

# Un primer estudio sobre el uso de aprendizaje sensible al coste con sistemas de clasificación basados en reglas difusas para problemas no balanceados

Victoria López

Dept. de Ciencias de la Computación  
e Inteligencia Artificial

Univ. de Granada

viclml@correo.ugr.es

Alberto Fernández

Dept. de Informática

Univ. de Jaén

alberto.fernandez@ujaen.es

Francisco Herrera

Dept. de Ciencias de la Computación  
e Inteligencia Artificial

Univ. de Granada

herrera@decsai.ugr.es

## Resumen

El problema de clasificación en dominios no balanceados se ha convertido en la actualidad en uno de los problemas más relevantes dentro del área de aprendizaje automático. Este problema ha ganado importancia debido a su presencia en muchas aplicaciones reales y ocurre cuando la distribución de ejemplos disponibles para realizar el aprendizaje es muy diferente entre sus clases. Habitualmente, la clase con menor representación es el objeto de interés del problema, por lo que el coste derivado de una mala clasificación de estos ejemplos es mayor que el de los ejemplos restantes.

En este trabajo analizamos el comportamiento de un método de aprendizaje para sistemas de clasificación basados en reglas difusas sensible al coste en un escenario de conjuntos de datos con un no balanceo alto. Se trata de utilizar una de las alternativas disponibles para resolver problemas de clasificación en este tipo de dominios y estudiar su potencial frente a otras posibilidades.

Los resultados experimentales muestran la competitividad del método de aprendizaje difuso sensible al coste propuesto obteniendo un buen equilibrio en la precisión de ambas clases, frente a otras propuestas específicas para problemas de no balanceo basadas en el preprocesamiento del conjunto de datos inicial para equilibrar la distribución de las clases.

## 1. Introducción

Dentro de los problemas de clasificación, un conjunto de datos no balanceado es aquel en el que el número de instancias para cada una de las clases difiere mucho entre ellas. El no balanceo en la distribución de las clases es dominante en un gran número de problemas reales como por ejemplo, en telecomunicaciones, WWW, finanzas, ecología, biología y medicina. La clase positiva o minoritaria suele ser la que tiene más interés desde el punto de vista del aprendizaje y además suele venir acompañada de un coste elevado cuando se cometen errores [7].

Las soluciones que se han presentado al problema se desarrollan tanto en el nivel de datos como en el nivel algorítmico. A nivel de datos, el objetivo consiste en equilibrar la distribución de las clases muestreando el espacio de datos [2, 3]. A nivel algorítmico, las soluciones tratan de adaptar diversos algoritmos de clasificación para reforzar el aprendizaje hacia la clase positiva [27]. Las soluciones sensibles al coste incorporan enfoques a nivel de datos, a nivel algorítmico, o a estos dos niveles conjuntamente, suponiendo costes mayores para la clasificación errónea de ejemplos de la clase positiva frente a la clase negativa buscando minimizar los errores de coste elevado [22, 28, 29].

En este trabajo nos centraremos en problemas de clasificación binarios no balanceados con un alto porcentaje de no balanceo.

Para diseñar el clasificador utilizaremos Sistemas de Clasificación Basados en Reglas Difusas (SCBRDs), siendo ésta una herramienta muy utilizada en el campo del aprendizaje automático ya que proporciona modelos interpretables para el usuario final [15]. Trabajos recientes han demostrado que los SCBRDs tienen un buen comportamiento trabajando con conjuntos de datos no balanceados mediante la aplicación de técnicas de preprocesamiento de instancias[9].

Dados estos precedentes, en este trabajo se pretende analizar el impacto del aprendizaje sensible al coste en dicha clase de problemas, para lo que desarrollaremos un método de aprendizaje difuso sensible al coste que nos permita obtener la base de reglas del SCBRDs. Este método se basa en el algoritmo propuesto en [4] (método de Chi en adelante), el cual se ve modificado para tener en cuenta el coste de clasificación errónea de los diversos ejemplos. Los resultados obtenidos con esta metodología se compararán con los obtenidos por el propio método de Chi original así como con el método de Chi al que se le aplica el método de preprocesamiento “Synthetic Minority Over-sampling Technique” (SMOTE) [3].

Para el estudio experimental hemos seleccionado un conjunto de 22 problemas con alto grado de no balanceo obtenidos del repositorio UCI [1]. Además, se llevará a cabo un estudio estadístico utilizando test no paramétricos para comprobar si existen diferencias significativas entre los resultados obtenidos [5, 10].

El trabajo está estructurado de la siguiente manera. Primero, en la Sección 2 se realiza una introducción a los problemas que plantean los conjuntos de datos no balanceados, describiendo sus características, cómo se puede trabajar con ellos, el aprendizaje sensible al coste y las métricas que se emplean en dicho campo. A continuación, en la Sección 3 presentaremos los aspectos fundamentales de un método de aprendizaje difuso para la construcción de SCBRDs a partir del cual proponemos una metodología de aprendizaje difuso sensible al coste. La sección 4 muestra el estudio experimental realizado. Finalmente, las conclusiones de este trabajo se exponen en la Sección 5.

## 2. Conjuntos de datos no balanceados en clasificación

En esta sección delimitamos el contexto en el que se trabaja introduciendo brevemente el problema de clasificación no balanceada así como el aprendizaje sensible al coste. Terminamos describiendo las métricas de evaluación que experimentan un cambio en este problema concreto con respecto a métricas habituales en clasificación.

### 2.1. El problema de clasificación no balanceada

El aprendizaje a partir de datos no balanceados es un tema importante que ha aparecido recientemente en el ámbito del aprendizaje automático [12, 20]. Concretamente, nos referimos a conjuntos de datos no balanceados cuando la distribución de clases no es uniforme, por lo que el número de ejemplos que representa a una de las clases es mucho menor que en las otras clases, además de que la caracterización de esa clase suele tener un mayor interés práctico. La importancia de este problema radica en que está presente en numerosos problemas reales de clasificación como por ejemplo en detección de parásitos de la malaria con imágenes [21], detección de minas mediante radar y sonar [26] o el estudio de las contracciones intestinales [23] por citar algunos de ellos.

Los algoritmos clásicos para clasificación a partir de ejemplos suelen favorecer a la clase negativa (clase predominante), ya que las reglas que clasifican correctamente un mayor número de ejemplos son seleccionadas en el proceso de aprendizaje al aumentar la métrica considerada (que suele estar basada en el porcentaje de ejemplos bien clasificados). Por tanto, las instancias de la clase positiva (clase minoritaria) son mal clasificadas con una frecuencia mayor que las que pertenecen a la clase negativa [25]. Otra característica importante de este tipo de problemas son los “small disjuncts”, es decir, una concentración de datos de una única clase en un pequeño espacio del problema siendo rodeada por ejemplos de la clase contraria, [17, 18]; este tipo de regiones son difíciles de detectar para la mayoría de

algoritmos de aprendizaje. Además, otro de los principales problemas de los conjuntos de datos no balanceados es la mayor probabilidad de solapamiento entre ejemplos de la clase negativa y de la clase positiva [11].

Existen diferentes grados de no balanceo entre los datos. En este trabajo utilizamos el “*imbalance ratio*” (IR) [17] para distinguir las diferentes categorías. Se define como la razón entre el número de ejemplos de la clase negativa y el número de ejemplos de la clase positiva. Consideraremos que un conjunto de datos presenta un alto grado de no balanceo cuando su IR es mayor que 9 (menos de un 10 % de instancias de la clase positiva).

## 2.2. Aprendizaje sensible al coste

El aprendizaje sensible al coste tiene en cuenta los costes variables de una clasificación errónea de diversas clases [6, 28]. Una matriz de costes codifica la penalización de clasificar ejemplos de una clase como si fueran de otra. Sea  $C(i, j)$  el coste de predecir una instancia de la clase  $i$  como si fuera de la clase  $j$ . Con esta notación,  $C(+, -)$  es el coste de errar al clasificar una instancia de la clase positiva como si fuera negativa y  $C(-, +)$  es el coste del caso contrario. Al abordar problemas no balanceados suele ser más interesante reconocer las instancias positivas que reconocer las instancias negativas. Por lo tanto, el coste al equivocarse en clasificar una instancia de la clase positiva es mayor que el coste al equivocarse en clasificar una instancia de la clase negativa ( $C(+, -) > C(-, +)$ ).

El proceso de aprendizaje sensible al coste trata de minimizar el número de errores de coste elevado y el error total de clasificación errónea, teniendo en cuenta para ello la matriz de costes durante la construcción del modelo para generar uno con el mínimo coste. Usualmente, se trata de minimizar el coste utilizando el Teorema de Bayes del mínimo riesgo asignando a cada ejemplo la clase con menor riesgo. El aprendizaje sensible al coste supone que hay una matriz de costes disponible para los distintos tipos de errores, sin embargo, dado un conjunto de datos, la matriz de costes no suele estar disponible [19, 20].

## 2.3. Métricas de evaluación

La mayoría de propuestas para aprendizaje automático de clasificadores utilizan alguna medida de precisión del modelo como el porcentaje de ejemplos bien clasificados. Sin embargo, ese tipo de medidas pueden llevar a conclusiones erróneas cuando se trabaja con datos no balanceados ya que no tienen en cuenta la proporción de ejemplos de cada clase ni se introduce el coste de clasificación errónea de instancias de una clase en otra. Por ese motivo, en este trabajo usaremos la medida denominada “*área bajo la curva ROC*” (en inglés *Area Under the Curve* o AUC) [13], que se define como:

$$AUC = \frac{1 + VP_{tasa} - FP_{tasa}}{2} \quad (1)$$

donde  $VP_{tasa}$  es el porcentaje de ejemplos de la clase positiva bien clasificados y  $FP_{tasa}$  es el porcentaje de ejemplos de la clase negativa mal clasificados.

## 3. Sistemas de clasificación basados en reglas difusas sensibles al coste

En esta sección introducimos un SCBRDs sensible al coste que tiene en cuenta los costes variables de una clasificación errónea durante la etapa de entrenamiento.

### 3.1. Sistemas de clasificación basados en reglas difusas

Un SCBRD tiene dos componentes principales: el Sistema de Inferencia y la Base de Conocimiento. En un SCBRD lingüístico, la Base de Conocimiento está compuesta de la Base de Reglas (BR), constituida por el conjunto de reglas difusas y la Base de Datos, que contiene las funciones de pertenencia de las particiones difusas asociadas a las variables de entrada. Si no existe información experta sobre el problema a resolver, es necesario utilizar algún proceso de aprendizaje automático de la Base de Conocimiento a partir de ejemplos.

Cualquier problema de clasificación está compuesto por  $m$  muestras de entrenamien-

to  $x_p = (x_{p1}, \dots, x_{pn})$ ,  $p = 1, 2, \dots, m$  de  $M$  clases donde  $x_{pi}$  es el valor del atributo  $i$  ( $i = 1, 2, \dots, n$ ) de la  $p$ -ésima muestra de entrenamiento. En este trabajo utilizamos reglas difusas de la siguiente manera para construir nuestro SCBRDs:

$$\begin{aligned} &\text{Rule } R_j : \\ &\text{If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ then} \\ &\text{Class} = C_j \text{ with } RW_j \end{aligned} \quad (2)$$

donde  $R_j$  es la etiqueta de la  $j$ -ésima regla,  $x = (x_1, \dots, x_n)$  es un vector de muestras  $n$ -dimensional (variables de entrada),  $A_{ji}$  es la  $i$ -ésima etiqueta del antecedente,  $C_j$  es la etiqueta de una clase, y  $RW_j$  es el peso de la regla [14]. Utilizamos funciones de pertenencia triangulares como particiones difusas asociadas a las variables de entrada. Para calcular el peso de la regla se han propuestos diversas alternativas, aunque se ha considerado como una buena elección el uso del método heurístico conocido como el Factor de Certeza Penalizado (FCP) [16]:

$$\text{FCP}_j = \frac{\sum_{x_p \in C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^m \mu_{A_j}(x_p)} \quad (3)$$

donde  $x_p$  es el ejemplo  $p$ -ésimo del conjunto de entrenamiento,  $C_j$  es la clase consecuente de la regla  $j$  y  $\mu_{A_j}(x_p)$  es el grado de emparejamiento del ejemplo con los antecedentes de la regla.

Los métodos de aprendizaje difusos son la base para construir un SCBRDs. El algoritmo utilizado en este trabajo es el método propuesto en [4], una extensión del conocido algoritmo de Wang y Mendel [24] para clasificación, al que hemos llamado el método de generación de reglas de Chi. Para generar la BR difusa este SCBRDs determina la relación entre las variables del problema y establece una asociación entre el espacio de características y el espacio de clases por medio de los siguientes pasos:

1. *Creación de las particiones lingüísticas.*  
Una vez que el dominio de variación de cada característica  $A_i$  se ha determinado, se calculan las particiones difusas.

2. *Generación de una regla difusa para cada ejemplo  $x_p = (x_{p1}, \dots, x_{pn}, C_p)$ .* Para hacer esto es necesario:
  - 2.1 Calcular el grado de emparejamiento  $\mu(x_p)$  del ejemplo a las diferentes regiones difusas utilizando un operador de conjunción.
  - 2.2 Adscribir cada ejemplo  $x_p$  a la región difusa con el mayor grado de pertenencia.
  - 2.3 Generar una regla para el ejemplo cuyo antecedente se determina con la región difusa seleccionada y cuyo consecuente es la etiqueta de la clase del ejemplo.
  - 2.4 Calcular el peso de la regla.

Se pueden generar varias reglas con el mismo antecedente durante el aprendizaje. Si tienen la misma clase en el consecuente se elimina una de las reglas duplicadas, no obstante, si la clase es diferente sólo la regla de mayor de peso se mantiene en la BR.

### 3.2. Aprendizaje sensible al coste

Notaremos nuestra propuesta como Chi-SC (Chi Sensible al Coste). El propósito principal de Chi-SC es la construcción de un SCBRDs que sea capaz de tener en cuenta durante el proceso de construcción de la BR los distintos costes asociados a una clasificación errónea de alguna de las muestras. Para ello se utiliza una solución a nivel algorítmico modificando el propio comportamiento del algoritmo en alguna de sus etapas.

Chi-SC mantiene el comportamiento del algoritmo Chi transformando el modo de calcular el peso de la regla. Hemos adaptado la heurística FCP construyendo lo que hemos denominado el Factor de Certeza Penalizado Sensible al Coste (FCP-SC) que ha sido lo que se ha utilizado para determinar el peso de la regla:

$$\text{FCP-SC}_j = \frac{\sum_{x_p \in C_j} \mu_{A_j}(x_p) \times \text{CS}_j}{\sum_{p=1}^m \mu_{A_j}(x_p) \times \text{CS}_j} - \frac{\sum_{x_p \notin C_j} \mu_{A_j}(x_p) \times \text{CS}_j}{\sum_{p=1}^m \mu_{A_j}(x_p) \times \text{CS}_j} \quad (4)$$

En este caso,  $\text{CS}_j$  es el coste de clasificar erróneamente un ejemplo de la clase  $j$ .

#### 4. Estudio experimental

Vamos a analizar el comportamiento de Chi-SC en una amplia selección de conjuntos de datos no balanceados con una alta tasa de no balanceo ( $IR > 9$ ). De manera más específica, hemos utilizado 22 conjuntos de datos del repositorio UCI [1] con diferente IR, como se muestra en el Cuadro 1, donde se indican el número de ejemplos (#Ej.), número de variables (#Var.), nombre de cada clase (positiva y negativa), la distribución de ejemplos en cada clase y la tasa de no balanceo (IR). El cuadro se muestra en orden ascendente de IR. Los problemas multiclase se han transformado para obtener problemas no balanceados con sólo dos clases, uniendo una o más clases dentro de la clase negativa o positiva. Las clases que se funden en los conjuntos de datos utilizados al modificarlos se pueden consultar en la columna "Clase (+, -)" del Cuadro 1. Es usual encontrar este tipo de construcción de datos sobre problemas del repositorio UCI al trabajar con datos no balanceados [8], [9].

Con objeto de analizar la bondad del modelo, compararemos los resultados obtenidos por Chi-SC con los obtenidos por el método de Chi básico, así como con los obtenidos por el método de Chi básico empleando el método de preprocesamiento SMOTE [3] para reducir los efectos negativos de usar datos altamente no balanceados, considerando solamente el vecino más cercano para generar nuevos ejemplos, hasta balancear ambas clases hacia una distribución uniforme (50%).

Las características de configuración de los algoritmos de aprendizaje automático de SCBRDs se presentan a continuación. Esta elección está justificada por resultados previos

Cuadro 1: Descripción resumida de los conjuntos de datos no balanceados

Conj. de datos	#Ej.	#Var.	Clase (+, -)	%Clase (+, -)	IR
<i>Conjuntos de datos altamente no balanceados (IR mayor que 9)</i>					
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2; mit, me3, exc, vac, er1)	(9.66, 90.34)	9.35
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
Class016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc, headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
Ecol4	336	7	(om; remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(vac; nuc)	(6.72, 93.28)	13.87
Shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)	13.87
Glass4	214	9	(containers; remainder)	(6.07, 93.93)	15.47
Page-blocks13vs2	472	10	(graphic; horiz.line.picture)	(5.93, 94.07)	15.85
Abalone9vs18	731	8	(18; 9)	(5.65, 94.25)	16.68
Class016vs5	184	9	(tableware; build-win-float-proc, build-win-non_float-proc, headlamps)	(4.89, 95.11)	19.44
Shuttle2vs4	129	9	(Fw; Open; Bypass)	(4.65, 95.35)	20.5
Yeast1458vs7	693	8	(vac; nuc, me2, me3, pos)	(4.33, 95.67)	22.10
Class5	214	9	(tableware; remainder)	(4.20, 95.80)	22.81
Yeast2vs8	482	8	(pos; cyt)	(4.15, 95.85)	23.10
Yeast4	1481	8	(me2; remainder)	(3.43, 96.57)	28.41
Yeast1289vs7	947	8	(vac; nuc, cyt, pos, er1)	(3.17, 96.83)	30.56
Yeast5	1481	8	(me1; remainder)	(2.96, 97.04)	32.78
Ecol0137vs26	281	7	(pp, imL; cp, im, imU, imS)	(2.49, 97.51)	39.15
Yeast6	1481	8	(exc; remainder)	(2.49, 97.51)	39.15
Abalone19	4174	8	(19; remainder)	(0.77, 99.23)	128.87

obtenidos con el método de Chi para clasificación con datos no balanceados [8, 9]: 3 etiquetas difusas, T-norma producto como operador de conjunción junto con la heurística de FCP [16] para el peso de las reglas (FCP-SC para Chi-SC) y método de razonamiento difuso de la regla ganadora.

Además, como parámetro adicional para Chi-SC hay que identificar los costes asociados a la clasificación errónea de ejemplos de entrenamiento: si se clasifica un ejemplo de la clase positiva como si fuera de la clase negativa, el coste asociado es el IR del conjunto de datos utilizado mientras que si se clasifica un ejemplo de la clase negativa como si fuera de la clase positiva el coste asociado es 1. El coste de realizar una clasificación acertada se considera 0 ya que en este caso clasificar correctamente no debe penalizar el modelo construido.

Para garantizar unos resultados no dependientes de la partición que se realice del conjunto de ejemplos, utilizamos un modelo de validación cruzada en la que se divide cada conjunto de ejemplos en cinco particiones de

igual tamaño. Se realizan entonces cinco experimentos para cada conjunto de datos reservando una de las particiones como conjunto de datos de test y la combinación de las otras cuatro restantes como datos de entrenamiento. Este modelo con cinco particiones es habitual en clasificación de datos no balanceados frente a otros modelos con más particiones ya que en esos casos quedarían muy pocos ejemplos de la clase positiva para realizar el entrenamiento. Para cada conjunto de datos consideramos los resultados medios del esquema de ejecución mencionado en las cinco particiones para el conjunto de entrenamiento y el conjunto de test.

Cuadro 2: Resultados detallados para los métodos analizados

Conjunto de datos	Chi		Chi + SMOTE		Chi-SC	
	$AUC_{entr}$	$AUC_{test}$	$AUC_{entr}$	$AUC_{test}$	$AUC_{entr}$	$AUC_{test}$
Yeast2vs4	66,43	66,49	89,68	<b>87,36</b>	87,63	86,19
Yeast0567vs4	53,62	51,58	82,65	79,17	81,25	<b>81,13</b>
Vowel0	93,75	92,78	98,57	98,39	98,58	<b>98,44</b>
Glass016vs2	50,00	48,57	62,71	54,17	62,79	<b>55,02</b>
Glass2	50,00	48,22	66,54	55,30	67,20	<b>56,45</b>
Ecoli4	75,00	72,34	94,06	91,51	94,89	<b>92,54</b>
Shuttle0vs4	100,0	99,12	100,0	99,12	100,0	<b>99,56</b>
Yeast1vs7	53,75	49,77	82,00	80,63	82,61	<b>81,88</b>
Glass4	66,45	55,67	95,27	85,70	96,21	<b>89,08</b>
Page-Blocks13vs4	73,22	71,22	93,68	92,05	94,49	<b>95,32</b>
Abalone9-18	50,00	49,99	71,44	63,94	70,27	<b>72,98</b>
Glass016vs5	56,89	48,57	90,57	79,71	95,86	<b>87,64</b>
Shuttle2vs4	92,00	93,38	95,00	90,78	100,00	<b>94,19</b>
Yeast1458vs7	50,00	50,00	71,25	<b>64,65</b>	71,41	63,71
Glass5	55,54	49,27	94,33	83,17	94,39	<b>85,55</b>
Yeast2vs8	77,42	74,78	78,61	77,28	77,39	<b>77,34</b>
Yeast4	50,00	50,00	83,58	<b>83,15</b>	83,17	82,07
Yeast1289vs7	50,00	50,00	74,70	<b>77,12</b>	65,87	65,04
Yeast5	53,98	52,33	94,68	<b>93,58</b>	92,98	92,83
Yeast6	50,00	50,00	88,48	<b>88,09</b>	86,95	86,84
Ecoli0137vs26	85,67	<b>84,45</b>	93,96	81,90	91,80	82,99
Abalone9	52,38	50,00	70,23	64,70	71,45	<b>65,75</b>
Media	63,91	61,75	85,09	80,52	84,87	<b>81,48</b>

El Cuadro 2 muestra los resultados obtenidos en precisión de los métodos (usando la medida AUC) para los algoritmos usados en la comparación (Chi básico y Chi preprocesado con SMOTE) así como para Chi-SC, siendo  $AUC_{entr}$ , el AUC sobre el conjunto de datos de entrenamiento y  $AUC_{test}$  el AUC sobre el conjunto de datos de test.

Como puede observarse, Chi-SC obtiene mejores resultados (en promedio) que el método de Chi básico, tanto en  $AUC_{entr}$  como  $AUC_{test}$ , demostrando la influencia positiva del aprendizaje sensible al coste en el comportamiento del clasificador. Además, Chi-SC

presenta mejores resultados en  $AUC_{test}$  que el método de Chi + SMOTE.

En el Cuadro 3 se presentan los resultados de un test estadístico no paramétrico por parejas (test de Wilcoxon) sobre los datos medios de test en cada conjunto de datos para dar soporte estadístico a los resultados anteriormente comentados. Con este test estadístico vemos que por un lado se muestran diferencias significativas entre Chi y Chi-SC a favor de nuestro modelo. Por otro lado, se observa el mejor comportamiento de Chi-SC frente a Chi + SMOTE representado estadísticamente con un mayor valor de rango.

Cuadro 3: Test de Wilcoxon para comparar Chi-SC con el método de Chi básico y el método de Chi con preprocesamiento SMOTE de acuerdo a su  $AUC_{test}$ .  $R^+$  corresponde a la suma de rangos de Chi-SC y  $R^-$  a la suma de rangos de Chi o Chi+SMOTE

Comparación	$R^+$	$R^-$	p-valor
Chi-SC vs Chi	243.0	10.0	0.000
Chi-SC vs Chi+SMOTE	169.0	84.0	0.168

Este comportamiento muestra que el aprendizaje sensible al coste es un modelo interesante ya que en términos de acierto es suficientemente competitivo respecto a soluciones a nivel de datos con preprocesamiento mientras que reduce el tiempo necesario para resolver estos problemas: no se hace necesaria la etapa previa de preprocesamiento en el aprendizaje sensible al coste y además la etapa de entrenamiento es más sencilla ya que el conjunto de datos con el que hay que tratar es de un menor tamaño.

## 5. Conclusiones finales

En este trabajo hemos estudiado el comportamiento de un método de aprendizaje difuso sensible al coste Chi-SC para sistemas de clasificación basados en reglas difusas en un escenario de conjuntos de datos con un no balanceo alto. Con ello hemos utilizado una aproximación diferente para resolver problemas no balanceados y hemos comprobado su efectividad.

Los resultados experimentales y estadísticos nos permiten afirmar que el aprendizaje sensible al coste es una alternativa viable para resolver problemas no balanceados modificando el comportamiento del propio algoritmo base ya que mejora sensiblemente dicha versión. Además, el aprendizaje sensible al coste es una propuesta suficientemente competitiva para trabajar en problemas con niveles altos de no balanceo comparada con soluciones al nivel de datos basadas en el preprocesamiento del conjunto de datos de entrada. El aprendizaje sensible al coste obtiene un buen equilibrio en la precisión de ambas clases al mejorar el acierto en la clase positiva (sensibilidad) sin perjudicar el acierto en la clase negativa (especificidad).

## Referencias

- [1] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007. University of California, Irvine, School of Information and Computer Sciences. URL <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [2] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard. A study of the behaviour of several methods for balancing machine learning training data. *SIGKDD Explorations*, 6(1):20–29, 2004.
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligent Research*, 16:321–357, 2002.
- [4] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific, 1996.
- [5] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
- [6] P. Domingos. Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD'99)*, pages 155–164, 1999.
- [7] C. Elkan. The foundations of cost-sensitive learning. In *Proceedings of the 17th IEEE International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 973–978, 2001.
- [8] A. Fernández, M. J. del Jesús, and F. Herrera. On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets. *Information Sciences*, 180(8):1268–1291, 2010.
- [9] A. Fernández, S. García, M. J. del Jesus, and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.
- [10] S. García and F. Herrera. An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2607–2624, 2008.
- [11] V. García, R. Mollineda, and J. S. Sánchez. On the k-NN performance in a challenging scenario of imbalance and overlapping. *Pattern Analysis Applications*, 11(3-4):269–280, 2008.
- [12] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.
- [13] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):299–310, 2005.
- [14] H. Ishibuchi and T. Nakashima. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 9(4):506–515, 2001.

- [15] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.
- [16] H. Ishibuchi and T. Yamamoto. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 13:428–435, 2005.
- [17] A. Orriols-Puig and E. Bernadó-Mansilla. Evolutionary rule-based systems for imbalanced datasets. *Soft Computing*, 13(3):213–225, 2009.
- [18] A. Orriols-Puig, E. Bernadó-Mansilla, D. E. Goldberg, K. Sastry, and P. L. Lanzi. Facetwise analysis of XCS for problems with class imbalances. *IEEE Transactions on Evolutionary Computation*, 13:260–283, 2009.
- [19] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12):3358–3378, 2007.
- [20] Y. Sun, A. K. C. Wong, and M. S. Kamel. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(4):687–719, 2009.
- [21] F. Tek, A. Dempster, and I. Kale. Parasite detection and identification for automated thin blood film malaria diagnosis. *Computer Vision and Image Understanding*, 114(1):21–32, 2010.
- [22] K. M. Ting. An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering*, 14(3):659–665, 2002.
- [23] F. Vilariño, P. Spyridonos, F. Deiorio, J. Vitria, F. Azpiroz, and P. Radeva. Intestinal motility assessment with video capsule endoscopy: Automatic annotation of phasic intestinal contractions. *IEEE Transactions on Medical Imaging*, 29(2):246–259, 2010.
- [24] L. Wang and J. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992.
- [25] G. M. Weiss. Mining with rarity: a unifying framework. *SIGKDD Explorations*, 6(1):7–19, 2004.
- [26] D. Williams, V. Myers, and M. Silvious. Mine classification with imbalanced data. *IEEE Geoscience and Remote Sensing Letters*, 6(3):528–532, 2009.
- [27] B. Zadrozny and C. Elkan. Learning and making decisions when costs and probabilities are both unknown. In *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01)*, pages 204–213, 2001.
- [28] B. Zadrozny, J. Langford, and N. Abe. Cost-sensitive learning by cost-proportionate example weighting. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*, pages 435–442, 2003.
- [29] Z.-H. Zhou and X.-Y. Liu. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):63–77, 2006.