

Funciones Núcleo en un Espacio de Órdenes de Magnitud Absolutos

¹F.J. Ruiz ¹C. Angulo, ²N. Agell, ²X. Rovira R. ²Campos, ³M. Sanchez

¹Depto. de Ingeniería de Sistemas. Universitat Politècnica de Catalunya
Av. Víctor Balaguer s/n. Vilanova i G. (Spain) fjruiz@mat.upc.es, cecilio.angulo@upc.es

²Depto. de Métodos Cuantitativos. ESADE- Universitat Ramon Llull
Av. Pedralbes 62-65. 08034 Barcelona (Spain) {agell,rovira,r.campos.e}@esade.edu

³Depto. de Matemática Aplicada II. Universitat Politècnica de Catalunya
C. Pau Gargallo, 5, 08028 Barcelona (Spain) monica.sanchez@upc.es

Resumen

Los algoritmos de aprendizaje basados en Funciones Núcleo, particularmente las Máquinas de Soporte Vectorial (MSV), han proporcionado buenos resultados en problemas de clasificación con patrones de entrada no separables linealmente. El uso de las Funciones Núcleo permite aplicar estos algoritmos de inferencia incluso sobre información proveniente de un conjunto sin estructura de espacio euclídeo. Al considerar una Función Núcleo, los datos se proyectan de forma implícita sobre un nuevo espacio de características cuya estructura es exportada hacia el espacio de origen.

En este trabajo se analiza una Función Núcleo que actúa sobre datos que pertenecen a un Espacio Cualitativo de Órdenes de Magnitud Absolutos. El diseño de esta Función Núcleo está inspirado en recientes métodos elaborados sobre Máquinas Núcleo para espacios discretos de trabajo. Como ilustración se presenta una aplicación de estos sistemas de aprendizaje en el campo financiero, concretamente en la modelización del riesgo de crédito. Se estudia los resultados de predicción de riesgo crediticio de un conjunto de empresas que entregan información pública al mercado. Para ello se utilizan variables económico-financieras de las compañías y su clasificación de riesgo emitida por una conocida evaluadora del mercado financiero.

Palabras clave: Aprendizaje Automático, Máquinas de Soporte Vectorial, Funciones Núcleo, Razonamiento Cualitativo.

1. Introducción

El diseño de algoritmos que sean capaces de captar de forma 'automática' la información relevante de un conjunto de patrones es uno de los objetivos principales de la Inteligencia Artificial (IA). Los Sistemas de Aprendizaje Au-

tomático, entre los que se encuentran las Redes Neuronales Artificiales (RNA) y, en concreto, las Máquinas de Soporte Vectorial (MSV), se han utilizado con éxito en problemas donde intervienen datos cuantitativos precisos. Para poder ampliar el rango de aplicación de estos sistemas, sería útil poder adaptarlos para que puedan trabajar con datos que presenten dis-

tintos niveles de precisión. La Lógica Borrosa, el Análisis Intervalar y el Razonamiento Cualitativo son dominios de la IA que tratan con este tipo de información desde distintos puntos de vista. Es el objetivo de este trabajo desarrollar algoritmos para sistemas de aprendizaje en el caso de tratar información con distintos niveles de precisión utilizando el enfoque del Razonamiento Cualitativo. Concretamente se propondrá y analizará una Función Núcleo definida en un Espacio de Órdenes de Magnitud Absolutos. Esta función, inspirada en la propuesta en (Cristianini y Shawe-Taylor 2000) para cadenas de caracteres, permitirá aplicar los algoritmos basados en este tipo de funciones, especialmente las MSV, a datos expresados por sus órdenes de magnitud.

Se presenta una aplicación de esta Función Núcleo para la medición del riesgo financiero de crédito. Se utilizan los datos de variables económico-financieras de compañías y su clasificación de riesgo (*rating*) emitida por *Standard & Poor's*. Los motivos principales que nos conducen a utilizar este tipo de aplicación son tres: Por una parte, el proceso de clasificación que siguen los evaluadores es un proceso complejo, donde concurren elementos cuantitativos, cualitativos y el juicio de los expertos financieros. En segundo lugar, diversos métodos numéricos, entre ellos, econométricos y redes neuronales, han sido sugeridos en la literatura para la resolución del problema sin que hayan permitido llegar a una clasificación satisfactoria. Por último, la capacidad de multclasificación de algunas MSV en comparación con los resultados dicotómicos obtenidos en clasificaciones ya utilizadas en relación a los *ratings* (Angulo (2001), Li e et al.(2002), Pang et al.(2002) y Fan y Palaniswami(2000)).

En la sección 2 se presenta una breve introducción a las MSV. En dicha introducción se pondrá especialmente de manifiesto el potencial de este tipo de algoritmos para tratar con cualquier tipo de datos. En la sección 3 se presentan, también de forma abreviada, los conceptos básicos extraídos de la teoría de Espacios Cualitativos de Órdenes de Magnitud Absolutos necesarios para la construcción de la Función Núcleo propuesta. Su construcción y estudio se discute en la sección 4, mientras que en la sección 5 se utiliza la metodología propuesta para analizar el riesgo crediticio de un conjunto de empresas en función de una serie de datos

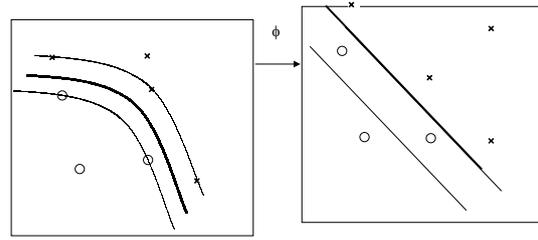


Figura 1: Interpretación del espacio de características.

económicos. Finalmente se realiza una discusión de resultados a modo de conclusión y se muestran líneas futuras de investigación.

2. Máquinas de Soporte Vectorial y Funciones Núcleo

En esta sección se hace una breve revisión de los conceptos básicos de las MSV y las Funciones Núcleo. Las MSV son algoritmos de aprendizaje capaces de clasificar un conjunto de patrones de entrada usando un proceso de inferencia de forma supervisada. En su definición original, para que la máquina de aprendizaje sea aplicable se requiere que los patrones pertenezcan a dos clases que sean linealmente separables. En concreto, el algoritmo busca el hiperplano $f(x) = \langle a \cdot x \rangle + b$ que mejor separa los patrones de entrenamiento, en el sentido de maximizar la distancia entre los patrones más cercanos de ambas clases. Este caso base, con los patrones de entrenamiento linealmente separables, es poco realista, y para poder aplicar esta técnica en el caso no separable, se introducen las Funciones Núcleo.

De forma implícita, una Función Núcleo provoca que los datos de entrada se proyecten sobre un espacio, normalmente de mayor dimensión (incluso infinita), donde los datos proyectados ya sean linealmente separables. Este espacio recibe el nombre de espacio de características. En la figura 1 se representa esquemáticamente este proceso para un caso de dimensión muy reducida.

Una ventaja de las Funciones Núcleo es que definen esa proyección de forma implícita. Esto

es así porque en la expresión del algoritmo de una MSV aparecen únicamente los productos escalares de todos los pares de patrones de entrenamiento y no ellos por separado¹. Por lo tanto, no es necesario utilizar explícitamente las proyecciones de cada patrón, sino únicamente los productos escalares de las proyecciones. Esto simplifica enormemente la implementación de la MSV y evita el problema de la dimensionalidad, característico de los algoritmos de aprendizaje cuando se consideran espacios de alta dimensión.

De hecho, toda posible aplicación f del conjunto inicial a un espacio euclídeo de características proporciona una única Función Núcleo que viene dada por la expresión $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, donde $\langle \cdot, \cdot \rangle$ indica un producto escalar en el espacio de características². Por otra parte, no cualquier función K es adecuada para ser una Función Núcleo. Las Funciones Núcleo vienen caracterizadas por el Teorema de Mercer (Vapnik 1995), (Cristianini y Shawe-Taylor 2000).

A pesar de que las Funciones Núcleo no necesitan trabajar de forma explícita con el espacio de características, también es posible tratar el problema a la inversa, es decir, se puede construir de forma explícita una función f para, a partir de ella, construir la Función Núcleo. De esta forma, no es necesario comprobar las condiciones de Mercer. Este es precisamente el método que se aplica en este trabajo.

Otra ventaja importante de las Funciones Núcleo es que no necesitan ser definidas sobre un espacio con estructura euclídea, ya que sólo serán utilizados los productos escalares en el espacio de características. Esta ventaja ya está siendo utilizada en diversas áreas de investigación para trabajar con patrones de distintos tipos: palabras, figuras, información genética, cadenas de proteínas, etc. De esta forma, no únicamente se utiliza la Función Núcleo para convertir en separables datos que no lo son de origen, sino que estas funciones permiten dotar de estructura euclídea a un conjunto que no la tenía.

¹El uso de una Función Núcleo K sólo implica, a nivel de formulación, sustituir los productos escalares $\langle x_i, x_j \rangle$ que aparecen en la algoritmia de la MSV por $K(x_i, x_j)$.

²La implicación inversa, sin embargo, no es cierta; es decir, una Función Núcleo no determina de forma unívoca una aplicación f .

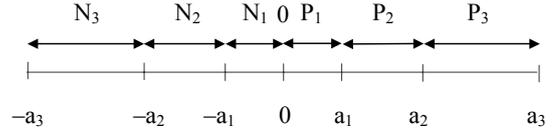


Figura 2: Partición simétrica en 7 clases.

En este trabajo se aplica esta metodología en el caso en que los patrones vengan dados por los órdenes de magnitud de sus descriptores.

3. El modelo de órdenes de magnitud absolutos

En esta sección se describe el modelo de Órdenes de Magnitud Absolutos descrito en (Agell (1998), Travé-Massuyès (1997)).

Uno de los objetivos del Razonamiento Cualitativo es tratar los problemas de forma que se conserve el principio de relevancia (Forbus1988), o sea que cada variable aparezca con el nivel de precisión requerido, para ello se introduce el modelo completo de Órdenes de Magnitud Absolutos de granularidad n , $OM(n)$. Este modelo se define a partir de una partición simétrica de la recta real en $2n + 1$ clases creada a partir de los números reales: $-a_{n-1}, \dots, -a_1, 0, a_1, \dots, a_{n-1}$. Cada clase recibe el nombre de descripción básica o elemento básico y se representa por una etiqueta del conjunto S_1 :

$$S_1 = \{N_n, \dots, N_1, 0, P_1, \dots, P_n\} \quad (1)$$

donde:

$$N_n =]-\infty, -a_{n-1}[, N_i = [-a_i, -a_{i-1}[\quad [i=2 \dots n-1,$$

$$N_1 = [-a_1, 0[, 0 = \{0\} , P_1 =]0, a_1],$$

$$P_i =]a_{i-1}, a_i] \quad [i=2 \dots n-1 , P_n =]a_{n-1}, \infty[$$

En algunos casos se considera únicamente intervalos acotados, imponiendo que todos los valores están en el intervalo $[-a_n, a_n]$. La figura 2 representa esta partición en el caso $n = 3$.

Este conjunto se amplía con todos los posibles subconjuntos conexos de la recta real que se

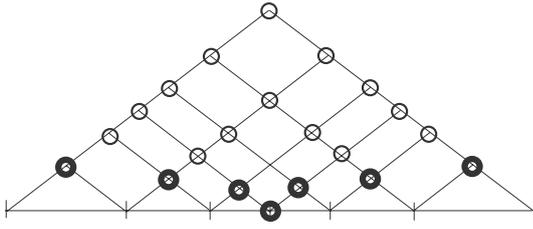


Figura 3: Representación de los elementos de S .

pueden formar a partir de los elementos básicos. De esta forma se obtiene el espacio de cantidades S formado por todas las etiquetas de la forma $I = [X, Y]; \forall X, Y \in S_1$ con $X < Y$ i.e. $x < y \forall x \in X, \forall y \in Y$.

Una representación útil del espacio de cantidades S consiste en identificar las etiquetas cualitativas por medio de un centro (C) y un radio (R). Cada etiqueta se representa como un punto del diagrama $C - R$. La abcisa de ese punto se corresponde con el punto medio del intervalo que representa y la ordenada con el radio de dicho intervalo. Todas las etiquetas posibles están inscritas en el triángulo isósceles de base $[-a_n, a_n]$ y de altura $R = a_n$. En la figura 3 se muestra esta representación en el caso $n = 3$. En esa figura se ha señalado con trazo más grueso los elementos básicos.

En este conjunto S se define la relación ser más preciso que (\leq_P) de la siguiente forma:

Definición: *Dados $X, Y \in S$, X es más preciso que Y , ($X \leq_P Y$) si $X \subseteq Y$.*

A partir de esta relación, puede definirse el concepto de base de una etiqueta cualitativa o elemento de S :

Definición: $\forall X \in S - \{0\}$, la base de X es el conjunto $B_X = \{B \in S_1 - \{0\} : B \leq_P X\}$

Es decir, B_X está formado por todas las etiquetas básicas no nulas contenidas en X .

En la figura 4 se observa la representación de la base de un elemento de S en el diagrama $C - R$ para la etiqueta cualitativa $X = [N_2, P_1]$. La base está formada por los descriptores básicos contenidos en el triángulo sombreado: $B_X = \{N_2, N_1, P_1\}$.

Por otra parte, la igualdad cualitativa o q -

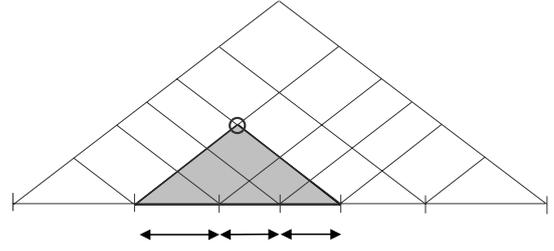


Figura 4: Representación de la base de un elemento de S .

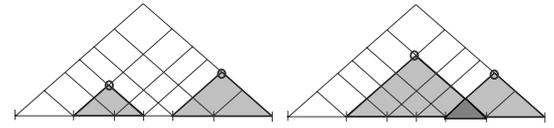


Figura 5: Representación de la q -igualdad.

igualdad se define como:

Definición: $X, Y \in S$ son q -iguales, ($X \approx Y$), si $X \cap Y \neq \emptyset$

La igualdad cualitativa refleja, de hecho, la posibilidad de que las etiquetas X e Y representen el mismo valor. En la figura 5 se muestra la interpretación sobre el diagrama $C - R$ del concepto de q -igualdad. En la parte izquierda se muestran dos etiquetas q -diferentes, mientras que en la parte de la derecha se muestran dos etiquetas q -iguales.

El par (S, \approx) se llama Espacio Cualitativo de Órdenes de Magnitud Absolutos de granularidad n . También nos referiremos a él como $OM(n)$.

Por último, con el propósito de transformar valores numéricos en etiquetas cualitativas, se define el concepto de expresión cualitativa de un subconjunto de la recta real.

Definición: Si $A \subset R$, la expresión cualitativa de A , que notaremos $[A]$, es el elemento de S más preciso que contiene a A .

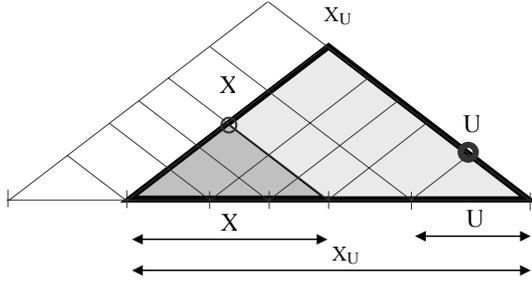


Figura 6: Representación de la U-expansión.

4. Construcción de una Función Núcleo en un espacio OM(n)

Tomando como base la metodología expuesta en (Cristianini y Shower-Taylor 2000) para la obtención de una Función Núcleo sobre un espacio discreto, se define una aplicación de S en $[0, 1]^{2n}$ que permitirá la creación de una Función Núcleo en S . Para ello se introducen los conceptos de expansión y alejamiento respecto de una etiqueta básica.

Definición: Dado $U \in S_1$, la U-expansión es la aplicación de S en S tal que si $X \in S$, entonces $\Psi_U(X) \equiv X_U = [X \cup U]$.

Definición: El alejamiento respecto a U es la aplicación de S en N tal que si $X \in S$, entonces $a_U(X) = \text{Card}(B_{X_U}) - \text{Card}(B_X)$, es decir, representa el número de elementos básicos en la U-expansión, sin incluir a X

En otras palabras, $a_U(X)$ es el número de etiquetas básicas no nulas que se han de añadir a la etiqueta X para que sea cualitativamente igual a la etiqueta básica U .

En la figura 6 se muestra la interpretación de la U-expansión en el caso $X = [N_2, P_1], U = P_3, X_U = [N_2, P_3]$. En ese caso se tiene que $a_U = 2$.

El concepto de alejamiento permite definir la aplicación ϕ de S al espacio de características $F \subseteq [0, 1]^{2n}$ de la siguiente forma:

Definición: Dado $X \in S_1 - \{0\}$, y $\lambda \in [0, 1]$, $\phi(X) = (\lambda^{a_{N_n}(X)}, \dots, \lambda^{a_{P_n}(X)})$.

Esta aplicación refleja el posicionamiento global de una etiqueta cualitativa respecto a todos los descriptores básicos. El factor λ comportará que un aumento del alejamiento entre etiquetas provoque una disminución de la componente correspondiente.

La función ϕ así definida permite construir una Función Núcleo en el espacio S de la forma: $K(X, Y) = \langle \phi(X), \phi(Y) \rangle$ donde $\langle \cdot, \cdot \rangle$ es el producto usual de R^{2n} .

Esta construcción es directamente generalizable a un espacio cualitativo k -dimensional $[OM(n)]^k$ para utilizarse en el caso en que los patrones estén dados por k descriptores. En tal caso, la función ϕ será una función definida del conjunto S^k a $[0, 1]^{2nk}$. Si $X = (X_1, \dots, X_k) \in S^k$, la función ϕ es $\phi(X) = (\phi(X_1), \dots, \phi(X_k))$

La inyectividad de la función ϕ permite considerar una distancia en S^k de manera que dados $X, Y \in S^k$:

$$\begin{aligned} d(X, Y) &= \sqrt{K(X, X) + K(Y, Y) - 2K(X, Y)} \\ &= \sqrt{\langle \phi(Y) - \phi(X), \phi(Y) - \phi(X) \rangle} \end{aligned} \quad (2)$$

Una vez construida la función ϕ , no únicamente es posible considerar la Función Núcleo $K(\cdot, \cdot)$ definida anteriormente, sino que tal y como se demuestra en Sánchez et al (2003), cualquier función $K' = G \circ (\phi \times \phi)$ es una Función Núcleo sobre S^k , siendo G una función núcleo cualquiera. Esta propiedad permite aplicar todo el conjunto de Funciones Núcleos (polinómicas, gaussianas, etc.) que se utilizan de forma habitual con MSV.

A modo de ilustración, tanto de la función ϕ como de la Función Núcleo correspondiente, se realiza el cálculo en el caso $k = 1$ para las etiquetas $[N_2, N_1]$ y $[N_1, P_1]$.

$$\begin{aligned} \phi([N_2, N_1]) &= (\lambda, 1, 1, \lambda, \lambda^2, \lambda^3) \\ \phi([N_1, P_1]) &= (\lambda^2, \lambda, 1, 1, \lambda, \lambda^2) \end{aligned} \quad (3)$$

$$\begin{aligned} K([N_2, N_1], [N_1, P_1]) &= \\ \lambda^3 + \lambda + 1 + \lambda + \lambda^3 + \lambda^5 \end{aligned} \quad (4)$$

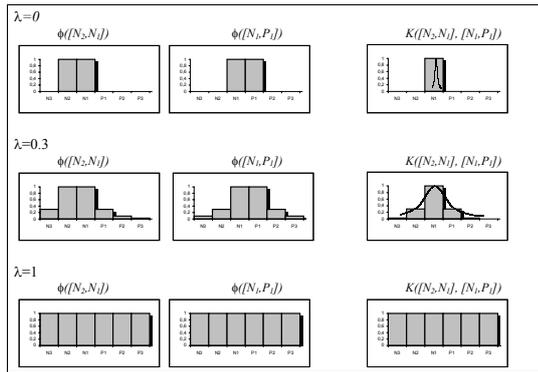


Figura 7: Interpretación del parámetro λ en la función ϕ y en la función núcleo K .

En la figura 7 se ilustran los resultados de este cálculo para tres valores distintos de λ . En cada gráfico se representa en forma de columna cada una de las componentes de la función ϕ . Para $\lambda \in (0, 1)$, las componentes respecto a los descriptores básicos que forman la base de la etiqueta cualitativa son iguales a 1 y las componentes respecto a descriptores básicos que no pertenecen a la base de la etiqueta cualitativa son menores cuanto más alejadas están de dicha etiqueta cualitativa.

En la misma figura también se muestra la contribución de cada una de las componentes en el valor de $K(\cdot)$. Si los rectángulos fuesen considerados de base 1, el valor de $K(\cdot)$ se obtendría como el área total de las barras de dicha figura. Se observa claramente que en el caso límite $\lambda = 0$, únicamente las etiquetas básicas comunes afectan al cálculo de $K(\cdot)$. El incremento del parámetro λ provoca que las etiquetas básicas no comunes pero cercanas a las comunes también contribuyan en el cálculo de $K(\cdot)$. Un valor de λ cercano a 1 implica disminuir la discriminación entre etiquetas. En el caso límite $\lambda = 1$, se tiene $K(X, Y) = 2n \forall X, Y$ y, por tanto $d(X, Y) = 0 \forall X, Y$.

A partir de esta interpretación, se analiza el caso límite, pero también más sencillo de $\lambda = 0$. En tal caso, el cálculo de K es sumamente simple: $K(X, X)$ es igual al número de descriptores básicos de X y $K(X, Y)$ es el número de descriptores básicos comunes de X e Y . Por este motivo, $K(X, Y)$ está relacionado con la q-igualdad de X e Y :

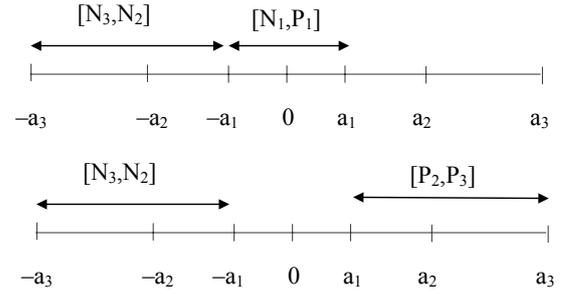


Figura 8: Representación de dos pares de etiquetas cuya distancia es 0 en el caso $\lambda = 0$.

$$\lambda = 0 \Rightarrow \{X \approx Y \Leftrightarrow K(X, Y) \neq 0\}. \quad (5)$$

Si $X \approx Y$, entonces $d^2(X, Y) = K(X, X) + K(Y, Y) - 2 \cdot K(X, Y)$, que coincide con el número de descriptores básicos de X o de Y no comunes. En caso contrario, es decir, si $X \not\approx Y$, entonces $d^2(X, Y) = K(X, X) + K(Y, Y)$ que es el número de descriptores básicos de X más el número de descriptores básicos de Y .

Se comprueba que en este caso límite la computación de K es muy simple, pero esta simplicidad se consigue a costa de una pérdida de información. Como ilustración, se puede observar que la distancia entre las etiquetas $[N_3, N_2]$ y $[N_1, P_1]$ coincide con la distancia entre las etiquetas $[N_3, N_2]$ y $[P_2, P_3]$, lo cual no parece muy intuitivo, según puede observarse en la figura 8.

Una forma de resolver este problema sin aumentar sobremanera el tiempo de computación es considerar un valor de λ distinto de 0, pero suficientemente pequeño como para poder despreciar los términos de orden superior o igual a λ^2 .

5. Una aplicación a la predicción del riesgo de crédito

Este trabajo se enmarca dentro de un proyecto más amplio en el que se pretende analizar y desarrollar técnicas innovadoras de *soft-computing* con integración de conocimiento experto para la medición del riesgo financiero de

AAA	Rating más alto asignado. Máxima capacidad de pago de intereses y del principal de la deuda.
AA	Fuerte capacidad de pago de intereses y del principal, aunque no tienen tantas garantías como los de la categoría anterior.
A	Alta capacidad de pago de intereses y del principal, aunque pueden verse negativamente afectados por un cambio en las circunstancias y condiciones económicas en mayor grado que en los casos anteriores.
BBB	Capacidad adecuada de pago de intereses y principal. En caso de un cambio de circunstancias, es probable que experimenten un mayor deterioro que en las categorías anteriores.
BB	Baja vulnerabilidad en el corto plazo, pero mayor riesgo en condiciones adversas que puedan llevar a reducir su capacidad de llevar a cabo el pago de los intereses y la devolución del principal.
B	Mayor vulnerabilidad de impago, pero en el corto plazo tienen la capacidad de pagar intereses y principal. Si se producen situaciones adversas, es probable que se debilite la capacidad y/o el deseo de pagar los intereses y el principal de la deuda.
CCC	Riesgo de impago en el corto plazo, y su seguridad depende de que se produzcan favorables circunstancias económicas y financieras. Si se producen situaciones adversas, no es probable que puedan afrontar el pago de intereses y principal.
CC	Deuda con un alto grado de especulación. En condiciones adversas, se espera que la entidad emisora tenga serios problemas de impago.
C	Deuda que en el momento tienen pendiente pago de intereses. Las perspectivas de pago futuro de interés y principal son pobres.
D	Deuda que tienen pendientes pago de intereses o devolución del principal en el momento actual.
+ o -	Para los ratings entre AA y CCC, el símbolo + o - se añade para reflejar la posición relativa dentro de la categoría

Figura 9: *Ratings* de deuda a largo plazo otorgados por S & P

crédito (proyecto MERITO). En la presente sección presentamos un primer ejemplo sencillo con datos reales que nos permitirá analizar y poner de manifiesto el potencial de la herramienta desarrollada en la sección anterior.

5.1. Presentación del problema

El *rating* es una valoración cualificada del nivel de riesgo de emisiones y empresas. Hay agencias especializadas en efectuar esta valoración, las más importantes son *Moody's* y *Standard & Poor's*. Estas agencias clasifican a las empresas según su nivel de riesgo, utilizando información cuantitativa y cualitativa. Los *ratings* de *Standard & Poors's*, que son los que se utilizan en este trabajo, están etiquetados por AAA, AA, A, BBB, BB, B, CCC, CC y D. De izquierda a derecha estos valores van de *muy alta* a *muy baja* calidad de crédito, es decir, de mucha a poca capacidad de la empresa para afrontar la deuda. En concreto se lista en la figura 9 la interpretación de cada una de las etiquetas en referencia a la deuda a largo plazo, a la que corresponden los datos usados en este ejemplo.

Los procesos usados por las agencias de *rating* son complejos. Las técnicas de decisión involucradas no están basadas en modelos puramente numéricos. Por un lado, utilizan la información dada por los datos financieros y los valores que les influyen, por otro, analizan el sector y el país o países donde opera la empresa, y tienen en

cuenta las posibilidades de crecimiento del negocio y de su posición competitiva. Finalmente, los analistas, para determinar el *rating* hacen una evaluación global basada en su propia experiencia.

En relación con la medición del riesgo de crédito, cabe citar los trabajos de Altman (1968) que proporcionaron en su momento un gran avance en este proceso. En esos trabajos se utilizaron funciones lineales discriminantes y se identificaron variables clave asociadas a la probabilidad de impago de las empresas seleccionadas. Más tarde este modelo se extendió a empresas que no cotizan en Bolsa (Altman 1993), a la calificación de deuda de mercados emergentes y a la calificación de créditos de consumo, entre otras aplicaciones. Finalmente algunos estudios se han centrado en la comparación de la efectividad de los distintos modelos de AI aplicados a muestras específicas (West(2000), Malhotra y Malhotra(2002), Baesens et al (2003).

En el ejemplo que se presenta en esta sección, las empresas vienen descritas a través de ocho variables que corresponden a ocho ratios financieras. Para cada una de estas variables se han considerado cuatro etiquetas cualitativas básicas, por tanto, cada empresa queda caracterizada por ocho etiquetas cualitativas. Los datos financieros utilizados en este trabajo provienen de 22 empresas contempladas en las carteras de los índices *Standard & Poor's 1500* y *Dow Jones Euro 500*, de Estados Unidos y Europa, respectivamente. Los ratios se han calculado teniendo en cuenta los estados financieros presentados por las compañías a 30 de diciembre de 2000. La descripción de los ratios utilizadas se presentan en la figura 10. La información se obtuvo de la base de datos que provee *WorldScope* de *Thompson*³. Los *ratings* para dichas empresas para este periodo de tiempo fueron proporcionados por *Standard & Poor's* y corresponden a la clasificación de deuda a largo plazo en moneda local.

5.2. Resultados

Se ha calculado los valores de la función $K(\cdot, \cdot)$ para todos los pares de un conjunto reducido de 22 empresas, con distintos valores del parámetro λ . Las figuras 11, 12 y 13 muestran

³Software *Datastream Advance 3.5*

	Variable	Ratio
X_1	Cobertura de intereses	(Intereses pagados) / (Flujos de Caja Operacionales),
X_2	Endeudamiento	(Total Pasivos) / (Total Activos),
X_3	Capacidad de generación de recursos	(Flujos de Caja Operacionales) / (Total Pasivos),
X_4	Valoración de mercado	(Valor de mercado del patrimonio) / (Total Pasivos),
X_5	Rentabilidad	(Utilidad antes de Imptos. e Intereses) / (Total Activos),
X_6	Autofinanciamiento	$1 - ((\text{Dividendos pagados}) / (\text{Utilidad disponible a distribuir}))$
X_7	Estructura de la deuda	(Total Pasivos Circulantes) / (Total Pasivos),
X_8	Variación de las ventas	$(\text{Ventas}) / (\text{Ventas}_{t-1} - 1)$

Figura 10: Descripción de las variables consideradas

respectivamente los resultados de este cálculo para $\lambda = 0$, $\lambda = 0,3$ y $\lambda = 0,8$.

En cada una de las figuras la casilla de posición (i, j) representa el valor obtenido por la función núcleo aplicada a las empresas E^i, E^j : $K(E^i, E^j) = \langle \phi(E^i), \phi(E^j) \rangle$.

Para este cálculo se han ordenado las empresas de izquierda a derecha en las tablas, de mayor a menor *ratings*. La diagonal principal de las tablas obtenidas corresponde, por tanto, a los valores de $K(E^i, E^i)$. Cerca de esta diagonal se tienen las imágenes de pares de empresas con calificaciones similares, y en las casillas más alejadas de esta diagonal las imágenes corresponden a pares de empresas de muy diferente valor de *rating*. Los valores más altos obtenidos en nuestro ejemplo, en cada caso, se han sombreado en las respectivas tablas.

En la figura 11 aparecen los valores de K , en la que $\lambda = 0$, los valores de K van desde 0 a 8 y se corresponden exactamente con el número de variables que toman el mismo valor en las dos empresas. Por tanto, $K(E, E) = 8$ para cualquier empresa E y, además, en virtud de la fórmula 2, se tiene que $d^2(E^i, E^j) = 16 - 2 \cdot K(E^i, E^j)$, o sea, la función K es directamente una medida de similitud de las empresas. Puede comprobarse que, efectivamente, los valores cercanos a la diagonal son sensiblemente superiores que el resto.

En la figura 12 se muestran los resultados para $\lambda = 0,3$. En este caso, la aportación a la función núcleo de las variables con valores distintos pero próximos hace que los valores de K sean superiores que en el caso anterior. No obstante, la distribución de los valores mayores coincide sustancialmente con las obtenidas en el caso anterior.

En la figura 13, correspondiente a $\lambda = 0,8$, la

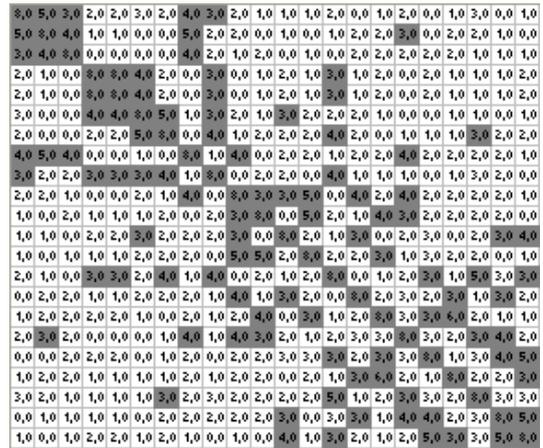


Figura 11: Representación de $K(.,.)$ para todas las parejas posibles de empresas consideradas con $\lambda \rightarrow 0$

aportación de las variables con valores distintos es excesiva y esta razón, junto con el hecho de utilizar un número muy reducido de etiquetas, hace que la función núcleo no se corresponde con una medida de similitud. Esto último sugiere que no es adecuado tomar un valor grande del parámetro λ .

6. Conclusión y Trabajo Futuro

El artículo se centra en la construcción de una función núcleo para trabajar en problemas en los que las variables de entrada vienen descritas mediante etiquetas cualitativas ordenadas. Este kernel está construido a partir del concepto de posicionamiento que permite capturar las ideas de alejamiento y proximidad entre valores cualitativos. La metodología empleada se puede utilizar en cualquier sistema de aprendizaje basado en funciones núcleo.

En primer lugar, como trabajo a desarrollar está el análisis de la influencia del alejamiento entre valores cualitativos en la función núcleo, en concreto se pretende utilizar distintos parámetros de decrecimiento λ en la expresión del producto escalar en función de las longitudes de los intervalos que definen el modelo cualitativo de órdenes de magnitud absolutos. En segundo lugar, se implementará este

8,1	5,5	3,9	2,5	2,5	3,2	2,3	4,2	3,7	2,5	2,4	1,3	1,9	2,5	0,5	2,1	2,7	0,5	1,9	3,5	0,5	1,3
5,5	8,1	4,4	1,3	1,3	0,6	0,4	5,5	2,4	2,9	1,2	1,3	1,0	1,5	2,5	2,9	3,7	0,6	2,9	2,5	1,5	0,6
3,9	4,9	8,1	0,6	0,6	0,9	0,6	4,8	2,5	2,2	2,9	0,5	2,1	0,9	2,7	3,1	2,8	2,3	3,1	1,9	1,5	1,3
2,5	1,3	0,6	8,1	8,1	4,7	3,0	0,2	3,9	0,7	2,2	2,9	1,9	3,8	1,7	2,7	0,8	2,9	1,5	1,9	1,8	2,7
2,5	1,3	0,6	8,1	8,1	4,7	3,0	0,2	3,9	0,7	2,2	2,9	1,9	3,8	1,7	2,7	0,8	2,9	1,5	1,9	1,8	2,7
3,2	0,6	0,9	4,7	4,7	8,2	5,7	1,2	3,7	2,7	2,4	3,9	3,1	2,9	2,9	2,1	0,9	1,3	2,1	1,9	1,3	2,3
2,3	0,4	0,6	3,0	3,0	5,7	8,1	0,4	4,3	1,9	2,9	3,1	3,0	4,5	2,9	0,9	1,5	1,9	1,9	3,5	2,7	3,1
4,2	5,5	4,8	0,2	0,2	1,2	0,4	8,1	1,3	4,5	1,0	2,1	2,7	1,3	2,7	2,9	4,5	2,2	3,1	2,3	2,4	1,5
3,7	2,4	2,5	3,9	3,9	3,7	4,3	1,3	8,1	0,9	3,2	2,5	1,1	4,5	1,5	2,1	1,7	0,8	1,9	3,3	2,5	0,8
2,5	2,9	2,2	0,7	0,7	2,7	1,9	4,5	0,9	8,1	3,5	3,5	5,1	0,4	4,7	2,5	4,5	2,3	2,7	2,3	2,5	1,5
2,4	1,2	2,9	2,2	2,2	2,4	2,9	1,0	3,2	3,5	8,1	0,9	5,3	2,7	1,7	4,5	3,5	2,2	2,6	2,7	2,2	0,4
1,3	1,3	0,5	2,9	2,9	3,9	3,1	2,1	2,5	3,5	0,9	8,1	2,5	1,7	3,5	0,8	2,5	3,5	1,0	2,5	3,7	4,7
1,9	1,0	2,1	1,9	1,9	3,1	3,0	2,7	1,1	5,1	5,3	2,5	8,1	2,5	2,9	3,9	1,5	3,4	2,9	2,5	0,4	1,4
2,5	1,5	0,9	3,8	3,8	2,9	4,5	1,3	4,5	0,4	2,7	1,7	2,5	8,1	0,9	2,1	2,5	3,7	1,9	5,2	3,3	3,7
0,5	2,5	2,7	1,7	1,7	2,9	2,9	2,7	1,5	4,7	1,7	3,5	2,9	0,9	8,1	2,7	3,3	2,7	3,9	1,9	3,7	2,9
2,1	2,9	3,1	2,7	2,7	2,1	0,9	2,9	2,1	2,5	4,5	0,8	3,9	2,1	2,7	8,1	3,3	3,5	6,2	2,9	1,3	1,5
2,7	3,7	2,8	0,8	0,8	0,9	1,5	4,5	1,7	4,5	3,5	2,5	1,5	2,5	3,3	3,3	8,1	3,1	2,5	3,7	4,3	2,5
0,5	0,6	2,3	2,9	2,9	1,3	1,9	2,2	0,8	2,3	2,2	3,5	3,4	3,7	2,7	3,5	3,1	8,1	1,5	3,5	4,1	5,1
1,9	2,9	3,1	1,5	1,5	2,1	1,9	3,1	1,9	2,7	2,6	1,0	2,9	1,9	3,9	6,2	2,5	1,5	8,1	2,7	2,5	3,5
3,5	2,5	1,9	1,9	1,9	3,5	2,3	3,3	2,3	2,7	2,5	2,5	5,2	1,9	2,9	3,7	3,5	2,7	8,1	3,1	3,5	
0,5	1,5	1,5	1,8	1,8	1,3	2,7	2,4	2,5	2,5	2,2	3,7	0,4	3,3	3,7	1,3	4,3	4,1	2,5	3,1	8,1	5,5
1,3	0,6	1,3	2,7	2,7	2,3	3,1	1,5	0,8	1,5	0,4	4,7	1,4	3,7	2,9	1,5	2,5	5,1	3,5	3,5	5,5	8,1

Figura 12: Representación de $K(,)$ para todas las parejas posibles de empresas consideradas con $\lambda = 0,3$

método para problemas de clasificación y multclasificación. Dentro del proyecto MERITO, se compararán los resultados obtenidos cuando los descriptores de los patrones de entrada son cualitativos con los obtenidos cuando son numéricos.

También, a partir de la metodología utilizada, se pretende trabajar con datos cualitativos y cuantitativos simultáneamente, representando los valores cualitativos mediante sus vectores posicionamiento y combinando éstos con los datos numéricos de que se disponga.

Por último, cabe también la posibilidad de definir otras funciones núcleo a partir diferentes valores que midan el grado de alejamiento o proximidad entre valores cualitativos, definidos no sólo por escalas ordinales si no también por atributos tales como el país y el sector en los que opera una compañía.

7. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto coordinado MERITO (análisis y desarrollo de técnicas innovadoras de soft-computing con integración de conocimiento experto: una aplicación a la medición del riesgo financiero de crédito), financiado por el Ministerio de Ciencia y Tecnología (TIC2002-04371-C02).

8. Bibliografía

Agell, N. *Estructures matemàtiques per al model qualitatiu d'ordres de magnitud absoluts*. Ph. D. Thesis Universitat Politècnica de Catalunya. 1998.

Altman, E.I. *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. Journal of Finance, Vol XXIII, Issue 4, Sept. 1968. p.589-609.

Altman, E.I. *Corporate Financial Distress and Bankruptcy : A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy* 2nd ed. New York. John Wiley & Sons, ISBN: 0-471-55253-4. 1993

20	19	19	19	19	20	19	19	19	19	20	18	19	19	18	19	19	18	19	18	18	18
19	19	19	18	18	18	19	18	19	19	19	19	18	18	19	19	19	18	19	18	18	18
19	19	20	19	19	20	19	19	19	19	20	18	20	19	19	20	19	18	20	19	18	18
19	18	19	20	20	20	18	20	19	20	19	19	19	19	19	19	18	19	19	18	19	19
19	18	19	20	20	20	18	20	19	20	19	19	19	19	19	19	18	19	19	18	19	19
20	19	20	20	22	21	19	20	20	21	20	20	20	20	20	19	19	20	20	19	20	20
19	18	19	20	20	21	20	18	20	19	20	20	19	19	19	19	20	19	19	19	19	19
19	19	19	18	18	18	18	19	18	19	18	19	18	18	19	18	17	19	18	18	18	18
19	19	19	20	20	20	18	20	19	20	19	19	19	19	19	19	18	19	18	19	18	18
19	19	19	19	19	20	19	19	19	20	19	18	18	20	19	18	19	18	19	19	19	19
20	19	20	20	21	20	19	20	20	21	19	20	19	20	19	19	20	19	19	20	19	19
18	18	18	19	19	20	18	19	19	19	20	19	19	19	19	18	18	18	18	18	18	19
19	19	20	19	19	20	20	19	19	19	20	19	20	19	20	19	20	19	20	19	20	18
19	18	19	19	19	20	18	19	18	18	18	19	19	20	19	19	18	19	18	19	19	18
18	18	19	19	19	20	19	19	20	19	19	20	19	19	20	19	19	19	19	19	19	19
19	19	20	19	19	20	19	19	19	20	19	20	19	19	19	19	18	18	18	18	18	18
19	19	19	18	18	19	18	19	18	19	18	19	18	18	19	18	18	18	18	18	18	18
18	18	18	19	19	19	17	18	18	18	18	19	18	19	19	18	19	18	18	18	18	18
19	19	20	19	19	20	20	19	19	19	20	19	20	19	20	20	19	19	20	19	19	19
18	18	19	19	19	20	19	18	18	19	19	18	18	19	19	19	19	18	18	18	18	18
18	18	18	18	18	19	18	18	18	19	19	18	18	18	18	18	18	18	18	18	18	18
18	18	18	19	19	20	19	18	18	18	19	19	19	19	19	18	18	18	18	18	18	18

Figura 13: Representación de $K(,)$ para todas las parejas posibles de empresas consideradas con $\lambda = 0,8$

Angulo, C. *Aprendizaje con máquinas núcleo en entornos de multclasificación*. Ph. D. Thesis Universitat Politècnica de Catalunya. 2001.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J.A.K., Vanthienen, J. *Benchmarking state-of-the-art classification algorithms for credit scoring*. Journal of the Operational Research Society. Oxford: June 2003, Vol.54. Iss.6, pp.627.

Cristianini, N., Shawe-Taylor, J. *An introduction to Support Vector Machines and other Kernel-based learning methods*. Cambridge University Press. 2000.

Fan, A., Palaniswami, M. *Selecting bankruptcy predictors using a support vector machine approach*. Neural Networks, IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on , Volume: 6 , 24-27 July 2000, Page(s): 354 -359 vol.6

Forbus, K.D. *Commonsense physics*. Annals Revue of Computer Science, 1988 pàg. 197-232.

Li, R.Z., Pang, S.L., Xu, J.M. *Neural network credit-risk evaluation model based on back-propagation algorithm*. Machine Learning and Cybernetics, Proceedings. 2002 International Conference on , Volume: 4 , 4-5 Nov. 2002, Page(s): 1702 -1706 vol.4

Malhotra, R., Malhotra, D.K. *Differentiating between goods credits and bad credits using neuro-fuzzy-systems*. European Journal of Operation Research. Amsterdam: Jan 1, Vol.136, Iss.1, 2002. PP. 190

Pang, S.L., Wang, Y.M., Bai, Y.H. *Credit scoring model based on neural network*. Machine Learning and Cybernetics. Proceedings. 2002 International Conference on , Volume: 4 , 4-5 Nov. 2002 Page(s): 1742 -1746 vol.4

Sánchez, M., Prats, F., Agell, N., Rovira, X. *Funciones kernel en espacios cualitativos de órdenes de magnitud: Aplicación a la evaluación del riesgo de crédito*. Actas de la X Conferencia de la Asociación Española para la Inteligencia Artificial CAEPIA'03. Nov 2003 vol I pp 69-78

Travé-Massuyès, L., Dague, P., Guerrin, F. *Le Raisonnement Qualitatif pour les Sciences de l'Ingenieur*. Hermès. 1997

Vapnik, V. *The nature of statistical learning theory*. Springer Verlag New York. 1995.

West, D. *Neural network credit scoring models*. Computers & Operations Research, New York: Sep/Oct 2000. Vol. 27, Iss.11, 12, pp. 1131



Ricardo A. Campos. He was born in Chile in 1969, he received in 1987 the title of Accountant (from Chilean Ministry of Education), and in 1996 the title of Public Accountant and Science Degree in Commerce and Economic Sciences, minor in Finance (from the Catholic University of Valparaiso, Chile). At the moment, he is Ph.D. candidate in the "Ph.D. in Management Sciences" program of ESADE and he is instructor professor at the Pontifical Catholic University of Valparaiso. Also, he is Research Assistant and Scholarship holder at ESADE. He participates actively in a research project about innovative techniques of Soft-Computing applied to the measurement of the credit risk.



Francisco J. Ruiz. Graduated in Physics (1988) from the University of Barcelona, Ph.D. candidate in Computer Science in the Technical University of Catalonia. He is currently associate professor of Mathematics at the Technical University of Catalonia and academic assistant in ESADE at the University Ramon Llull. He is presently active in investigating Machine Learning, in particular statistical learning theory and its application to finances.



Xari Rovira. Received her degree in Mathematics at the Universitat de Barcelona (1984). She has an MBA from ESADE (1993) and she was awarded a Ph.D. from the Universidad Politécnica de Madrid (1988) for a dissertation on Functional Equations and Inequalities. She has also a Master degree on Mathematics for financial instruments from the Universitat Autònoma de Barcelona (2000). She is currently Professor in the Quantitative Methods Department and Associate Director of the Ph.D. Program at ESADE, Universitat Ramon Llull (Barcelona). She is member of the research groups GREC and ARCA. Her fields of interest are the applications of Qualitative and Fuzzy Reasoning to Management Sciences, specially to Finance, as well as the study of mathematical aspects of Qualitative Calculus on Orders of Magnitude Models.



Mónica Sánchez. Graduated in Mathematics and with Ph. D. in Computer Science, is a Senior Lecturer in the Applied Mathematics 2 Department of the Technical University of Catalonia in Barcelona and the responsible person of this University's Research Line "Qualitative Reasoning and Decision Technologies". Her research work is focused on the applications of Qualitative Reasoning and Qualitative Calculus on Orders of Magnitude, Fuzzy Logic and in general Artificial Intelligence techniques.



Núria Agell. Was born in Barcelona in 1958. Graduated from the University of Barcelona in Mathematics. She was awarded a Ph.D. in Mathematics from the Technical University of Catalonia. Full professor in the Quantitative Methods department, ESADE, University Ramon Llull, where at the present time she is the coordinator of the research group "Artificial Intelligence Applications to Management". Her main research interests are Qualitative Reasoning formalisms and Learning Algorithms. Her work is now focused on the application of Artificial Intelligence techniques to marketing and finance. She is member of the research groups GREC (Research Group on Engineering Knowledge) and ARCA (Qualitative Reasoning Automatization and Applications).



Cecilio Angulo. Received his M.Sc. degree in Mathematics in 1993 from the University of Barcelona and the Ph.D. degree in 2001 from the Technical University of Catalonia. He is currently an Aggregated Professor at the Technical University of Catalonia and a member of the Knowledge Engineering Research Group, GREC, a partner of the European Research Laboratory LEA SICA. His research interest include machine learning, intelligent control and financial data modelling