

Nuevas Métricas de Poda basadas en Ordenamiento para Conjuntos de Clasificadores en Problemas con Clases No Balanceadas

Mikel Galar¹, Alberto Fernández², Edurne Barrenechea^{1,4}, Humberto Bustince^{1,4}, and Francisco Herrera³

¹ Departamento de Automática y Computación, Universidad Pública de Navarra, Pamplona, España

{mikel.galar,edurne.barrenechea,bustince}@unavarra.es

² Departamento de Informática, Universidad de Jaén, Jaén, España

alberto.fernandez@ujaen.es

³ Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, España

herrera@decsai.ugr.es

⁴ Institute of Smart Cities (ISC), Universidad Pública de Navarra, Pamplona, España

Resumen En los últimos años, la tarea de clasificación con datos no balanceados ha atraído un interés creciente en los investigadores. Este hecho es debido a la gran cantidad de aplicaciones y problemas reales que presentan esta característica, causando que los algoritmos de aprendizaje estándar no alcancen el rendimiento esperado. De acuerdo con lo anterior, se han diseñado muchas aproximaciones para tratar este problema desde diferentes perspectivas, por ejemplo, preprocesamiento de datos, modificación algorítmica, y aprendizaje sensible al coste.

La extensión de las técnicas anteriores a los llamados conjuntos de clasificadores o “ensembles” ha mostrado ser muy efectiva en términos de la calidad de los modelos obtenidos. Sin embargo, no es sencillo determinar el valor óptimo para el número de clasificadores a priori, lo cuál puede alterar el comportamiento del sistema. Por esta razón, se han propuesto las técnicas de poda basadas en ordenamiento para abordar este tema en problemas de aprendizaje en clasificación estándar. El problema es que estas métricas no están específicamente diseñadas para clasificación no balanceada, por lo que se degradaría el rendimiento en este contexto.

En este trabajo, proponemos dos nuevas adaptaciones para las métricas de poda basadas en el ordenamiento en el ámbito de clasificación no balanceada, concretamente la “minimización de distancias marginales” y la aproximación “basada en boosting”. A través de un completo estudio experimental, nuestro análisis muestra la bondad de ambos esquemas en contraste con los ensembles sin poda, y con las métricas de poda estándar en ensembles tipo Bagging.

Palabras clave: Conjuntos de datos no balanceados, Ensembles, Poda basada en ordenamiento, Bagging

1. Introducción

La distribución desigual entre ejemplos de diferentes clases en las tareas de clasificación es conocida como el problema de los conjuntos de datos no balanceados [8,18]. El uso de los clasificadores estándar en este marco de trabajo conlleva a soluciones no deseables, ya que el modelo estará normalmente sesgado en favor de los conceptos con mayor representación [10]. De este modo, varias propuestas han sido desarrolladas para abordar esta temática, que pueden dividirse en tres grandes grupos que incluyen el preprocesamiento para remuestrear el conjunto de datos de entrenamiento [15], la adaptación algorítmica de los métodos estándar [2], y el aprendizaje sensible al coste [21]. Adicionalmente, todos estos esquemas pueden integrarse en un algoritmo de aprendizaje basado en “ensembles” (conjuntos de clasificadores), incrementando el rendimiento de las propuestas base [7,6,10].

Un ensemble es un conjunto de clasificadores donde los clasificadores individuales que lo componen se complementan unos a otros, de tal forma que la capacidad de generalización del modelo final mejore respecto al uso de un único clasificador base [17]. Para clasificar una nueva instancia, la salida dada por todos los miembros del conjunto es considerada, obteniendo la decisión final en consenso. En contra de lo que pueda parecer, el número total de clasificadores que componen un ensemble no es sinónimo de su calidad y rendimiento [23], ya que se deben tener en cuenta diferentes cuestiones que pueden degradar su comportamiento: (1) el tiempo empleado en las etapas de aprendizaje y predicción; (2) los requisitos de memoria; y (3) contradicciones y/o redundancia entre los componentes del ensemble.

De acuerdo a lo anterior, se han desarrollado varias propuestas para llevar a cabo una poda de los clasificadores dentro un ensemble [22]. En concreto, la poda basada en ordenamiento se fundamenta en un enfoque voraz que añade clasificadores iterativamente al conjunto final de acuerdo a la maximización de una métrica o heurística, hasta que un número preestablecido de clasificadores ha sido seleccionado [9,12].

En esta contribución, nuestro objetivo es el desarrollo de una adaptación de dos métricas populares de poda para el escenario de la clasificación con conjuntos de datos no balanceados. En concreto, consideramos dos métricas diferentes para la poda: Minimización de Distancia Marginal (MDM) y métrica Basada en Boosting (BB) [5,13]. Desde nuestro punto de vista, el efecto de cada clasificador en ambas clases debe ser analizado también después de la construcción del clasificador y no sólo antes (por ejemplo, mediante el reequilibrio del conjunto de datos) lo cual obliga también a adaptar los modelos existentes para la poda de ensembles estándar.

La bondad de esta nueva propuesta es analizada por medio de un completo estudio experimental, incluyendo 66 conjuntos de datos no balanceados. Hemos seleccionado SMOTE-Bagging [19] y Under-Bagging [1] como sistemas de aprendizaje tipo ensemble ya que, a pesar de ser enfoques relativamente simples, han demostrado lograr un rendimiento más alto que otros muchos algoritmos más complejos [7]. Al igual que en otros estudios relacionados, hemos seleccionado el

algoritmo C4.5 como clasificador base [16]. Por último, nuestros resultados están apoyados por el estudio estadístico correspondiente basado en tests estadísticos no paramétricos [4].

El resto de este trabajo está organizado de la siguiente manera. La Sección 2 introduce brevemente el problema de los conjuntos de datos no balanceados. Posteriormente, la Sección 3 presenta la metodología de poda basada en ordenamiento, describiendo las métricas estándar para la realización de este proceso y nuestras adaptaciones a la clasificación no balanceada. A continuación, los detalles sobre el marco experimental son descritos en la Sección 4. El análisis y la discusión de los resultados experimentales se lleva a cabo en la Sección 5. Finalmente, la Sección 6 resume y concluye este trabajo.

2. Conceptos Básicos sobre Clasificación con Conjuntos de Datos No Balanceados

La clasificación con conjuntos de datos no balanceados aparece cuando la distribución de instancias entre las clases de un problema dado difiere notablemente [10,15]. Por lo tanto, este problema de clasificación requiere de un tratamiento especial con el fin de llevar a cabo una discriminación precisa entre ambos conceptos, con independencia de su representación.

La presencia de clases con pocos datos puede generar modelos sub-óptimos de clasificación, debido a que hay un sesgo hacia la clase mayoritaria cuando se guía el proceso de aprendizaje mediante métricas estándar de precisión. Además, estudios recientes han demostrado que otras características intrínsecas de los datos tienen una influencia significativa al tratar de llevar a cabo una correcta identificación de los ejemplos de la clase minoritaria [10]. Algunos ejemplos son el solapamiento, pequeños subconjuntos de datos disjuntos, ruido y el llamado “dataset shift”.

Las soluciones desarrolladas para hacer frente a este problema se pueden clasificar en tres grandes grupos [10]: (1) *a nivel de datos* [15], (2) *a nivel algorítmico* [2] y (3) *sensibles al coste* [21]. Además, cuando los anteriores enfoques se integran en un conjunto de clasificadores (ensemble), se ha demostrado que su eficacia se ve reforzada [7,10].

Por último, con el fin de evaluar el rendimiento en este escenario particular de clasificación, las métricas utilizadas deben ser diseñadas para tener en cuenta la distribución de clases. Una alternativa comúnmente considerada es el área bajo la curva ROC (AUC). En aquellos casos en los que el clasificador utilizado da como salida una sola solución, esta medida puede calcularse simplemente mediante la siguiente fórmula:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (1)$$

donde $TP_{rate} = \frac{TP}{TP+FN}$ y $FP_{rate} = \frac{FP}{TN+FP}$.

3. Una Propuesta sobre Poda Basada en Ordenamiento para Ensembles en Dominios No Balanceados

Los clasificadores basados en Ensembles [17] están generalmente compuestos por un conjunto de los llamados *clasificadores débiles*, es decir, aquellos en los que pequeños cambios en los datos producen cambios destacados en los modelos inducidos. Esta propiedad es interesante ya que permite alcanzar la diversidad en el ensemble, lo que resulta importante para poder alcanzar un buen rendimiento. Entre los diferentes modelos de ensemble, dos de las técnicas más conocidas y utilizadas son Bagging [3] y Boosting [5].

En este trabajo, nos centraremos en el primer esquema debido su simplicidad para combinarlo con las técnicas de preprocesamiento de datos [7]. En esta metodología, cada clasificador del conjunto es entrenado con diferentes conjuntos aleatorios de ejemplos obtenido de los iniciales de entrenamiento. Al clasificar una nueva muestra, todos los clasificadores individuales son consultados, utilizando un método de voto mayoritario o ponderado para inferir la clase final.

El primer parámetro a tener en cuenta en la construcción de este tipo de modelos es el número de clasificadores considerados en el conjunto. En este sentido, los métodos de poda fueron diseñados para obtener el número “óptimo” de clasificadores, llevando a cabo una selección a partir de un grupo determinado de componentes del ensemble. La hipótesis es que la precisión aumenta generalmente de forma monótona a medida que se añaden más elementos al conjunto [9,12,13]. La mayor parte de las técnicas de poda hacen uso de una función heurística para buscar este conjunto reducido de clasificadores. En el caso de *poda basada en ordenamiento*, se ha de definir una métrica para medir la bondad que resulta de añadir cada clasificador al ensemble, de modo que el clasificador con el valor más alto es incluido al subconjunto final. Este mismo proceso se lleva a cabo hasta que el tamaño del subconjunto alcanza un valor de un parámetro especificado a priori.

En este trabajo, estudiamos dos métricas de poda populares, MDM y BB [5,13]. A continuación, describimos ambos esquemas y nuestra adaptación a la clasificación no balanceada:

- *MDM* se basa en ciertas distancias entre los vectores de salida de los ensembles. Estos vectores de salida tienen la longitud igual al tamaño del conjunto de entrenamiento, y su valor en la posición i -ésimo es 1 o -1 dependiendo de si el i -ésimo ejemplo se clasifica correctamente o incorrectamente por el clasificador. El vector signatura de un subconjunto se calcula como la suma de los vectores de los clasificadores seleccionados. En resumen, el objetivo es sumar los clasificadores con el objetivo de obtener un vector signatura del subconjunto en el que todos los componentes sean positivos, es decir, todos los ejemplos sean correctamente clasificados. Para una descripción más amplia consulte [12].

Este método selecciona el clasificador que se añade en función de la distancia euclídea más cercana entre un punto objetivo (donde cada uno de los componentes son positivos) y el vector signatura del subconjunto después

de haber agregado el clasificador correspondiente. Como consecuencia, cada ejemplo tiene el mismo peso en el cálculo de la distancia, lo cual puede sesgar la selección a aquellos clasificadores que favorecen la clase mayoritaria. Por lo tanto, nuestra adaptación consiste en calcular la distancia de los ejemplos de la clase mayoritaria y los ejemplos de la clase minoritaria de forma independiente. Luego, las distancias son normalizadas por el número de ejemplos utilizados para calcularlas y son posteriormente agregadas. Es decir, damos el mismo peso a ambas clases en el cálculo de la distancia. Este nueva métrica la denominamos *MDM-NB*.

- *BB* selecciona el clasificador que minimiza el coste con respecto al esquema de boosting. Esto significa que el algoritmo de boosting se aplica para calcular los pesos (costes) para cada ejemplo en cada iteración. Sin embargo, en vez de entrenar un clasificador con estos pesos, se añade al subconjunto el clasificador que obtiene el menor coste de los seleccionables del conjunto original, y los pesos se actualizan consecuentemente. Por lo tanto, resulta indiferente el hecho de que los clasificadores individuales hayan sido aprendidos mediante boosting o no. A diferencia del método original de boosting, cuando ningún clasificador obtiene un error de entrenamiento ponderado por debajo del 50 %, los pesos se reinician (pesos iguales para todos los ejemplos) y el método continúa (mientras que en boosting este sería detenido). Una vez seleccionados el subconjunto de clasificadores final, los pesos asignados a cada clasificador durante el proceso de boosting son olvidados, no teniéndose en cuenta en la fase de agregación.

Es sabido que el boosting por sí mismo no es capaz de gestionar el problema del no balanceo de clases [7]. Por esta razón, también hemos adaptado este enfoque de una manera similar al caso de *MDM*. En boosting, cada ejemplo tiene inicialmente el mismo peso y éstos se actualizan en función de si están correctamente clasificados o no. A pesar que las instancias de la clase minoritaria deberían recibir pesos mayores si están mal clasificadas, estos pesos pueden ser insignificantes en comparación con los de los ejemplos de la clase mayoritaria. Por tanto, antes de encontrar el clasificador que minimice el coste total, los pesos de los ejemplos de cada clase se normalizan por medio de su suma, de modo que ambas clases tendrán la misma importancia en el momento de seleccionar el clasificador (a pesar de que cada ejemplo de cada clase tendrá un peso diferente). Esto solo se hace antes de seleccionar el clasificador, y posteriormente los pesos se actualizan de acuerdo con los originales (los no normalizados). Este procedimiento trata de ser similar a los aplicados con éxito en varios modelos de boosting como EUS-Boost [6]. Este segundo enfoque de ponderación lo denominamos *BB-NB*.

4. Marco Experimental

La Tabla 1 muestra los problemas de referencia seleccionados para nuestro estudio, en la que se muestra el nombre, número de ejemplos, número de atributos e IR (relación entre las instancias de la mayoría y de clase minoritaria). Los conjuntos de datos están ordenados con respecto a su grado de no

balanceo. Una descripción más amplia de estos problemas se puede encontrar en <http://www.keel.es/datasets.php>.

Tabla 1. Resumen de los conjuntos de datos no balanceados utilizados

Nombre	#Ej.	#Atrs.	IR	Nombre	#Ej.	#Atrs.	IR
glass1	214	9	1.82	glass04vs5	92	9	9.22
ecoli0vs1	220	7	1.86	ecoli0346vs5	205	7	9.25
wisconsin	683	9	1.86	ecoli0347vs56	257	7	9.28
pima	768	8	1.87	yeast05679vs4	528	8	9.35
iris0	150	4	2.00	ecoli067vs5	220	6	10.00
glass0	214	9	2.06	vowel0	988	13	10.10
yeast1	1484	8	2.46	glass016vs2	192	9	10.29
vehicle2	846	18	2.52	glass2	214	9	10.39
vehicle1	846	18	2.52	ecoli0147vs2356	336	7	10.59
vehicle3	846	18	2.52	led7digit02456789vs1	443	7	10.97
haberman	306	3	2.78	ecoli01vs5	240	6	11.00
glass0123vs456	214	9	3.19	glass06vs5	108	9	11.00
vehicle0	846	18	3.25	glass0146vs2	205	9	11.06
ecoli1	336	7	3.36	ecoli0147vs56	332	6	12.28
newthyroid2	215	5	4.92	cleveland0vs4	1771	13	12.62
newthyroid1	215	5	5.14	ecoli0146vs5	280	6	13.00
ecoli2	336	7	5.46	ecoli4	336	7	13.84
segment0	2308	19	6.01	shuttle0vs4	1829	9	13.87
glass6	214	9	6.38	yeast1vs7	459	8	13.87
yeast3	1484	8	8.11	glass4	214	9	15.47
ecoli3	336	7	8.19	pageblocks13vs4	472	10	15.85
pageblocks0	5472	10	8.77	abalone918	731	8	16.68
ecoli034vs5	200	7	9.00	glass016vs5	184	9	19.44
yeast2vs4	514	8	9.08	shuttle2vs4	129	9	20.50
ecoli067vs35	222	7	9.09	yeast1458vs7	693	8	22.10
ecoli0234vs5	202	7	9.10	glass5	214	9	22.81
glass015vs2	506	8	9.12	yeast2vs8	482	8	23.10
yeast0359vs78	172	9	9.12	yeast4	1484	8	28.41
yeast0256vs3789	1004	8	9.14	yeast1289vs7	947	8	30.56
yeast02579vs368	1004	8	9.14	yeast5	1484	8	32.73
ecoli046vs5	203	6	9.15	yeast6	1484	8	41.40
ecoli01vs235	244	7	9.17	ecoli0137vs26	281	7	39.15
ecoli0267vs35	244	7	9.18	abalone19	4174	8	129.44

Las estimaciones de la medida AUC se obtienen por medio de una validación cruzada por medio de una Distribución Óptima Balanceada (DOB-SCV) [14], tal y como se sugiere en la literatura especializada para trabajar en clasificación no balanceada [11]. El procedimiento de validación cruzada se realiza con 5 particiones, con el objetivo de incluir suficientes instancias de la clase positiva en los diferentes bloques. De acuerdo con la naturaleza estocástica de los métodos de aprendizaje, estos 5 bloques se generan con 5 semillas diferentes. Por lo tanto, los resultados experimentales se calculan con el promedio de 125 ejecuciones.

Como técnicas de ensembles, haremos uso de SMOTE-Bagging [19] y Under-Bagging [1]. Con el fin de aplicar el procedimiento de poda, vamos a aprender un total de 100 *clasificadores* por ensemble, seleccionando un subconjunto de sólo 21 *clasificadores* como se sugiere en la literatura especializada [12]. Los modelos base de ensembles para la comparación (en los que no se realiza la poda) utilizarán 40 clasificadores como se recomienda en [7]. Para *SMOTE-Bagging*, la configuración de SMOTE será la estándar con una distribución de *clases* al 50%, 5 vecinos para generar las muestras sintéticas y la métrica HVDM para calcular la distancia entre los ejemplos. Por último, ambos enfoques de aprendizaje incluyen el árbol de decisión C4.5 [16] como clasificador de referencia, con un nivel de confianza

de 0, 25, y 2 como el número mínimo de la item-sets por hoja, usando la poda para obtener el árbol final. El lector puede referirse a [7] con el fin de obtener una descripción detallada de los anteriores métodos de ensemble.

Finalmente, haremos uso del test de signos-rangos de Wilcoxon [20] para detectar si existen diferencias significativas entre una pareja de algoritmos.

5. Estudio Experimental

Nuestro análisis se centra en determinar si las nuevas métricas propuestas, diseñadas específicamente para hacer frente al no balanceo de clases se adaptan bien a este problema con respecto a las métricas originales, es decir, *BB* y *MDM*. Además, vamos a analizar la mejora en el rendimiento de los resultados con respecto a los modelos de ensemble originales. Los resultados completos en términos de la medida AUC en el conjunto de test se muestran en la Tabla 2, mientras que los valores promedio de estos resultados experimentales se muestran en la Tabla 3.

Atendiendo a la comparación entre los esquema de poda, en el caso de *BB* y *BB-NB* encontramos que para SMOTE-Bagging la métrica adaptada para la clasificación no balanceada logra un rendimiento medio superior. Respecto a Under-Bagging, las diferencias relativas son inferiores al 1 % a favor del enfoque estándar. Por otro lado, el análisis de las métricas *MDM* y *MDM-NB* muestra la necesidad del uso de un enfoque para clases no balanceadas de acuerdo a los resultados experimentales. Por último, la robustez de las métricas desequilibradas debe subrayarse de acuerdo con la baja desviación estándar que se muestra con respecto al caso estándar.

Para determinar la métrica más adecuada de forma estadística, realizamos un test de pares de Wilcoxon en la Tabla 4. Hemos incluido un símbolo para destacar si las diferencias significativas se encuentran en el 95 % de grado de confianza (*) o al 90 % (+). Los resultados de estos test están de acuerdo con nuestras anteriores observaciones. Las diferencias en el caso de MDM son claras en favor de la versión no balanceada. En el caso de BB el comportamiento varía dependiendo de la técnica de ensemble, obteniéndose diferencias significativas para SMOTE-Bagging mientras que no se encuentra ninguna en Under-Bagging.

Por último, cuando contrastamos estos resultados frente al enfoque estándar de ensemble, también observamos un doble comportamiento: en el caso de SMOTE-Bagging nuestro enfoque de poda permite la definición de un conjunto más simple (al disponer de menos clasificadores) con la mínima disminución del rendimiento, especialmente cuando se selecciona la métrica BB-NB. En el caso de la métrica MDM-NB, la sinergia con SMOTE-Bagging no es tan positiva, mostrando una pérdida de precisión significativa. Por otro lado, para Under-Bagging observamos una notoria mejora de los resultados en todos los casos en que se aplica la poda basada en ordenamiento, mostrando un mejor comportamiento con MDM-NB y especialmente con BB-NB (ver Tablas 2 y 3). Estos resultados se complementan mediante un test de Wilcoxon (mostrado en la Tabla 5), en el que observamos

Tabla 2. Resultados en test para la propuesta estándar de ensemble (Base) y la poda basada en orden con las métricas originales (BB y MDM) y no balanceadas (BB-NB) y MDM-NB) usando la medida AUC.

Problema	SMOTE-Bagging					Under-Bagging				
	Base	BB	BB-NB	MDM	MDM-NB	Std.	BB	BB-NB	MDM	MDM-NB
glass1	.7675	.8021	.7925	.7866	.7900	.7686	.7979	.7927	.7918	.7928
ecoli0vs1	.9812	.9750	.9763	.9802	.9788	.9805	.9806	.9764	.9826	.9809
wisconsin	.9707	.9692	.9700	.9662	.9666	.9691	.9698	.9704	.9678	.9672
pima	.7568	.7451	.7546	.7500	.7558	.7598	.7561	.7548	.7532	.7539
iris0	.9880	.9888	.9880	.9880	.9880	.9900	.9900	.9900	.9900	.9900
glass0	.8347	.8517	.8464	.8413	.8430	.8264	.8469	.8438	.8399	.8352
yeast1	.7312	.7192	.7321	.7301	.7315	.7304	.7333	.7310	.7331	.7307
vehicle2	.9723	.9752	.9734	.9686	.9691	.9704	.9750	.9744	.9680	.9686
vehicle1	.7848	.7691	.7918	.7898	.7934	.8016	.8020	.7983	.7959	.7985
vehicle3	.7784	.7593	.7827	.7795	.7808	.8060	.7979	.7976	.7966	.7974
haberman	.6627	.6517	.6476	.6500	.6498	.6627	.6616	.6486	.6488	.6620
glass0123vs456	.9405	.9318	.9357	.9308	.9378	.9335	.9432	.9379	.9264	.9337
vehicle0	.9635	.9630	.9636	.9609	.9614	.9492	.9558	.9595	.9539	.9544
ecoli1	.9053	.8988	.9067	.9044	.9107	.8988	.8981	.9101	.9043	.9123
newthyroid2	.9642	.9540	.9586	.9567	.9577	.9605	.9572	.9696	.9614	.9692
newthyroid1	.9558	.9460	.9486	.9456	.9467	.9490	.9479	.9550	.9594	.9613
ecoli2	.9145	.9153	.9128	.9131	.9099	.9054	.9057	.8996	.9017	.8996
segment0	.9917	.9917	.9924	.9922	.9926	.9866	.9881	.9887	.9872	.9878
glass6	.9291	.9164	.9213	.9157	.9203	.9096	.9277	.9248	.9228	.9190
yeast3	.9330	.9308	.9325	.9315	.9329	.9311	.9326	.9305	.9311	.9295
ecoli3	.8462	.8508	.8560	.8506	.8514	.8830	.8702	.8670	.8793	.8707
pageblocks0	.9580	.9552	.9585	.9572	.9581	.9610	.9631	.9626	.9612	.9615
ecoli034vs5	.9129	.9032	.9018	.9029	.8948	.8922	.9148	.9203	.8701	.9037
yeast2vs4	.9277	.9192	.9155	.9123	.9223	.9445	.9408	.9482	.9383	.9536
ecoli067vs35	.8576	.8651	.8626	.8653	.8630	.8582	.8624	.8578	.8670	.8523
ecoli0234vs5	.9007	.9008	.9036	.8935	.8939	.8641	.9053	.9027	.8404	.8784
glass015vs2	.7041	.7004	.7015	.7052	.7025	.7412	.7117	.7604	.7553	.7628
yeast0359vs78	.7173	.7023	.7174	.7016	.7134	.7373	.7414	.7386	.7394	.7387
yeast02579vs368	.8028	.7982	.7995	.7927	.7993	.8159	.8090	.8068	.8136	.8075
yeast0256vs3789	.9183	.9173	.9176	.9150	.9185	.9149	.9136	.9099	.9140	.9098
ecoli046vs5	.9132	.9086	.9114	.9046	.9083	.8869	.9188	.9238	.8666	.9123
ecoli01vs235	.8988	.8665	.8815	.8789	.8883	.8815	.9031	.9047	.8893	.8942
ecoli0267vs35	.8617	.8544	.8611	.8664	.8642	.8573	.8623	.8556	.8662	.8483
glass04vs5	.9910	.9836	.9879	.9876	.9869	.9940	.9900	.9940	.9940	.9940
ecoli0346vs5	.8921	.8888	.8929	.8762	.8884	.8799	.8961	.9051	.8618	.8956
ecoli0347vs56	.8595	.8701	.8707	.8590	.8643	.8762	.8875	.8897	.9009	.8800
yeast05679vs4	.8177	.8152	.8133	.8088	.8124	.8209	.8287	.8189	.8018	.8182
ecoli067vs5	.8897	.8894	.8888	.8909	.8886	.8820	.8883	.8888	.9028	.8779
vowel0	.9878	.9874	.9880	.9838	.9853	.9588	.9671	.9684	.9689	.9685
glass016vs2	.7009	.7083	.7176	.7168	.7214	.7025	.7185	.7291	.7265	.7323
glass2	.7425	.7390	.7436	.7458	.7458	.7569	.7394	.7691	.7452	.7702
ecoli0147vs2356	.8685	.8637	.8719	.8673	.8793	.8328	.8625	.8536	.8665	.8468
led7digit02456789vs1	.8466	.8547	.8407	.8500	.8383	.8268	.8397	.8322	.8449	.8399
ecoli01vs5	.8881	.8786	.8782	.8688	.8755	.8726	.9142	.9174	.8705	.8937
glass06vs5	.9926	.9954	.9954	.9916	.9912	.9151	.9910	.9940	.9940	.9940
glass0146vs2	.6961	.7161	.7295	.7189	.7254	.7214	.7335	.7336	.7323	.7434
ecoli0147vs56	.8703	.8848	.8804	.8682	.8750	.8738	.9035	.8870	.8819	.8756
cleveland0vs4	.7894	.7933	.8004	.7815	.7835	.8492	.8714	.8305	.7917	.8069
ecoli0146vs5	.8875	.9037	.9022	.8828	.8994	.8933	.9197	.9273	.8639	.8988
ecoli4	.9245	.9220	.9247	.9094	.9135	.8952	.9357	.9349	.9017	.8969
shuttle0vs4	.9999	.9999	.9999	.9999	.9999	1.0000	1.0000	1.0000	1.0000	1.0000
yeast1vs7	.7458	.7354	.7349	.7368	.7303	.7661	.7869	.7852	.7463	.7824
glass4	.9069	.8795	.8788	.8716	.8675	.9065	.9182	.8903	.8943	.8882
pageblocks13vs4	.9952	.9932	.9964	.9963	.9963	.9804	.9937	.9946	.9928	.9928
abalone9vs18	.7120	.7140	.7076	.7090	.7085	.7560	.7490	.7388	.7222	.7354
glass016vs5	.9865	.9493	.9747	.9675	.9674	.9429	.9698	.9675	.9670	.9663
shuttle2vs4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
yeast1458vs7	.6330	.6175	.6144	.6059	.6153	.6374	.6530	.6263	.6009	.6315
glass5	.9769	.9533	.9619	.9586	.9626	.9488	.9596	.9639	.9631	.9621
yeast2vs8	.8064	.7916	.7946	.8014	.8068	.7526	.7846	.7608	.7579	.7629
yeast4	.8211	.8117	.8114	.8046	.8124	.8420	.8534	.8537	.8416	.8543
yeast1289vs7	.7046	.6818	.6905	.6831	.7004	.7370	.7194	.7392	.6918	.7433
yeast5	.9622	.9536	.9581	.9525	.9585	.9593	.9689	.9673	.9623	.9625
yeast6	.8375	.8354	.8446	.8369	.8431	.8673	.8736	.8570	.8706	.8514
ecoli0137vs26	.8347	.8273	.8336	.8363	.8400	.7807	.8774	.7874	.8060	.7789
abalone19	.5432	.5380	.5447	.5375	.5462	.7121	.7034	.7251	.7213	.7307
Promedio	.8645	.8602	.8635	.8596	.8625	.8647	.8755	.8734	.8653	.8699

Tabla 3. Resultados promedio en test para la propuesta estándar de ensemble (Base) y la poda basada en orden con las métricas originales (BB y MDM) y no balanceadas (BB-NB) y MDM-NB).

Ensemble	Base	BB	BB-NB	MDM	MDM-NB
SMOTE-Bagging	.8645 ± .0587	.8602 ± .0632	.8635 ± .0610	.8596 ± .0629	.8625 ± .0622
Under-Bagging	.8647 ± .0516	.8755 ± .0564	.8734 ± .0544	.8653 ± .0563	.8699 ± .0558

Tabla 4. Test de Wilcoxon para métricas de poda: estándar [R^+] y no balanceado [R^-].

Ensemble	Comparativa	R^+	R^-	p -valor
SMOTE-Bagging	BB vs. BB-NB	540.0	1671.0	0.00028*
	MDM vs. MDM-NB	436.0	1775.0	0.00002
Under-Bagging	BB vs. BB-NB	1277.0	934.0	0.27939
	MDM vs. MDM-NB	831.5	1379.5	0.07246+

Un "+" o un "*" junto al p -valor indica que existen diferencias estadísticas significativas con un nivel de confianza del 90% y el 95%, respectivamente.

diferencias significativas a favor de la poda basada en ordenamiento para el enfoque Under-Bagging.

Tabla 5. Test de Wilcoxon para comparar el enfoque estándar de ensemble (Std.) [R^+] y aquél con la poda no balanceada basada en orden [R^-].

Ensemble	Comparativa	R^+	R^-	p -valor
SMOTE-Bagging	Std. vs. BB-NB	1261.5	883.5	0.215579
	Std. vs. MDM-NB	1386.5	758.5	0.039856*
Under-Bagging	Std. vs. BB-NB	502.0	1709.0	0.000114*
	Std. vs. MDM-NB	637.0	1574.0	0.002735*

Un "+" o un "*" junto al p -valor indica que existen diferencias estadísticas significativas con un nivel de confianza del 90% y el 95%, respectivamente.

6. Conclusiones

La poda basada en ordenamiento para ensembles consiste en llevar a cabo una selección de aquéllos elementos del conjunto de los que espera que trabajen con una mejor sinergia. El proceso anterior se guía mediante una métrica de rendimiento que se centra en diferentes propiedades del ensemble. Sin embargo, no se ha considerado ningún desarrollo previo de las mismas dentro del escenario de los conjuntos de datos no balanceados.

En este trabajo, hemos propuesto dos adaptaciones de las métricas para poda basada en ordenamiento en el ámbito de la clasificación no balanceada, en

particular BB-NB y MDM-NB. El análisis experimental ha demostrado el éxito de estas nuevas métricas con respecto a su definición original, especialmente en el caso del enfoque SMOTE-Bagging. Además, hemos resaltado la obtención de una mejora significativa en el comportamiento del modelo Under-Bagging por medio de la aplicación de la poda basada en orden, superando los resultados con respecto al modelo original.

Como trabajo futuro planeamos incluir un mayor número de métricas de poda y metodologías de aprendizaje de ensembles para conjuntos de datos no balanceados con el objetivo de dar un mayor soporte y fuerza a los resultados obtenidos en esta contribución.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología de España con los proyectos TIN-2011-28488, TIN-2012-33856, TIN2013-40765-P; los planes de investigación de Andalucía P11-TIC-7765 y P10-TIC-6858; y tanto por la Universidad de Jaén y la Caja Rural Provincial de Jaén bajo el proyecto UJA2014/06/15.

Referencias

1. Barandela, R., Valdovinos, R., Sánchez, J.: New applications of ensembles of classifiers. *Pattern Analysis Applications* 6(3), 245–256 (2003)
2. Barandela, R., Sánchez, J.S., García, V., Rangel, E.: Strategies for learning in class imbalance problems. *Pattern Recognition* 36(3), 849–851 (2003)
3. Breiman, L.: Bagging predictors. *Machine Learning* 24(2), 123–140 (1996)
4. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7, 1–30 (2006)
5. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1), 119–139 (1997)
6. Galar, M., Fernández, A., Barrenechea, E., , Herrera, F.: EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition* 46(12), 3460–3471 (2013)
7. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: A review on ensembles for class imbalance problem: Bagging, boosting and hybrid based approaches. *IEEE Transactions on System, Man and Cybernetics Part C: Applications and Reviews* 42(4), 463–484 (2012)
8. He, H., Garcia, E.A.: Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 21(9), 1263–1284 (2009)
9. Hernández-Lobato, D., Martínez-Muñoz, G., Suárez, A.: Statistical instance-based pruning in ensembles of independent classifiers. *IEEE Transactions On Pattern Analysis And Machine Intelligence* 31(2), 364–369 (2009)
10. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences* 250(20), 113–141 (2013)

11. López, V., Fernández, A., Herrera, F.: On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences* 257, 1–13 (2014)
12. Martínez-Muñoz, G., Hernández-Lobato, D., Suárez, A.: An analysis of ensemble pruning techniques based on ordered aggregation. *IEEE Transactions On Pattern Analysis And Machine Intelligence* 31(2), 245–259 (2009)
13. Martínez-Muñoz, G., Suárez, A.: Using boosting to prune bagging ensembles. *Pattern Recognition Letters* 28(1), 156–165 (2007)
14. Moreno-Torres, J.G., Sáez, J.A., Herrera, F.: Study on the impact of partition-induced dataset shift on k-fold cross-validation. *IEEE Transactions On Neural Networks And Learning Systems* 23(8), 1304–1313 (2012)
15. Prati, R.C., Batista, G.E.A.P.A., Silva, D.F.: Class imbalance revisited: a new experimental setup to assess the performance of treatment methods. *Knowledge and Information Systems in press*, 1–25 (2014)
16. Quinlan, J.: *C4.5: Programs for Machine Learning*. Morgan Kaufman (1993)
17. Rokach, L.: Ensemble-based classifiers. *Artificial Intelligence Review* 33(1), 1–39 (2010)
18. Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence* 23(4), 687–719 (2009)
19. Wang, S., Yao, X.: Diversity analysis on imbalanced data sets by using ensemble models. In: *Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining (CIDM'09)*. pp. 324–331 (2009)
20. Wilcoxon, F.: Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6), 80–83 (1945)
21. Zadrozny, B., Langford, J., Abe, N.: Cost-sensitive learning by cost-proportionate example weighting. In: *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*. pp. 435–442 (2003)
22. Zhang, Y., Burer, S., Street, W.N.: Ensemble pruning via semi-definite programming. *Journal of Machine Learning Research* 7, 1315–1338 (2006)
23. Zhou, Z.H., Wu, J., Tang, W.: Ensembling neural networks: Many could be better than all. *Artificial Intelligence* 137(1-2), 239–263 (2002)