



CUADERNOS DE LA FUNDACIÓN

Nº 97

*** * * ***

**APLICACIÓN DE MÉTODOS DE
INTELIGENCIA ARTIFICIAL PARA
EL ANÁLISIS DE LA SOLVENCIA EN
ENTIDADES ASEGURADORAS**

**Autores: María Jesús Segovia Vargas
 Carlos Bousoño Calzón
 Sancho Salcedo Sanz**

Julio, 2005

ISBN: 84-89429-87-1

Depósito Legal: M-30.849-2005

Copyright: Fundación MAPFRE Estudios

Prohibida la reproducción total o parcial de este trabajo sin el permiso escrito de los autores o de la FUNDACIÓN MAPFRE ESTUDIOS

LISTA DE CUADERNOS DE LA FUNDACIÓN MAPFRE ESTUDIOS EDITADOS:

1. Filosofía Empresarial
2. Resultados de la Encuesta sobre "Altos Profesionales de Seguros" (A.P.S.)
3. Dirección y Gestión de la Seguridad
4. Los Seguros en una Europa cambiante: 1990-1995 (No disponible)
5. La Distribución Comercial del Seguro: Sus Estrategias y Riesgos
6. Elementos de Dirección Estratégica de la Empresa
7. Los Seguros de Responsabilidad Civil y su Obligatoriedad de Aseguramiento
8. La Implantación de un Sistema de Controlling Estratégico en la Empresa
9. Técnicas de Trabajo Intelectual
10. Desarrollo Directivo: Una Inversión Estratégica
11. El Concepto de Seguridad en la Ciencia y la Ciencia de la Seguridad
12. Los Seguros de Salud y la Sanidad Privada
13. Calidad Total y Seguridad
14. El Reaseguro de Exceso de Pérdidas
15. El Coste de los Riesgos en la Empresa Española 1991
16. La Legislación Española de Seguros y su Adaptación a la Normativa Comunitaria

Número Especial: Informe sobre el Mercado de Seguros 1993

17. Medio Ambiente Seguro: Desarrollo Futuro
18. El Seguro de Crédito a la Exportación en los países de la OCDE (Evaluación de los resultados de los aseguradores públicos)
19. Una Teoría de la Educación
20. El Reaseguro en los Procesos de Integración Económica

Número Especial: Informe sobre el Mercado de Seguros 1994

21. La Nueva Regulación de las Provisiones Técnicas en la Directiva de Cuentas de la C.E.E. Provisiones Técnicas de Seguros de Vida en las Directivas Comunitarias
22. Rentabilidad y Productividad de Entidades Aseguradoras
23. Análisis de la Demanda de Seguro Sanitario Privado
24. El Seguro: Expresión de Solidaridad desde la Perspectiva del Derecho
25. El Reaseguro Financiero
26. El Coste de los Riesgos en la Empresa Española 1993
27. La Calidad Total como Factor para elevar la Cuota de Mercado en Empresas de Seguros
28. La Naturaleza Jurídica del Seguro de Responsabilidad Civil
29. Ruina y Seguro de Responsabilidad Civil Decenal

Número Especial: Informe sobre el Mercado de Seguros 1995

30. El Tiempo del Directivo
31. Tipos Estratégicos, Orientación al Mercado y Resultados Económicos: Análisis Empírico del Sector Asegurador Español

32. Decisiones Racionales en Reaseguro
33. La función del Derecho en la Economía
34. El Coste de los Riesgos en la Empresa Española 1995
35. El Control de Riesgos en Fraudes Informáticos
36. Cláusulas Limitativas de los Derechos de los Asegurados y Cláusulas Delimitadoras del Riesgo Cubierto. Las Cláusulas de Limitación Temporal de la Cobertura en el Seguro de Responsabilidad Civil

Número Especial: Informe sobre el Mercado de Seguros 1996

37. La Responsabilidad Civil por Accidente de Circulación. Puntual Comparación de los Derechos Francés y Español
38. Legislación y Estadísticas del Mercado de Seguros en la Comunidad Iberoamericana
39. Perspectiva Histórica de los Documentos Estadístico-Contables del Órgano de Control: Aspectos Jurídicos, Formalización y Explotación
40. Resultados de la Encuesta sobre la Organización y Gestión de la Seguridad en la Empresa (1996)
41. De Maastricht a Amsterdam: Un paso más en la integración europea

Número Especial: Informe sobre el Mercado de Seguros 1997

42. La Responsabilidad Civil por contaminación del entorno y su aseguramiento
43. Resultados de la Encuesta sobre Disponibilidad de Instalaciones de Protección contra Incendios en la Empresa 1997"
44. Resultados de la Encuesta sobre Implantación en la Empresa de la Ley de Prevención de Riesgos Laborales
45. Los Impuestos en una Economía Global

46. Evolución y Predicción de las Tablas de Mortalidad Dinámicas para la Población Española
47. El Fraude en el Seguro del Automóvil: Cómo detectarlo
48. Matemática Actuarial no Vida con MapleV
49. Solvencia y Estabilidad Financiera en la Empresa de Seguros: Metodología y Evaluación Empírica mediante Análisis Multivariante
50. Mixturas de Distribuciones: Aplicación a las variables más relevantes que modelan la siniestralidad en la Empresa Aseguradora
51. Seguridades y Riesgos del joven en los grupos de edad
52. La Estructura Financiera de las Entidades de Seguros
53. Habilidades Directivas: Estudio de sesgo de género en instrumentos de evaluación
54. El Corredor de Reaseguro y su legislación específica en América y Europa
55. Resultados de la Encuesta: "La Seguridad contra Intrusión (Seguridad Privada) en la Empresa. 1999"
56. Análisis económico y estadístico de los factores determinantes de la demanda de los seguros privados en España
57. Informe final. Encuesta: "La Organización y Gestión de la Seguridad en la Empresa. 1999"
58. Problemática contable de las operaciones de reaseguro
59. Estudios sobre el Euro y el Seguro
60. Análisis Técnico y Económico del conjunto de las empresas aseguradoras de la Unión Europea

61. Sistemas Bonus-Malus generalizados con inclusión de los costes de los siniestros
62. Seguridad Social. Temas generales y régimen de clases pasivas del Estado
63. Análisis de la repercusión fiscal del seguro de vida y los planes de pensiones. Instrumentos de previsión social individual y empresarial
64. Fundamentos técnicos de la Regulación del Margen de Solvencia
65. Ética Empresarial y Globalización
66. Encuesta: "Seguridad contra Incendios en la empresa. 2000"
67. Gestión Directiva en la Internacionalización de la Empresa
68. Los seguros de crédito y de caución en Iberoamérica
69. Provisiones para prestaciones a la luz del Reglamento de Ordenación y Supervisión de los Seguros Privados: Métodos Estadísticos de Cálculo
70. El Cuadro de Mando Integral para las entidades aseguradoras
71. Gestión de activos y pasivos en la cartera de un fondo de pensiones
72. Análisis del proceso de exteriorización de los compromisos por pensiones
73. Financiación del capital-riesgo mediante el seguro
74. Estructuras de propiedad, organización y canales de distribución de las empresas aseguradoras en el mercado español
75. Incidencia de la Nueva Ley de Enjuiciamiento Civil en los procesos de responsabilidad civil derivada del uso de vehículos a motor
76. La incorporación de los sistemas privados de pensiones en las pequeñas y medianas empresas

77. Resultados de la Encuesta sobre *"El Coste de los Riesgos en la Empresa Española. 2001"*
78. Nuevas perspectivas de la educación universitaria a distancia
79. La actividad de las compañías aseguradoras de vida en el marco de la gestión integral de activos y pasivos
80. Los Planes y Fondos de Pensiones en el contexto europeo: la necesidad de una armonización
81. El Seguro de Dependencia. Una visión general
82. Informe Final. Encuesta: "La Organización y Gestión de la Seguridad en la Empresa 2002"
83. La teoría del valor extremo: fundamentos y aplicación al seguro, ramo de responsabilidad civil autos
84. Estudio de la estructura de una cartera de pólizas y de la eficiencia de un Sistema Bonus-Malus
85. La Matriz Valor-Fidelidad en el Análisis de los Asegurados en el Ramo del Automóvil
86. El Margen de Solvencia de las Entidades Aseguradoras en Iberoamérica
87. Dependencia en el modelo individual, aplicación al riesgo de crédito
88. Análisis Multivariante Aplicado a la Selección de Factores de Riesgo en la Tarificación
89. Mercados Aseguradores en el Área Mediterránea y Cooperación para su Desarrollo
90. Centros de atención telefónica del sector asegurador

91. Problemática de la reforma de la Ley de Contrato de Seguro
92. Medición de la esperanza de vida residual según niveles de dependencia en España y costes de cuidados de larga duración
93. Informe Final. Encuesta: "Seguridad contra Intrusión y Actos Antisociales en la Empresa (Seguridad Privada). 2003"
94. La renovación del Pacto de Toledo y la reforma del sistema de pensiones: ¿Es suficiente el pacto político?
95. Papel del docente universitario: ¿Enseñar o ayudar a aprender?
96. El sistema ABC - ABM: su aplicación en las entidades aseguradoras
97. Aplicación de métodos de inteligencia artificial para el análisis de la solvencia en entidades aseguradoras

**APLICACIÓN DE MÉTODOS DE INTELIGENCIA
ARTIFICIAL PARA EL ANÁLISIS DE LA
SOLVENCIA EN ENTIDADES ASEGURADORAS**

Autores: María Jesús Segovia Vargas
Universidad Complutense de Madrid

Carlos Bousoño Calzón
Universidad Carlos III de Madrid

Sancho Salcedo Sanz
Universidad de Alcalá de Henares

Para la realización de este trabajo la Fundación MAPFRE ESTUDIOS concedió a sus autores una Beca de Investigación Riesgo y Seguro.

-ÍNDICE-

INTRODUCCIÓN.....	1
CAPÍTULO I.- DEFINICIÓN DEL PROBLEMA Y DE LAS	
VARIABLES.....	10
I.1.- Concepto de crisis empresarial e insolvencia: definición del problema y de la variable independiente	11
I.2.- Los estados financieros como instrumentos de la predicción de la insolvencia: los ratios.....	13
I.2.1.- Enfoque metodológico seguido en la selección de los ratios.....	18
I.2.2.- Selección de la muestra.....	22
I.2.3.- Análisis del Balance y la Cuenta de Resultados de las entidades aseguradoras.....	25
I.2.4.- Elección de las variables: Los Ratios Contables.....	28
I.2.5.- Análisis estadístico descriptivo de los ratios.....	45
ANEXO CAPÍTULO I.....	49
CAPÍTULO II.- EL MÉTODO ROUGH SET. APLICACIÓN	
A LA PREDICCIÓN DE INSOLVENCIAS.....	56
II.1.- El método Rough Set.....	57
II.1.1.- Aproximación al concepto de Rough Set.....	57
II.1.2.- Tabla de información y relación de no diferenciación.....	60
II.1.3.- Aproximación de conjuntos.....	63
II.1.4.- Reducción y dependencia de atributos.....	65

II.1.5.- Reglas de decisión.....	67
II.1.6.- Toma de decisiones utilizando reglas de decisión.....	71
II.2.- El método rough set aplicado a la predicción insolvencias en empresas españolas de seguros no vida.....	73
II.2.1.- Fundamentación del Análisis.....	73
II.2.2.- Aplicación del Análisis y Resultados.....	74
II.2.2.1.- Aplicación de la metodología rough set.....	74
II.2.2.2.- Resultados.....	83
II.2.3.- Validación de los resultados: Clasificación de empresas utilizando los dos algoritmos generados.....	86
II.2.4.- Comparación del análisis rough set con el análisis discriminante.....	88

CAPÍTULO III.- ALGORITMIA HEURÍSTICA HÍBRIDA.

APLICACIÓN A LA PREDICCIÓN DE INSOLVENCIAS 92

III.1.- Introducción a la algoritmia heurística.....	93
III.2.- El problema de clasificación multiatributo.....	95
III.2.1.- Predicción de insolvencia en empresas de seguros no vida.....	96
III.2.2.- Selección de características en problemas de clasificación.....	97
III.3.- Algoritmos heurísticos de búsqueda.....	99
III.3.1.- Algoritmos genéticos.....	99
III.3.2.- El algoritmo de enfriamiento simulado.....	104
III.4.- Aplicación de los algoritmos descritos a un problema real de selección de características.....	107
III.4.1.- Análisis de los resultados.....	109
APÉNDICE: Breve introducción a las máquinas de vectores soporte (SVM)...	111

RESUMEN Y CONCLUSIONES.....115

BIBLIOGRAFÍA.....122

INTRODUCCIÓN

A diferencia de otros problemas financieros, el gran número de agentes e intereses afectados cuando se produce una insolvencia hacen que el estudio de la misma se mantenga siempre actual. Es más, la viabilidad y, en consecuencia, el riesgo de fracaso empresarial (entendido como culminación de un proceso de insolvencia) ha sido reconocido como un factor importante en el área de la macroeconomía o en los análisis industriales. De hecho, tanto el número de empresas quebradas como su tendencia relativa, se consideran indicadores de la solidez de las industrias y del conjunto de la economía.

La actividad aseguradora, conjuntamente con la bancaria y de inversión, constituye una proporción muy significativa de las transacciones financieras. Derivado de este hecho, se desprende la importancia social y económica del sector asegurador. Por consiguiente, la detección precoz de insolvencias o de las condiciones que pueden llevar a que ésta acaezca, es una cuestión clave para las autoridades reguladoras, inversores, gestores, analistas financieros, bancos, auditores y asegurados. Es más, la detección de insolvencias en las aseguradoras es una cuestión de interés general dado que los costes asociados a este problema no se limitan a las responsabilidades que pudieran tener, en su caso, gestores y auditores, si no que recaen en el conjunto de la economía, por ejemplo, a través de los fondos de garantía existentes en algunos países.

En definitiva, dada la importancia de este problema, la actividad aseguradora ha sido objeto desde hace tiempo de supervisión a través de la articulación de diversos sistemas de control: obligación de informar sobre su situación económico-financiera mediante la publicidad de las cuentas anuales y auditoría; imposición de normas encaminadas a garantizar su solvencia como pueden ser el cumplimiento de los coeficientes de garantía y de solvencia; provisiones obligatorias; inversión de las mismas; capital mínimo; regulación de los precios; remisión a la autoridad supervisora de los estados confidenciales; o, finalmente, realización de inspecciones *in situ*.

En la actualidad el marco legislativo en materia de solvencia para entidades aseguradoras está siendo objeto de un profundo estudio. El proyecto *Solvencia II*¹ pretende marcar las directrices que reformen las reglas existentes en la Unión Europea en relación con la solvencia.

Antes del proyecto Solvencia II, ha habido otros proyectos previos en el ámbito europeo. Tras la finalización en el año 1997 del informe Muller, comienzan a concretarse las reflexiones sobre la solvencia de empresas de seguros. Este informe constituyó el punto de partida para las modificaciones a las normas relativas a la solvencia de las entidades aseguradoras. Estas modificaciones se materializaron, en un primer lugar en el proyecto Solvencia I que finalizó con la aprobación de las Directivas 2002/12/CE y 2002/13/CE que modifican los requisitos o normas relativas al margen de solvencia para este tipo de entidades.

¹ http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm

La finalidad de Solvencia II es revisar la normativa existente para la valoración y supervisión de la situación financiera global de las entidades aseguradoras y su modo de actuación interno. Este proyecto está dividido en dos fases. La primera de ellas ha tenido como objetivo determinar la forma general de un futuro sistema de solvencia de la UE. La segunda fase trata de concretar los detalles del sistema.

Solvencia II se concreta en tres objetivos (los denominados *tres pilares*):

- Desarrollar un sistema para establecer los mecanismos para el cálculo de los recursos propios mínimos y de las provisiones de las compañías en base a su exposición final a los riesgos.
- Establecer las competencias y mecanismos de los supervisores para que pudieran anticiparse a las situaciones de riesgo sin aumentar los niveles de solvencia exigidos.
- Determinar la información que deben proporcionar las entidades aseguradoras, fundamentalmente en relación a su política de gestión de riesgos, facilitando así la transparencia del mercado (*disciplinar el mercado*), y en consecuencia, los agentes intervinientes tendrían información suficiente sobre el nivel de solvencia de estas entidades en base a la cual podrían tomar sus decisiones.

Por tanto, y en línea con estos objetivos, desarrollar y aplicar nuevos métodos en relación con la supervisión de compañías aseguradoras o con el análisis de la información, tanto cuantitativa como *cualitativa*, que suministren las entidades aseguradoras relativa a su gestión de riesgos es una cuestión de gran actualidad e importancia.

El objetivo de nuestro trabajo es aplicar nuevas metodologías que permitan analizar o anticipar el riesgo de insolvencia de una entidad aseguradora (analizando variables cuantitativas y cualitativas) y que se conviertan en herramientas útiles, en modelos empíricos, para la toma de decisiones en este sector.

Los modelos son una herramienta más para los usuarios, que apoyados en otras informaciones, pueden ayudarles a dictaminar sobre una determinada empresa como empresa en crisis. Queremos señalar, que no pretendemos sustituir la capacidad o intuición de los distintos entes interesados en conocer la evolución de la empresa, pero sí proporcionar información adicional sobre la solvencia de la misma. Es decir, complementar la capacidad de juicio y experiencia de las personas, y añadir una sistemática o procedimientos objetivos comunes de análisis.

Estos modelos, en términos prácticos, pueden usarse como “sistema de alerta temprana” para directivos, autoridades y demás usuarios interesados en prevenir la crisis empresarial. Es decir, los analistas y otros usuarios podrían establecer con suficiente tiempo de antelación políticas que minimicen el riesgo

de posibles situaciones de crisis empresarial no deseadas. Otra utilidad sería el empleo del modelo como sistema de diagnóstico automático para preseleccionar aquellas empresas que necesiten una atención especial, de una manera rápida y a un coste bajo, y de este modo se lograría una gestión más eficiente del tiempo dedicado por el analista financiero, auditor, autoridad supervisora, etc. Finalmente, señalar que otra aplicación del mismo podría ser la de servir de guía para dar uniformidad a los juicios y métodos efectuados sobre una empresa.

Llegados a este punto queremos mencionar que, pese a la importancia de la predicción de la insolvencia en el sector asegurador, la mayoría de estudios y trabajos llevados a cabo para predecir la insolvencia o el fracaso empresarial no han considerado este sector. Por eso existen escasos modelos relativos al mismo. Este hecho se explica en parte por las siguientes razones: una equivocada falsa importancia debido al escaso número de empresas que componen este sector en comparación con otros y, en segundo lugar, las peculiaridades del negocio asegurador llevan a la necesidad de poseer unos conocimientos técnicos específicos sobre el mismo que lleva a muchos investigadores a no acometer su análisis. Estas peculiaridades hacen que variables y modelos que son válidos en otros sectores no lo sean tanto en el sector asegurador o deban ser redefinidos y adaptados a estas especiales características que presenta el mismo.

Evidenciada la importancia y actualidad de la cuestión que vamos a abordar, trataremos de justificar el planteamiento de nuestro trabajo.

La teoría financiera tradicional analiza sus problemas de decisión (entre los que se encuentra la predicción de la insolvencia empresarial), tanto a corto como a largo plazo, según una perspectiva de búsqueda de soluciones óptimas considerando un único objetivo y un conjunto de soluciones posibles. Este hecho, ha propiciado que muchos investigadores hayan aplicado métodos de investigación operativa a los problemas financieros. Esta perspectiva es demasiado simplista y dificulta su aplicación a complejos problemas financieros reales. El análisis de estos problemas financieros se debe producir en un marco más realista basado en el paradigma de la decisión multiatributo. Este nuevo escenario se fundamenta sobre la base de que los decisores, en el ámbito económico, buscan un equilibrio entre una serie de objetivos que normalmente están en conflicto, teniendo en cuenta todo un conjunto de atributos que reflejan sus preferencias, su conocimiento y experiencia. Es decir, se pasa de un contexto de búsqueda de soluciones óptimas a otro que busca soluciones satisfactorias.

El principal objetivo del paradigma multiatributo es proveer a los decisores de herramientas que les permitan resolver, sobre la base de múltiples atributos, complejos problemas de decisión donde hay varios objetivos en conflicto.

Como una de las *ventajas* que presentan los métodos multiatributo cabría mencionar la posibilidad de estructurar problemas de decisión complejos al permitir modelizarlos incorporando simultáneamente varios objetivos y varios criterios. Otra ventaja es que permiten la utilización tanto de variables cuantitativas (ratios financieros, por ejemplo) como variables de tipo cualitativo. Los problemas reales obedecen a un conjunto de causas complejas y múltiples que

justifican la necesidad de incorporar ambos tipos de variables para poder analizarlos. De este modo, los criterios cualitativos tales como la capacidad de la dirección, la tendencia del mercado, la posición en el mercado, etc., juegan un papel fundamental en el proceso de decisión; incluso en algunos casos las variables cuantitativas son consecuencia de algunos factores cualitativos. Por tanto, y dada la complejidad del problema que estamos considerando, los métodos multiatributo se adaptan perfectamente a él.

Además, estos métodos permiten que el usuario final, el decisor, se pueda implicar en todo el proceso de toma de decisiones incorporando su modo de pensar, conocimiento y experiencia en forma de criterios y evaluando los resultados que se obtienen, lo que le permite utilizar dichos resultados para justificar las decisiones que va a tomar. En consecuencia, las decisiones que proponen los modelos obtenidos con estos métodos, tienen que estar de acuerdo con las decisiones de los centros decisores que son sus naturales usuarios. Es más, los decisores deberían ser capaces de comprender y explicar sus decisiones y, razón por la que en la elaboración de los modelos multiatributo se utiliza su conocimiento y experiencia. En definitiva, este enfoque permite a los centros decisores comprender el funcionamiento del modelo y enseñarles para que puedan incluir nuevos conocimientos y experiencia; de esta manera se pueden modificar los modelos para hacer frente a los cambios del medio y explorar otros nuevos caminos.

Por estas y otras razones, muchos investigadores y usuarios han encontrado los métodos multiatributo interesantes y atractivos y los han utilizado en sus investigaciones. Nosotros hemos acudido, en consecuencia con lo expuesto, a la búsqueda de métodos multiatributo que se ajusten a la investigación que vamos a realizar. La elección de métodos multiatributo se hace por las ventajas que presenta su utilización así como por la problemática que plantea la utilización de otros métodos, como los estadísticos.

Dentro de los métodos multiatributo, las herramientas surgidas de la *Inteligencia Artificial* se han mostrado muy apropiadas para tratar problemas financieros. La filosofía de la que parten es diferente de las aproximaciones convencionales: concretamente pretenden ayudar a la toma de decisiones pero simplificando la tarea del usuario final, de tal forma que a este último no se le requiera prácticamente ningún conocimiento técnico para interpretar los resultados del modelo o para su aplicación.

Tradicionalmente se han venido utilizando en el sector asegurador técnicas estadísticas para el análisis de la solvencia tomando como base datos contables en forma de ratios financieros. La técnica más utilizada ha sido el *Análisis Discriminante*, pionera en los estudios sobre predicción del fracaso empresarial. Esta técnica presenta algunas limitaciones solventadas en parte por la *Regresión Logística (Logit)* y el *Análisis Probit*, métodos también utilizados en este sector. No obstante, el tipo de variables usadas en este sector no suelen satisfacer las hipótesis estadísticas que requieren dichos métodos, lo que dificulta su aplicación o cuestiona los resultados obtenidos. A veces en este tipo de técnicas la presencia de observaciones atípicas (*outliers*) afecta en gran medida a los

resultados finales y/o los resultados obtenidos son difíciles de interpretar. Además los métodos estadísticos difícilmente analizan variables cualitativas, variables que, dado lo expuesto anteriormente, van a cobrar gran importancia en el futuro escenario en que se enmarque la supervisión una vez que concluya la actual reforma en esta materia.

Por tanto, a diferencia de otros trabajos, y dado los buenos resultados obtenidos en estudios preliminares, pretendemos desarrollar *modelos basados en herramientas de la Inteligencia Artificial*. Dichos modelos versan sobre la *adquisición del conocimiento inductiva o aprendizaje inductivo*. Tratan de establecer similitudes entre ejemplos positivos, que representan el mismo concepto, y comprobar las no similitudes entre los ejemplos positivos y negativos, siendo estos últimos los que representan otros conceptos. El conocimiento basado en la similitud también se denomina *conocimiento empírico* para subrayar el hecho de que se basa en la inducción del conocimiento que subyace a unos datos empíricos. De este modo, para la ingeniería del conocimiento basado en la similitud no es necesario ningún conocimiento *a priori*.

El enfoque *rough set* es un representante típico de esta categoría; así como *algoritmia heurística híbrida* (AHH) donde se consideran combinaciones y mezclas de paradigmas tales como *las redes neuronales, el enfriamiento simulado y los algoritmos evolutivos*: son ambas metodologías las que utilizaremos en esta investigación.

La **Teoría Rough Set** utiliza la experiencia de una forma objetiva, a través del estudio de la experiencia histórica de una manera cuantitativa, para así explicitar reglas que, resumiendo y objetivando esa experiencia acumulada, ayuden en las decisiones futuras. El problema que nos ocupa, la predicción de las crisis empresariales, no permite establecer fácilmente las relaciones causa-efecto entre los atributos que están relacionados o que pueden llevar a una quiebra empresarial y el acaecimiento efectivo de la misma, por varias razones (McKee, 2000). Una de ellas es que en la actualidad la teoría no nos permite identificar claramente todos los posibles atributos relevantes, y, en consecuencia, algunos podrían ser omitidos. Una segunda razón es que algunos de los atributos pueden ser cualitativos y otros pueden ser cuantitativos, lo que provoca un problema adicional de medida. Una tercera razón es que los atributos pueden ocurrir en uno o más periodos previos a la quiebra y de este modo surge el problema de la medición temporal. Cuando se trabaja con compañías reales, estas dificultades se traducen en que el análisis de las relaciones causa-efecto se compone normalmente de inconsistencias en las clasificaciones. Es decir, una empresa sana puede tener los mismos atributos que una quebrada. Por todo ello, el enfoque *Rough Set* parece una opción razonable para tratar esta problemática.

Esta metodología no requiere que los datos satisfagan ningún tipo de requisito o hipótesis, no se ve afectada por la presencia de observaciones atípicas y el modelo final obtenido consiste en un conjunto de reglas de decisión en forma de sentencias lógicas que las hace fácilmente comprensibles por cualquier usuario. Además dichas reglas están basadas en decisiones pasadas (experiencia del decisor), y por tanto, son casos reales lo que justificaría su utilización para futuras

decisiones. Además se han obtenido unos resultados muy satisfactorios que han sido comparados con una técnica estadística, el Análisis Discriminante, consiguiendo resultados similares pero evitando los inconvenientes que implican el incumplimiento de las hipótesis estadísticas que se da en los ratios contables. Estos resultados demuestran que esta teoría es una herramienta de gran utilidad y una alternativa o complemento a los métodos estadísticos ya que mejora en algunos aspectos algunos de sus problemas. Es un método rápido para procesar gran cantidad de información tanto cualitativa como cuantitativa; lo hace, además, de manera sencilla porque las reglas de decisión no requieren para su interpretación la ayuda de expertos, y elimina las variables que no son necesarias. Todas estas características hacen idónea su utilización por un gran número de usuarios (auditores, autoridades supervisoras, inversores y analistas) que pueden beneficiarse de una herramienta que les ayude a la hora de justificar sus decisiones con un gran ahorro en coste y en tiempo.

La **AHH** está basada en una mezcla entre algoritmos heurísticos de búsqueda local (redes neuronales) y de búsqueda global (algoritmos evolutivos y enfriamiento simulado). La búsqueda local afina en una determinada región del espacio de soluciones al problema de insolvencias pero la optimalidad de sus resultados se limita a esa región. La búsqueda global es capaz de otear un horizonte más amplio, a costa de menor precisión local. Estas dos tendencias de búsqueda son miscibles en esquemas con sinergia global-local que mejora sustancialmente las prestaciones de ambas por separado. En otro orden de cosas, el adjetivo *heurístico* apela a los símiles de la naturaleza que se toman como principios rectores de la búsqueda: las redes neuronales naturales para las artificiales, la genética darwinista para los algoritmos de ese nombre y los procesos de ordenación cristalina en sólidos que se enfrían en el otro de los algoritmos empleados.

Las *Redes Neuronales* y, en particular, las *Máquinas de Vectores Soporte* (SVM) constituyen potentes esquemas de clasificación dadas unas variables concretas. Es precisamente en la elección de esas variables donde se ha observado la posibilidad de conseguir la sinergia global-local antes aludida. En este sentido, los *algoritmos genéticos* y el *enfriamiento simulado* se apoyan en los resultados locales conseguidos por las SVM para elegir las combinaciones –resultados globales- que minimicen las dispersiones en la predicción, mejorando sustancialmente los resultados de la misma.

Una vez aplicadas ambas metodologías –Rough Set y AHH-, y comprobada su idoneidad, las hemos comparado en cuanto a selección de variables y/o resultados resaltando las capacidades de cada una de las mismas.

Finalmente queremos mencionar que cualquier modelo necesita la especificación de las **variables** que van a formar parte del mismo. El problema de predicción del fracaso empresarial es uno de los problemas más complejos dentro del área de gestión empresarial ya que obedece a multitud de causas que pueden únicamente describirse por un gran número de variables. Los *ratios* financieros han sido las principales variables utilizadas en relación a la predicción de crisis empresariales. Aunque las metodologías que vamos a emplear permiten la

utilización de variables cualitativas (además de cuantitativas) como pueden ser la experiencia de los directivos, el nicho de mercado en el cual se encuentra la empresa, el entorno y sus cambios, etc., no hemos considerado en nuestra investigación características cualitativas debido a la imposibilidad de conseguirlas, aunque su incorporación a los modelos que hemos obtenido sería sencillo y los mejoraría notablemente. Por tanto, nos hemos centrado fundamentalmente en mostrar la idoneidad de ambas metodologías para el problema que vamos a acometer.

La utilización de documentos contables nos permite obtener una información cuantificable, objetiva y sistemática de la realidad económica y financiera de la empresa. Y, a través de su análisis, mediante determinadas técnicas eficientes, se puede evaluar e interpretar la información contable disponible por el analista, para ofrecer un diagnóstico sobre la situación pasada, presente o futura. Es más, los modelos de predicción del fracaso empresarial proporcionan una forma de probar el contenido informativo de los datos contables sobre la solvencia de la empresa, al medir la relación entre los datos y el fracaso. A la vez, son un mecanismo para pronosticar el fracaso empresarial, de manera que proveen de una herramienta útil para la toma de decisiones por parte de los múltiples usuarios del análisis de la solvencia (Calvo-Flores y García, 1998a, p. 8). En este sentido, los *ratios* financieros individualmente considerados ofrecen una valiosa información sobre aspectos diferentes de la situación empresarial y además, pueden agruparse en estructuras de análisis con el fin de sumar la información que cada ratio contiene sobre la empresa. De esta forma, conoceremos la valoración global de la situación empresarial y los aspectos concretos a mejorar para superar el deterioro global, en caso de que dicha valoración advierta del mismo, o en cualquier caso para corregir desequilibrios. De hecho, muchos de los trabajos realizados en relación con la predicción del fracaso empresarial han demostrado que existe una elevada correlación entre los datos contables y la futura solvencia de la empresa, corroborando la utilidad de dichos datos para la toma de decisiones.

Es más, el análisis contable es una disciplina viva y en constante evolución que se nutre de diferentes herramientas para cumplir el objetivo de convertir los datos contenidos en los estados económico-financieros en información útil para la toma de decisiones. Esta disciplina ha ido evolucionando en cuanto a la utilización de los instrumentos técnicos que emplea paralelamente a la evolución de otras ciencias, como la Estadística, la Investigación Operativa y la Computación. El desarrollo de nuevas técnicas de computación procedentes de la Inteligencia Artificial ha revolucionado el análisis de estados financieros y se revela como un área en constante crecimiento (Serrano y Martín del Brío, 1993).

En resumen, nuestro trabajo trata de comprobar empíricamente la utilidad e idoneidad de los métodos Rough Set y AHH para el análisis de la solvencia de las empresas de seguros, mediante el desarrollo de unos modelos que han sido aplicados a empresas aseguradoras españolas en el ramo no-vida, y que incorpora como factores de dicho análisis ratios financieros elaborados a partir de los Balances y Cuentas de Resultados de dichas entidades.

Para abordar este cometido, hemos estructurado nuestro trabajo en las siguientes secciones: en el **capítulo primero** explicamos la definición del problema y las variables que vamos a emplear; el **capítulo segundo** ilustra los principales conceptos del método *Rough Set* y desarrolla la aplicación práctica de dicha metodología además de su comparación con el *Análisis Discriminante*; el **capítulo tercero** expone, tras una introducción a la algoritmia heurística, los fundamentos esenciales de los algoritmos que se han empleado para la discusión, presentando los resultados obtenidos. Finalmente, terminamos arriesgando unas **conclusiones** acerca de la viabilidad de estas tecnologías en el ámbito de las aseguradoras.

Capítulo 1
DEFINICIÓN DEL PROBLEMA
Y
DE LAS VARIABLES

I.1.- CONCEPTO DE CRISIS EMPRESARIAL E INSOLVENCIA: DEFINICIÓN DEL PROBLEMA Y DE LA VARIABLE INDEPENDIENTE.

Sin ánimo de ser exhaustivos, queremos revisar brevemente en este apartado qué entendemos por crisis empresarial, su relación con la insolvencia, sus efectos y sus causas, con la intención de poder justificar las variables (ratios) que vamos a introducir en los modelos que se aplicarán para la predicción de la misma.

La expresión crisis empresarial, según las fuentes que consideremos, está sujeta a distintas acepciones. Nosotros nos referiremos a la crisis de carácter financiero fundamentalmente. Esta concepción financiera de crisis empresarial es muy amplia. Esta amplitud tanto en su significado como en su utilización lleva a la ambigüedad, ya que el término fracaso o crisis empresarial se emplea tanto para referirse a la *insolvencia técnica*, cuando la empresa es incapaz de atender sus obligaciones financieras, o a la *insolvencia definitiva*, cuando el valor total de los activos de la empresa es menor que el valor de sus pasivos. Por tanto, el concepto de solvencia se refiere normalmente a dos componentes relacionados pero distintos. El primer componente se refiere a la disponibilidad de medios líquidos necesarios para hacer frente a las obligaciones derivadas de la gestión corriente y de inversiones en una empresa puntualmente, y el segundo, a la capacidad para obtener beneficios y generar tesorería. En consecuencia, nos estamos refiriendo no sólo a la solvencia necesaria para un ejercicio, sino también a la garantía de solvencia (reflejada en una adecuada estructura patrimonial y una conveniente dimensión cuantitativa de fondos generados) que tiene una empresa de manera permanente.

Quizás, la voz *insolvencia* sea menos ambigua en su utilización, y se podría definir, en términos exclusivamente financieros, como la incapacidad de una entidad para pagar sus deudas. Un primer periodo de insolvencia puede ser superado mediante aplazamientos o medidas de reestructuración empresarial, o por el contrario desembocar en una crisis definitiva que termine en venta o absorción de la empresa, en situaciones concursales, o en cierre definitivo. La crisis empresarial es el colofón del proceso de agravación de la insolvencia, que culmina en el fracaso empresarial, excluyendo la venta o absorción, para entrar en procesos judiciales de quiebra, suspensión de pagos o en cierre empresarial.

Al margen de las posibles discusiones sobre estos significados, cabría preguntarse sobre cómo llega una empresa a ser insolvente. La variedad de situaciones por las que puede transitar una empresa insolvente agrava aún más la ya de por sí difícil tarea de dar de manera unívoca y concreta una definición de *fracaso empresarial*. La realidad nos muestra cómo situaciones y procesos similares pueden terminar en soluciones de venta de una empresa, mientras que otras terminan en cierre. También existen casos en los que la declaración por parte de los tribunales de una suspensión de pagos supone la confirmación del fracaso e inicio de la liquidación de la empresa, mientras que en otros casos, la empresa declarada en suspensión de

pagos continúa sus actividades y consigue un acuerdo de quita y/o aplazamiento en el pago de las deudas, para seguir operando con normalidad.

Esta diversidad de situaciones nos obliga a precisar un concepto de insolvencia para nuestro trabajo en función tanto de nuestros objetivos como de la disponibilidad de datos. Nuestra definición de insolvencia está basada en situaciones concursales ya que esta perspectiva aporta un concepto riguroso, ajeno a interpretaciones diversas y presente en bases de datos asequibles, lo que supone un añadido de objetividad para cualquier investigación empírica. En concreto, y dado que nuestra investigación se centra en el sector asegurador, una *empresa fracasada* será aquella que esté en una situación de dificultad grave, de desequilibrio patrimonial y financiero que afecte negativamente a la continuidad de la empresa. Por tanto, todas ellas habrán sido intervenidas por la C.L.E.A (Comisión Liquidadora de Entidades Aseguradoras); lo que nos da una medida, desde la perspectiva jurídica, objetivamente determinable de las empresas que fracasan.

La Comisión Liquidadora de Entidades Aseguradoras era un organismo autónomo de la Administración General del Estado, vinculado al Ministerio de Economía cuyo control de eficacia corresponde a la Dirección General de Seguros y Fondos de Pensiones. Entre sus funciones estaban las siguientes: asumir la liquidación de entidades aseguradoras cuando le sean encomendadas por el Ministro de Economía o, en su caso, por el Órgano competente de la respectiva autoridad autonómica; actuar como Interventor único en los procedimientos de Suspensión de Pagos y como Comisario, Síndico y Depositario en los de quiebra.

Se enfatiza que la definición de este concepto es un factor que influye con fuerza en los resultados obtenidos, porque las empresas que aun estando en situación de insolvencia extrema son vendidas, no están en el bloque de las fracasadas, al mismo tiempo que otras en situación legal de quiebra o suspensión de pagos, y que por tanto, se encuentran en el bloque de las quebradas, continúan sus actividades y consecuentemente son empresas vivas. Ello implica que la propia naturaleza del problema de predicción o clasificación del fracaso empresarial muestra un límite a la precisión que cabe pedir a sus resultados, y que se tiene en cuenta en forma de un coeficiente a la cifra de los mismos (Sanchis, 2000).

I.2.- LOS ESTADOS FINANCIEROS COMO INSTRUMENTOS DE LA PREDICCIÓN DE LA INSOLVENCIA: LOS RATIOS

Las características propias del sector seguros motiva indagación adicional acerca de los factores que hay que tener en cuenta a la hora de analizar la insolvencia, y como consecuencia de ella, el fracaso y la crisis dentro del mismo, es decir, para analizar la solvencia de las compañías aseguradoras hay que tener en cuenta los factores comunes a cualquier sector, pero además habría que añadir los específicos del sector asegurador.

Vamos a extendernos en esta cuestión ya que, al habernos centrado en los ratios financieros por los motivos que argumentaremos con posterioridad, el resto de factores que se podrían contemplar se excluyen de nuestro estudio.

Siguiendo a Bannister (1997), un primer conjunto de factores a considerar serían *los factores macroeconómicos* que vienen determinados al operar la compañía en un país determinado:

- *La práctica contable.* Las compañías de seguros están sometidas a dos tipos de normas en este terreno: las aplicadas a todas las compañías en general y las específicas para el sector. La razón de estas últimas es la protección de los asegurados, lo que lleva a un incremento de exigencias de información que tienen su impacto en la contabilidad.
- *La supervisión a que está sometido el sector.* Es necesario revisar si la compañía en cuestión está cumpliendo con todos los requisitos y la legislación aplicable al mundo del seguro.
- *La práctica fiscal.* La fiscalidad de un país tiene una enorme relevancia en las operaciones de seguro. Una fiscalidad favorable (la existencia de incentivos fiscales) a la actividad aseguradora puede llevar a un aumento de los fondos procedentes de suscripciones de primas y viceversa.
- *Factores monetarios.* Las fluctuaciones en el tipo de cambio de las monedas pueden influir en la industria aseguradora internacional ya que puede haber un aumento en la siniestralidad derivada de una diferencia de cambio negativa. Es necesario además revisar la convertibilidad de las reservas o del efectivo cuando los pagos se han de hacer en monedas diferentes a la nacional.
- *Factores económicos y políticos.* Una amplia gama de factores económicos y políticos pueden afectar a la capacidad de una aseguradora para hacer frente al pago de siniestros; entre ellos podemos distinguir: la inflación y su impacto en los resultados o en los fondos, la situación bursátil que puede afectar al valor de las inversiones y provisiones,

cambios legales que lleven como consecuencia un aumento de las reclamaciones.

- *El conjunto de la industria aseguradora de un determinado país.* Es necesario analizar la competitividad dentro del sector y su capacidad financiera.

Un segundo grupo de factores de interés serían los denominados *factores corporativos*, es decir todos aquellos que pueden afectar a un asegurador o reasegurador considerado de manera individual. En otras palabras, aquellos que pretenden medir la calidad de un asegurador o reasegurador a partir de factores cuantitativos (ratios, etc.) y cualitativos (todos aquellos factores no financieros que pueden afectar a la capacidad o voluntad de pagar). En este segundo grupo destacaríamos:

- *La estructura de grupo al que pertenece la compañía de seguros que estamos considerando.* Desde el punto de vista del análisis de la solvencia la propiedad puede dividirse en tres grupos:
 - a) Grupo asegurador cotizado en bolsa donde el negocio asegurador es la principal actividad y la propiedad no está concentrada. En este caso habrá que centrarse no tanto en los cambios de la propiedad sino en los directores y gestores, sus cambios y su trayectoria.
 - b) Grupo privado asegurador. Aquí al no estar tan “diluida” la propiedad sí es necesario ver los cambios en la misma y la presión que la propiedad ejerce sobre los directivos y ejecutivos, así como su conocimiento del sector asegurador.
 - c) Grupo no asegurador ya sea cotizado en bolsa o privado. Aquí es importante desde el punto de vista de la solvencia analizar la marcha del negocio principal, ya que si éste sufre una recesión, no estará en disposición de respaldar a la compañía aseguradora (si ésta se halla en crisis) por no ser la actividad aseguradora su actividad principal.
- *Los directivos.* El equipo directivo de una empresa aseguradora o reaseguradora varía enormemente en cuanto a experiencia, competencia y grado de responsabilidad. Cualquier analista debe prestar atención a la composición del equipo directivo, sus cambios, su trayectoria en la compañía, el tiempo que llevan en la misma o dentro del sector, etc. Es, además, necesario comprobar cuál es la capacidad del equipo directivo en relación a los principales competidores.
- *La imagen.* Cada asegurador tiene una imagen frente a sus clientes, reaseguradores y con el mundo del seguro en general. La imagen que se tenga de una compañía puede ser un factor positivo a la hora de valorar la

solvencia o no. Lo importante es considerar si esta variable es relevante para una compañía a la hora de analizar su solvencia y qué factores pueden dañar seriamente dicha imagen.

- *La estrategia de mercado.* Al igual que ocurre en otros mercados, en éste cada compañía compite por su cuota de mercado y su éxito en esta estrategia depende de factores que incluyen cualquier mínima ventaja en su habilidad para hacer frente a los siniestros, su relación con los clientes y la tenencia de cuentas históricas, la efectividad de un programa de reaseguro, un buen marketing, etc. Todas estas cuestiones se deben tener en consideración de cara a evaluar futuros resultados financieros y la probabilidad de que éstos realmente acaezcan, es decir, una mala política de negocio es un indicador de que en cualquier momento los resultados pueden caer y afectar al pago de siniestros y en consecuencia, a la solvencia.
- *Los agentes de seguros.* Es necesario un adecuado control tanto de los resultados que producen como de la imagen que dan de la compañía.

Finalmente, el tercer grupo de factores que tenemos que tener en cuenta son los *factores específicos* de la compañía. Dentro de esta categoría estarían:

- *Los fondos propios.* Desde el punto de vista de la solvencia, es necesario que los mismos sean adecuados ya que estos fondos pueden ser exigidos ante una eventualidad derivada de una mala tarificación, depreciación del capital o una mala política reaseguradora.
- *La rentabilidad.* Existen tres factores que ejercen una gran presión sobre la solvencia y sobre los fondos propios y que pueden ser medidos en términos financieros utilizando ratios.
 - a) Rentabilidad del negocio. Se puede medir a través del ratio combinado.
 - b) Rentabilidad de las inversiones. Existen varios ratios para medir la rentabilidad de las inversiones y que son susceptibles de ser, en consecuencia, utilizados. Por citar algunos de ellos: los ingresos financieros sobre activo de inversiones, los ingresos financieros sobre gastos técnicos directos y los ingresos financieros sobre primas y recargos del seguro directo.
 - c) La depreciación o revalorización del capital. Se puede medir a través del cambio en el valor de las inversiones (valor de mercado normalmente) sobre el valor original de las mismas (es decir, el valor al principio del periodo que estemos considerando).

- *La viabilidad financiera.* La continuidad con éxito de las operaciones financieras exige la generación de beneficios suficientes para satisfacer (después del pago de impuestos) las expectativas de los accionistas en forma de dividendos y obtener suficientes reservas para aumentar el capital base. Como mínimo debería generarse suficiente beneficio para mantener el capital base en un crecimiento al menos igual a la inflación. Por tanto, es particularmente importante desde el punto de vista de la solvencia el análisis de la tendencia de los resultados, ya que cualquier caída o variación en su continuidad es una señal de peligro.
- *El margen de solvencia.* El margen de solvencia obligatorio normalmente toma un valor arbitrario y de acuerdo a una fórmula legislada. El analista debe recalcular ese margen y comprobar si al menos cumple con los requisitos mínimos legales o por el contrario excede de dichos requisitos.
- *El reaseguro.* En cualquier análisis de la solvencia de una compañía de seguros una variable clave es el reaseguro.

Este tercer grupo de factores a tener en cuenta, como vemos, se pueden medir directamente a través de ratios financieros proporcionando así un marco para el análisis financiero. Es muy importante de cara a cualquier análisis financiero el estar familiarizado tanto con los ratios como con los factores que subyacen bajo ellos y que les afectan. La comparación de los cambios anuales en estos ratios ayudará a determinar la tendencia y a través de ésta determinar la proyección de la empresa. Cuando se analizan los valores de varios años de los ratios es necesario comprobar la tendencia de las pérdidas y ganancias del negocio asegurador, la rentabilidad de las inversiones, y el riesgo de tipo de cambio.

Una vez que se han considerado los factores que podríamos considerar relevantes para el análisis de la solvencia de una entidad aseguradora, éstos deberían ser incorporados a cualquier modelo que queramos utilizar para predecir el fracaso empresarial en forma de variables, ya sean de tipo cualitativo o de tipo cuantitativo. Sin embargo, y ante la imposibilidad de recabar información sobre los dos primeros grupos de factores en nuestro trabajo sólo incluimos el último grupo. No obstante, queremos destacar la idoneidad potencial de las metodologías que empleamos ya que permiten la incorporación de variables de tipo cualitativo. En futuros trabajos trataremos de incorporar, si se dispone de información, factores adicionales que consideremos relevantes y de esta forma mejorar el modelo aquí presentado.

Veamos cuáles son las principales *fuentes de información* de las compañías aseguradoras. Podríamos enumerar las siguientes: a) *Los datos suministrados por la propia compañía*, particularmente sus cuentas anuales y los estados complementarios que las acompañan (incluyendo los estados financieros de carácter interno); b) *Los informes remitidos por la compañía aseguradora a la autoridad supervisora*. Normalmente no son públicos pero si se dispone de ellos, porque los facilite la

compañía, ofrecen una gran ventaja para cualquier análisis ya que presentan una mejor consistencia y preparación, e incluso un mayor detalle que las cuentas presentadas y publicadas de la compañía; c) *Otros informes oficiales*, principalmente los datos suministrados a las rectoras de las bolsas; d) *Cuestionarios o charlas informales y formales con la compañía*; e) *Material publicado*, principalmente comentarios sobre la compañía y el mercado en el que opera tanto en revistas del sector como en la prensa general y financiera; f) *Datos sobre la solvencia realizados por especialistas*, mediante servicios de suscripción (por ejemplo, Standard & Poor, etc.); g) *Análisis de los corredores de bolsa, de los departamentos de inversiones de las compañías aseguradoras, etc.*

En nuestro análisis no hemos tenido acceso a la totalidad de las fuentes de información que hemos mencionado: nuestras fuentes de información han sido, fundamentalmente, los balances y las cuentas de pérdidas y ganancias de las empresas; considerado relevantes, de cara al análisis, los ratios.

1.2.1.- Enfoque metodológico seguido en la selección de los ratios.

En línea con el planteamiento financiero de crisis empresarial, y de acuerdo con nuestro objetivo de predicción, debemos:

- a- en primer lugar, efectuar un análisis de las causas, como paso previo y fundamental para intentar detectar las señales tempranas de alerta de posibles insolvencias
- b- y, en segundo lugar, dado que estos indicios suelen tener su reflejo en la información financiera, hacer un análisis de la misma al objeto de seleccionar aquellas variables que puedan ser introducidas en el modelo.

Las causas del fracaso de las empresas son muy diversas y generalmente son varias las que concurren para llegar a una situación de crisis. En ocasiones el origen del problema es externo (indicios del entorno); también puede ser interno (indicios internos) y a menudo influyen ambos tipos, aunque siempre hay que señalar que con independencia del número y origen de los problemas el equipo directivo de la empresa tiene la responsabilidad de abordarlos y encontrar soluciones. Lo que resulta innegable es que las causas o factores del fracaso empresarial son muy numerosos y/o variados. Estas señales pueden ayudarnos a predecir -en primer lugar- para -en segundo lugar- prevenir; así pues, es necesario analizarlas cuando se presentan.

Como resultado, y desde nuestra mentalidad matemática, pensamos que es muy difícil establecer una relación funcional tipo causa-efecto, aunque en principio pudiera parecer el enfoque más directo. Esto nos conduce (al igual que en otros muchos trabajos de investigación) a plantear un acercamiento fenomenológico, abandonando discusión pormenorizada de las causas para centrarnos en la consideración de las consecuencias de los problemas indicados en la información económica y financiera de las empresas. Nuestra intención consiste en detectar indicios o indicadores en dicha información que permitan posteriormente inferir, explicar o predecir el fracaso empresarial.

Por tanto, el fracaso empresarial (como culminación de un proceso de insolvencia) surge como resultado de un proceso económico complejo con múltiples causas y aparece en las consecuencias que tienen su reflejo en los estados financieros presentados por las sociedades (balance, cuenta de resultados y memoria). Así pues, tanto el análisis interno como el externo pueden contribuir a diagnosticar anticipadamente el riesgo de fracaso de una empresa. Constituyen, por tanto, una herramienta esencial para los distintos usuarios de la información financiera.

Hay que añadir otra dificultad a las mencionadas anteriormente, y es que la predicción está envuelta en una gran incertidumbre derivada, en este caso y según lo expuesto en el párrafo precedente, de factores tales como la inexistencia de información fiable y oportuna sobre la situación y perspectivas de la empresa y el

sector, o el hecho de que existan acontecimientos imprevisibles que lleven la quiebra de manera indetectable y sin reflejo en la información financiera (Jiménez et al., 2000).

Todo esta complejidad hace difícil que exista como tal una “Teoría de sobre la insolvencia o fracaso empresarial”, por lo que las investigaciones acuden al terreno de lo empírico para ver qué efectos tienen en la información financiera las situaciones de insolvencia. Y, todo ello, y de acuerdo con lo que hemos indicado, considerando que el deterioro financiero no sobrevenga de manera repentina sino que se produzca de manera gradual hacia una situación cada vez más grave.

Si se efectúa un primer análisis de la información financiera relativa a las empresas insolventes (en el sentido explicado), nos encontramos una serie de variables que están relacionadas con la insolvencia: pérdidas, flujos de caja negativos, endeudamiento alto -o mal estructurado- que conduce a una carga financiera excesiva, fondos propios insuficientes, fondo de maniobra negativo, incumplimiento o no-renovación de deudas al vencimiento, etc. Estas variables nos informan de problemas presentes y futuros; aún así, la experiencia nos demuestra que no son razones suficientes para el fracaso. En consecuencia, es necesario incorporar al análisis *el grado de intensidad del problema*. En este nivel de intensidad estaría una primera explicación de por qué hay muchas empresas que tienen alguno de los problemas citados, con carácter circunstancial o en períodos largos, e incluso de forma permanente (fondo de maniobra negativo de los grandes almacenes), sin que peligre su futuro.

Por tanto, para caracterizar las empresas desde el punto de vista de la predicción de la insolvencia, es necesario identificar las variables *más* afectadas por la situación y cuantificarlas de forma que resulten discriminantes para establecer el comportamiento propio de las empresas solventes y el de las no solventes.

Los estados financieros contables de las empresas se consideran como una de las más valiosas fuentes de datos para la previsión de las crisis. Su análisis nos permite lograr nuestro primer objetivo: *identificar las variables*.

A través de los estados contables se examina información económica y financiera que caracteriza a una empresa, para poner de manifiesto sus aspectos estructurales y sus movimientos financieros y, de esta manera, poder emitir un diagnóstico sobre ella. Sin embargo, aún cuando el análisis de los estados financieros permite obtener conclusiones sólidas sobre la rentabilidad y la solvencia de una empresa y algo más limitadas para la liquidez, no suelen ser suficientes para poder analizar adecuadamente la evolución. En este sentido, no hay que olvidar que los estados financieros en general y la contabilidad en particular no nacieron ni están pensados con el objetivo básico de que un tercero pueda analizar adecuadamente la situación actual y mucho menos la futura. Son principalmente, el medio que tiene los administradores para rendir cuentas de su actuación ante terceros.

Una vez identificadas las variables en los estados financieros la forma que presenta más ventajas de tratar esas variables y cuantificarlas es mediante el cálculo de determinados ratios financieros en los que intervengan las mismas.

Antes de dar unas breves notas sobre los ratios financieros, queremos ya manifestar que cualquier modelo que sólo utilice la información contenida en los estados financieros (máxime cuando se utilizan sólo el balance y la cuenta de pérdidas y ganancias) presenta una serie de limitaciones. Pero, sin embargo, estos datos que podríamos denominar *externos*, son datos disponibles para cualquier interesado, por ser datos de obligada publicidad. Con lo que es fácil desarrollar un modelo (aunque limitado) en función de los datos disponibles para poder mejorarlo mediante la incorporación de otra información, si se dispone, de carácter interno y/o no publicable y añadir dichas variables al modelo. Por otro lado, y hablando de los ratios, la utilización de los mismos permite comparar con otros valores que, el mismo ratio, haya tomado en otros tiempos (análisis histórico o de tendencia) o con ciertos estándares propuestos, tales, como por ejemplo, la media del sector industrial (análisis normativo). El análisis histórico o de tendencia lleva al estudio de las series temporales de ratios de una misma empresa cuyo objetivo suele ser la predicción del valor futuro de los mismos (análisis predictivo). Cuando realizamos un análisis de la crisis empresarial mediante ratios contables, esperamos que los ratios, de un modo u otro, se deterioren conforme se aproxima la fecha de la quiebra y que dicho deterioro sea perceptible o medible en un intervalo de tiempo relativamente amplio, a veces de hasta cinco años (Calvo-Flores y García (coord.), 1998b)

Los modelos de predicción de la insolvencia que obtengamos, en relación con los ratios contables, cumplen una doble misión:

- a) prueban el contenido informativo de los datos contables sobre la solvencia de la empresa, al medir la relación entre los datos y el fracaso,
- b) y, en segundo lugar, derivado del anterior, son un mecanismo para pronosticar el fracaso empresarial. Constituyen una herramienta útil para la toma de decisiones por parte de los múltiples usuarios del análisis de la solvencia.

Queremos anticipar que los indicadores o variables financieras relacionados con la medición de la insolvencia son muy variados, habiendo sido contrastados muchísimos. Es frecuente en numerosos estudios empíricos partir de una gran cantidad de ratios para que el modelo propuesto reduzca el número de los mismos. En nuestro caso, el enfoque metodológico que hemos utilizado para encontrar la relación entre las causas o síntomas y situaciones de insolvencia crediticia o total ha consistido en seleccionar diversas variables consideradas a priori como apropiadas (para las cuales justificaremos su utilización), para contrastarlas utilizando nuestros métodos. Es decir, hemos considerado no introducir una gran cantidad de ratios sino

previamente seleccionar una serie de ellos. Pensamos que introducir mucha información puede incidir negativamente en los objetivos propuestos, por lo que preferimos no introducir información superflua en el modelo, que, lejos de contribuir a los fines predictivos, los dificulte. Este enfoque permite seleccionar las variables más significativas para explicar la situación en función de los datos manejados, así como identificar si las conclusiones obtenidas son eficientes también cuando se aplican a otras empresas de la primera muestra estudiada, y a los datos correspondientes a las empresas uno, dos, tres o más años anteriores al del estudio inicial (año de fracaso).

I.2.2.- Selección de la muestra.

La muestra utilizada es la seleccionada para la aplicación del análisis discriminante en la predicción de la insolvencia en entidades aseguradoras españolas no vida (Sanchis, 2000 y 2003). Consta, en el ámbito nacional, de 36 empresas sanas y 36 empresas fracasadas, emparejadas fundamentalmente por tamaño y sector, para minimizar el efecto de estas variables en el estudio sobre los ratios que haga que diferencias en alguno de ellos se deban a otros factores ajenos a su situación financiera. Aunque para un mayor detalle sobre la selección de la muestra se puede consultar la bibliografía citada anteriormente, queremos resaltar brevemente las características de la muestra.

La definición de *empresa fracasada*, tal y como ya se ha indicado, es aquella empresa que fue intervenida por la C.L.E.A. Todas ellas son Sociedades Anónimas, por tanto se han excluido del análisis otro tipo de formas societarias (Mutuas, Cooperativas).

El tipo de muestreo ha sido por *emparejamiento* controlado por unos determinados factores. Con el *muestreo por parejas*, se pretende que las dos submuestras (empresas sanas y empresas fracasadas) tengan características similares para aquellos factores que no son objeto del estudio con la finalidad de que, cuando se obtengan los resultados de clasificación, se puedan atribuir estos al factor que queremos explicar: el fracaso de las empresas. Por tanto, la muestra está formada por el mismo número de empresas fracasadas que sanas y son muestras no demasiado heterogéneas ya que el emparejamiento ha tenido en cuenta el factor tamaño (medido a través del volumen de primas) y el tipo de negocio.

La información utilizada en el planteamiento de este trabajo, ha sido eminentemente contable. Estos datos están recogidos en los estados financieros de las empresas, de los cuales sólo se ha podido obtener el Balance y la Cuenta de Pérdidas y Ganancias. Los datos están extraídos de la publicación anual de "Balances y cuentas. Seguros privados" de la Dirección General de Seguros. La utilización de documentos contables permite obtener una información cuantificable, objetiva y sistemática de la realidad económica y financiera de la empresa.

Para ver el poder explicativo de los ratios, se han tomado los datos a lo largo de cinco años antes de la quiebra, tomando como año base el primer año anterior a la misma en el caso de las fracasadas, y por extensión, también para su pareja ese será el año base. Sin embargo, dado que en 1982 se introdujo una nueva estructura en las cuentas anuales, no se ha utilizado información anterior a esa fecha para que la misma fuese lo más homogénea posible. Por tanto la serie temporal abarca datos desde el 83 al 94.

Las empresas utilizadas en el análisis aparecen en las Tablas 1.1 y 1.2.

TABLA 1.1: EMPRESAS INTERVENIDAS POR LA CLEA

Nº	NOMBRE	CODIGO	AÑOS
1	Kairos, cia. De seguros y reaseg.	C-043	93-92-91-90-89
2	Igualatorio Médico Palentino de seguros	C-130	93-92-91-90-89
3	Asistencia Sanitaria 2000 (1)	C-454	93-92-91-90-89
4	Sociedad Andaluza de Seguros	C-507	93-92-91-90-89
5	Conseguar, S.A. de Seguros Generales	C-598	93-92-91-90-89
6	Unión social de Seguros	C-638	92-91-90-89-88
7	Mundi Seguros (2)	C-663	92-91-89
8	Apolo	C-008	91-90-89-88-87
9	Unión Europea de Seguros	C-568	91-90-89-88-87
10	Segurauto	C-573	91-90-89-88-87
11	Reunión Grupo 86	C-440	90-89-88-87-86
12	Servicios Médicos	C-450	90-89-88-87-86
13	Larra	C-561	90-89-88-87-86
14	Unasyr, Unión alicantina de Seguros	C-567	90-89-88-87-86
15	Mades Fondo Asegurador(3)	C-664	90-89-88
16	Técnica Aseguradora	C-352	90-89-88-87-86
17	Unión peninsular de seguros	C-555	90-89-88-87-86
18	Mas Grupo 86 Fondo Asegurador	C-581	89-88-87-86-85
19	Munauto (4)	C-608	89-88-87
20	Unión Ibérica Grupo 86	C-523	89-88-87-86-85
21	Sociedad occidental de Seguros (5)	C-615	89-88
22	España Vitalicia	C-071	89-88-87-86-85
23	Instituto Médico Quirúrgico	C-422	88-87-86-85-84
24	Madrid, S.A. de seguros Generales	C-111	86-85-84-83
25	Cía. Mercantil de Seguros	C-560	86-85-84-83
26	Médica Riojana	C-460	86-85-84-83
27	Igualatorio Vallisoletano Médico Quirur. y de Espec.	C-328	86-85
28	Igualatorio Médico Ntra. Sra. Del Rosario	C-321	86-85-84
29	Argüelles	C-289	86-85-84-83
30	Palace	C-250	85-84-83
31	Asociación clínica Española	C-283	84
32	Cosmos	C-564	84-83
33	Alianza Previsora	C-220	84-83
34	Clínica Argüeso	C-224	84-83
35	Labor	C-330	84-83
36	La Gloria Eterna	C-474	84-83

Notas

- (1)- Asistencia Sanitaria 2000, antes Policlínica Santiago
- (2) - Mundi Seguros empezó a operar en noviembre del 88
- (3)- Mades Fondo Asegurador, antes mutua.
- (4)- Munauto, antes Mutua
- (5)- Sociedad Occidental de Seguros, solo operó dos años

TABLA 1.2. EMPRESAS SANAS

Nº	NOMBRE	CODIGO	AÑOS
101	Metrópolis	C-121	93-92-91-90-89
102	Igualatorio Médico Leonés	C-403	93-92-91-90-89
103	Alergia	C-286	93-92-91-90-89
104	Seguros Mercurio	C-630	93-92-91-90-89
105	Génesis Seguros Generales (1)	C-695	93-92-91-90
106	Lagún-Aro	C-572	92-91-90-89-88
107	La Unión Alcoyana	C-188	92-91-90-89-88
108	Athena, cía. Ibérica de seguros y reaseg. (2)	C-228	91-90-89-88-87
109	Lepanto (3)	C-108	91-90-89-88-87
110	Federación Ibérica (4)	C-076	91-90-89-88-87
111	La patria Hispana	C-139	90-89-88-87-86
112	Asociación Médica conquense	C-313	90-89-88-87-86
113	Aseguradora Universal	C-012	90-89-88-87-86
114	Sur	C-186	90-89-88-87-86
115	Munat (5)	C-665	90-89-88
116	Europa Seguros Diversos (6)	C-508	90-89-88-87-86
117	Hispano Alsaciana (7)	C-061	90-89-88-87-86
118	Nortehispania	C-275	89-88-87-86-85
119	Andalucía y Fenix agrícola (8)	C-004	89-88-87-86-85
120	Compañía Astra	C-468	89-88-87-86-85
121	La Alianza española	C-002	89-88-87-86-85
122	La humanitaria (9)	C-318	89-88-87-86-85
123	Boreal Médica	C-027	88-87-86-85-84
124	ADEA (10)	C-378	86-85-84-83
125	ASEFA	C-522	86-85-84-83
126	Igualatorio Médico Quirurg. Pilarista (11)	C-390	86-85-84-83
127	Asistencia Clínica Univ. De Navarra	C-325	86-85-84
128	Sanitaria Médico Quirúrgico	C-515	86-85-84-83
129	La Antártida	C-506	86-85-84-83
130	La Gloria	C-229	85-84-83
131	Federación Médica	C-434	84-83
132	Le Mans S.E.	C-552	84
133	Clinos	C-226	84-83
134	Salus, Asistencia Sanitaria	C-485	84-83
135	El Paraíso Universal	C-238	84-83
136	Seguro Europeo	C-319	84-83

Notas

- (1)- Génesis, año 91 primer balance completo. Cenit sustituye a Génesis
- (2)- Athena sustituye a DAPA
- (3)- Lepanto sustituye a Regal Insurance
- (4)- Federación Ibérica sustituye a AMIC
- (5)- Munat, antes Mutua
- (6)- Europa Seguros Diversos, antes Previsur
- (7)- Hispano Alsaciana, absorbe Azur y cambia de denominación
- (8)- Andalucía y Fenix Agrícola, actual Hispasalud
- (9)- La Humanitaria, actual Al-Andalus
- (10)- ADEA, transformación en Gan-España, C-686
- (11)- Igualatorio M. Q. Pilarista, actual Pilarista

1.2.3.- Análisis del balance y la cuenta de resultados de las entidades aseguradoras

Tras la obtención de la muestra de empresas hemos procedido a la homogeneización de la información contable para tener una reclasificación operativa de los estados contables.

Es necesario mencionar que, derivado del periodo que comprende la muestra, tanto el balance como la cuenta de resultados que vamos a reclasificar se corresponden con el modelo vigente hasta 1997 (Plan General de Contabilidad adaptado a las entidades de seguros, reaseguros y capitalización aprobado por la O.M. de 30 de julio de 1981). En la actualidad el Plan de Cuentas que contiene los modelos de Cuentas Anuales que están en vigor es el aprobado por el Real Decreto 2014/1997.

También queremos resaltar que, dadas las peculiaridades que presenta el sector asegurador, existen ciertas dificultades conceptuales a la hora de asignar determinadas partidas al circulante o al fijo (por ejemplo las provisiones técnicas o las inversiones). Este es el motivo por el cual, aunque en nuestro trabajo hemos seguido el criterio tradicional de partir de la división del activo y pasivo de una sociedad aseguradora según el balance propuesto por el plan sectorial de contabilidad y ordenar el balance en distintas masas patrimoniales con el criterio de menor a mayor liquidez en el activo y de menor a mayor exigibilidad en el pasivo, tal y como expondremos a continuación, caben otros criterios a la hora de clasificar las partidas entre fijo y circulante que afectarían a la definición de las masas patrimoniales y, en consecuencia a la definición de los ratios (Linares, 1998, p. 45-50).

La utilización de otras clasificaciones (Millán, 2000, p.82-85; Fernández-Palacios y Maestro, 1991, p. 518-527; García y Ballester, 1994, p. 162-163) dependerá del usuario del modelo y afectará a los resultados del modelo y a su interpretación, pero el objeto de este trabajo es mostrar la utilidad de la utilización de métodos de la inteligencia artificial en la predicción de las solvencia en entidades aseguradoras con independencia de los criterios utilizados en la elaboración de los datos de entrada de los modelos, ya sean éstos contables o no.

Reclasificación del Balance

Activo

I. Accionistas II. Inmovilizado III. Inversiones 1. Materiales 2. Financieras 3. Empresas del grupo	Activo Fijo I. Accionistas II. Inmovilizado III. Inversiones 1. Materiales 2. Financieras (1) 3. Empresas del grupo
IV. Provisiones de reaseguro V. Créditos	Circulante exigible IV. Provisiones de reaseguro V. Créditos
VI. Ajustes	Circulante realizable III. Inversiones 2. Financieras (2) VI. Ajustes
VII. Efectivo	Circulante Disponible VII. Efectivo

(1) Dentro de las inversiones financieras sólo las de control (cobertura de provisiones técnicas) no considerando las especulativas. Si las inversiones en empresas del grupo cotizasen en el mercado organizado pasarían a circulante exigible.

(2) Sólo las que coticen en mercados organizados.

Nota: A los efectos del cálculo de los ratios y debido a que no tenemos el desglose de las inversiones (únicamente disponemos del dato que figura en el balance) las incluiremos siempre como parte del activo circulante, es decir como si no existiese cartera de control o que todas las inversiones cotizaran en mercados organizados.

Pasivo

	Neto Patrimonial
I. Capitales propios	I. Capitales propios
II. Provisiones técnicas 1. Riesgos en curso 2. Matemáticas 3. Para prestaciones 4. Otras	Exigible a largo plazo II. Provisiones técnicas 1. Riesgos en curso 2. Matemáticas 4. Otras III. Provisión de responsabilidades IV: Depósitos de reaseguros V. Deudas (1)
III. Provisión de responsabilidades IV: Depósitos de reaseguros V. Deudas	
VI. Ajustes	Exigibles a corto plazo II. Provisiones técnicas 3. Para prestaciones V. Deudas a corto plazo VI. Ajustes

(1) Deudas a largo plazo

Nota: A los efectos del cálculo de los ratios y debido a que no tenemos el desglose de las deudas (únicamente disponemos del dato que figura en el balance) consideraremos que son a largo plazo los empréstitos, las deudas con empresas del grupo y las deudas con empresas asociadas. Es necesario mencionar que las deudas y establecimientos de crédito (epígrafe V.2.) han sido incluidos en el corto plazo pero podrían haberse considerado a largo plazo. La arbitrariedad viene derivada de, como en el caso anterior, la falta de desglose de las deudas.

Una vez estructurado el balance de la sociedad, se puede efectuar un análisis financiero comparativo (con una empresa modelo, una empresa ideal o con un balance agregado sectorial) o calcular ratios.

1.2.4.- Elección de las variables: Los Ratios Contables

La elección de las variables más adecuadas a utilizar en la elaboración del modelo de predicción es parte fundamental del éxito final del mismo (García *et al.*, 1997). Además, al tratarse de un modelo de predicción del fracaso empresarial se deben considerar posibles variables significativas que discriminen la insolvencia. De esta forma, hemos configurado un conjunto de variables, que, a nuestro juicio, y, desde un punto de vista teórico y práctico, forman un modelo integral de análisis financiero de una entidad aseguradora.

Aunque las metodologías empleadas en este trabajo permiten el uso de variables cualitativas, y pensamos que *a priori* puede ser muy positiva su introducción en cualquier modelo, no las hemos considerado debido a la imposibilidad de conseguir los correspondientes datos. En cuanto a la selección de los ratios, hemos acudido a los más utilizados en el análisis financiero para el pronóstico de la insolvencia de entidades aseguradoras, teniendo en cuenta nuestro criterio personal, dadas las propias características del sector seguros y teniendo en cuenta un análisis preliminar de las particularidades de los dos grupos de empresas (sanas y fracasadas). Por otra parte, la consideración de muchas variables en un modelo puede acarrear el efecto contrario al que buscamos y, en vez de incrementar la capacidad predictiva, se introduce incertidumbre (o *ruido*) en el modelo que normalmente disminuye dicha capacidad. Pretendemos, con las metodologías que emplearemos, obtener el más pequeño subconjunto de ratios con el mayor poder discriminante en relación con el fracaso empresarial.

Es necesario recalcar que las peculiaridades sectoriales de las empresas de seguros son tan significativas que no resultan aplicables, al menos parcialmente, los modelos y contenidos de las Cuentas anuales elaboradas para empresas industriales y comerciales (Prieto, 1992, p.15). Por tanto, junto con los ratios aplicables al conjunto de empresas industriales y comerciales se consideran todos aquellos específicos para el sector asegurador.

La inversión del proceso productivo que caracteriza al sector asegurador implica la determinación *a priori* del precio del seguro o de la prima en base a los valores esperados tanto de la siniestralidad como de los gastos. Al hablar de valores esperados, cabe la posibilidad de que se produzcan desviaciones negativas entre los datos reales y los previstos. Por tanto la empresa debe disponer de unos fondos patrimoniales o margen de solvencia para hacer frente a las posibles desviaciones. El margen de solvencia se convierte, de este modo, en una variable clave en cualquier estudio que implique la viabilidad de las empresas del sector asegurador. En nuestra propuesta de ratios existen varios que incluyen dicha variable en su definición pero, debido al período que abarca la muestra, dicho dato se encuentra en los estados confidenciales remitidos a la Dirección General de Seguros que no son públicos y, en consecuencia, el acceso a ellos con fines de investigación es imposible. Este es el motivo por el que en ninguno de nuestros ratios aparece el margen de solvencia pero creemos es una variable clave que en muestras que abarquen periodos posteriores en los que dicho dato ya es público debería aparecer en al menos uno de los ratios.

A lo largo de la literatura contable se han propuesto muchos ratios, tanto generales como específicos para el sector asegurador, y por tanto, podrían haberse utilizado otros alternativos ¹a nuestra elección. A continuación vamos a describir brevemente dichos ratios y justificar por qué forman parte de los modelos.

¹ Hay ratios financieros convencionales que se utilizan en muchos estudios sobre solvencia empresarial y que no aparecen en nuestro trabajo. Por citar alguno de ellos estaría el ratio Activo circulante/Pasivo Circulante o Activo Fijo/Deudas a largo plazo pero por las razones que indican García y Ballester (1994), carecen de utilidad operativa para un correcto análisis de la solvencia empresarial por basarse en una concepción patrimonialista de la empresa, es decir sin tener en cuenta el principio de gestión continuada y comparando partidas, si bien homogéneas en cuanto al periodo medio de maduración, son heterogéneas para el análisis financiero.

Ratio del capital circulante o fondo de maniobra sobre total activo

$$R1 = \frac{\text{Fondo de Maniobra}}{\text{Total Activo}}$$

En ocasiones recibe la denominación de liquidez la relación existente entre el capital circulante y el activo total de la empresa. El objetivo de este ratio es medir la liquidez neta de los activos sobre el total de la capitalización, significando la importancia que representa el capital circulante en relación con el total activo y, por tanto, la garantía que ofrece la empresa con su activo circulante para afrontar a un pasivo de igual grado (Álvarez y Urías, 1987).

Cuanto mayor sea el valor de este ratio indicará una mejor situación respecto a la liquidez. El problema de liquidez es bastante importante en una empresa aseguradora ya que dada la inversión del proceso productivo los cobros se realizan antes que los pagos luego no debería existir este problema. Además si una empresa aseguradora no es capaz de pagar los siniestros que se producen puede acarrear una pérdida de la confianza de sus clientes.

Además, la empresa debe ofrecer un equilibrio entre su estructura económica y financiera. Para obtener este equilibrio las fuentes financieras deben ajustarse, de acuerdo a un plan financiero, a las inversiones. Los elementos acíclicos y circulantes necesarios para la continuidad del ciclo de explotación tienen que estar financiados con recursos propios y recursos ajenos a largo plazo en su correcta proporción. El desequilibrio financiero más frecuente, surge, generalmente, del inadecuado diseño de la financiación de los inmovilizados, lo que puede originar a corto plazo serios problemas de liquidez o tensiones financieras innecesarias (García *et al.*, 1997, p. 141).

El equilibrio financiero lo mediremos a través del cálculo del fondo de maniobra (diferencia entre los activos circulantes y pasivos circulantes), y para facilitar la comparación determinamos el porcentaje que representa sobre el total del activo.

Cuando el fondo de maniobra es positivo, significa que parte del activo circulante está siendo financiado con recursos permanentes, ya sean propios o ajenos a largo plazo. Por el contrario, un capital circulante negativo, implica que una parte del inmovilizado está siendo financiado por deudas a corto plazo.

Ratio de rentabilidad financiera de los recursos propios

$$R2 = \frac{\text{Beneficio antes de Impuestos}}{\text{Capitales Propios}}$$

Con carácter general, la rentabilidad financiera de una empresa trata de comparar el rendimiento obtenido, a lo largo del ejercicio económico considerado, con los capitales propios invertidos (Rivero, 1987, p. 239-240).

El marco genérico en el que se envuelve toda empresa gira en torno a dos premisas fundamentales, asegurar su funcionamiento en el futuro bajo hipótesis de normalidad, y conseguir que ésta se desarrolle en las mejores condiciones de rentabilidad posibles. Para la empresa, el objetivo del beneficio debe cumplir una doble función, por un lado, retribuir convenientemente a los suministradores del capital, y por otro capitalizar la empresa de forma que se asegure un correcto equilibrio financiero (García *et al.*, 1997, p. 146).

En consecuencia, para una empresa es necesario conseguir una adecuada rentabilidad financiera no sólo por razones de imagen que le permitan colocar las ampliaciones de capital o conseguir créditos sino también porque un beneficio suficiente en una firma con autonomía financiera lleva consigo el mantenimiento de la misma.

Este ratio informa sobre la eficiencia de la empresa en la utilización de los capitales invertidos por accionistas (o de los acreedores ya que también se considera como rentabilidad financiera al ratio que relaciona el beneficio con el pasivo).

Si la rentabilidad financiera es mayor que cero, se dice que existe una rentabilidad financiera *positiva* en relación con los recursos propios (o con el pasivo, según se considere).

Además, de acuerdo con el estudio realizado por Gabás (1997, p. 26), en el que se analizan distintos trabajos empíricos sobre predicción de la insolvencia empresarial, la variable más repetida en cuanto a eficiencia en su capacidad de predicción es la rentabilidad o capacidad de obtener beneficios por lo que habría que destacar su importancia desde el punto de vista de aportación a una teoría general de la solvencia.

Ratio de ingresos financieros por activo de inversiones

$$R3 = \frac{\text{Ingresos Financieros}}{\text{Activo de Inversiones}}$$

Es un ratio que se considera que puede modificar la posición de solvencia ya que informa sobre la eficiencia en el empleo e inversión de los recursos financieros y la proporción que éstos suponen respecto a activos. Por tanto expresa la eficiencia en el aprovechamiento de los activos dedicados a inversiones, y en consecuencia, la medida de la bondad en la política de inversiones de los recursos disponibles (Martín *et al.*, 1999, p. 80).

Este ratio cobra aún más importancia si tenemos en cuenta que los rendimientos financieros de las entidades aseguradoras son en la actualidad la fuente principal de sus beneficios, o lo que es lo mismo, representan la posibilidad de obtener una rentabilidad adecuada para los recursos propios y también una fuente de financiación importante de los incrementos del Margen de Solvencia (Prieto, 1993, p. 95).

Además, es uno de los ratios utilizado por la NAIC (National Association Insurance Commissioners) en su sistema IRIS (Insurance Regulatory Information System), sistema de alarma para la detección de posibles insolvencias. Nos indica una medida de la adecuación del rendimiento obtenido por la cartera de inversiones.

Ratio de cash flow sobre total pasivo

$$R4 = \frac{\text{Cash Flow}}{\text{Total Pasivo}}$$

La justificación de utilizar el cash-flow (como recursos generados) viene motivada por la capacidad que tienen las empresas para flexibilizar, alisar e incluso manipular el resultado contable. Por ello, se ha desarrollado una línea de investigación para contrastar si los fondos de tesorería generados son más eficaces para la predicción de la solvencia, ya que son menos manipulables si los comparamos con los resultados. En *casos específicos*, la evolución de los flujos de tesorería es *mucho más explicativa* de los problemas reales de la empresa que el beneficio (Gabás, 1997, p.27).

La caída final que supone un desplome de beneficios a cuantiosas pérdidas, que hemos podido comprobar en algunas cuentas anuales nos hace sospechar de una política de manipulación de los resultados en los últimos años de la vida de las empresas, aunque tanto en nuestro estudio como en otros trabajos de investigación, partimos de la hipótesis de adecuada calidad de la información contable, por lo que si esto no sucede se produce un sesgo imposible de controlar.

Además según se constata en los informes de auditoría que acompañan las cuentas anuales de las empresas en suspensión de pagos, en un alto porcentaje significativo de casos, se dan incumplimientos de principios contables (Roqueta, 2000, p. 2-5). Ante las dificultades económicas, la empresa disminuye o dota insuficientemente las cuentas de amortizaciones y provisiones y suelen realizar operaciones extraordinarias. Estas últimas son las que han motivado que hayamos considerado el cashflow añadiendo los recursos generados por los resultados extraordinarios.

En el trabajo realizado (ver referencia anterior), este ratio detecta precozmente la insolvencia (al situarla por debajo del promedio del mercado, representado por el conjunto de todas las empresas) motivo por el cual hemos considerado incluirlo en nuestro estudio. Además podemos mencionar otros trabajos (Laitinen, 1992, p. 323-340; Karels y Prakash, 1987, p. 573-592), en los cuales este ratio se ha mostrado como el más eficiente a la hora de predecir el fracaso empresarial. El argumento que justifica esta eficiencia en cuanto a su capacidad de predicción es, que una financiación escasa debida a ingresos insuficientes que aparezcan de manera inesperada puede ser la causa del comienzo de proceso de quiebra. Los motivos que la pueden originar son varios: baja rentabilidad, crecimiento demasiado rápido o una velocidad del capital demasiado lento (entendiendo por esta última la capacidad de los activos para generar ingresos).

Ratios de solvencia en sentido estricto

Según hemos visto, la solvencia de una sociedad es un concepto difícil de definir con precisión. El mercado asegurador, al tratarse de un mercado institucional y depender de la confianza del consumidor en la institución y de la eficacia del marco jurídico en el que se desenvuelve, hace que la noción de solvencia en la entidad aseguradora tenga matices propios (Linares, 2000, p.59). De una parte es función de los activos realizables y, de otra, de las obligaciones ciertas que hubiera asumido, pero también de conceptos como la rentabilidad de la cartera o desarrollo futuro de la misma.

En las entidades aseguradoras distinguimos dos tipos de solvencia:

Solvencia estática, que indica la disponibilidad del asegurador de medios suficientes para hacer frente a sus obligaciones contractuales. Este concepto podría, en términos generales, identificarse con el volumen de provisiones técnicas y su correcta inversión.

Solvencia dinámica, que se basa en medir la capacidad de la sociedad para hacer frente, no ya a sus obligaciones contractuales, sino a las futuras que pudiera adquirir.

Se refiere a la capacidad de la empresa para seguir desempeñando en el futuro la función de cobertura de riesgos que le es propia (Fernández-Palacios y Maestro, 1991).

Al analizar una entidad aseguradora debemos considerar la posibilidad de que se produzca una siniestralidad superior a la técnicamente prevista. Es decir, dado que la prima representa el valor medio de la siniestralidad esperada y es una variable aleatoria, ésta puede tomar infinitos valores en torno a ese valor medio que den lugar a que, por mucho que esté bien calculada, la siniestralidad real no coincida con la esperada, sino que en unos ejercicios se halle por debajo y en otros por encima. En este último caso, si la entidad no contara con más recursos que las primas recaudadas, la entidad entraría en insolvencia porque con ellas no podría hacer frente a todos los siniestros (Maestro, 2000).

En consecuencia con lo expuesto en el párrafo precedente, surge además otra necesidad: la de tener un patrimonio o fondo de cuantía suficiente, cuya finalidad sea la de compensar la insuficiencia de los ingresos percibidos por la entidad en relación con los gastos, cualquiera que sea la causa que los provoque (Fernández-Palacios y Maestro, 1991).

Plantear estas posibilidades, es plantear el problema de la solvencia dinámica de una entidad aseguradora.

Por tanto se hace necesario que la entidad disponga de un patrimonio complementario, es decir, un patrimonio no afecto a obligación alguna. Estas garantías financieras adicionales que persiguen que el asegurador cumpla el

principio de solvencia, en su acepción dinámica, son el margen de solvencia y la provisión de estabilización.

$$R5 = \frac{\text{Total Primas adquiridas de seguro directo}}{\text{Capitales Propios}}$$

$$R6 = \frac{\text{Total Primas adquiridas de negocio neto}}{\text{Capitales Propios}}$$

$$R7 = \frac{\text{Total Primas adquiridas de seguro directo}}{\text{Capitales Propios+ Provisiones Técnicas}}$$

$$R8 = \frac{\text{Total Primas adquiridas de negocio neto}}{\text{Capitales Propios+ Provisiones Técnicas}}$$

Estos cuatro ratios (al igual que **R11-R12-R13-R14** que comentaremos a continuación) se pueden considerar como indicadores de solvencia en un momento determinado del tiempo, solvencia estática, y también, estudiando su variación a lo largo de un determinado periodo, serían indicadores de la solvencia dinámica (Martín *et al.*, 1999, p. 77-78).

Estos ratios recogen en sus numeradores la medida de los riesgos anuales, basándose en la cuantía de las primas destinadas a cubrir los riesgos asegurados en el año (primas devengadas en el período en que se trate). Se diferencia entre seguro directo y seguro neto, con el objetivo de comprobar las diferencias entre ambas posiciones y observar su repercusión en la solvencia.

Los denominadores muestran el soporte financiero de las empresas. En los dos primeros se contemplan los capitales propios en sentido estricto lo que sugiere un soporte global del riesgo independiente de las características anuales. En los dos últimos se suman a los capitales propios las provisiones técnicas de las que se dispone (totales o netas según el tipo de gasto técnico con el que se compare), logrando de este modo el soporte financiero real para el periodo analizado.

Estos ratios, que utilizan las primas como una medida del riesgo en que incurre la empresa de seguros, son usados en Estados Unidos por la *National Association of Insurance Comisioners* (N.A.I.C.), dentro de sus ratios de alarma ante una posible situación de insolvencia.

Además los dos primeros ratios (el segundo es el primero expresado neto de reaseguro cedido) se denominan *Ratios de Apalancamiento Asegurado* (Millán, 2000, p. 103-104). Reflejan el apalancamiento del volumen de negocio neto de la compañía en relación con sus recursos propios. Mide el grado de absorción por la compañía de los errores en el precio de sus productos. Estos dos ratios no es aconsejable que presenten una cuantía muy elevada, ya que el nivel de apalancamiento incrementa el retorno del capital (con pocos recursos se está generando un elevado volumen de negocios, *rotación de los recursos permanentes* (Linares, 2000, p. 55), pero también el riesgo de inestabilidad.

$$R11 = \frac{\text{Gastos técnicos de seguro directo}}{\text{Capitales Propios}}$$

$$R12 = \frac{\text{Gastos técnicos de negocio neto}}{\text{Capitales Propios}}$$

$$R13 = \frac{\text{Gastos técnicos de seguro directo}}{\text{Capitales Propios} + \text{Provisiones Técnicas}}$$

$$R14 = \frac{\text{Gastos técnicos de negocio neto}}{\text{Capitales Propios} + \text{Provisiones Técnicas}}$$

Estos cuatro ratios, **R11-R12-R13-R14**, se diferencian de los anteriores en que recogen en sus numeradores la medida de los riesgos anuales, basándose en la valoración de los riesgos que realmente han ocurrido (siniestros del año) y que se recogen en la Cuenta de Pérdidas y Ganancias como Gastos Técnicos. Al igual que en los anteriores, se diferencia entre seguro directo y neto para poder analizar la influencia del reaseguro en la estructura de solvencia.

Ratio de autonomía financiera

$$R9 = \frac{\text{Capitales Propios}}{\text{Pasivo Total}}$$

El objetivo de este ratio es informar sobre la composición estructural de las fuentes de financiación y medir el grado de autonomía o independencia financiera de la entidad (en el sentido de si ante una decisión de inversión puede elegir libremente los recursos que más le interesen) en virtud de la procedencia de las fuentes financieras utilizadas (Rivero *et al.*, 1998, p. 202-203).

La autonomía financiera implica la posibilidad de elegir, por parte de la empresa, de entre las distintas fuentes de financiación, aquellas que considera más adecuadas para acometer la explotación de un proceso productivo. En el caso de no poder tomar los fondos que le resultaran más convenientes por tener cerradas otras fuentes de financiación, la empresa habrá perdido su independencia y no tendrá libertad financiera que garantice la autonomía de sus decisiones. La firma, en este caso, actuará de una manera forzada.

A través de este ratio se intenta conocer cuál es el nivel óptimo (proporción) de endeudamiento de una empresa, o sea, cuál es el grado de independencia con respecto al exterior, es decir, con respecto a terceros. Evalúa la cantidad de recursos que son propiedad de los accionistas. Cuanto mayor sea el ratio que analizamos menor será la probabilidad de insolvencia a largo plazo, así como menores serán las necesidades de la empresa para hacer frente al pago de la carga financiera y al reembolso del principal. En consecuencia, un valor bajo del ratio significa, en principio, que la empresa no estará en condiciones de elegir, ya que los prestamistas entenderán que existe un alto riesgo de impago.

Un alto índice de endeudamiento puede motivar que la firma no tenga capacidad de adecuar las decisiones de inversión y financiación, no pudiendo elegir entre distintas acepciones financieras por haber perdido su autonomía con los riesgos asociados a esta pérdida.

Un esquema correcto de financiación debería permitir la cobertura del inmovilizado con capitales propios y con recursos exigibles a plazos por lo menos iguales a su vida útil, para evitar problemas asociados a la refinanciación y además, atender a las necesidades del fondo de maniobra.

La autonomía financiera está ligada al apalancamiento financiero. Una entidad que tenga autonomía financiera puede optar a los beneficios derivados del apalancamiento financiero positivo, hecho difícil de conseguir para una empresa que haya perdido su autonomía.

Además la independencia financiera también se haya unida a la solvencia, entendida ésta en sentido amplio; esto es, garantía suficiente por los activos y por la generación de flujos de tesorería para hacer frente a la remuneración de los capitales propios y ajenos y a la restitución de estos últimos.

La independencia financiera se alcanza cuando se da una relación óptima entre los recursos propios y ajenos. No obstante, la significación del ratio estará en función de la empresa a que se aplica y del sector donde opere; a nivel general, se suele situar en torno a 1, entre 0,7 y 1,5. Un valor superior a los citados transmite la imagen de que son los propietarios de la empresa los que verdaderamente soportan el riesgo que comporta su negocio. Este dato debe, no obstante, ser contrastado y aceptado o no por cada empresa en función de sus condiciones particulares.

El inverso de la tasa de autonomía financiera es *la tasa de endeudamiento* que informa sobre la relación existente entre el pasivo exigible total y el patrimonio neto.

Ratio de exposición aseguradora

$$R10 = \frac{\text{Provisiones Técnicas}}{\text{Capitales Propios}}$$

Es otro ratio de apalancamiento. Este ratio refleja el grado de apalancamiento de las obligaciones técnicas de la compañía en relación con los recursos propios. Mide el grado de absorción por la compañía de los errores de estimación en sus provisiones. También puede expresarse neto de reaseguro (Millán, 2000, p. 104).

Con este ratio se pretende medir la relación entre recursos administrados y recursos propios. Si bien su estudio se debe efectuar por comparación, un criterio de prudencia puede aconsejar que no sea muy elevado ya que significaría un volumen de recursos administrados (provisiones técnicas) muy superior al de los recursos propios, situándose la empresa en los límites de su capacidad de gestión, con menos margen de maniobra para enfrentarse a eventuales problemas de equilibrio financiero.

Ratio combinado y sus variantes

Los tratadistas del análisis contable aplicado al sector seguros han establecido, con carácter general, un indicador para evaluar la gestión global de la actividad aseguradora denominado ratio combinado (Millán, 2000, p.146-148).

Se define de forma genérica como la suma de otros dos ratios:

- *Ratio de Siniestralidad* definido como

$$\frac{\text{Gastos Técnicos}}{\text{Ingresos adquiridos}}$$

- *Ratio Global de gastos de gestión* definido como:

$$\frac{\text{Gastos de Gestión}}{\text{Ingresos suscritos}}$$

Este ratio según se ha definido presenta dos grandes críticas:

- Se determina como suma de relaciones con distinto denominador: El mejor indicador de la calidad de la gestión de siniestros es el ratio de siniestralidad, ya que relaciona los ingresos de un período debidamente periodificados con sus gastos técnicos, por lo que es lógica su inclusión en el ratio combinado. Sin embargo, medir la calidad de aplicación de los gastos de gestión es más complicado por su heterogeneidad. Habrá que buscar algún indicador adecuado que cumpla la condición de relacionar estos gastos con los ingresos que los han generado, siempre teniendo en cuenta el reaseguro cedido. En consecuencia, en el numerador deben figurar los gastos de gestión del ejercicio netos de las comisiones de reaseguro y, en el denominador, como ingresos que han generado dichos gastos, deben incluirse los ingresos brutos no periodificados, puesto que su consumo principal (devengo de comisiones, gastos de agencia y apertura de pólizas), se efectúa en el momento de la suscripción y no a lo largo de la vida de la póliza.
- La segunda crítica viene derivada de que en su diseño no se recoge el reaseguro cedido. Para responder a esta cuestión se incorpora, como variante del ratio combinado el reaseguro cedido.

R15= Ratio de Siniestralidad de seguro directo + Ratio de Gastos

R16= Ratio de Siniestralidad de negocio neto + Ratio de Gastos Netos

Ratios de liquidez

$$R17 = \frac{\text{Gastos técnicos directos} + \text{Comis. y otros gastos de explotación}}{\text{Primas y Recargos de seguro directo}}$$

$$R18 = \frac{\text{Gastos técnicos netos} + \text{Comis. y otros gastos de explotación}}{\text{Primas y Recargos de negocio neto}}$$

$$R20 = \frac{\text{Gastos técnicos directos}}{\text{Primas y Recargos de seguro directo}}$$

$$R21 = \frac{\text{Gastos técnicos netos}}{\text{Primas y Recargos de negocio neto}}$$

Se considera que pueden modificar la posición de solvencia a largo plazo. Aunque, en principio, informan de la capacidad de afrontar los gastos anuales con las primas anuales correspondientes, es decir, de la vertiente de liquidez (Martín *et al.*, 1999, p. 78-79).

Un primer grupo de ratios de liquidez serían los ratios de gastos, llamados por la N.A.I.C., *Expense Ratios*, y son nuestros R17 y R18. En ellos se refleja qué proporción suponen los gastos de explotación respecto de las primas y recargos.

El **R17** es uno de los ratios utilizado por la NAIC (National Association Insurance Commissioners) en su sistema IRIS (Insurance Regulatory Information System) para detectar qué compañías aseguradoras pueden necesitar atención reguladora. Es el denominado *Ratio combinado, base comercial* (Pozo, 1998, p. 37-48), que une los gastos por siniestros con los demás gastos en un único ratio al unir el ratio de siniestros con el ratio de gastos en base comercial.

Un segundo grupo de ratios de liquidez serían los ratios de siniestralidad, llamados *Loss Ratios* por la N.A.I.C. y se corresponden con el R20 y R21.

Con estos últimos dos ratios se obtiene la proporción que suponen los siniestros del año (riesgo incurrido) sobre las primas destinadas a la protección o cobertura de dichos riesgos. Esto es, en qué medida las primas financian los riesgos para los cuales estaban destinadas. Hay separación entre seguro directo y neto.

Estos dos últimos ratios, tanto el que considera el seguro directo o el neto, se denominan *ratio de siniestralidad* o *ratio de siniestralidad neta*. Ambos consideran en su numerador los gastos técnicos que son, en una compañía de seguros, lo equivalente al coste industrial de una empresa de transformación, de ahí la importancia que presenta su análisis en el contexto de estas entidades (Millán, 2000, p.123). Es el indicador por excelencia de la gestión de los gastos técnicos; a través de su evolución y contrastación con el sector se puede determinar la calidad de la gestión técnica. Mediante su cotejo con la base técnica de la sociedad se puede apreciar el nivel de desviación respecto a las previsiones de coste y, en su caso, el déficit en los cálculos del precio de la póliza, permitiendo su reajuste a la mayor rapidez, siempre y cuando no esté regulado por

la Administración. Esta regularización, en caso de existir deteriora a la larga el sentido y uso de este ratio.

En pocas palabras, la eficiencia económica de un sistema de seguros se mide por su ratio de siniestralidad que es el cociente entre primas y siniestros.

Ratio de provisiones técnicas del reaseguro cedido sobre provisiones técnicas

$$R19 = \frac{\text{Provisiones Técnicas de reaseguro cedido}}{\text{Provisiones Técnicas}}$$

Ya mencionamos en la introducción que en cualquier análisis de una empresa de seguros es necesario considerar una variable clave en la misma como es el reaseguro. El reaseguro es una práctica común en cualquier compañía aseguradora ya que es una forma de diversificar el riesgo, pero al mismo tiempo también es necesario que éste no sea desproporcionado para evitar una excesiva dependencia del mismo, que en caso de insolvencia de la entidad reaseguradora, se comprometa la estabilidad de la aseguradora directa.

Es decir, las entidades aseguradoras tienen unas características técnico-financieras que le permiten cubrir las consecuencias de los riesgos hasta un límite determinado (Linares, 1998, p.208). Cuando los riesgos contratados tienen una cuantía superior a este límite, la entidad sólo tiene capacidad para asumir una parte del mismo, estando técnicamente obligada a ceder el resto. Una forma de hacer frente a este problema es compartir la cuantía entre varios aseguradores, ésta es la figura del *coaseguro*; pero otra forma es la de asumir frente al asegurado la totalidad del riesgo, asegurando a su vez con otra entidad, llamada *reasegurador*, la parte del riesgo no asumible, cubriendo sólo una parte de él, denominada *pleno de retención*. Cuando una entidad acepta ciertos riesgos y éstos exceden su pleno de retención, se ve obligada a su vez, a ceder los excesos a otras entidades. Esta figura se denomina *retrocesión*.

Una vez distinguida la figura del reaseguro de otras con las que guarda similitud, es necesario comprender, desde el punto de vista del análisis de la solvencia de una compañía de seguros, cómo funciona el reaseguro y cuál es su impacto en el balance y cuenta de resultados de la compañía. Describiremos brevemente en qué consiste.

El reaseguro es la operación por la cual una compañía de seguros transfiere a otra empresa aseguradora, denominada *reasegurador*, una parte de los riesgos del asegurado que ha aceptado y asumido, bien directamente por contratación directa (mediante suscripción de una póliza de seguro), o bien a través de otro asegurador por cesión de éste (*retrocesión*), a cambio de un precio o *prima de reaseguro* (Fernández, 2000).

Es un desplazamiento de una parte de los riesgos inicialmente transferidos por el asegurado a la compañía aseguradora (*asegurador directo*), quien a su vez los cede o transfiere en parte (de ahí que en estas operaciones se la denomine *empresa cedente* o también *reasegurado*) hacia otra empresa aseguradora (*reasegurador*), quien los acepta por la parte cedida (denominada *aceptante*).

De acuerdo con el Profesor Prieto (Prieto, 1973, p. 50) la necesidad del reaseguro para la entidad aseguradora puede tener diferentes orígenes, entre otros:

“Cuando los siniestros acaecidos dan lugar a una indemnización total de gran cuantía, cosa que puede deberse a la no homogeneidad de la cartera desde el punto de vista de la suma asegurada o sencillamente al azar. La cuestión es que la cuantía de la indemnización de todos o parte de los siniestros acaecidos fue superior a la cuantía media de la distribución de la correspondiente variable aleatoria asociada a la cuantía de la indemnización por siniestro para la clase o clases de riesgos considerados... El ente asegurador busca, mediante el reaseguro, la cobertura de estos riesgos, en cuanto que no resulten absorbibles por el mismo y, en consecuencia, comprometan su supervivencia... El desequilibrio, en orden a los resultados ofrecidos por una cartera de seguros, puede aparecer por insuficiencia de las primas cobradas por la entidad aseguradora para la cobertura del riesgo. Aquí no consideramos este caso, admitiendo siempre primas equitativas, modeladas al riesgo cubierto.”

Por tanto, de igual manera que el asegurado transfiere riesgos al asegurador directo, éste realiza una transferencia o cesión de una parte de los riesgos que tiene asumidos por las pólizas de seguro suscritas al reasegurados, en función de su capacidad económica y financiera. Con ello cambia un coste de pérdida probable incierta, cuyo montante total para el conjunto de la cartera es de cuantía imprevisible, por un coste cierto y fijo determinado en la prima de reaseguro.

Cuando los riesgos son transferidos por un reasegurados a otro, se trata de una operación de *reaseguro retrocedido*, o simplemente *retrocesión* (al reasegurados se le llama retrocesionario).

La contratación de reaseguro se convierte de esta forma en uno de los objetivos principales de la compañía aseguradora para controlar y reducir las pérdidas económicas potenciales derivadas del negocio asegurador. Contribuye, por tanto, a reforzar la confianza de los asegurados en la capacidad de la compañía de hacer frente a grandes riesgos para los cuales ha demandado su cobertura.

Al hilo de esta última cuestión, y siguiendo al Profesor Prieto (Prieto, 1973, p. 53), el reaseguro supone:

“Que la pérdidas o siniestros reasegurados sean en principio a cargo del asegurador directo. El asegurador directo se obliga a facilitar unas ciertas garantías a sus asegurados, de modo que, acaecido un siniestro, el asegurador hasta los límites de las garantías responde del pago de las correspondientes indemnizaciones. Si parte del riesgo asumido por el asegurador directo fuera cedido a uno o varios reaseguradores, esto no significaría otra cosa que el asegurador directo cubre una parte de las responsabilidades contraídas ante los asegurados, frente a los cuales continúa siendo el único responsable... El reaseguro es una relación entre dos aseguradores, careciendo el reasegurador de relación inmediata con los asegurados, ante los cuales es responsable solamente el asegurador directo. Ante éste, lo son el reasegurador o reaseguradores, de acuerdo con las características del convenio o convenios del reaseguro establecidos.”

1.2.5.- Análisis estadístico descriptivo de los ratios

Hemos efectuado previamente una *análisis descriptivo* de los ratios obteniendo de todos ellos las principales medidas estadísticas (media, varianza desviación típica, rango, valores máximos, valores mínimos, coeficiente de asimetría y de curtosis, percentiles; ver ANEXO CAPÍTULO 1).

No hemos calculado la matriz de correlación para el conjunto de ratios, aunque sabemos que algunos de ellos están muy correlacionados por dos motivos:

1.- No son modelos estadísticos con lo que los resultados no se ven afectados por las posibles correlaciones entre las variables. Es más, al poder incluir todas las variables no perdemos el contenido informativo que puede existir en ratios semiindependientes aumentando así el poder predictivo, principal objetivo de la construcción del modelo.

2.- Dado que las metodologías que vamos a emplear reducen el conjunto inicial de ratios, no deberían haber ratios muy correlacionados.

En consecuencia, en el caso de la metodología rough set, una vez seleccionados de entre todos los reductos (concepto que veremos más adelante), algunos que a nuestro juicio contienen los atributos más significativos, para poder elegir el reducto definitivo se han calculado algunos coeficientes de correlación y así asegurarnos de que el reducto que elijamos no contenga ratios muy correlacionados. Por tanto, en algunos casos sí hemos calculado el coeficiente de correlación aunque no la matriz completa.

Después, hemos calculado las *medias* de cada ratio en los cinco años considerados para ver su evolución en este período, distinguiendo entre empresas sanas y fracasadas, al objeto de comprobar el comportamiento de los ratios y sus posibles diferencias entre los casos de fracaso y no fracaso. Es decir, el análisis a través de las medias de cada variable en un período de tiempo, puede dar una idea de la evolución y diferenciación, entre los ratios de las empresas consideradas sanas, respecto a los de aquellas que fracasaron por el motivo que fuese.

En la tabla 1.3 se indican los valores de las medias:

TABLA 1.3.- MEDIAS PARA LOS DISTINTOS HORIZONTES

Ratios	Horizonte	1	2	3	4	5
	Tipo Empre.	Año	Años	Años	Años	Años
R1	Sanas	0.39	0.62	0.41	0.38	0.37
	Fracasadas	0.15	0.23	0.24	0.26	0.23
R2	Sanas	0.06	0.06	0.09	0.09	0.10
	Fracasadas	0.08	-0.23	0.02	0.06	0.07
R3	Sanas	0.1	0.08	0.09	0.10	0.10
	Fracasadas	0.09	0.07	0.12	0.18	0.07
R4	Sanas	0.11	0.12	0.13	0.14	0.13
	Fracasadas	0.20	0.18	0.24	0.14	0.20
R5	Sanas	2.12	-3.42	2.64	2.49	3.66
	Fracasadas	38.7	8.59	2.61	2.39	10.05
R6	Sanas	1.69	-1.65	2.20	2.16	3.23
	Fracasadas	37.08	5.64	2.33	2.11	1.96
R7	Sanas	0.79	0.88	1.06	1.02	1.10
	Fracasadas	2.94	1.25	0.88	0.85	8.94
R8	Sanas	0.78	0.87	1.04	1.01	1.09
	Fracasadas	2.53	1.25	0.88	0.88	1.11
R9	Sanas	0.49	0.49	0.43	0.42	0.38
	Fracasadas	0.29	0.39	0.37	0.39	0.41
R10	Sanas	2.05	-4.33	2.12	2.03	2.67
	Fracasadas	45.98	8.04	2.27	1.92	1.19
R11	Sanas	1.49	-4.00	1.66	1.58	2.28
	Fracasadas	34.74	7.23	1.98	1.77	2.55
R12	Sanas	1.12	-2.11	1.29	1.32	1.94
	Fracasadas	34.26	5.10	1.80	1.44	1.32
R13	Sanas	0.49	0.54	0.59	0.59	0.63
	Fracasadas	1.95	0.90	0.67	0.61	1.93
R14	Sanas	0.47	0.52	0.56	0.58	0.61
	Fracasadas	1.59	0.89	0.68	0.59	0.75
R17	Sanas	1.12	1.30	1.28	1.28	1.05
	Fracasadas	1.39	1.18	1.19	1.11	1.43
R18	Sanas	1.18	1.35	1.34	1.42	1.10
	Fracasadas	1.42	1.22	1.25	1.26	2.52
R19	Sanas	0.16	0.15	0.17	0.17	0.16
	Fracasadas	0.12	0.12	0.15	0.20	0.18
R20	Sanas	0.63	0.74	0.86	0.62	0.62
	Fracasadas	0.66	0.72	0.74	0.67	0.63
R21	Sanas	0.62	0.72	0.83	0.61	0.62
	Fracasadas	0.62	0.70	0.73	0.68	0.65

	Nº Empresas Sanas	Nº Empresas Fracasadas
Año 1	36	36
Año 2	35	35
Año 3	27	30
Año 4	24	27
Año5	19	21

El uso de la media como medida representativa de una muestra puede causar problemas de significación, cuando la dispersión en torno a dicha media es grande, hecho que puede producirse al aparecer algunos valores atípicos en la muestra. En nuestra muestra en el cálculo de los valores de los ratios aparecieron algunos valores extremos o atípicos, que son conocidos con la denominación de valores *outliers*. Dichos valores no han sido eliminados en la aplicación de las distintas metodologías. En concreto en el método rough set, su valor se ha discretizado al igual que para el resto de los datos. Esto es una de las ventajas de estos métodos: toman la información tal cual se presenta en la realidad, con sus inconsistencias y sin que sea necesario que los datos verifiquen ninguna hipótesis como las que se requieren para aplicar la mayoría de las técnicas estadísticas. Sin embargo la presencia de estos valores atípicos, sí afecta al cálculo de algunas de las medias, que están distorsionadas por la presencia de los mismos.

Del análisis de la evolución de las medias, y pese al inconveniente de la presencia de outliers, se deduce que para la mayoría de los ratios, las medias presentan mejores valores para las empresas sanas que para las fracasadas, y en éstas los valores se van deteriorando a medida que se acerca el fracaso. Esto demuestra que la capacidad de los ratios para distinguir entre empresas sanas y fracasadas va disminuyendo a medida que nos alejamos del momento de la crisis. Así para el año 1 antes del fracaso los ratios presentan en la mayoría de los casos valores muy diferenciados entre empresas sanas y fracasadas, diferencia que va disminuyendo para los años posteriores.

Por consiguiente, y como es lógico de esperar, una conclusión que se puede extraer del análisis del nivel de significación de los ratios es que a mayor cercanía al momento del fracaso, el poder diferenciador de los ratios es superior.

En general, y tal como hemos indicado, de los diecinueve ratios considerados, se aprecian diferencias entre unas y otras empresas, a favor de las sanas. Los ratios más significativos son el R1, R2, R4, R9, R11, R12, R13 y R14 y donde más se decanta es en el R1 que es un ratio de liquidez. En consecuencia del análisis de la significatividad de los ratios en la muestra podemos extraer estas conclusiones de cara al análisis de la solvencia:

- *La importancia de la liquidez (R1)*: Se aprecia una mayor liquidez en las empresas sanas que en las que se van a liquidar en todos y cada uno de los años, lo cual indica problemas de liquidez desde mucho antes de la liquidación.

- *La necesidad de obtener una adecuada rentabilidad financiera (R3)* no solo por razones de imagen que le permitan colocar las ampliaciones de capital o conseguir créditos, sino también porque un beneficio suficiente en una firma con autonomía financiera lleva consigo el mantenimiento de la misma.

- *Una conveniente política de financiación (R4)*. Una financiación escasa debida a ingresos insuficientes que aparezcan de manera inesperada puede ser la causa del comienzo de proceso de quiebra.

- *Autonomía Financiera (R9)*: Las empresas sanas muestran una mayor autonomía financiera lo cual les posibilita el beneficiarse del apalancamiento financiero y muestran por ello un mayor grado de este último.

- *La importancia de evaluar la solvencia estática y dinámica (R11, R12, R13, R14)*.

Además algunos de ellos también son indicativos de:

- Una mala fijación de las primas: La inversión del proceso productivo del sector asegurador, hace que el proceso de tarificación sea clave. Este proceso no es una ciencia exacta y primas adecuadamente estimadas en el pasado pueden no estarlo en el presente si las condiciones cambian y llevan consigo un aumento de los siniestros. Es evidente que si la mala tarificación da como resultado una insuficiencia de la prima para cubrir los siniestros, este hecho llevaría a resultados técnicos negativos de difícil absorción por los resultados financieros, e incluso por los capitales libres a la larga. Este factor gana en importancia según nos alejamos al momento de la crisis.

- Una infradotación de las provisiones que lleva consigo a que el beneficio del asegurador no sea el real si no que esté distorsionado y en concreto sea superior al real. Esto conduce a que dicho beneficio pueda ser repartido con la consecuente descapitalización de la empresa o a que se pueda aumentar más el volumen de negocio de lo que en realidad se debería si el beneficio no estuviese sobrestimado, lo cual provoca el lógico incremento en los riesgos que no estarían cubiertos.

- Las empresas fracasadas concentran un mayor volumen de riesgo con vencimiento en el ejercicio siguiente al de cobro de las primas.

ANEXO CAPITULO 1

TABLA 1.4. RATIOS UTILIZADOS EN EL ANÁLISIS

RATIO	DEFINICIÓN
R1	Fondo de maniobra/Activo Total $= (AC-PC)/Activo\ Total$ AC= Activo Circulante = Inversiones financ. + Créditos + Ajustes periodif.+ Efectivo + Prov. Reaseguro cedido PC= Prov. Prestaciones+ Deudas -(Empréstitos + Emp. Grupo + Emp.Asociadas) +Ajustes periodificación
R2	Beneficio antes de Impuestos(BAI)/Capitales propios (CP) CP = Capitales propios balance - Gastos Establecimiento
R3	Ingresos Financieros/ Total Inversiones
R4	BAI*/ Pasivo Total BAI* = BAI+ Amortizaciones + Provisiones + Rdos. Extraordinarios
R5	Total Primas adquiridas de seguro directo / Capitales propios
R6	Total Primas adquiridas de negocio neto / Capitales propios
R7	Total Primas adquiridas de seguro directo / Capitales propios + Prov. Técnicas
R8	Total Primas adquiridas de negocio neto / Capitales propios + Provisiones Técnicas
R9	Capitales Propios / Pasivo Total
R10	Provisiones Técnicas / Capitales Propios
R11	Gastos Técnicos de seguro directo / Capitales propios
R12	Gastos Técnicos de negocio neto / Capitales propios
R13	Gastos Técnicos de seguro directo / Capitales propios + Prov. Técnicas
R14	Gastos Técnicos de negocio neto / Capitales propios + Provisiones Técnicas
R15	Ratio Combinado 1 = Ratio Siniestralidad de seguro directo (RSD)+ Ratio de Gastos (RG) RSD = Gastos Técnicos de seguro directo/ Total Primas adquiridas de seguro directo RG = Comisiones y otros gastos de explotación/ Otros ingresos explotación
R16	Ratio Combinado 2 = Ratio Siniestralidad de negocio neto (RSN)+ Ratio de Gastos (RG) RSN = Gastos Técnicos de negocio neto/ Total Primas adquiridas de negocio neto RG = Comisiones y otros gastos de explotación/ Otros ingresos explotación
R17	(Gastos Técnicos de seguro directo + Comisiones y otros gastos de Explotación)/ Total Primas adquiridas de seguro directo
R18	(Gastos Técnicos de negocio neto + Comisiones y otros gastos de Explotación)/ Total Primas adquiridas de negocio neto
R19	Provisiones Técnicas de reaseguro cedido / Provisiones Técnicas
R20	RSD = Gastos Técnicos de seguro directo/ Total Primas adquiridas de seguro directo
R21	RSN = Gastos Técnicos de negocio neto/ Total Primas adquiridas de negocio neto

Análisis Estadístico de los ratios

Summary Statistics for R1

Count = 72
 Average = 0.270556
 Variance = 0.110941
 Standard deviation = 0.333077
 Minimum = -0.59
 Maximum = 0.98
 Range = 1.57
 Stnd. skewness = -1.50046
 Stnd. kurtosis = 0.694964

Percentiles for R1

1.0% = -0.59
 5.0% = -0.32
 10.0% = -0.21
 25.0% = 0.115
 50.0% = 0.295
 75.0% = 0.475
 90.0% = 0.64
 95.0% = 0.85
 99.0% = 0.98

Summary Statistics for R2

Count = 72
 Average = 0.0701389
 Variance = 0.0293056
 Standard deviation = 0.171189
 Minimum = -0.08
 Maximum = 1.2
 Range = 1.28
 Stnd. skewness = 16.1148
 Stnd. kurtosis = 47.0338

Percentiles for R2

1.0% = -0.08
 5.0% = 0.0
 10.0% = 0.0
 25.0% = 0.0
 50.0% = 0.01
 75.0% = 0.07
 90.0% = 0.18
 95.0% = 0.35
 99.0% = 1.2

Summary Statistics for R3

Count = 72
 Average = 0.0959722
 Variance = 0.0155934
 Standard deviation = 0.124874
 Minimum = 0.0
 Maximum = 0.73
 Range = 0.73
 Stnd. skewness = 11.7927
 Stnd. kurtosis = 23.8401

Percentiles for R3

1.0% = 0.0
 5.0% = 0.0
 10.0% = 0.01
 25.0% = 0.03
 50.0% = 0.06
 75.0% = 0.11
 90.0% = 0.18
 95.0% = 0.33
 99.0% = 0.73

Summary Statistics for R4

Count = 72
 Average = 0.155972
 Variance = 0.0459709
 Standard deviation = 0.214408
 Minimum = -0.37
 Maximum = 1.15
 Range = 1.52
 Stnd. skewness = 6.54386
 Stnd. kurtosis = 10.9518

Percentiles for R4

1.0% = -0.37
 5.0% = -0.02
 10.0% = 0.0
 25.0% = 0.03
 50.0% = 0.08
 75.0% = 0.26
 90.0% = 0.38
 95.0% = 0.57
 99.0% = 1.15

Summary Statistics for R5

Count = 72
 Average = 20.4122
 Variance = 24907.0
 Standard deviation = 157.819
 Minimum = -10.91
 Maximum = 1340.66
 Range = 1351.57
 Stnd. skewness = 29.3736
 Stnd. kurtosis = 124.592

Percentiles for R5

1.0% = -10.91
 5.0% = -4.62
 10.0% = 0.06
 25.0% = 0.565
 50.0% = 1.565
 75.0% = 3.29
 90.0% = 6.6
 95.0% = 7.58
 99.0% = 1340.66

Summary Statistics for R6

Count = 72
 Average = 19.3869
 Variance = 22927.3
 Standard deviation = 151.418
 Minimum = -10.91
 Maximum = 1286.11
 Range = 1297.02
 Stnd. skewness = 29.3758
 Stnd. kurtosis = 124.604

Percentiles for R6

1.0% = -10.91
 5.0% = -4.28
 10.0% = 0.06
 25.0% = 0.525
 50.0% = 1.38
 75.0% = 2.715
 90.0% = 4.94
 95.0% = 6.98
 99.0% = 1286.11

Summary Statistics for R7

Count = 72
 Average = 1.86431
 Variance = 40.5553
 Standard deviation = 6.3683
 Minimum = 0.01
 Maximum = 53.7
 Range = 53.69
 Stnd. skewness = 27.1696
 Stnd. kurtosis = 111.189

Percentiles for R7

1.0% = 0.01
 5.0% = 0.09
 10.0% = 0.29
 25.0% = 0.455
 50.0% = 0.725
 75.0% = 1.22
 90.0% = 2.7
 95.0% = 3.74
 99.0% = 53.7

Summary Statistics for R8

Count = 72
 Average = 1.65278
 Variance = 40.3128
 Standard deviation = 6.34924
 Minimum = -5.26
 Maximum = 53.7
 Range = 58.96
 Stnd. skewness = 27.5966
 Stnd. kurtosis = 114.5

Percentiles for R8

1.0% = -5.26
 5.0% = 0.05
 10.0% = 0.22
 25.0% = 0.46
 50.0% = 0.7
 75.0% = 1.18
 90.0% = 1.9
 95.0% = 3.13
 99.0% = 53.7

Summary Statistics for R9

Count = 72
 Average = 0.389722
 Variance = 0.161465
 Standard deviation = 0.401827
 Minimum = -1.16
 Maximum = 1.0
 Range = 2.16
 Std. skewness = -3.77911
 Std. kurtosis = 5.27407

Percentiles for R9

1.0% = -1.16
 5.0% = -0.16
 10.0% = 0.06
 25.0% = 0.14
 50.0% = 0.35
 75.0% = 0.68
 90.0% = 0.91
 95.0% = 0.96
 99.0% = 1.0

Summary Statistics for R10

Count = 72
 Average = 24.0114
 Variance = 36804.4
 Standard deviation = 191.845
 Minimum = -11.17
 Maximum = 1629.05
 Range = 1640.22
 Std. skewness = 29.3821
 Std. kurtosis = 124.64

Percentiles for R10

1.0% = -11.17
 5.0% = -1.66
 10.0% = 0.0
 25.0% = 0.04
 50.0% = 0.545
 75.0% = 2.97
 90.0% = 6.04
 95.0% = 6.67
 99.0% = 1629.05

Summary Statistics for R11

Count = 72
 Average = 18.1181
 Variance = 20493.5
 Standard deviation = 143.155
 Minimum = -7.81
 Maximum = 1215.8
 Range = 1223.61
 Std. skewness = 29.3817
 Std. kurtosis = 124.638

Percentiles for R11

1.0% = -7.81
 5.0% = -3.01
 10.0% = 0.0
 25.0% = 0.27
 50.0% = 1.095
 75.0% = 2.43
 90.0% = 4.94
 95.0% = 5.48
 99.0% = 1215.8

Summary Statistics for R12

Count = 72
 Average = 17.6903
 Variance = 20132.4
 Standard deviation = 141.889
 Minimum = -6.89
 Maximum = 1204.81
 Range = 1211.7
 Std. skewness = 29.3843
 Std. kurtosis = 124.653

Percentiles for R12

1.0% = -6.89
 5.0% = -2.8
 10.0% = 0.0
 25.0% = 0.27
 50.0% = 0.845
 75.0% = 1.815
 90.0% = 2.93
 95.0% = 4.54
 99.0% = 1204.81

Summary Statistics for R13

Count = 72
Average = 1.22292
Variance = 16.238
Standard deviation = 4.02964
Minimum = 0.0
Maximum = 33.65
Range = 33.65
Std. skewness = 26.4092
Std. kurtosis = 106.154

Percentiles for R13

1.0% = 0.0
5.0% = 0.01
10.0% = 0.14
25.0% = 0.27
50.0% = 0.49
75.0% = 0.82
90.0% = 1.57
95.0% = 2.23
99.0% = 33.65

Summary Statistics for R14

Count = 72
Average = 1.03278
Variance = 15.993
Standard deviation = 3.99912
Minimum = -4.46
Maximum = 33.65
Range = 38.11
Std. skewness = 27.1092
Std. kurtosis = 112.157

Percentiles for R14

1.0% = -4.46
5.0% = 0.0
10.0% = 0.1
25.0% = 0.225
50.0% = 0.435
75.0% = 0.765
90.0% = 1.54
95.0% = 2.02
99.0% = 33.65

Summary Statistics for R17

Count = 72
Average = 1.25361
Variance = 1.03725
Standard deviation = 1.01846
Minimum = 0.66
Maximum = 9.5
Range = 8.84
Std. skewness = 26.6412
Std. kurtosis = 108.716

Percentiles for R17

1.0% = 0.66
5.0% = 0.87
10.0% = 0.91
25.0% = 0.98
50.0% = 1.055
75.0% = 1.27
90.0% = 1.5
95.0% = 1.74
99.0% = 9.5

Summary Statistics for R18

Count = 72
Average = 1.30111
Variance = 1.04942
Standard deviation = 1.02441
Minimum = 0.73
Maximum = 9.5
Range = 8.77
Std. skewness = 25.7814
Std. kurtosis = 103.575

Percentiles for R18

1.0% = 0.73
5.0% = 0.92
10.0% = 0.94
25.0% = 1.0
50.0% = 1.09
75.0% = 1.29
90.0% = 1.57
95.0% = 1.74
99.0% = 9.5

Summary Statistics for R19

Count = 68
Average = 0.148529
Variance = 0.0453172
Standard deviation = 0.212878
Minimum = 0.0
Maximum = 0.89
Range = 0.89
Std. skewness = 5.91814
Std. kurtosis = 4.46797

Percentiles for R19

1.0% = 0.0
5.0% = 0.0
10.0% = 0.0
25.0% = 0.0
50.0% = 0.065
75.0% = 0.19
90.0% = 0.48
95.0% = 0.54
99.0% = 0.89

Summary Statistics for R20

Count = 72
Average = 0.643889
Variance = 0.0664495
Standard deviation = 0.257778
Minimum = 0.0
Maximum = 1.22
Range = 1.22
Std. skewness = -2.15329
Std. kurtosis = 1.01208

Percentiles for R20

1.0% = 0.0
5.0% = 0.06
10.0% = 0.29
25.0% = 0.515
50.0% = 0.68
75.0% = 0.785
90.0% = 0.95
95.0% = 1.05
99.0% = 1.22

Summary Statistics for R21

Count = 72
Average = 0.619306
Variance = 0.0671305
Standard deviation = 0.259096
Minimum = -0.01
Maximum = 1.38
Range = 1.39
Std. skewness = -1.46278
Std. kurtosis = 1.93368

Percentiles for R21

1.0% = -0.01
5.0% = 0.02
10.0% = 0.28
25.0% = 0.515
50.0% = 0.655
75.0% = 0.75
90.0% = 0.94
95.0% = 0.99
99.0% = 1.38

Capítulo 2

EL MÉTODO ROUGH SET. APLICACIÓN A LA PREDICCIÓN DE INSOLVENCIAS

II.1. EL MÉTODO ROUGH SET

II.1.1.- Aproximación al concepto de rough set

Pawlak (1982), introdujo a comienzos de los años ochenta una herramienta matemática para tratar con la incertidumbre inherente en un proceso de decisión para el caso en que se presenten inconsistencias en un sistema de información. Esta herramienta es la denominada teoría del *Rough Set* y por inconsistencia entendemos objetos que poseen los mismos valores para un conjunto de atributos y sin embargo pertenecen a distintas clases o categorías. En nuestro caso, se trataría de dar solución al caso en que dos empresas caracterizadas por los mismos ratios financieros u otras variables, una sin embargo pertenece a las solventes y la otra no.

Esta teoría implica el cálculo de particiones, divisiones o clases, según queramos. Es algo diferente tanto de la teoría estadística de la probabilidad como de la teoría *fuzzy set*. Para comprender en qué se diferencian describiremos las tres categorías generales de imprecisión que encontramos en el análisis científico. La primera categoría ocurre, cuando un acontecimiento es aleatorio en naturaleza (por ejemplo, lanzar una moneda) y por lo tanto no se sabe qué resultado en particular (de todos los posibles) es el que se va a materializar. Por lo tanto, la imprecisión asociada con este tipo de acontecimientos puede describirse por la teoría estadística de la probabilidad. La segunda surge del hecho de que los objetos pueden no pertenecer exclusivamente a una única categoría pero pueden pertenecer a varias categorías aunque con diferentes grados, en este caso la imprecisión toma forma de pertenencia difusa a un conjunto (el ejemplo más claro es la clasificación de los objetos de color gris en las categorías “negro” o “blanco”). Esta segunda categoría es el objeto de la lógica *fuzzy*. La tercera categoría es la teoría rough set que es útil cuando las clases en las que han de clasificarse los objetos son imprecisas, pero sin embargo pueden aproximarse mediante conjuntos precisos (Nurmi et al., 1996).

Estas diferencias muestran una de las principales ventajas de la teoría rough set y es que no necesita ninguna información adicional acerca de los datos como puede ser una distribución de probabilidad en estadística o el grado o probabilidad de pertenencia en la teoría *fuzzy set*. Otras ventajas serían las siguientes:

- analiza los hechos escondidos en los datos,
- no necesita ninguna información adicional o preliminar sobre los datos como probabilidades a priori en la teoría de la probabilidad o grado de pertenencia en la teoría *fuzzy set*,
- reduce los atributos y objetos superfluos para obtener los denominados *reductos* (o *conjuntos mínimos de atributos*) que aseguran la misma calidad de clasificación que el conjunto total,

- obtiene la mínima representación del conocimiento en forma de reglas de decisión teniendo en consecuencia, una interpretación sencilla,
- está especialmente indicada para tratar las inconsistencias de los datos; no corrige las inconsistencias que se presentan en los ejemplos o en los casos sino que calcula aproximaciones "por arriba" y "por abajo". En consecuencia, y basándose en estas aproximaciones, se inducen dos conjuntos de reglas: determinísticas (ciertas) y no-determinísticas (posibles),
- conceptualmente es simple y por tanto requiere algoritmos simples.

Las ideas que vamos a exponer en las secciones siguientes de este capítulo son algunos conceptos básicos de la teoría Rough Set descritos por Pawlak 1982, 1984 y 1991. Aunque en la actualidad existen extensiones de esta teoría (Greco et al., 1998), expondremos el enfoque clásico, que es el que hemos seguido en nuestra aplicación práctica.

Básicamente, el *nuevo enfoque*, incorpora información adicional acerca de las propiedades de ordenación de los atributos analizados, en consecuencia se obtiene potencialmente como resultado un modelo final más compacto ya que algunas reglas contradictorias para determinados casos son de este modo eliminadas (McKee, 2000). En otras palabras, utiliza información adicional para generar un modelo final más simple.

El enfoque original no ordena el dominio de los atributos. Asume que los diferentes valores de los mismos atributos son igualmente preferibles y que únicamente el valor predictivo del atributo, tal y como es revelado por los datos, será tenido en cuenta en el modelo. Por lo tanto, esto le hace menos restrictivo en cuanto a las hipótesis sobre los datos que el nuevo enfoque.

Rough Set podría traducirse como conjunto aproximado o impreciso. La hipótesis fundamental en la teoría rough set es la siguiente: los objetos del universo se perciben únicamente a través de la información que hay accesible acerca de ellos, esto es, a través de valores de atributos es como podemos evaluar estos objetos.

La filosofía del método se basa en la suposición de que con cada objeto del universo que estamos considerando se puede asociar alguna información (datos, conocimiento). Los objetos caracterizados por la misma información no se pueden diferenciar en vista de la información disponible. La relación de "no diferenciación" generó de este modo la base matemática para esta teoría y es debido a la existencia de información imprecisa.

De manera intuitiva, un rough set es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de valores de un conjunto de atributos. Hablaremos de aquí en adelante de rough set.

Por lo tanto, si pretendemos clasificar objetos se hará en base a la información que hay accesible sobre ellos y no en los objetos en sí mismos.

Normalmente junto con la información existente sobre los objetos se da una clasificación suministrada por un decisor. El problema de clasificación en este caso consiste en qué medida es posible reflejar mediante atributos (de condición) la clasificación efectuada por el decisor.

Pero nos encontramos con un problema para realizar la clasificación, y es que la no diferenciación de objetos por medio de atributos impide generalmente su asignación precisa a un conjunto, tal y como se ha mencionado. De este modo, si no podemos asignar de manera precisa los objetos a sus clases o categorías trataremos de hacer una clasificación lo más aproximada posible. Es decir *aproximaremos las clases o categorías*.

Basándonos en esta idea de la aproximación podemos dar otro concepto de rough set: es un par de aproximaciones por arriba y por abajo de un conjunto en términos de objetos no discernibles. En otras palabras, un rough set es una colección de objetos que, en general, no pueden ser clasificados de manera precisa en términos de los valores del conjunto de atributos, mientras que las aproximaciones por arriba y por abajo sí pueden.

En consecuencia, cada rough set tiene casos fronterizos, esto es objetos que no pueden clasificarse con certeza como miembros del conjunto o de su complementario.

En definitiva, un rough set puede ser reemplazado o representado por un par de conjuntos precisos, llamados *la aproximación por encima y por debajo*. La aproximación por debajo consiste en todos los objetos que con seguridad pertenecen al conjunto y la aproximación por encima contiene los objetos que posiblemente pertenecen al conjunto.

II.1.2.- Tabla de información y relación de no diferenciación

La teoría del rough set asume la representación del conocimiento de los objetos en forma de una *tabla de información*, que es un caso especial de un sistema de información. En las filas de la tabla se indican los *objetos* (acciones, alternativas, candidatos, pacientes, empresas, etc.), mientras que las columnas se corresponden con los *atributos* (ratios financieros, síntomas de enfermedades,...). Las entradas en la tabla son los *valores del atributo*. Es decir, la entrada en columna q y en fila x tiene el valor $f(x, q)$. Cada fila en la tabla representa información sobre un objeto del sistema de información.

La teoría rough set utiliza la noción de atributo en lugar de la de criterio porque el primero es más general que el segundo debido a que el dominio (escala) de un criterio ha de ordenarse de menor a mayor preferencia mientras que el dominio de los atributos no ha de ser ordenado.

Por tanto, para cada par (objeto, atributo) se conoce un valor denominado *descriptor*. Cada fila de la tabla contiene descriptores que representan información correspondiente a un objeto del universo, es decir, cada fila en la tabla representa la información acerca de un objeto. Por ejemplo, un fichero de datos relativo a pacientes que sufren de una determinada enfermedad.

Además, y en general, el conjunto de atributos se puede dividir en dos subconjuntos: un subconjunto de *atributos de condición* (características, variables, rasgos,...) y en subconjunto de *atributos de decisión* (opiniones, clases, diagnósticos,...). Si se distingue entre *atributos de condición* y *atributos de decisión* en un sistema de información, obtenemos, como veremos una tabla de decisión.

Expresados es tos conceptos la relación de *no diferenciación* (no olvidemos que es la base de esta teoría) ocurrirá cuando dados dos objetos, x e y , decimos que x e y *no son discernibles* si todos los descriptores toman los mismos valores. Es decir, si para todos los atributos toman los mismos valores, por tanto si, y solo si, $f(x, q) = f(y, q)$

Para clarificar los conceptos que iremos explicando, vamos a ilustrar este capítulo con un ejemplo sobre la selección de alumnos para un colegio (Moscarola, 1978 y Slowinski, 1993).

Los candidatos han remitido sus solicitudes con sus calificaciones académicas, su curriculum vitae y las referencias del colegio anterior. Todos estos documentos son examinados por un comité que decidirá su admisión o no. Basándose en los documentos anteriores, los candidatos son *descritos* utilizando siete criterios (atributos de condición). A continuación presentamos la lista de estos criterios y sus escalas correspondientes, ordenadas de mejor a peor:

C_1 – nota de matemáticas, {5, 4, 3}

C_2 – nota de física, {5, 4, 3}

C_2 – nota de física, {5, 4, 3}

C_3 – nota de inglés, {5, 4, 3}

C_4 – nota media en el resto de asignaturas, {5, 4, 3}

C_5 – tipo colegio donde realizó la educación secundaria, {1, 2, 3}

C_6 – motivación, {1, 2, 3}

C_7 – opinión del colegio anterior, {1, 2, 3}

Quince candidatos son los que han presentado las solicitudes y han sido clasificados por el comité después de la debida consideración. Estos candidatos son los ejemplos que vamos a utilizar para conocer cuales son las *preferencias* de dicho comité a la hora de admitir o rechazar a un candidato.

El atributo de decisión es d y en base al mismo se puede hacer una clasificación dicotómica de los candidatos: $d = A$ significa admisión y $d = R$ significa no admisión.

Al haber distinguido entre atributos de condición y atributos de decisión en este sistema de información, obtenemos, una tabla de decisión.

La tabla de decisión sería la siguiente:

Crterios	C_1	C_2	C_3	C_4	C_5	C_6	C_7	D
Candidatos								
X_1	4	4	4	4	2	2	1	A
X_2	3	3	4	3	2	1	1	R
X_3	3	4	3	3	1	2	2	R
X_4	5	3	5	4	2	1	2	A
X_5	4	4	5	4	2	2	1	A
X_6	3	4	3	3	2	1	3	R
X_7	4	4	5	4	2	2	2	A
X_8	4	4	4	4	2	2	2	A
X_9	4	4	4	4	2	2	2	R
X_{10}	5	3	5	4	2	1	2	A
X_{11}	5	4	4	4	1	1	2	A
X_{12}	5	3	4	4	2	2	2	A
X_{13}	4	3	3	3	3	2	2	R
X_{14}	3	3	4	3	2	3	3	R
X_{15}	4	5	5	4	2	1	1	A

Esta claro que el conjunto de los atributos de condición es:

$$C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$$

y el conjunto de los atributos de decisión es:

$$D = \{d\}$$

Denominemos como Y_A al conjunto de candidatos admitidos y a Y_N al conjunto de los rechazados por el comité. Por tanto, $Y_A = \{x_1, x_4, x_5, x_7, x_{10}, x_{11}, x_{12}, x_{15}\}$, $Y_N = \{x_2, x_3, x_6, x_8, x_9, x_{13}, x_{14}\}$, $Y = \{Y_A, Y_N\}$. Los conjuntos Y_A e Y_N son clases definidas por el conjunto D en la tabla de decisión.

II.1.3.- Aproximación de conjuntos

Ya se ha mencionado al definir el concepto de rough set que dado un sistema de información, para cada categoría en que puede clasificarse un objeto podemos establecer dos subconjuntos, las denominadas aproximaciones por encima y por debajo.

La aproximación por debajo de una determinada categoría está formada por el conjunto de elementos que, de acuerdo al conjunto de atributos, pueden con certeza clasificarse en dicha categoría. La aproximación por encima de una determinada categoría está formada por el conjunto de elementos que, de acuerdo con el conjunto de atributos, *posiblemente* puedan clasificarse en dicha categoría. Es decir, no los podemos clasificar con certeza porque habrá casos de duda que vamos a incorporar a la aproximación por arriba. Es precisa estos “casos dudosos” (si es que existen) los que diferencian la aproximación por encima y por debajo y constituyen la denominada *frontera* o *región de duda*. Puede ocurrir que el sistema no tenga casos de duda porque la información sea muy precisa y permita asignar sin problemas los objetos a cada categoría.

Usando la aproximación por encima y por debajo de un conjunto se puede definir la *precisión* de la aproximación y la *calidad* de la misma, que son números del intervalo $[0, 1]$ y que definiremos a continuación.

Para cada categoría o clase se puede calcular la denominada *precisión* dicha categoría. La precisión es el cociente entre el número de elementos que tiene la aproximación por debajo de una categoría dividido por el número de elementos que tienen la aproximación por encima de dicha categoría. Si es igual a 1 indicará que dicha categoría no tiene casos de duda (las dos aproximaciones son iguales) y por lo tanto todos los elementos pertenecientes a la misma se pueden clasificar de manera precisa.

Continuando con el ejemplo iniciado en las secciones anteriores en nuestra tabla existen 13 conjuntos elementales que son: dos pares de candidatos que no se diferencian, $\{x_4, x_{10}\}$, $\{x_8, x_9\}$ y 11 candidatos que sí se diferencian. Las aproximaciones por arriba y por debajo de los conjuntos Y_A e Y_N son: respectivamente:

$$\underline{C}Y_A = \{x_1, x_4, x_5, x_7, x_{10}, x_{11}, x_{12}, x_{15}\} \text{ (aprox. por debajo clase A).}$$

$$\overline{C}Y_A = \{x_1, x_4, x_5, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{15}\} \text{ (aprox. por encima clase A)}$$

$$Bn_p(Y_A) = \{x_8, x_9\} \text{ (frontera clase A)}$$

$$\underline{C}Y_R = \{x_2, x_3, x_6, x_{13}, x_{14}\},$$

$$\overline{C}Y_R = \{x_2, x_3, x_6, x_8, x_9, x_{13}, x_{14}\}$$

$$Bn_p(Y_R) = \{x_8, x_9\}$$

La precisión de la aproximación para cada conjunto es la siguiente:

$$\alpha(Y_A) = 8/10 = 0.8 \text{ y } \alpha(Y_N) = 5/7 = 0.71.$$

Observemos que la región de duda o frontera está compuesta por dos candidatos, el 8 y el 9. Ambos tienen la misma puntuación de acuerdo a los criterios definidos por el conjunto C pero el comité ha admitido al 8 y rechazado al 9. Esto implica que la decisión no es consistente con la evaluación de los candidatos en base a los criterios considerados. Por tanto, en apariencia, el comité debería recabar información adicional de las solicitudes de los candidatos o entrevistarlos. Esta conclusión lleva consigo que el comité o bien adopta un criterio adicional discriminatorio o bien, si su definición es demasiado difícil, debería crear una tercera categoría de candidatos: aquellos a los que se debe entrevistar.

Si en nuestro sistema de información hay varias clases o categorías podemos calcular la denominada *calidad de clasificación* como el cociente entre el número de todos los elementos que componen todas las aproximaciones por debajo de todas las categorías, dividido por el número total de objetos del sistema. Expresa el ratio de todos los objetos correctamente clasificados respecto a todos los objetos del sistema (Slowinski y Stefanowski, 1994a, 6-8).

En nuestro ejemplo la calidad tomaría el siguiente valor:

$$\gamma_p(Y) = (8+5)/15 = 0.87.$$

Con los conceptos de precisión y calidad de la aproximación vistos en las dos últimas secciones podemos definir en qué medida está descrito el conjunto (o clasificación) de objetos examinado utilizando la información disponible. En consecuencia, podemos evaluar la importancia de cada atributo en particular para la clasificación ya que si la calidad toma un valor bajo significa que hay que tomar en consideración otros atributos no considerados a la hora de poder tomar una decisión.

II.1.4.- Reducción y dependencia de atributos

Una de las funciones más importantes de la investigación rough set es el descubrimiento de las dependencias entre atributos al analizar una tabla de información. Mediante dicho descubrimiento se pretende lograr la reducción del conocimiento, es decir, el proceso de expresar un conocimiento dado del modo más eficiente.

Descubrir las *dependencias* implica que puede ocurrir que empleando un subconjunto de atributos de la totalidad que existen en nuestra tabla, podamos clasificar los objetos del sistema de la misma manera que si empleásemos la totalidad de atributos que hay en el mismo.

Es decir, en las aplicaciones prácticas estamos interesados en reducir aquellos atributos que son redundantes (es decir, estamos interesados en obtener lo que hemos denominado *reductos*). La reducción de atributos se consigue mediante la búsqueda de un modelo tal que el conjunto reducido de atributos proporcione la misma calidad de clasificación que el conjunto original de atributos. El mínimo subconjunto de atributos que asegure la misma calidad de clasificación que el conjunto de todos los atributos es denominado *reducto* o *conjunto mínimo*.

En el caso en que la tabla de información tenga más de un reducto, la intersección de todos ellos se llama *Núcleo*.

El núcleo es la colección de los atributos más relevantes en la tabla que no pueden ser eliminados sin que disminuya la calidad de la clasificación.

En otras palabras, dentro de la reducción del conocimiento, los conceptos de núcleo y reducto pueden tomar la siguiente interpretación. El núcleo se compone de aquellas clasificaciones que son las más esenciales en el conocimiento. Ninguna relación del núcleo puede ser eliminada en la reducción del conocimiento sin distorsionar el mismo. Un reducto, por otro lado, proporciona un conjunto de relaciones que es suficiente para caracterizar el conocimiento sin perder nada esencial. Es decir, los reductos son intercambiables (se puede considerar uno u otro) pero el núcleo no.

Sin entrar en la casuística de modelos propuestos para hallar los reductos y el núcleo, presentaremos brevemente uno de los más utilizados.

Skowron propuso representar el conocimiento en forma de una *matriz de diferenciación* (Skowron, 1991a y 1991b). Esta representación tiene muchas ventajas, en particular permite de manera simple el cálculo del núcleo, reductos y otros conceptos considerados. La idea es la siguiente:

Por matriz de diferenciación de S, que denominaremos M(S), tenemos la matriz nxn definida por las entradas c_{ij} . La entrada c_{ij} es el conjunto de todos los atributos que diferencian todos los objetos x_i y x_j .

El núcleo puede definirse ahora como el conjunto de todas las entradas de un único elemento de la matriz de diferenciación y el reducto es el mínimo conjunto de atributos que diferencian todos los objetos a distinguir por la totalidad de todos los atributos.

La matriz de diferenciación es simétrica lo cual simplifica y facilita su cálculo.

Para aclarar el concepto de matriz de diferenciación proponemos el siguiente ejemplo (Pawlak, 1982). Consideremos que nuestra tabla de información es la siguiente:

U	a	b	c	d
1	0	1	2	0
2	1	2	0	2
3	1	0	1	0
4	2	1	0	1
5	1	1	0	2

Donde a, b, c, y d son los criterios y 1, 2, 3, 4, 5 son los candidatos evaluados por los criterios.

La matriz sería la siguiente:

	1	2	3	4	5
1					
2	a, b, c, d				
3	a, b, c,	b, c, d			
4	a, c, d	a, b, d	a, b, c, d		
5	a, c, d	b	b, c, d	a, d	

De la tabla se deduce fácilmente que el núcleo es el atributo *b* y que en la tabla hay dos reductos el {a, b} y el {d, b}.

En nuestro ejemplo de los candidatos y, procediendo de la misma manera, obtendríamos los siguientes reductos:

$$RED_{y1}(C) = \{c_2, c_3, c_6, c_7\}$$

$$RED_{y2}(C) = \{c_1, c_3, c_7\}$$

$$RED_{y3}(C) = \{c_2, c_3, c_5, c_7\}$$

Por tanto, podría decirse que el comité toma las decisiones teniendo en cuenta únicamente los criterios de los reductos y descartando el resto. Por ejemplo, el criterio c_4 no influye en ninguna decisión puesto que no aparece en ningún reducto.

El núcleo está formado por la intersección de los tres reductos, es decir:

$$CORE_y(C) = \{c_3, c_7\}$$

II.1.5.- Reglas de decisión

La obtención de las reglas de decisión es la cuestión más importante de la metodología rough set dado que el modelo que obtenemos con la aplicación de esta metodología es un conjunto de reglas de decisión o, en nuestro caso, de clasificación que nos permitan clasificar los objetos. Para ello, en las secciones anteriores hemos reducido los atributos son redundantes para que nuestra tabla de información sea lo más precisa posible.

Por tanto, en esta sección vamos a tratar la inducción de reglas de decisión de un conjunto de datos que representan la experiencia en un dominio. Se supone que el conjunto de datos contiene información de un conjunto de *objetos* descritos por un conjunto de *atributos*. El problema, por tanto, consiste en encontrar reglas que determinen si un objeto pertenece a una determinada *clase de decisión* o a un *concepto*. También vamos a suponer que la definición de esta clase es conocida porque por ejemplo lo ha definido el experto o un usuario. Las reglas objeto de estudio, como comprobaremos más adelante, son sentencias lógicas de la siguiente forma:

SI <se cumplen las condiciones> **ENTONCES** <el objeto pertenece a la clase de decisión indicada>

Donde las condiciones se componen de una conjunción de valores de los atributos.

Formalizando el problema: supongamos que los objetos se describen mediante un conjunto de atributos que a su vez se dividen en dos subconjuntos de atributos de condición y atributos de decisión, a los que vamos a denotar por C y D, respectivamente. Estamos ante una *tabla de decisión*. D a menudo es un conjunto unitario (porque la decisión puede consistir en aceptar o rechazar; o solvente o insolvente, por ejemplo).

Es decir, es suficiente considerar tablas de decisión con una única decisión ya que siempre por medio de una simple codificación se puede transformar cualquier tabla de decisión con más de una decisión en una tabla de decisión con exactamente una sola decisión. Se puede interpretar un atributo de decisión como un tipo de clasificación del universo de los objetos que ha sido formulada por un experto decisor, operador, físico... Observemos, como resultado, que el atributo de decisión *d* sirve para clasificar los objetos de la tabla de información.

De una tabla de decisión se pueden derivar un *conjunto de reglas de decisión*. De una tabla de decisión podemos obtener dos tipos de reglas. Reglas *o consistentes* que son aquellas en que si un objeto verifica el conjunto C (atributos de condición) es asignado de manera *unívoca* a una clase que ha determinado el atributo D. Las reglas de decisión determinísticas describen *unívocamente* las decisiones (acciones) que deben tomarse cuando algunas *condiciones* se satisfacen. En otro caso las reglas son *no determinísticas* o *inconsistentes*. En el caso de una tabla no determinística, las decisiones no se determinan unívocamente por las condiciones (es decir, las condiciones pueden conducir a varias posibles

decisiones o a varias clases). En su lugar, se define un conjunto de decisiones, el cual sería apropiado bajo circunstancias determinadas por condiciones. Estas últimas reglas son consecuencia de la información imperfecta de los objetos. En definitiva, las reglas derivadas de utilizar la aproximación por debajo son exactas, mientras que las reglas son aproximadas cuando utilizamos la aproximación por encima o la región frontera. Si no hay región frontera, las dos aproximaciones coinciden y todas las reglas que se obtienen son exactas.

El conjunto de reglas para todas las clases de decisión se denomina *algoritmo de decisión*.

Un algoritmo de decisión puede entenderse como la representación más compacta y sin redundancias correspondiente a un sistema de información. La primera condición se entiende en el sentido del menor número de reglas de decisión y, la segunda como, el menor número de atributos que aparezcan en la definición de todas las reglas.

El algoritmo de decisión muestra todas las relaciones importantes utilizando un mínimo número de reglas y/o un mínimo número de atributos que aparezcan en las reglas. Por tanto, el algoritmo de decisión es más legible para el usuario que el sistema de información original.

Cada regla de decisión se caracteriza por su *fuerza*, esto es el número de objetos que satisfacen la parte de la condición de la regla (en el argot se dice *cubiertos* por la regla) y que pertenecen a la clase de decisión sugerida. En el caso de reglas aproximadas, la fuerza se calcula para cada posible clase de decisión de forma separada. Las reglas más fuertes son normalmente más generales, es decir, su parte de condición es más corta y menos especializada.

Por tanto no todas las reglas de decisión del algoritmo de decisión son igualmente importantes o fiables para el decisor. Las reglas se crean sobre la base de diferentes números de objetos del sistema de información. Una regla puede crearse sobre la base de un único objeto mientras que otra puede crearse sobre la base de muchos objetos. En consecuencia, la primera regla es *más débil* que la segunda, o sea, menos *fiable* como ayuda a la hora de tomar una decisión. Esta es la razón por la que junto a la regla se indican el número de objetos que en el sistema de información han dado lugar a la misma, o sea, su fuerza.

Puede también resultar interesante buscar reglas *parcialmente discriminantes* (Mienko et al., 1996). Estas son reglas que cubrirían un número limitado de ejemplos negativos además de los positivos. Se caracterizan por un coeficiente denominado *nivel de discriminación*, que especificaría en qué medida la regla es consistente, esto es, cuál es el ratio de ejemplos positivos con respecto al total de ejemplos cubiertos por la regla.

La ingeniería del conocimiento ha dado lugar a un gran número de algoritmos para inducir generalizaciones a través de los casos o ejemplos. Algunos algoritmos dan esas descripciones en forma de árboles de decisión,

mientras que otros proveen reglas de decisión que son las que en este capítulo estamos considerando.

Los procedimientos para generar reglas de decisión a partir de una tabla de decisión operan sobre los principios del aprendizaje inductivo. No vamos a entrar en la casuística de algoritmos de inducción de reglas que operan en base a la teoría rough set solamente indicaremos que los algoritmos de inducción existentes utilizan una de las siguientes estrategias:

- 1) generación de un conjunto mínimo de reglas que cubran todos los objetos de una tabla de decisión,
- 2) generación de un conjunto exhaustivo de reglas consistentes en todas las reglas posibles de una tabla de decisión,
- 3) generación de un conjunto de reglas de decisión “fuertes”, incluso en parte discriminantes, cubriendo relativamente muchos objetos pero no necesariamente todos los objetos de la tabla de decisión.

Vamos a ver cuáles serían las reglas que se derivarían del ejemplo de los candidatos que hemos ido desarrollando a lo largo de este capítulo.

Supongamos que de los tres reductos que teníamos el comité elige el segundo (en un caso real debería justificarse dicha elección). Por tanto el resto de criterios podría eliminarse y nos quedaría la siguiente tabla reducida:

Criterios	C ₁	C ₃	C ₇	D
Candidatos				
X ₁	4	4	1	A
X ₂	3	4	1	R
X ₃	3	3	2	R
X ₄	5	5	2	A
X ₅	4	5	1	A
X ₆	3	3	3	R
X ₇	4	5	2	A
X ₈	4	4	2	A
X ₉	4	4	2	R
X ₁₀	5	5	2	A
X ₁₁	5	4	2	A
X ₁₂	5	4	2	A
X ₁₃	4	3	2	R
X ₁₄	3	4	3	R
X ₁₅	4	5	1	A

De esta tabla se obtiene las siguientes reglas de decisión:

Regla 1: si $c_1 = 5$ entonces $d = A$

Regla 2: si $c_3 = 5$ entonces $d = A$

Regla 3: si $c_1 = 4$ y $c_7 = 1$ entonces $d = A$

Regla 4 si $c_1 = 4$ y $c_3 = 4$ y $c_7 = 2$ entonces $d = A$ o R

Regla 5: si $c_1 = 3$ entonces $d = R$

Regla 6: si $c_3 = 3$ entonces $d = R$.

Hemos de destacar que cinco de las reglas son determinísticas y una de ellas no lo es. La regla no determinística es la 4 y es la que se deriva de los candidatos 8 y 9 que pertenecen a diferentes clases de decisión. Define a una tercera categoría de decisión que incluiría los candidatos a los que se debería entrevistar. Las reglas nos indican cuál es la política que sigue el colegio al admitir o rechazar un candidato.

II.1.6.- Toma de decisiones utilizando reglas de decisión

Las reglas de decisión derivadas de una tabla con el enfoque rough set se expresan en términos de los atributos más significativos sin las redundancias típicas de los datos originales y, en consecuencia, pueden ser utilizadas para recomendaciones sobre la clasificación de nuevos objetos.

Un *problema de clasificación* es un tipo de problema de decisión. Consiste en determinar la asignación de un objeto, descrito por un conjunto de atributos, a una clase de decisión conocida a priori. La ayuda para clasificar los nuevos objetos se basa en el uso de reglas de decisión derivadas de la experiencia pasada (manifestada a través de un conjunto de ejemplos de aprendizaje) y sistematizada en forma de dichas reglas a través del enfoque rough set.

Los problemas de clasificación son el punto central para la metodología rough set.

Una vez que hemos obtenido el algoritmo de las reglas de decisión, el cual representa el conocimiento que hemos obtenido sobre los casos que teníamos en un sistema de información, sería interesante y deseable utilizar este conocimiento para justificar la clasificación de nuevos objetos.

Por nuevos objetos entendemos aquellos que no estén contemplados en el sistema de información del experto y que están descritos únicamente por valores de atributos de condición. La asignación de estos objetos a clases es desconocida y el objetivo del experto es descubrir su asignación sobre la base del conocimiento de su experiencia previa.

Por tanto, se trata, en definitiva, de encontrar en el algoritmo de decisión aquella regla cuya parte de la condición coincida con la descripción del nuevo objeto. Si se encuentra dicha regla la decisión es clara y si no se encuentra, tal y como ya hemos indicado, se buscan las reglas más cercanas.

Asimismo, existen otros problemas de decisión en los que se puede aplicar el enfoque rough set (además del de clasificación). Dado que nuestro objetivo será clasificar empresas nos detendremos en este último tipo de problemas.

La clasificación de cualquier *nuevo objeto* se logra comparando su descripción con cada una de las reglas de decisión obtenidas, es decir, comparando dicho objeto descrito por medio de atributos con las condiciones de las reglas de decisión. Esta comparación puede conducir a una de las siguientes situaciones (Slowinski, 1993 y Slowinski y Stefanowski, 1994):

- a) el nuevo objeto encaja exactamente con una regla de clasificación determinística (exacta),
- b) el nuevo objeto encaja con más de una regla determinística (exacta) indicando, sin embargo la misma clase de decisión,

c) el nuevo objeto encaja con una regla aproximada (no determinística) o más de una regla aproximada que indican diferentes clases de decisión, o más de una regla determinística que también conduzcan a diferentes clases de decisión. Hemos de recordar que la filosofía rough set aplicada a un problema de clasificación implica que las inconsistencias mostradas en los ejemplos de clasificación no se corrigen sino que las reglas producidas se clasifican en precisas o aproximadas. Dentro del contexto de clasificación, las reglas no determinísticas significan que bajo determinadas condiciones, no es posible asignar los objetos de manera unívoca a las categorías a menos que se solicite información adicional.

d) el nuevo objeto no encaja con ninguna de las reglas.

En los casos a) y b), la recomendación es unívoca mientras que en el caso c) de comparación no clara o ambigua, el usuario es informado de la *fuerza* total (el número de objetos clasificados que soportan cada categoría posible, en otras palabras, que verifican cada posible decisión indicada por la regla) de todas las reglas de decisión con las que encaja con respecto a las clases de decisión sugeridas. Si la fuerza de una categoría es mayor (esto es, el número de objetos es mayor que el 51% de todos los objetos que han dado lugar a dicha regla) que la fuerza de otras categorías dentro de las que existen en la regla no-determinística, se puede concluir que de acuerdo a esta regla, el objeto considerado lo más probable es que pertenezca a la categoría *más fuerte*. La información acerca de la fuerza de cada regla puede darse de manera extendida presentando al decisor los objetos que soportan una regla dada.

Si ninguna regla de decisión encaja con el nuevo objeto (caso d), el decisor puede conseguir el conjunto de reglas "más cercanas" a la descripción del objeto. Las reglas más cercanas son reglas que están próximas a la descripción de un objeto clasificado en el sentido de una determinada medida de distancia (en otras palabras, no difieren del objeto de un modo significativo). El motivo de introducir las reglas "más cercanas" resulta de la creencia de que ofrecer al decisor más información sobre la vecindad de un objeto clasificado puede darle alguna sugerencia en lugar de no ofrecerle nada.

La noción de *proximidad* implica, tal y como se ha indicado, el uso de una medida de distancia. El software que hemos utilizado en la parte empírica del trabajo (ROSE) incorpora la medida de distancia denominada métrica L_p (Stefanowski 1993 a y b, Slowinski y Stefanowski 1992).

Antes de finalizar esta sección queremos señalar que los resultados obtenidos por el enfoque rough set se expresan de *forma similar al lenguaje natural humano*. Así, estas representaciones son muy fáciles de comprender por el usuario/analista y permiten justificar y explicar las conclusiones derivadas de análisis de los datos. Además, da la posibilidad al analista de controlar el proceso de análisis de datos de manera simple. Tal posibilidad no es normalmente ofrecida por las técnicas tradicionales de análisis de datos, unido además a su baja carga computacional y posibilidad de interacción con un usuario a modo de diálogo.

II.2. EL MÉTODO ROUGH SET APLICADO A LA PREDICCIÓN INSOLVENCIAS EN EMPRESAS ESPAÑOLAS DE SEGUROS NO VIDA

II.2.1.- Fundamentación del análisis

Se ha mencionado a lo largo de la exposición que en nuestra aplicación práctica vamos a seguir el enfoque original de la teoría rough set y no las extensiones de dicha teoría, que presuponen algún orden en el dominio de los atributos.

En nuestro trabajo vamos a utilizar ratios financieros. El nuevo enfoque permite tener en cuenta las propiedades de ordenación de los atributos analizados, si éstas existen y son conocidas. Por ejemplo, si se sabe que un ratio financiero que toma un valor “alto” es preferible que tome un valor bajo, la firma con el valor del ratio alto se preferirá a la que tiene un valor del ratio bajo, y por tanto se considera que los valores del ratio se podrían ordenar (McKee, 2000, p.162).

La teoría convencional ha considerado a los ratios financieros como atributos que pueden ser ordenados. Por ejemplo, se considera que un ratio como el que relaciona el beneficio neto con el activo total es preferible que tome un valor alto a uno bajo.

Sin embargo cabe cuestionarse, si para los fines de predicción del fracaso empresarial, los ratios financieros pueden considerarse que tengan un dominio ordenado. Un estudio de 1999 (McKee y Lensberg, 1999, p. 23) en el que se utiliza un algoritmo genético para desarrollar un modelo de predicción del fracaso empresarial encontró que la probabilidad de quiebra podría predecirse a través de una función de tres ratios financieros; activo total, beneficio neto/activo total, cash/pasivo circulante. Más aún, se contrastó que el que un valor alto o bajo para un ratio determinado fuese considerado como bueno o malo dependía del valor que tomaran los otros dos ratios. En otras palabras, la teoría convencional de “cuanto mayor sea el valor de un ratio, mejor” quedó cuestionada ya que el estudio mostraba que una rentabilidad alta bajaba las probabilidades de quiebra, excepto cuando esa rentabilidad era inusualmente alta. En este último caso, un aumento en la rentabilidad aumentaba la probabilidad de quiebra, excepto para el caso en que la empresa en cuestión fuese muy pequeña.

De acuerdo a este resultado, sería aconsejable, no asumir (al menos de forma generalizada), que los ratios financieros tienen una relación de dominio. De acuerdo con esto, nuestra investigación utiliza el enfoque original de la teoría rough set basado en la relación de no diferenciación.

II.2.2.- Aplicación del análisis y resultados

Antes de comenzar con la explicación del análisis hemos de mencionar que el software utilizado para efectuar todo el análisis es el programa ROSE facilitado por la Universidad de Poznan (Polonia) en concreto por el *Institute of Computing Science*. Queremos mencionar que cualquier ordenador personal conectado a internet puede acceder a la página web www-idss.cs.put.poznan.pl/rose. donde se puede descargar el programa ROSE, su manual e informarse brevemente de algunas de sus características.

II.2.2.1.- Aplicación De La Metodología Rough Set

Tabla de Información

Para la aplicación de esta metodología lo primero que hemos construido es una tabla de información con las 72 empresas, 36 sanas y 36 fracasadas. Las empresas han sido evaluadas de acuerdo con los valores que toman los 21 ratios elegidos para cada una de ellas. Estas empresas han sido clasificadas en una de las dos categorías, sana o fracasada, para el año que denominaremos *base* (año 1), que es el año anterior a la quiebra. Por tanto estamos considerando un atributo de decisión, d , que efectúa en esta tabla de información (a partir de ahora tabla de decisión) una partición dicotómica: si $d=0$ la firma pertenece al grupo de las fracasadas y si $d=1$, la empresa pertenece al grupo de las sanas.

Hemos de mencionar que aunque se calcularon los ratios 15 y 16, no han sido utilizados en este ni en otros análisis porque las peculiaridades de la muestra hacían que tomasen valores que carecían de sentido económico, al no tener la mayoría de las empresas otros ingresos de explotación. Con lo que solo hemos utilizado los 19 restantes en nuestro sistema de información.

Discretización

Antes de comenzar la aplicación hemos de adecuar la información que contiene nuestro sistema de información.

En las aplicaciones prácticas, los sistemas de información a menudo contienen datos de carácter continuo y discreto. El análisis rough set de los sistemas de información da buenos resultados cuando los dominios de los atributos son conjuntos finitos cuyo número cardinal es relativamente bajo. Este requisito a menudo se satisface cuando los atributos tienen un carácter cualitativo (discreto). Sin embargo, los atributos que toman valores arbitrarios de unos intervalos dados, es decir, tienen un carácter cuantitativo, deben ser tratados para traducir sus valores a términos cualitativos (por ejemplo bajo, medio y alto) utilizando los correspondientes valores numéricos como "1, 2, 3". Esta manipulación implica una división del dominio original en algunos subintervalos y la asignación de códigos cualitativos a dichos subintervalos.

No existe un modo general de definir de manera óptima los valores frontera de los subintervalos. La definición de los valores frontera de los subintervalos debería tener en cuenta la experiencia, el conocimiento, los hábitos y costumbres utilizados por los expertos e, incluso, un error de medida en algunos atributos. Muchas veces los valores frontera surgen normalmente de convencionalismos, hábitos o de asignación subjetiva.

Es necesario resaltar que la definición de los valores frontera de los subintervalos puede influir en los resultados, y en particular, en la calidad de la clasificación, por tanto se hace necesario revisar la sensibilidad de los resultados frente a cambios en la definición de estos valores.

Por tanto, vamos a traducir los valores cuantitativos de los ratios financieros a términos cualitativos. Por esta razón vamos a considerar dos tipos de sistema de información: *el sistema de información original* y *el sistema de información codificado*. El primero contiene los ratios financieros (atributos cuantitativos) con los valores de sus dominios originales. El segundo contiene atributos con valores codificados. La teoría rough set analiza posteriormente el sistema de información codificada.

La codificación de los valores de los ratios la hemos efectuado sustituyendo los valores originales por códigos que, en nuestro caso, son números que designan a los subintervalos que los contienen. Es decir, se han codificado utilizando la denominada *tabla de normas*.

La tabla de normas es una estructura que contiene las definiciones de todos los valores frontera de los subintervalos definidos por el usuario.

En consecuencia, para cada ratio se ha buscado en cada grupo de empresas (sanas y fracasadas) cuales son sus niveles más representativos, con el objetivo final de establecer varios niveles de corte en los valores del ratio, que separen distintos estados de equilibrio o desequilibrio respecto a ese ratio.

Para el establecimiento de esos puntos de separación, es decir, las *normas*, hemos calculado los cuartiles de cada ratio. Es decir, los percentiles 25, 50, y 75 de cada ratio nos informan, respectivamente, de las empresas que toman un valor para dicha variable por debajo de ese percentil. Además la utilización de los cuartiles evita cualquier subjetividad y no excluye la posibilidad de que la experiencia o conocimiento de un experto modifiquen dichos valores. La tabla 2.2 muestra los valores utilizados para discretizar los distintos ratios.

La utilización de percentiles para establecer intervalos en los ratios es frecuente en los trabajos de investigación. Podemos citar otras investigaciones que utilizan este procedimiento: Laitinen (1992), García *et al.*, (1997) o McKee, (2000).

TABLA 2.1. INTERVALOS

Ratio	1°	2°	3°	4°
R1	$(-\infty, 0.115]$	$(0.115, 0.295]$	$(0.295, 0.475]$	$(0.475, +\infty)$
R2	$(-\infty, 0]$	$(0, 0.1]$	$(0.1, 0.07]$	$(0.07, +\infty)$
R3	$(-\infty, 0.03]$	$(0.03, 0.06]$	$(0.06, 0.11]$	$(0.11, +\infty)$
R4	$(-\infty, 0.03]$	$(0.03, 0.08]$	$(0.08, 0.26]$	$(0.26, +\infty)$
R5	$(-\infty, 0.565]$	$(0.565, 1.565]$	$(1.565, 3.29]$	$(3.29, +\infty)$
R6	$(-\infty, 0.525]$	$(0.525, 1.38]$	$(1.38, 2.715]$	$(2.715, +\infty)$
R7	$(-\infty, 0.455]$	$(0.455, 0.725]$	$(0.725, 1.22]$	$(1.22, +\infty)$
R8	$(-\infty, 0.46]$	$(0.46, 0.7]$	$(0.7, 1.18]$	$(1.18, +\infty)$
R9	$(-\infty, 0.14]$	$(0.14, 0.35]$	$(0.35, 0.68]$	$(0.68, +\infty)$
R10	$(-\infty, 0.04]$	$(0.04, 0.545]$	$(0.545, 2.97]$	$(2.97, +\infty)$
R11	$(-\infty, 0.27]$	$(0.27, 1.095]$	$(1.095, 2.43]$	$(2.43, +\infty)$
R12	$(-\infty, 0.27]$	$(0.27, 0.845]$	$(0.845, 1.815]$	$(1.815, +\infty)$
R13	$(-\infty, 0.27]$	$(0.27, 0.49]$	$(0.49, 0.82]$	$(0.82, +\infty)$
R14	$(-\infty, 0.225]$	$(0.225, 0.435]$	$(0.435, 0.765]$	$(0.765, +\infty)$
R17	$(-\infty, 0.98]$	$(0.98, 1.055]$	$(1.055, 1.27]$	$(1.27, +\infty)$
R18	$(-\infty, 1]$	$(1, 1.09]$	$(1.09, 1.29]$	$(1.29, +\infty)$
R19	$(-\infty, 0]$	$(0, 0.065]$	$(0.065, 0.19]$	$(0.19, +\infty)$
R20	$(-\infty, 0.515]$	$(0.515, 0.68]$	$(0.68, 0.785]$	$(0.785, +\infty)$
R21	$(-\infty, 0.515]$	$(0.515, 0.655]$	$(0.655, 0.75]$	$(0.75, +\infty)$

Hemos de resaltar que la utilización de *normas* que transformen los atributos cuantitativos en términos cualitativos no se impone por el rough set pero sí para una interpretación práctica de los atributos cuantitativos. Incluso si un atributo representa una medida continua, como es un ratio financiero, o la presión sanguínea en medicina, el experto normalmente interpreta los valores de los atributos en términos cualitativos, esto es, bajo, medio o alto. Es más, este proceso se hace necesario debido a que la precisión de los ratios financieros es bastante dudosa, y dificulta el extraer conclusiones generales de los datos en términos de dependencias, reductos o reglas de decisión.

Además, y dado que las normas se utilizan desde el principio del problema y hasta el final en la explicación de las reglas de decisión, no falsean la imagen original de la situación.

Las empresas para las que los valores de estos ratios financieros están en los mismos subintervalos se considera que tienen el mismo comportamiento y características. Además, los códigos que se utilizan para referirse a un

subintervalo no representan ningún orden de preferencias. La teoría rough set no tiene en cuenta el orden, ya que trabaja con la relación de no diferenciación, es más la selección de los códigos no afecta a los resultados obtenidos.

Una vez determinados los intervalos hay que asignar los códigos a los mismos.

Con la tabla anterior y el principio “cuanto más alto sea el código, mejor subintervalo”, tenemos las bases para obtener la tabla de asignación de códigos. Dado el principio que hemos aplicado, para los atributos decrecientes (aquellos en que los valores más bajos son mejores) se les ha asignado un código en orden inverso. Además, hemos efectuado algunas correcciones en la escala en el caso en que, nuestro juicio y/o preferencias, no estuvieran en concordancia con la secuencia de aumento o disminución de los subintervalos.

La siguiente tabla muestra la asignación de códigos a los subintervalos:

TABLA 2.2. CÓDIGOS ASIGNADOS A LOS INTERVALOS

Ratios	Intervalos			
	1°	2°	3°	4°
R1	1	2	3	4
R2	1	2	3	4
R3	1	2	3	4
R4	1	2	3	4
R5	1	3	4	2
R6	1	3	4	2
R7	1	3	4	2
R8	1	3	4	2
R9	1	3	4	2
R10	1	3	4	2
R11	1	4	3	2
R12	1	4	3	2
R13	1	4	3	2
R14	1	4	3	2
R17	1	4	3	2
R18	1	4	3	2
R19	1	3	3	2
R20	4	3	2	1
R21	4	3	2	1

Las siguientes tablas han sido obtenidas aplicando los códigos a los valores de los ratios, obteniendo así las tablas de información codificada.

TABLA 2.3. EMPRESAS FRACASADAS

FIRMS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R17	R18	R19	R20	R21	D
E1	3	1	3	3	4	4	4	4	3	4	3	3	4	3	4	3	3	3	2	0
E2	2	1	2	2	3	3	4	2	4	3	4	3	2	2	4	1	1	2	2	0
E3	1	1	1	2	3	3	4	4	3	3	4	3	2	2	3	3	1	1	1	0
E4	2	3	1	3	2	2	4	4	3	2	2	2	3	3	4	4	3	3	3	0
E5	1	2	3	1	1	1	1	1	2	3	1	1	1	1	2	2	2	4	4	0
E6	1	1	3	4	1	1	4	4	1	1	1	1	3	3	2	2	3	2	1	0
E7	1	4	1	4	2	2	2	2	1	2	2	2	3	2	1	1	3	3	3	0
E8	2	1	3	4	2	2	4	4	1	2	2	2	3	3	3	4	3	2	2	0
E9	3	1	2	4	2	2	4	4	1	2	2	2	3	3	4	4	3	2	2	0
E10	1	1	1	4	1	1	4	4	1	1	1	1	3	3	3	3	1	2	2	0
E11	1	1	1	2	1	1	4	4	1	1	1	1	3	3	4	4	3	3	3	0
E12	1	1	4	4	1	1	2	2	1	1	1	1	2	2	3	3	1	3	3	0
E13	2	1	2	2	2	2	4	4	1	2	2	2	2	3	2	3	2	1	3	0
E14	2	4	1	3	2	2	4	4	3	2	3	3	4	4	1	1	3	3	3	0
E15	2	3	4	3	2	4	4	3	1	2	2	3	3	3	4	3	2	3	2	0
E16	1	1	4	1	1	1	2	1	1	1	1	1	2	1	2	2	3	1	1	0
E17	1	1	2	4	1	1	2	2	1	1	1	1	2	2	2	2	3	1	1	0
E18	2	4	4	3	1	1	1	1	4	3	1	1	1	1	2	2	3	2	2	0
E19	1	1	1	4	2	2	4	4	1	2	2	2	3	3	2	2	3	1	1	0
E20	2	1	2	3	4	4	1	1	3	4	3	3	4	4	3	3	3	1	1	0
E21	3	1	4	4	3	3	3	4	4	3	1	1	1	1	2	2	1	4	4	0
E22	1	2	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	4	4	0
E23	3	4	2	4	4	2	2	2	4	3	3	2	2	2	1	1	1	2	2	0
E24	2	1	2	2	3	3	3	3	3	4	4	4	4	4	1	4	2	4	3	0
E25	2	3	4	1	4	4	1	3	1	2	3	4	4	1	2	3	2	1	4	0
E26	4	4	3	3	4	4	2	2	2	3	4	3	2	2	1	1	1	3	2	0
E27	2	1	2	2	2	2	2	2	4	3	2	2	2	2	4	4	1	2	2	0
E28	1	2	1	1	1	3	3	3	4	1	4	4	1	4	1	1	1	3	3	0
E29	4	1	4	1	1	1	1	1	2	3	1	1	1	4	3	3	2	2	1	0
E30	4	4	1	3	4	2	2	2	4	4	3	2	2	2	1	1	1	2	2	0
E31	2	1	1	4	1	1	1	1	2	1	1	1	1	1	2	2	1	4	4	0
E32	2	1	2	3	3	3	3	1	4	4	4	4	4	4	2	2	2	2	3	0
E33	4	1	1	1	2	2	2	2	4	1	2	2	2	2	4	4	1	3	3	0
E34	1	3	1	3	2	2	2	2	4	4	2	2	2	2	4	1	1	4	4	0
E35	3	4	4	3	4	4	2	2	2	3	3	3	2	2	1	1	1	3	3	0
E36	4	3	2	2	4	4	2	2	2	3	3	3	2	2	1	1	1	1	1	0

TABLA 2.4. EMPRESAS SANAS

FIRMS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R17	R18	R19	R20	R21	D
E101	4	3	3	1	1	1	1	1	3	4	4	4	1	1	2	2	3	1	1	1
E102	4	1	3	1	3	3	3	3	2	1	1	1	1	1	3	4	1	4	4	1
E103	4	4	3	4	3	3	4	4	2	1	4	4	3	3	1	1	1	4	4	1
E104	3	4	2	3	4	4	1	3	1	2	3	3	1	4	3	4	2	2	3	1
E105	4	4	3	4	3	3	3	3	4	4	4	4	1	1	1	1	2	4	4	1
E106	3	1	4	2	2	2	4	4	1	2	2	2	3	3	4	3	2	1	1	1
E107	4	4	3	3	4	4	3	3	3	4	3	3	4	4	1	4	3	3	2	1
E108	3	1	4	4	2	2	3	3	1	2	2	2	3	3	2	2	3	1	1	1
E109	4	3	3	1	4	3	3	1	3	2	3	4	4	1	1	3	2	3	3	1
E110	3	3	3	1	3	3	1	1	3	4	4	4	4	4	2	3	2	1	2	1
E111	3	3	4	3	2	2	3	3	1	2	2	2	4	3	3	3	3	1	1	1
E112	3	4	4	3	4	4	2	2	4	3	3	2	2	2	3	3	1	1	1	1
E113	4	3	4	1	3	3	1	1	3	4	3	4	4	4	3	2	2	1	1	1
E114	2	1	3	2	4	4	3	3	3	2	2	3	3	4	3	3	3	1	2	1
E115	1	1	2	1	4	4	3	3	3	4	3	3	4	3	2	2	3	1	1	1
E116	3	1	2	4	2	2	4	4	1	2	2	2	3	3	3	3	3	2	2	1
E117	3	1	3	4	4	4	1	1	3	2	3	3	4	4	2	2	3	1	1	1
E118	4	4	3	3	4	4	2	2	4	4	4	4	4	4	1	1	1	4	4	1
E119	3	1	4	2	2	4	3	3	1	2	2	3	3	4	1	2	2	2	3	1
E120	3	1	3	1	3	3	3	3	3	4	4	4	4	4	3	3	3	2	2	1
E121	2	1	1	2	2	2	2	2	4	4	2	2	2	2	4	4	1	4	4	1
E122	2	2	3	1	1	1	3	3	2	1	1	1	1	1	3	4	1	4	4	1
E123	4	3	2	2	3	3	4	4	2	3	4	4	3	3	4	4	1	3	3	1
E124	2	2	3	1	4	4	1	3	3	4	3	3	4	4	3	3	2	3	3	1
E125	1	3	2	2	3	3	3	3	3	4	4	4	1	1	1	1	3	4	4	1
E126	4	3	2	2	3	3	3	3	2	3	4	4	4	4	4	4	1	2	2	1
E127	3	4	3	4	4	4	2	2	4	3	3	3	3	2	4	4	1	3	3	1
E128	1	1	1	3	3	3	4	4	2	1	4	4	3	3	3	3	1	2	2	1
E129	3	2	1	2	3	3	3	3	4	3	4	4	1	4	1	4	3	4	4	1
E130	1	2	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	4	4	1
E131	4	3	2	1	3	4	2	2	2	1	3	3	2	2	4	4	1	2	1	1
E132	3	3	3	2	3	1	1	1	4	4	4	4	1	1	3	2	2	2	2	1
E133	3	3	1	1	2	2	2	2	2	3	3	2	2	2	4	4	1	3	3	1
E134	1	2	1	1	1	1	1	1	2	1	1	1	1	1	4	1	1	4	4	1
E135	2	1	1	2	1	1	1	1	2	1	1	1	1	1	2	2	3	4	4	1
E136	4	1	1	1	1	1	1	1	2	1	1	1	1	1	2	2	1	4	4	1

La tabla codificada, que es la que se hemos utilizado para todo el análisis posterior, constaba de las 72 empresas descritas por los 19 atributos codificados (ratios financieros), utilizando los datos del año previo a la quiebra (año -1) y asignadas de forma binaria a una clase de decisión (sana o fracasada, representada por 1 ó 0 respectivamente).

Experimentos Realizados

Hemos realizado dos tipos de experimentos, obteniendo de cada uno de ellos dos modelos, por lo que nos referiremos a ellos como **modelo 1** y **modelo 2**.

Las reglas de decisión del modelo 1 se han obtenido empleando los datos del año 1 (año anterior a la quiebra) y verificando la capacidad predictiva de las mismas utilizando los datos de los años 2, 3, 4 y 5 previos a la quiebra. Posteriormente se ha efectuado un análisis discriminante siguiendo los mismos criterios que para el análisis rough set, es decir, estimar la función discriminante con el año 1 y comprobar la capacidad predictiva de dicha función utilizando los años 2, 3, 4 y 5 anteriores a que las firmas quebraran.

Dado que utilizar las mismas empresas para validar y para estimar puede condicionar los resultados obtenidos hemos obtenido un segundo modelo siguiendo este procedimiento: De las 72 empresas de que consta la muestra original (36 sanas y 36 fracasadas), y para el año 1 (año previo a la quiebra) hemos utilizado únicamente 27 empresas de cada una de las submuestras para estimar las reglas de decisión, dejando las 9 restantes para comprobar la validez de las mismas. Pretendemos de esta forma aplicar el modelo a datos que no hubieran sido utilizados para la elaboración del mismo. En consecuencia, en este segundo modelo tendremos una muestra de *entrenamiento* para obtener las reglas formada por 54 empresas y una muestra de *validación* para verificar el modelo formada por 18 empresas. La selección de la muestra de validación se ha efectuado aleatoriamente tomando las empresas numeradas en Sanchis (2000, 2002) de la 19 a la 27, para las fracasadas, y de la 119 a la 127 para las sanas como submuestras de validación.

Dado que muchos resultados son comunes a ambos experimentos los iremos comentando conjuntamente y refiriéndonos a ellos como modelo 1 y modelo 2 en el caso en que sean comentarios específicos a cada uno de ellos.

Precisión y Calidad de la Aproximación

El primer resultado que obtenemos con el análisis rough set es la aproximación para cada una de las clases que estamos considerando, sana y fracasada, por el conjunto de atributos (ratios financieros). En ambos modelos la totalidad del conjunto de atributos dio una aproximación perfecta de las clases de decisión, esto es igual a uno. Además y, en consecuencia, la calidad de la clasificación es también igual a uno.

El resultado es muy significativo porque a pesar de tener un gran sistema de información, las empresas están muy bien discriminadas entre ellas. Es decir, no hay casos de *duda* o casos *fronterizos*, por tanto no hay dos empresas caracterizadas por los mismos ratios y que una pertenezca a las sanas y otra a las fracasadas.

Núcleo y Reductos

El siguiente paso del análisis es la construcción del mínimo subconjunto de atributos independientes que aseguren la misma calidad de clasificación que la totalidad del conjunto de todos ellos, esto es los reductos, a partir del sistema de información codificado.

Para el primer modelo hemos obtenido 241 reductos con una longitud de los mismos de entre 8-5 atributos por reducto lo que indica que como mínimo, en este primer modelo, 11 atributos son redundantes, y en consecuencia podrían ser eliminados.

Los atributos que con más frecuencia aparecen en los reductos (al menos en un 40% de ellos) son, por orden de mayor frecuencia: R4, R17, R9, R18, R3, R1, R19.

En el segundo experimento hemos obtenido 229 reductos de la tabla de información codificada cuya longitud varía entre 4 y 7 atributos por reducto lo que indica que como mínimo, en este segundo modelo, 12 atributos son redundantes, y en consecuencia podrían ser eliminados.

Los atributos que aparecen con mayor frecuencia (más del 40%) en los reductos por orden de mayor frecuencia son: R9, R4, R3, R1, R19, R17.

De los resultados que hemos obtenido se pueden extraer dos primeras conclusiones.

La primera de ellas es que en ambos casos los ratios que con más frecuencia aparecen en los reductos son R1, R3, R4, R9, R17, R18 y R19 lo cual indica que son variables muy discriminatorias en la muestra entre empresas solventes y no solventes. Esto demuestra que a la hora de evaluar la solvencia de una empresa de seguros hay que tener en cuenta las siguientes cuestiones:

- a) La *liquidez* medida a través del ratio R1: Una de las cuestiones más importantes para asegurar el adecuado funcionamiento de cualquier empresa es la necesidad de tener suficiente liquidez. Pero en el caso de una empresa de seguros, este problema no debería aparecer, y si aparece reviste una mayor importancia que en otro tipo de empresas, ya que la inversión del proceso productivo implica que las primas se cobran antes de que acaezcan los eventuales siniestros. Si una aseguradora no puede pagar los siniestros que ocurran, sus clientes y el público en general perderían la confianza en dicha compañía. Por otro lado, este ratio es una medida del equilibrio financiero, si es positivo, ya que indicaría que el capital circulante también lo es.
- b) La *rentabilidad* de las inversiones medida a través del ratio R3. Ya hemos comentado que los rendimientos financieros de las entidades aseguradoras son en la actualidad la fuente principal de sus beneficios, o lo que es lo mismo, representan la posibilidad de obtener una rentabilidad adecuada para los recursos propios y también una fuente

- de financiación importante de los incrementos del Margen de Solvencia.
- c) Una política *apropiada* de *financiación* medida a través del ratio R4. Una financiación escasa debida a ingresos insuficientes que aparezcan de manera inesperada puede ser la causa del comienzo de un proceso de quiebra.
 - d) Unos adecuados *capitales propios* medidos a través del ratio R9. Se confirma la importancia, desde el punto de vista de la solvencia, de estos fondos destinados a satisfacer, en caso necesario, futuros siniestros.
 - e) La importancia de una conveniente *tarificación* medida a través de los ratios R17 y R18. Una mala tarificación lleva a una insuficiencia de las primas en relación con los costes tanto de los siniestros como del resto.
 - f) Una *adecuada gestión del reaseguro* (R19).

El segundo comentario es que al menos 11 ó 12 atributos, según se trate de primer experimento o del segundo, son redundantes y podrían ser eliminados de la tabla. Esto demuestra la fuerza de esta metodología para la selección de las variables más significativas.

La intersección de los reductos nos da el núcleo. El núcleo de atributos en las dos pruebas realizadas estaba vacío. Esto indica que ningún único atributo es absolutamente necesario para la aproximación perfecta de las dos clases de decisión. Si el núcleo no hubiese sido vacío indicaría que existen atributos en el sistema que son indispensables desde el punto de vista discriminante, porque la no inclusión de cualquiera de los atributos contenidos en el núcleo conduce inmediatamente a una disminución de la calidad de la aproximación. Por otra parte, un núcleo no vacío ayuda a determinar los atributos más importantes además de la aproximación de las clases con las que está relacionado.

El siguiente paso consiste en la elección del reducto. En ambos casos para la selección del mismo hemos tenido en cuenta los siguientes criterios:

- a) El reducto debía contener el menor número de atributos como fuese posible.
- b) Debía contener los atributos considerados, en nuestra opinión, más significativos para la evaluación de la solvencia de las empresas.
- c) Después de seleccionar unos pocos reductos que cumpliesen los dos primeros requisitos, la tercera condición es que el reducto que se fuese a seleccionar no tuviera ratios que estuviesen correlacionados. Por tanto hemos calculado algunos coeficientes de correlación.

Los reductos que hemos elegido para cada uno de las pruebas son el formado por **R1, R3, R9, R14, R17 y R19** para el primer modelo y el formado por **R3, R4, R9, R14, R17** para el segundo modelo.

Elegidos los reductos, el resto de atributos se pueden eliminar de la tabla de información codificada. Por tanto en el primer caso, con el reducto se pasó de

un sistema inicial de información codificado de 19 columnas a otro de 6, y en el segundo caso se pasó a otro de 5 columnas.

II.2.2.2.- Resultados

El sistema de información reducido puede verse como una tabla de decisión de la cual derivaremos las reglas de decisión.

De las posibles estrategias que incorpora ROSE para derivar las reglas de decisión de la tabla de información codificada y reducida, hemos seleccionado la estrategia consistente en la obtención del mínimo conjunto de reglas que cubren todos los objetos de la tabla de decisión.

El primer modelo obtenido se compone de 25 reglas y el segundo modelo obtenido se compone de 27 reglas. Las tablas siguientes muestran los dos algoritmos obtenidos:

TABLA 2.5. REGLAS DE DECISIÓN –MODELO 1

Nº Regla	R1	R3	R9	R14	R17	R19	Decisión	Fuerza	Empresas
1	2	2					0	6	2,13,20,24,27,32
2				2	1		0	6	7,23,26,30,35,36
3		1	1				0	8	6,7,10,11,12,16,17,19
4	1		4				0	3	22,28,34
5		1	3				0	3	3,4,14
6					4	3	0	4	1,4,9,11
7			2			2	0	2	5,29
8		4		1			0	4	16,18,21,25
9	2				2	1	0	1	31
10	4		4		4		0	1	33
11	2			3			0	4	4,8,13,15
12		3		4			1	7	43,46,50,53,54,56,60
13			2		4		1	5	59,62,67,69,70
14	4			1			1	5	37,38,41,45,72
15	3	4	1				1	4	42,44,47,55
16	3				3		1	6	40,47,48,52,56,68
17	1		2			1	1	3	64,66,70
18	1		3			3	1	2	51,61
19			2	3			1	3	39,59,64
20		1		1		3	1	1	71
21	2	1		2			1	1	57
22	3		4		4		1	1	63
23	4		3				1	4	37,43,45,49
24				1	3		1	3	38,58,68
25	3			4			1	6	40,46,53,55,56,65

TABLA 2.6. REGLAS DE DECISIÓN –MODELO 2

Nº Regla	R3	R4	R9	R14	R17	Decisión	Fuerza	Empresas
1	2		4			0	4	2, 14, 18, 23
2		3		2		0	4	17, 21, 25, 26
3	1	4				0	3	7, 10, 22
4		1	4			0	3	13, 19, 24
5			3	3		0	2	1, 4
6	3		1			0	2	6, 8
7	2		3	4		0	2	11, 15
8	4			1		0	2	12, 16
9	4	1				0	2	16, 20
10	2		1	3		0	1	9
11			3	2		0	1	3
12	3		2		2	0	1	5
13	2	2		2		0	3	2, 18, 27
14				1	3	1	3	29, 40, 50
15			2		4	1	5	41, 44, 49, 51, 52
16		1	3			1	4	28, 36, 38, 42
17	4	2				1	2	33, 37
18	3	4			1	1	2	30, 32
19	1	2	4			1	2	39, 47
20		2			2	1	1	53
21		3	1			1	1	31
22	2			1		1	2	43, 48
23	3			4		1	3	34, 38, 42
24	4			3		1	2	33, 35
25	1	3	2			1	1	46
26	1	1			2	1	1	54
27		4		2	4	1	1	45

Es necesario recordar que las reglas son sentencias lógicas. Así por ejemplo la regla número 1 del *modelo 2* se leería: *Si* el ratio R3 toma el valor 2 (0.03, 0.06] y el ratio R9 toma el valor 4 (0.35, 0.68), *entonces* la empresa pertenece al grupo de las fracasadas. La fuerza de esta regla es de 4 porque son 4 las empresas que verifican este hecho en nuestra muestra.

Las reglas de decisión derivadas de la tabla de información codificada son, en ambos casos, reglas exactas dado que la calidad de la clasificación es igual a uno y, en consecuencia, las fronteras de las clases de decisión estaban vacías. Es decir, dado que los reductos dan una calidad de clasificación igual a la original, es decir a uno, todas las reglas son determinísticas, esto es, describen únicamente la decisión que ha de tomarse cuando se satisfacen algunas condiciones.

En el primer modelo la fuerza de las reglas varía entre una longitud máxima de 7 a una mínima de 1, mientras que en el segundo modelo la longitud máxima es de 5 y la mínima vuelve a ser de 1.

Las reglas obtenidas constituyen para los dos modelos un algoritmo de clasificación. Son una representación no redundante del conocimiento contenido en la muestra seleccionada y están reunidas en un sistema de información codificado.

Si observamos en los algoritmos obtenidos sólo se han utilizado 58 y 63 descriptores (ratios calculados) para cada modelo, lo que supone menos del 5% (4.24% y 4.6%, respectivamente) de los que aparecían en el sistema de información inicial (1.368).

En general, cuanto más corta es una regla más general es. En los dos conjuntos de reglas de decisión, el número de reglas que describen a las empresas no quebradas es mayor que el número de reglas que describen a las quebradas. Esto significa que, en nuestra muestra, cuesta más generalizar la descripción de las empresas sanas que la de las quebradas.

La muestra de 72 empresas utilizada en el estudio para derivar los algoritmos de decisión puede considerarse como una muestra de prueba utilizada para revelar las características financieras que discriminan las empresas solventes de las insolventes.

II.2.3.- Validación de los resultados: clasificación de empresas utilizando los dos algoritmos generados

Los dos algoritmos de clasificación pueden ser utilizados para evaluar cualquier otra empresa. Por tanto, la precisión predictiva debe contestarse en el curso de la verificación práctica ya que los algoritmos de clasificación obtenidos revelan únicamente qué criterios hay que tener en cuenta para evaluar la solvencia.

Por tanto las reglas de decisión obtenidas a partir del primer modelo han sido probadas con los datos de los años 2, 3, 4 y 5 antes del año base o año 1 (año antes de la quiebra). El segundo algoritmo obtenido ha sido validado con la muestra independiente formada por las 18 empresas que constituían la submuestra de validación.

Las siguientes tablas muestran los porcentajes de clasificaciones correctas obtenidas con la aplicación del método rough set para cada uno de los experimentos realizados:

TABLA 2.7. MODELO 1

Clasificaciones correctas	Año 1	Año 2	Año 3	Año 4	Año 5
Clase 0 (empresas quebradas)	100%	77.14%	55.56%	70.83%	63,16%
Clase 1 (empresas sanas)	100%	80%	76,67%	59,26%	76,19%
Total	100%	78.57%	66.67%	64,71%	70%

TABLA 2.8. MODELO 2

Clasificaciones correctas	Año1
Clase 0 (empresas quebradas)	77.78%
Clase 1 (empresas sanas)	77.78%
Total	77.78%

Debemos recordar que el programa ROSE utiliza una medida de distancia para reclasificar aquellas empresas que no encajan con ninguna regla y que es aplicada directamente, por lo que estos resultados ya incluyen aquellas empresas que no coincidiendo con ninguna regla son clasificadas con las reglas más cercanas según la medida de distancia ya comentada.

En general y respecto a los dos modelos podemos comentar que los resultados obtenidos son bastante satisfactorios. Además, en el primer modelo, también se puede observar como lógicamente el porcentaje de clasificaciones correctas va disminuyendo conforme avanzamos hacia atrás en el tiempo y nos alejamos del año 1 o año previo a la quiebra excepto para el año 5. Una posible explicación puede ser que 5 años antes de la quiebra las empresas no tratan de manipular sus estados contables para ocultar los problemas financieros que atraviesan en vistas a que la situación puede ser reversible. Pero conforme esa situación, y tal y como hemos podido comprobar al analizar los balances y cuentas

de resultados de las empresas que constituyen la muestra, se va complicando, empiezan a aparecer errores e irregularidades en las cuentas en aras de ocultar los problemas financieros por los que atraviesan. Pero llega un momento en que esa problemática situación es tan crítica que no se puede ocultar volviendo por tanto a mejorar en el año 2 la capacidad predictiva del modelo.

II.2.4.- Comparación del análisis rough set con el análisis discriminante

En general, la discriminación de tipo estadístico tiene gran aplicación cuando se trata de clasificar personas, entidades o cosas como pertenecientes a dos o más poblaciones dadas, en base a la observación de los valores que alcanzan unas variables, previamente elegidas, de carácter comúnmente cuantitativo.

En el caso de las finanzas aparecen aplicaciones obvias, por ejemplo:

- Clasificar, a los solicitantes de un crédito, en clientes potencialmente buenos o malos en función de determinados datos contables como pueden ser el grado de endeudamiento o la liquidez de su negocio.

- Clasificar, a un grupo de empresas, en candidatas potenciales al fracaso o al éxito en los negocios, según ciertos ratios contables tales como rentabilidad, endeudamiento, liquidez, proporción de beneficios retenidos, etc.

El análisis discriminante puede ser una buena herramienta de trabajo en el diagnóstico y pronóstico de crisis empresariales.

Aunque la filosofía de las dos técnicas es muy diferente, ambas metodologías se pueden aplicar a la predicción del fracaso empresarial sobre un conjunto de datos comunes.

El análisis discriminante es una técnica estadística que se utiliza para clasificar objetos en diferentes grupos basándose en la observación de algunas características de dichos objetos. Básicamente consiste en estimar una función lineal discriminante que calcula un resultado para cada objeto. Esta función es una combinación lineal ponderada de los valores que toman las características observadas del objeto. Las ponderaciones representan, en esencia, la importancia relativa y el impacto de las características que forman parte de la función discriminante. El objeto se clasifica en un determinado grupo o clase basándose en la puntuación obtenida con la función discriminante.

El análisis discriminante está sujeto a un gran número de hipótesis tales como: cada grupo debe seguir una distribución normal multivariante; las matrices de covarianzas de cada grupo son idénticas; y, las matrices de covarianzas, las probabilidades a priori y los errores de clasificación han de ser conocidos. Si no se cumplen estas hipótesis los resultados que obtenemos pueden ser erróneos.

Sin embargo, en la práctica, estas hipótesis no se suelen verificar y muchos investigadores han cuestionado su aplicabilidad (Eisenbeis, 1977; Ohlson, 1980; Zavgren, 1985).

Desde un punto de vista estadístico formal las principales críticas se han formulado por parte de Eisenbeis (1977) que señala los siguientes problemas:

- La distribución de las variables empleadas debería ser normal multivariada para que el análisis discriminante funcionase

correctamente. Sin embargo, en la práctica, los ratios contables no suelen presentar distribuciones normales.

- En el análisis de funciones lineales, como las empleadas por casi todos los modelos, se supone la igualdad de las matrices de covarianzas que indican el grado de dispersión de los grupos. Si dicha igualdad no se da, caso probable, habrían de emplearse mejor funciones cuadráticas.
- La significatividad de las variables del modelo no es fácil de dilucidar. No existen pruebas para detectar el impacto absoluto de una variable determinada. Solo pueden proponerse pruebas par determinar la importancia relativa de las variables individuales, tales como el test F a escala univariada y el ajuste de los coeficientes de la función discriminatoria al multiplicarlos por la diagonal de la matriz de covarianzas correspondiente.
- Cuando se reduce la dimensión del modelo eliminando ciertas variables dicha reducción afectará a los resultados de la clasificación, aunque se eliminen variables aparentemente no significativas.
- El análisis no puede aplicarse correctamente cuando una variable, intrínsecamente continua, se segmenta y se utiliza como una base de formación de grupos. Solo tiene sentido formar grupos, basados en una determinada variable, si aparece una ruptura o discontinuidad natural como en el caso de empresas en quiebra y empresas sanas.
- Las reglas de clasificación deben incorporar las probabilidades “a priori” de fracaso y no fracaso y los costes relativos de los errores. En la práctica, estos aspectos no se tienen en cuenta.
- Si para estimar los errores del modelo se reclasifica la muestra original, la estimación conducirá a resultados altamente sesgados y en exceso optimistas. Conviene, por ello, emplear muestras o submuestras diferentes, una para estimar la función y otra para evaluar los errores de clasificación.

Además debemos señalar como puntos débiles evidentes, en aplicaciones financieras del análisis discriminante, los siguientes:

- Limitaciones sectoriales o industriales. Los modelos funcionan mejor con muestras homogéneas de empresas en cuanto a sector o tipo de industria.
- Limitaciones dimensionales. Los predictores del fracaso de una gran empresa pueden no ser exactamente los mismos que en una pequeña, por ello suele acudir al emparejamiento de empresas, saneada y fracasada, por dimensión. Motivo por el cual la muestra de la que partimos está emparejada por dimensión.

- Aplicaciones a fenómenos extremos. La crisis observable ha de ser de carácter terminal, quiebra o liquidación, embotándose el análisis ante situaciones no extremas como problemas de liquidez o dificultades financieras pasajeras.

Por lo tanto, en general, podemos hablar de falta de validación externa de los modelos expuestos aunque su validez interna suele ser bastante alta.

Como ventajas tendríamos:

- Obtención de una puntuación única que permite una jerarquización de las empresas analizadas, así como una clasificación dicotómica entre saneadas y candidatas potenciales al fracaso.
- Utilización de los modelos, en el seno de una determinada empresa, para evaluar la propia situación financiera y poner en marcha las consiguientes acciones preventivas. Estaríamos ante el uso normativo de los modelos de predicción.

Por tanto, pese a estas últimas ventajas que hemos citado, en la práctica las hipótesis estadísticas que requiere el discriminante raramente se satisfacen. En consecuencia, en la mayoría de las ocasiones los ratios financieros no satisfacen los supuestos estadísticos mencionados por lo que cabría preguntarse por las razones que justifican su elección como método para comparar con la teoría rough set. La elección del discriminante se justifica porque ha sido el primer método que se aplicó y, en consecuencia, el más utilizado para problemas de clasificación y predicción dentro del campo financiero y de gestión. Además, ha dado muy buenos resultados empíricos en problemas financieros en los que intervienen como variables ratios contables pese a que estos no satisfagan las hipótesis estadísticas.

A diferencia del discriminante, el enfoque rough set no necesita que los datos verifiquen a priori ningún tipo de hipótesis pero una descripción más detallada y comparativa de los dos métodos se encuentra en Krusinnska et al. (1992) y Stefanowski (1992).

Para comparar los dos métodos hemos derivado una función discriminante utilizando los ratios del reducto seleccionado en su forma original (sin codificar). La siguiente tabla muestra los coeficientes obtenidos de la función discriminante:

Variable	Coficiente
Término Constante	-0,62358
R1	4,68606
R3	-0,29266
R9	0,24881
R14	-0,03074
R17	-0,76597
R19	1,86133

En cuanto a los porcentajes de clasificaciones correctas obtenidas con la aplicación del discriminante son las siguientes:

TABLA 2.9. RESULTADOS ANÁLISIS DISCRIMINANTE

Clasificaciones correctas	Año 1	Año 2	Año 3	Año 4	Año 5
Clase 0 (empresas quebradas)	76.47%	78.79%	87.5%	72.72%	81.25%
Clase 1 (empresas sanas)	55.88%	64.71%	61.53%	43.47%	50%
Total	66.17%	71.75%	74.51%	58.09%	65.62%

Estos resultados provienen de asumir una determinada relación de probabilidades de clasificación errónea a la hora de estimar las probabilidades de pertenencia a cada grupo para cada unidad. Estas probabilidades estimadas de pertenencia a un grupo se utilizaron para clasificar las unidades en cada grupo, tomando como punto de corte una probabilidad de pertenencia a un grupo de 0.5, dado que en nuestro caso trabajamos con dos grupos. La idea es la siguiente normalmente se compensa el hecho de lo que interesa captar (la insolvencia) tiene una probabilidad de ocurrencia reducida, con el hecho de que la valoración que se hace de cometer un error en el sentido de clasificar una unidad de ese grupo reducido (insolventes) en el grupo más amplio (solventes) es mucho peor que el error ocurra en sentido contrario.

La siguiente tabla muestra los resultados comparativos de los dos enfoques, siempre en porcentajes de clasificaciones correctas:

TABLA 2.10. RESULTADOS AMBOS MÉTODOS

Clasificaciones correctas	Año 1	Año 2	Año 3	Año 4	Año 5
Rough Set	100%	78.57%	66.67%	64,71%	70%
Análisis Discriminante	66.17%	71.75%	74.51%	58.09%	65.62%

Tal y como puede verse en la tabla anterior los resultados del enfoque rough set, excepto para el año 3, son mejores que los del análisis discriminante. Estos resultados muestran la capacidad de esta metodología para responder de manera eficiente al problema de predicción del fracaso empresarial, siendo una alternativa muy fiable al análisis discriminante.

Es más, desde el punto de vista del usuario, las reglas de decisión expresan las dependencias entre las características financieras de una empresa y su situación en un lenguaje más accesible que una función discriminante. Una función discriminante solo da una visión global, mientras que las reglas de decisión hablan el lenguaje natural de las decisiones tomadas por un experto o decisor y que han servido como ejemplos.

Capítulo 3

ALGORITMIA HEURÍSTICA HÍBRIDA APLICACIÓN A LA PREDICCIÓN DE INSOLVENCIAS

III.1-. INTRODUCCIÓN A LA ALGORITMIA HEURÍSTICA

Una parte muy importante de las herramientas de inteligencia artificial está formada por los llamados *algoritmos heurísticos*, para la resolución de problemas en los que aproximaciones de tipo tradicional no son adecuadas. La gran mayoría de estos algoritmos tienen una componente de aleatoriedad importante, pero, al mismo tiempo, poseen una estructura de búsqueda que les hace ser muy potentes, y que ha contribuido a su popularidad en los últimos años.

Sin duda, el paradigma de algoritmo heurístico está formado por los algoritmos basados en la teoría de la evolución, más conocidos como *algoritmos evolutivos*, y entre estos, los más populares y estudiados, los *algoritmos genéticos* (Goldberg, 1989). Los algoritmos genéticos son heurísticos utilizados para optimización y búsqueda, basados en la aplicación de las leyes de la evolución en la naturaleza (básicamente la supervivencia del más fuerte). A su popularidad han contribuido, en parte, la elegancia de las ideas que hay detrás, y sin duda también su sencillez, flexibilidad y aplicabilidad a una enorme variedad de problemas distintos en campos muy diferentes entre sí. Otro tipo de algoritmos como el *enfriamiento simulado* (Kirpatrick, 1983), (Kirpatrick, 1984), basado en el proceso físico de enfriamiento de la materia, la *búsqueda tabú* (Glover, 1989), (Glover, 1990), basado en la restricción de caminos de búsqueda mediante el marcado de soluciones como prohibidas, la programación genética (Koza, 1992), en la que se evolucionan árboles de decisión, o la conocida como *swarm particle optimization* (Kennedy, 1995) en la que se construye un conjunto de soluciones que se mueven como un enjambre (swarm), son igualmente aplicables a muchos problemas de búsqueda y optimización.

La aplicación de algoritmos heurísticos en economía ha sido tradicionalmente menor que en otras disciplinas, posiblemente porque en economía la mayoría de los problemas en los que se han aplicado técnicas de inteligencia artificial son problemas de decisión, más apropiados para redes neuronales y distintos tipos de clasificadores, y no tanto para algoritmos de optimización (O'leary, 1998). Esta tendencia se matiza, sin embargo, en los últimos años, con la aparición de múltiples problemas de optimización relacionados con el estudio del mercado de valores, y la previsión de insolvencias empresariales (Sanchis, 2003), (Salcedo, 2004), donde a la batería de herramientas anteriores se le añade la de los algoritmos evolutivos.

Este capítulo estudio tiene por objetivo presentar la aplicación de técnicas heurísticas en problemas de optimización que surgen en la previsión de insolvencias en compañías de seguros, concretamente en seguros no-vida. Específicamente se presentará la aplicación de algoritmos genéticos y enfriamiento simulado como algoritmos básicos que permiten mejorar la decisión (compañía solvente o insolvente) de un clasificador dado. A lo largo del estudio se presentarán de forma sencilla y estructurada las ideas fundamentales de los algoritmos heurísticos utilizados, de manera que el lector novel encuentre en este estudio una guía básica para entender su funcionamiento y aplicabilidad. El lector

interesado podrá encontrar en este estudio la descripción matemática del problema de predicción de insolvencias en compañías de seguros, como un caso concreto del problema general de clasificación. Asimismo, se darán detalles del clasificador utilizado en los experimentos realizados, concretamente una *máquina de vectores soporte*, que actualmente se considera como el clasificador con mejores prestaciones en términos de generalización de la solución.

La estructura del resto de este capítulo estudio es la siguiente: en primer lugar se recordará presentará brevemente el problema de la predicción de insolvencias en empresas de seguros no vida, a partir de un problema general de clasificación de muestras. A partir de este problema inicial se describe el problema asociado de la *selección de características*, en el cual se aplican los algoritmos heurísticos probados en este estudio, salvo el caso particular de la programación genética, que será abordado al final del estudio. Los algoritmos heurísticos de optimización serán descritos a continuación, dando paso, por último a un ejemplo de su aplicación en un problema real de predicción de insolvencia en empresas de seguros no-vida españolas.

III. 2.- EL PROBLEMA DE CLASIFICACIÓN MULTIATRIBUTO

Como ya se ha visto, este estudio se centra en el la predicción de la insolvencia de empresas de seguros no vida, que, como ha sido mencionado anteriormente es un caso particular del problema de clasificación multiatributo. Este problema consiste en la asignación de un *objeto* o *muestra*, descrito por una serie de atributos (también llamados características), a una clase de categorías (o clases) predefinidas. Generalmente un objeto es definido mediante un número elevado de características, mientras que las clases predefinidas suelen ser dos en la mayoría de los casos. En caso de que haya más de dos clases, el problema es conocido como problema de clasificación multiclase.

Matemáticamente, un problema de clasificación puede ser definido como sigue:

Sea $\{x_i\}$, $x_i \in \mathcal{X}$, $i \in \{1, \dots, l\}$ un conjunto de observaciones (muestras), extraídas de una distribución de probabilidad en general desconocida $P(x, y)$, e $\{y_i\} \in \{-1, 1\}$, (clases), un conjunto de etiquetas asociadas que describen la correcta pertenencia de las muestras a una clase en particular.

Una máquina de clasificación (clasificador) está definida entonces como un conjunto de funciones $x \rightarrow f(x, \alpha)$, donde una elección particular de parámetros α genera lo que se llama una “máquina entrenada”. Por ejemplo, en el caso general de una red neuronal con una arquitectura fija, α correspondería a los pesos de la red, en el caso de una máquina de vectores soporte (SVM), α serían los parámetros ajustables de la SVM. El entrenamiento de la máquina se lleva a cabo mediante un conjunto de muestras conocido como conjunto de entrenamiento.

Un error empírico en test puede ser definido como:

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x_i, \alpha)| \quad (1)$$

con $f(x, \alpha)$ el resultado del clasificador. Este error debe ser calculado en otro conjunto de muestras (diferente al conjunto de entrenamiento), usualmente conocido como conjunto de test.

Hay que notar que $R_{emp}(\alpha)$ es un valor fijo para una elección particular de α utilizando un conjunto de entrenamiento concreto $\{x_i, y_i\}$. En el caso general, el problema de clasificación multiatributo consiste en encontrar la máquina que aprenda el mapeo $x_i \rightarrow y_i$ con la mayor habilidad de generalización posible, esto es, minimizando $R_{emp}(\alpha)$ para cualquier conjunto de prueba.

III.2.1.- Predicción de insolvencia en empresas de seguros no vida

El problema de clasificación multiatributo es aplicable de forma inmediata al problema de la predicción de la insolvencia en empresas de seguros, sin más que identificar los objetos $\{x_i\}$, $i \in \{1, \dots, l\}$, como un conjunto de empresas descrito por n ratios financieros (cada característica x_{ij}), e $\{y_i\} \in \{-1, 1\}$ una etiqueta que describe el estado de la compañía como solvente ($y_i = 1$) o insolvente ($y_i = -1$).

Así pues, dado un conjunto de empresas de seguros no vida descrito por una serie de ratios, la predicción de la insolvencia se lleva a cabo entrenando un clasificador (en general puede ser cualquier tipo de clasificador, una red neuronal, una máquina de vectores soporte etc.) en un conjunto de entrenamiento (encontrando los parámetros α que describen el clasificador). Cuando aparezca una nueva muestra (nueva empresa de seguros para ser evaluada), la aplicación del clasificador nos dirá si la empresa es solvente o insolvente con un determinado error, dado por el error empírico de clasificación en test $R_{emp}(\alpha)$.

III.2.2.- Selección de características en problemas de clasificación

En un problema de clasificación es muy común que no todas las características que definen las muestras sean completamente necesarias. En general, por el contrario, es frecuente encontrarse con determinadas características completamente superfluas para un problema de clasificación dado. Este tipo de características son en general perjudiciales, porque añaden ruido al sistema, sin aportar absolutamente ninguna información que ayude al clasificador a clasificar correctamente las muestras. La inclusión de este tipo de características digamos “dañinas” es en la práctica inevitable, puesto que cuando se generan los datos no se conoce cuales son las características importantes para un determinado problema, de manera que en principio las muestras son descritas con el mayor número de características posibles. Para dar un ejemplo sencillo, imagínese el lector un problema de clasificación en el que se intente discernir entre personas con riesgo de padecer una enfermedad coronaria y personas que no es probable que vayan a padecerla. Entre las características que definirán a las personas que pidan créditos al banco aparecerán características importantes para el problema en cuestión (edad, fumador o no, antecedentes familiares, etc.) y se podrían incorporar otras completamente indiferentes para la resolución del problema (altura de la persona, color de los ojos, etc.). El cometido de la selección de características es eliminar estas características superfluas que sólo incorporan ruido al sistema y pueden llegar a afectar el rendimiento de un clasificador en un problema dado.

La selección de características se puede definir y resolver, en general, como un problema de optimización combinatoria. Normalmente este tipo de problemas son difíciles de resolver con métodos tradicionales, y para ellos, los algoritmos heurísticos son una alternativa muy interesante. La siguiente definición describe matemáticamente un problema de selección de características (FSP) como un problema de optimización combinatoria:

Dado un conjunto de muestras etiquetadas $(x_1, y_1), \dots, (x_l, y_l)$, donde $x_i \in \mathcal{R}^n$ e $y_i \in \{-1, 1\}$, el FSP consiste en elegir un subconjunto de m características ($m < n$), tal que, utilizando dichas características, un clasificador consiga el menor error de clasificación en test.

Siguiendo la notación en (Weston, 2000), el FSP puede definirse como encontrar el vector binario de n componentes (σ) , óptimo σ° , que define el subconjunto de características que cumple:

$$\sigma^\circ = \arg \min_{\sigma, \alpha} R_{emp}(\alpha, x^* \sigma) \quad (2)$$

donde se ha definido $x^* \sigma = (x_1 \sigma_1, \dots, x_n \sigma_n)$.

En otras palabras: hay que encontrar el vector binario σ° entre todos los vectores binarios σ con n componentes, tal que una vez hecha la operación $x * \sigma = (x_1 \sigma_1, \dots, x_n \sigma_n)$, se obtenga el menor error de clasificación en test. Hay que notar que con la operación $*$, los vectores σ actúan como plantillas, eliminando las características que coinciden con los 0 del vector σ , y dejando las características que coinciden con los 1 del vector σ . En el caso particular de este estudio, el vector $x_i \in \mathcal{R}^n$ estará formado por ratios financieros que describan una compañía de seguros no-vida, y el vector σ° nos dará el mejor conjunto de características posible para un clasificador dado.

El cálculo del vector óptimo σ° es un problema de optimización combinatoria. Una posible primera idea para hallar este vector sería hacer una búsqueda exhaustiva de todos los posibles vectores σ , reteniendo como solución el que nos dé menor error de clasificación en test. Nótese que, debido a que el vector σ° tiene n componentes, el número de posibles vectores σ es 2^n , con lo que la opción de la búsqueda exhaustiva se reduce a valores de n muy pequeños, ya que cuando n crece, el número de posibles vectores σ crece exponencialmente con n . Así pues, es necesaria la introducción de algoritmos de búsqueda que nos proporcionen una solución aceptable en un tiempo razonable, y eso es precisamente lo que los heurísticos de búsqueda heurísticos como algoritmos genéticos y el enfriamiento simulado proporcionan.

III.3.- ALGORITMOS HEURÍSTICOS DE BÚSQUEDA

III.3.1.- Algoritmos genéticos

Los algoritmos genéticos (Goldberg, 1989) son una clase de técnicas de búsqueda y resolución de problemas basados en la generación de una *población* de posibles soluciones (normalmente codificadas como cadenas de números binarios), llamados *individuos*. La población es en principio generada al azar, y evolucionada mediante la aplicación de *operadores evolutivos* (básicamente selección, cruce y mutación), que producen la aparición de nuevos individuos (nuevas soluciones al problema) a la vez que sólo las mejores soluciones (con respecto a alguna función objetivo o *fitness*) sobreviven, eliminándose con mucha mayor probabilidad aquellos individuos con peor *fitness*.

Codificación de una solución e inicio del algoritmo

Gran parte de la utilidad de los algoritmos genéticos radica en que son capaces de soportar una gran cantidad de posibles codificaciones para un determinado problema. Por supuesto, no todas las codificaciones son igual de efectivas, luego el rendimiento del algoritmo puede verse afectado por la elección de la codificación de las soluciones. En principio, en un algoritmo genético siempre es necesario codificar las soluciones como cadenas de números, debido al funcionamiento del operador de cruce, que será analizado posteriormente. La codificación más usual usada en un algoritmo genético es la binaria, ya que los operadores “estándar” de cruce y mutación fueron definidos en su día para este tipo de codificación. Si se requiere otro tipo de codificación es muy probable que haya que modificar los operadores, convirtiéndolos en operadores “ad-hoc” para la codificación elegida y el problema a resolver. De ahora en adelante este estudio sólo considera la codificación binaria de soluciones en el algoritmo, y los operadores de cruce y mutación serán descritos teniendo en cuenta esta codificación.

En cuanto a la inicialización del algoritmo, lo usual es generar la primera población de forma aleatoria, y dejar evolucionar el algoritmo desde ese punto. En principio, y salvo en problemas con muy fuertes restricciones, los algoritmos genéticos funcionan bien desde cualquier punto de inicialización. A pesar de esto, puede ser conveniente en ocasiones “sembrar” el algoritmo con soluciones buenas (si se conocen), ya que esto puede mejorar el tiempo de convergencia del algoritmo. En este estudio consideraremos la inicialización aleatoria del algoritmo genético en la primera generación. El tamaño de la población a utilizar por el algoritmo genético es también un parámetro variable. En general se utilizan poblaciones de entre 50 y 100 individuos.

El operador de selección

El operador de selección tiene como objetivo preservar para la siguiente generación los mejores individuos, con respecto a la función de fitness del algoritmo genético. Normalmente siempre se incluye algún tipo de aleatoriedad en la selección, pero siempre de manera que los individuos con mayor fitness tengan más oportunidades de sobrevivir que los individuos peores. El operador de selección más comúnmente usado es el llamado *método de la ruleta*, en el que cada individuo tiene una probabilidad de pasar a la siguiente generación proporcional a su fitness:

$$\frac{f_i}{\sum_{v_i} f_i} \tag{3}$$

siendo f_i el fitness asociado al individuo i .

El conjunto de nuevos individuos seleccionados de la población anterior conformarán la nueva población, que será del mismo tamaño que la población anterior. Es muy importante darse cuenta de que el operador de selección no produce nuevos individuos, y por tanto no contribuye a la búsqueda dentro del algoritmo genético. Este operador sólo selecciona los mejores individuos entre los existentes en una población dada.

El operador de cruce

El operador de cruce ha sido descrito como la clave de los algoritmos genéticos. Este operador produce nuevos individuos a partir de los existentes en la población, mediante el cambio de “material genético” entre los individuos. Este cambio de material es realizado de la siguiente forma. Primero se emparejan los individuos que vayan a cruzarse. A continuación se elige al azar el punto de cruce, y se intercambian las subcadenas binarias desde ese punto hasta el final del individuo.

Ejemplo:

Considérense dos individuos que han sido elegidos para cruzarse en un algoritmo genético con codificación binaria $n=10$:

```
1 0 0 1 1 1 0 1 0 0
1 1 0 0 1 0 1 0 1 1
```

si, por ejemplo, escogemos el bit número 5 como punto de cruce obtendremos:

```
1 0 0 1 1 0 1 0 1 1
1 1 0 0 1 1 0 1 0 0
```

Nótese que se han generado dos nuevos individuos a partir de dos existentes previamente en la población. Un punto importante en el operador de cruce es que no siempre se realiza, sino que los individuos emparejados tienen una probabilidad de cruce P_c , siempre estrictamente menor que 1, y normalmente de 0.6 y 0.7. Esto es debido a que si todos los individuos se cruzaran siempre ($P_c=1$), el algoritmo genético tendría problemas de convergencia.

El operador de mutación

El operador de mutación actúa sobre los componentes de una población de manera individual, evitando que la población se quede estancada en soluciones subóptimas, debido a que todos los individuos sean el mismo (esto es frecuente que ocurra al final del algoritmo, cuando el operador de selección ha sido aplicado muchas veces). La mutación es un proceso completamente aleatorio de generación de nuevos individuos. Dado un individuo, se escoge un bit (o varios bits) y se cambian de 0 a 1 o viceversa.

Ejemplo:

Dado el individuo:

1 1 0 0 1 0 0 1 0 1

si se elige por ejemplo el bit número 3 como bit a mutar tendríamos:

1 1 1 0 1 0 0 1 0 1

El operador de mutación se aplica con una probabilidad muy pequeña a cada individuo, P_m , aproximadamente un valor estándar de esta probabilidad es $P_m=0.01$. Probabilidades de mutación grandes introducen demasiada aleatoriedad al algoritmo genético, y conllevan la convergencia de ésta a soluciones subóptimas.

Elitismo y finalización del algoritmo

En general, la solución aportada a un determinado problema por un algoritmo genético se considera que es el mejor individuo de la última generación. Debido a errores de tipo estocástico en el operador de selección, es posible perder soluciones que sean mejores que la solución final del algoritmo. Para evitar esto se considera el llamado *elitismo*, que consiste en pasar siempre el mejor individuo de una generación a la siguiente. Esto evita la pérdida de este individuo en el proceso de selección.

La finalización del algoritmo se realiza usualmente después de un número dado de generaciones. Este número es variable dependiendo del problema y de lo costoso que sea el cálculo de la función de *fitness*.

Aplicación de un algoritmo genético al FSP

Para aplicar un algoritmo genético al FSP, cada individuo de la población codifica una posible solución (vector σ) mediante una cadena binaria. A la población de vectores σ , inicializada de forma aleatoria, se le aplican los operadores de selección, cruce y mutación descritos anteriormente, con un proceso de elitismo para asegurar que el mejor individuo de la población no se elimine en la aplicación de los operadores.

La función de fitness utilizada para el operador de selección consiste en el valor de la probabilidad de error en test proporcionada por algún tipo de clasificador (en este estudio hemos utilizado máquinas de vectores soporte, pero cualquier tipo de clasificador podría ser utilizado) sobre los datos $(x^*\sigma, y)$. Cómo ésta probabilidad de error puede depender de los conjuntos de entrenamiento y test escogidos, normalmente se lleva a cabo un proceso de *validación cruzada*, en el que los datos disponibles se dividen en varios conjuntos, utilizando uno de ellos como test y el resto como entrenamiento, y promediando el resultado obtenido sobre todos los conjuntos de test posibles. La Figura 1 muestra cómo sería un proceso de validación cruzada con 4 conjuntos. Cada vez se escoge un conjunto de test diferente, utilizando el resto como conjunto de entrenamiento. El resultado que se utiliza como fitness para el algoritmo genético es el promedio de las distintas probabilidades de error en cada uno de los conjuntos de test.

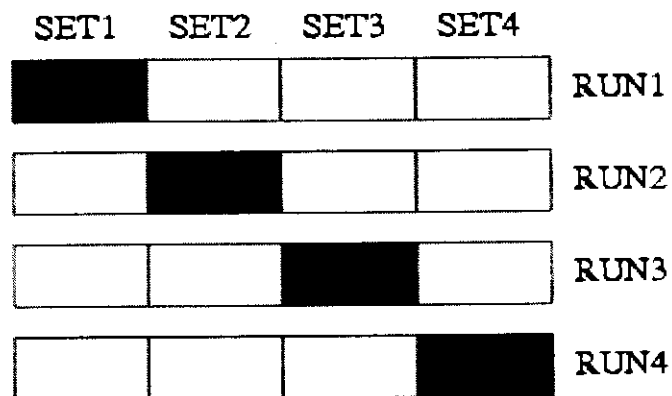


Figura 1. Ejemplo del proceso de validación cruzada. Se tienen 4 conjuntos distintos (SET1-SET4), y se hacen 4 entrenamientos distintos del clasificador utilizando como conjunto de entrenamiento los conjuntos en blanco de la figura, resultando 4 probabilidades de test (en principio diferentes) calculadas en el conjunto restante, en negro en la figura. El valor del fitness que será utilizado por el algoritmo genético se obtiene promediando las distintas probabilidades de error en el total de conjuntos de test.

Un aspecto importante a destacar es que si utilizamos el método de la ruleta para realizar la selección el algoritmo genético tratará de encontrar siempre

el máximo de la función de fitness. Como nosotros estamos interesados en soluciones que minimicen la probabilidad de error en test hay que introducir una función auxiliar que sustituya a la probabilidad de error. Esta función puede ser definida como:

$$F=100(1-R_{emp}(\alpha, \sigma)) \quad (4)$$

Nótese que el máximo de esta función se produce en el mínimo de la probabilidad de error en test, que es nuestro objetivo. El mejor individuo de la última generación será considerado como el vector σ^0 que define el vector óptimo de características.

III.3.2.- El algoritmo de enfriamiento simulado

El enfriamiento simulado (Kirpatrick, 1983), (Kirpatrick, 1984) es un heurístico de búsqueda y optimización basado en el proceso natural conocido como *temple*. El proceso de temple consiste en calentar una sustancia y posteriormente enfriarla poco a poco, hasta que una estructura cristalina fuerte sea obtenida. Este proceso puede ser simulado en un ordenador y aplicado a la obtención de soluciones óptimas de un problema de optimización combinatoria. El proceso sería el siguiente: Se parte de una única solución al problema, generada aleatoriamente. Se define entonces una *temperatura del sistema* que irá siendo disminuida poco a poco a lo largo del algoritmo. Mediante un operador de mutación se van obteniendo distintas soluciones a partir de la solución actual. Si la nueva solución encontrada es mejor que la actual, ésta última es tomada como nueva solución actual. Si la nueva solución encontrada es peor, ésta puede ser tomada como solución actual probabilísticamente, y la probabilidad de aceptación de esta solución peor como la actual depende de la temperatura del sistema en ese momento. La temperatura del sistema, que controla la probabilidad de aceptación de soluciones peores que la actual, debe ser tal que al principio la probabilidad de aceptación sea grande, mientras que al final del algoritmo sólo las soluciones que sean mejores que la actual puedan ser aceptadas (probabilidad de aceptación de soluciones peores nula). El procedimiento de enfriamiento simulado es mejor que otros heurísticos escapando de soluciones subóptimas, sin embargo, la solución que aporta no tiene por qué ser el óptimo global, sino que será una solución “suficientemente buena” que podemos considerar la solución a nuestro problema.

La estructura general de un algoritmo de enfriamiento simulado está formada por la selección de la codificación de las soluciones, la función objetivo a ser optimizada y el operador de mutación aplicado. A continuación detallamos estas componentes del algoritmo de enfriamiento simulado utilizado en el FSP.

Codificación del FSP

La codificación utilizada en el enfriamiento simulado para resolver el FSP es análoga a la del algoritmo genético: básicamente una cadena binaria de longitud n representando el vector σ^o . Es muy importante destacar que, mientras el algoritmo genético está basado en una población de posibles soluciones, el enfriamiento simulado sólo codifica una solución, que irá cambiando a lo largo del algoritmo. El simulated annealing no es, por lo tanto, un algoritmo basado en una población de soluciones.

El operador de mutación

El algoritmo de enfriamiento simulado está basado en un único operador, el operador de mutación. En el caso del FSP se considera la mutación básica de un conjunto de K bits en la cadena binaria que representa la solución actual. Este número K puede ser mayor o menor dependiendo de la longitud de la cadena binaria n . Este operador de mutación no se considera con una cierta probabilidad

P_m como en el caso del algoritmo genético, sino que se aplica siempre a la solución actual del algoritmo.

La función objetivo

La función objetivo del enfriamiento simulado para el FSP es idéntica a la utilizada por el algoritmo genético. Al igual que en el caso anterior, se considera la probabilidad de error en test de un clasificador dado como la función a minimizar, y se utiliza un procedimiento de validación cruzada para tratar de hacer independiente el valor de la función de fitness de los conjuntos de validación y test escogidos. Más detalles a cerca de este procedimiento se dan en la sección anterior.

Implementación del enfriamiento simulado

El procedimiento de enfriamiento simulado se inicia con una solución aleatoria que se considera la primera solución actual. La temperatura inicial (T_0) del sistema se fija de tal manera que la probabilidad de aceptación de nuevas soluciones tenga un valor de 0.8 (valor estándar utilizado típicamente en una gran cantidad de aplicaciones del enfriamiento simulado). La probabilidad de aceptación de soluciones viene dada por la función

$$P(\sigma^{mut}, T) = e^{\left(\frac{-a}{T}\right)} \quad (5)$$

Con σ^{mut} la nueva solución conseguida mutando la solución actual, y T la temperatura actual del sistema, a es un parámetro del algoritmo fijado para que la probabilidad inicial sea de 0.8.

Para una temperatura dada, se muta la solución actual un número M de veces, aceptando la solución si es mejor o probabilísticamente con una probabilidad dada por la ecuación (5) si es peor. El descenso de la temperatura se hace gradualmente, siguiendo una función del tipo:

$$f_T = \frac{T_0}{1+k} \quad (6)$$

con k el número de veces que se ha bajado la temperatura hasta el momento.

La finalización del algoritmo está basada en un número máximo de bajadas de temperatura. Es decir, se plantea un número máximo de posibles bajadas de temperatura mediante un parámetro del sistema, $numIt$, y a partir de este número se calcula la temperatura mínima a la que el sistema va a llegar mediante la siguiente ecuación:

$$T_{min} = f_T(T_0, numIt) \quad (7)$$

La solución final del enfriamiento simulado será considerada como el vector σ^0 que define el vector óptimo de características.

Pseudo-código del Enfriamiento simulado.

```
k = 0;
T = T0
Initialize the current configuration  $\sigma$  at random;
Run the SVM  $\rightarrow \sigma = P_e(test)$ 
repeat
  for j=0 to M
     $\sigma^{mut} = mutate(\sigma)$ 
    Run the SVM  $\rightarrow f(\sigma^{mut}) = P_e(test)$ 
    if ((  $f(\sigma^{mut}) < f(\sigma)$  ) OR (  $random(0,1) < e^{\frac{-a}{T}}$  )) then
       $\sigma = \sigma^{mut}$ 
    endif
  endfor
k = k + 1
T =  $f_T(T_0, k)$ 
until (  $T < T_{min}$  )
```

III.4.- APLICACIÓN DE LOS ALGORITMOS DESCRITOS A UN PROBLEMA REAL DE SELECCIÓN DE CARACTERÍSTICAS

En esta sección se aplican los dos algoritmos descritos en las secciones previas a una aplicación real de selección de características en empresas de seguros no vida. El conjunto de empresas utilizado, así como los ratios empleados ha sido descrito en el capítulo primero.

Como ya ha sido mencionado anteriormente, tanto el algoritmo genético como el enfriamiento simulado para selección de características pueden ser utilizados con cualquier tipo de clasificador. En este estudio se ha utilizado una Máquina de Vectores Soporte (SVM), debido a que, actualmente, este es el clasificador que mejores propiedades de generalización posee. Hemos incluido una pequeña introducción a las SVM en los apéndices de este estudio. El lector más interesado puede además acudir a las referencias (Burges, 1998), (Scholkopf, 2002).

El objetivo de los experimentos desarrollados ha sido intentar conseguir un conjunto de características que describan las 72 empresas consideradas, que sea menor que el inicial (19 ratios), y que proporcione un error medio de clasificación en test menor o igual que el proporcionado por el conjunto inicial de características. En primer lugar es necesario entonces evaluar la probabilidad de error en test obtenida con la SVM utilizando los 19 ratios del conjunto inicial de características. La probabilidad de error obtenida con estos 19 ratios utilizando el proceso de validación cruzada fue $P_e = 0.33$. Esto significa que la SVM, utilizando los 19 ratios acertaría en media dos tercios de las veces que evaluara la solvencia o insolvencia de una determinada empresa.

Una vez tomada esta referencia del conjunto inicial de ratios, se realiza la selección de características mediante el algoritmo genético y el enfriamiento simulado. Los parámetros del algoritmo genético fueron fijados en los estándar, $P_c = 0.6$ y $P_m = 0.01$ (ver (Goldberg, 1989), con una población de 30 individuos. Los parámetros del enfriamiento simulado fueron fijados como $M = 30$ y $numIt = 100$.

Los resultados obtenidos en la ejecución de ambos algoritmos fueron los siguientes. La solución proporcionada por el algoritmo genético mejoró sustancialmente la solución con 19 ratios. La solución del algoritmo genético fue de una probabilidad de error media en clasificación de 0.23, frente al 0.33 obtenido con el conjunto inicial. Esta mejor solución dada por el algoritmo genético contenía tan sólo 3 ratios, R1, R9 y R13, lo que significa que el algoritmo genético reduce sustancialmente el número de características iniciales. Así las empresas consideradas pueden ser descritas sin ninguna pérdida de información por estos 3 ratios, en lugar de los 19 iniciales.

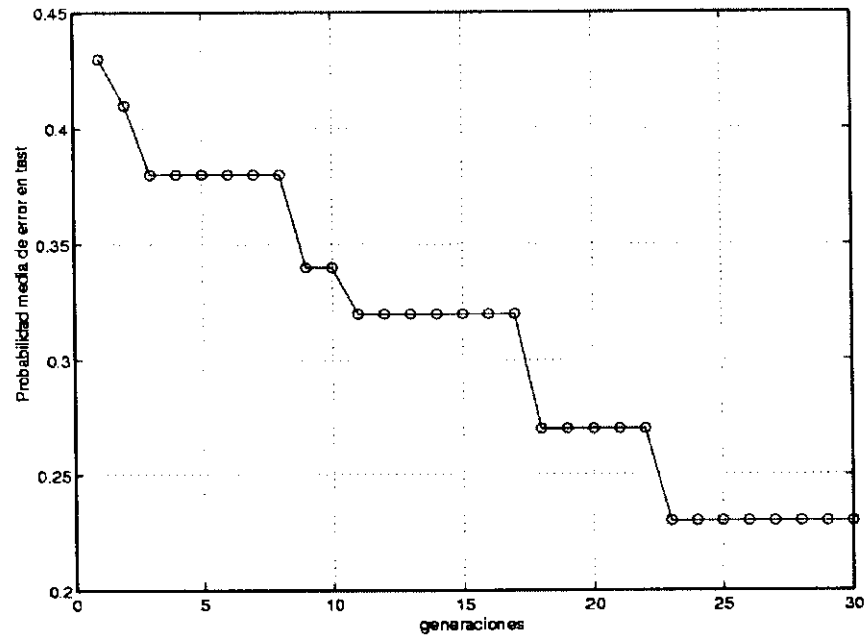


Figura 2. Evolución del algoritmo genético en la ejecución en la que fue obtenida la mejor solución.

La Figura 2 muestra la evolución del algoritmo cuando la mejor solución fue obtenida. El número de generaciones que fue necesario para llegar a la convergencia fue pequeño, alrededor de 20, lo que muestra que el coste computacional del algoritmo genético fue bajo en este problema.

El algoritmo de enfriamiento simulado proporcionó 2 conjuntos diferentes de 3 características cada uno con los que se obtenía una probabilidad media de error en test de 0.23. Estos conjuntos fueron {R1, R9, R13}, y {R3, R9, R19}. El enfriamiento simulado llegó a uno de estos conjuntos en todas las simulaciones lanzadas. Ningún otro conjunto de 3 características dio mejor resultado en términos de probabilidad de error. Nótese que, al igual que el algoritmo genético, el enfriamiento simulado es capaz de encontrar 3 características que describen las empresas mejor que las 19 iniciales.

III.4.1.- Análisis de los resultados

La primera conclusión a la que se puede llegar viendo los resultados obtenidos es una selección de características adecuada puede reducir la descripción de las empresas de 19 ratios a solamente 3, sin perder en absoluto capacidad de clasificación lo cual es un avance muy interesante. Estos datos nos dicen también que, en este caso, la mayoría de los ratios utilizados inicialmente en la descripción de las compañías de seguros analizadas eran superfluos, y solamente añadían ruido al sistema, sin ninguna información.

Un análisis pormenorizado de los ratios obtenidos por los algoritmos de selección de características utilizados puede ayudar a comprender por qué estos ratios son tan importantes en la descripción de compañía de seguros no-vida. Aunque en el capítulo primero se pueden ver las definiciones de los ratios y lo que implica el que algunos ratios hayan sido seleccionados como variables fuertemente discriminatorias entre empresas solventes e insolventes, brevemente recordaremos algunas de sus implicaciones:

R1- Una de las cuestiones más importantes a la hora de evaluar el buen funcionamiento de una compañía de seguros es la necesidad de tener suficiente liquidez. Sin embargo, en el caso de una compañía de seguros no-vida, la falta de liquidez no debería aparecer, ya que las primas se pagan antes de que ocurran los siniestros. Por tanto en este sector la falta de liquidez es un problema más acuciante que en otros sectores dada la inversión del proceso productivo. Por otro lado, este ratio es una medida de equilibrio financiero: si es positivo implica que el capital circulante también es positivo.

R9- Este ratio indica la proporción de los capitales propios sobre el pasivo total (capital y reservas). Confirma, desde el punto de vista de la solvencia, la importancia de que estos fondos sean los adecuados, ya que si se producen algunos imprevistos, estos fondos, podrían necesitarse en el futuro para hacer frente a las obligaciones contraídas.

R13- Este ratio es uno de los ratios considerados como *ratios de solvencia en sentido estricto*. El numerador muestra el riesgo debido a la siniestralidad y el denominador representa el soporte financiero real de las empresas ya que se considera junto con los capitales propios las provisiones técnicas. La selección de este ratio demuestra la importancia de poseer suficientes fondos propios así como la necesidad de cumplir adecuadamente con las exigencias, en cuanto a provisiones técnicas, para garantizar la viabilidad de la entidad aseguradora.

El enfriamiento simulado, por su parte, encontró otro conjunto de ratios que generan la misma probabilidad de error en test que los anteriores. Este conjunto incluye el ratio R9, descrito anteriormente, y los siguientes ratios:

R3- Este ratio indica que la obtención de suficientes ingresos financieros es un aspecto crítico, debido a que actualmente estos ingresos son la principal fuente de beneficios de una empresa de seguros.

R19- Este ratio confirma la importancia de comprobar la idoneidad de la política de reaseguro seguida por una firma de seguros a la hora de evaluar su solvencia.

En resumen, los algoritmos propuestos en este capítulo proporcionan una nueva e interesante forma de abordar la predicción de insolvencias en empresas de seguros no-vida. Tanto el algoritmo genético como el algoritmo de enfriamiento simulado utilizados para abordar un problema real de selección de características en empresas de seguros no-vida obtienen una importante reducción en el número de ratios utilizados en la clasificación, mientras que la SVM clasifica las empresas en solventes e insolventes basándose en estos ratios.

APÉNDICE: Breve introducción a las máquinas de vectores soporte (SVM)

Esta sección proporciona una breve descripción de la máquina de vectores soporte, SVM, para clasificación. Un análisis más detallado puede ser consultado en (Burges, 1998). En primer lugar describiremos la SVM para problemas linealmente separables, introduciendo después la SVM no lineal.

Considérese un conjunto de firmas de seguros, representadas por sus ratios $\{x_i\}$, $i \in \{1, \dots, l\}$, y un conjunto de etiquetas $\{y_i\} \in \{-1, 1\}$, que describen cada firma como solvente o insolvente. En primer lugar vamos a estudiar el caso en que este conjunto de muestras sea linealmente separable, esto es, el conjunto de entrenamiento puede ser separado por un hiperplano. Si éste es el caso, la SVM resuelve el siguiente problema:

encontrar $w \in \mathfrak{R}^n$, $b \in \mathfrak{R}$ y x_i , $i = 1, \dots, l$, que minimice

$$\frac{1}{2} \|w\|^2, \quad (8)$$

sujeto a:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, \dots, l \quad (9)$$

Una vez que tales w y b han sido encontradas, la regla de clasificación para firmas solventes o insolventes vendría dada por $\text{sign}(w^T x + b)$. Esto significa que las empresas localizadas en un lado del hiperplano serían solventes, y las empresas del otro lado serían insolventes. El error de clasificación asociado definido como el porcentaje de firmas mal clasificadas estaría dado por $R_{emp}(w, b)$ (ver Sección 2).

Considérese ahora el caso en el que los puntos del conjunto de entrenamiento no sean linealmente separables. Si esto es así, la restricción (9) no puede ser satisfecha. Se introducen entonces una serie de variables de corrección ξ_i para solventar esto. La formulación de la SVM incluyendo estas variables de corrección es la siguiente:

encontrar $w \in \mathfrak{R}^n$, $b \in \mathfrak{R}$ y x_i , $i = 1, \dots, l$, que minimice

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i, \quad (10)$$

bajo las restricciones:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, l \quad (11)$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, l \quad (12)$$

donde C es un parámetro del clasificador a estimar.

La Figura 3 muestra un ejemplo de un problema de clasificación formado por 2 clases (cruces y puntos). La solución mostrada está obtenida utilizando una SVM lineal. Nótese que en este caso la frontera de decisión no es capaz de separar completamente ambas clases. Las líneas punteadas representan los márgenes, es decir, el conjunto de puntos que satisface la ecuación (11). Las muestras sobre los márgenes, que están rodeadas de un círculo son los Vectores Soporte. Ésta es la única información necesaria para dibujar la frontera de clasificación.

La SVM no lineal mapea las variables de entrada a un espacio de muchas dimensiones (a menudo infinitas), y aplica la SVM lineal en este espacio. Computacionalmente, esto se consigue mediante la aplicación de una función de *kernel*. La SVM no lineal con un kernel K es equivalente a al siguiente problema de regularización en un espacio de Hilber H_K :

encontrar $\phi(x) = h(x) + b$ con $h \in H_K$, $b \in \mathfrak{R}$ y ξ_i , $i = 1, \dots, l$, que minimice

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (13)$$

sujeto a

$$y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, l \quad (14)$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, l \quad (15)$$

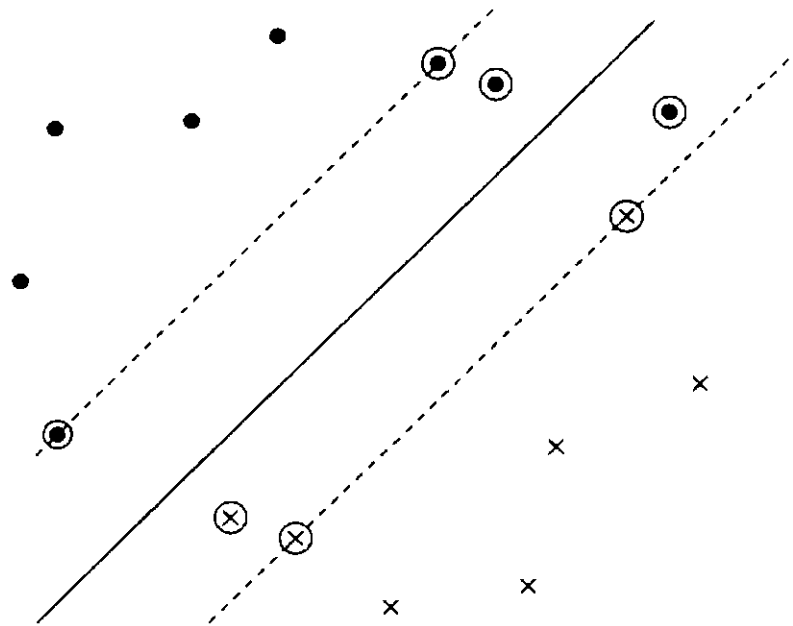


Figura 3. Ilustración de una SVM lineal para resolver un problema de 2 clases.

La SVM no lineal es capaz de clasificar cualquier conjunto de firmas en solventes o insolventes, con una probabilidad de error dada por $R_{emp}(w, \alpha)$. La Figura 4 ilustra una SVM no lineal para un problema de clasificación, utilizando un kernel Gaussiano. El conjunto a clasificar es el mismo que el de la Figura 3, pero, en esta ocasión la SVM no lineal si que es capaz de clasificar todas las muestras correctamente. La resolución del problema es obtenida mediante una frontera lineal en un espacio de Hilbert generado por el kernel. Ésta frontera lineal es la curva dibujada en la Figura 4 cuando se pasa del espacio de Hilbert al espacio de las muestras de entrada.

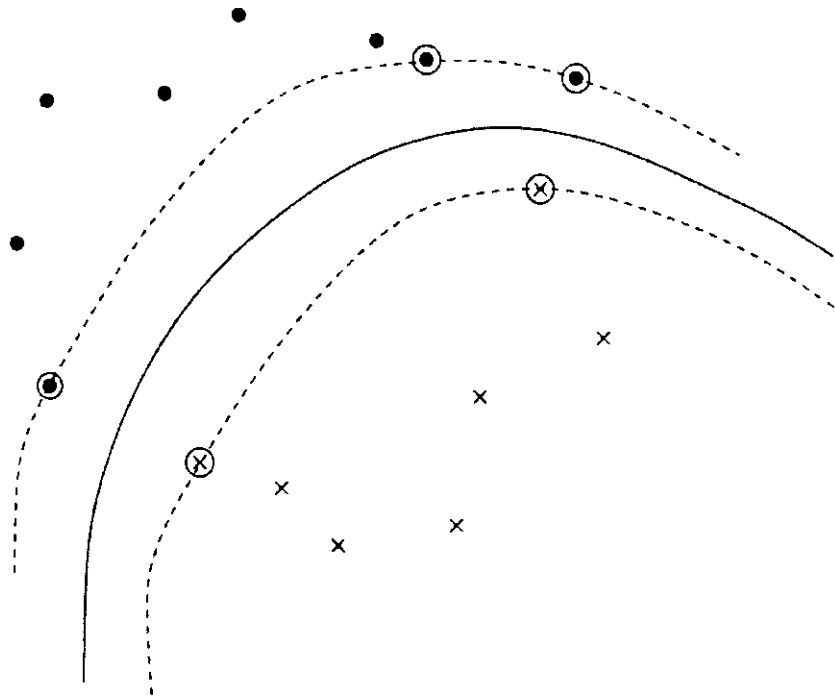


Figura 4. Ilustración de una SVM no lineal (kernel Gaussiano) para resolver un problema de 2 clases.

RESUMEN Y CONCLUSIONES

Para este trabajo, de entre todos los retos en el ámbito financiero que podríamos acometer, hemos elegido la **predicción de insolvencias en empresas aseguradoras**. Tal elección, como hemos mencionado en la introducción, queda justificada -en primer lugar- por su gran actualidad, derivada del proyecto *Solvencia II*, y -en segundo lugar- porque, a diferencia de otros problemas financieros, el número de agentes e intereses afectados cuando se produce una insolvencia en este sector justifica la importancia de su regulación, manteniendo este campo siempre actual. Además, hemos decidido abordar esta cuestión utilizando metodologías derivadas de la Inteligencia Artificial, en concreto la *Teoría Rough Set* y la *Algoritmia Heurística Híbrida*. La comparación de los resultados obtenidos en este y otros ámbitos de modelado empírico y la disponibilidad de recursos computacionales potentes a bajo coste, avalan esta última decisión.

Para la aplicación de estas metodologías, hemos incorporado como factores del análisis de insolvencia (directa o indirectamente) ratios financieros. Estos ratios calculados han sido las *variables explicativas* del riesgo de insolvencia y, aunque es una información demasiado agregada, presenta la ventaja de ser obligatoria para las empresas, homogénea y pública. Al hilo de esta cuestión queremos insistir en la imposibilidad práctica de haber introducido variables de tipo cualitativo u otras informaciones como el margen de solvencia (aspectos que hubiesen enriquecido en gran manera nuestros modelos), pero la dificultad de obtener las primeras y la falta de publicidad del margen de insolvencia y de otros datos técnicos en el periodo analizado han impedido su consideración. Aunque hubiese sido posible elegir otras variables, habitualmente empleadas en la literatura contable para la medida de la insolvencia, nuestro énfasis se centra en mostrar la viabilidad de las metodologías, por lo que es innecesaria -a nuestro juicio- discusión más pormenorizada sobre dichas alternativas. Sí conviene, no obstante, hacer notar que nuestros modelos permiten considerar variables tanto cuantitativas como cualitativas, ofreciendo amplia flexibilidad ante esas alternativas no cubiertas.

Hemos insistido en recordar que la elección de las variables más adecuadas a utilizar en la elaboración de un modelo de predicción es parte fundamental del éxito final del mismo. En nuestro caso, hemos configurado un conjunto de variables que, a nuestro juicio y desde puntos de vista teórico y práctico, forman un modelo integral de análisis financiero de una entidad aseguradora; considerando por nuestra parte que son las más relevantes en relación con la solvencia. Para ello hemos acudido a los ratios más utilizados en el análisis financiero para el pronóstico de la insolvencia de entidades aseguradoras (tanto generales como específicos) y tenido en cuenta nuestro criterio personal, las características propias del sector seguro y un análisis preliminar de las características de los dos grupos de empresas (sanas y fracasadas).

Con respecto a la Teoría Rough Set como método a aplicar para la predicción del fracaso empresarial en el sector seguros no-vida cabría recordar que, cuando se trabaja con compañías reales, el análisis de las relaciones causa-efecto relativas a la predicción de insolvencias adolece normalmente de

inconsistencias en las clasificaciones. Es decir, una empresa sana puede tener los mismos atributos que una quebrada. Un enfoque adecuado para tratar esta problemática es el que ofrece la Teoría de los Rough Set. Además, este método está especialmente concebido para problemas de clasificación multiatributo, grupo al que la predicción de insolvencias al pretender asignar empresas (objetos), descritas por los valores de un conjunto de ratios financieros (atributos), a una clase o categoría (sana o fracasada).

Por otro lado, en comparación con otros métodos hemos constatado que el enfoque rough set ofrece las siguientes ventajas:

- Analiza los hechos escondidos en los datos, sin necesitar ninguna información adicional como probabilidades (en estadística) o grado de pertenencia (en la teoría fuzzy set). Es conceptualmente simple y emplea algoritmos sencillos.

- Reduce los atributos y objetos superfluos para obtener los denominados *reductos* (o *conjuntos mínimos de atributos*) que aseguran la misma calidad de clasificación que el conjunto total. A través de ellos se obtiene la mínima representación del conocimiento en forma de reglas de decisión teniendo, en consecuencia, una interpretación sencilla. Es decir, el modelo que obtenemos se desarrolla en forma de reglas de decisión que tienen en cuenta las preferencias del decisor que toma parte en el proceso de construcción de las mismas. Es más, las reglas de decisión revelan los atributos más relevantes que debería considerar un decisor para evaluar el riesgo de fracaso de una empresa.

El modelo de preferencia global, en forma de reglas de decisión derivadas de un conjunto de objetos, tiene ventajas explicativas sobre un modelo en forma de función ya que explicita las preferencias a través de los hechos importantes en términos –únicamente- de los criterios más significativos y de manera fácilmente comprensible. Las reglas se justifican a través de los objetos de las que se derivan y, aún más, las inconsistencias que se manifiestan en los objetos ni se corrigen ni se agregan (no se corrigen) sino que se calculan aproximaciones "por arriba" y "por abajo". En consecuencia, y basándose en estas aproximaciones, se inducen dos conjuntos de reglas: determinísticas (ciertas) y no-determinísticas (posibles). Estos conjuntos de reglas se pueden utilizar para justificar las decisiones que se tomen en relación a un nuevo objeto mediante el emparejamiento de éste a una de las reglas; si ello no es posible, se presenta al decisor el conjunto de reglas "más cercanas". Por tanto, descubre los hechos importantes escondidos en los datos y los expresa en el lenguaje natural de las reglas de decisión.

- Acepta tanto atributos cualitativos como cuantitativos (incluyendo aquellos cuyos dominios no están ordenados) y especifica su importancia mediante una medida de la aproximación a la clasificación. Además los modelos clásicos para la predicción del fracaso empresarial (por ejemplo los modelos basados en el análisis discriminante) proporcionan una medida financiera del riesgo, ignorando otras medidas importantes del mismo como

la experiencia de los directivos, el nicho de mercado o las ventajas competitivas que posea la empresa. Es decir, un modelo construido a través de reglas parece más cercano a una medida real del riesgo potencial de quiebra al poder tener en consideración variables de tipo cualitativo que, aunque excluidas de nuestro estudio falta de datos, podrían introducirse sin mayor problema para obtener esa evaluación global del riesgo.

- Contribuye a la minimización de tiempo y de coste del proceso de toma de decisión (el enfoque rough set es un sistema que procesa la información en tiempo real).
- Da transparencia a las decisiones de clasificación, lo que permite su discusión y refrendo.
- Toma en consideración todo el conocimiento adquirido del decisor.
- Finalmente, la utilización de un método multicriterio para abordar decisiones financieras en tiempo real implica la necesidad de disponer de un software que permita su aplicación práctica (en nuestro caso, se ha empleado ROSE).

Se ha comparado el análisis discriminante con la metodología rough set. Ésta presenta unas ventajas muy importantes sobre el primero, como son: a) no requiere especificar ningún dato a priori o que las variables empleadas satisfagan ningún tipo de hipótesis estadística; b) acepta variables de tipo cualitativo sin ningún tipo de limitación; c) se puede aplicar a sistemas de información con escasos números de datos; y, d) las reglas de decisión son muy flexibles con los cambios en el modelo a lo largo del tiempo y son fácilmente interpretables y aplicables por el decisor. En definitiva, es un método rápido para procesar gran cantidad de información tanto cualitativa como cuantitativa. Lo hace, además, de manera sencilla porque las reglas de decisión no requieren para su interpretación la ayuda de expertos. Todas estas características permiten su utilización por un gran número de usuarios (auditores, autoridades supervisoras, inversores y analistas) que pueden beneficiarse de una herramienta que les ayude a la hora de justificar sus decisiones con un gran ahorro en coste y en tiempo.

Con la metodología Rough Set, a partir del sistema de información inicial (72 empresas descritas mediante 19 ratios) realizamos dos tipos de experimentos obteniendo de cada uno de ellos dos modelos (dos conjuntos de reglas de decisión), a los que nos hemos referido como *modelo 1* y *modelo 2*. Las reglas de decisión del *modelo 1* se obtuvieron empleando los datos del año 1 (año anterior a la quiebra) y verificando la capacidad predictiva de las mismas utilizando los datos de los años 2, 3, 4 y 5 previos a la quiebra. Posteriormente estimamos una *función discriminante* con el año 1 y comprobamos su capacidad predictiva utilizando los años 2, 3, 4 y 5 anteriores a que las firmas quebraran, al objeto de comparar ambos métodos.

Dado que utilizar las mismas empresas para validar y para estimar podía condicionar los resultados obtenidos, obtuvimos un *segundo modelo* seleccionado

para el año 1 (año previo a la quiebra), 27 empresas sanas y 27 fracasadas de manera aleatoria para estimar las reglas de decisión, dejando las 18 restantes (9 sanas y 9 fracasadas) para comprobar la validez de las mismas. Pretendíamos de esta forma aplicar el modelo a datos que no hubieran sido utilizados para la elaboración del mismo.

Mediante el empleo de la metodología rough set obtuvimos las variables más significativas a través de los reductos, y el resto de atributos se pueden eliminar de la tabla de información codificada. Por tanto en el primer caso, con el reducto se pasó de un sistema inicial de información codificado de 19 columnas a otro de 6, y en el segundo caso se pasó a otro de 5 columnas. Los dos sistemas reducidos proporcionan la misma calidad de clasificación que la totalidad del conjunto de atributos. Demostramos, así, la fuerza de esta teoría en cuanto a selección de las características o variables más significativas de un sistema de información. Esto permite centrarse en aquellas variables clave reduciendo el tiempo y el coste del proceso de decisión.

En este estudio, el análisis rough set ha recalcado la importancia de los ratios financieros de liquidez, rentabilidad, cashflow, capacidad de endeudamiento, además de una adecuada tarificación, suficiencia de las provisiones y una conveniente política de reaseguro.

A partir de los reductos, obtuvimos los dos modelos consistentes en dos algoritmos compuestos, el primero de ellos por 25 reglas y el segundo de ellos por 27 reglas. Las reglas obtenidas constituyen para los dos modelos un algoritmo de clasificación. Son una representación no redundante del conocimiento contenido en la muestra seleccionada y están reunidas en un sistema de información codificado. Es decir, la muestra de 72 empresas utilizada en el estudio para derivar los algoritmos de decisión puede considerarse como una muestra de prueba empleada para revelar las características financieras que discriminan las empresas solventes de las insolventes.

Si observamos en los algoritmos obtenidos solo se han utilizado 58 y 63 descriptores para cada modelo, lo que supone menos del 5% (4.24% y 4.6%, respectivamente) de los que aparecían en el sistema de información inicial (1.368). Los algoritmos de clasificación obtenidos únicamente muestran qué criterios hay que tener en cuenta para evaluar la solvencia, pero es necesario validarlos de acuerdo con lo expuesto anteriormente para cada modelo. Los resultados del primer modelo de la teoría Rough Set, en comparación con los obtenidos con la función discriminante estimada son, en general, superiores en nuestra muestra.

Esta conclusión es muy importante porque convierte al enfoque rough set en el futuro, en una fuerte alternativa para el análisis de los problemas de gestión financiera.

Para el segundo modelo de esta teoría obtuvimos un porcentaje de aciertos del 77.78%, lo cual es un resultado bastante bueno ya que el modelo se ha validado sobre una muestra independiente.

Es necesario mencionar que las reglas se han derivado de unos datos particulares y representan una descripción general de la experiencia de un decisor en particular. De acuerdo con esto, las reglas no pueden aplicarse de manera general: para su aplicación a otros sectores deberían utilizarse nuevos datos, analizarlos y generar las reglas adecuadas, que representarán la experiencia del nuevo decisor.

Respecto a la **Algoritmia Heurística Híbrida** se ha presentado la aplicación de dos de estas técnicas a un problema de optimización combinatoria dentro de la predicción de insolvencias en compañías de seguros no-vida. Específicamente, se ha estudiado la aplicación de algoritmos genéticos y enfriamiento simulado para la selección de características o atributos (ratios financieros) y así mejorar la decisión (compañía solvente o insolvente) de un clasificador local (máquina de vectores soporte).

A lo largo de la exposición se han presentado las principales características de los algoritmos heurísticos utilizados, así como una descripción del problema de predicción de insolvencias en compañías de seguros, y un caso concreto del problema general de clasificación.

Hemos comprobado la aplicabilidad de las técnicas heurísticas descritas en un problema de selección de características en datos reales de empresas de seguros no-vida, obteniéndose resultados muy positivos (77% de aciertos sobre una muestra independiente) en la clasificación de estas empresas en solventes e insolventes. Además ese porcentaje de aciertos se han obtenido a través de dos conjuntos de tres ratios sin perder en absoluto capacidad de clasificación lo que representa una interesante simplificación. Estos datos nos dicen también que, en este caso, la mayoría de los ratios utilizados inicialmente en la descripción de las compañías de seguros analizadas eran superfluos, y solamente añadían ruido al sistema, sin ninguna información. Además los ratios seleccionados por estas dos últimas técnicas son prácticamente los mismos que los que aparecían como ratios significativos para la metodología rough set lo cual vuelve a recalcar la importancia de determinados ratios, tal y como se ha mencionado, para evaluar la solvencia en este sector.

Sin embargo, no debemos olvidar en cuanto a la verificación de los modelos que es evidente que no podemos obtener conclusiones significativas al disponer de un escaso número de empresas analizadas; sí podemos observar que el nivel de aciertos es muy superior al que cabría esperar de un modelo aleatorio y que existe un mayor nivel de aciertos cuanto más próximo está el momento de la crisis.

Por otro lado, el estudio tiene una serie de **limitaciones**. Una de ellas es que las compañías seleccionadas son sociedades anónimas con lo que se han excluido otras formas societarias del sector seguro y, en consecuencia, no es representativo del conjunto. Otra limitación es la falta de generalización derivada del escaso número de empresas analizadas.

Como *aplicaciones prácticas*, esta metodología puede ser muy útil para establecer un “sistema de alerta temprana” de detección de insolvencia para

entidades aseguradoras ya que la selección de ratios efectuada por cualquiera de las técnicas nos permiten centrarnos en un conjunto mínimo de variables. Este sistema podría ser muy útil para las autoridades reguladoras de este tipo de empresas, inversores, directivos, analistas financieros, bancos, auditores, asegurados y consumidores ya que centrándose en muy pocos ratios pueden controlar y revisar gran cantidad de firmas reduciendo por tanto el tiempo y los costes empleados en el proceso. Es más, los modelos obtenidos pueden usarse como sistemas de diagnóstico automático para preseleccionar por ejemplo aquellas empresas que necesiten una atención especial, de una manera rápida y a un coste bajo, y de este modo habría una gestión más eficiente del tiempo dedicado por el analista financiero o la autoridad supervisora dando, además, uniformidad a los juicios emitidos sobre una empresa.

En definitiva, y pese a las limitaciones mencionadas, nuestra investigación apoya la validez de estas metodologías en relación a la problemática derivada de la predicción de insolvencias en el sector seguros, configurando un conjunto de herramientas de razonable utilidad para investigadores y decisores del sector asegurador y con previsible proyección en otros problemas financieros.

BIBLIOGRAFÍA

Altman, E. I., Marco, G. y Varetto, F. (1994). Corporate distress diagnosis: Comparisons using discriminant analysis and neural networks (the Italian experience), *Journal of Banking and Finance* 18, 505-529.

Álvarez, S. y Urías, J. (1987). *Análisis Contable Superior*, Universidad Nacional de Educación a Distancia.

Ambrose, J.M. y Carroll, A.M. (1994). Using Best's Ratings in Life Insurer Insolvency Prediction, *The Journal of Risk and Insurance* Vol. 61, 2, 317-327.

Ambrose, J.M., y Seward, J.A. (1988). Best's Ratings, Financial Ratios and Prior Probabilities in Insolvency Prediction, *The Journal of Risk and Insurance*, 55, 229-244.

Bannister, J. (1997). *Insurance Solvency Analysis*, LLP limited, Segunda Edición.

Bar-Niv, R. y Smith, M.L. (1987). Underwriting, Investment and Solvency, *Journal of Insurance Regulation*, 5, 409-428.

Bar-Niv, R. (1990). Accounting Procedures, Market Data, Cash-Flow Figures, and Insolvency Classification: The case of the Insurance Industry, *The Accounting Review*, Vol. 65, 3, July, 578-604.

Best's, A.M. Company (1991). *Best's Insolvency Study, Property and Casualty Insurers 1969-1990*.

Browne, M. J. y Hoyt, R. E. (1996). Economic and market predictors of insolvencies in the property-liability insurance industry, *The journal of Risk and Insurance*, 309-327.

Burges, C. J. (1998). A tutorial on Support Vector Machines for pattern recognition, *Knowledge Discovery and Data Mining*, vol. 2, no. 2, 121-167.

Calvo-Flores, A. y García, D. (coord.) (1998a). *El riesgo financiero*, Asociación Española de Contabilidad y Administración de Empresas.

Calvo-Flores, A. y García, D. (coord.) (1998b). *Predicción de la Insolvencia Empresarial*, Asociación Española de Contabilidad y Administración de Empresas

Dimitras, A.I., Zopounidis, C. y Hurson, Ch. (1995). A multicriteria decision aid method for the assessment of business failure risk, *Foundations of Computing and Decision Sciences* 20, 2, 99-112.

Dimitras, A.I., Zanakis, S.H. y Zopounidis, C. (1996). A survey of business failures with and emphasis on prediction methods and industrial applications, *European Journal of Operational Research* 90, 487-513.

Dimitras, A.I., Slowinski, R., Susmaga, R. y Zopounidis, C. (1999). Business failure prediction using Rough Sets, *European Journal of Operational Research*, 114, 263-280.

Eisenbeis, R.A. (1977). Pitfalls in the application of discriminant analysis in business and economics, *The Journal of Finance* 32, 875-900.

Fayyad, K.B. y Irani, K.B. (1992). On the Handling of Continuous-Valued Attributes in Decision Tree Generation, *Machine Learning*, vol. 8, 87-102.

Fernández-Palacios, J. y Maestro J.L. (1991). *Manual de Contabilidad y Análisis Financiero*, Centro de Estudios del Seguro, S.A.

Fernández, A.J. (2000). *Análisis de los Estados Financieros de las Entidades Aseguradoras*, Aula Universitaria de Economía Actuarial y Financiera.

Gabás, F. (1990). *Técnicas Actuales de Análisis Contable. Evaluación de la Solvencia Empresarial*, Instituto de Contabilidad y Auditoría de Cuentas.

Gabás, F. (1997). Predicción de la insolvencia empresarial, en Calvo-Flores, A. y García, D. (coord.), *Predicción de la Insolvencia Empresarial*, Asociación Española de Contabilidad y Administración de Empresas, 11-32.

García, V. y Ballester, M. (1994). Limitaciones de los ratios financieros convencionales: una alternativa, *Actualidad Financiera*, 8/ 21-27 febrero, 159-167.

García, V. y García-Pérez, A. (coords.) (2000). *Decisiones Financieras y Fracaso Empresarial*, Asociación Española de Contabilidad y Administración de Empresas.

García, D., Calvo-Flores, A. y Arques, A. (1997) Factores discriminantes del riesgo financiero en la industria manufacturera española, en Calvo-Flores, A. y García, D. (coord.), *Predicción de la Insolvencia Empresarial*, Asociación Española de Contabilidad y Administración de Empresas, 125-156.

Glover, F. (1989). Tabu search - part I, *ORSA Journal on Computing*, vol. 1, nº 3, 190-206.

Glover, F. (1990). Tabu search - part II, *ORSA Journal on Computing*, vol. 2, nº. 1, 4-32.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley..

Greco, S., Matarazzo, B., y Slowinski, R. (1998). A new rough set approach to evaluation of bankruptcy risk, in C. Zopounidis (ed.), *New Operational Tools in the Management of Financial Risks*, Kluwer Academic Publishers, Dordrecht, 121-136.

Jiménez, S., García-Ayuso, M. y Sierra, G. (2000). *Análisis Financiero*, Editorial Pirámide.

Karels, G.V., y Prakash, A.J. (1987). Multivariate Normality and Forecasting of Business Bankruptcy, *Journal of Business Finance and Accounting*, Winter, 573-592.

Kennedy, J. y Eberhart, R.C. (1995). Particle swarm optimization, *Proceedings of the 1995 IEEE International Conference on Neural Networks*, Perth, Australia, 1942-1948.

Kirpatrick, S. y Gerlatt, C. D. y Vecchi, M. P. (1983). Optimization by simulated annealing, *Science*, vol. 220, 671-680.

Kirpatrick, S. (1984). Optimization by simulated annealing--Quantitative studies, *Journal of Statistical Physics*, vol. 34, 975-986.

Koza, J. (1992). *Genetic Programming*, MIT press, Cambridge, MA.

Krusinska, K., Slowinski, R. y Stefanowski, J., (1992). Discriminant versus rough sets approach to vague data analysis, *Applied Stochastic Models and Data Analysis* 8, 43-56.

Laitinen, E. K., (1992). Prediction of failure of a newly founded firm, *Journal of Business Venturing*, julio, 323-340.

Ley 33/1984, de 2 de agosto, sobre *Ordenación del Seguro Privado*, BOE nº 186, 4 de agosto de 1984.

Ley 30/1995, de 8 de noviembre, de *Ordenación y Supervisión de los seguros privados*, BOE nº 268, 9 de noviembre de 1995.

Linares, A. (1998). *Tratado de Auditoría de las Entidades Aseguradoras*, editorial Mapfre, S.A.

Linares, A., (1998). *Contabilidad de Entidades Aseguradoras*, editorial Mapfre, S.A.

Linares, A. (2000). *Auditoría y Control Interno de las Entidades Aseguradoras*, Editorial Fundación Mapfre Estudios.

López, D., Moreno, J. y Rodríguez, P. (1994). Modelos de Previsión del fracaso empresarial: aplicación a entidades de seguros, *Revista española de Seguros* 54, abril-junio, 71-110.

Maestro, J.L. (2000). *Garantías Técnico-Financieras de las Entidades Aseguradoras*, Editorial Grupo Winterthur.

Martín, M.L., Leguey, S., Sánchez, J. M. (1999). Solvencia y estabilidad financiera en la empresa de seguros: Metodología y evaluación empírica mediante análisis multivariante, *Cuadernos de la Fundación Mapfre Estudios*, 49.

McKee, T.E y Lensberg, T. (1999). Using a genetic algorithm to obtain a causally ordered model from a rough sets derived bankruptcy prediction model. Paper presented at *The International Symposium on Audit Research*, The University of Southern California.

McKee, T. (2000). Developing a Bankruptcy Prediction Model via Rough Sets Theory, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 9, 159-173.

Mienko, R., Slowinski, R., Stefanowski, J. y Susmaga, R. (1996a). RoughFamily – Software implementation of the rough set based data analysis and rule discovery techniques, in Tsumoto S. (ed), *Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, Tokyo Nov. 6-8, 437-440.

Mienko, R., Stefanowski, J., Toumi, K y D. Vanderpooten (1996b). Discovery-Oriented Induction of Decision Rules, *Cahier du Lamsade 141*, Paris, Université de Paris Dauphine, september.

Millán, A., (2000). *Análisis Contable de Entidades Aseguradoras*, Asociación Española de Contabilidad y Administración de Empresas.

Mora, A. (1994). Los modelos de predicción del fracaso empresarial: una aplicación empírica del logit, *Revista Española de Financiación y Contabilidad*, 78, enero-marzo, 203-233.

Moscarola, J. (1978). Multicriteria decision aid – two applications in education management. In: Multiple Criteria Problem Solving (S. Zionts, ed.). *Lecture Notes in Economics and Mathematical Systems*, vol. 155, Springer-Verlag, Berlin, 402-423.

Nurmi, H., Kacprzyk, J. y Fedrizzi, M. (1996). Probabilistic, fuzzy and rough concepts in social choice. *European journal of Operational Research* 95, 264-277.

Ohlson, J.A., (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18, 109-131.

O'Leary, D.E. (1998). Using neural networks to predict corporate fealure, *International Journal of Intelligent Systems in Accounting Finance and Management*, vol. 7, 187-197.

Orden Ministerial de 30 de julio de 1981, por la que se aprueban las normas de adaptación del Plan General de Contabilidad a las empresas de seguros, reaseguros y capitalización.

Pawlak, Z. (1982). Rough Sets. *International Journal of Information & Computer Sciences* 11, 341-356.

Pawlak, Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht/ Boston/ London.

Pawlak, Z. y Slowinski, R. (1994). Rough set approach to multi-attribute decision analysis, *European Journal of Operational Research* 72, 443-459.

Predki, B., Slowinski, R., Stefanowski, J., Susmaga, R. y Wilk, S. (1998). ROSE – Software Implementation of the Rough Set Theory, in L. Polkowski, A. Skowron, eds. *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, vol. 1424. Springer-Verlag, Berlin, 605-608.

Predki, B. y Wilk, S. (1999). Rough Set Based Data Exploration Using ROSE System. In: Z.W. Ras, A. Skowron eds. *Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 1609, Springer-Verlag, Berlin, 172-180.

Prieto, E. (1973). *El Reaseguro: Función Económica*, Ediciones ICE.

Prieto, E. (1992). La estructura del Balance de las entidades aseguradoras en la CEE, *Revista Española de Seguros* 69, enero-marzo, 15-28.

Prieto, E., (1993). Los riesgos de inversión y gestión de carteras en las entidades aseguradoras, *Análisis Financiero*, 59, 1^{er} cuatrimestre, 95.

Pozo, E.M. (1998). Modelos de Ratios para la predicción de Insolvencias en seguros no vida, *Previsión y Seguro*, 68, 37-48.

Real Decreto 1348/1985, de 1 de agosto, por el que se aprueba el *Reglamento de Ordenación del Seguro Privado*. BOE nº. 185, 186, 187 de 3, 5 y 6 de agosto de 1985.

Real Decreto 2020/1986, de 22 de agosto, por el que se aprueba el *Reglamento de la Comisión Liquidadora de Entidades Aseguradoras*. BOE nº 235, de 1 de octubre.

Real Decreto 2014/1997, de 26 de diciembre, por el que se aprueba del Plan de Contabilidad de Entidades Aseguradoras y normas de formulación de las cuentas de los grupos aseguradores, BOE, 30 de diciembre de 1997.

Real Decreto 2486/1998, de 20 de noviembre, por el que se aprueba el *Reglamento de Ordenación y supervisión de los Seguros Privados*. BOE nº 282 de 25 de noviembre de 1998.

Rivero, P. (1987). *Análisis de balances y estados complementarios*, Editorial Pirámide.

Rivero, P., Banegas, R., Sánchez, F. y Nevado, D. (1998). *Análisis por Ratios de los Estados Contables Financieros (Análisis Externo)*, Editorial Civitas.

Rodríguez, M.C. (1990). *La Predicción de las Crisis Empresariales. Modelos para el sector Asegurador*, Secretariado de Publicaciones Universidad de Valladolid.

Romero, C. (1993). *Teoría de la decisión multicriterio: Conceptos, técnicas y aplicaciones*, Alianza Universidad Textos.

Roqueta, A. (2000). El principio de empresa en funcionamiento: a través de un ratio financiero, *El Auditor* 8- Diciembre 2000, 2-5.

Salcedo-Sanz, S., Fernández-Villacañas, J. L., Segovia-Vargas M.J. y Bousoño-Calzón, C. (2004). Genetic programming for the prediction of insolvency in non-life insurance companies", *Computers and Operations Research*, in press.

Sanchis, A. (2000). *Una aplicación del Análisis Discriminante a la previsión de la Insolvencia en las empresas españolas de seguros no-vida*, Tesis Doctoral, Universidad Complutense de Madrid.

Sanchis, A., Gil, J.A. y Heras, A. (2003). El análisis discriminante en la previsión de la insolvencia en las empresas de seguros no vida, *Revista Española de Financiación y Contabilidad*, vol. 116.

Schölkopf, B. y Smola, A.J. (2002). *Learning with Kernels*, MIT press, Cambridge, MA.

Serrano, C. y Martín del Brío, B., (1993). Predicción de la Crisis Bancaria Mediante el Empleo de Redes Neuronales Artificiales, *Revista Española de Financiación y Contabilidad XXII*, nº 74, p.153-176.

Skowron, A. (1991a). *The Implementation of Algorithms based on Discernibility Matrix*, Manuscript.

Skowron, A. (1991b). The Discernibility Matrices and Functions in Information Systems, *Institute of Computer Sciences Reports*, Warsaw Technical University and Fundamental Informaticae.

Slowinski, R. (1993). Rough set learning of preferential attitude in multicriteria decision making, in: J. Komorowski and Z. W. Ras (eds.), *Methodologies for Intelligent Systems*. Lecture Notes in Artificial Intelligence vol. 689, Springer-Verlag, Berlin, 642-651.

Slowinski, R. (1995). Rough set approach to decision analysis, *AI Expert Magazine* 10, 3, 18-25.

Slowinski R. y Stefanowski J. (1992). "RoughDAS" and "RoughClass" software implementations of the rough set approach. Chapter III-8 in Slowinski (1992), 445-456.

Slowinski R. y Stefanowski J. (1994a). *RoughDas: Rough Set Based Data Analysis System-Version 2.0- User's Guide book*, Poznan, Poland.

Slowinski R. y Stefanowski J. (1994b). Rough classification with valued closeness relation, in: Diday E. et al (eds.), *New Approaches in Classification and Data Analysis*, Springer-Verlag, Berlin, 482-488.

Slowinski R. y Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk, *International Journal of Intelligent Systems in Accounting, Finance and Management* 4, nº 1, 27-41.

Stefanowski, J. (1992). Rough Sets Theory and discriminant methods as tools for analysis of information systems. A comparative study, *Foundations of Computing and Decision Sciences* vol 17 (2), 81-98.

Stefanowski J. (1993a). Classification and decision supporting based on the rough set theory, *Foundations of Computing and Decision Sciences*, vol. 18, nº 3-4, 371-380.

Stefanowski J. (1993 b). Classification support based on the rough set theory, in J. Wessels, A.P. Wierzbicki (eds.): *User-Oriented Methodology and Techniques of Decision Analysis and Support, LNEMS 397*, Springer Verlag, Berlin, 185-192.

Tam, K.Y. y Kiang, M.Y. (1992). Managerial applications of neural networks: The case of bank failure predictions, *Management Science*, vol. 38, nº 7, 926-947.

Urias, J. (1999). *Análisis de Estados Financieros* (Segunda Edición), Editorial McGraw-Hill.

Weston, H., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, M. y Vapnik, V. (2000). Feature selection for SVMs, *Advances in NIPS*, 12, MIT press, Cambridge, MA.

Zavgren, C.V. (1985). Assessing the vulnerability to failure of American industrial firms. A logistic analysis, *Journal of Business Finance and Accounting* 12 (1), 19-45.

Zopounidis, C. y Dimitras A. (1998). *Multicriteria Decision Aid Methods for the Prediction of Business Failure*, Kluwer Academic Publishers, Dordrecht.

MAP 6 SEG APL
37795



© FUNDACIÓN MAPFRE

Prohibida la reproducción total o parcial de esta obra sin el permiso escrito del autor o de FUNDACIÓN MAPFRE