

NUEVOS DISEÑOS EXPERIMENTALES PARA ALGORITMOS DE EXTRACCIÓN DE CONOCIMIENTO CON DATOS DE BAJA CALIDAD

José Otero¹

¹ Edificio Departamental 1, 33204 (Gijón) Asturias, jotero@uniovi.es

Resumen

El estudio de los diseños experimentales para algoritmos que usan datos de baja calidad y particularmente, datos intervalares ha recibido la atención por parte de la comunidad de Machine Learning en los últimos años. Hay consenso en que no se pueden extraer conclusiones de las técnicas habituales como las basadas en validación cruzada cuando la varianza de los puntos centrales de los intervalos es del mismo orden que la imprecisión de los datos. En este trabajo, se muestra un estudio empírico preliminar en donde se usa la extensión intervalar habitual de un test de hipótesis no paramétrico para incorporarlo a un diseño experimental basado en validación cruzada con el fin de solventar este problema. Usando esta configuración se analizan los resultados de una experimentación con datos intervalares, realizando una estimación del error de tipo I y de la potencia de los tests utilizados.

Palabras Clave: Datos de baja calidad, tests estadísticos, diseño experimental.

1 INTRODUCCIÓN

Existe una gran cantidad de material bibliográfico en donde se recopilan o analizan la mayor parte de los diseños experimentales y tests estadísticos de aplicación en experimentaciones de Machine Learning [10], ya que la inclusión de estos en el análisis de los resultados se ha hecho obligatoria [3]. En la casi totalidad de los casos, estos estudios se ciñen a las situaciones en las que los datos son nítidos, sin incertidumbre. Sin embargo es habitual en aplicaciones reales encontrarse

con esta en los datos por distintos motivos, tolerancias en las medidas, datos perdidos o censurados, etc. La temática del análisis estadístico de datos con imprecisión está también muy desarrollada [1, 8], pero no así su aplicación a experimentaciones de Machine Learning. En este trabajo pretendemos mostrar algunos resultados preliminares de la aplicación directa de las técnicas existentes en Estadística a las experimentaciones de Machine Learning con datos de baja calidad. Los resultados provienen de una experimentación con datos sintéticos que nos permite controlar la tolerancia de los datos y la dificultad del problema (para el test estadístico). El objetivo es observar como se comportan las extensiones habituales de los tests estadísticos cuando se aplican a experimentaciones de Machine Learning en las que los resultados poseen imprecisión. En la siguiente sección se describe el problema de las fuentes de imprecisión en los algoritmos y se citan las contribuciones más destacadas en el campo del análisis estadístico de datos imprecisos. En la sección 3 se detallan los experimentos realizados y se muestran los resultados obtenidos. Finalmente en 4 recopilamos sobre lo discutido en este trabajo en la forma de conclusiones y posible líneas futuras de trabajo.

2 DESCRIPCIÓN DEL PROBLEMA

Existen diversas situaciones del mundo real en donde la utilización de datos con una imprecisión intrínseca es implícita. En [12] se mencionan varios ejemplos algunos de los cuales reproducimos a continuación:

- Dispositivos digitales de medición con un número de cifras significativas determinado. Por ejemplo, una báscula digital con un sólo decimal para indicar el peso en kilogramos. Evidentemente, al pesar objetos desde x.00... hasta x.09... la báscula indicará la misma medida.
- Dispositivos que proporcionan una indicación de la tolerancia de la medición asociada a cada nivel

de confianza. Por ejemplo, un GPS, dispositivo que proporciona la posición del receptor sobre la corteza terrestre con la tolerancia (CEP, Círculo de Error Probable) a distintos niveles de probabilidad. Típicamente se da un CEP de 3 metros al 50% y 5 metros al 90%. Esto quiere decir que con la probabilidad mencionada el receptor estará *realmente* dentro de un círculo del radio correspondiente, centrado en la posición *medida*.

- Observación por parte de un ser humano de un instrumento analógico. Por ejemplo los instrumentos con indicación óptica y escala, todos ellos son susceptibles de ofrecer distintas medidas en función del ángulo de observación, debido al efecto del paralaje.

Cuando se tratan datos de esta clase mediante algoritmos que no los convierten de algún modo en datos nítidos, es posible obtener resultados que a su vez sean imprecisos. Si se deben comparar los resultados obtenidos por distintos algoritmos de este tipo, surge al necesidad de utilizar las herramientas estadísticas desarrolladas para este tipo de datos.

El estudio de los tests estadísticos y diseños experimentales para datos crisp ha sido objeto de estudios pormenorizados por parte de la comunidad de Machine Learning recientemente [10]. En el caso de datos de baja calidad o con imprecisión, el material bibliográfico es más escaso. Uno de los artículos más exhaustivos y extensos sobre el tema en general es [8]. En [7] se muestra como extender un test no paramétrico para poder ser usado con datos intervalares o fuzzy, de modo que los resultados del mismo sean un intervalo o un conjunto borroso, respectivamente. Este trabajo guarda cierta relación con [5], en el que se define el concepto de región de confianza para datos borrosos. Por otra parte en [4] se propone un método para defuzzificar p-valores borrosos y en [6] se da una definición de potencia para test de hipótesis para datos con imprecisión. Un texto completamente dedicado al análisis de datos fuzzy es [1], en donde se recopilan una serie de trabajos en los que se muestra como realizar tests de hipótesis con datos fuzzy, regresión, etc. En [14] se muestra una estimación intervalar de la media de una población con datos imprecisos.

La obtención de modelos y clasificadores a partir de datos de este tipo tiene sus problemas asociados, sin embargo, en este trabajo, sólo nos ocuparemos de la etapa de test. Es decir, como obtener una muestra del error del clasificador o modelo que responda a la cualidad imprecisa de los datos y como utilizar las herramientas estadísticas existentes para comparar distintos clasificadores o modelos. Es decir, en primer lugar obtendremos una muestra del error de test de

cada clasificador, de modo que cada elemento de las muestras sea un intervalo. En segundo lugar, compararemos esas muestras intervalares mediante un test de hipótesis adaptado a datos de este tipo. Este test de hipótesis proporcionará un intervalo de p-valores y analizaremos su potencia y error de Tipo I, los cuales serán también intervalares [6].

3 ESTUDIO EXPERIMENTAL

3.1 METODOLOGÍA

Con el fin de estudiar el comportamiento de las técnicas estadísticas existentes más representativas para evaluar modelos y clasificadores sobre datos con imprecisión se han realizado una serie de experimentos con un dataset sintético parametrizable. Este dataset se ha construido partiendo de la idea original de Haykin [9], consistente en dos gaussianas solapadas. De esta forma se han construido dos muestras de tamaño n , la primera se compone de pares (x^1, y^1) en donde x^1 sigue una $N(x_c^1, v^1)$ e y^1 sigue una $N(y_c^1, v^1)$ (siendo $(x^1$ e $y^1)$ independientes) y la segunda de pares (x^2, y^2) en donde x^2 sigue una $N(x_c^2, v^2)$ e y^2 sigue una $N(y_c^2, v^2)$ ($(x^1$ e $y^1)$ también independientes). Las dos muestras se concatenan y los índices de la muestra resultante se permutan. Sobre la muestra con estos índices permutados se aplica 10-cv, con el fin de obtener una muestra de tamaño 10 del error de los clasificadores. Más adelante explicaremos como obtener muestras del error intervalares, considerando una tolerancia en los datos de test.

Esto nos ha permitido realizar experimentos sobre condiciones controladas:

- La dificultad del problema (para el test) se gradúa mediante la distancia entre los centros de dichas gaussianas.
- La tolerancia de los datos de test incide en las diferencias entre el error máximo y mínimo, a mayor tolerancia mayores son las diferencias entre estos, lo que a su vez incide sobre la diferencia entre los p-valores máximos y mínimos.
- Es posible establecer a priori una serie de clasificadores que se puedan ordenar según su error para este problema concreto. Para este trabajo hemos escogido los clasificadores lineal, cuadrático y el óptimo bayesiano. No existen diferencias significativas entre el cuadrático y el óptimo y si entre el lineal y el óptimo. De esta forma se puede estimar el error de Tipo I comparando los test cuadrático y óptimo Bayesiano (H_0 cierta) y el de Tipo II (y por lo tanto la potencia) comparando el lineal y el óptimo (H_0 falsa).

En el caso del dataset sintético utilizado, el error de test máximo y mínimo se determina mediante un análisis de Montecarlo, como se sugiere en [8]. Sean $\{(x_1, y_1, c_1), (x_2, y_2, c_2), \dots, (x_i, y_i, c_i), \dots, (x_k, y_k, c_k)\}$ los datos de la partición de test, en donde cada (x_i, y_i) es un punto que pertenece a la clase c_i , en el caso de este trabajo $c_i \in \{1, 2\}$. A cada (x_i, y_i) se le añade una cantidad aleatoria dentro de la tolerancia especificada, de modo que se obtiene $(x_i + d_{x_i}, y_i + d_{y_i})$ con $d_{x_i} \in (-t, t)$ y $d_{y_i} \in (-t, t)$. Se evalúa el clasificador mediante la partición de test modificada de esta forma y se obtiene el error de test, se actualiza el máximo y el mínimo para esa partición. Este proceso se realiza un cierto número (evidentemente suficientemente elevado) de veces.

En lo que se refiere al cálculo de los p-valores máximo y mínimo, hemos utilizado el test de Wilcoxon adaptado a datos intervalares usando el método que se sugiere en la referencia citada.

Dadas dos muestras intervalares $X^1 = \{(e_{min\ 1}^1, e_{max\ 1}^1), (e_{min\ 2}^1, e_{max\ 2}^1), \dots, (e_{min\ n}^1, e_{max\ n}^1)\}$, $X^2 = \{(e_{min\ 1}^2, e_{max\ 1}^2), (e_{min\ 2}^2, e_{max\ 2}^2), \dots, (e_{min\ n}^2, e_{max\ n}^2)\}$ se enfrenta el cálculo de los p-valores máximo y mínimo como un proceso de optimización de la función $W(e_1^1, e_2^1, \dots, e_n^1, e_1^2, e_2^2, \dots, e_n^2)$ que devuelve el p-valor correspondiente al test de Wilcoxon para las muestras $\{e_1^1, e_2^1, \dots, e_n^1\}, \{e_1^2, e_2^2, \dots, e_n^2\}$.

El objetivo es encontrar los valores nítidos $X_{min}^1 = \{e_{min\ 1}^{*1}, e_{min\ 2}^{*1}, \dots, e_{min\ n}^{*1}\} / e_{min\ i}^{*1} \in (e_{min\ i}^1, e_{max\ i}^1)$, $X_{min}^2 = \{e_{min\ 1}^{*2}, e_{min\ 2}^{*2}, \dots, e_{min\ n}^{*2}\} / e_{min\ i}^{*2} \in (e_{min\ i}^2, e_{max\ i}^2)$, que proporcionan el valor mínimo de W y $X_{max}^1 = \{e_{max\ 1}^{*1}, e_{max\ 2}^{*1}, \dots, e_{max\ n}^{*1}\} / e_{max\ i}^{*1} \in (e_{min\ i}^1, e_{max\ i}^1)$, $X_{max}^2 = \{e_{max\ 1}^{*2}, e_{max\ 2}^{*2}, \dots, e_{max\ n}^{*2}\} / e_{max\ i}^{*2} \in (e_{min\ i}^2, e_{max\ i}^2)$ que proporcionan el valor máximo de la función W .

En [7] se sugiere de nuevo emplear el Método de Montecarlo para este proceso de optimización y en [6] se comenta la dificultad del problema de optimización. Por esta razón, en este trabajo hemos escogido utilizar el método L-BGFS-B [2] del cual existe una implementación para el paquete R en lugar del método de Montecarlo. El motivo es que para muestras de tamaño 10, como las que se obtienen al aplicar 10-cv, el número de variables que intervienen es 20, con lo cual el problema es abordable por este método en menos tiempo del que se necesitaría usando Montecarlo.

Los valores de los p-valores máximos y mínimos obtenidos se utilizan para estimar la potencia y el error máximos y mínimos. La estimación de la potencia se realiza como se propone en [6] para este caso concreto de imprecisión.

Cada vez que se comparan los clasificadores lineal y

exacto (H0 falsa) se estiman los p-valores máximos y mínimos. La fracción de veces en que el p-valor máximo es mayor que el nivel escogido es la potencia máxima. La fracción de veces en que el p-valor mínimo es mayor que el nivel escogido es la potencia mínima. Análogamente, cuando se comparan los clasificadores cuadrático y exacto (H0 cierta), la fracción de veces en que el p-valor mínimo es menor que el nivel es el error de Tipo I máximo y la fracción de veces en que el p-valor máximo es menor que el nivel es el error de Tipo I mínimo.

A modo de resumen gráfico de lo que se pretende realizar, véase la figura 1. En dicha figura se representan parte de los puntos del conjunto de datos utilizado, así como las superficies de decisión, para los clasificadores lineal (izquierda) y cuadrático (derecha). En cada figura se muestra, alrededor de cada punto un cuadrado del tamaño de la tolerancia. En la parte superior la tolerancia es de 0.05 y en la inferior de 0.1. Como puede verse, debido a la tolerancia de los datos, durante el proceso de determinación del error de test máximo y mínimo, es posible que los puntos utilizados para estimar el error estén a uno u otro lado de la superficie de decisión. Cuanto mayor es la tolerancia, mayor será el número de puntos susceptibles de estar a uno u otro lado de la superficie de decisión y por lo tanto la diferencia entre el error máximo y mínimo será mayor. Consiguientemente la diferencia entre los p-valores menor y mayor también será mayor.

De esta forma, se pueden poner de relieve diferencias entre clasificadores con un error nítido similar cuando la superficie de decisión de uno de ellos pasa más cerca de zonas con mayor densidad de ejemplos [13].

3.2 RESULTADOS

En la tabla 1 se detallan los parámetros de los experimentos realizados. Como puede observarse, se ha generado el dataset con puntos procedentes de dos Gaussianas bidimensionales con matrices de covarianza 1.0,0,0,1.0) y (2.0, 0,0,2.0), siendo el dataset balanceado. La tolerancia utilizada varía desde 0.05 a 0.10, en pasos de 0.01. El centro de una de las Gaussianas está fijo en el punto (0,0) mientras que el otro se mueve de (2.0,0) a (2.1,0) y finalmente a (2.2,0).

En la tabla 2 se muestran las estimaciones de potencia y error de Tipo I, máximos y mínimos, que se han obtenido para las combinaciones de parámetros de la tabla 1. Realmente, para analizar los resultados de forma objetiva, habría que utilizar alguna herramienta estadística clásica a los resultados de potencia y error de tipo I y para ello se necesitaría un conjunto de resultados mayor. Como no es el caso se tratará de extraer alguna conclusión cualitativa con los resultados de que

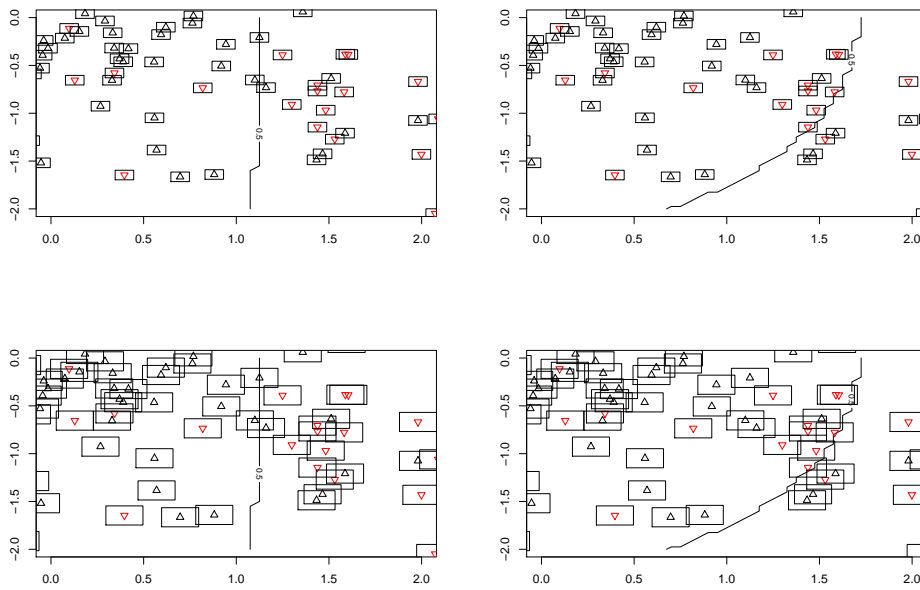


Figura 1: Ejemplo gráfico de la influencia de la tolerancia en el conjunto de test. Arriba tolerancia 0.05, abajo tolerancia 0.01. Derecha clasificador cuadrático, izquierda clasificador lineal. Los puntos están etiquetados con triángulos que apuntan hacia arriba o hacia abajo según la clase a la que pertenezcan. Obsérvese como al aumentar la tolerancia el número de puntos imprecisos susceptibles de intersecar la superficie de decisión aumenta.

Tabla 1: Valores de los parámetros utilizados en los experimentos

Rep. Mont.	50
Rep. 10-cv	50
Var. Gaus.	1.0 y 2.0
Centros Gaus.	(0,0) y (2.0,0), (2.1,0), (2.2,0)
Tolerancia	0.05, 0.06, 0.07, 0.08, 0.09, 0.10
Tam. dataset	500+500

se dispone.

Si se examinan las primeras cuatro columnas de la tabla 2, se puede observar como el intervalo obtenido para la potencia se hace gradualmente mayor. Este comportamiento es coherente con la variación de la tolerancia, como se dijo en la sección anterior, cuanto mayor es la tolerancia más fácil es (durante el proceso de optimización que calcula los p-valores) que los datos (imprecisos) del conjunto de test intersequen con la superficie de decisión.

En cuanto al error de Tipo I, no se aprecia variación alguna relacionada con las combinaciones de parámetros tolerancia vs. distancia entre centros. Como es habitual, se aprecia un incremento del error de Tipo I

cuando el nivel de significación aumenta.

Si se examina la variación de los resultados a través de las cuatro primeras filas, se observa como aumenta el tamaño del intervalo para la potencia a medida que las gaussianas están más separadas. Parece coherente con la experimentación planteada ya que las diferencias entre los clasificadores lineal y cuadrático disminuyen al estar las gaussianas más separadas. De forma dual a este efecto, cuanto más separadas están las gaussianas, menos errores de Tipo I se producen. Los efectos comentados se degradan rápidamente cuando la tolerancia de los datos aumenta, obsérvese las siguientes cuatro filas.

Todos estos efectos son más difíciles de observar cuando se examinan los resultados variando simultáneamente filas y columnas. Es posible que los resultados no sean concluyentes debido a que el número de repeticiones en el análisis de Montecarlo del error máximo y mínimo deba de ser más elevado.

4 CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha mostrado un estudio experimental preliminar en el que las técnicas más recientes, que

Tabla 2: Potencias y Errores de Tipo I, máximos y mínimos para cada combinación de parámetros, a los niveles 0.1, 0.05 y 0.01

t=0.03											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.840	1.000	0.000	0.380	0.960	1.000	0.000	0.300	0.260	0.980	0.000	0.100
1.000	1.000	0.000	0.380	1.000	1.000	0.000	0.300	0.760	1.000	0.000	0.100
1.000	1.000	0.000	0.920	1.000	1.000	0.000	0.880	0.920	1.000	0.000	0.820
t=0.04											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.9800	1.000	0.000	0.440	1.000	1.000	0.000	0.100	0.260	0.960	0.000	0.040
1.000	1.000	0.000	0.440	1.000	1.000	0.000	0.100	0.760	1.000	0.000	0.040
1.000	1.000	0.000	0.920	1.000	1.000	0.000	0.700	0.920	1.000	0.000	0.420
t=0.05											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.033	1.000	0.000	0.100	0.133	1.000	0.000	0.400	0.000	1.000	0.000	0.1667
0.667	1.000	0.000	0.100	0.900	1.000	0.000	0.400	0.433	1.000	0.000	0.1667
0.800	1.000	0.000	0.800	0.933	1.000	0.000	0.400	0.700	1.000	0.000	0.867
t=0.06											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.533	1.000	0.000	0.233	0.000	0.933	0.000	0.067	1.000	1.000	0.000	0.300
0.967	1.000	0.000	0.233	0.000	1.000	0.000	0.067	1.000	1.000	0.000	0.300
1.0000	1.000	0.000	0.900	0.000	1.000	0.000	0.933	1.000	1.000	0.000	0.833
t=0.07											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.067	1.000	0.000	0.233	0.400	1.000	0.000	0.300	0.000	1.000	0.000	0.267
0.567	1.000	0.000	0.233	0.867	1.000	0.000	0.300	0.033	1.000	0.000	0.267
0.767	1.000	0.000	0.733	0.967	1.000	0.000	0.800	0.233	1.000	0.000	0.767
t=0.08											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.000	1.000	0.000	0.133	0.000	1.000	0.000	0.067	0.000	1.000	0.000	0.167
0.233	1.000	0.000	0.133	0.000	1.000	0.000	0.067	0.233	1.000	0.000	0.167
0.233	1.000	0.000	0.9	0.167	1.000	0.000	0.733	0.600	1.000	0.000	0.867
t=0.09											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.000	1.000	0.000	0.133	0.267	1.000	0.000	0.100	0.000	1.000	0.000	0.367
0.200	1.000	0.000	0.133	0.800	1.000	0.000	0.100	0.167	1.000	0.000	0.367
0.400	1.000	0.000	0.867	0.967	1.000	0.000	0.833	0.333	1.000	0.000	0.933
t=0.10											
xc1=2.0				xc1=2.1				xc1=2.2			
Potencia		Err. I		Potencia		Err. I		Potencia		Err. I	
min	max	min	max	min	max	min	max	min	max	min	max
0.967	1.000	0.000	0.133	0.267	1.000	0.000	0.267	0.200	1.000	0.000	0.200
1.000	1.000	0.000	0.133	0.733	1.000	0.000	0.267	0.767	1.000	0.000	0.200
1.000	1.000	0.000	0.933	0.800	1.000	0.000	0.767	0.933	1.000	0.000	0.900

extienden los test de hipótesis habituales a datos intervalares, se aplican a una experimentación de Machine Learning con datos sintéticos. Variando de forma controlada los parámetros de la experimentación se ha comprobado (informalmente) que los resultados obtenidos, en lo que se refiere a error de Tipo I y potencia, parecen ser coherentes con la variación en tolerancia de los datos de test y con la dificultad del problema en si, el grado de solapamiento entre las gaussianas. De esta forma, se puede afirmar que un análisis estadístico de este tipo puede complementar las propuestas de los trabajos en los que se usan datos de baja calidad con técnicas de Soft computing [12] [13].

En relación a los métodos usados para las estimaciones del error de test máximo y mínimo y para los p-valores máximos y mínimos, es necesario afinar lo más posible estos, ya que la estimación de la potencia y del error de Tipo I se cimentan en estas estimaciones. En este sentido se baraja sustituir el método de optimización "L-BFGS-B" por otro basado en un algoritmo genético [11]. Otro factor a estudiar es el número de particiones empleadas en la validación cruzada.

Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Ciencia y Tecnología en el marco del proyecto TIN2008-06681-C06-04.

Referencias

- [1] C. Bertoluzza, M. A. Gil, D. A. Ralescu (Eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Series: Studies in Fuzziness and Soft Computing , Vol. 87 2002, XIV
- [2] R. H. Byrd, P. Lu and J. Nocedal. A Limited Memory Algorithm for Bound Constrained Optimization, (1995), SIAM Journal on Scientific and Statistical Computing , 16, 5, pp. 1190-1208.
- [3] O. Cerdón, F. A. C. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, Genetic fuzzy systems. New developments, *Fuzzy Sets Syst.*, vol. 141, no. 1, pp. 173, 2004.
- [4] I. Couso, L. Sánchez. Defuzzification of Fuzzy p-Values. *Advances in Soft Computing* (2008) 126-132
- [5] I. Couso, L. Sánchez. Inner and outer fuzzy confidence regions determined by low quality data. Eurofuse Workshop 09: Preference Modelling and Decision Analysis
- [6] I. Couso. Como Hacer Inferencia Estadística a partir de datos de baja calidad. ESTYLF 2010.
- [7] T. Denoeux, M. H. Masson, P.A. Hbert, Nonparametric Rank-Based Statistics and Significance Test for Fuzzy Data, *Fuzzy Sets and Systems*, 153 (2005) 1-28
- [8] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, L. Ginzburg. Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. Technical Report SAND2007-0939. Sandia National Laboratories, Albuquerque, New Mexico
- [9] S. Haykin. Neural Networks, A Comprehensive Foundation. Prentice Hall. 1999
- [10] S. García, F. Herrera. An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of Machine Learning Research* 9 (2008) 2677-2694
- [11] W. R. Mebane Jr., J. S. Sekhon. The rgenoud Package. 2002.
- [12] L. Sánchez, I. Couso. Advocating the use of Imprecisely Observed Data in Genetic Fuzzy Systems. em IEEE Transactions on Fuzzy Systems 15:4 (2007) 551-562
- [13] L. Sánchez, J. Otero. Learning Fuzzy Linguistic Models from Low Quality Data by Genetic Algorithms. Proceedings of the 16th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE07). London (United Kingdom, 2007) 1921-1926
- [14] S.H. Ong, Rahul Mukerjee. Interval estimation of the population mean under model uncertainty: Robust versus empirical statistics. *Journal of Statistical Planning and Inference*, Volume 138, Issue 12, 1 December 2008, Pages 3696-3704.