

Mejorando los Sistemas Difusos Evolutivos para Problemas Multi-Clase: Ponderando la Competencia en el Modelo Uno-contra-Uno con Truncado de las Confianzas

Alberto Fernández¹, Mikel Elkano², Mikel Galar², José Antonio Sanz²,
Humberto Bustince², and Francisco Herrera³

¹ Departamento de Informática, Universidad de Jaén, Jaén, España
alberto.fernandez@ujaen.es

² Departamento de Automática y Computación, ISC (Institute of Smart Cities),
Universidad Pública de Navarra, Pamplona, España

{mikel.elkano,mikel.galar,joseantonio.sanz,bustince}@unavarra.es,
³ Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad
de Granada, Granada, España
herrera@decsai.ugr.es

Resumen Las ventajas de los sistemas de multi-clasificación basados en estrategias de descomposición, y especialmente el esquema Uno-contra-Uno, han destacado por su buen comportamiento incluso para aquellos algoritmos que pueden abordar múltiples clases. Sin embargo, hay un problema inherente a la estrategia Uno-contra-Uno relacionado con el proceso de decisión: los clasificadores no competentes. Este caso se muestra cuando un clasificador binario emite un voto positivo para un par de clases que no están relacionadas con el ejemplo de entrada. Este hecho aporta “ruido” en la matriz de puntuación, degradando por tanto la precisión final. Por esta razón, se han desarrollado varios enfoques para hacer frente a la influencia de esta “no competencia”. Entre ellos, la estrategia de combinación basada en la distancia se ha destacado como una solución muy robusta. En esta contribución, proponemos una extensión sobre el modelo anterior truncando los grados de confianza antes de realizar el ajuste basado en distancia. Nuestro objetivo es mejorar el comportamiento de los Sistemas Difusos Evolutivos usados como clasificadores base, cuyas propiedades difusas pueden implicar en mayor grado el problema de la no competencia. Finalmente, mostraremos que la aplicación de esta nueva metodología permite una mejora significativa de los resultados obtenidos en contraste con los modelos base.

1. Introducción

Los problemas con múltiples clases están considerados como más difíciles que los binarios de acuerdo a la complejidad inherente tanto al mayor número de fronteras de decisión, como a la superposición entre las clases [17]. Algunos clasificadores sólo pueden hacer frente a problemas binarios, e incluso si el

algoritmo de aprendizaje es capaz de abordar problemas multi-clase, se ha demostrado que el uso de estrategias de descomposición puede llevar a una mejora significativa del rendimiento [8].

Uno de los métodos de descomposición más comunes es el llamado Uno-contra-Uno (en inglés “One-vs-One” u OVO) [12]. Este enfoque divide el problema original en tantos pares de clases como sea posible, haciendo caso omiso de los ejemplos que no pertenecen a las clases relacionadas. La simplificación de la tarea de aprendizaje implica que debe centrarse una mayor atención en la etapa de combinación. En ésta, los clasificadores que aprendieron a resolver cada problema binario deben agregarse a partir de una matriz de puntuación, para tomar una decisión final sobre la etiqueta de la clase [15].

En concreto, este proceso tiene un problema conocido como “clasificadores no competentes” [13]. Dicha situación se produce porque el ejemplo de entrada se consulta a todos los clasificadores binarios, incluso si no han sido entrenados para reconocer la clase real de esta instancia. Por lo tanto, estos clasificadores presentarán un grado de confianza no relevante, lo que podría alterar el proceso de decisión y degradar la precisión del clasificador.

Se observa que la competencia de un clasificador en el enfoque OVO no se puede establecer a priori. Por lo tanto, las técnicas propuestas actúan como una metodología de post-procesamiento dinámica con el fin de adaptar los valores de confianza de la matriz de puntuación antes de la etapa de decisión. En concreto, los nuevos valores se establecen de acuerdo a un cierto grado de importancia calculado para cada clasificador con respecto al ejemplo de entrada. Un enfoque desarrollado en [9], conocido como Ponderación de Competencia Basada en Distancia (PCBD), modifica los grados de confianza de los clasificadores ponderándolos de acuerdo con la distancia calculada a partir de los vecinos más cercanos de cada clase al ejemplo.

Entre los diferentes algoritmos de clasificación, los Sistemas de Clasificación basados en Reglas Difusas (SCBRDs) [14] han mostrado un buen comportamiento al modelar problemas complejos debido a la correcta gestión de la incertidumbre lograda por conjuntos difusos. Además, los SCBRDs pueden mejorar aún más hacia sistemas más precisos mediante la inclusión de las capacidades de aprendizaje y adaptación de la optimización evolutiva, lo que lleva a los llamados Sistemas Difusos Evolutivos (SDEs) [7].

El uso de clasificadores difusos puede influir negativamente en la “no competencia”. Ello es debido a la gran cobertura y generalización de las reglas con etiquetas lingüísticas. Nuestra hipótesis es que los grados de confianza calculados por los algoritmos de clasificación SDE dentro del enfoque OVO pueden tener un mayor beneficio en cooperación con el esquema PCBD. De esta manera, el objetivo de este trabajo es doble: (1) analizar experimentalmente si la combinación entre PCBD y los SDEs permite obtener una mejora significativa en el rendimiento del clasificador; y (2) proponer una modificación del modelo estándar PCBD truncando los valores de confianza en la matriz de puntuación a 0.0 y 1.0. La hipótesis para este último punto es buscar un comportamiento más

estable de acuerdo al observado por el esquema de agregación de Voto Binario en OVO para SCBRDs [5].

Para ello, llevamos a cabo un estudio experimental con 23 conjuntos de datos de clasificación seleccionados desde el repositorio KEEL [2]. Finalmente, este estudio experimental se consolidará mediante el uso de pruebas estadísticas, tal como se sugiere en la literatura especializada [3,11,10].

El esquema de esta contribución es el siguiente. En primer lugar, la Sección 2, presenta los preliminares de este trabajo mediante la introducción del enfoque de aprendizaje OVO y la metodología PCBD que aborda el problema de los clasificadores no competentes. A continuación, la Sección 3 describe nuestra extensión al modelo anterior que hemos denominado como PCBD-VB. En la Sección 4, se lleva a cabo el análisis experimental, en el que se destacan las principales conclusiones de esta contribución. Finalmente, la Sección 5 resume este estudio y concluye el documento.

2. Estrategia Uno-contra-Uno y Tratamiento del Problema de la No-Competencia

En esta sección primero recordaremos las propiedades básicas del esquema de aprendizaje OVO (Subsección 2.1). Posteriormente, describiremos la metodología PCBD para resolver el problema de la no competencia (Subsección 2.2).

2.1. Estrategias de descomposición: Uno-contra-Uno

El uso de estrategias de descomposición en multi-clasificación ha demostrado ser de gran interés en la comunidad investigadora [8,17], incluyendo los SCBRDs [4,5,6]. La idea principal de este esquema de aprendizaje es afrontar un problema de múltiples clases por medio de clasificadores binarios, a través de un paradigma tipo divide y vencerás.

En este trabajo nos centramos en la metodología de descomposición tipo OVO [12], consistente en el aprendizaje de clasificadores binarios encargados de discernir entre toda pareja de clases. En concreto, OVO divide un problema de m -clases en $m(m - 1)/2$ subproblemas binarios independientes, resultado de enfrentar todas las clases entre sí, donde cada uno ellos se aprende por un único clasificador. En la fase de clasificación, la instancia de entrada se presenta a todos los clasificadores, por lo que cada uno de ellos emite un grado de confianza r_{ij} y $r_{ji} \in [0, 1]$ a favor de su par de clases C_i y C_j (normalmente $r_{ji} = 1 - r_{ij}$). Estos grados de confianza se unifican dentro de una matriz de puntuación:

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix} \quad (1)$$

Se han desarrollado diferentes agregaciones con el fin de calcular la clase de salida [8]. La agregación más simple aunque potente, es la estrategia de voto,

donde cada clasificador vota para su clase predicha y la clase predicha será la que obtenga el mayor número de votos:

$$Clase = \arg \max_{i=1,\dots,m} \sum_{1 \leq j \neq i \leq m} s_{ij} \quad (2)$$

Cuando s_{ij} equivale a r_{ij} hablamos de *Voto Ponderado*, mientras que el enfoque de *Voto Binario* se aplica cuando valoramos $s_{ij} = 1$ si $r_{ij} > r_{ji}$ y 0 en otro caso.

2.2. PCBD: Ponderación de Competencia Basada en Distancia

En un esquema de aprendizaje OVO, un clasificador se dice que es “no competente” para clasificar una instancia siempre que la clase real de este ejemplo no se corresponda con ninguna de la pareja de clases aprendidas por el clasificador binario [13]. Aunque este problema implica “ruido” en la la matriz de puntuación para la decisión, no implica necesariamente una disminución en el rendimiento si todos los clasificadores competentes alcanzan un alto grado de confianza. Sin embargo, cuando los clasificadores competentes fallan, la decisión final está influenciada por los no competentes, lo que podría conducir a una clasificación errónea de la instancia.

Con el fin de aliviar el efecto negativo de los clasificadores no competentes, se desarrolló recientemente el enfoque PCBD [9]. Esta metodología consiste en llevar a cabo una adaptación dinámica de la matriz de puntuación, alterando los grados de confianza mediante la asignación de un peso superior a los clasificadores cuyas clases de salida estén más cerca de la instancia consultada, suponiendo que son más competentes que aquéllos que están más lejanos. Esta distancia se calcula utilizando el enfoque estándar de los k -vecinos más cercanos. De esta manera, una vez que se ha obtenido la matriz de puntuación, el procedimiento de PCBD es como sigue:

1. Calcular los k vecinos más cercanos de cada clase para la instancia dada, y almacenar las distancias promedio de los k vecinos de cada clase en un vector $\mathbf{d} = (d_1, \dots, d_m)$. El valor por defecto para k será 1.
2. Se crea una nueva matriz de puntuación R^w donde la salida r_{ij} de un clasificador que distingue entre las clases i, j se pondera como:

$$r_{ij}^w = r_{ij} \cdot w_{ij}, \quad (3)$$

donde w_{ij} es la competencia relativa del clasificador en la salida correspondiente, calculado como

$$w_{ij} = \frac{d_j^2}{d_i^2 + d_j^2} \quad (4)$$

siendo d_i la distancia de la instancia al vecino más cercano de la clase i .

- Usar la estrategia de Voto Ponderado sobre la matriz de puntuación modificada R^w para obtener la clase final:

$$Clase = \arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} r_{ij}^w \quad (5)$$

Este procedimiento completo será denominado como PCBD-WV. Con el fin de entender el comportamiento de PCBD-WV, un ejemplo ilustrativo para su procedimiento de trabajo se puede encontrar en [9].

3. PCBD-VB: Una Nueva Aproximación Dinámica basada en Puntuaciones Truncadas

Cuando se utilizan clasificadores difusos en un esquema OVO, las puntuaciones obtenidas por medio de los grados de confianza no son del todo adecuadas para la agregación basada en Voto Ponderado. En lugar de ello, muestran un mejor comportamiento con el “Voto binario” [5]. En concreto, debemos destacar dos puntos:

- Ambas clases de salida de los clasificadores binarios tienen la tendencia de alcanzar un grado de activación positivo, especialmente en áreas de alto solapamiento.
- Los clasificadores no-competentes obtendrán erróneamente una alta puntuación debido al nivel de confianza asociado a la gran cobertura y generalización de las reglas con etiquetas difusas.

Por los motivos anteriores, y con objeto de mejorar el comportamiento de los SDEs en el contexto del aprendizaje tipo OVO, proponemos una metodología en dos etapas, que hemos denominado como “PCBD-VB”:

- Truncar los valores de confianza a los valores $\{0.0, 1.0\}$ en la matriz de puntuación. El objetivo será limitar la influencia de los clasificadores no-competentes al menos en una de las dos clases asociadas.
- A continuación, utilizar la técnica PCBD para realizar la agregación a partir de la matriz de puntuación “truncada”. De este modo, se unifica la bondad mostrada por el modelo de Voto binario con el uso del proceso de decisión dinámico para reducir mayormente el problema de la no-competencia.

4. Estudio Experimental

En esta sección vamos a mostrar el estudio del comportamiento del enfoque PCBD-VB en SDEs. Para ello, primero introduciremos el clasificador base difuso seleccionado, así como los parámetros utilizados para la etapa de aprendizaje (Subsección 4.1). Posteriormente, vamos a listar los conjuntos de datos utilizados y la metodología de evaluación (Subsección 4.2). Por último, presentaremos los resultados experimentales y una breve discusión destacando los principales resultados obtenidos de este análisis (Subsección 4.3).

4.1. Clasificador base y parámetros

Para la realización de este estudio se ha seleccionado el algoritmo FARC-HD como un SDE robusto y representativo [1]. Concretamente, utilizaremos la extensión del mismo aplicando funciones de solapamiento durante el proceso de inferencia, tanto para la agregación de los antecedentes de la regla, como la combinación con el peso de la misma. Destacar que esta propuesta ha mostrado un mejor comportamiento para el esquema de aprendizaje OVO, tal como se sugiere en [5]. Los parámetros de configuración considerados, comunes para todos los problemas, se muestran en el Cuadro 1.

Cuadro 1. Especificación de parámetros para el algoritmo de aprendizaje FARC-HD.

<i>Parámetro</i>	<i>Valor</i>
Número de etiquetas	5 etiquetas
Operador de agregación	Media armónica
Inferencia	Combinación aditiva
Soprote mínimo	0.05
Confianza mínima	0.8
Profundad max. del árbol	3
k para pre-selección	2
Agregación OVO	VOTO BINARIO

4.2. Conjuntos de datos y evaluación de los clasificadores

Hemos utilizado 23 conjuntos del repositorio de datos KEEL [2], de tal modo que las mismas particiones puedan ser utilizadas por otros investigadores. Adicionalmente, en lugar de la validación cruzada estándar, y con motivo de corregir el “dataset shift” [16], (los datos de entrenamiento y test no siguen la misma distribución), utilizaremos un procedimiento de particionado recientemente publicado llamado Validación Cruzada con Distribución Equilibrada Óptima [18]. El Cuadro 2 resume las propiedades de estos conjuntos de datos. Finalmente, consideramos el uso del porcentaje de acierto para evaluar el rendimiento.

Con el fin de llevar a cabo la comparación de los clasificadores adecuadamente, consideramos tests no paramétricos [3,11]. En esta contribución, vamos a considerar un test Alineado de Friedman tanto para calcular el ranking de los algoritmos de acuerdo a su desempeño, así como los p -valores que determinan diferencias significativas entre los resultados. En ese caso, vamos a proceder con un procedimiento estadístico no paramétrico de Holm para comparaciones $1 \cdot n$, donde la obtención de los p -valores ajustados (APV) asociada a cada comparación representará el nivel más bajo para rechazar la hipótesis de igualdad. Además, con el fin de realizar comparaciones entre dos algoritmos, vamos a utilizar un test de Wilcoxon [19].

Cuadro 2. Descripción resumen de los conjuntos de datos.

Conjunto	id	#Ej.	#Atrs.	#Num.	#Nom.	#Cl.
Balance	bal	625	4	4	0	3
Contraceptive	con	1473	9	9	0	3
Hayes-roth	hay	132	4	4	0	3
Iris	iri	150	4	4	0	3
New-thyroid	new	215	5	5	0	3
Tae	tae	151	5	5	0	3
Thyroid	thy	7200	21	21	0	3
Wine	win	178	13	13	0	3
Vehicle	veh	846	18	18	0	4
Cleveland	cle	297	13	13	0	5
Page-blocks	pag	5472	10	10	0	5
Shuttle	shu	58000	9	9	0	5
Autos	aut	159	25	15	10	6
Glass	gla	214	9	9	0	7
Satimage	sat	6435	36	36	0	7
Segment	seg	2310	19	19	0	7
Ecoli	eco	336	7	7	0	8
Penbased	pen	10992	16	16	0	10
Yeast	yea	1484	8	8	0	10
texture	tex	5500	40	40	0	11
Vowel	vow	990	13	13	0	11
wine-quality-red	wqr	1599	11	11	0	11
wine-quality-white	wqw	4898	11	11	0	11

4.3. Resultados experimentales y análisis

Los resultados de la experimentación se muestran en el Cuadro 3. A partir de estos valores, observamos en primer lugar que el enfoque de aprendizaje por pares mejora el comportamiento del clasificador difuso original en un alto grado. Asimismo, el uso del esquema dinámico original PCBD muestra un mejor comportamiento con un valor promedio por encima de los anteriores.

Sin embargo, el principal hallazgo extraído a partir de la tabla es la clara superioridad del modelo PCBD-VB propuesto. La calidad mostrada por esta metodología se resalta siendo la mejor propuesta en 11 de los 23 conjuntos de datos, especialmente cuando tenemos un mayor número de clases, lo que implica una importante robustez.

Para dar mayor soporte a las conclusiones extraídas en este estudio, los Cuadros 4 y 5 muestran los resultados de aplicar los tests estadísticos.

En el primer caso, comprobamos que nuestra propuesta PCBD-VB alcanza el mejor ranking y que se muestran diferencias significativas con respecto tanto al algoritmo base como a la metodología OVO. En el segundo análisis, al realizar la comparativa entre los dos modelos PCBD-WV y PCBD-VB, se observa claramente un mejor comportamiento del nuevo esquema, tanto a nivel de ranking como en número de “victorias”. Por estos motivos, podemos destacar la calidad de este esquema y su sinergia positiva con los SDEs.

5. Conclusiones

En esta contribución hemos realizado una nueva propuesta para abordar el problema de los clasificadores no competentes en la fase de decisión de la estrategia de OVO. El procedimiento, denominado PCBD-VB, consiste en truncar los grados de confianza de los clasificadores binarios, y posteriormente ponderar

Cuadro 3. Resultados experimentales. De izquierda a derecha mostramos el algoritmo FARC-HD con funciones de solapamiento (FARC-HD-Ov), el modelo de aprendizaje por parejas (OVO), el modelo de agregación dinámico original (PCBD-WV) y nuestra propuesta (PCBD-VB). El mejor resultado por problema se marca en negrita.

data	#cl	FARC-HD-Ov	OVO	PCBD-WV	PCBD-VB				
bal	3	92.12	88.81	92.07	86.57	89.17	86.88	92.05	87.68
con	3	62.53	53.15	66.36	54.37	56.02	52.48	64.33	53.69
hay	3	91.60	77.97	90.56	76.20	85.83	79.27	90.49	78.52
iri	3	99.41	96.00	99.41	95.33	96.00	95.33	99.41	95.33
new	3	99.54	96.26	99.95	96.69	97.16	97.66	99.95	97.16
tae	3	77.78	56.82	82.42	62.72	60.72	56.95	74.17	57.85
thy	3	93.13	93.06	93.18	93.10	93.35	93.33	93.56	93.53
win	3	99.75	92.15	99.94	93.84	99.06	96.20	99.94	96.07
veh	4	82.60	71.03	88.11	71.16	78.92	74.44	88.06	72.69
cle	5	87.17	59.62	93.94	57.14	77.21	55.27	88.03	54.64
pag	5	96.13	95.43	96.66	96.04	95.46	95.43	97.16	96.60
shu	5	99.58	99.58	99.69	99.68	99.21	99.21	99.88	99.87
aut	6	98.61	81.32	100.00	79.50	89.37	82.73	93.57	82.80
gla	7	84.16	70.68	88.37	71.99	78.02	73.35	85.98	70.95
sat	7	64.29	63.81	88.88	86.83	88.51	88.49	89.78	87.99
seg	7	94.43	92.73	97.88	95.67	97.03	96.45	98.21	96.84
eco	8	92.33	82.16	95.64	82.94	85.68	82.27	91.70	84.06
pen	10	97.29	96.42	99.53	97.86	99.19	99.08	99.66	99.18
yea	10	64.14	59.22	70.10	59.36	59.34	59.35	65.38	60.15
tex	11	90.66	89.71	98.43	95.36	98.41	98.24	99.12	98.11
vow	11	73.83	66.36	99.70	91.72	98.89	98.28	99.71	99.09
wqr	11	65.60	61.15	69.68	61.47	62.11	60.40	69.05	62.97
wqw	11	55.38	53.38	58.52	55.03	56.54	56.13	59.91	57.13
Media	—	85.31	78.12	89.96	80.89	84.40	81.62	88.66	81.87

Cuadro 4. Ranking promedio (Friedman alineado) y APVs (test de Holm) para todos los clasificadores. El método de control se indica con asteriscos. El símbolo * implica diferencias significativas al 95%; el símbolo + marca el grado de confianza al 90%.

Método	Ranking	APV (Holm)
Baseline	65.6087 (4)	.000056*
OVO	48.1957 (3)	.076782+
PCBD-WV	40.3043 (2)	.285314
PCBD-VB	31.8913 (1)	*****

Cuadro 5. Test de Wilcoxon para comparar PCBD-VB [R^+] y el modelo estándar PCBD-WV [R^-]. Los valores de Victorias/Empates/Perdidas se calculan con respecto al nuevo modelo PCBD-VB.

PCBD-VB (R^+)	PCBD-WV (R^-)	p-valor	V/E/P
185.0	91.0	.148539	15/0/8

dinámicamente estos valores de acuerdo a la distancia del ejemplo de entrada a cada clase de problema. En particular, hemos aplicado este nuevo diseño en sinergia con el uso de los SDEs como algoritmos base.

Los resultados experimentales mostraron la validez de este nuevo enfoque, ya que permitió alcanzar un salto de calidad significativo en contraste tanto con el clasificador base, el modelo OVO estándar, e incluso la metodología de ponderación dinámica estándar. Adicionalmente, se observó que se alcanzaba un mayor rendimiento para problemas con un mayor número de clases, lo cual implica una alta robustez.

A partir de este estudio se abre camino para trabajo futuro. Entre otros, debemos estudiar si esta sinergia positiva entre SDEs y PCBD-VB también se muestra para otros SDEs. Además, podría ser interesante investigar esquemas adicionales para ponderar la competencia, tales como modelos de “clustering” en lugar del enfoque de vecindad simple propuesto.

6. Agradecimientos

This work was supported by the Spanish Ministry of Science and Technology under projects TIN-2011-28488, TIN-2012-33856, TIN2013-40765-P; the Andalusian Research Plans P11-TIC-7765 and P10-TIC-6858; and both the University of Jaén and Caja Rural Provincial de Jaén under project UJA2014/06/15.

Referencias

1. Alcalá-Fdez, J., Alcalá, R., Herrera, F.: A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems* 19(5), 857–872 (2011)
2. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multi-Valued Logic and Soft Computing* 17(2-3), 255–287 (2011)
3. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7, 1–30 (2006)
4. Elhag, S., Fernandez, A., Bawakid, A., Alshomrani, S., Herrera, F.: On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems. *Expert Systems with Applications* 42(1), 193–202 (2015)
5. Elcano, M., Galar, M., Sanz, J., Fernandez, A., Barrenechea, E., Herrera, F., Bustince, H.: Enhancing multi-class classification in farc-hd fuzzy classifier: On the synergy between n-dimensional overlap functions and decomposition strategies. *IEEE Transactions on Fuzzy Systems* in press, 1–1 (2015)
6. Fernandez, A., Calderon, M., Barrenechea, E., Bustince, H., Herrera, F.: Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations. *Fuzzy Sets and Systems* 161(23), 3064–3080 (2010)

7. Fernandez, A., Lopez, V., del Jesus, M.J., Herrera, F.: Revisiting evolutionary fuzzy systems: Taxonomy, applications, new trends and challenges. *Knowledge-Based Systems* 80, 109–121 (2015)
8. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition* 44(8), 1761–1776 (2011)
9. Galar, M., Fernandez, A., Barrenechea, E., Herrera, F.: DRCW-OVO: Distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems. *Pattern Recognition* 48(1), 28–42 (2015)
10. García, S., Fernández, A., Luengo, J., Herrera, F.: Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences* 180(10), 2044–2064 (2010)
11. García, S., Herrera, F.: An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research* 9, 2607–2624 (2008)
12. Hastie, T., Tibshirani, R.: Classification by pairwise coupling. *The Annals of Statistics* 26(2), 451–471 (1998)
13. Hüllermeier, E., Vanderlooy, S.: Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting. *Pattern Recognition* 43(1), 128–142 (2010)
14. Ishibuchi, H., Nakashima, T., Nii, M.: Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining. Springer-Verlag (2004)
15. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 1st edn. (2004)
16. Lopez, V., Fernandez, 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)
17. Lorena, A.C., Carvalho, A.C., Gama, J.M.: A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review* 30(1–4), 19–37 (2008)
18. 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)
19. Sheskin, D.: *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC (2006)