al mejor en 62 de las 70 configuraciones diferentes estudiadas, y por su parte, los métodos *Ensemble6* lo hacen en 64 de las 70 configuraciones. Así, al utilizar el método *Chi-Cuadrado* como parte del *ensemble* final se mejora la independencia del dominio del problema.

4. Conclusiones

Los conjuntos de datos pueden ser muy grandes, tanto en muestras como en número de características, y también pueden ser redundantes, ruidosos, multi-

variantes y no lineales. La mayoría de los métodos existentes por sí solos no son capaces de hacer frente a problemas con estas características, y 'el mejor método de selección de características' no existe en general, por lo que es difícil para los usuarios seleccionar un método u otro. Con el fin de hacer una elección correcta, un usuario no sólo necesita conocer bien el dominio y las características de cada conjunto de datos, sino que también debe saber detalles técnicos de los algoritmos [11]. Por tanto, los *ensembles* de selección de características surgen con el objetivo de liberar al usuario de la tarea de saber qué método es el más apropiado para un problema dado. Esto se logra gracias a la regularidad de resultados que se ha demostrado en la Sección 3.2, donde se observa como los métodos *ensemble* obtienen mejores resultados en media que los métodos individuales.

En los *ensembles* propuestos se han utilizado métodos de selección de características de tipo *ranker*, y por lo tanto ha sido necesario realizar una unión de los diferentes *rankings* aportados por cada uno de estos métodos. En este sentido, se ha visto que no existe un método de unión que aporte resultados relevantes respecto al resto, por lo que los resultados finales son independientes de la técnica de unión utilizada, al menos para los siete métodos usados en este trabajo. Se propone realizar un estudio en mayor profundidad sobre este tema en trabajos futuros.

Finalmente, los experimentos realizados han demostrado que los métodos *ensemble* propuestos obtienen de media los mejores resultados, consiguiendo una mayor independencia del dominio concreto del problema a medida que se aumenta el número y la diversidad de métodos que forman el *ensemble*.

A la vista de estos resultados, se puede liberar al usuario de decidir qué método de selección de características es más apropiado para cada problema.

5. Agradecimientos

Esta investigación ha sido respaldada en parte por el Ministerio de Economía y Competitividad del Gobierno de España a través del proyecto de investigación TIN 2012-37954, parcialmente financiada por los fondos FEDER de la Unión Europea; y por la Consellería de Industria de la Xunta de Galicia a través del proyecto de investigación GRC2014/035. V. Bolón-Canedo agradece el apoyo de la Xunta de Galicia bajo el código de la beca postdoctoral ED481B 2014/164-0.

Referencias

1. Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B., et al.: Gene prioritization through genomic data fusion. Nature biotechnology 24(5), 537–544 (2006)
2. Asuncion, A., Newman, D.: UCI machine learning repository. http://archive.ics.uci.edu/ml/datasets.html (2007)
3. Breiman, L.: Bagging predictors. Machine learning 24(2), 123–140 (1996)
4. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning 46(1-3), 389–422 (2002)

5. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.: Feature extraction. Foundations and applications (2006)
6. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 133–142. ACM (2002)
7. Kolde, R., Laur, S., Adler, P., Vilo, J.: Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 28(4), 573–580 (2012)
8. Kononenko, I.: Estimating attributes: analysis and extensions of relief. In: Machine Learning: ECML-94. pp. 171–182. Springer (1994)
9. Kuncheva, L.: Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience (2004)
10. Liu, H., Setiono, R.: Chi2: Feature selection and discretization of numeric attributes. In: 2012 IEEE 24th International Conference on Tools with Artificial Intelligence. pp. 388–388. IEEE Computer Society (1995)
11. Liu, H., Yu, L.: Toward integrating feature selection algorithms for classification and clustering. Knowledge and Data Engineering, IEEE Transactions on (2005)
12. Mejía-Lavalle, M., Sucar, E., Arroyo, G.: Feature selection with a perceptron neural net. In: Proceedings of the international workshop on feature selection for data mining. pp. 131–135 (2006)
13. Olsson, J., Oard, D.W.: Combining feature selectors for text classification. In: Proceedings of the 15th ACM international conference on Information and knowledge management. pp. 798–799. ACM (2006)
14. Park, J., Sandberg, I.W.: Universal approximation using radial-basis-function networks. Neural computation 3(2), 246–257 (1991)
15. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27(8), 1226–1238 (2005)
16. Quinlan, J.: Induction of decision trees. Machine learning 1(1), 81–106 (1986)
17. Rodríguez, D., Ruiz, R., Cuadrado-Gallego, J., Aguilar-Ruiz, J.: Detecting fault modules applying feature selection to classifiers. In: Information Reuse and Integration, 2007. IEEE International Conference on. pp. 667–672. IEEE (2007)
18. Schapire, R.: The strength of weak learnability. Machine learning 5(2) (1990)
19. Tsymbal, A., Pechenizkiy, M., Cunningham, P.: Diversity in search strategies for ensemble feature selection. Information fusion 6(1), 83–98 (2005)
20. Tukey, J.W.: Comparing individual means in the analysis of variance. Biometrics pp. 99–114 (1949)
21. Wang, H., Khoshgoftaar, T.M., Napolitano, A.: A comparative study of ensemble feature selection techniques for software defect prediction. In: Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on. IEEE (2010)
22. Willett, P.: Combination of similarity rankings using data fusion. Journal of chemical information and modeling 53(1), 1–10 (2013)
23. Yang, F., Mao, K.: Robust feature selection for microarray data based on multicriterion fusion. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 8(4), 1080–1092 (2011)
24. Yu, L., Liu, H.: Efficient feature selection via analysis of relevance and redundancy. The Journal of Machine Learning Research 5, 1205–1224 (2004)