Feature Selection for Bankruptcy Prediction: A Multi-Objective Optimization Approach

F. Mendes*, J. Duarte**, A. Vieira**, A. Gaspar-Cunha*

* IPC/I3N - Institute of Polymers and Composites, University of Minho, Guimarães, Portugal, e-mail: {fmendes,agc}@dep.uminho.pt}

** Department of Physics, Instituto Superior de Engenharia do Porto, R. S. Tomé, 4200 Porto, Portugal, e-mail: {jmmd,asv}@isep.ipp.pt

Abstract In this work a Multi-Objective Evolutionary Algorithm (MOEA) was applied for feature selection in the problem of bankruptcy prediction. The aim is to maximize the accuracy of the classifier while keeping the number of features low. A two-objective problem - minimization of the number of features and accuracy maximization – is fully analyzed using two classifiers: Support Vector Machines and Logistic Function. A database containing financial statements of 1200 medium sized private French companies was used. It was shown that MOEA is a very efficient feature selection approach. Furthermore, it can provide very useful information for the decision maker in characterizing the financial health of a company.

1 Introduction

Financial bankruptcy prediction is of high importance for banks, insurance companies, creditors and investors. Different approaches have been used in this type of prediction like linear discriminant analysis and non-linear models. However, these methods have some limitations. Discriminant analysis is limited due to its linearity, restrictive assumptions, for treating financial ratios as independent variables and can only be used with continuous independent variables. In non-linear models the choice of the regression function creates a bias that restricts the outcome, they are very sensitive to exceptions, and most conclusions have an implicit Gaussian distribution on data, which is inappropriate in many cases [1, 2].

More recently other approaches have been applied for bankruptcy classification, such as Artificial Neural Networks (ANN) [3], Evolutionary Algorithms (EA) and Support Vector Machines (SVM) [4]. ANN, EA and SVM are used as complementary tools to classify credit risk. Some of the studies performed show that ANN outperforms discriminant analysis in bankruptcy prediction [5-7]. These

models have shown promising results. However it is generally recognized that further research is needed to achieve higher predictive capabilities, which is the avenue of our research.

Since these methods can handle a large number of variables, and some of them are highly correlated, it is of crucial importance to have a feature selection algorithm to reduce the number of features considered for analysis. In this work a methodology based on MOEA to accomplish simultaneously two objectives: the minimization of the number of features used and the maximization of the accuracy of the classifier used is proposed. The use of MOEA for solving this problem is not new. Merelo et al. {8] applied a MOEA to take into account individually the errors of type I (false positive) and type II (false negative). Hamdani et al. [9] used the NSGA-II [10] algorithm to optimize simultaneously the number of features and the global error obtained by a neural network classifier.

A database containing financial statements of 1200 small and medium size private French companies during 2006 was been used. Each company is characterized by a set of 30 features measuring its most important ratios, from profitability to debt.

This text is organized as follows. In section 2 the problem to solve will be explained in more detail as well the classification methods used here (RG and SVM). Then, in section 3 the MOEA used will be described in more detail. The method proposed will be applied to a case study and the results will be presented and discussed in section 4. Finally, the conclusion will be stated in section 5.

2 Bankruptcy prediction

Bankruptcy prediction is a problem stated as follows: given a set of financial statements from a company over one or several previous years, predict the changes that it will become distressed over a given period, normally the next year or two years ahead.

The first step consists in "cleaning up" the database in order to create a well balanced and unbiased sample. The dataset must be simplified in order to the problem be more understandable by the decision maker. This can be accomplished by reducing the number of features necessary. On the other hand it must be assured that this reduction does not decrease the performance of the classifier.

The methodology proposed in this work uses the data classifiers to measure the accuracy of the data (or the features used), while a MOEA will be used to determine the best compromise between the refereed two objectives. To test the ability of the method proposed to accomplish these objectives two classifiers will be used: Logistic Regression (LR) and Support Vector Machines (SVM).

LR is a well known generalized linear method, allowing the prediction of a discrete outcome (generally binary, such as success/failure), from a set of variables that may be continuous, discrete, binary, or a mix of any of these [11]. In the present case the LR was trained by Stochastic Gradient Decent, which is able to esti-

mates the maximum likelihood logistic regression coefficients from sparse input data.

Support Vector Machines (SVMs) are a set of supervised learning methods based on the use of a kernel, which can be applied to classification and regression. In the SVM a hyperplane or set of hyperplanes is (are) constructed in a high-dimensional space. In this case, a good separation is achieved by the hyperplane that has the largest distance to the nearest training datapoints of any class. Thus, the generalization error of the classifier is lower when this margin is larger. SVMs can be seen an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik [12].

3 Multi-Objective Optimization

MOEAs have been recognized in the last decade as good methods to explore and find an approximation to the Pareto-optimal front for multi-objective optimization problems. This is due to the difficulty of traditional exact methods to solve this type of problems and by their capacity to explore and combine various solutions to find the Pareto front in a single run. A MOEA must provide a homogeneous distribution of the population along the Pareto frontier, together with an improvement of the solutions along successive generations [13, 14]. In this work, the Reduced Pareto Set Genetic Algorithm (RPSGA) is adopted [15, 16], where a clustering technique is applied to reduce the number of solutions on the efficient frontier. Detailed information about this algorithm can be found elsewhere [15, 16].

In the present study the RPSGA algorithm was adapted to deal with the features selection problem. First it was necessary to identify clearly which are the decision variables, and then the RPSGA was to be linked with the classification algorithms in order to obtain the information about the performance of the classifier.

In what concerns the definition of the decision variables, two different possibilities were considered. Initially, a problem where only the features are to be chosen was studied. In this situation the parameters of the classifiers, such as type of training (holdout or k-fold cross validation), learning rate and training fraction, for both LR and SVM, kernel type and other parameter only for SVM, must be previously defined. In a second approach, these parameter where also included as variables to be optimized. The advantage of the latter approach consists in obtaining in a single run the best results. The methodology proposed will be better illustrated with the example presented in the next section.

4 Results and discussion

4.1 Case studies

Let us now illustrate the use of above MOEA methodology to solve a feature extraction problem. The problem to be addressed is a classification problem. The aim is to find the minimum number of features needed to obtain the maximum accuracy of the companies' evaluation. Accuracy is defined as the number of companies correctly predicted divided by the total number of companies in the sample.

Table 1 shows the features definitions considered in the database. Regarding the company evaluation, the classifier output predicts whether the company has survived or gone into bankruptcy.

In the case of LR, several runs were performed using the gradient descent method and various combinations of the training method (holdout and 10-fold validation), the learning rate (0.001, 0.01, 0.02 and 0.1) and the training fraction (0.5, 0.7 and 0.7). For the case of holdout training method the data used for training is the same for all runs, depending on the training fraction used. One run was performed considering these three variables as decision variable ranging in the same interval.

Similarly, for the case of SVMs two different kernels were tested (μ-SVC and C-SVC). Different combinations of training method, learning rate, training fraction and other kernel parameters were tested. The runs performed used the better RPSGA parameters as defined in reference [15]. The comparison between the performances of the different runs was made using the attainment functions methodology [17].

4.2 Influence of classification methods parameters

The results presented in Figures 1, 2 and 3 were obtained using LR with the training method, the learning rate and the training fraction as decision variables. As shown in Figure 1, after 100 generations, the evolution leads to a considerable gain in accuracy while decreasing significantly the number of features needed. On the final population 8 non-dominated solutions exist having, respectively 1, 2, 3, 4, 7, 8, 9 and 10 features, which are identified in Figure 2. Figure 3 shows the values obtained for the training method, the learning rate and the training fraction for the same non-dominated solutions.

Identical results were obtained for the SVM classifier. More results can be seen on www.dep.uminho.pt/agc/results. Finally, the attainment functions methodology

was used to compare the performance of the methodology proposed when using different parameters of the classifiers (see the same web site for more details).

Table 1. Set of features considered in the problem to solve

| Feature	Designation
F1	Number of employees
F2	Capital Employed / Fixed Assets
F3	Financ. Debt / Capital Employed (\%)
F4	Depreciation of Tangible Assets (\%)
F5	Working capital / current assets
F6	Current ratio
F7	Liquidity ratio
F8	Stock Turnover days
F9	Collection period
F10	Credit Period
F11	Turnover per Employee (thousands euros)
F12	Interest / Turnover
F13	Debt Period days
F14	Financial Debt / Equity (\%)
F15	Financial Debt / Cashflow
F16	Cashflow / Turnover (\%)
F17	Working Capital / Turnover (days)
F18	Net Current Assets/Turnover (days)
F19	Working Capital Needs / Turnover (\%)
F20	Export (\%)
F21	Value added per employee
F22	Total Assets / Turnover
F23	Operating Profit Margin (\%)
F24	Net Profit Margin (\%)
F25	Added Value Margin (\%)
F26	Part of Employees (\%)
F27	Return on Capital Employed (\%)
F28	Return on Total Assets (\%)
F29	EBIT Margin (\%)
F30	EBITDA Margin (\%)
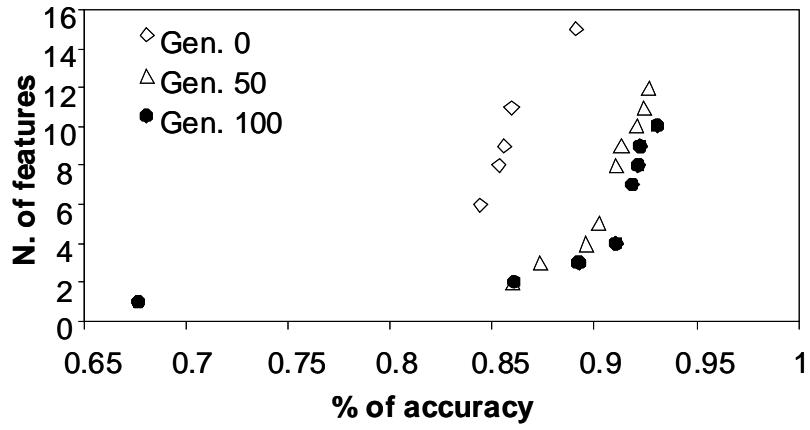
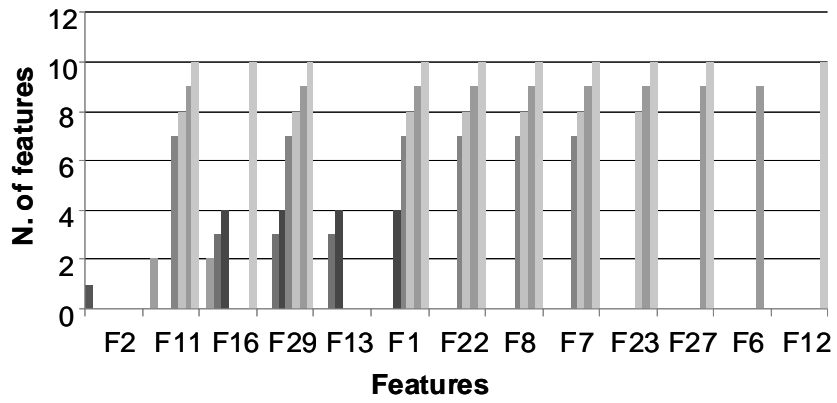**Fig. 1.** Pareto fronts for the initial population and after 50 and 100 generations.

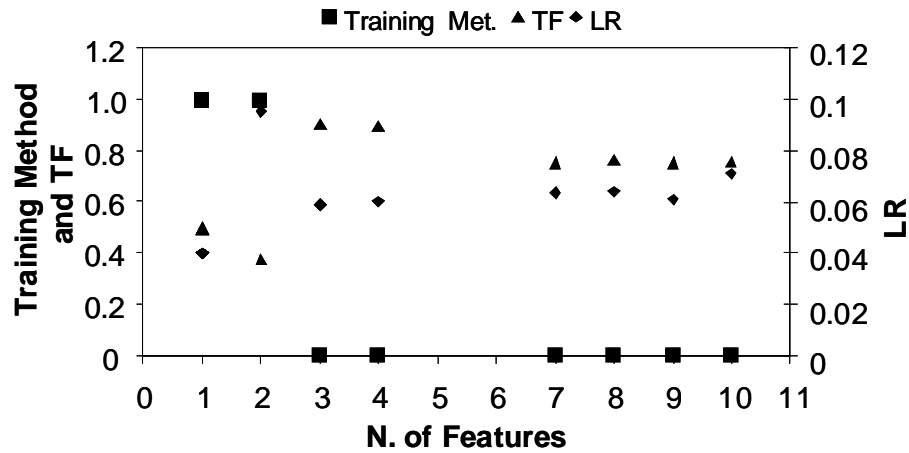**Fig. 2.** Features obtained for the non-dominated solutions after 100 generations.

Fig. 3. Training method, learning rate and training fraction values for the 8 non-dominated solutions.

5 Conclusions

In the present study the LR and SVM methods with different parameters values have been used. The results obtained allow one to conclude that the proposed methodology is able to solve this problem and, simultaneously, made available information to the decision maker. The information provided concerns not only with the best features to be used but, also, with the best parameters of the classifier.

References

1. E. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23, pp. 589-609, 1968.
2. S. Nanda, P. Pendharkar, Linear models for minimizing misclassification costs in bankruptcy prediction, Intelligent Systems in Accounting, Finance & Management, 10, pp. 155 – 168, 2001.
3. A. Charitou, E. Neophytou, C. Charalambous, Predicting Corporate Failure: Empirical Evidence for the UK. European Accounting Review, 13, pp. 465-497, 2004.
4. A.F. Atiya, Bankruptcy prediction for credit risk using neural networks A survey and new results, IEEE Transactions on Neural Networks, 12, pp. 929-935, 2001.
5. P.K. Coats, L.F. Fant, Recognizing Financial Distress Patterns Using a Neural Network Tool. Financial Management, 22, pp. 142-155, 1993.
6. Z.R. Yang, Probabilistic Neural Networks in bankruptcy prediction, Journal of Business Research, 44, pp. 67–75, 1999.

8

7. C. Tan, H. Dihardjo, A study on using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model, Managerial Finance, 27(4), pp. 56–77, 2001.

8. E. Alfaro-Cid, P. A. Castillo, A. Esparcia, K. Sharman, J.J. Merelo, A. Prieto, A.M. Mora, J.L.J. Laredo, Comparing Multiobjective Evolutionary Ensembles for Minimizing Type I and II Errors for Bankruptcy Prediction, 2008 Congress on Evolutionary Computation (CEC'2008), pp. 2907-2913, 2008.

9. T.M. Hamdani, J.-M. Won, A.M. Alimi, F. Karray, Multi-objective Feature Selection with NSGA II, In B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro, editors, Adaptive and Natural Computing Algorithms, 8th International Conference, ICANNGA 2007, Part I, Springer-Verlag. Lecture Notes in Computer Science Vol. 4431,pp. 240-247, 2007.

10. K.Deb, A. Pratap. S. Agarwal, T. Meyarivan, A fast and elitist multi-objective genetic algorithm: NSGA-II, IEEE Transaction on Evolutionary Computation, 6(2), pp. 181-197, 2002.

11. A. Agresti, Categorical Data Analysis. New York: Wiley-Interscience, 2002.

12. V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.

13. A. Gaspar-Cunha, P. Oliveira, J.A. Covas, Use of Genetic Algorithms in Multicriteria Optimization to Solve Industrial Problems, Seventh Int. Conf. on Genetic Algorithms, Michigan, USA, 1997.

14. K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, Wiley, 2001.

15. A. Gaspar-Cunha, J.A. Covas, J.A. - RPSGAe - A Multiobjective Genetic Algorithm with Elitism: Application to Polymer Extrusion. In X. Gandibleux, M. Sevaux, K. Sörensen and V. T'kindt, editors, *Metaheuristics for Multiobjective Optimisation*, volume 535 in Lecture Notes in Computer Science, 221-249, Springer Verlag, Berlin, 2004.

16. A. Gaspar-Cunha, Modelling and Optimization of Single Screw Extrusion, PhD Thesis, University of Minho, Guimarães, Portugal, 2000 (Downloadable from website http://www.lania.mx/~ccoello/EMOO/).

17. V. G. da Fonseca, C. Fonseca, A. Hall, Inferential performance assessment of stochastic optimisers and the attainment function. In E. Ziztler, K. Deb, L. Thiele, C.A. Coello Coello, and D. Corne, editors, Evolutionary Multi-Criterion Optimization (EMO 2001), Lectures Notes in Computer Science 1993, 213-225. Springer Verlag, 2001.