INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. **12**, 000–000 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/isaf.255

FEATURE SELECTION METHODS INVOLVING SUPPORT VECTOR MACHINES FOR PREDICTION OF INSOLVENCY IN NON-LIFE INSURANCE COMPANIES

SANCHO SALCEDO-SANZ,[a] MARIO DEPRADO-CUMPLIDO,[a]* MARÍA JESÚS SEGOVIA-VARGAS,[b] FERNANDO PÉREZ-CRUZ[a] AND CARLOS BOUSOÑO-CALZÓN[a]

[a] *Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Madrid, Spain*
[b] *Department of Financial Economy and Accounting I, Universidad Complutense de Madrid, Madrid, Spain*

SUMMARY

We propose two novel approaches for feature selection and ranking tasks based on simulated annealing (SA) and Walsh analysis, which use a support vector machine as an underlying classifier. These approaches are inspired by one of the key problems in the insurance sector: predicting the insolvency of a non-life insurance company. This prediction is based on accounting ratios, which measure the health of the companies. The approaches proposed provide a set of ratios (the SA approach) and a ranking of the ratios (the Walsh analysis ranking) that would allow a decision about the financial state of each company studied. The proposed feature selection methods are applied to the prediction the insolvency of several Spanish non-life insurance companies, yielding state-of-the-art results in the tests performed. Copyright © 2004 John Wiley & Sons, Ltd.

1. INTRODUCTION

Financial risk assessment is one of the key issues when controlling the insurance market sector, due to the high amount of money payable by the state insurance guaranty funds when an insurance company has gone bankrupt. At the same time, it is not desirable to perturb the market excessively with unneeded interventions. Also, protecting society and the whole insurance sector against insolvent insurance companies is of great concern to auditors, governments and managers in the sector, because a bankruptcy reduces public confidence in all insurance companies. These facts explain the increasing interest in accurately predicting insurance company failures. The European Union, through the *Solvency II Project* has taken an active role in redefining a set of rules to provide society with information about how healthy the insurance companies are.

Many insolvency[1] cases appeared after the insurance cycles of the 1970s and 1980s in the USA and the European Union. Several surveys have been devoted to identifing the main causes of insurers' insolvency. In particular, the Müller Group Report (Müller Group, 1997) analyses the main identified causes of insurance insolvencies in the European Union. These main causes can be

* Correspondence to: Mario DePrado-Cumplido, Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Madrid, Spain. E-mail: mprado@tsc.uc3m.es
Contract/grant sponsor: Comunidad de Madrid; Contract/grant number: 07T/0016/2003.
[1] In general financial terms, insolvency can be referred to as the inability of a firm to pay its debts. A prior period of insolvency could be overcome, for example, by means of the postponement in the payments of the debts. If the firm is unable to overcome this first period, it can become bankrupt. Therefore, bankruptcy could be interpreted as the culmination of the insolvency process.

summarized as follows: operational risks (operational failure related to inexperienced or incompetent management, fraud); underwriting risks (inadequate reinsurance programme and failure to recover from reinsurers, higher losses due to rapid growth, excessive operating costs, poor underwriting process); insufficient provisions and imprudent investments. On the other hand, many insurance companies, especially large companies, have developed internal risk models for a number of purposes. In general, these models are partial in nature and do not cover all possible risks. Therefore, developing new methods to tackle prudential supervision in insurance companies is a highly topical question, especially for countries that belong to the European Union, such as Spain.

Most approaches to insolvency research in insurance companies use statistical methods such as discriminant or logit analysis (Barniv, 1990; Ambrose and Carol, 1994; Sanchis *et al.*, 2003), and use financial ratios as explicative variables. This kind of variable does not usually satisfy statistical assumptions and, therefore, results obtained by these statistical techniques could be erroneous. Moreover, models obtained by these techniques are difficult to interpret by the decision maker if he/she is not an expert. Thus, in order to avoid these problems, we propose an approach to predicting insolvency of insurance companies based on emergent and machine learning methods. We use these methods to select the set of financial ratios that best classifies companies as solvent on insolvent. A sample of Spanish non-life insurance firms is used, and general financial ratios, as well as those that are specifically proposed for evaluating insolvency of the insurance sector, are employed.

There have been several previous proposals to apply operational research and pattern recognition methods to predict insolvency (Barniv, 1990; Tam, 1991; Ambrose and Carol, 1994), but few have been applied to the insurance sector. Among these approaches to the prediction of insolvency in insurance companies the emergent classifiers algorithms such as neural networks (Serrano-Cinca, 1996; O'Leary, 1998), rough set (Dimitras *et al*., 1998), support vector machines (SVMs) or genetic programming (GP; Li and Tsang, 2000; Salcedo-Sanz *et al.*, 2002), have gained importance in the last few years as powerful methods which provide very good results in classification performance and generalization. This paper is focused on the application of SVM-based methods for improving the prediction of the insolvency of non-life insurance companies.

SVMs are state-of-the-art tools for linear and nonlinear knowledge discovery problems (Vapnik, 1998; Schölkopf and Smola, 2002). SVMs are designed to work in high-dimensional spaces even when very few input patterns are available, as in the situation at hand. The SVM is based on the maximum margin idea, which states that, without any prior knowledge, the best classifications boundary must correctly classify every given sample and be situated as far as possible from all the samples, reducing the risk of misplacing a new unseen pattern. But in many applications there are irrelevant features that, if the data are scarce, may bias the solution of the pattern classifier employed and will make it perform poorly with new unseen examples.

Feature selection (FS) is an open problem in machine learning, which basically consists of finding a subset of input features that describes the underlying system structure as well as, or better, than all available features. In the ranking problem, the point is to weigh the features according to relevance. The significance of FS appears when irrelevant features reduce classification accuracy, and when the given features are used to explain the results achieved. Note that in some applications, such as financial ones, being able to explain the solution obtained (in terms of the input features selected) becomes as relevant as obtaining the best possible answer (accuracy of the subsequent classifier or regressor).

In this paper we propose two new FS procedures that work jointly with an SVM. The first proposal is based on a simulated annealing (SA) algorithm (Kirpatrick *et al.*, 1983; González *et al.*, 2002), which uses the SVM validation error to select the most relevant features for the problem at hand. Our second proposal works with the Walsh expansion (Vose and Wright, 1998) of a particular binary function, which relates binary strings to the SVM soft output. This expansion will not be used to select the best features, but to rank them, giving richer information about the problem. The ranking will allow us to discuss the relevance of each of the features.

The rest of the paper is organized as follows. We provide background material on SVMs, FS, SA and Walsh analysis in Section 2. Section 3 is devoted to the proposed FS schemes. We show the actual validity of the proposed approaches with non-life insurance companies in Section 4, and we end the paper in Section 5 with some concluding remarks and suggestions for further work.

2. BACKGROUND

2.1. SVMs

The SVM has been reported as a powerful method for classification problems, with very good generalization properties (Burges, 1998). This section provides a brief summary of the standard SVM for classification[2] applied to insolvency, starting from the simple linear SVM and moving on to the nonlinear SVM.

Consider a set of l firms represented by the value of their n ratios $\{\mathbf{x}_i\}$, $i = 1, \ldots, l$, with $\mathrm{x}_i \in \mathbb{R}^n$, and a set of associated labels $y_i \in \{-1, 1\}$ that describe the firm as failed or healthy respectively. First, imagine that this training set can be separated by a linear hyperplane (a line in two dimensions, a plane in three dimensions, and so on). The SVM solves the following problem. Find $\mathbf{w} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, to minimize $1/2\|\mathbf{w}\|^2$, subject to

$$y_i(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}_i + b) \geq 1 \qquad \forall i = 1, \ldots, l \tag{1}$$

Once such $\mathbf{w}$ and b have been found,[3] our classification rule for a new firm $\mathbf{x}$ is given by sign $(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + b)$. Thus, firms located on one side of the hyperplane will be healthy and on the other side will be failed. The associated error of this classification, $R(\mathbf{w}, b)$, is defined as the percentage of misclassified firms.

Consider now the case when the points in the training set $\{\mathbf{x}_i\}$ are not linearly separable; then, constraint (1) cannot be satisfied. We can then introduce some nonnegative slack variables ξ_i in order to overcome this difficulty. The following SVM formulation results in this case. Find $\mathbf{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$ and ξ_i, $i = 1, \ldots, l$, to minimize $1/2\|\mathbf{w}\|^2 + C\,\Sigma_{i=1}^{l}\,\xi_i$, under the constraints

$$y_i(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \qquad \forall i = 1, \ldots, l \tag{2}$$

$$\xi_i \geq 0 \qquad \forall i = 1, \ldots, l \tag{3}$$

where C is a parameter of the classifier to be estimated.

Figure 1 shows an example of a classification problem formed by two classes (crosses and dots). The solution in figure 1(a) is obtained with a linear SVM, which defines a linear decision boundary

[2] A more complete analysis and further results about SVMs can be found in Burges (1998) and Schölkopf *et al*. (2002).
[3] Note that $\mathbf{w}$ and b are the parameters that characterize the hyperplane.

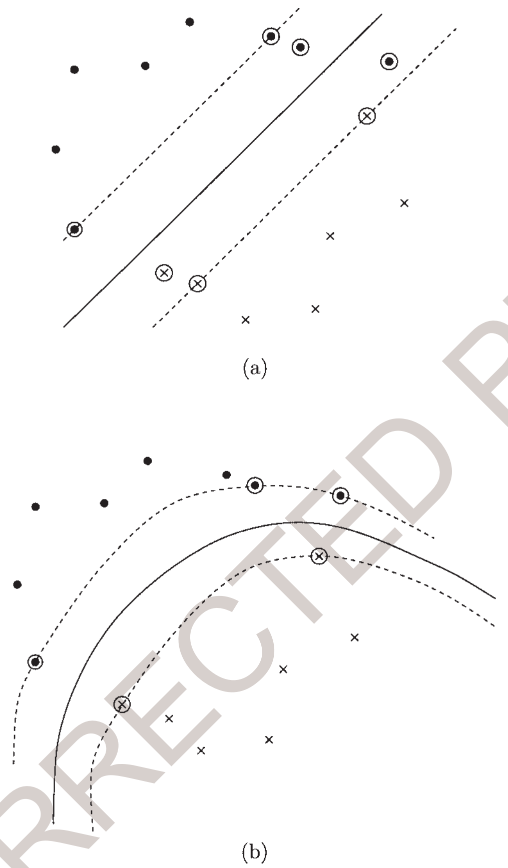


Figure 1. (a) Illustration of linear SVM for solving a two-class classification problem. Note that some samples of dot class are misclassified. (b) Illustration of the same problem using a nonlinear SVM (radial basis function kernel). Now the boundary obtained is capable of completely separing both sample sets

(**w**) unable to separate both classes completely. The dashed lines represent the margins, i.e. the set of points that satisfies equation (2). The samples over the margins, which are surrounded by circles, are the support vectors, the only information needed to plot the boundary. The samples located out of their regions are misclassifications (consequently with ξ_i greater than zero).

The classification obtained with the introduction of the slack variables ξ_i is still given by a linear frontier. The nonlinear SVM maps the input sample into a higher dimensional (often infinite dimensional) feature space, and applies the linear SVM in this feature space. All the appearances of the mapping ϕ are within dot products, which can be substituted by a kernel function. The nonlinear SVM with kernel K is equivalent to a regularization problem in the reproducing kernel Hilbert space H_K:

Find the mapping $\phi(\mathbf{x}) = h(\mathbf{x}) + b$ with $h \in H_K$, $b \in \mathbb{R}$ and ξ_i, $i = 1, \ldots, l$, to minimize

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \qquad (4)$$

subject to

$$y_i(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \qquad \forall i = 1, \ldots, l \tag{5}$$

$$\xi_i \geq 0 \qquad \forall i = 1, \ldots, l \tag{6}$$

Figure 1(b) illustrates a nonlinear SVM classification with a Gaussian kernel. The sample set is identical to figure 1(a), but in this case all samples are correctly classified. The resolution of the problem is obtained by a linear boundary in the Hilbert space generated by the kernel, which in the input space (the one represented in the figure) is a simple curve.

The kernel used in this article is the well-known Gaussian kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\gamma}\right)$$

where γ controls the width of the Gaussian.[4] The nonlinear SVM is able to classify any set of firms as healthy or failed, with a probability of error given by $R(\mathbf{w}, b)$.

Given a training set, the selection of the input variables (financial ratios) is an important issue to be considered, since irrelevant or redundant ratios can affect in a negative way the result given by the SVM. This is the so-called FS *problem* (FSP Weston *et al.*, 2000), in which the features are the financial ratios. In the next subsection we give a brief review of the FSP focused on non-life insurance insolvency prediction.

2.2. FS

The FSP in a learning from samples scheme can be addressed as follows: given a set of labelled data points $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \{\pm 1\}$, choose a subset of m features ($m < n$), that achieves the lowest classification error (Weston *et al.*, 2000; Salcedo-Sanz *et al.*, 2002). Following Weston *et al*. (2000), we will define the ESP as finding the optimum n-column vector $\boldsymbol{\sigma}$, where $\sigma_i \in \{0, 1\}$, that defines the subset of selected features, which is found as

$$\boldsymbol{\sigma}^{\mathrm{o}} = \arg\min_{\boldsymbol{\sigma}, \boldsymbol{\alpha}} \left(\int V(y, f(\mathbf{x} * \boldsymbol{\sigma}, \boldsymbol{\alpha}))\, \mathrm{d}P(\mathbf{x}, y) \right) \tag{7}$$

where $V(\cdot,\cdot)$ is a loss functional, $P(\mathbf{x}, y)$ is the unknown probability function that the data were sampled from, and we have defined $\mathbf{x} * \boldsymbol{\sigma} = (x_1\sigma_1, \ldots, x_n\sigma_n)$. The function $y = f(\mathbf{x}, \boldsymbol{\alpha})$ is the classification engine that is evaluated for each subset selection $\boldsymbol{\sigma}$ and for each set of its hyper-parameters $\boldsymbol{\alpha}$.

In this problem, the objective is to process the data in order to extract a valid, novel, potentially useful, and ultimately understandable structure in data by identifying relevant and meaningless features (Bradley *et al.*, 1999). This is the first step in a *knowledge discovery* learning scheme. In this context, two main approaches can be followed:

[4] This value and C are the two tunable parameters of the SVM.

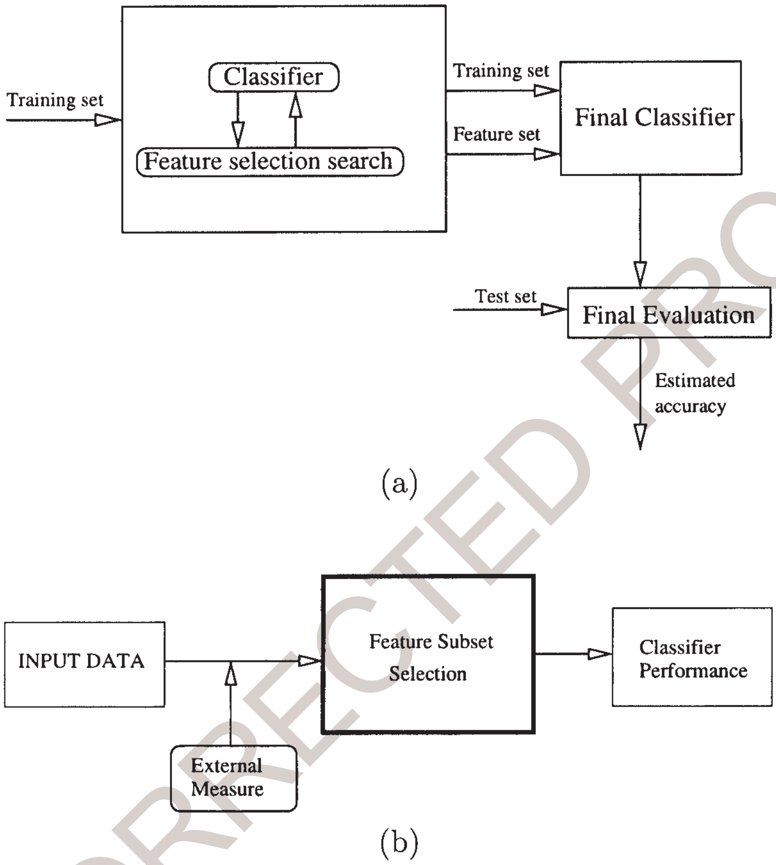


Figure 2. (a) Outline of a wrapper method; (b) outline of a filter method

- The *wrapper approach* to the FSP was introduced in John *et al*. (1994). In this approach, the FS algorithm conducts a search for a good subset of features using the classifier itself as part of the evaluation function. Figure 2(a) shows the idea behind the wrapper approach: the classifier is run on the training dataset with different subsets of features. The feature subset that produces the lowest estimated error in an independent but representative test set is chosen as the final feature set. For further consideration of wrapper methods, see Kohavi and John (1997) and Salcedo-Sanz *et al*. (2002).
- In the *filter approach* to the FSP, the FS is performed based on the data, and ignoring the classifier algorithm. An external measure calculated from the data must be defined in order to select a subset of features. After the search, the best feature subset found is evaluated on the data by means of the classifier algorithm. Note that the filter algorithm's performance depends completely on the measure selected for comparing feature subsets. Figure 2(b) shows an example of how a filter algorithm works. Filter methods are usually faster than wrapper methods. However, their main drawback is that they totally ignore the effect of the selected feature subset on the performance of the classification algorithm. Further analysis and application of filter methods can be found in Chen *et al*. (1999) and Salcedo-Sanz *et al*. (2002).

For both wrapper and filter methods, a binary representation can be used for the FSP, where a 1 in the ith position of the binary vector $\boldsymbol{\sigma}$ means that the feature i is considered within the subset of features, and a 0 in the jth position of the binary vector means that feature j is not considered within the subset of features. Note that using this notation is equivalent to encoding the problem as the vector $\boldsymbol{\sigma}$ included in expression (7). Note also that there are 2^n different subsets of features (n total number of features), and the problem is to select the best one in terms of a certain measure, which can be either internal (wrapper methods) or prior (filter methods) to the classifier.

2.3. SA

SA is a powerful solution technique that has been successfully applied to a wide variety of optimization problems (Kirpatrick *et al.*, 1983; Wang and Ansari, 1997; Gonzalez *et al.*, 2000). It is inspired by the physical process of heating a substance and then cooling it slowly until a strong crystalline structure is obtained. This process is simulated by lowering an initial temperature by slow stages until the system reaches an equilibrium point, and no more changes occur. Each stage of the process consists of changing the configuration several times, until a thermal equilibrium is reached, and a new stage starts with a lower temperature. The solution of the problem is the configuration obtained in the last stage. In the standard SA, the changes in the configuration are performed in the following way: a new configuration is built by a random displacement of the current one. If the new configuration is better, then it replaces the corrent one; if not, it may replace the current one probabilistically. This probability of replacement is high in the beginning of the algorithm, and decreases at every stage. This procedure allows the system to move toward the best configuration. However, SA is not guaranteed to find the global optimum, though it is better than other algorithms at escaping from local optima. The solution found by SA can be considered a 'good enough' solution, but it is not guaranteed to be the best.

The most important parts in an SA algorithm are the chosen representation for solutions, the objective function to be minimized during the process, and the mutation or configuration change operator.

2.4. Walsh Analysis and Spectrum

Walsh analysis is a commonly used method for studying the internal structure of binary functions (Vose and Wright, 1998). Walsh analysis is equivalent to a Fourier expansion of a function in binary search space $\{0, 1\}^n$. The Walsh expansion of a function associates a Walsh coefficient w_j with a binary vector $\mathbf{j}$ (*partition*). The function can be completely reconstructed from partition[5] $\mathbf{j}$s and Walsh coefficients w_j. Continuing with the notation, we give the main steps used to define the Walsh expansion of a function.

The Walsh basis function for a partition $\mathbf{j}$, $\psi_{\mathbf{j}} : \{0, 1\}^n \to \mathbb{R}$ is defined as

$$\psi_{\mathbf{j}}(\mathbf{x}) = \prod_{i=1}^{n} (-1)^{x_i j_i} \tag{8}$$

where x_i and j_i are the components of binary vectors $\mathbf{x}$ and $\mathbf{j}$.

[5] Hereafter, we will denote in bold a partition indexed as a binary vector, $\mathbf{j}$, and in italics, j, for its corresponding integer value.

Walsh functions form a complete orthogonal set of basis functions (Goldberg, 1989). Every function $f : \{0, 1\}^n \to \mathbb{R}$ can be expanded as

$$f(\mathbf{x}) = \sum_{j=0}^{2^n-1} w_j \psi_{\mathbf{j}}(\mathbf{x}) \tag{9}$$

where

$$w_j = \frac{1}{2^n} \sum_{x=0}^{2^n-1} f(\mathbf{x}) \psi_{\mathbf{j}}(\mathbf{x}) \tag{10}$$

The Walsh expansion captures the internal structure of a function: if this function has dependencies among variables, then its Walsh coefficients for partitions involving non-dependent variables are zero. For example, let $f(x_1, x_2, x_3, x_4) = f_1(x_1, x_2) + f_2(x_3, x_4)$, then $w_{0111} = w_{1110} = w_{0110} = 0$ (Kargupta, 1999). For further analysis and details of Walsh analysis see Vose and Wright (1998).

The *spectrum* of a function (Hordijk and Stadler, 1998) is a graphic representation of the most
1 important partitions of the function which is obtained from its Walsh expansion. The *order* of a partition $\mathbf{j}$ is defined as the number of 1s on it. Note that, in the FSP with binary representation, the order of a partition is equivalent to the number of features selected.

Using the definitions above, the spectrum of a function is defined starting from its Walsh coefficients as follows: let $\wp$ be the set of all partitions belonging to the search space $S = \{0, 1\}^n$. Let $\wp_p$ be the set of partitions belonging to S with order p. The total energy for the function is defined as

$$\sigma^2 = \sum_{j \in \wp} w_j^2 \tag{11}$$

The energy for the partitions with order p is

$$\beta_p^2 = \sum_{j \in \wp_p} w_j^2 \tag{12}$$

and their normalized energy

$$B_p = \frac{\beta_p^2}{\sigma^2} \tag{13}$$

Vector B = $\{B_1 \ldots B_n\}$ is the *spectrum* associated with the Walsh expansion of $f(\cdot)$. It can be readily shown that $B_p \geq 0$ and $\sum_p B_p = 1$. In this paper we will use a modification of the spectrum in order to perform a ranking of the best features for a given FSP problem.

3. PROPOSED FS METHODS

3.1. SA with SVMs

The first method we present in this paper of FS is based on an SA algorithm, which performs a search over the space $\boldsymbol{\sigma}$ of binary strings (See definition of FSP in Section 2.2). As was mentioned above, the main parts of an SA algorithm are the chosen representation for solutions, the objective function to be minimized during the process, and the mutation or configuration change operator. We describe them as follows:

Problem representation. We encode every solution to the FS as a binary string $\boldsymbol{\sigma}$, where $\sigma_i \in \{0, 1\}$ defines the subset of features selected. The length of binary vector $\boldsymbol{\sigma}$ will be equal to the total number of features in the problem.
Objective Function. We use the probability of error in testing given by a SVM as the objective function to be minimized by the SA. The SA algorithm will look for configurations (feature subsets) which provide the least error probability in the test set.
Mutation Operator. In this paper we consider a classical *random flip mutation* (RFM) operator, where N_f bits are randomly selected and flipped to obtain a configuration in the neighbourhood of the current one.

The SA we use has the following pseudo-code:

Pseudo-code of the SA algorithm.

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

---

where k counts the number of iterations performed, T is the current temperature, T_0 is the initial temperature, $T_{\min}$ is the minimum temperature to be reached, $\boldsymbol{\sigma}$ stands for the current configuration, and $\boldsymbol{\sigma}^{\text{mut}}$ stands for the new configuration after the mutation operator is applied. $f(\boldsymbol{\sigma})$ represents the objective function (probability of error in testing provided by the SVM in this case), M is the number of changes performed with a given temperature T, f_T is the freezer function, and a is a previously fixed constant. Parameter a and the initial temperature T_0 are calculated in order to set the initial acceptance probability to be 0.8, which is the value usually used. The freezer function is defined as

$$f_T = \frac{T_0}{1 + k} \tag{14}$$

The minimum temperature $T_{\min}$ is calculated on the basis of the desired number of iterations as

$$T_{\min} = f_T(T_0, \text{ numIt}). \tag{15}$$

The current configuration of the SA algorithm in the last iteration is interpreted as the solution of the problem $\boldsymbol{\sigma}^{o}$.

3.2. Walsh Analysis for Ranking of Features

The second approach to the FS we present consists of obtaining a ranking of features analysing the intrinsic structure of an SVM classifier by means of a Walsh analysis procedure.

The spectrum of a function, defined in Hordijk and Stadler (1998) (also summarized in Section 2.4), represents the distribution of energies among different orders (number of 1s) of the partitions that form the search space. If the function analysed is the objective function of an FSP, the spectrum gives a measure of what are the most important features.

The definition of the Spectrum from the Walsh analysis of a function given in Hordijk and Stadler (1998) calculates the energy associated with partitions of the same order. For example, in the search space $\{0, 1\}^4$, partitions {1011}, {1110}, {0111} contribute to the same component of the spectrum (B_3). However, they represent three different sets of features. Thus, the 'classical' definition of the spectrum cannot be used for ranking the features according to their relevance. In order to solve this problem, we propose a slight modification of the definition of the spectrum.

Let ζ_i be the set of partitions **j** with a 1 in the *i*th position. In the example above, partitions {1011} and {1110} belong to ζ_1 (they also belong to ζ_3) whereas partition {0111} does not belong to ζ_1 (but it belongs to ζ_3). Therefore, a modified spectrum, termed the *prime spectrum*, can be defined as follows:

$$B_i' = \frac{\sum_{j \in \zeta_i} w_j^2}{\sum_{\forall j} o(\mathbf{j}) \cdot w_j^2} \tag{16}$$

where $o(\mathbf{j})$ is the order of partition **j**.

The prime spectrum fulfils $B_i' \geq 0$ and $\Sigma_i B_i' = 1$. In fact, it can be interpreted as the associated energy of every feature in the binary search space and, thus, features with large values of B_i' are more relevant than features with small values of B_i'. Consequently, we propose using vector $\mathbf{B}'$ to perform a ranking of features in the FSP.

Vector $\mathbf{B}'$ depends on the fitness function selected for the FSP through the values of its Walsh coefficients. Note that, in large search spaces, the calculation of the Walsh expansion can be computationally infeasible, and estimation methods such as the one proposed in Hordijk and Stadler (1998) should be used. We briefly describe this method in the Appendix.

In this paper we propose the calculation of the vector $\mathbf{B}'$ using as the objective function the test error provided by an SVM in the test set. Vector $\mathbf{B}'$ provides a measure of the intrinsic structure of the SVM when different features are removed from the data. The calculation of $\mathbf{B}'$ in pseudocode is as follows:

Pseudo-code of the calculation of B′ spectrum.

for $j = 0$ **to** $2^n - 1$
  $j \rightarrow \sigma$ (Decimal to binary step)
    Run the SVM $\rightarrow f(\boldsymbol{\sigma}) = P_e(\text{test})$;
    Calculate $w_j = \frac{1}{2^n} \sum_{x=0}^{2^n-1} f(\boldsymbol{\sigma})\psi_{\mathbf{j}}(\mathbf{x})$
endfor
for $i = 1$ **to** n
  $$B'_i = \frac{\sum_{j \in \zeta_i} w_j^2}{\sum_{\forall j} o(\mathbf{j}) \cdot w_j^2}$$
endfor

4. EXPERIMENTS AND RESULTS

4.1. Test Data and Input Variables

In this section we describe the main characteristics of the data and identify the variables that will be used to test the approaches presented in this paper. We have used the sample of firms also used by Sanchis *et al*. (2003). This data sample consists of Spanish non-life insurance firms data (balance sheets and income statements) 5 years prior to failure. The firms were in operation or went bankrupt between 1983 and 1994. In each period, 72 firms (36 failed and 36 non-failed) are selected. As a control measure, a failed firm is matched with a non-failed one in terms of premium volume. In our study we have used data 1 year prior to the firms being declared bankrupt. Therefore, the prediction of the insolvency achieved by our method is 1 year in advance.

Because we are looking for those financial ratios that could help the decision maker anticipate possible insolvencies, our definition of insolvency has been made in a strict sense. The insolvent group consists of those firms that have been taken over by the Spanish ministry of economy, and have disappeared due to permanent financial problems. In this way we avoid working with firms that have temporary financial problems or that have disappeared voluntarily. We have checked that the firms in the solvent group continued working for several years after the sample period, because it is possible that a firm that has not been taken over by the Spanish ministry of economy can become bankrupt.

In this research, each firm is described by 21 financial ratios (features) that have come from a detailed analysis of previous bankruptcy studies for non-life insurance. We have to pay particular attention to the fact that financial characteristics of insurance companies require general financial ratios and those that are specifically proposed for evaluating insurance sector insolvency. Table I shows the 21 ratios that describe the firms. Ratios R1, R2, R3, R4 and R9 are general financial ratios, and the rest are specific to the insurance sector.

An important variable in this sector is *reinsurance*. It is possible to select ratios by distinguishing between ratios involving *earned premiums* and ratios involving *earned premiums net of reinsurance*, and between claims incurred and claims incurred net of reinsurance. Thus, we can separate the following couples of ratios: R5–R6, R7–R8, R11–R12, R13–R14, R15–R16, R17–R18 and R20–R21

Table I. Definition of the ratios

Ratio	Definition
R1	$\frac{\text{Working Capital}}{\text{Total Assets}}$
R2	$\frac{\text{Earnings Before Taxes (EBT)}}{\text{Capital + Reserves}}$
R3	$\frac{\text{Investment Income}}{\text{Investments}}$
R4	$\frac{\text{EBT + Reserves for Depreciation + (Extraordinary Income} - \text{Extraordinary Charges)}}{\text{Total Liabilities}}$
R5	$\frac{\text{Earned Premiums}}{\text{Capital + Reserves}}$
R6	$\frac{\text{Earned Premiums Net of Reinsurance}}{\text{Capital + Reseves}}$
R7	$\frac{\text{Earned Premiums}}{\text{Capital + Reserves + Technical Provisions}}$
R8	$\frac{\text{Earned Premiums Net of Resinsurance}}{\text{Capital + Reserves + Technical Provisions}}$
R9	$\frac{\text{Capital + Reserves}}{\text{Total Liabilities}}$
R10	$\frac{\text{Technical Provisions}}{\text{Capital + Reserves}}$
R11	$\frac{\text{Claims Incurred}}{\text{Capital + Reserves}}$
R12	$\frac{\text{Claims Incurred Net of Reinsurance}}{\text{Capital + Reserves}}$
R13	$\frac{\text{Claims Incurred}}{\text{Capital + Reserves + Technical Provisions}}$
R14	$\frac{\text{Claims Incurred Net of Reinsurance}}{\text{Capital + Reserves + Technical Provisions}}$
R15	$\frac{\text{Claims Incurred}}{\text{Earned Premiums}} + \frac{\text{Other Charges and Commisions}}{\text{Other Income}}$
R16	$\frac{\text{Claims Incurred Net of Reinsurance}}{\text{Earned Premiums Net of Reinsurance}} + \frac{\text{Other Charges and Commissions}}{\text{Other Income}}$
R17	$\frac{\text{Claims Incurred + Other Charges and Commisions}}{\text{Earned Premiums}}$
R18	$\frac{\text{Claims Incurred Net of Reinsurance + Other Charges and Commissions}}{\text{Earned Premiums Net of Reinsurance}}$
R19	$\frac{\text{Technical Provisions of Assigned Reinsurance}}{\text{Technical Provisions}}$
R20	$\frac{\text{Claims Incurred}}{\text{Earned Premiums}}$
R21	$\frac{\text{Claims Incurred Net of Reinsurance}}{\text{Earned Premiums Net of Reinsurance}}$

(see Table I). The majority of firms in our study do not have reinsurance, so it is expected that the results for both sets of ratios will be similar since general financial ratios are common for both sets. Ratios 15 and 16 have been removed, due to most of the firms not having 'other income'. This reduce the number of ratios used to 19.

4.2. Experiments

Using the data defined above, we test the performance of the two approaches presented in this paper, by means of several experiments and comparisons. The performance of the SA with an SVM algorithm described in Section 3 is tested in the selection of the best set of features from the 19 initial ones. First, we choose the tunable SVM parameters C and γ (see Section 2.1) by means of a cross-validation scheme, following Bishop (1995). In this paper we maintain these parameters fixed in each simulation. Note that, since the number of features varies in each iteration of the SA and in the calculation of the Walsh spectrum, it would be desirable to recalculate C and γ values each time. However, the computational cost of this process is very high.

The set of 72 firms are split in to four sets, every set is formed by 18 firms (nine failed and nine non-failed). The cross-validation procedure consists of training the SVM with three of the four sets, and validating the result with the remaining set. This process is repeated for each of the four possible combinations. The final result is the average of the four results obtained. Figure 3 shows an example of this process. Once C and γ are fixed, the SA described in Section 3 is run, using the test error of the SVM as the objective function value in the algorithm. Note that a different subset of features, indicated by $\boldsymbol{\sigma}$, is used in each SA step.

In this paper we compare the features obtained by this method with the features achieved by a *rough set algorithm.* Rough set theory was first developed by Pawlak (1991) as a mathematical tool to deal with the uncertainty or vagueness inherent in a decision-making process. This methodology has provided very good results in insolvency problems (Slowinski and Zopounidis, 1995).

Briefly, the rough set approach works by discovering dependencies between attributes in an information table, and reducing the set of attributes by removing those that are not essential to

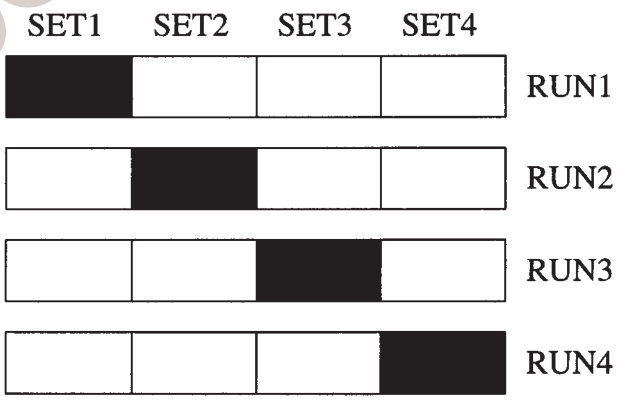


Figure 3. Schematic illustration of the cross-validation procedure. The neural network is trained four times, each time using a version of the data set in which one of the segments (shaded) is omitted. Each trained network is then tested on the data from the set which was omitted during training. The final result is the average over the four sets

characterize knowledge. A *reduct* is defined as the minimal subset of attributes that provides the same quality of classification as the set of all attributes. A reduced information table may provide decision rules of the form ‘if conditions then decisions’. These rules specify what decisions (actions) should be undertaken when some conditions are satisfied, and can be used to assign new objects to a decision class by matching the condition part of one of the decision rules to the description of the object. We have performed the rough set analysis using the rough set system ROSE. For a more detailed description of the rough set theory and the ROSE software, see Pawlak (1991) and Predki *et al*. (1998; Predki and Wilk, 1999).

The performance of the ranking provided by the Walsh analysis (spectrum $\mathbf{B}'$) is tested as follows. The function involved in the calculation of $\mathbf{B}'$ is given by the probability of error of an SVM for all possible binary strings $\boldsymbol{\sigma}$. In order to perform a better testing of the ranking provided by means of the prime spectrum $\mathbf{B}'$, we consider two groups of 13 features each. In the first group we include ratios G_1 = [R1, R2, R3, R4, R5, R7, R9, R10, R11, R13, R17, R19, R20]. In the second group we include ratios G_2 = [R1, R2, R3, R4, R6, R8, R9, R10, R12, R14, R18, R19, R21]. The difference between the groups is that G_2 contains the general ratios and the specific ones expressed net of reinsurance (see Table I). Recall that due to most of the firms considered not having reinsurance, both groups of ratios are similar, and therefore it is expected that the results obtained for both sets of ratios, G_1 and G_2 should be similar as well.

4.3. Results

Feature Selection Through SA and SVMs

The best sets of features obtained with the SA algorithm are formed by ratios {R1, R9, R13}, and {R3, R9, R19}, both with a probability of error in test P_e = 0.233. The SA algorithm reached one of these sets in all simulations run. No other combination of features provides a better value of probability of error.

The reducts obtained using a rough set approach for all training sets contain from four to six attributes (ratios). The ratios that have the highest frequency of occurrence (more than 40%) in reducts are R1, R3, R4, R9, R17, R18 and R19. Note that there are several coincidences among variables provided by the SA with an SVM and the rough set approaches. This indicates that these variables are highly discriminatory between solvent and insolvent firms in our sample. Note also that the reducts obtained using the SA algorithm contain fewer ratios than the reducts provided by the rough set approach, but give the same quality as using the whole set of ratios. These results show that the SA algorithm provides a good FS, comparable with those obtained by existing methods, such as the rough set algorithm.

Ranking of Features

It is important to note that FS methods in general provide very specific results, without information about the possible relationships among features. These relationships, which may be very important in some applications, can be obtained by performing a ranking of features. A simple and commonly used method to tackle the ranking of variables is the so-called *Fisher ranking*, where features are linearly classified following the value of a *Fisher score*. In this section we compare the proposed alternative ranking given by the prime spectrum with the Fisher ranking.

Figure 4 shows the ranking obtained by means of the prime spectrum $\mathbf{B}'$ for sets of ratios G_1 (a) and G_2 (b). Recall that these two sets of ratios have similar properties due to most of the firms not have reinsurance, so the rankings in both sets should also be similar. It is easy to see that the ranking

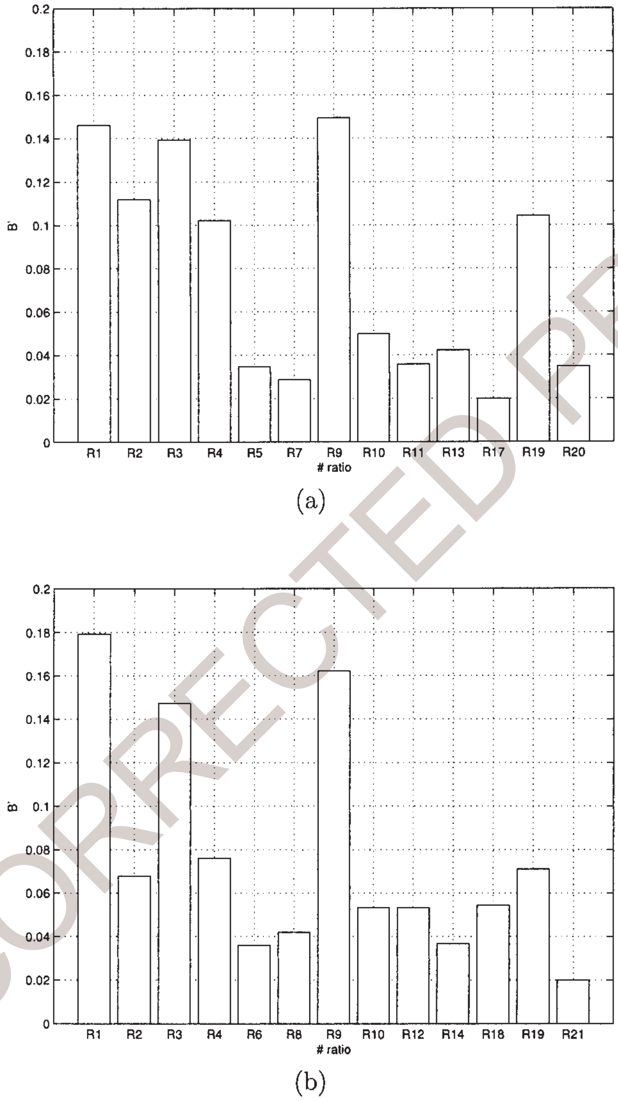


Figure 4. (a) Ranking provided by the prime spectrum **B′** using ratios in set G_1; (b) ranking provided by the prime spectrum **B′** using ratios in set G_2

B′ for the set of features G_1 highlights ratios R1, R9, R3 and R19, and also R2 and R4 with a slightly lower value of **B′**. On the other hand, the anking for set G_2 highlights ratios R1, R3 and R9. R19 and R4 are the following ratios in the ranking, but with a smaller value of **B′**.

Figure 5 shows a Fisher ranking for the same sets G_1 (a) and G_2 (b). This ranking gives much importance to ratio R1, whereas other ratios, such as R3 or R19 for example, are completely ignored. This indicates that our ranking algorithm performs better than the Fisher ranking for this problem.

Figure 6 shows the test error probability when the SVM classifier is trained with all the samples and an increasing number of features, sorted according to the ranking obtained with the Walsh

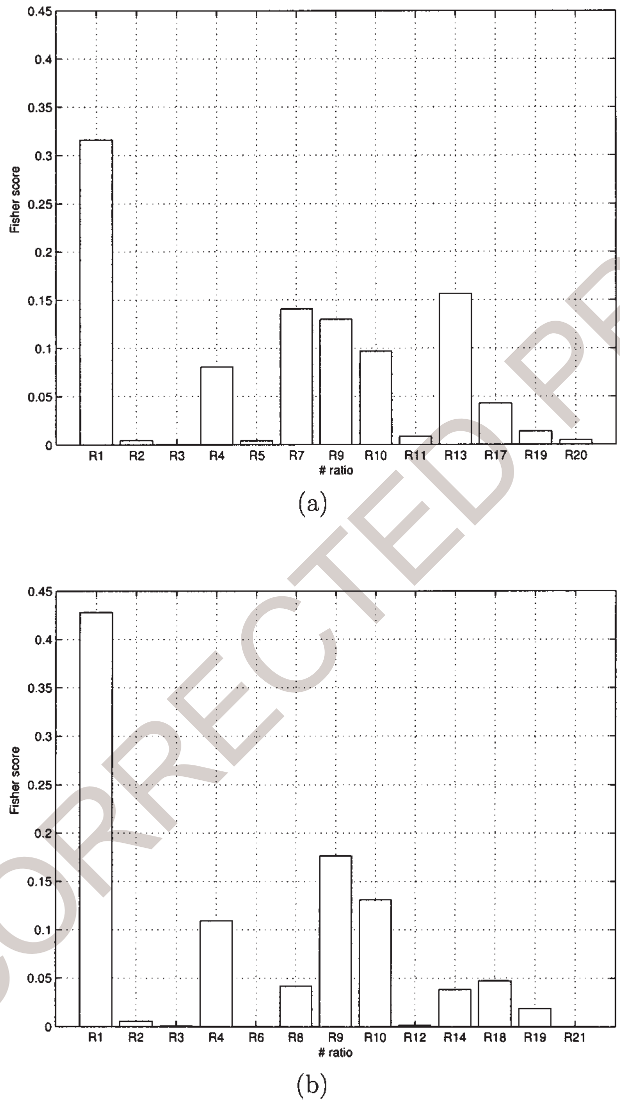
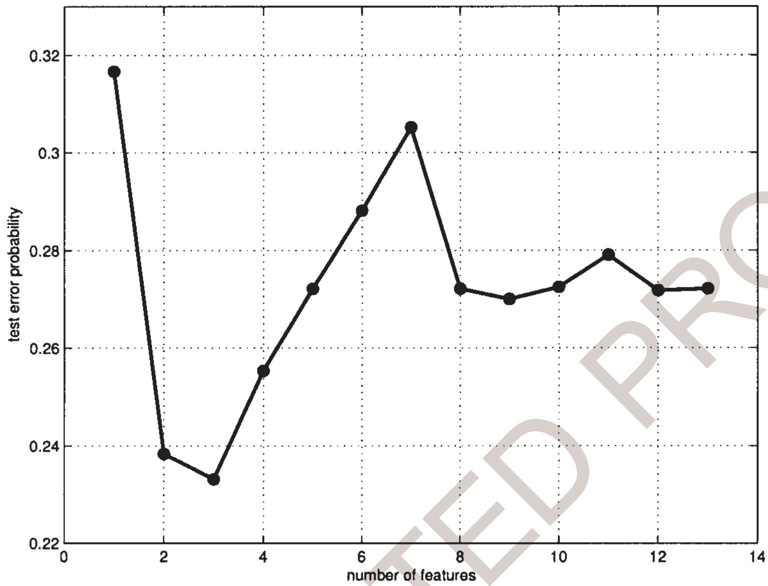


Figure 5. (a) Ranking provided by the Fisher score using ratios in set G_1; (b) ranking provided by the Fisher score using ratios in set G_2

analysis for set G_1 (see Figure 4(a)). Each point has been obtained by averaging 200 classification error trials. This way, the first point in the plot corresponds to the mean classification error when the most representative feature is chosen, and the last point is the mean test error using all the features in the problem. The best feature, provides a 5% decrease in accuracy. This is due to the dimension reduction, i.e. from a 13-dimensional space to a single-dimension space. However, by using two and three features the performance of the classifier improves accuracy compared with the full set of features case. Note that the best result, an error probability of 0.233, is obtained with just three ratios, which is consistent with the results provided by the SA.



(a)

Figure 6. Test error probability of set G_1 provided by an SVM classifier according to the number of features. Features are sorted according to the ranking provided by the Walsh method

4.4. Results Analysis and Discussion

The results obtained show the importance of the feature selection procedure used. In our sample, despite 19 financial attributes being available, we need consider just three characteristics in order to check to solvency of a firm (solution provided by the SA algorithm). We have found two sets of three ratios each (five different ratios, since R1 is common to both sets), which seem sufficient to analyse the solvency of a non-life insurance company {R1, R3, R9, R13, R19}.

The analysis of the results obtained by the ranking **B′** applied to sets G_1 and G_2 provides very interesting conclusions. First, since the majority of firms in our sample do not have reinsurance, the ranking of features provided by **B′** is similar in both sets G_1 and G_2. Second, the higher values of **B′** are for ratios R1, R9, R3 and R19 (order of importance). Note the coincidence with the ratios provided by the SA algorithm. Note also that the SA algorithm only chooses these variables, but **B′** in addition ranks them in importance. R1 is the most important, then R9, then R3 and finally R19.

A brief resumé of the financial meaning of these ratios is as follows:

- R1. One of the most important questions in order to assure the proper functioning of any firm is the need to have sufficient liquidity. But, in the case of an insurance firm, the lack of liquidity should not arise due to 'productive activity inversion', which implies that premiums are paid before claims occur. If an insurance firm cannot pay the claims incurred, then the clients and public in general could lose faith in that company. On the other hand, this ratio is a measure of financial equilibrium. If it is positive, then implies that working capital is also positive.
- R3. This ratio indicates that obtaining enough income is a critical issue because, nowadays, income is the main source of benefit for an insurance company.

- R9. This ratio shows what proportion of the total liabilities represent, the shareholders' funds (capital and reserves). This confirms the importance, from a solvency viewpoint, of the adequacy of these funds. These resources could be called on to meet the future claims obligations of the insurer.
- R19. This ratio shows the importance of a proper reinsurance in evaluating the solvency of insurance firms.

5. CONCLUSIONS AND FURTHER RESEARCH

In this paper we have presented two FS methods based on SVMs, and we have applied them to the prediction of insolvency in non-life insurance companies. We have chosen SVM-based methods because an SVM is considered a fast and robust classifier capable of obtaining accurate classifications in high-dimensional problems with very few examples.

First, we presented an SA algorithm for FS. Second, we considered an algorithm for ranking features using the prime spectrum obtained from a Walsh analysis of a function that involves an SVM. The SA algorithm presented in this work can be useful in problems where the number of features is large and an exhaustive analysis of all combinations of features is infeasible. In problems where the number of features allows an exact calculation or a reasonable estimation of the SVM. Walsh analysis, the prime spectrum gives a powerful tool for analysing relations among features, ranking them in importance for the problem.

The FS algorithms presented have been used in order to estimate which financial ratios are the most useful in the prediction of the insolvency of Spanish non-life insurance companies. We have obtained good results, with the SVM discarding noisy and redundant ratios.

In practical terms, the models generated can be used to preselect companies to be examined more thoroughly. They can also be used to check and monitor insurance firms as a 'warning system' for insurance regulators, investors, management, financial analysts, banks, auditors, policy holders and consumers.

This work opens new lines of research, such as the application of the SVM to the estimation of missing data or ratios. This would be especially useful in cases where it is costly to obtain new data, as in most economics applications.

APPENDIX

In this Appendix we briefly describe the method of Hordijk and Stadler (1998) for estimating the spectrum **B** associated with a binary fitness function *f*.

The estimated autocorrelation $\hat{r}(s)$ of a function from sampling points $\{f_t\}$ along a random walk over the search space can be defined as

$$\hat{r}(s) = \frac{\sum_{t=1}^{T-s} (f_t - \bar{f})(f_{t+s} - \bar{f})}{\sum_{t=1}^{T} (f_t - \bar{f})^2} \quad \text{(A.1)}$$

where

$$\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$$

This autocorrelation measure is known to be related to the energy coefficients by (Hordijk and Stadler, 1998)

$$r(s) = \sum_{p \neq 0} B_p \left(1 - \frac{2p}{l}\right)^s \quad \text{(A.2)}$$

which gives the following system of linear equations to calculate the estimated spectrum:

$$\hat{r}(0) = \hat{B}_1 + \hat{B}_2 + \ldots + \hat{B}_l$$

$$\hat{r}(1) = \left(1 - \frac{2}{l}\right)\hat{B}_1 + \left(1 - \frac{4}{l}\right)\hat{B}_2 + \cdots + (1 - 2)\hat{B}_l \quad \text{(A.3)}$$

$$\vdots$$

$$\hat{r}(k) = \left(1 - \frac{2}{l}\right)^k \hat{B}_1 + \left(1 - \frac{4}{l}\right)^k \hat{B}_2 + \cdots + (1 - 2)^k \hat{B}_l \quad \text{(A.4)}$$

This system can be solved using a steepest descent algorithm that iteratively minimizes the sum of squared errors in the above equations, provided that constraints $B_p \geq 0$ are imposed (Hordijk and Stadler, 1998).

ACKNOWLEDGEMENTS

We would like to thank the valuable comments and hints of the anonymous referees, which have contributed to the quality of this paper.

REFERENCES

Altman EI, Marco G, Varetto F. 1994. Corporate distress diagnosis: comparisons using discriminant analysis and neural networks (the italian experience). *Journal of Banking and Finance* **18**: 505–529. 2

Ambrose JM, Carol AM. 1994. Using best ratings in life insurer insolvency prediction. *Journal of Risk and Insurance* **61**: 317–327. 3

Bannister J. 1997. *Insurance Solvency Analysis*, second edition. LLP.

Barniv R. 1990. Accounting procedures, market data, cash-flow figures and insolvency classification: the case of the insurance industry. *The Accounting Review* **65**(3): 578–604.

Bishop CM. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press.

Bradley PS, Fayyad UM, Mangasarian OL. 1999. Mathematical programming for data mining: formulations and challenges. *INFORMS Journal on Computing* **11**(3): 217–238.

Burges CJ. 1998. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining* **2**(2): 121–167.

Chen HY, Chen TC, Min D, Fischer G, Wu YM. 1999. Prediction of tacrolimus blood levels by using the neural network with genetic algorithm in liver transplantation patients. *Therapeutic Drug Monitoring* **21**(1): 50–56.

4 Dimitras AI, Zanakis SH, Zopounidis C. 1996. A survey of businers failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research* **90**(3): 487–513.

Dimitras AI, Slowinski R, Susmaga R, Zopounidis C. 1998. Business failure using rough set. *European Journal of Operational Research* **114**(2): 263–280.

Goldberg DE. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley: Reading, MA.

González J, Rojas I, Pomares H, Salmerón M, Merelo JJ. 2002. Web newspaper layout optimization using simulated annealing. *IEEE Transactions on Systems, Man and Cybernetics* **32**(5): 686–691.

Hordijk W, Stadler PF. 1998. Amplitude spectra of fitness landscapes. *Advances in Complex Systems* **1**(1): 39–66.

John G, Kohavi R, Pfleger K. 1994. Irrelevant features and the subset selection problem, In *Proceedings of the 11th International Conference on Machine Learning*, August. Morgan Kaufmann: 121–129.

Kargupta H. 1999. A striking property of genetic code-like transformations. Technical Report EECS-99-04, EECS School, Washington State University, Washington.

Kirpatrick S, Gerlatt CD, Vecchi MP. 1983. Optimization by simulated annealing. *Science* **220**: 671–680.

5 Kohavi R, John GH. 1997. Wrappers for features subset selection. *International Journal of Digital Libraries* **1**: 108–121.

Li J, Tsang EPK. 2000. Reducing failures in investments recommendations using genetic programming. In *Proceedings of the 6th International Conference on Computing in Economics*, Barcelona, July.

Müller Group. 1997. Müller Group Report. Solvency of insurance undertakings. *In Conference of Insurance Supervisory Authorities of the Member States of the European Union*.

O'Leary DE. 1998. Using neural networks to predict corporate fealure. *International Journal of Intelligent Systems in Accounting Finance and Management* **7**: 187–197.

6 Patuwo E, Wu MY, Hung MS. 1993. Two group classification using neural networks. *Decision Sciences* **23**: 899–916.

Pawlak Z. 1991. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer: London.

7 Predki B, Wilk S. 1999. Rough set based data exploration using ROSE system. In *Foundations of Intelligent Systems*. Lecture Notes in Artificial Intelligence, vol. 1609. Springer-Verlag: 172–180.

8 Predki B, Slowinski R, Stefanowski J, Susmaga R, Wilk S. 1998. ROSE: software implementation of the rough set theory. In *Rough Sets and Current Trends in Computing*. Lecture Notes in Artificial Intelligence, vol. 1424. Springer-Verlag: 605–608.

Salcedo-Sanz S, DePrado-Cumplido M, Pérez-Cruz F, Bousoño-Calzón C. 2002. Feature selection via genetic optimization. In *Proceedings of the International Conference on Artificial Neural Networks, ICANN2002*, Madrid, Spain, August, 547–552.

9 Salcedo-Sanz S, Fernández-Villacañas JL, Segovia-Vargas MJ, Bousoño-Calzón C. 2003. Genetic programming for the prediction of insolvency in non-life insurance companies. *Computers & Operations Research* in press.

10 Sanchis A, Gil JA, Heras A. 2003. El análisis discriminante en la previsión de la insolvencia en la empresa de seguros no-vida. *Revista Española de Financiación y Contabilidad* **116**.

11 Segovia-Vargas MJ, Salcedo-Sanz S, Bousoño-Calzón C. 2003. Prediction of insolvency in non-life insurance companies using support vector machines and genetic algorithms. In *Proceedings, X SIGEF Congress in Emergent Solutions for the Information and Knowledge Economy*, León, Spain, October.

12 Serrano-Cinca C. 1996. Self organizing neural networks for financial diagnosis. *Decision Support Systems* **17**: 227–238.

Schölkopf B, Smola AJ. 2002. *Learning with Kernels*. MIT Press: Cambridge, MA.

Slowinski R, Zopounidis Z. 1995. Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management* **4**: 27–41.

Tam KY. 1991. Neural network models and the prediction of bankruptcy. *Omega* **19**(5): 429–445.

13 Tam KY, Kiang MY. 1992. Managerial applications of neural networks: the case of bank failure predictions. *Management Science* **38**(7): 926–947.

14 Vafaie H, De Jong KA. 1992. Genetic algorithms as a tool for features selection in machine learning. In *Proceedings of the 4th International Conference on Tools with Artificial Systems*. IEEE Computer Society Press. Arlington, VA; 200–204.

Vapnik VN. 1998. *Statistical Learning Theory*. John Wiley: New York.

Vose MD, Wright AH. 1998. The simple genetic algorithm and the Walsh transform: part I, theory. *Evolutionary Computation* **6**(3): 253–273.

Wang G, Ansari, N. 1997. Optimal broadcast scheduling in packet radio networks using mean field annealing. *IEEE J, Selected Areas in Communications* **15**(2): 250–259.

Weston H, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. 2000. Feature selection for SVMs. In 15
Advances in NIPS 12. MIT Press: 526–532.

Wilson RL, Sharda R. 1994. Bankruptcy prediction using neural networks. *Decision Support Systems* **11**: 545–557. 16

Zopounidis C. 1999. Multicriteria decision aid in financial management. *European Journal of Operational Research* **119**(2): 404–415. 17

Zopounidis C, Dimitras A. 1998. *Multicriteria Decision Aid Methods for the Prediction of Business Failure*. Kluwer. 18

From: Production Department
John Wiley & Sons Ltd
The Atrium, Southern Gate
Chichester
West Sussex
PO19 8SQ

Queries Form Publisher

Title of Journal: International Journal of Intelligent Systems in Accounting, Finance & Management

Paper Ref. No.: ISAF_255

During the copyediting of your article the following queries have arisen. Please mark your corrections and answers to these queries directly on to the proof at the relevant place – do not mark your corrections on this query sheet.

Many thanks for your assistance.

Query Refs.	**Query**	**Remarks**
1 *(On p. 8)*	You mean 'and' 'that' or '. . . on, which'	
2 *(On p. 19)*	not cited in text	
3 *(On p. 19)*	not cited in text	
4 *(On p. 20)*	not cited in text	
5 *(On p. 20)*	Confirm/clarify journal title	
6 *(On p. 20)*	not cited in text	
7 *(On p. 20)*	missing eds?	

Query Refs.	**Query**	**Remarks**
8 *(On p. 20)*	missing eds?	
9 *(On p. 20)*	update?	
10 *(On p. 20)*	Is this a Journal? pp?	
11 *(On p. 20)*	not cited in text	
12 *(On p. 20)*	title change OK?	
13 *(On p. 20)*	not cited in text	
14 *(On p. 20)*	not cited in text	
15 *(On p. 21)*	Confirm book title is as set. Or missing contribute title?	
16 *(On p. 21)*	not cited in text	
17 *(On p. 21)*	not cited in text	
18 *(On p. 21)*	not cited in text	