

EL PROBLEMA DE SELECCIÓN DE VARIABLES. APLICACIÓN A LA PREDICCIÓN DEL RIESGO DE QUIEBRA EN LAS EMPRESAS

CASADO YUSTA, SILVIA

Departamento de Economía Aplicada
Universidad de Burgos
correo-e: scasado@ubu.es

GOMEZ PALACIOS OLGA

Instituto de empresa
correo-e: olga_gomezpalacios@yahoo.es

NÚÑEZ LETAMENDIA LAURA

Departamento de Finanzas
Instituto de Empresa
correo-e: laura.nunez@ie.edu

PACHECO BONROSTRO, JOAQUÍN

Departamento de Economía Aplicada
Universidad de Burgos
correo-e: jpacheco@ubu.es

RESUMEN

En este trabajo se hace uso de la estrategia metaheurística GRASP para resolver el problema de selección de variables en el ámbito de la clasificación. El problema consiste en dado un conjunto de variables usadas en la clasificación, seleccionar el subconjunto que lleve a cabo la tarea de forma óptima.

Este caso tiene su aplicación en el ámbito financiero. La trascendencia de la quiebra por sus efectos nocivos para accionistas, acreedores, empleados, etc., justifica los esfuerzos que desde los 60 – con los trabajos de Beaver y Altman - se vienen realizando por encontrar los ratios financieros que permitan anticipar consistentemente la misma.

Palabras clave: *selección de variables, metaheurístico, quiebra, ratios financieros*

1. Introducción

Desde los pioneros trabajos de Beaver (1966) y Altman (1968), se ha desarrollado una considerable literatura sobre la predicción de la insolvencia empresarial¹, o incapacidad de las empresas de hacer frente al pago de sus deudas, a través de la información contable disponible sobre las mismas. La trascendencia de la insolvencia, por sus efectos nocivos para accionistas, acreedores, empleados, etc., justifica este esfuerzo por encontrar variables contables o financieras con capacidad predictiva en este sentido.

Los primeros estudios sobre insolvencia utilizaban técnicas univariantes – este es el caso de Beaver (1966). Fue Altman, en 1968, quien introdujo el análisis multivariante discriminante, que constituyó la técnica predominante durante la década de los 70. Posteriormente, en los 80, el análisis discriminante – cuyo supuesto de normalidad para las variables independientes es habitualmente violado por las distribuciones de los ratios financieros - fue reemplazado por el análisis de regresión logística, que perdura hasta nuestros días. Previo a la realización de este tipo de análisis (análisis discriminante, logit...etc.) algunos de los paquetes estadísticos que se usan para ello realizan una primera selección de variables con objeto de eliminar del análisis aquellas variables que son menos significativas. En este trabajo nos vamos a centrar en la resolución de este problema de selección de variables aplicado al mundo financiero, considerando por tanto como variables de nuestro estudio los ratios financieros.

El problema de selección de variables consiste en encontrar un subconjunto de variables con las que se pueda llevar a cabo la tarea de clasificar de forma óptima. Reducir la dimensionalidad conlleva diversas ventajas como la reducción del coste en la adquisición de datos, mejora en la comprensión del modelo final del clasificador, incremento en la eficiencia del clasificador y mejora en la eficacia del clasificador. La búsqueda del subconjunto de variables es un problema NP-duro (Kohavi, 1995), de

¹ A lo largo de este trabajo cuando hablamos de insolvencia nos referimos indistintamente a los estados de suspensión de pagos y quiebra.

modo que es razonable el uso de metaheurísticas que permiten obtener soluciones buenas sin explorar todo el espacio de soluciones.

Las técnicas metaheurísticas realizan una exploración de todo el campo de soluciones posible buscando aquellas regiones con buenas soluciones, para después centrar la búsqueda en estas regiones seleccionadas. Actualmente estas técnicas están siendo usadas para resolver muchos tipos de problemas de optimización aunque en su origen la mayoría de ellos fueron diseñados para la resolución de determinados problemas de optimización combinatoria. En esta expresión podemos incluir la mayoría de los problemas que tienen un número finito o al menos numerable de soluciones alternativas.

Dentro de las estrategias metaheurísticas se han aplicado al problema de selección de variables los Algoritmos Genéticos (AG) (Bala et al., 1996 y Jourdan et. al., 2001) y más recientemente la Búsqueda Dispersa (BD) (García et al., 2003) alcanzándose, en ambos casos, resultados prometedores.

El problema de selección de variables no es por tanto un problema muy estudiado en la literatura, y como se ve las pocas aportaciones que hay son relativamente recientes. En nuestro caso, a diferencia de las anteriores referencias, se va a tratar este problema en el caso especial en el que las variables que se usan para clasificar son todas cuantitativas. El uso de todas las variables cuantitativas permite una mejor medición y comparación de su capacidad clasificatoria o discriminante. Esto hace que se puedan desarrollar métodos de selección variables, adaptados a este tipo de variables, en los que se incorpore estas medidas discriminantes, y por tanto mas eficaces que los anteriores (para este caso).

En este trabajo se resuelve el problema de selección de variables mediante un algoritmo basado en la estrategia metaheurística denominada GRASP. Concretamente se comparan los resultados que obtiene GRASP con los obtenidos por otros dos

algoritmos constructivos: el algoritmo en el que si inspiran paquetes estadísticos como BMDP y SPSS para realizar la selección de variables (lo hemos denominado “*algoritmo constructivo determinístico*”) y otro algoritmo constructivo en el que, sobre la base del anterior, se introduce cierto grado de aleatoriedad (denominado en este trabajo “*algoritmo ávido-aleatorio*”).

El trabajo se estructura de la siguiente forma. En la sección 2 se describe el modelo y la formulación del problema y en la sección 3 se describe el procedimiento GRASP usado. En la sección 4 se describe la muestra de empresas y los ratios que se consideran. Posteriormente en la penúltima sección se muestran los resultados de las experiencias computacionales realizadas y finalmente en la sección 6 se exponen las principales conclusiones obtenidas.

2. Modelización y formulación del problema de selección de variables

Considérese un conjunto de n casos o instancias $A = \{ a_1, a_2, \dots, a_n \}$, y un conjunto de m variables $V = \{ v_1, v_2, \dots, v_m \}$; (para simplificar también se identificará indistintamente V con los coeficientes, es decir, $V = \{ 1, 2, \dots, m \}$). Cada instancia a_i viene definida de la forma

$$a_i = (a_{i1}, a_{i2}, \dots, a_{im} \mid c_i),$$

es decir por el valor que toman las variables y la clase a la que pertenece.

Sea un valor predefinido $p \in N$, se trata de encontrar un subconjunto $S \subset V$, de tamaño p de mayor capacidad clasificatoria.

Para medir la capacidad clasificatoria de los diferentes subconjuntos S considérese k particiones previamente definidas del conjunto A . En cada una de ellas hay dos subconjuntos A_1 (conjunto de entrenamiento) y A_2 (conjunto de validación). Es decir

$A = A_1 \cup A_2$, donde A_1 y A_2 tienen aproximadamente la misma proporción de elementos de cada clase que A . El cardinal de todos los subconjuntos A_1 es el mismo (y por tanto el de los A_2). Para cada subconjunto de variables S , y para cada par de instancias a_i y a_t se define la siguiente distancia

$$d(a_i, a_t) = \sum_{j \in S} d_j^2(a_i, a_t)$$

donde

$$d_j(a_i, a_t) = \frac{|a_{ij} - a_{tj}|}{\max_j - \min_j}$$

con \max_j y \min_j los valores máximos y mínimos de la variable v_j observados en el conjunto de entrenamiento.

Para determinar el valor de la bondad $f(S)$ de cada subconjunto de variables S se actúa como se indica a continuación. Para cada una de las particiones consideradas se realiza el siguiente proceso: Para cada instancia a_i del conjunto de validación A_2 se determina la instancia más cercana del conjunto de entrenamiento A_1 , a_{i^*} y se asigna a a_i a la clase a la que pertenece a_{i^*} . El porcentaje de aciertos total es la medida de la bondad $f(S)$ de cada subconjunto S .

3.- Descripción de un algoritmo GRASP

Nuestro método está basado en constructivos GRASP. GRASP, o Greedy Randomized Adaptive Search Procedure, es una estrategia metaheurística que construye soluciones usando una aleatoriedad controlada mediante una función voraz. La mayoría de las implementaciones GRASP, como la que se propone en este trabajo, además incluyen una búsqueda local que se usa para mejorar las soluciones generadas con el método ávido-aleatorio. GRASP fue propuesto originalmente para el set covering problem (Feo and Resende, 1989). Detalles de esta metodología y aplicaciones recientes se pueden encontrar en Feo and Resende (1995) y Pitsoulis and Resende (2002).

Un esquema del funcionamiento de nuestro algoritmo GRASP es el siguiente

Repetir

Construir una solución por el método Avido-aleatorio

Mejorar dicha solución mediante búsqueda local

Actualizar la mejor solución obtenida hasta ese momento

hasta alcanzar un criterio de parada

El criterio de parada se alcanza cuando transcurren un predeterminado número de iteraciones (*max_iter*) sin mejora. A continuación se describen los dos procedimientos principales: el método ávido aleatorio y la búsqueda local.

3.1.- Procedimiento ávido-aleatorio

La función ávida que guía la entrada de variables en la solución esta basada en conocidísimos resultados sobre la descomposición de la varianza. Más concretamente, sea x una variable cualquiera definida sobre los n casos considerados, es decir, $x' = (x_1, x_2, x_3, \dots, x_n)$, ng el número de clases y nn_i el número de casos del grupo i , $i = 1 \dots ng$. Sea así mismo

\bar{x} : media de la variable x en el conjunto de los n casos;

\bar{x}_i : media de la variable x en los casos de la clase i ; $i = 1 \dots ng$;

$cl(j)$: clase a la que pertenece el individuo j .

Se definen

$$VT(x) = \sum_{j=1}^n (x_j - \bar{x})^2 \quad (\text{Variabilidad Total})$$

$$VE(x) = \sum_{i=1}^{ng} nn_i (\bar{x}_i - \bar{x})^2 \quad (\text{Variabilidad entregrupos})$$

$$VI(x) = \sum_{j=1}^n (x_j - \bar{x}_{cl(j)})^2 \quad (\text{Variabilidad intragrupos})$$

$$y \quad F(x) = \frac{VE(x)}{VI(x)}$$

Se sabe que $VT(x) = VE(x) + VI(x)$; también es conocido que la función $F(x)$ es un buen medidor de la capacidad discriminatoria de cada variable.

Sea S la solución que se va a construir, el procedimiento ávido-aleatorio se describe de la forma siguiente

1. Iniciar: Hacer $S = \emptyset$
2. Calcular $F_j = F(v_j), j = 1..m$
3. Determinar $Fmax = \max \{F_j / j = 1..m\}$ y $Fmin = \min \{F_j / j = 1..m\}$
4. Construir $L = \{j / F_j \geq \alpha \cdot Fmax + (1-\alpha) \cdot Fmin\}$
5. Elegir $j^* \in L$ aleatoriamente y hacer $S = \{j^*\}$
6. Mientras $|S| < p$ hacer:
 - a. Sea $S = \{j_1, j_2, \dots, j_t\}$ (las variables que ya están en la solución)

$\forall j \notin S$: - Determinar los valores de la variable r_j en el siguiente modelo lineal por mínimos cuadrados ordinarios $v_j = \alpha + \beta_1 \cdot v_{j_1} + \beta_2 \cdot v_{j_2} + \dots + \beta_t \cdot v_{j_t} + r_j$

- Calcular $F_j = F(r_j)$
 - b. Determinar $Fmax = \max \{F_j / j \notin S\}$ y $Fmin = \min \{F_j / j \notin S\}$
 - c. Construir $L = \{j / F_j \geq \alpha \cdot Fmax + (1-\alpha) \cdot Fmin\}$
 - d. $j^* \in L$ aleatoriamente y hacer $S = S \cup \{j^*\}$

Como se observa la función F , antes definida, es la que guía el procedimiento de selección de variables. Sin embargo en cada paso no se elige necesariamente la variable correspondiente al mayor valor de F , $Fmax$. En este caso se construye un conjunto L (denominado “lista de candidatos”), formado por los de mayor valor, y se elige aleatoriamente uno de esa lista.

Inicialmente la función guía es el valor de la función F en las variables originales. Posteriormente se usa el valor de la F pero no en las variables originales candidatas a entrar, sino en los residuos que se obtienen al quitar en dichas variables la información que nos suministran las variables que ya están en la solución S . Esta idea la usan algunos programas estadísticos conocidos como BMDP y SPSS en sus procedimientos para seleccionar variables previos a la ejecución de los métodos de discriminación propiamente dichos. La diferencia con nuestro método GRASP, es que la selección de variables que usan estos programas es determinística, es decir, siempre se selecciona la variable correspondiente a $Fmax$.

Precisamente, una de las ideas para el uso de estos procedimientos ávido aleatorios, es que la mejor solución a la que se llega repitiendo varias veces la ejecución de este procedimiento, suele ser mejor que la obtenida con la selección determinística. Como se verá en los apartados siguientes esto también ocurre en este caso.

El parámetro α sirve para controlar el grado de aleatoriedad del procedimiento. A mayor valor de α menor aleatoriedad. Si $\alpha = 0$, el procedimiento es totalmente aleatorio, ya que L o “Lista de candidatos”, estaría formada por todos los variables que no están en la solución; Si $\alpha = 1$ L estaría únicamente formado por la variable correspondiente a $Fmax$. En adelante, en este trabajo denominaremos *constructivo determinístico* al método propuesto cuando $\alpha = 1$.

3.2.- Procedimiento de búsqueda local

A cada solución completa S generada por el procedimiento ávido aleatorio se la mejora posteriormente por un sencillo procedimiento de búsqueda local. En este caso, cada paso de la búsqueda local va a consistir en intercambiar una variable que este dentro de la solución por otra que este fuera. Mas concretamente, sea S una solución se define

$$N(S) = \{ S' / S' = S \cup \{j'\} - \{j\}, \forall j \in S, j' \notin S \}$$

El procedimiento de búsqueda local se puede describir como sigue

Leer Solución Inicial S

Repetir

Hacer valor_ant = f(S)

Buscar $f(S^) = \max \{ f(S') / S' \in N(S) \}$*

Si $f(S^) > f(S)$ entonces hacer $S = S^*$*

hasta $f(S^) \leq \text{valor_ant}$*

Como se observa el procedimiento finaliza cuando ningún intercambio produce mejora.

4. Muestra de empresas y ratios

4.1 Empresas

La muestra empleada esta compuesta de 198 empresas españolas, de las que aproximadamente un tercio - 67 - suspendieron pagos o instaron la quiebra en el año 2003. El resto - 131 - se encontraban activas en dicho año. Las empresas fueron seleccionadas de la Base de Datos SABI, que comprende todas las empresas que depositan sus cuentas en el Registro Mercantil español. La selección se hizo de forma aleatoria para cada grupo (activas / en suspensión de pagos o quiebra), contemplando sólo las formas jurídicas de sociedad limitada y sociedad anónima, y aceptando sólo aquellas que tuvieran la información completa o a falta de muy pocos valores para los tres años anteriores². Para la mayor parte de las empresas insolventes - cuyo estado de insolvencia se produjo en el año 2003 - no estaban disponibles los datos del 2002, por lo que la muestra se tomó bajo el requisito de que existieran datos bien para los tres años 2002, 2001, y 2000, bien para los tres años 2001, 2000, y 1999. Este factor introduce un cierto sesgo, ya que en todos los casos del grupo de empresas activas – cuya situación de activa se mantiene hasta diciembre de 2003 al menos - se toman datos

² La tasa de valores no disponibles - 244 - para el conjunto de datos recogidos – que ascendió a 27.918 - se ha situado por debajo del 1%. En estos casos, se ha mantenido el dato referente al periodo anterior, o de no existir este tampoco, el referente al periodo posterior.

de los años 2002, 2001 y 2000, mientras que en el grupo de empresas insolventes, los datos referentes a dichos años resultan estar sólo en 7 de las empresas - habiéndose tomado para el resto como se ha comentado datos de 2001, 2000, y 1999. No obstante, el hecho de que cuando se habla de empresas activas, su situación se refiere a diciembre de 2003, mientras que cuando se habla de empresas insolventes, su situación hace referencia a cualquier momento del año 2003 – y de hecho un 67% de ellas adquiere la situación de insolvencia en la primera mitad del año 2003, y el 100% en los primeros 9 meses - mitiga en gran medida este sesgo.

La tabla 1 muestra los datos de ambas distribuciones por sectores. Aunque las muestras han sido seleccionadas de forma aleatoria, sin tener en cuenta su pertenencia a uno u otro sector, para cada grupo - solventes e insolventes - es interesante señalar que 55 de las 67 empresas insolventes han resultado emparejadas a nivel sectorial, para la CNAE a dos dígitos. La tabla 2 muestra la media de las distribuciones por tamaño, medido este por el número de empleados, por forma jurídica y por antigüedad de la empresa. Como era de esperar, el tamaño medio de las empresas solventes resulta superior al de las empresas insolventes. No obstante, la eliminación de las empresas solventes que cuentan con más de 100 empleados – que resultan ser sólo 10 – de la muestra, conlleva una reducción del tamaño medio de las empresas solventes al nivel del grupo de las insolventes, e incluso ligeramente por debajo. Es interesante observar que la forma jurídica de las empresas se reparte por igual en ambas distribuciones, siendo un 60% de las mismas sociedades limitadas y el restante 40% sociedades anónimas. Por otra parte resulta asombroso que la antigüedad media de las empresas de ambos grupos resulte idéntica, 18 años, siendo sus desviaciones estándar muy similares, dado que suele argumentarse que muchas de las empresas que suspenden pagos o quiebran lo hacen en sus primeros años de vida. En este sentido nuestro análisis introduce un sesgo de supervivencia, que parcialmente puede estar explicando estos datos, en la medida en la que la selección de las muestras de ambos grupos de empresas se ha realizado aleatoriamente, pero con la condición de que hubiera datos disponibles para los 3 ejercicios anteriores al año 2003 para las empresas solventes y a este o al año 2002, en caso de que los datos del 2002 no estuvieran disponibles, lo que ha ocurrido con frecuencia, para las empresas insolventes. A partir de estos datos de antigüedad,

parecería que una vez que las empresas superan esos 2 o 3 primeros años de funcionamiento, su probabilidad de insolvencia no mantiene relación con su antigüedad.

CNAE-93 Rev. 1	Insolventes	Solventes
01 - Agricultura y Ganadería	0	3
02 - Forestal	0	1
15 - industria Alimentación y Bebidas	5	6
17 - Ind. Textil	3	1
18 - Ind. Confección	1	1
19 - Fabr. Calzado	0	1
20 - Ind. Madera y Corcho	1	2
21 - Ind. del Papel	0	1
22 - Edición y Artes Gráficas	2	3
24 - Industria Química	0	4
25 - Fabric. Prod. Caucho y Plástico	1	2
26 - Fabr. Otros Prod. Minerales	0	1
27 - Metalúrgica	2	1
28 - Fabr. Pro. Metálicos	4	3
29 - Construcc. Maquinaria	5	3
31 - Fabr, Maquinaria Eléctrica	2	0
33 - Fabr. Equipo Médico	0	1
34 - Fabr,. Vehículos Motor	0	1
35 - Fabr. Otro Material Transporte	1	0
36 - Fabr. Muebles; Otras Industrias	4	3
41 - Captación, Depuración y Distrib. Agua	0	1
45 - Construcción	10	16
50 - Venta y Repara. Vehículos Motor	0	5
51 - Comercio al Por Mayor	12	16
52 - Comercio al Por Menor	7	11
55 - Hostelería	0	4
60 - Transporte Terrestre	0	2
61 - Transporte Marítimo	0	1
63 - Activ. Auxiliares de Transporte	1	1
65 - Intermed. Financiera (excepto seguros)	0	1
70 - Promoción Inmobiliaria	2	16
74 - Otras activ. Empresariales	2	12
80 - Educación	2	0
85 - Activ. Sanitarias y Veterinarias	0	2
92 - Act. Recreativas, culturales y deportivas	0	4
93 - Act. De Servicios Personales	0	1
Total	67	131

Tabla 1. *Distribución de las muestras de empresas por sectores*

	Insolventes (67)	Solventes (131)
Tamaño*		
Número medio de empleados	22 (22)	36 (65)
Número medio de empleados (sin >100)**	22 (22)	20 (23)
Forma Jurídica***		
Sociedades Anónimas	27 (40%)	52 (40%)
Sociedad Limitada	40 (60%)	79 (60%)
Antigüedad*		
Número de años de existencia	18 (15)	18 (13)

* Entre paréntesis la desviación estándar

** Eliminando de la muestra de empresas solventes las 10 con un número de empleados superior a 100

*** Numero de empresas (entre paréntesis en porcentaje de cada muestra)

Tabla 2. *Medias de tamaño, forma jurídica y antigüedad de las distribuciones*

4.2 Ratios.

Se seleccionaron para cada empresa 36 ratios de los publicados en SABI – la base de datos empleada en el estudio – para cada uno de los tres años anteriores al 2003, o 2002 en su caso, lo que supone un total de 108 datos por empresa. Prácticamente se incluyeron todos los ratios publicados en SABI para las empresas españolas, a excepción de algunos para los que recurrentemente la información no estaba disponible, como por ejemplo el ratio de “plazo de pago a proveedores”, que desgraciadamente tuvo que excluirse. A su vez, se añadieron 11 nuevos ratios relativos a variaciones temporales para 11 de los 36 ratios previamente seleccionados – se calcularon tres variaciones para cada uno de los estos ratios: entre el año t y t-1, entre t-1 y t-2 y entre t y t-2. Por tanto los datos totales seleccionados para cada empresa ascienden a 141. La introducción de las variaciones temporales de los ratios en el análisis de insolvencia no es una práctica habitual, sin embargo, puede resultar de gran interés, dado que estudios previos, como Beaver (1966), detectan que la distribución de los ratios de las empresas solventes permanece constante en el tiempo, mientras que la de las empresas insolventes varía significativamente, a causa del deterioro que sufren estos. Teniendo en cuenta este factor la variación en determinados ratios podría tener mayor poder predictivo que el propio nivel del ratio y por otra parte, parece a priori, que dicha variación pueda tener también mayor independencia con relación a los sectores de actividad de las empresas,

que los niveles absolutos de los propios ratios. Las tablas 3 presenta la definición del conjunto de ratios financieros empleados.

Ratios de Actividad	
Crecimiento de la Cifra de ventas (%)	$[(Ventas_t - Ventas_{t-1}) / Ventas_{t-1}] \times 100\%$
Rotación de activos	$Ventas_t / Activo\ Total$
Productividad	$[Ingresos\ Explot. - Consumos\ y\ Gastos\ Explot.] / Gastos\ de\ Personal$
E_Costs of employee / Oper. Rev. (%)	$[Gastos\ de\ Personal / Ingresos\ de\ Explotación] \times 100\%$
Crecimiento Valor Añadido (%)	100%
E_Profit Margin (%)	$[Resultado\ antes\ de\ Impuestos / Ingresos\ de\ Explotación] \times 100\%$
E_Net Assets Turnover	$Ingresos\ de\ Explotación / Fondos\ Permanentes$
Ratios de Rentabilidad	
E_Return on Capital Employed (%)	$[(Resultado\ antes\ de\ Impuestos + Gastos\ Financieros) / Fondos\ Permanentes] \times 100\%$
Rentabilidad económica (%) (ROA)	$[Resultado\ del\ Ejercicio / Activo\ Total] \times 100\%$
E_Return on Total Assets (%)	$[Resultado\ antes\ de\ Impuestos / Activo\ Total] \times 100\%$
Rentabilidad financiera (%) (ROE)	$[Resultado\ del\ Ejercicio / Fondos\ Propios] \times 100\%$
E_Return on Shareholders Funds (%)	$[Resultados\ antes\ de\ Impuestos / Fondos\ Propios] \times 100\%$
Gastos financieros (%)	$[Gastos\ Financieros / Ventas] \times 100\%$
Ratios de Equilibrio	
Fondo maniobra (FM) (Mil)	$Fondos\ Propios + Provisiones\ R\ y\ G + Acreedores\ LP - Inmovilizado$
Necesidad fondo maniobra (NFM)(Mil)	$[ADNE + Gastos\ a\ Distribuir + (Activo\ Circulante - IFT - Tesorería)] - [Ingresos\ a\ Distribuir + (Acreedores\ CP - Entidades\ de\ Crédito)]$
Tesorería (Mil)	$IFT + Tesorería - Entidades\ de\ Crédito$
Equilibrio	$[Fondos\ Propios + Provisiones\ R\ y\ G + Acreedores\ LP] / Inmovilizado$
Ratios de Equilibrio Cinéticos	
Fondo de maniobra (días)	$[FM / Ventas] \times 360$
Necesidad fondo de maniobra (días)	$[NFM / Ventas] \times 360$
Tesorería (días)	$[Tesorería / Ventas] \times 360$
E_Collection period (días)	$[Deudores / Ingresos\ de\ Explotación] \times 360$
Credito de Clientes (días)	$[Deudores / Ventas] \times 360$

Ratios de Solvencia	
Endeudamiento (%)	$[(\text{Pasivo Total} - \text{Fondos Propios}) / \text{Pasivo Total}] \times 100\%$
E_Solvency Ratio (%)	$[\text{FP} / \text{Activo Total}] \times 100\%$
FP sobre Permanentes (%)	$[[\text{FP} / [\text{FP} + \text{Acreedores LP} + \text{Provisiones R y G}]] \times 100\%$
Capacidad devolución	$[\text{Acreedores LP y CP} / [\text{Ventas} + \text{Dotaciones} + \text{Var. Provisiones} + \text{PF}]]$
Ratios de Liquidez	
Liquidez inmediata	$[\text{IFT} + \text{Tesorería}] / \text{Acreedores a CP}$
E_Current Ratio	Activo Circulante / Pasivo Líquido
E_Liquidity Ratio	$[\text{Activo Circulante} - \text{Existencias}] / \text{Pasivo Líquido}$
E_Interest Cover	Resultado de Explotación / Gastos Financieros
Ratios por empleado	
E_Profit per Employee (Mil)	Resultado antes de Impuestos / Número de Empleados
E_Oper. Rev. per Employee (Mil)	Ingresos de Explotación / Número de Empleados
E_Aver. Cost of Empl. / Year (Mil)	Gastos de Personal / Número de Empleados
E_Share Funds per Employee (Mil)	Fondos Propios / Número de Empleados
E_Work. Capital per Employee (Mil)	Fondo de Maniobra / Número de Empleados
E_Total Assets per Employee (Mil)	Activo Total / Número de Empleados

Tabla 3. Ratios seleccionados (SABI). Sombreadas variaciones anuales incorporadas

5. Experiencias computacionales

Se han realizado una serie de experiencias computacionales para la muestra de empresas descrita anteriormente. El número de casos (empresas) que se consideran es 198, divididos en 2 clases (solventes y no solventes), con 131 y 67 elementos respectivamente. Se considera una partición, obtenida aleatoriamente $A = A_1 \cup A_2$, donde A_1 tiene 100 elementos (66 solventes y 34 no) y A_2 98 (65 y 33).

A continuación se muestran las tablas de los resultados obtenidos. (tablas 5, 6, 7, 8, 9 y 10). En cada tabla se han seleccionado un número diferente de ratios ($p=10, \dots, 15$). Cada tabla cuenta con tres columnas, una para cada estrategia usada: *constructivo determinístico*, *constructivo ávido aleatorio* (se ejecuta 20 veces y $\alpha = 0.85$), y finalmente *GRASP* ($\alpha = 0.85$ y $max_iter = 20$). Así mismo, la segunda fila de cada tabla

después del título, es para el valor de la función objetivo (f), que mide el porcentaje de aciertos (expresado en tanto por uno).

Constructivo Determinístico	Avido-aleatorio	GRASP (Avido-aleatorio+b. local)
$f. = 0.66938776$	$f.= 0.71020408$	$f= 0.80612245$
E_Return on total assets_t	E_Return on total assets_t	E_Return on total assets_t
FP sobre permanentes_t-1	Crecimiento valor añadido_t	Crecimiento valor añadido_t
Crecimiento valor añadido_t	FP sobre permanentes_t-1	E_Return on Capital Employed_t
E_solvency ratio_t-1	FP sobre permanentes_t	Rotación de activos_t-2
Productividad_t	Productividad_t	Equilibrio_t-2
Crecimiento valor añadido_t-1	Credito clientes dias (t-2)	E_Solvency Ratio_t
Productividad_t-2	Productividad_t-2	Crecimiento valor añadido_t-1
E_cost of employee/Oper. Rev. t-1	E_cost of employee/Oper. Rev. t-1	E_cost of employee/Oper. Rev. t-1
FP sobre permanentes_t	Endeudamiento_t vs_t-1	E_solvency Ratio_t vs_t-2
E_return on total Assets_t-2	Crecimiento valor añadido_t-1	E_Return on Shareholders Funds_t

Tabla 4. Ratios seleccionados para $p=10$

Constructivo determinístico	Avido- aleatorio	GRASP (Avido-aleatorio+b. local)
$f = 0.67959184$	$f = 0.70408163$	$f= 0.80204082$
E_Return on total assets_t	E_Return on total assets_t	E_Return on total assets_t
FP sobre permanentes_t-1	FP sobre permanentes_t-1	E_solvency ratio_t-1
Crecimiento valor añadido_t	Crecimiento valor añadido_t	Crecimiento valor añadido_t
E_solvency ratio_t-1	E_solvency ratio_t-1	E_Return on Shareholders Funds_t
Productividad_t	Crecimiento valor añadido_t-1	Productividad_t
Crecimiento valor añadido_t-1	Productividad_t	E_Aver. Cost of Empl. / Year_t-2
Productividad_t-2	Productividad_t-2	E_cost of employee/Oper. Rev. t-1
E_cost of employee/Oper. Rev. t-1	E_cost of employee/Oper. Rev. t-1	Rotación de activos_t-2
FP sobre permanentes_t	Endeudamiento_t vs_t-1	E_solvency Ratio_t-1 vs_t-2
E_return on total Assets_t-2	E_Net Assets Turnover_t-1	Endeudamiento_t vs_t-2
E_Net Assets Turnover_t-1	E_solvency Ratio_t vs_t-2	E_solvency Ratio_t vs_t-2

Tabla 5. Ratios seleccionados para $p=11$

Constructivo determinístico	Constructivo Ávido-aleatorio	GRASP (Ávido-aleatorio+b. local)
f=0.67959184	f=0.7	f=0.80612245
E_Return on total assets_t	E_Return on total assets_t	E_Return on total assets_t
FP sobre permanentes_t-1	E_Solvency Ratio_t-1	E_Solvency Ratio_t-1
Crecimiento valor añadido_t	Crecimiento valor añadido_t	Crecimiento valor añadido_t
E_solvency ratio_t-1	FP sobre permanentes_t	E_Return on Shareholders Funds_t
Productividad_t	Productividad_t	Rentabilidad económica_t
Crecimiento valor añadido_t-1	Crédito clientes (días) t-2	Productividad_t
Productividad_t-2	Endeudamiento_t vs t-1	E_cost of employee/Oper. Rev. t-1
E_cost of employee/Oper. Rev. t-1	E_cost of employee/Oper. Rev. t-1	E_solvency ratio_t-1_vs_t-2
FP sobre permanentes t	Productividad_t-2	Rotación de activos_t-2
E_return on total Assets t-2	Crecimiento valor añadido_t-1	E_Net Assets Turnover_t-1
E_Net Assets Turnover_t-1	E_solvency ratio_t	Endeudamiento_t vs t-2
Endeudamiento_t vs t-2	E_Net Assets Turnover_t-1	E_Solvency Ratio_t vs t-2

Tabla 6. Ratios seleccionados para $p=12$

Constructivo determinístico	Ávido-aleatorio	GRASP (Ávido-aleatorio+b. local)
f=0.68775510	f=0.71020408	f=0.81632653
E_Return on total assets_t	E_Return on total assets_t	E_Return on total assets_t
FP sobre permanentes_t-1	FP sobre permanentes_t-1	FP sobre permanentes_t-1
Crecimiento valor añadido_t	Crecimiento valor añadido_t	Equilibrio_t-1
E_solvency ratio_t-1	E_solvency ratio_t-1	E_solvency ratio_t-1
Productividad_t	Crecimiento valor añadido_t-1	Crecimiento valor añadido_t-1
Crecimiento valor añadido_t-1	Productividad_t	Liquidez inmediata_t-1_vs_t-2
Productividad_t-2	E_cost of employee/Oper. Rev. t-1	Endeudamiento_t-1_vs_t-2
E_cost of employee/Oper. Rev. t-1	Productividad_t-2	E_Profit Margin_t_vs_t-1
FP sobre permanentes t	Endeudamiento_t vs t-1	Endeudamiento_t vs t-1
E_return on total Assets t-2	E_Net Assets Turnover_t-1	E_Net Assets Turnover_t-1
E_Net Assets Turnover_t-1	E_Return on Shareholders Funds_t-2	Rentabilidad económica_t
Endeudamiento_t_vs_t-2	E_Return on Capital Employed_t-2	E_Return on Capital Employed_t-2
E_Solvency Ratio_t_vs_t-2	Rentabilidad Financiera_t-2	E_solvency ratio_t_vs_t-1

Tabla 7. Ratios seleccionados para $p=13$

Constructivo determinístico (BMDP)	Ávido- aleatorio	GRASP (Ávido-aleatorio+b. local)
f =0.67551020	f =0.71428571	f =0.82040816
E_Return on total assets_t	E_Return on total assets_t	E_Return on total assets_t
FP sobre permanentes_t-1	FP sobre permanentes_t-1	Rentabilidad económica_t
Crecimiento valor añadido_t	Crecimiento valor añadido_t	FP sobre permanentes_t-1
E_solvency ratio_t-1	E_solvency ratio_t-1	E_solvency ratio_t-1
Productividad_t	Crecimiento valor añadido_t-1	Crecimiento valor añadido_t-1
Crecimiento valor añadido_t-1	Productividad_t	Equilibrio_t-1
Productividad_t-2	E_cost of employee/Oper. Rev. t-1	Endeudamiento_t-1_vs_t-2
E_cost of employee/Oper. Rev. t-1	Productividad_t-2	Fondo Maniobra_t_vs_t-1
FP sobre permanentes t	Endeudamiento_t_vs_t-1	Necesidad fondo Maniobra_t_vs_t-1
E_return on total Assets t-2	E_Net Assets Turnover_t-1	Endeudamiento_t_vs_t-1
E_Net Assets Turnover_t-1	E_Return on Shareholders Funds_t-2	E_Net Assets Turnover_t-1
Endeudamiento_t_vs_t-2	E_Return on Capital Employed_t-2	E_solvency ratio_t_vs_t-1
E_Solvency Ratio t_vs_t-2	E_Aver. Cost of Empl. / Year_t-2	E_Return on Capital Employed_t-2
E_Return on Capital Employed_t-1	E_Oper. Rev. per Employee_t-2	E_Profit_Margin_t

Tabla 8. Ratios seleccionados para p=14

Constructivo determinístico	Ávido- aleatorio	GRASP (Ávido-aleatorio+b. local)
f = 0.66530612	f =0.70816327	f = 0.81224490
E_Return on total assets_t	E_Return on total assets_t	E_Return on Capital Employed_t
FP sobre permanentes_t-1	Crecimiento valor añadido_t	Equilibrio_t-1
Crecimiento valor añadido_t	FP sobre permanentes_t-1	E_Solvency Ratio_t-1
E_solvency ratio_t-1	E_solvency ratio_t-1	FP sobre permanentes_t
Productividad_t	Crecimiento valor añadido_t-1	Tesorería(días)_t-1
Crecimiento valor añadido_t-1	Productividad_t	E_Oper Rev. per Employee_t
Productividad_t-2	Productividad_t-2	Crecimiento Valor Añadido_t-1
E_cost of employee/Oper. Rev. t-1	E_cost of employee/Oper. Rev. t-1	Necesidad fondo maniobra_t_vs_t-1
FP sobre permanentes t	Endeudamiento_t_vs_t-1	E_Solvency Ratio_t-1_vs_t-2
E_return on total Assets t-2	E_Net Assets Turnover_t-1	E_Return on Capital Employed_t-2
E_Net Assets Turnover_t-1	E_Aver. Cost of Empl. / Year_t-2	E_Net Assets Turnover_t-1
Endeudamiento_t_vs_t-2	E_Oper. Rev. per Employee_t-1	E_Solvency Ratio_t_vs_t-2
E_Solvency Ratio t_vs_t-2	E_Profit_Margin_t_vs_t-1	Endeudamiento_t_vs_t-1
E_Return on Capital Employed_t-1	E_Return on Shareholders Funds_t-2	E_Profit_Margin_t
Crecimiento de la Cifra de ventas_t-2	E_Return on Capital Employed_t-2	Liquidez inmediata_t-1_vs_t-2

Tabla 9. Ratios seleccionados para p=15

Para cada valor de p se observa como el valor de la función objetivo, o lo que es lo mismo el porcentaje de aciertos, es mejor con el procedimiento GRASP (tercera columna de cada tabla) que con los otros dos. Esto es lógico dado que en el procedimiento GRASP la solución obtenida por el constructivo se mejora con el procedimiento de búsqueda local. Vemos, por tanto, como la estrategia metaheurística nos permite obtener las mejores soluciones. Por otro lado, también se observa como el constructivo ávido-aleatorio (2ª columna en la tabla) nos permite obtener mejores resultados que el constructivo determinístico en el que se basan distintos paquetes estadísticos. Esto nos permite concluir que los resultados obtenidos por estos paquetes estadísticos pueden ser mejorados, ya sea simplemente incorporando aleatoriedad al constructivo o mediante una estrategia metaheurística más compleja, como GRASP, que mejora las soluciones que obtiene el constructivo aleatorizado con un procedimiento de búsqueda local.

En cuanto al número de ratios que se seleccionan vemos que va de 10 a 15. Se observa como al aumentar el número de ratios seleccionados no aumenta necesariamente el valor de la función objetivo. Si que vemos como los mejores valores de la función objetivo se obtienen cuando $p=14$ y $p=15$ y para el procedimiento GRASP, pero también se observa como mientras que el valor de f para $p=10$ usando el constructivo aleatorizado es de $f=0.71020408$, para $p=11$ y con el mismo procedimiento $f=0.70408163$, es decir, que el porcentaje de aciertos es menor a pesar de que hemos aumentado el valor de p .

Finalmente, en la tabla 11 se muestran los resultados globales obtenidos. En las columnas 2, 3 y 4 de esta tabla se observa el número de veces que cada ratio ha sido seleccionado por las distintas estrategias usadas: *constructivo*, *C_aleatorio* y *Grasp*. En la primera columna de la tabla 11 se muestra el número total de veces que ese ratio ha sido seleccionado para el conjunto de estrategias y en la última columna se indica el tipo de ratio de que se trata: A (actividad), R (rentabilidad), E (equilibrio), S (solvencia), L (liquidez), E_C (equilibrio cinético) y PE (por empleado).

TOTAL	Determinístico	Avido-aleatorio	Grasp	Ratios seleccionados	
31	12	12	7	Crecimiento Valor añadido	A
1	1	0	0	Crecimiento Ventas	A
26	12	12	2	Productividad	A
15	6	6	3	E_cost of employee/Oper. Rev.	A
4	0	1	3	E_Profit Margin	A
7	2	2	3	Rotación de Activos	A
10	3	3	4	E_Net Assets Turnover	A
3	0	0	3	Rentabilidad Económica	R
23	12	6	5	E_Return on Total Assets	R
1	0	1	0	Rentabilidad Financiera	R
6	0	3	3	E_Return on Shareholders Funds	R
10	2	3	5	E_Return on Capital Employed	R
31	9	7	15	E_Solvency Ratio	S
22	12	7	3	FP sobre Permanentes	S
17	4	6	7	Endeudamiento	S
4	0	0	4	Equilibrio	E
1	0	0	1	Fondo Maniobra	E
1	0	0	2	Necesidad Fondo Maniobra	E
2	0	2	0	Crédito de Clientes (días)	E_C
1	0	2	1	E_Oper. Rev. Per Employee	PE
3	0	2	1	E_Aver. Cost of Empl. / year	PE
2	0	0	2	Liquidez Inmediata	L
1	0	0	1	Tesorería (Mill.)	E
	Determinístico	Avido-aleatorio	Grasp	Ratios no seleccionados	
	0	0	0	Gastos Financieros %	R
	0	0	0	Fondo de maniobra (días)	E_C
	0	0	0	Necesidad fondo de maniobra (días)	E_C
	0	0	0	Tesorería (días)	E_C
	0	0	0	E_Collection period (días)	E_C
	0	0	0	Capacidad devolución	S
	0	0	0	E_Current Ratio	L
	0	0	0	E_Liquidity Ratio	L
	0	0	0	E_interest Cover	L
	0	0	0	E_Profit per employee	PE
	0	0	0	E_Share Funds per employee	PE
	0	0	0	E_Work. Capital per Employee	PE
	0	0	0	E_Total Assets per Employee	PE
224	75	75	75		

Tabla 10. Número de veces que cada ratio es seleccionado por los diferentes modelos

Ahora, centrándonos en los ratios seleccionados, es decir , aquellos que mejor pueden predecir una situación de insolvencia en la empresa podemos concluir lo siguiente:

- Con carácter general los ratios más seleccionados son ratios de actividad, solvencia y en menor medida los de rentabilidad. concretamente destacan los ratios de Crecimiento del Valor añadido, E_Solvency Ratio, Productividad, E_Return on Total Assets , FP

sobre permanentes, endeudamiento y $E_cost\ of\ employee/Oper.\ Rev.$. En conjunto estos ratios nos permiten tener un buen conocimiento sobre el funcionamiento de la empresa.

- Por otro lado vemos como entre los ratios seleccionados cobran protagonismo los ratios relativos a variaciones temporales, ya sea entre el año t y $t-1$, como entre el t y $t-2$ y $t-1$, $t-2$. Por tanto, a pesar de que de forma habitual en este tipo de análisis no se introduzcan variaciones temporales de los ratios, vemos como si que tienen gran importancia. En Beaver (1966), se indica como la distribución de los ratios de las empresas solventes permanece constante en el tiempo, mientras que la de las empresas insolventes varia significativamente. En este caso destaca sobre todo la variación de los ratios de endeudamiento y de solvencia, cosa por otro lado lógica dado que cuando estos ratios sufren un empeoramiento en el tiempo puede indicar que la empresa se encuentra próxima a una situación de insolvencia.

6. Conclusiones

En este trabajo se trata el problema de selección de variables en el ámbito de la clasificación aplicado al mundo financiero, concretamente a la predicción de la insolvencia financiera en las empresas. Los métodos de solución que se proponen son tres: un algoritmo constructivo determinístico (en el que se inspiran algunos paquetes estadísticos), un algoritmo constructivo ávido-aleatorio y un procedimiento GRASP (ávido-aleatorio + búsqueda local). Tras la realización de las experiencias computacionales se observa como el procedimiento GRASP es el que nos permite obtener los mejores resultados. Por otro lado también se puede concluir que la selección de un número de ratios mayor no implica una mejora en el valor de la función objetivo.

Referencias bibliográficas

Altman, E., (1968): “*Financial Ratios, Discriminant analysis and the prediction of corporate bankruptcy*” in *Journal of Finance*, September, pp. 189-209.

Bala J., Dejong K., Huang J., Vafaie H. y Wechsler H. (1996): “*Using Learning to Facilitate the Evolution of Features for Recognizing Visual Concepts*”. *Evolutionary Computation*, 4, 3, 297-311.

Beaver, W., (1966): “*Financial ratios as predictors of failures, in Empirical Research*” in *Accounting, selected studies*, pp. 71-111.

Feo T.A. y Resende M.G.C., (1989): “*A Probabilistic Heuristic for a Computationally Difficult Set Covering Problem*” in *Operations Research Letters*. Vol. 8. Pp. 67-71.

Feo T.A. y Resende M.G.C., (1995): “*Greedy Randomized Adaptive Search Procedures*” in *Journal of Global Optimization*. Vol. 2. Pp. 1-27.

García F.C., García-Torres M., Moreno Pérez J.M. y Moreno-Vega J.M. (2003): “*Búsqueda Dispersa para el Problema de la Selección de Variables*”. CAEPIA-2003.

Jourdan L., Dhaenens C. y Talbi E. (2001): “*A Genetic Algorithm for Feature Subset Selection in Data-Mining for Genetics*”. MIC 2001 Proceedings, 4th Metaheuristics International Conference, 29-34.

Kohavi R. (1995): “*Wrappers for Performance Enhancement and Oblivious Decision Graphs*”. Stanford University, Computer Science Department.

Pitsoulis L.S. y Resende M.G.C. (2002): “*Greedy Randomized Adaptive Search Procedures*” in *Handbook of Applied Optimization*, P. M. Pardalos and M. G. C. Resende (Eds.), Oxford University Press, pp. 168-182.