



# CEDI 2010 VALENCIA

7 A 10 DE SEPTIEMBRE DE 2010

III CONGRESO ESPAÑOL DE INFORMÁTICA

UNIVERSIDAD POLITÉCNICA DE VALENCIA

Actas del V Simposio de Teoría y Aplicaciones  
de Minería de Datos  
(TAMIDA2010)

EDITORES

Alicia Troncoso, José C. Riquelme



# Actas del V Simposio de Teoría y Aplicaciones de Minería de Datos [TAMIDA 2010]

## Editores:

Alicia Troncoso  
José C. Riquelme

**Simposio organizado por**  
Red Española de Minería de Datos y Aprendizaje

## Entidades Colaboradoras



**Garceta**  
grupo editorial

**Actas del V Simposio de Teoría y Aplicaciones de Minería de Datos  
(TAMIDA 2010)**

**Editores:** Alicia Troncoso, José C. Riquelme

**ISBN:** 978-84-92812-60-8

**IBERGARCETA PUBLICACIONES, S.L., Madrid, 2010**

**Edición:** 1ª

**Impresión:** 1ª

**Nº de páginas:** 400

**Formato:** 17 x 24

**Materia CDU:** 004 Ciencia y tecnología de los ordenadores. Informática

Reservados los derechos para todos los países de lengua española. De conformidad con lo dispuesto en el artículo 270 y siguientes del código penal vigente, podrán ser castigados con penas de multa y privación de libertad quienes reprodujeran o plagiaran, en todo o en parte, una obra literaria, artística o científica fijada en cualquier tipo de soporte sin la preceptiva autorización. Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede ser reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste electrónico, químico, mecánico, el electro-óptico, grabación, fotocopia o cualquier otro, sin la previa autorización escrita por parte de la editorial.

Dirijase a CEDRO (Centro Español de Derechos Reprográficos), [www.cedro.org](http://www.cedro.org), si necesita fotocopiar o escanear algún fragmento de esta obra.

COPYRIGHT © 2010 IBERGARCETA PUBLICACIONES, S.L.  
[info@garceta.es](mailto:info@garceta.es)

**Actas del V Simposio de Teoría y Aplicaciones de Minería de Datos (TAMIDA 2010)**

Derechos reservados ©2010 respecto a la primera edición en español, por LOS AUTORES

Derechos reservados ©2010 respecto a la primera edición en español, por IBERGARCETA PUBLICACIONES, S.L.

1ª Edición, 1ª Impresión

ISBN: 978-84-92812-60-8

Depósito legal: M-35543-2010

**Maquetación:** Los Editores

**Coordinación del proyecto:** @LIBROTEX

**Portada:** Estudio Dixi

**Impresión y encuadernación:**

OI: 23/2010

PRINT HOUSE, S.A.

**IMPRESO EN ESPAÑA -PRINTED IN SPAIN**

*Nota sobre enlaces a páginas web ajenas:* Este libro puede incluir referencias a sitios web gestionados por terceros y ajenos a IBERGARCETA PUBLICACIONES, S.L., que se incluyen sólo con finalidad informativa. IBERGARCETA PUBLICACIONES, S.L., no asume ningún tipo de responsabilidad por los daños y perjuicios derivados del uso de los datos personales que pueda hacer un tercero encargado del mantenimiento de las páginas web ajenas a IBERGARCETA PUBLICACIONES, S.L., y del funcionamiento, accesibilidad y mantenimiento de los sitios web no gestionados por IBERGARCETA PUBLICACIONES, S.L., directamente. Las referencias se proporcionan en el estado en que se encuentran en el momento de publicación sin garantías expresas o implícitas, sobre la información que se proporcione en ellas.

## Contenido

### V Simposio de Teoría y Aplicaciones de Minería de Datos [TAMIDA 2010]

---

#### Sección I Metodologías

---

Group association rules: Interestingness measures and group ranking criteria.....	3
Aída Jiménez, Fernando Berzal, Juan-Carlos Cubero, Matías Gámez	
Multirelational data mining using frequent tree patterns.....	11
Aída Jiménez, Fernando Berzal, Juan-Carlos Cubero, Francisco Cortijo	
Efficient algorithms for XML data mining.....	19
Aída Jiménez, Fernando Berzal, Juan-Carlos Cubero	
Hacia una integración de Minería de Datos y Sistemas Multiagente: Una destacada y desafiante área de investigación y desarrollo.....	29
Guillermo Gilberto Molero Castillo, María Elena Meda Campaña	
Elección de la técnica de minería de datos: Mapa conceptual de técnicas.....	37
Karina Gibert, Miquel Sánchez-Marrè, Víctor Codina	
Comparative Analysis of Articulated and Behavioural Social Networks: A case study at Meneame.net.....	45
Andreas Kaltenbrunner, Gustavo González, Ricard Ruiz de Querol, Yana Volkovich	
Una técnica de minería de datos para la inducción de redes de restricciones temporales cuantitativas entre eventos.....	55
Miguel R. Álvarez, Paulo Félix, Purificación Cariñena	
Influencia del Ruido en Sistemas de Clasificación con Múltiples Clases: Análisis sobre la estrategia Uno-contra-Uno.....	65
Mikel Galar, José A. Sáez, Julián Luengo, Francisco Herrera	
Intersection kernel for Interval-valued Feature Selection.....	75
Albert Samà, Francisco J. Ruiz, Andreu Català, Cecilio Angulo	
Algunas Consideraciones sobre Aprendizaje de Métricas mediante Maximización del Margen.....	83
Adrián Pérez-Suay, Francesc J. Ferri	
Sobre el uso de algoritmos de búsqueda local en el espacio de órdenes en el Aprendizaje estructural de Redes Bayesianas.....	93
Juan I. Alonso-Barba, Luis de la Ossa, José M. Puerta Callejón	

# Influencia del Ruido en Sistemas de Clasificación con Múltiples Clases: Análisis sobre la estrategia Uno-contra-Uno

Mikel Galar

Dept. de Automática y Computación  
Universidad Pública de Navarra  
31006 Pamplona  
mikel.galar@unavarra.es

José A. Sáez, Julián Luengo, Francisco Herrera

Dept. de Ciencias de la Computación e Inteligencia Artificial  
Universidad de Granada  
18071 Granada  
smja@correo.ugr.es, {julianlm,herrera}@decsai.ugr.es

## Resumen

El ruido está presente en cualquier conjunto de datos real y su presencia afecta a la precisión, tiempo de construcción y complejidad de los modelos extraídos en el ámbito del aprendizaje automático. Estos problemas se pueden ver amplificados cuando el conjunto de datos contiene múltiples clases. Una forma habitual de afrontar los problemas multi-clase son las técnicas de binarización.

El objetivo de esta contribución es analizar el comportamiento de los sistemas de clasificación en el ámbito de múltiples clases ante la presencia de ruido en los conjuntos de datos. Por ello, se considera el análisis de la estrategia de binarización uno-contra-uno. Nuestra intención es estudiar la robustez dicha estrategia en presencia de ruido y contrastarla con la de los clasificadores base. Para ello empleamos tres clasificadores (C4.5, SVM y  $k$ NN) con buen comportamiento en un gran abanico de problemas reales. Perturbamos los conjuntos de datos con diferentes niveles de ruido y analizamos apoyados por test estadísticos no paramétricos la conveniencia de las técnicas de binarización.

## 1. Introducción

En el ámbito del aprendizaje automático supervisado, los sistemas de clasificación [6] son una de las técnicas más utilizadas por su capacidad para ser aplicadas en problemas reales. Muchos de estos problemas reales contienen

más de dos clases, se suelen denominar multi-clase [18]. Los problemas multi-clase son en general más difíciles de resolver que los binarios, dado que las fronteras son más complejas y el solapamiento entre clases suele ser mayor. Por esta razón surgieron las técnicas de binarización (una revisión muy completa puede verse en [11]) que tratan de reducir el problema multi-clase original en problemas de dos clases más sencillos de resolver. Así, cada problema binario es afrontado por un clasificador que trata de distinguir entre el par de clases consideradas. Estos clasificadores suelen denominarse *clasificadores base* [7]. Finalmente, la salida de cada uno de ellos ante una nueva instancia debe ser agregada o combinada para decidir cual es la clase a la que pertenece el ejemplo.

Por otro lado, la presencia de ruido en los conjuntos de datos afecta directamente a los modelos extraídos. Los datos reales recopilados nunca son perfectos y pueden estar sujetos a errores de medición (humanos o por falta de precisión), corrupciones etc. El concepto de ruido engloba a todas esas imperfecciones en los datos. Se distinguen principalmente dos tipos de ruido: de clase o de atributos, dependiendo de las características afectadas [21]. La presencia de ruido en problemas reales es habitual y su tratamiento es actualmente un problema abierto [20]. Además, este tipo de mecanismos no se pueden aplicar de manera general ya que pueden producir imperfecciones en los datos e incluso, pueden producir pérdidas importantes de información. Por ello, es

importante conocer la robustez de los sistemas de clasificación ante la presencia de ruido en los datos.

Los problemas multi-clase, al igual que los problemas binarios, están afectados generalmente por la presencia de ruido. Por esta razón, es interesante extender el análisis sobre la presencia de diferentes niveles de ruido en los sistemas de clasificación al ámbito de los problemas multi-clase. En esta contribución, analizaremos el comportamiento de las técnicas de binarización frente al de los algoritmos originales que soportan varias clases de manera nativa. Mostraremos la bondad de las primeras en el manejo de datos con ruido en cuanto a robustez y precisión.

Para llevar a cabo el análisis, realizaremos un estudio experimental con clasificadores base provenientes de diferentes paradigmas, tales como, máquinas vector soporte (SVM) [19], árboles de decisión [14] y aprendizaje basado en instancias [1]. Trataremos con 11 conjuntos de datos con múltiples clases obtenidos del repositorio UCI [3]. Perturbaremos estos conjuntos con diferentes niveles de ruido siguiendo las pautas presentadas en [21]. Apoyados por test estadísticos tal y como se sugiere en la literatura [5, 9, 8], analizaremos la robustez de la estrategia uno-contra-uno ante el ruido frente a la de los métodos base y métodos de ensemble [15].

El resto de la contribución está organizada como sigue. En la Sección 2 describimos la estrategia de binarización utilizada. Introducimos el concepto de ruido en los conjuntos de datos en la Sección 3. Presentamos el marco experimental utilizado en el análisis en la Sección 4. En la Sección 5 mostramos los resultados obtenidos en el estudio experimental y analizamos la presencia de ruido tanto de clases, como de atributos. Finalmente, en la Sección 6 señalamos las conclusiones obtenidas.

## 2. Problemas Multi-clase con Técnicas de Binarización

En la literatura existen diferentes formas de descomponer un problema multi-clase en pro-

blemas binarios a priori más sencillos de resolver [11]. Las estrategias más comúnmente utilizadas son las denominadas uno-contra-uno (“*One-vs-One*”, OVO) y uno-contra-todos (“*One-vs-All*”, OVA).

- OVO divide el problema en tantos problemas binarios como enfrentamientos entre clases podamos realizar. Cada clasificador aprende a discriminar entre un par de clases. Luego, las salidas de todos los clasificadores son combinadas para obtener la clase final.
- OVA aprende un clasificador para distinguir a cada clase frente al resto de las clases. El clasificador que muestre la mayor confianza en la predicción de la clase que intenta discriminar, señala la clase final.

En este trabajo hemos optado por la primera descomposición dadas las ventajas que ofrece respecto a la segunda [7]:

- Mayor precisión.
- Produce fronteras más simples.
- Menor tiempo de entrenamiento (aunque hay más clasificadores, son más sencillos).
- Tiene menor tendencia a crear conjuntos de datos no balanceados [17] que pueden resultar contraproducentes.

En la literatura existen diferentes métodos para combinar las salidas de cada clasificador. En este trabajo, consideramos la estrategia de voto simple [7], donde cada clasificador da un voto a la clase que predice y la clase con más votos es la elegida como salida. Los empates se decantan por la clase mayoritaria, en caso de continuar se resuelven de manera aleatoria.

## 3. Presencia de Ruido en los Conjuntos de Datos

El ruido es un problema que afecta a cualquier conjunto de datos real [21]. Cuando está presente en los conjuntos de entrenamiento afecta directamente a la precisión del clasificador construido, también, el tiempo de construcción, el tamaño y por consiguiente su comprensión pueden verse afectados. Por otro lado,

conjuntos de test perturbados provocan clasificaciones erróneas. Aunque existen métodos para el filtrado de ruido [4], es una línea de investigación abierta y con futuro principalmente en cuanto al ruido en atributos. En todo caso, estas técnicas deben aplicarse teniendo en cuenta que no se obtienen datos equivalentes a los originales [20]. Por este motivo resulta interesante conocer de antemano la robustez que ofrecen los métodos de clasificación frente a datos imprecisos, imperfectos o erróneos.

El ruido en clases se produce cuando la etiqueta de clase de alguna instancia no es la adecuada, puede ocurrir por ejemplos mal etiquetados o contradictorios (instancias que aparecen más de una vez con diferentes etiquetas de clase). El ruido en clases solo tiene sentido en el conjunto de entrenamiento. Existen diversas aproximaciones para reducirlo [4] y han demostrado que en muchos casos su eliminación mejora la precisión obtenida por los clasificadores construidos.

El manejo de ruido sobre atributos es más complejo [20]. Además, Quinlan en [13] concluyó que la precisión en test de los árboles de decisión tras la eliminación del ruido en atributos en el conjunto de entrenamiento era menor que la obtenida sobre el árbol imperfecto construido a partir de los datos con ruido. De esa manera indicaba que podía ser contraproducente su eliminación si los datos de test también iban a contener el mismo ruido en los atributos. En cualquier caso, el ruido en atributos suele estar mucho más presente en los conjuntos de datos. La clase se define con una única característica, mientras que los atributos suelen corresponderse con varias, por lo que es más fácil encontrar información incompleta, errónea o no disponible. La presencia de ruido en atributos debe estudiarse desde tres puntos de vista atendiendo al lugar donde está presente [21]:

- LR: Conjunto de entrenamiento libre de ruido y el de test perturbado.
- RL: Conjunto de entrenamiento con ruido pero las muestras de test libres de él.
- RR: Ruido presente en ambos conjuntos.

## 4. Marco Experimental

En esta sección, comenzamos mostrando en la Subsección 4.1 los conjuntos de datos reales escogidos para la experimentación. En la Subsección 4.2 damos los detalles correspondientes a la inserción de ruido tanto de clases como de atributos. Indicamos los clasificadores base utilizados para el estudio junto con los parámetros utilizados en su ejecución en la Subsección 4.3. En la Subsección 4.4 introducimos los test estadísticos aplicados para la comparación de resultados a lo largo del estudio experimental. Finalizamos la sección describiendo la metodología utilizada para el análisis de la robustez de los métodos en la Subsección 4.5.

### 4.1. Conjuntos de Datos

En el Cuadro 1 mostramos las propiedades de los conjuntos de datos seleccionados para la experimentación. Para cada conjunto presentamos el número de ejemplos (#Ej.), el número de atributos (#Atrs.) junto al número de cada tipo, numéricos (#Num.) o nominales (#Nom.), y el número de clases (#Cl.).

La estimación de la precisión (porcentaje de acierto) de cada clasificador la obtenemos por medio de una validación cruzada de 5 particiones. Dividimos el conjunto de datos en 5 partes, cada una con el 20% de las instancias del conjunto y manteniendo la proporción entre clases. Para cada partición, se aprende el clasificador con el resto de las particiones y se utiliza la primera para validar el modelo obtenido. En el análisis únicamente consideraremos la precisión en test.

Cuadro 1: Descripción de los conjuntos de datos.

Conjunto	#Ej.	#Atrs.	#Num.	#Nom.	#Cl.
Contraceptive	1473	9	9	0	3
Iris	150	4	4	0	3
Thyroid	7200	21	6	15	3
Wine	178	13	13	0	3
Page-blocks	5472	10	10	0	5
Glass	214	9	9	0	6
Satimage	6435	36	36	0	7
Segment	2310	19	19	0	7
Ecoli	336	7	7	0	8
Penbased	10992	16	16	0	10
Yeast	1484	8	8	0	10

#### 4.2. Perturbación de los Conjuntos de Datos con Ruido

La inserción de ruido en los conjuntos de datos la hemos realizado de acuerdo con el modelo presentado en [21] para el análisis de su influencia en los sistemas de clasificación. Así pues, introducimos ruido de la siguiente forma, dependiendo del tipo de ruido:

- Ruido en clases: Siendo  $X$  la clase mayoritaria e  $Y$  la segunda clase con mayor número de instancias, fijamos un nivel de ruido  $x$ , como la probabilidad de que una instancia de  $X$  sea incorrectamente etiquetada como perteneciente a la clase  $Y$ . Por tanto, el  $x \cdot 100\%$  de las instancias de  $X$  pasan a ser instancias de  $Y$  de manera aleatoria.
- Ruido en atributos: Insertamos un nivel de ruido  $x$  en cada atributo de manera aleatoria. Cada atributo de una instancia tiene una probabilidad  $x \cdot 100\%$  de ser modificado. Para los nominales intercambiamos la etiqueta por otra elegida de manera aleatoria, en el caso de los numéricos, modificamos el valor por uno aleatorio en el rango del atributo.

Tanto las probabilidades como los nuevos valores introducidos de manera aleatoria los obtenemos mediante una distribución uniforme. En este trabajo consideramos niveles de ruido  $x = 5\%$ ,  $10\%$  y  $20\%$ . Para realizar una estimación de la pérdida ( $P$ ) producida en un clasificador ante la presencia de ruido, utilizamos la siguiente medida:

$$P = \frac{Prec_{0\%} - Prec_{x\%}}{Prec_{0\%}} \quad (1)$$

donde  $Prec_{x\%}$  es la precisión media del clasificador sobre los conjunto de datos con un nivel de ruido  $x$ . Construiremos los clasificadores sobre los diferentes niveles sin considerar la utilización de técnicas de filtrado o limpieza de ruido.

#### 4.3. Clasificadores base y configuración de parámetros

Para el estudio empírico hemos seleccionado una serie de clasificadores de diferentes para-

digmas muy utilizados en el ámbito del aprendizaje automático:

- SVM [19] trasladan el espacio de entradas original en un espacio de características de mayores dimensiones. En el nuevo espacio de características, buscan el hiperplano de separación óptimo con un margen de separación máximo entre clases, mejorando la generalización.
- C4.5 [14] es un algoritmo de generación de árboles de decisión. El árbol se construye descendientemente utilizando la ganancia de información normalizada (diferencia en entropía) que se obtiene al elegir un atributo para dividir el árbol en el siguiente nodo. El atributo con la mayor ganancia de información es el elegido para dividir el árbol (tomar la decisión).
- $k$ NN [12] *k-Nearest Neighbours* o *k-Vecinos* más cercanos, busca las  $k$  instancias más próximas en el conjunto de entrenamiento respecto al nuevo patrón de test. Elige la clase predominante en ese vecindario. La distancia utilizada y el número de vecinos considerados resultan clave para este algoritmo.

Hemos realizado esta selección en base al buen comportamiento de estos algoritmos en un amplio número de problemas reales. Además, remarcar que en el caso de las SVM, originalmente no tiene soporte para problemas multi-clase, por lo que en ese caso, la utilización de técnicas de binarización es obligada.

Por último, además de los tres algoritmos mencionados, queremos estudiar el comportamiento y la robustez de ensembles [15] formados por estos tres clasificadores. Utilizaremos por tanto una combinación de clasificadores basada al igual que la estrategia OVO, en el método de voto simple. En caso de empate, consideraremos el patrón como mal clasificado.

Los algoritmos han sido ejecutados haciendo uso de la herramienta KEEL<sup>1</sup> [2] utilizando los parámetros mostrados en el Cuadro 2 siguiendo las configuraciones por defecto de la herra-

<sup>1</sup>www.keel.es



mienta, obtenidas por recomendación del autor para C4.5 y ajustadas experimentalmente para un buen comportamiento global en SVM.

Cuadro 2: Especificación de parámetros utilizada en la fase de aprendizaje.

Algoritmos	Parámetros
SVM	C = 1.0 Tolerance Parameter = 0.001 Epsilon = 1.0E-12 Kernel Type = Polynomial Polynomial degree = 1 Fit Logistic Models = True
C4.5	Prune = True Confidence level = 0.25 Minimum number of item-sets per leaf = 2
1NN	k = 1 Distance metric = HVDM
3NN	k = 3 Distance metric = HVDM
5NN	k = 5 Distance metric = HVDM

#### 4.4. Test Estadísticos

En esta contribución, utilizamos las técnicas de contraste de hipótesis para dar un soporte estadístico al análisis de los resultados [16]. Utilizaremos los test estadísticos no paramétricos, dado que las condiciones iniciales que garantizan la fiabilidad de los test paramétricos pueden no satisfacerse [5, 9]. Concretamente, utilizaremos el test de Iman-Davenport [16] para detectar diferencias significativas entre grupos de resultados y el test post-hoc de Holm [10] para encontrar qué algoritmos rechazan la hipótesis de igualdad con respecto al método de control seleccionado.

#### 4.5. Metodología de análisis de la robustez de los sistemas de clasificación

Para el análisis de la robustez hemos considerado una metodología guiada por los resultados de pérdida y precisión. Nos basamos en el uso de test estadísticos no paramétricos para la comparativa de resultados. Queremos analizar la robustez de cada método de aprendizaje respecto a su variante OVO y a su vez observar el comportamiento de la combinación de los tres algoritmos mediante técnicas de ensemble.

Por esta razón, realizamos el análisis de los resultados por grupos o conjuntos de métodos. Un grupo está formado por todos aquellos algoritmos que utilizan un mismo clasificador base. Incluimos en un grupo el algoritmo con y sin la utilización de OVO en sus cuatro ejecuciones sobre diferentes niveles de ruido. Esta división forma 8 grupos (SVM, C4.5, 1NN, 3NN y 5NN, y los tres ensembles), con 8 métodos cada uno (niveles de ruido de 0%, 5%, 10% y 20% para OVO y sin OVO), salvo para SVM que son 4 ya que no dispone de versión multi-clase. Así, queremos analizar estadísticamente tres aspectos:

1. El nivel de ruido a partir del cual encontramos diferencias estadísticas significativas en la precisión de los clasificadores respecto al mejor caso sin ruido.
2. Contrastar la robustez de la estrategia OVO frente a la de los algoritmos base.
3. Observar comportamiento de los ensembles ante el ruido.

En cada grupo, utilizamos el test de Iman-Davenport para identificar diferencias significativas entre los resultados. En caso de que existan, aplicamos el test post-hoc de Holm para conocer el nivel de ruido a partir del cual se encuentran estas diferencias. Compararemos la robustez de los métodos originales y la estrategia OVO en base a esos niveles.

## 5. Análisis del comportamiento de las técnicas OVO frente al ruido.

En esta sección analizamos la robustez de los algoritmos de aprendizaje frente al ruido. En él, apoyados por diversos test estadísticos, nuestra intención es estudiar el comportamiento de la estrategia de binarización OVO respecto a los algoritmos de aprendizaje base. Comenzamos estudiando el ruido en clases en la Subsección 5.1. Posteriormente, analizamos los tres casos de ruido en atributos. En la Subsección 5.2 analizamos la situación de ruido únicamente en test. Estudiamos el caso de ruido en el conjunto de entrenamiento en la Subsección 5.3. Finalizamos analizando en la

Subsección 5.4 el caso más general con ruido en ambos conjuntos.

### 5.1. Ruido en Clases

Iniciamos el análisis mostrando en el Cuadro 3 los resultados obtenidos para las ejecuciones de cada caso. En él mostramos la pérdida (Ecuación (1)) junto a la precisión (tasa de aciertos) obtenida por el clasificador de la fila, con el nivel de ruido indicado en la columna. Las columnas marcadas como (Med.) muestran la media del método en todos los niveles de ruido. Los métodos EN corresponden a los ensembles, indicando el número que lo sigue la cantidad de vecinos utilizados en  $k$ NN. Por último, cuando el clasificador es construido mediante la estrategia de binarización, al clasificador base le sigue la extensión *ovo*.

Cuadro 3: Pérdida y tasa de aciertos para niveles de ruido en clases de 5, 10 y 20 %.

Método	Pérdida				Precisión				
	5%	10%	20%	Med.	0%	5%	10%	20%	Med.
SVM	0.79	1.53	3.33	1.88	82.72	82.06	81.45	79.96	81.55
SVM <sub>ovo</sub>	0.78	1.18	2.78	1.58	83.44	82.79	82.45	81.12	82.45
C4.5	0.18	0.77	2.74	1.23	83.98	83.83	83.33	81.68	83.20
C4.5 <sub>ovo</sub>	2.13	3.30	7.65	4.36	82.82	81.05	80.08	76.49	80.11
1NN	2.03	3.45	7.81	4.43	83.08	81.40	80.22	76.60	80.32
1NN <sub>ovo</sub>	0.63	1.27	4.85	2.25	83.62	83.10	82.57	79.57	82.21
3NN	0.32	1.07	4.87	2.09	83.93	83.66	83.03	79.84	82.61
3NN <sub>ovo</sub>	0.14	0.60	3.26	1.33	83.72	83.61	83.22	80.99	82.88
5NN	0.17	0.73	3.21	1.37	84.18	84.04	83.57	81.48	83.32
5NN <sub>ovo</sub>	0.61	0.94	2.15	1.23	84.67	84.15	83.88	82.85	83.89
EN1	0.21	0.53	2.09	0.95	84.67	84.49	84.22	82.90	84.07
EN1 <sub>ovo</sub>	0.61	0.94	2.15	1.23	84.67	84.15	83.88	82.85	83.89
EN3	<b>-0.03</b>	<b>-0.09</b>	<b>1.64</b>	<b>0.51</b>	84.63	84.65	<b>84.71</b>	83.24	84.31
EN3 <sub>ovo</sub>	0.22	0.83	2.12	1.06	84.77	84.58	84.06	82.97	84.10
EN5	0.32	0.41	1.99	0.91	<b>85.00</b>	<b>84.73</b>	84.66	<b>83.31</b>	<b>84.42</b>
EN5 <sub>ovo</sub>									

Cabe destacar que el ruido en clases, salvo para un clasificador como 1NN que depende directamente de la clase a la que pertenece el ejemplo más cercano, el ruido en clases no produce grandes pérdidas a niveles bajos de ruido. Además, la utilización de OVO es beneficiosa, ya que en la gran mayoría de los casos la pérdida obtenida es menor y la precisión final aumenta respecto al caso sin OVO. Resaltar también la capacidad de los ensembles para soportar el ruido, obteniendo pérdidas muy reducidas y una precisión superior a los clasificadores base que los forman. Esta mejora queda todavía más patente en los ensemble formados por los OVO. Para realizar

un análisis más exhaustivo sobre los resultados y siguiendo la metodología propuesta, hemos llevado a cabo el test no paramétrico de Iman-Davenport para cada grupo.

En la Figura 1 mostramos los rangos medios obtenidos para cada caso, necesarios para la ejecución del test. Para calcularlos en cada conjunto de métodos, a cada clasificador le asignamos la posición respecto a su rendimiento en cada conjunto de datos y a continuación obtenemos el valor medio sobre todos los conjuntos. Observando la figura, vemos que los rankings más bajos vienen dados por niveles de ruido menores. Es interesante observar la diferencia entre los OVO y su clasificador base, ya que en todos los casos (salvo para 1NN con niveles bajos de ruido) la aproximación OVO es mejor. En los 8 casos, el test de Iman-Davenport rechaza la hipótesis de igualdad y por tanto realizamos el test post-hoc de Holm para cada conjunto.

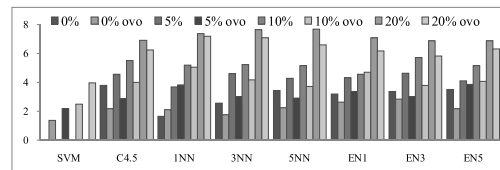


Figura 1: Rangos para ruido en clases.

En el Cuadro 4 mostramos los resultados de estos test. Cada columna representa a un grupo y por tanto un test. Cada fila representa el nivel de ruido y la técnica utilizada (agregamos *ovo* para indicar el caso con la estrategia de binarización). Hemos realizado el test para niveles de significancia  $\alpha = 0,05$  y  $\alpha = 0,1$ .

De este cuadro destacamos cuatro puntos:

1. Inicialmente OVO tiene mejor comportamiento (es elegido generalmente como método de control).
2. Niveles de ruido en clases de hasta 5% pueden ser soportados por la mayoría de los clasificadores sin que su pérdida sea estadísticamente significativa (aunque el comportamiento sí se ve afectado).
3. Las técnicas OVO soportan mejor el ruido de clases. Observamos los niveles del 5% y



### 5.3. Ruido en Atributos en el Conjunto de Entrenamiento (RL)

En el Cuadro 7 mostramos los resultados para el caso en el que añadimos ruido únicamente en el conjunto de entrenamiento (RL). La Figura 3 muestra los rangos medios para cada grupo. Se observa un comportamiento similar al caso de ruido en clases y ruido en atributos LR respecto a las estrategias OVO. En este caso, las mejoras son más destacadas. Destaca la gran capacidad de C4.5<sub>ovo</sub> para soportar el ruido en el entrenamiento, dadas las características de C4.5 en la elección de los atributos (sin alcanzar el mismo nivel, la versión sin OVO también obtiene buenos resultados). Por otro lado, la precisión obtenida por los ensembles de clasificadores a niveles bajos de ruido no puede ser superada por ninguno de los clasificadores base. Al contrario que en el caso LR, ahora sí permiten manejar el ruido de manera adecuada siempre que los clasificadores base no vean significativamente afectados sus resultados.

Cuadro 7: Pérdida y tasa de aciertos con ruido en atributos RL de 5%, 10% y 20%.

Método	Pérdida				Precisión				
	5%	10%	20%	Med.	0%	5%	10%	20%	Med.
SVM <sub>ovo</sub>	1,74	3,20	6,71	3,88	82,68	81,24	80,04	77,14	80,28
C4.5	1,83	1,51	4,09	2,48	83,44	81,91	82,18	80,03	81,89
C4.5 <sub>ovo</sub>	<b>0,72</b>	<b>0,94</b>	<b>2,44</b>	<b>1,36</b>	83,98	83,38	83,19	<b>81,93</b>	83,12
1NN	1,77	3,31	6,10	3,73	82,85	81,38	80,11	77,79	80,53
1NN <sub>ovo</sub>	1,97	3,81	6,55	4,11	83,13	81,48	79,96	77,68	80,56
3NN	1,55	2,04	5,18	2,92	83,63	82,34	81,93	79,30	81,80
3NN <sub>ovo</sub>	1,36	2,06	4,54	2,65	83,95	82,81	82,22	80,14	82,28
5NN	1,07	2,22	3,91	2,40	83,73	82,84	81,87	80,46	82,23
5NN <sub>ovo</sub>	1,48	2,74	4,23	2,82	84,17	82,93	81,87	80,61	82,40
EN1	1,88	2,34	5,48	3,23	84,78	83,19	82,79	80,14	82,73
EN1 <sub>ovo</sub>	0,95	1,72	4,09	2,26	84,67	<b>83,86</b>	83,21	81,20	83,23
EN3	1,72	2,46	4,94	3,04	84,68	83,23	82,60	80,50	82,75
EN3 <sub>ovo</sub>	1,29	1,51	3,61	2,14	84,68	83,59	<b>83,40</b>	81,63	83,33
EN5	1,85	2,68	4,29	2,94	84,72	83,15	82,45	81,08	82,85
EN5 <sub>ovo</sub>	1,69	2,17	3,78	2,55	<b>84,99</b>	83,55	83,14	81,78	<b>83,36</b>

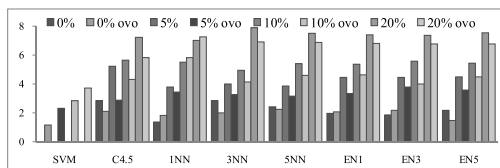


Figura 3: Rangos para ruido en atributos RL.

La ejecución del test de Iman-Davenport re-

chaza las hipótesis de igualdad en todos los casos, por lo que realizamos el test post-hoc de Holm para cada grupo. El Cuadro 8 muestra los resultados del test. Podemos observar como para este tipo de ruido, los algoritmos originales soportan peor el ruido respecto a la estrategia de binarización. Para los clasificadores base, un ruido del 10% provoca diferencias significativas en todos los casos y un 5% en varios de ellos. Para OVO la hipótesis se rechaza en menos ocasiones. Con 5% de ruido y con  $\alpha = 0,1$  nunca se rechaza. Con un 10% de ruido, la mitad de los casos tampoco la rechazan. En general, debemos destacar que el ruido únicamente en test frente al ruido solo en entrenamiento es mucho menos perjudicial. Las diferencias significativas en la mayoría de los métodos se dan a partir del 10%.

Cuadro 8: Test de Holm para ruido en atributos RL. Casilla vacía: método de control por columna; "+": hipótesis de igualdad rechazada; "=": no rechazada; ("="/+"): no se rechaza la hipótesis para  $\alpha = 0,05$  y sí para  $\alpha = 0,1$ .

Método	SVM	C45	1NN	3NN	5NN	EN1	EN3	EN5
0%	×	=	=	=	=	=	=	=
0% ovo								
5%	×	+	=/+	=	=	=/+	=/+	+
5% ovo	+	=	=/+	=	=	=	=	=/+
10%	×	+	+	+	+	+	+	+
10% ovo	+	=/+	+	=	=/+	+	=	+
20%	×	+	+	+	+	+	+	+
20% ovo	+	+	+	+	+	+	+	+

### 5.4. Ruido en Atributos en Ambos Conjuntos (RR)

Finalizamos el estudio analizando la última casuística de ruido en atributos que considera el ruido en ambos conjuntos de datos (RR). En el Cuadro 9 mostramos los resultados de pérdida y precisión. La Figura 4 muestra los rangos obtenidos en cada conjunto de algoritmos. De nuevo la metodología OVO obtiene menores pérdidas y mejores rangos salvo para 1NN. Los ensembles a niveles altos de ruido empeoran considerablemente su rendimiento. A niveles bajos, cuando SVM todavía no está afectado por el ruido, son realmente competitivos, siendo los métodos que mayor precisión

obtienen.

Cuadro 9: Pérdida y tasa de aciertos con ruido en atributos RR de 5%, 10% y 20%.

Método	Pérdida				Precisión				
	5%	10%	20%	Med.	0%	5%	10%	20%	Med.
SVM <sub>ovo</sub>	6,15	9,73	17,09	10,99	82,68	77,60	74,64	68,55	75,87
C4.5	5,79	8,44	15,08	9,77	83,44	78,60	76,39	70,86	77,32
C4.5 <sub>ovo</sub>	3,69	6,84	12,87	7,80	83,98	80,87	78,23	73,17	79,06
1NN	4,47	8,36	14,60	9,15	82,85	79,14	75,92	70,75	77,17
1NN <sub>ovo</sub>	4,90	8,90	15,14	9,65	83,13	79,05	75,72	70,54	77,11
3NN	4,14	6,93	12,11	7,73	83,63	80,17	77,84	73,50	78,79
3NN <sub>ovo</sub>	3,83	6,63	11,74	7,40	83,95	80,73	78,38	74,09	79,29
5NN	<b>3,59</b>	6,68	<b>10,58</b>	<b>6,95</b>	83,73	80,73	78,14	74,88	79,37
5NN <sub>ovo</sub>	3,93	<b>6,55</b>	10,64	7,04	84,17	80,87	78,66	<b>75,22</b>	79,73
EN1	5,04	8,58	15,74	9,79	84,78	80,51	77,51	71,44	78,56
EN1 <sub>ovo</sub>	4,13	7,81	14,40	8,78	84,67	<b>81,17</b>	78,05	72,47	79,09
EN3	4,91	8,33	14,68	9,31	84,68	80,53	77,63	72,25	78,77
EN3 <sub>ovo</sub>	4,16	7,50	12,93	8,20	84,68	81,16	78,34	73,74	79,48
EN5	4,80	8,31	14,20	9,10	84,72	80,66	77,67	72,69	78,93
EN5 <sub>ovo</sub>	4,51	7,18	12,73	8,14	<b>84,99</b>	81,15	<b>78,88</b>	74,17	<b>79,80</b>

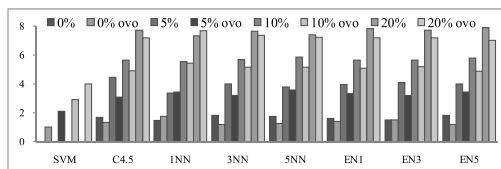


Figura 4: Rangos para ruido en atributos RR.

Aplicamos el test post-hoc de Holm para observar las diferencias entre cada grupo de métodos, ya que la hipótesis de igualdad de los test de Iman-Davenport se rechaza para todos los grupos. Aquí queda patente la bondad de los métodos OVO ante este caso más general de ruido. Los algoritmos originales reducen significativamente su rendimiento con un 5% de ruido. Los que utilizan la binarización en la mayoría de los casos no muestran diferencias significativas con  $\alpha = 0,05$  y nunca  $\alpha = 0,1$ .

## 6. Conclusiones

En esta contribución hemos estudiado la influencia del ruido en los sistemas de clasificación cuando consideramos múltiples clases. Para ello, hemos llevado a cabo un análisis empleando tres algoritmos clásicos (SVM, C4.5 y  $k$ NN) y hemos utilizado una de las técnicas de binarización más comunes a la hora de afrontar los problemas multi-clase como es la estrategia OVO. Hemos realizado un análisis

Cuadro 10: Test de Holm para ruido en atributos RR. Casilla vacía: método de control por columna; “+”: hipótesis de igualdad rechazada; “=”: no rechazada; (“= /+”: no se rechaza la hipótesis para  $\alpha = 0,05$  y sí para  $\alpha = 0,1$ ).

Método	SVM	C45	1NN	3NN	5NN	EN1	EN3	EN5
0%	×	=		=	=	=	=	=
0% ovo			=					
5%	×	= /+	=	+	= /+	+	+	+
5% ovo	= /+	=	=	=	= /+	=	=	= /+
10%	×	+	+	+	+	+	+	+
10% ovo	+	+	+	+	+	+	+	+
20%	×	+	+	+	+	+	+	+
20% ovo	+	+	+	+	+	+	+	+

desde el punto de vista de ruido en clases y en atributos, perturbando los conjuntos de datos originales con niveles de ruido de 5%, 10% y 20%.

A partir de los resultados y de su análisis hemos obtenido tres conclusiones principales. Primero, la bondad de utilizar los métodos OVO en presencia de ruido frente a los clasificadores base. En segundo lugar, destacar que en general, un 5% de ruido es suficiente para que los clasificadores sean significativamente afectados por el ruido. Aunque, dependiendo del tipo de ruido, el nivel soportado varía. Por último, hemos visto que el comportamiento de los ensembles de clasificadores ante el ruido es mejor que el de los clasificadores base. Permiten amortiguar el ruido, logrando en algunos casos mantenerse sin diferencias significativas hasta niveles de ruido del 10%. Logran mantener una buena precisión hasta que el ruido afecta a los clasificadores base significativamente, perjudicando la estrategia de voto.

## Agradecimientos

Este trabajo ha sido posible gracias a la subvención del Ministerio de Educación y Ciencia bajo los Proyectos TIN2007-65981 y TIN2008-06681-C06-01. J. Luengo está subvencionado por una beca FPU del Ministerio de Ciencia e Innovación.

## Referencias

- [1] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
- [2] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J.C. Fernández, and F. Herrera. KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3):307–318, 2008.
- [3] A. Asuncion and D. J. Newman. UCI machine learning repository, 2007.
- [4] C. E. Brodley and M. A. Friedl. Identifying mislabeled training data. *Journal of Artificial Intelligence Research*, 11:131–167, 1999.
- [5] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
- [6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley, 2nd edition, 2001.
- [7] J. Fürnkranz. Round robin classification. *Journal of Machine Learning Research*, 2:721–747, 2002.
- [8] S. García, A. Fernández, J. Luengo, and F. Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180:2044–2064, 2010.
- [9] 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*, 9:2677–2694, 2008.
- [10] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.
- [11] A. C. Lorena, A. C. Carvalho, and J. M. Gama. A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review*, 30(1-4):19–37, 2008.
- [12] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley and Sons, 2004.
- [13] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.
- [14] J. R. Quinlan. *C4.5: Programs for Machine Learning*. San Mateo-California: Morgan Kaufmann Publishers, 1st edition, 1993.
- [15] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.
- [16] D. Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, 2nd edition, 2006.
- [17] 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.
- [18] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 29(5):854–869, 2007.
- [19] V. Vapnik. *Statistical Learning Theory*. New York: Wiley, 1998.
- [20] X. Wu and X. Zhu. Mining with noise knowledge: Error-aware data mining. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 38(4):917–932, 2008.
- [21] X. Zhu and X. Wu. Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, 22(3):177–210, 2004.

## ÍNDICE DE AUTORES

- Agell, Núria, 303  
Aguilar-Ruíz, Jesús S., 251  
Alba-Elías, F., 211  
Albisua, Iñaki, 111  
Alonso-Barba, Juan I., 93  
Álvarez, Eva, 319  
Álvarez, Miguel R., 55  
Alzate, Leandro, 129  
Angulo, Cecilio, 75  
Arbelaitz, O., 293  
Baena-García, M., 283  
Balcázar, José L., 329  
Banchs, Rafael, 201  
Bermúdez, José D., 377  
Berzal, Fernando, 3, 11, 19  
Bustillo, Andrés, 275  
Cadenas, J. M., 265  
del Campo-Ávila, José, 165  
Cariñena, Purificación, 55  
Carmona-Cejudo, J. M., 283  
Català, Andreu, 75  
Codina, Joan, 201  
Codina, Víctor, 37  
Cordón, Óscar, 219  
Cortijo, Francisco, 11  
del Coz, Juan José, 319  
Cubero, Juan-Carlos, 3, 11, 19  
Cugota, L., 345  
Derrac, Joaquín, 155  
Díaz Díaz, Norberto, 251  
Diez, Jorge, 319  
Diez-Pastor, José F., 275  
Domínguez, Juan Luis, 173  
Escribano, Rubén, 147  
Estruch, Vicent, 183  
Félix, Paulo, 55  
Fernández, Julio, 147, 311  
Fernández, Roberto, 311  
Ferri Francesc J., 83  
Ferri, César, 183  
Flores, Iván, 193  
Florido, J. P., 369  
Fornells, A., 345  
Gago, Mónica, 361  
Galar, Mikel, 65  
Gámez, Matías, 3  
García, Jesús, 121  
García, Salvador, 155  
García-Escudero, L. A., 237  
García Gutiérrez, Jorge, 243  
García-Osorio, César, 275  
García-Piquer, A., 345  
Gibaja, Eva, 103  
Gibert, Karina, 37  
Golobardes, E., 345  
Gómez Vela, Francisco, 251  
González, Ana, 147  
González, Gustavo, 45  
González, Pablo, 319  
González-Arteaga, M. T., 237  
González-Marcos, A., 211  
González-Quirós, Rafael, 319  
Grivolla, Jens, 201  
Gurrutxaga, I., 293  
Gutiérrez, Germán, 355  
Hernández-Orallo, José, 183  
Herrera, Francisco, 65, 155  
Herrera, L. J., 369  
Inza, Iñaki, 129  
del Jesús, María José, 383  
Jiménez, Aida, 3, 11, 19  
Juaristi, Elizabet, 361  
Kaltenbrunner, Andreas, 45, 201  
López-Urrutia, Ángel, 319  
Lozano, José A., 129  
Luis Ávila, José, 103  
Marbán, Óscar, 139  
Mariscal, Gonzalo, 139  
Martínez de Pisón, F. Javier, 311  
Martínez-Plumed, Fernando, 183  
Mata Vázquez, Jacinto, 173, 339  
Mateos García, Daniel, 243

- Mayo, A., 237  
Meda Campaña, María Elena, 29  
Meza, Rodrigo, 201  
Molero Castillo, Guillermo Gilberto, 29  
Molina, José M., 121  
Morales-Bueno, Rafael, 165, 283  
Moreno, Gabriel, 227  
Muguerza, Javier, 111, 293  
Muñoz-Ledesma, A., 265  
Nepomuceno, Isabel, 251  
Nogueira, Enrique, 319  
Núñez, Haydemar, 193  
Ordieres-Meré J., 211  
Orriols-Puig, A., 345  
Ortigosa-Hernández, Jonathan, 129  
de la Ossa, Luis, 93  
Pachón Álvarez, Victoria, 339  
Pantoja, Liliana, 227  
Peralta, Juan, 355  
Pérez, Gonzalo, 121  
Pérez, Jesús M., 111  
Pérez, Nelson, 227  
Pérez Godoy, María Dolores, 383  
Pérez-Recuerda, Pedro, 383  
Pérez-Suay, Adrián, 83  
Perona, I., 293  
Pomares, H., 369  
Puerta Callejón, José M., 93  
Quirin, Arnaud, 219  
Ramírez-Quintana, María José, 183  
Ramos, Esmeralda, 193  
Ramos-Jiménez, Gonzalo, 165  
Reñones, Anibal, 275  
Riquelme Santos, José C., 243  
Rivera, Antonio J., 383  
Rodríguez, Juan Diego, 129  
Rodríguez, Juan J., 275  
Rodríguez, Vidal, 193  
Rodríguez Baena, Domingo S., 251  
Rojas, I, 369  
Ruiz, Francisco J., 75  
Ruiz de Querol, Ricard, 45  
Sáez, Julián Luengo, José A., 65  
Samà, Albert, 75  
Sanabria, José Antonio, 303  
Sánchez, Germán, 303  
Sánchez-Marrè, Miquel, 37  
Sanchís, Araceli, 355  
Santos, José Antonio, 339  
Sanz, Andrés, 147, 311  
Sayeras Josep M., 303  
Segovia, F. Javier, 139  
Segura, José Vicente, 377  
Shelokar, Prakash, 219  
Tirnauca, Cristina, 329  
Urquiza, J. M, 369  
Uruburu, A., 211  
Ventura, Sebastián, 103  
Vercher, Enriqueta, 377  
Victoriano, Manuel, 103  
Vitelli, Juana, 193  
Volkovich, Yana, 45  
Zarco, Carmen, 219  
Zorrilla, Marta E., 329