Using Support Vector Machine approach for forecasting the failure of the Tunisian companies *

Yousri HENCHIRI and Salwa BENAMMOU and Taher HAMZA

Abstract In this study we look to analyze, the financial determinants of failure and the ability of predictive models to anticipate the occurrence of this risk. This gait grants a preventive character through the establishment of precocious alert system rather than that of making-decision. The background in this field suggested various conceptions tools for the prediction of failure, such as discriminate analysis, logistic regression, artificial neural network, genetic algorithms... As part of this research, we propose to test the predictive capacity of the SVM model concerning distressed companies. This model, that is a learning algorithm class, initially conceived for discrimination, has been applied on two samples of small and medium Tunisian companies: a training sample and a test sample. The grid-search technique, using the cross-validation, is used to find out the best parameter value of SVM's kernel function. Finally, we empirically show that the SVM model gives high accuracy rate than other approaches being tested in this domain.

Key words: Failure, Prediction model, financial determinants, Support Vector Machine, Grid-search

1 Introduction

The analysis of the determinants of failure shows that these latter's are of various natures: macroeconomic, strategic and organizational and finally financial. Besides, we noticed that most studies adopt a multi-referential methodology enrolling in a double optical: the one

* Publié in the Book of Short Papers, 7th Scientific Meeting of CLAssification and Data Analysis Group (CLADAG) of the Italian Statistical Society, Catania 2009.

Yousri HENCHIRI and Taher HAMZA
Institut Supérieur de gestion, Université de Sousse, Tunisia., e-mail: henchiriyousri@yahoo.fr and e-mail: taher.hamza@topnet.tn

Salwa BENAMMOU
Faculté de Droit & des Sciences Economiques & Politiques, Université de Sousse, Tunisia., e-mail: saloua.benammou@fdseps.rnu.tn

is that of optimization forecasting financial distress approach and the other is that of conceptualization. In this study we look to analyze, the financial determinants of failure and the ability of predictive models to anticipate the occurrence of this risk. The background in this field suggested various conceptions tools for the prediction of failure, such as multiple discriminate analysis (MDA) *[3]-[5]*, logistic regression (LOGIT) *[9]-[5]*, artificial neural network (ANN) *[3]-[5]*, decision tree *[4]-[11]*, genetic algorithms *[11]-[7]*. Recently, a learning algorithm class, initially conceived for discrimination under the name of, Support Vector Machine (SVM) has been also introduced *[8]-[6]-[10]*. We look to apply this forecasting approach of the risk of the financial distress and provide a new tool improving its prediction accuracy. Developed by *Vapnik (1998)*, SVM method is gaining popularity thanks to many attractive features and excellent generalization performance on a wide range of problems. In addition, since the optimal parameters search of SVM plays a crucial role in building a distress prediction model with high accuracy and stability; we apply 5-fold cross validation and a grid-search technique in order to identify the correct value parameter of the kernel function of SVM. We use LIBSVM software *[2]* to conduct SVM experiment.

2 Research methodology and analysis results

TThe database used in this study, is composed of 120 firms *[see [5], were the sample is composed of two groups of firms : The first is the small and medium Tunisian companies (SMTC) such in a state of failure and the second group is sane SMTC]*. The research methodology goes around 5 phases: a selection of discriminatory 28 financial ratios (inspired by previous researches). over a period of two successive years, a test of Equality of Group means ratios between the two companies populations (sane and distressed), developing a matrix of correlations between ratios, using the principal component analysis (PCA) to reduce the size of each feature and to sequentially optimize a choice more limited of ratios and finally a SVM process is applied in an optical forecasting financial distress to 1 and 2 years of the event.following this process, we did a set of experiments with different kernel functions such as the linear, RBF, polynomial, and sigmoïd in order to see the quality of generalization for each kernel function. We apply a grid-search to find the best pair (C, gamma) for each kernel function using 5-fold cross-validation. In order to increase efficiency, we try exponentially growing sequences of (C, gamma) to identify good parameters $(C = 2^{-5}, 2^{-3}, \ldots, 2^{15}; gamma = 2^{-15}, 2^{-12}, \ldots, 2^{12})$.
After the optimal (C, gamma) is found, the whole training data is trained using the SVMs with different kernels and the best parameters to generate the final models.
Table 1 compares prediction performance of the SVM models using four different kernel functions. We note, in case of the RBF kernel, the prediction accuracy of tested data is turned out to be 95%, while that of the training data is 100%. As shown in *Table 1*, in terms of recognition accuracy, the RBF kernel gets the best result (100%) followed by the polynomial kernel (100% when d = 2), the linear kernel (97.5%) and finally the Sigmoïd kernel (80%). The highest prediction accuracy (97.5%) obtained by the linear kernel and the polynomial kernel followed by the RBF kernel (95%) and the sigmoïd kernel (92.5%).
Table 1 also shows that the underfitting or the overfitting problem has occurred in case of the polynomial kernel function. This is a typical overfitting phenomenon in machine learn-

Table 1 Performance of SVM Kernel on each optimal (C, γ).

Kernel Function	*C*	*Gamma*	*d*	*Prediction accuracy (%)*	
				Training data	*Test data*
Linear	2^{-5}	2^{6}	N/A	97.5% (78/80)	97.5% (39/40)
RBF	2^{9}	2^{-9}	N/A	100% (80/80)	95% (38/40)
Polynomial[a]	2^{11}	2^{-12}	1	98.75% (79/80)	97.5% (39/40)
			2	100% (80/80)	97.5% (39/40)
			3	100% (80/80)	97.5% (39/40)
			4	100% (80/80)	95% (38/40)
Sigmoïde[a]	2^{-1}	2^{-6}	N/A	80% (64/80)	92.5% (37/40)

[a] Parameter ”*r*” in set to 0.

ing. Therefore, we need to make extra efforts to find the best value of the degree ”d” in the polynomial kernel SVM model. As shown in *Table 1*, the RBF kernel achieves the highest prediction accuracy better than that in other kernels. In addition, overfitting is unlikely to occur with the RBF kernel function.

In the same perspective followed in the analysis of the data 1-year before the failure that we have processed the data 2-year the failure. *Table 2* compares prediction performance of the SVM models using four different kernel functions. In case of the RBF kernel, the prediction accuracy of tested data is turned out to be, 92.5% while that of the training data is 98.75%. As shown in *Table 2*, the RBF kernel gets the best recognition accuracy (98.75%) followed by the polynomial kernel (98.75% when d = 2), the linear kernel (97.5%) and finally the Sigmoïd kernel (95%). Also, the linear kernel and the Sigmoïd kernel obtained the best prediction accuracy of tested data (97.5%), followed by the polynomial kernel (95%) and the RBF kernel (92.5%). The obtained results are relatively satisfactory for a

Table 2 Performance of SVM Kernel on each optimal (C, γ).

Kernel Function	*C*	*Gamma*	*d*	*Prediction accuracy (%)*	
				Training data	*Test data*
Linear	2^{3}	2^{12}	N/A	97.5% (78/80)	97.5% (39/40)
RBF	2^{11}	2^{-12}	N/A	98.75% (79/80)	92.5% (37/40)
Polynomial[a]	2^{11}	2^{-12}	1	98.75% (79/80)	95% (38/40)
			2	100% (80/80)	90% (36/40)
			3	100% (80/80)	90 % (36/40)
			4	100% (80/80)	90% (36/40)
Sigmoïde[a]	2^{7}	2^{-6}	N/A	95% (76/80)	97.5% (39/40)

[a] Parameter ”*r*” in set to 0.

near horizon (1-year) for which recognition accuracy varies respectively from 80% and 100% according to SVM model applied respectively on a training sample. Also, the forecasting accuracy varies respectively from 92.5% and 97.5% according to SVM model applied respectively on a test sample. These results are better via the rates that are found with MDA and LOGIT model. On the basis of the same sample, we compare the results of the SVM model of our study with those obtained by the ANN, MDA and LOGIT models from the study of *[5]*. *Table 3* also summarizes the prediction performance of SVM approach, ANN, MDA and LOGIT model. As shown in *Table 3*, SVM and ANN are slightly outperforms LOGIT model and MDA for tested data.

Table 3 The best prediction accuracy of SVM, ANN, LOGIT model and MDA (%).

	Training data			Test data		
	Data (T-1)	Data (T-2)	*Average*	Data (T-1)	Data (T-2)	*Average*
SVM	100%	98,5%	*99.25%*	95%	92,5%	*93.75%*
ANN	95%	100%	*97.5%*	95%	92.5%	*93.75%*
LOGIT	93.75%	95%	*93.175%*	92,5%	80%	*86.25%*
MDA	93.75%	96.75%	*95%*	72.5%	62.5%	*66.875%*

Table 3, compares the best performance of SVM, ANN, MDA and LOGIT in the training data and tested data and show a superiority of the SVM method in relation to ANN and to the traditional approaches parametric. In relation to the remoteness of the forecasting horizon, all models tested in our study experienced know, an average, improvement in the phase and deterioration in their external validation on a test sample. We can conclude, nevertheless, that the discrimination algorithms are a good preacher of the risk of distress to a close horizon to the failure, especially if the forecasting is made on the horizon of 1-year. Another particularity of our results is the performance of the SVM for kernel RBF rather than other functions kernel to a close horizon to the failure.

3 Conclusion

The application of this technique to two samples of firms (sane and distressed) permitted us to obtain significant results and to propose a forecasting model more appropriate in terms of quality of generalization. The explanation for these results is related to causes rather than economic statistics since the SVM modeling does not require restrictive assumptions than those for traditional approaches. Concerning the results we noticed a significant difference from firms distressed and sane in terms of the business activity, the needs cleared by their exploitation, in debts rate, profitability, financial imbalance, liquidity and solvency. These reports are consistent with 1 and 2 years before the failure. Finally, the anticipation of the risk of financial distress of firms with SVM enabled us to achieve significant results and propose a suitable prediction model with a predictive capacity between 92.5% and 97.5% a 1-year before the failure.

References

1. Vapnik, V.N.: Statistical learning theory. New York: Springer (1998)
2. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A practical guide to support vector classification. Tech. rep., Department of Computer Science and Information Engineering, National Taiwan University(2008)

3. Altman, E.I., Marco, G., Varetto, F.: Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking and Finance. Appl. **18(3)**, 505–529 (1994)
4. Frydman, H.E., Altman, E.I., Kao, D.: Introducing recursive partitioning for financial classification: The case of financial distress. The Journal of Finance. Appl. **40(1)**, 269–291 (1985)
5. Hamza, T., Baghdadi, K.: Profil et déterminants financiére de la défaillance des P.M.E Tunisiennes (1999-2003). Banque et Marchés. Appl. **93(Mars-Avril)**, 45–62 (2008)
6. Hua, Z.-S., Wang, Y., Xu, X.-Y., al.: Predicting corporate financial distress based on integration of support vector machine and logistic regression. Expert Systems. Appl. **33**, 434–440 (2007)
7. Liang, L., Wu, D.-S.: An application of pattern recognition on scoring Chinese corporations financial conditions based on back propagation neural network. Computers & Operations Research. Appl. **32(5)**, 1115–1129 (2005)
8. Min, J.-H., Lee, Y.-C.: Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems. Appl. **28**, 603–614 (2005)
9. Mossman, C.E., Bell, G.G., Swartz, L.M., Turtle, H.: An empirical comparison of bankruptcy model. Financial Review. Appl. **33**, 35–54 (1998)
10. Sun, J., Lin, H.: Data mining method for listed companies financial distress prediction. Journal of Knowledge-Based systems. Appl. **21**, 1–5 (2008)
11. Varetto, F.: Genetic Algorithms applications in the analysis of insolvency risk. Journal of Banking and Finance. Appl. **22(10-11)**, 1421–1439 (1998)