

Estudio de la influencia de las medidas de complejidad de los datos en los Sistemas de Clasificación Basados en Reglas Difusas: Análisis de la Razón Discriminante de Fisher

Julián Luengo¹ Salvador García¹ José Ramón Cano² Francisco Herrera¹

¹ Dept. CCIA, ETSIT, Univ. Granada, Granada, 18071, España, {julianlm,salvag1,herrera}@decsai.ugr.es

² Dept. Ciencias de la Computación, CPS, Univ. Jaén, Linares, 23700, España, jrcano@ujaen.es

Resumen

El uso de Sistemas de Clasificación Basados en Reglas Difusas es bien conocido y tiene multitud de aplicaciones actualmente. No obstante, los problemas de clasificación pueden ser complejos, y esta complejidad ha sido tipificada y medida por diversos autores. La capacidad de predicción de un clasificador, y por tanto de un Sistema de Clasificación Basado en Reglas Difusas, posee una fuerte dependencia con la complejidad inherente de los datos usados.

En este trabajo presentamos un estudio preliminar sobre el comportamiento de los sistemas de clasificación basados en reglas difusas para la medida de complejidad denominada "razón discriminante de Fisher". Queremos estudiar en que medida las medidas de complejidad nos pueden aportar conocimiento a priori sobre el comportamiento de los sistemas basados en reglas difusas para estos problemas. Para ello comenzamos con esta primera medida y con un conjunto de bases de datos artificiales que se han generado para disponer de valores de la medida de Fisher en todo su rango, que nos permiten extraer unas primeras conclusiones y plantear diferentes trabajos a realizar de cara a tener un completo estudio para esta medida.

Palabras Clave: Clasificadores, Medidas de Complejidad, Medidas de Solapamiento, Sistemas de Clasificación Basados en Reglas Difusas.

1 INTRODUCCIÓN

Los Sistemas de Clasificación Basados en Reglas Difusas (SCBRD) [4] son métodos de clasificación que emplean reglas difusas para representar el conocimiento. Los SCBRDs se encuentran muy extendidos en la actualidad, con numerosas aplicaciones y estudios de su comportamiento y efectividad [9].

En la literatura especializada existen precedentes recientes que intentan caracterizar los datos sobre los que trabaja un método de clasificación utilizando las llamadas medidas de complejidad [1, 8]. Las medidas de complejidad se han aplicado en análisis similares para algunos Sistemas de Clasificación que toman como base las medidas de Ho y Basu[1]. Entre ellos podemos encontrar selección de características [6] y algoritmos de inducción de reglas[11, 3]. Sánchez et al. [16] analizan el efecto de la complejidad de datos en un clasificador K-Vecinos más Cercanos. Sin embargo, no se ha aplicado ningún análisis de este tipo a los SCBRDs.

Para efectuar este estudio preliminar utilizaremos una medida de complejidad para el *solapamiento* de las clases [8]: La razón discriminante de Fisher. Queremos analizar la relación existente entre los resultados obtenidos por los SCBRD y los valores de esta medida de solapamiento. Para ello, hemos analizado el comportamiento de modelos SCBRDs sencillos considerando conjuntos de datos generados de acuerdo a la medida establecida en [8]. A su vez, lo hemos comparado con un algoritmo clásico de clasificación como es C4.5 [15], de forma que sea posible ver las diferencias entre modelos de clasificación de diferentes familias.

Esta contribución se organiza de la siguiente manera. En la Sección 2 se describen los SCBRDs, y en particular el modelo usado en nuestro estudio. En la Sección 3 se define y describe la medida de solapamiento que emplearemos. Asimismo, en la Sección 4 se establece el marco experimental y los resultados obtenidos. Fi-

nalmente, en la Sección 5 expresamos las conclusiones de nuestro trabajo.

2 SISTEMAS DE CLASIFICACIÓN BASADOS EN REGLAS DIFUSAS

Un SCBRDs está compuesto por una Base de Conocimiento (BC) y un Motor de Razonamiento Difuso (MRD) que, utilizando la información de la BC, determina una clase para cualquier patrón de datos admisible que llegue al sistema. La potencia del razonamiento aproximado reside en la posibilidad de obtención de un resultado (una clasificación) incluso cuando no tengamos compatibilidad exacta (con grado 1) entre el ejemplo y el antecedente de las reglas.

2.1 Base de Conocimiento

La BC está formada por dos componentes:

- La *Base de Datos* (BD), que contiene la definición de los conjuntos difusos asociados a los términos lingüísticos utilizados en la Base de Reglas.
- La *Base de Reglas* (BR), formada por un conjunto de reglas de clasificación

$$R = \{R_1, \dots, R_L\} \quad (1)$$

En nuestro caso vamos a utilizar reglas difusas con una clase y un grado de certeza asociado a la clasificación para esa clase en el consecuente.

$$R_k : \text{ Si } X_1 \text{ es } A_1^k \text{ y } \dots \text{ y } X_N \text{ es } A_N^k \text{ entonces } Y \text{ es } C_j \text{ con grado } r_k \quad (2)$$

Donde X_1, \dots, X_N son las variables asociadas a los diferentes atributos del sistema de clasificación, A_1^k, \dots, A_N^k son las etiquetas lingüísticas utilizadas para discretizar los dominios continuos de las variables, e Y es la variable que indica la clase C_j a la cual pertenece el patrón. r_k es el grado de certeza asociado a la clasificación de la clase C_j para los ejemplos pertenecientes al subespacio difuso delimitado por el antecedente de la regla.

2.2 Motor de Razonamiento Difuso

El MRD es un procedimiento de inferencia que utiliza la información de la BC para predecir una clase ante un ejemplo no clasificado. Tradicionalmente en la literatura especializada [5] se ha utilizado el MRD del máximo, también denominado MRD clásico o de la regla ganadora, que considera la clase indicada por una

sola regla teniendo en cuenta el grado de asociación del consecuente de la regla sobre el ejemplo. Otros MRD que combinan la información aportada por todas las reglas que representan el conocimiento de la zona a la que pertenece el ejemplo se estudian en [5]. En este trabajo utilizaremos, el MRD clásico.

A continuación presentamos el modelo general de razonamiento difuso que combina la información proporcionada por las reglas difusas compatibles con el ejemplo.

En el proceso de clasificación del ejemplo $e = (e_1, \dots, e_N)$, los pasos del modelo general de un MRD son los siguientes:

1. **Calcular el grado de emparejamiento del ejemplo con el antecedente de las reglas.**
2. **Calcular el grado de asociación del ejemplo a la clase consecuente de cada regla mediante una función de ponderación entre el grado de emparejamiento y el grado de certeza de la regla con la clase asociada.**
3. **Determinar el grado de asociación del ejemplo con las distintas clases.**
4. **Clasificación.** Para ello aplicaremos una función de decisión F sobre el grado de asociación del ejemplo con las clases que determinará, en base al criterio del máximo, la etiqueta de clase v a la que corresponda el mayor valor.

En el punto (3) es donde se distinguen los dos métodos mencionados.

2.3 Algoritmo de Wang y Mendel aplicado a problemas de clasificación

Para nuestra experimentación hemos utilizado la extensión del Algoritmo de Wang y Mendel [19] a problemas de clasificación realizada por Chi et al.[4]. Este método de diseño de SCBRDs determina las relaciones entre las variables del problema y establece una correspondencia entre el espacio de características y el de clases en un proceso que sigue esta serie de pasos:

1. *Establecimiento de las particiones lingüísticas.* Una vez determinado el dominio de variación de cada característica X_i , se determinan las particiones difusas.
2. *Generación de una regla difusa para cada ejemplo* $e^h = (e_1^h, \dots, e_N^h, C_h)$. Para ello es necesario:
 - 2.1 Calcular los grados de pertenencia del ejemplo e^h a las distintas regiones difusas.

- 2.2 Asignar el ejemplo e^h a la región difusa con mayor grado de pertenencia.
- 2.3 Generar una regla para el ejemplo, cuyo antecedente está determinado por la región difusa seleccionada y con la etiqueta de clase del ejemplo en el consecuente.
- 2.4 Calcular el grado de certeza. Para ello se determinará el cociente S_j/S , siendo S_j la suma del grado de pertenencia de los ejemplos de entrenamiento de la clase C_j a la región difusa determinada por el antecedente de la regla, y S la suma del grado de pertenencia a la misma región de todos los ejemplos independientemente de la clase a la que pertenezcan.

3 MEDIDAS DE COMPLEJIDAD

Los problemas de clasificación pueden ser difíciles por tres razones diferentes:

- Ciertos problemas son conocidos por tener un error de Bayes no nulo [7]. Algunas clases pueden ser ambiguas intrínsecamente o debido a medidas incorrectas de los atributos.
- Algunos problemas pueden presentar límites de decisión complejos, por lo que no se puede ofrecer una descripción compacta de los mismos [16].
- Muestras de tamaño reducido y la dispersión inducida por la alta dimensionalidad afecta a las reglas [13, 14].

El uso de medidas es reciente. En [8] Ho y Basu definen medidas de complejidad para conjuntos de datos de dos clases. Singh en [17] ofrece una revisión de medidas de complejidad y propone dos más. Sotoca et al. en [18] extienden algunas medidas de Ho y Basu para problemas multiclase. Para ello, analizan las medidas generalizadas en dos problemas clásicos de Selección de Prototipos.

En nuestro estudio, a partir de las medidas de Ho y Basu [8] hemos empleado la *razón discriminante de Fisher*. Se trata de una medida geométrica de solapamiento, cuya versión más sencilla indicada por Ho y Basu calcula cómo de separadas están dos clases de acuerdo a una característica en concreto. Compara la diferencia entre la media de las clases con la suma de la varianza de las clases. La razón discriminante de Fisher para un atributo se define como:

$$f = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (3)$$

donde μ_1 , μ_2 , σ_1 y σ_2 son las medias y varianzas de las dos clases respectivamente. Calculamos f para cada uno de los atributos y se toma el máximo como la medida F1. Valores pequeños indican que las clases tienen un alto grado de solapamiento. Las Figuras 1 a 4 muestran un ejemplo ilustrativo generado artificialmente con 2 variables en el rango $[0.0, 1.0]$ y dos clases como ejemplo.

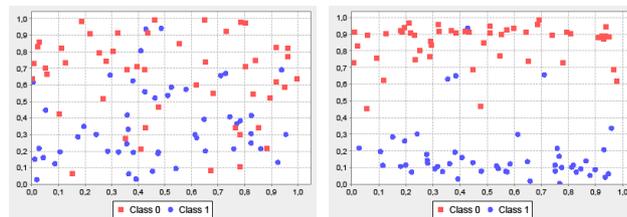


Figura 1: $F1 = 0.6994$

Figura 2: $F1 = 9.69$

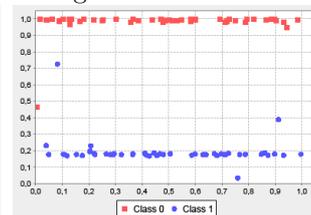
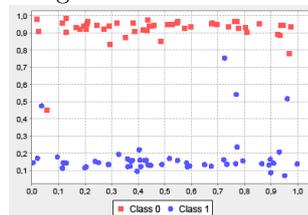


Figura 3: $F1 = 26.16$

Figura 4: $F1 = 48.65$

4 ESTUDIO EXPERIMENTAL

En esta sección describiremos los datos y metodología usadas en la experimentación de nuestro estudio. Se mostrarán los resultados obtenidos y se analizarán obteniendo conclusiones en base a los mismos para cada una de las medidas de complejidad consideradas.

4.1 Marco Experimental

Para comprobar el comportamiento del método de Chi con diferentes valores de la medida F1, hemos empleado una serie de conjuntos de datos. Hemos considerado tres tipos de conjuntos de datos: con 2, 4 y 8 atributos de entrada. Cada tipo consta de cien conjuntos de datos distintos, en los cuales se incrementa progresivamente el valor de la medida F1, para un total de 300 conjuntos de datos. Estos conjuntos de datos disponen de atributos reales de entrada en el rango $[0.0, 1.0]$, y un atributo nominal de salida con dos posibles etiquetas diferentes (clases), existiendo un total de 100 ejemplos por conjunto de datos. De estos 100 ejemplos, hay 50 ejemplos para cada clase. En nuestros conjuntos, la medida F1 está en el intervalo $[0.699, 74.1]$. Partimos de una distribución de ejemplos no totalmente ruidosa, cuyo valor F1 es 0.699. Conforme aumenta el valor de la medida de solapamiento,

las clases estarán identificadas por atributos menos solapados.

Para realizar un estudio comparativo vamos a utilizar un modelo de validación cruzada de orden 5 (5-fcv). Los métodos de clasificación se han ejecutado en cada uno de los 100 conjuntos de datos, de 2, 4 y 8 atributos, con este método de validación. La media de las tasas de acierto de las 5 particiones se toma como medida significativa del rendimiento del modelo.

Los modelos SCBRD empleados para medir y comparar su rendimiento son dos: un SCBRD Chi con 3 etiquetas por atributo y un SCBRD Chi con 5 etiquetas por atributo. Para los modelos de Chi emplearemos como *forma de combinación del grado de emparejamiento* y del *grado de certeza* la T-norma producto. El *tipo de inferencia* será el método clásico de la regla ganadora.

Asimismo, y con el fin de comparar con otro algoritmo clásico en el ámbito de clasificación de diferente naturaleza, hemos realizado el mismo estudio con los mismos datos y particiones para el árbol de decisión C4.5 [15]. Esto permitirá comprobar si los SCBRDs tienen comportamientos diferenciados con otras familias de clasificadores.

4.2 Resultados para F1: Razón discriminante de Fisher

El porcentaje de acierto en clasificación tanto en entrenamiento como en test para la medida F1 y cada uno de los tipos de conjuntos (con 2,4 y 8 atributos) se encuentra reflejado en las Figuras 5 a 7.

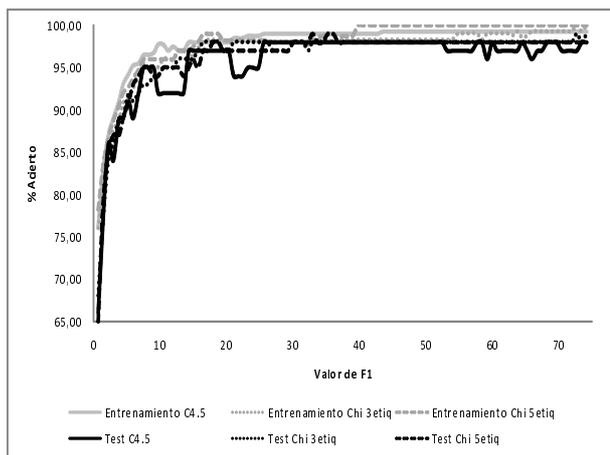


Figura 5: Porcentaje de acierto para F1 con 2 atr.

Para valores bajos de la medida F1 (alto solapamiento), tanto C4.5 como Chi con 3 etiquetas tienen un porcentaje de acierto bajo en entrenamiento, mien-

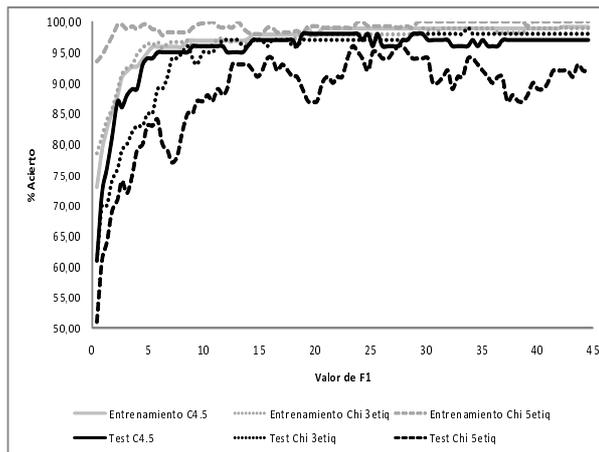


Figura 6: Porcentaje de acierto para F1 con 4 atr.

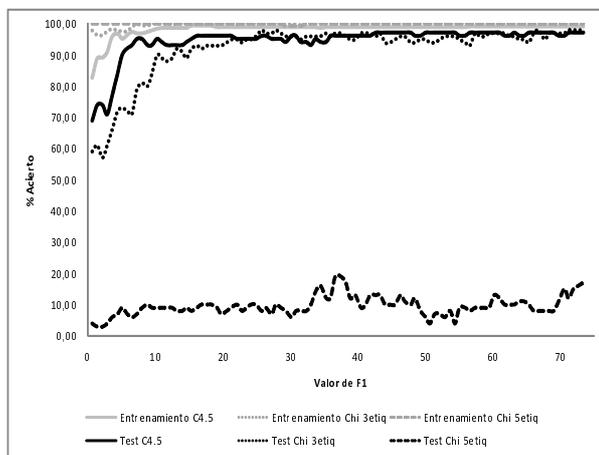


Figura 7: Porcentaje de acierto para F1 con 8 atr.

tras que Chi con 5 etiquetas obtiene mejores resultados, con motivo del sobre-aprendizaje que se produce, por el alto número de reglas. Conforme el valor de F1 aumenta los tres métodos son capaces de mejorar su porcentaje de acierto en entrenamiento. En test para F1 queda patente que el valor en test mejora cuando aumenta la separabilidad. Para valores bajos de la medida el porcentaje de acierto en test es bajo.

En las Figuras 8 a 10 se ha dispuesto la cantidad de nodos de C4.5 y de reglas de los métodos de Chi para cada conjunto de datos (ordenados según F1 de forma ascendente) según el número de atributos. Mientras que C4.5 disminuye el número de reglas al aumentar el valor de F1, los métodos de Chi mantienen un número alto de reglas para cubrir los ejemplos separados. Este alto número de reglas para Chi con 5 etiquetas es otro indicativo de su sobre-aprendizaje, sobre todo en el caso de 8 atributos donde siempre tiene 80 reglas (una

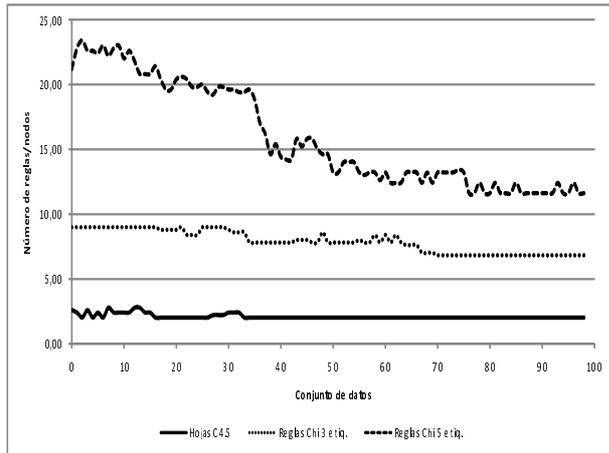


Figura 8: Evolución de reglas y nodos para 2 atributos

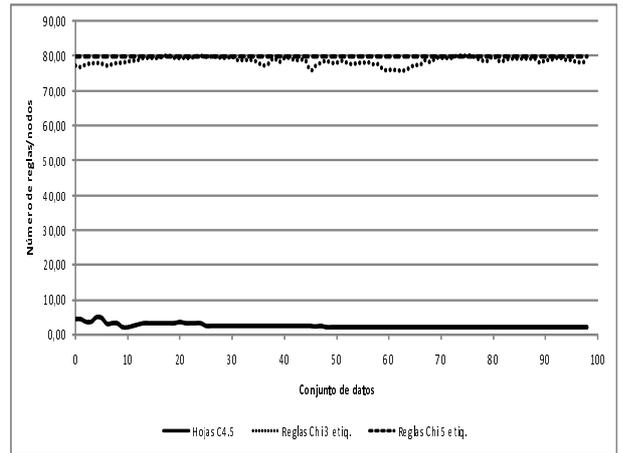


Figura 10: Evolución de reglas y nodos para 8 atributos

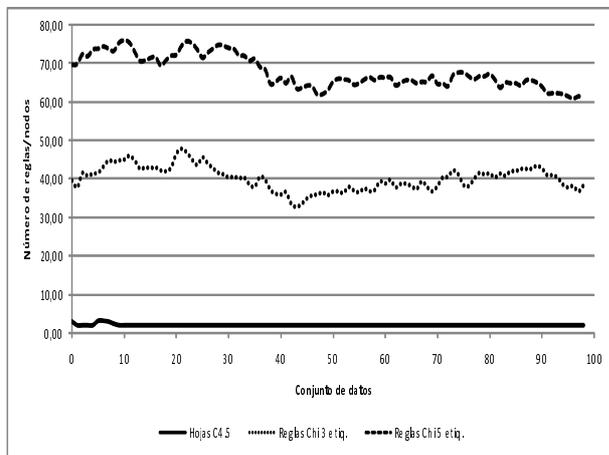


Figura 9: Evolución de reglas y nodos para 4 atributos

regla por ejemplo de entrenamiento). Chi con 3 etiquetas aumenta el número de reglas conforme aumenta también el número de atributos. C4.5 no obstante, obtiene siempre un número bajo de reglas sin importar el número de atributos considerado.

El uso de métodos de selección de características podría evitar el aumento del número de reglas (sobre todo en el caso de un alto número de atributos), puesto que C4.5 es capaz de seleccionar los atributos interesantes en su entrenamiento y es capaz de simplificar el número de reglas que produce.

4.3 Análisis de selección de características para F1: Razón discriminante de Fisher

Hemos analizado todas las combinaciones posibles de atributos a la hora de entrenar y evaluar los modelos Chi de 3 y 5 etiquetas para la medida F1 con la

variante de 8 atributos. La mejor combinación se establece en base a dos criterios posibles: la que mejor precisión en entrenamiento tenga o la que mejor acierto en test ofrezca. De esta manera comprobamos el criterio del SCBRD de Chi para elegir la combinación de atributos que considera óptima con el mejor valor de entrenamiento (notada como [Mejor Entren.]), frente a la que realmente sería óptima (notada como [Mejor Test]) que obtiene mejor valor en test. Los resultados se encuentran en la Figura 11 para Chi con 3 etiquetas y en la Figura 12 para Chi con 5 etiquetas.

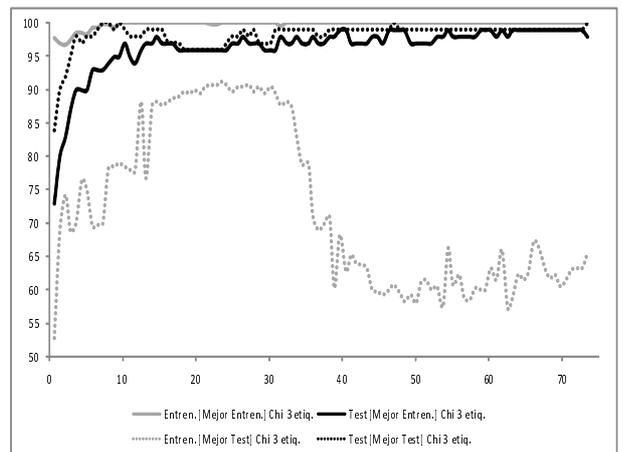


Figura 11: Porcentaje de acierto de Chi 3 etiq. para F1 con la mejor combinación de atributos

La combinación obtenida por el mejor entrenamiento ofrece sobre-aprendizaje para Chi con 5 etiquetas para este número de atributos, pero no en Chi con 3 etiquetas. El sobre-aprendizaje de Chi con 5 etiquetas se ve paliado con la mejor combinación de atributos para

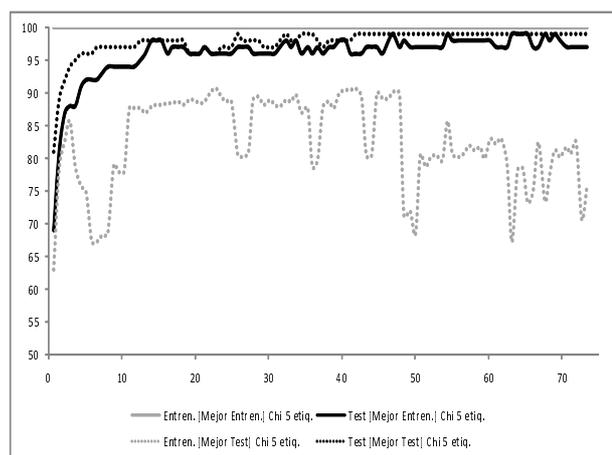


Figura 12: Porcentaje de acierto de Chi 5 etiq. para F1 con la mejor combinación de atributos

test, donde el valor de entrenamiento es similar al de Chi con 3 etiquetas para esta mejor combinación de atributos, y siempre por debajo del propio valor de test.

La combinación para el mejor entrenamiento produce valores de acierto en entrenamiento más altos, y siempre superiores a la combinación óptima para test. El porcentaje de acierto en test para ambos criterios es similar para ambos métodos, aunque el de la combinación óptima para test siempre es ligeramente mejor. Las razones de esta diferencia en entrenamiento podemos encontrarlas en la Figura 13, donde se muestra el número de medio de atributos para el cual cada estrategia (mejor entrenamiento o mejor test) ha sido óptima para cada uno de los 100 conjuntos de datos.

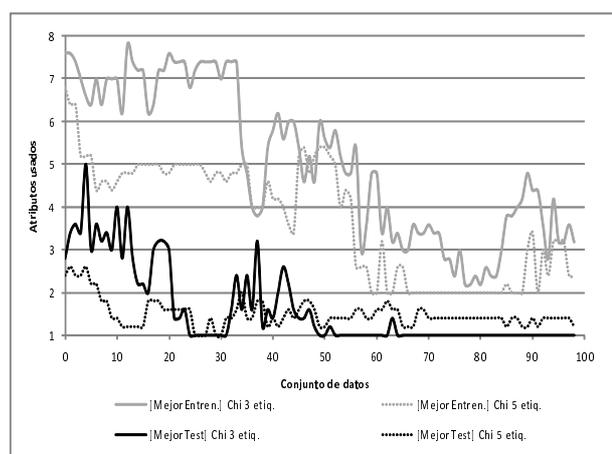


Figura 13: Media de atributos usados para F1 con 8 atr. con cada estrategia

Chi con 3 etiquetas emplea un mayor número de atributos que Chi con 5 etiquetas. Sin embargo, con la estrategia que obtendría el mejor valor de test se aprecia el hecho de que Chi con 3 etiquetas necesita menos atributos que Chi con 5 etiquetas (llegando incluso a usar sólo un atributo). Para obtener el mejor valor de test siempre se emplea un menor número de etiquetas que para obtener el mejor porcentaje de acierto en entrenamiento. Además, observamos que se disminuye la necesidad de utilizar todas las variables cuando nos alejamos de los valores bajos de la función F1 para ambas estrategias, esto permite evitar el sobre-aprendizaje y reafirma la necesidad de utilizar algoritmos de selección de características dentro del proceso de aprendizaje. Puede ser interesante el uso de métodos de selección de características basados en técnicas de filtro. No obstante, emplear la precisión del SCBRD Chi en entrenamiento para obtener la combinación óptima no es adecuado, y se plantea la necesidad de usar otros criterios.

5 CONCLUSIONES

En este trabajo se realiza un primer estudio sobre la influencia de la medida de complejidad llamada razón discriminante de Fisher para analizar el comportamiento de los SCBRDs. Observamos que el solapamiento de las clases (valor de F1 bajo) produce sobre-aprendizaje si utilizamos muchas etiquetas por variables (en este estudio 5 frente a 3) y que se mejora el comportamiento cuando aumenta el valor de F1 y por tanto la separabilidad de las clases. Esta primera experimentación también muestra que cuando aumenta la separabilidad puede ser conveniente el uso de técnicas de selección de características, y que para evitar el sobre-aprendizaje puede ser conveniente que estas técnicas sean de filtro.

Hemos de destacar que éste es un estudio muy preliminar donde hemos trabajado con bases de datos artificiales con un número de variables fijo en 5, un número fijo de instancias y cuatro clases. Este estudio es necesario extenderlo a bases de datos con diferente número de variables, instancias y clases, para poder obtener más información sobre la utilidad de la medida F1.

Como trabajo a realizar para mejorar este estudio, estamos abordando los siguientes estudios:

- Analizar el uso de técnicas de filtro, como los métodos LVF [12] y MIFS [2].
- Analizar el comportamiento de varios métodos de aprendizaje de reglas difusas, en particular aquellos que incluyan la selección de características dentro de las propias reglas, como el método propuesto en [10].

- c) Estudiar posibles extensiones de la medida F1 atendiendo a otros parámetros como son el número de instancias, variables y clases que intervienen en el problema.

Finalmente hemos de indicar que extenderemos el presente estudio a otras medidas de complejidad presentes en la literatura especializada.

Agradecimientos

Este trabajo de investigación ha sido posible gracias a la subvención del proyecto TIN2005- 08386-C05-01.

Referencias

- [1] M. Basu and T.K. Ho. Data complexity in pattern recognition., *Springer*, 2006.
- [2] R. Battiti. Using Mutual Information For Selection Features In Supervised Neural Net Learning. *IEEE Transactions on Neural Networks*, 5:4, Pág.537-550, 1994.
- [3] E. Bernardó-Mansilla and T.K. Ho. Domain of competence of XCS classifier system in complexity measurement space. *IEEE Transactions on Evolutionary Computation*, 9:1, Pág.82-104, 2005.
- [4] Z.Chi, H.Yan, and T.Pham. Fuzzy algorithms with applications to image processing and pattern recognition. *World Scientific*, 1996.
- [5] O.Cordón, M.J. del Jesus, and F.Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 20:1, Pág.21-45, 1999.
- [6] M. Dong and R. Kothari. Feature subset selection using a new definition of classifiability. *Pattern Recognition Letters*, 24:9-10, Pág.1215-1225, 2003.
- [7] T.K. Ho and H.S. Baird. Large-scale simulation studies in image pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:10, Pág.1067-1079, 1997.
- [8] T.K. Ho and M. Basu. Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:3, Pág.289-300, 2002.
- [9] H. Ishibuchi, T. Nakashima and M. Nii. Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining. *Springer-Verlag*, 2004.
- [10] H. Ishibuchi, T. Yamamoto and T. Nakashima. Hybridization of Fuzzy GBML Approaches for Pattern Classification Problems. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 35:2, Pág.359-365, 2005.
- [11] Y.-H. Li, M. Dong and R. Kothari. Classifiability-based omnivariate decision trees. *IEEE Transactions on Neural Networks*, 16:6, Pág.1547-1560, 2005.
- [12] H. Liu, R. Setiono. Incremental Feature Selection. *Applied Intelligence*, 9:3, Pág.217-230, 1998.
- [13] M. Liwicki and H. Bunke. Handwriting recognition of whiteboard notes - studying the influence of training set size and type. *International Journal of Pattern Recognition and Artificial Intelligence*, 21:1, Pág.83-98, 2007.
- [14] X. Qiu and L. Wu. Nearest neighbour discriminant analysis. *International Journal of Pattern Recognition and Artificial Intelligence*, 20:8, Pág.1245-1259, 2006.
- [15] J.R. Quinlan. C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, 1993.
- [16] J.S. Sánchez, R.A. Mollineda and J.M. Sotoca. An analysis of how training data complexity affects the nearest neighbours classifiers. *Pattern Analysis & Applications*, 10, Pág.189-201, 2007.
- [17] S. Singh. Multiresolution estimates of classification complexity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:12, Pág.1534-1539, 2003.
- [18] J.M. Sotoca, R.A. Mollineda and J.S. Sánchez. A meta-learning framework for pattern classification. *Revista Iberoamericana de Inteligencia Artificial*, 29, Pág.31-38, 2006.
- [19] L.X. Wang and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 25:2, Pág.353-361, 1992.