PREDICTION OF INSOLVENCY IN NON-LIFE INSURANCE COMPANIES USING SUPPORT VECTOR MACHINES, GENETIC ALGORITHMS AND SIMULATED ANNEALING

María Jesús Segovia-Vargas [a], Sancho Salcedo-Sanz [b] and Carlos Bousoño-Calzón [b]

[a] *Department of Financial Economy and Accounting I, Universidad Complutense de Madrid (mjsegovia@ccee.ucm.es)*

[b] *Department of Signal Theory and Communications, Universidad Carlos III de Madrid (sancho,cbousono@tsc.uc3m.es)*

Abstract

In this paper we propose an approach to predict insolvency of non-life insurance companies based on the applications of Support Vector Machines (SVMs), hybridized with two global search heuristics: a Genetic Algorithm (GA) and a Simulated Annealing (SA). A SVM is used to classify firms as failed or non-failed, whereas a GA and a SA are used to perform on-line feature selection in the ratios space of the SVM, in order to improve its performance. We use general financial ratios and also other specific ratios which have been proposed for evaluating insolvency of insurance sector. In the simulations section, we compare the performance of the GA and SA as part of the proposed algorithm. The results obtained with both techniques show that the proposed algorithm can be a useful tool for parties interested in evaluating insolvency of non-life insurance firms.

Keywords: Insolvency, Non-life insurance companies, Support Vector Machines, Genetic Algorithms, Simulated Annealing.

1 INTRODUCTION

Business failure risk assessment is one of the key issues when controlling the insurance market sector, due to the high amount of money to pay by the state insurance guaranty funds when an insurance company has gone bankrupt. Also protecting the 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 the public confidence in all the insurance companies. These facts explain the increasing interest in accurately predicting insurance companies failures.

In this paper we focus on controlling the solvency of non-life insurance companies, a pattern recognition problem in which we have to decide whether a non-life insurance company is solvent or insolvent, or to predict if it would be insolvent in the years to come, from a given set of inputs. Unlike other financial problems, there are a great number of agents (auditors, management, analysts, creditors, government) facing business failure, so research in this topic has been of growing interest in the last decades, see works by [Tam et al. (1992)] and [Dimitras et al. (1998)]. Following [Zopounidis et al. (1998)], the number of bankrupt firms and the relative trend indicate the robustness for the economy of a country in a worldwide scale, and the substantial costs associated to insolvency can become a national political issue. On the other hand, insolvency of non-life insurance companies have been a concern of parties such as insurance regulators, investors, management, financial analysts, banks, auditors, policy holders and consumers. This concern has arisen from the necessity of protecting the general public against the consequences of insurers insolvencies, as well as minimizing the responsibilities for management and auditors.

A large number of methods have been proposed to predict business failure, see works by [Ambrose et al. (1994)], [Barniv(1990)] and [Sanchis et al. (2003)]; however, the special characteristics of the insurance sector have made most of them unfeasible, and just a few have been applied to this sector [Salcedo et al. (2003)]. Most approaches to prediction of failure in non-life insurance companies are statistical methods, such as discriminant or logic analysis, which use financial ratios as explicative variables. However, this kind of variables does not usually satisfy any statistical assumptions, what introduces an extra difficulty for the use of statistical methods in insurance business failure prediction.

In this paper, we propose an approach to predict insolvency of non-life insurance companies based on a Support Vector Machine (SVM) [Burges (1998)] for classifying companies into solvent or insolvent, and on a Genetic Algorithm (GA) [Goldberg (1989)], and a Simulated Annealing (SA) [Kirpatrick (1983)], which perform on-line feature selection in the financial ratios space.

The SVM has been shown as a powerful method for classification problems, with very good properties of versatility [Burges (1998)]. The feature selection process, implemented through the GA or the SA heuristics, eliminates irrelevant or redundant ratios, for improving the SVM performance.

We will show that the proposed algorithm is a robust and powerful approach, which completes previous research in prediction of business failure, developing a prediction model for non-life insurance companies. The results obtained are very encouraging and show that the proposed approach can be a useful tool for parties interested in evaluating insolvency of a non-life insurance firm.

The rest of the paper is structured as follows: In Section 2 the main features of insolvency prediction in non-life insurance companies, introduced as a particular case of the *multi-attribute classification problem*, are given. Section 3 describes the algorithm proposed, providing a brief description of SVMs in classification problems and GA and SA applied to the feature selection problem. Section 4 includes the analysis of the test data used, the experiments performed in order to test the proposed algorithm and the results obtained. Finally, in Section 5 some concluding remarks can be found.

2   PROBLEM DEFINITION

In this paper we tackle the prediction of non-life insurance firms failure, which can be considered a particular example of the so called *multi-attribute classification problem*. This problem consists in the assignment of an *object*, described by values of *attributes*, to a predefined class of category.

Mathematically a multi-attribute classification problem can be stated as follows:

Let $\{\mathbf{x}_i\}$, $\mathbf{x}_i \in \mathbb{R}^n$, $i \in \{1, \ldots, l\}$ a set of observations (objects) drawn from some unknown probability distribution $P(\mathbf{x}, y)$, and $\{y_i\} \in \{-1, 1\}$ (categories), a set of associated true labels.

A classification machine is defined by a set of possible mappings $\mathbf{x} \mapsto f(\mathbf{x}, \alpha)$, where a particular choice of parameters α generates what is called a "trained machine". As example, in a general neural network with fixed architecture, α corresponds to the weights and biases of the neural network, in the case of GP α are the number of nodes, branches and functions in nodes of the final tree.

The expectation of the test error for a trained machine can be defined as:

3

$$R(\alpha) = \left(\int V(y, f(\mathbf{x}, \alpha)) dP(\mathbf{x}, y) \right) \tag{1}$$

where $V(\cdot, \cdot)$ is a loss functional, $P(\mathbf{x}, y)$ is the unknown probability function the data was sampled from and the function $y = f(\mathbf{x}, \alpha)$ is the classification engine.

For the multi-attribute classification problem the loss functional can be defined as:

$$V(y, f(\mathbf{x}, \alpha)) = \frac{1}{2}|y - f(\mathbf{x}, \alpha)| \tag{2}$$

and the expected test error for a trained machine yields:

$$R(\alpha) = \left(\int \frac{1}{2}|y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, y) \right) \tag{3}$$

An *empirical* test error can be defined as:

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(\mathbf{x}_i, \alpha)| \tag{4}$$

Note that $R_{emp}(\alpha)$ is a fixed number for a particular choice of α and for a particular training set $\{\mathbf{x}_i, y_i\}$, since no probability distribution is involved in this calculation.

In the general case, the multi-attribute classification problem consists in finding the machine which learn the mapping $\mathbf{x}_i \mapsto y_i$ with the highest generalization ability possible. According to the statistical learning theory, the generalization error of a learning machine can be analyzed considering the machine's capacity and its empirical risk [Bis95]. The capacity factor represents the machine's complexity, whereas the empirical risk factor measures its quality. To ensure high generalization ability, the tradeoff between these two factors should be addressed.

2.1   PREDICTION OF BUSINESS FAILURE

The multi-attribute classification problem is applicable in a straight forward manner to the business failure prediction problem, with $\{\mathbf{x}_i\}$, $i = 1, \ldots, l$, a set of firms described by a set of n financial ratios (every component x_{ij}), and

$y_i \in \{-1, 1\}$ a label which describes the state of the firm as "healthy" $y_i = 1$ or failed $y_i = -1$.

In addition, the number of ratios which describe the training set should be as low as possible, due to irrelevant and redundant attributes, used as part of the classification system, can make poorer its performance. Thus, the complete problem this paper faces consists in characterizing a firm as healthy or failed, as well as reducing the number of financial ratios involved in the classification process, by removing those that are not essential to characterize a firm.

3   THE PROPOSED ALGORITHM

The use of classification machines, mainly neural networks, in prediction of business failure is not a new trend. Works as the one by [O'leary (1998)], [Serrano (1996)] or [Wilson et al. (1994)] have shown that non-linear classifiers are a competitive and robust approach in this field. Thus, the application of SVMs to the prediction of insolvency in non-life insurance firms is a natural extension of previous works.

3.1   BRIEF INTRODUCTION TO SUPPORT VECTOR MACHINES

This section provides a brief summary of the standard SVM for classification [1] applied to business failure, starting from the simple linear SVM and moving on to the nonlinear SVM.

Consider a set of firms represented by the value of their ratios $\{\mathbf{x}_i\}$, $i = 1, \ldots, l$, and a set of associated labels $y_i \in \{-1, 1\}$ which describe the firm as failed or healthy. First imagine that this training set can be separated by a linear hyperplane. The Support Vector Machine solves the following problem:

Find $\mathbf{w} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, to minimize $\frac{1}{2}\|\mathbf{w}\|^2$, subject to:

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

Once such $\mathbf{w}$ and b are found, our classification rule for firms healthy or failed is given by $sign(\mathbf{w}^T \cdot \mathbf{x} + b)$, and the associated error to this classification is $R_{emp}(\mathbf{w}, b)$.

---

[1] A more complete analysis as well as further results about SVMs can be found in [Burges (1998)].

Consider now the case when the points in the training set are not linearly separable; then constraint (5) cannot be satisfied. We can introduce then some nonnegative slack variables ξ's in order to overcome this difficulty, the 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 $\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i$, under the constraints:

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \qquad \forall i = 1, \ldots, l \qquad (6)$$
$$\xi_i \geq 0 \qquad \forall i = 1, \ldots, l \qquad (7)$$

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

The nonlinear SVM maps the input variable into a high dimensional (often infinite dimensional) feature space, and applies the linear SVM in this feature space. Computationally, this can be achieved by the application of a kernel function. The nonlinear SVM with kernel K is equivalent to a regularization problem in the reproducing kernel Hilbert space H_K:

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

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \qquad (8)$$

subject to

$$y_i(\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i) + b) \geq 1 - \xi_i \qquad \forall i = 1, \ldots, l \qquad (9)$$
$$\xi_i \geq 0 \qquad \forall i = 1, \ldots, l \qquad (10)$$

The nonlinear SVM is able to classify any set of firms as healthy or failed, with a probability of error given by $R_{emp}(\mathbf{w}, \alpha)$. However, given a training set, the selection of the input variables (financial ratios) is an important issue to be considered, due to irrelevant or redundant ratios can affect in a negative way to the result given by the SVM. This is the so called *Feature Selection 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 in non-life insurance insolvency prediction.

In its more general form, FSP for a learning problem from samples can be addressed in the following way: 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 $y_i \in \{+1, -1\}$, choose a subset of m features $(m < n)$, that achieves the lowest classification error, see [Weston et al. (2000)] for details.

Following [Weston et al. (2000)], we will define the FSP as finding the optimum n-column vector $\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}_i \in \{1, 0\}$, that defines the subset of selected features, which is found as:

$$\boldsymbol{\sigma}^o = arg \ \min_{\boldsymbol{\sigma}, \alpha} \left(\int V(y, f(\mathbf{x} * \boldsymbol{\sigma}, \alpha)) dP(\mathbf{x}, y) \right) \simeq arg \ \min_{\boldsymbol{\sigma}, \mathbf{w}, b} R_{emp}(\mathbf{w}, \mathbf{b}, \boldsymbol{\sigma}) \quad (11)$$

where $V(\cdot, \cdot)$, $P(\mathbf{x}, y)$ where defined in (2) and we have defined $\mathbf{x} * \boldsymbol{\sigma} = (x_1 \sigma_1, \ldots, x_n \sigma_n)$.

The general FSP can be particularized for insolvency of non-life insurance companies prediction by considering that the components of vectors $\mathbf{x}_i \in \mathbb{R}^n$ are the financial ratios which define the state of a non-life insurance company, and vector $\boldsymbol{\sigma}^o$ defines the best set of ratios to be used by the SVM. Feature selection eliminates irrelevant and redundant ratios of every firm, what improves the performance of the SVM. The remaining ratios are used by the SVM for the classification process.

In this paper we propose solving the FSP by means of a Genetic Algorithm and a Simulated Annealing.

3.3   A GENETIC ALGORITHM FOR THE FSP

Genetic Algorithms (GAs) are a class of robust problem solving techniques based on a population of solutions (binary strings), called *individuals*, which evolves through successive generations by means of the application of the so called genetic operators: selection, crossover and mutation, see [Goldberg (1989)]. Selection is the process by which individuals in the population are randomly sampled with probabilities proportional to their fitness values. An elitist strategy, consisting in passing the highest fitness string to the next generation, is applied in order to preserve the best solution encountered so far in the evolution. The selected set, of the same size of the initial population, is subjected to the crossover operation. Firstly, the binary strings are coupled at random. Second, for each pair of strings, an integer position along the string is selected uniformly at random. Two new strings are composed by swapping all bits between the selected position and the end of the string. This operation is applied to the couples with probability P_c less than one. By means of the mutation

operation, every bit in every string of the population may be changed from 1 to 0, or vice versa, with a very small probability, P_m.

The population of the GA for the FSP is formed by a number ξ of binary strings $\boldsymbol{\sigma} \in \{0,1\}^n$, which evolves by the iterative procedure of the genetic operators described above. A component $\sigma_i = 1$ means that the correspondent ratio i has to be taken into account for the SVM, and if the component $\sigma_i = 0$ means that the correspondent ratio i has to be removed from the set of ratios. Note that every individual of the GA population (a binary vector $\boldsymbol{\sigma}$) stands for a different set of ratios to be used by the SVM. The fitness function associated the each individual is the classification error obtained classifying l training points $(\mathbf{x}*\boldsymbol{\sigma}, y)$, that can be estimated as $R_{emp}(\mathbf{w}, b, \boldsymbol{\sigma})$. Due to GAs maximizes the fitness function, and the objective function in the FSP is minimizing the error probability, a modified fitness function is introduced:

$$F = 100(1 - R_{emp}(\mathbf{w}, b, \boldsymbol{\sigma})) \tag{12}$$

3.4 A SIMULATED ANNEALING ALGORITHM FOR THE FSP

Simulated Annealing (SA) is a powerful solving technique which has been successfully applied to a wide variety of optimization problems [Kirpatrick (1983)], [González (2002)]. 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 to 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 current one, and if not, it may replace the current one probabilistically. This probability of replacement is high in the beginning of the algorithm, and decreases in every stage. This procedure allows the system to move toward the best configuration. However, SA is not guaranteed to find the global optima, it is better than others algorithms 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 a SA algorithm are: the chosen representation for solutions, the objective function to be minimized during the process and the mutation or configuration change operator.

Problem representation:

We encode every solution to the FSP as a binary string $\boldsymbol{\sigma}$, where $\sigma_i \in \{0, 1\}$ defines the subset of selected features. The length of binary vector $\boldsymbol{\sigma}$ will be equal to the total number of features in the problem.

Mutation Operator:

In this paper we consider a classical *Random Flip Mutation* operator, where N_f bits are randomly selected and flipped to obtain a configuration in the neighborhood of the current one.

Objective Function:

We use the probability of error in test 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.

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 $\rightarrow f(\boldsymbol{\sigma}) = P_e(test)$;

repeat

    for $j = 0$ **to** M

    $\boldsymbol{\sigma}^{mut} = \textbf{mutate}(\boldsymbol{\sigma})$;

    Run the SVM $\rightarrow f(\boldsymbol{\sigma}^{mut}) = P_e(test)$;

        if$((f(\boldsymbol{\sigma}^{mut}) < f(\boldsymbol{\sigma}))$ OR $(random(0, 1) < e^{(\frac{-a}{T})}))$ **then**

        $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{mut}$;

        endif

    endfor

  $T = f_T(T_0, k)$;

  $k = k + 1$;

until$(T < T_{min})$;

---

where k counts the number of iterations performed; T keeps 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, $\boldsymbol{\sigma}^{mut}$ stands for the new configuration after mutation operator is applied. $f(\boldsymbol{\sigma})$ represents the objective

function (probability of error in test 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 the initial acceptance probability to be 0.8, which is the value usually used. The freezer function is defined as

$$f_T = \frac{T_0}{1+k}. \tag{13}$$

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

$$T_{min} = f_T(T_0, numIt). \tag{14}$$

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

3.5   THE COMPLETE ALGORITHM

The complete algorithm for the prediction of non-life insurance companies consists in SVM for the classification process hybridized with a GA or a SA, which perform a search over the space of vectors $\boldsymbol{\sigma}$. Note that both global algorithms (GA and SA) might have associated any other type of classifier, such a neural network.

The algorithm starts by randomly choosing an initial population in the GA or SA:

If we use the GA as a global search heuristic, every binary vector representing a set of ratios (individual in the GA) is evaluated through the SVM, getting the test error as the fitness associated to the individual. This fitness value is used in the Selection process as was indicated in Section 3.3. The Crossover and Mutation operators of the GA are also performed as described in Section 3.3. This process is repeated until a given stop criterion (number of generations or convergence reached) is hold.

If we use the SA algorithm as global heuristic, there is only one binary vector representing the solution of the problem. A search in the neighborhood of the current solution is carried out as shown in Section 3.4. The SVM is launched in order to get the objective function associated to the new binary vector, and compared with the objective function of the current solution. If it is better, the new solution is accepted as the current one, and if it is worse, it is accepted with a probability depending on the temperature of the system. The

SA algorithm is also stopped with a criterion based of a maximum number of generations reached.

4 EXPERIMENTS

4.1 TEST DATA AND INPUT VARIABLES

In this Section, we show the main characteristics of the data and variables that will be used to test our algorithm. We have used the sample of firms used by [Sanchis et al. (2003)] and [Salcedo et al. (2003)]. This data sample consists of Spanish non-life insurance firms data five 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 industry and size (premiums volume). In our study we have used data one year prior to the firms declare bankruptcy, due to this, it has to be noted that the prediction of the insolvency achieved by our method will be one year in advance.

In order to test the predictive accuracy of the SVM, it is necessary to split the set of original data to form a set to train the SVM, i.e. the training set, and a holdout sample to validate the obtained model, i.e. the testing set. Due to the decision of splitting the original 72 firms could affect the results of the evaluation and validation, this study utilizes a cross-validation procedure, as the one described by [Bishop (1995)], in order to generate multiple subsets from the original firms. The set of 72 firms are split in four sets, every set consists of 18 firms (9 failed and 9 non-failed). The cross-validation procedure consists in 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 elections of the set which is omitted from the training process. The final result, which will be used as fitness value in the GA, is the average of the four obtained results. Figure 1 illustrates the cross-validation procedure used.

The selection of initial variables for bankruptcy prediction is a critical issue. These variables could influence the quality of the obtained results. In this research, each firm is described by 21 financial ratios that have come from a detailed analysis of the variables and previous bankruptcy studies for non-life insurance.
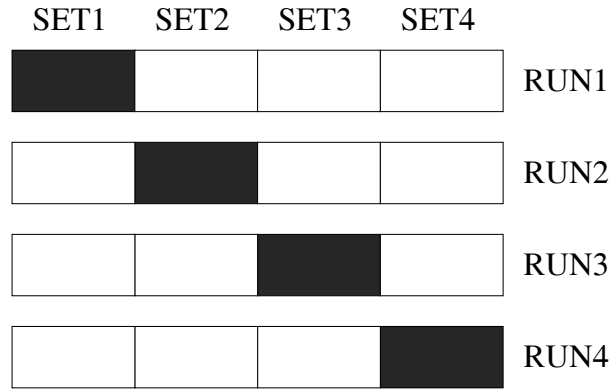
Figure 1. Schematic illustration of the cross-validation procedure. The SVM is trained four times, each time using a version of the data set in which one of the segments (shaded) is omitted. Each trained network is then tested on the data from the set which was omitted during training. The final result is the average value obtained over the four sets.

Table 1 shows the 21 ratios which describe the firms. Ratios 15 and 16 have been removed in our study, due to most of the firms have not "other income"; this reduce the total number of ratios (and therefore the length of the individuals in the GA) to 19. Note that the special financial characteristics of insurance companies require general financial ratios as well as those that are specially proposed for evaluating insolvency of insurance sector. The ratios have been calculated from the last financial statements (balance sheets and income statements) issued before the firms declared bankruptcy.

4.2 RESULTS

We have tested our algorithm in the data introduced above. First, the SVM was run using the training data without feature selection, i.e. using 19 ratios in every training point. The average probability of error using the cross-validation procedure was $P_e = 0.33$ in this case. After that, we ran the complete algorithm using the GA and SA global heuristics. The GA parameters was fixed to the standard ones $P_c = 0.6$ and $P_m = 0.01$ (see [Goldberg (1989)]), with a population of $\xi = 30$ individuals. Whereas the SA parameters are fixed to $M = 50$ and $numIt = 300$.

The solution with 19 ratios ($P_e = 0.33$), is poorer than the best solution found by the GA involving few number of features. Note that the best solution achieved by the GA involves 3-features, R1, R9 and R13, with an average P_e in the cross validation procedure of 0.23.

The best evolution of the GA is shown in Figure 2. Note that the number of

Table 1
Definition of the Ratios

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

generations needed for convergence is small, about 20, what shows that the algorithm's computational cost is low.

The best sets of features obtained with the SA algorithm are formed by ratios {R1, R9 , R13}, and {R3, R9, R19} both of them with a probability of error in test $P_e = 0.23$. The SA algorithm reached one of this sets in all simulations run. No other combinations of features provided a better value of probability of error. Again, the probability of error provided by the solutions found by the GA is better than the probability of error achieved with the complete set of 19 ratios.
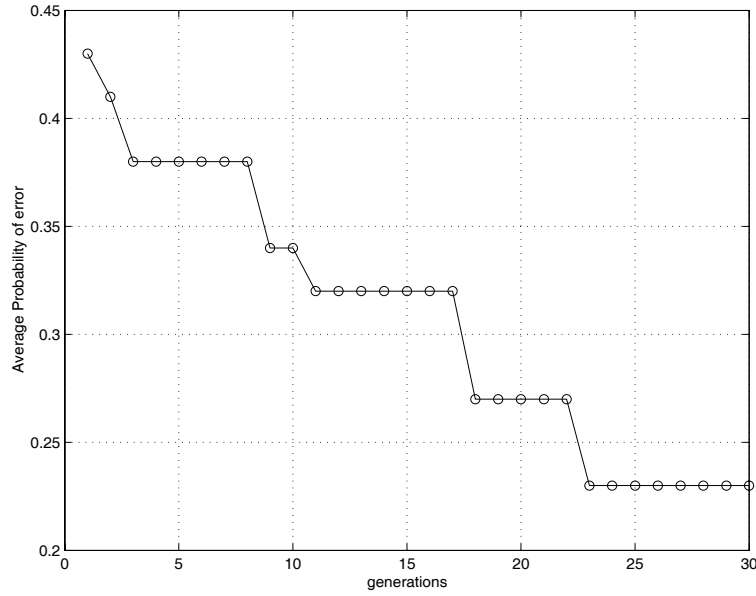
Figure 2. Evolution of the Genetic Algorithm in the run when the best solution was obtained.

4.3 ANALYSIS OF THE RESULTS

The obtained results show the importance of the feature selection procedure used: in spite of the initial large information system of 19 financial attributes, we can consider just three characteristics in order to check the solvency of a firm. Both global search heuristics, the GA and the SA obtain the same set of three ratios, which are essential to analyze the solvency of a non-life insurance company:

- R1- One of the most important questions in order to assure the proper functioning of any firm is the need of having sufficient liquidity. However, in the case of a non-life insurance firm, the lack of liquidity should not arise, due to premiums are paid in before claims occur. If an insurance firm cannot pay the incurred claims, the clients and general public could lose confidence in that company. On the other hand, this ratio is a measure of financial equilibrium: if it is positive it implies that the working capital is also positive.
- R9- This ratio shows what proportion of the total liabilities represents the shareholders' funds (capital and reserves). It confirms the importance, from a solvency viewpoint, of the adequacy of the mentioned funds, due to these resources could be required to meet the future claims obligations of the insurer in some eventualities.
- R13- This rate is one of the ratios considered as *solvency ratios in strict sense*. The numerator shows the risk exposure through incurred claims. The denominator shows the real financial support because technical provisions

14

are considered together with capital and reserves. This demonstrates the need of having sufficient shareholders' funds and the need of complying correctly with the technical provisions to guarantee the financial viability of the insurance company.

The SA algorithm achieved another set of three ratios which provides the same probability of error as the previous ones. This set includes ratio R9, described above, and ratios:

- R3- This ratio indicates that to obtain enough financial incomes is a critical issue because nowadays these incomes are the main source of benefit for an insurance company.
- R19- This ratio shows the importance of a proper reinsurance to evaluate the solvency in insurance firms.

Summarizing, the algorithm proposed in this paper provides a very good performance on prediction of business failure in non-life insurance companies, both with GA and SA. It obtains very good results in terms of probability of error in the classification procedure, with very few ratios.

5   CONCLUSIONS

In this paper, an algorithm for prediction of insolvency in non-life insurance companies has been presented. The approach consists of a Support Vector Machine (SVM) which classifies a firm as healthy or failed, depending on the value of a set of financial ratios which characterize every firm. The SVM is hybridized with a Genetic Algorithm (GA) and a Simulated Annealing (SA) in order to perform on-line feature selection in the space of financial ratios. The resulting approach allows a very accurate classification of firms into healthy or failed, based on very few financial ratios. This improves the generalization performance of the SVM while reduces cost of training the classifier. In addition using our algorithm it is not necessary that the data fulfil any statistical assumptions or preprocessing them, what is an important advantage with respect to statistical methods.

We have tested our approach in a real problem of prediction of insolvency of Spanish non-life insurance companies, formed by 72 firms, described by 21 financial ratios. Our approach, using GA and SA search algorithms, has achieved very good results, obtaining the lowest probability of error with only 3 ratios, what means that, in spite of the initial large information system with of 21 financial attributes, we can consider just three characteristics in order to check the solvency of a firm.

References

[Altman et al. (1994)] Altman E. I., Marco, G. and Varetto, F. (1994): "Corporate distress diagnosis: comparisons using discriminant analysis and neural networks (the italian experience)", *Jounal of Banking and Finance*, 18:505-529.

[Ambrose et al. (1994)] Ambrose, J. M., Carol, A. M. (1994): "Using best ratings in life insurer insolvency prediction", *Journal of Risk and Insurance*, 61:317-327.

[Bannister (1997)] Bannister, J. (1997): *Insurance solvency analysis*, LLP limited, second edition.

[Barniv(1990)] 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 (1995)] Bishop, C. M. (1995): *Neural Networks for pattern recognition*, Oxford University press.

[Burges (1998)] Burges, C. J. (1998): "A tutorial on Support Vector Machines for pattern recognition", *Knowledge Discovery and Data Mining*, 2(2):121-167.

[Dimitras et al. (1996)] Dimitras, A. I., Zanakis, S. H., and 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 et al. (1998)] Dimitras, A. I., Slowinski, R., Susmaga, R. and Zopounidis, C. (1998): "Business failure using rough set", *European Journal of Operational Research*, 114(2):263-280.

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

[González (2002)] González, J., Rojas, I., Pomares, H., Salmerón, M. and Merelo, J. J. 2002. Web newspaper layout optimization using Simulated Annealing, *IEEE Trans. Systems, Man and Cybernetics*, 32(5):686-691.

[Kirpatrick (1983)] Kirpatrick, S., Gerlatt, C. D. and Vecchi, M. P. 1983. Optimization by simulated annealing, *Science*, 220:671-680.

[O'leary (1998)] O'Leary, D. E. (1998): "Using neural networks to predict corporate fealure", *International Jounal of Intelligent Systems in Accounting Finance and Management*, 7:187-197.

[Patuwo et al. (1993)] Patuwo, E., Wu, M. Y., and Hung, M. S. (1993): "Two group classification using neural networks", *Decision Sciences*, 23:899-916.

[Salcedo et al. (2003)] Salcedo-Sanz, S., Fernández-Villacañas, J. L., Segovia-Vargas, M. J. and Bousoño-Calzón, C. 2002. Genetic programming for the prediction of insolvency in non-life insurance companies, *Computers & Operations Research*, in press.

[Sanchis et al. (2003)] Sanchis, A., Gil, J. A., and 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* 115.

[Serrano (1996)] Serrano-Cinca, C. (1996): "Shelf organizing neural networks for financial diagnosis", *Decision Support Systems*, 17:227-238.

[Tam (1991)] Tam, K. Y. (1991): "Neural network models and the prediction of bankruptcy", *Omega*, 19(5):429-445.

[Tam et al. (1992)] Tam, K. Y., and Kiang, M. Y. (1992): "Managerial applications of neural networks: the case of bank failure predictions", *Management Science*, 38(7):926-947.

[Vafaie et al. (1992)] Vafaie, H., and De Jong, K. A. (1992): "Genetic Algorithms as a Tool for Features Selection in Machine Learning". *Proc. of the 4th Intl. Conf. on Tools with Artificial Systems*, IEEE computer society press, Arlintong, VA, 200-204.

[Weston et al. (2000)] Weston, H., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. (2000): "Feature Selection for SVMs", Advances in NIPS 12, MIT Press, 526-532.

[Wilson et al. (1994)] Wilson, R. L., and Sharda, R. (1994): "Bankruptcy prediction using neural networks", *Decision Support Systems*, 11:545-557.

[Zopounidis et al. (1998)] Zopounidis. C., and Dimitras, A. (1998): *Multicriteria decision aid methods for the prediction of business failure*, Kluwer.

[Zopounidis (1999)] Zopounidis, C. (1999): "Multicriteria decision aid in financial management", *European Journal of Operational Research*, 119(2):404-415.