Dissimilarity-based Linear Models for Corporate Bankruptcy Prediction

**Vicente García · Ana I. Marqués ·
J. Salvador Sánchez · Humberto J.
Ochoa-Domínguez**

Abstract Bankruptcy prediction has acquired great relevance for financial institutions due to the complexity of global economies and the growing number of corporate failures, especially since the world financial crisis of 2008. In this paper, the problem of corporate bankruptcy prediction is faced by means of four linear classifiers (Fisher's linear discriminant, linear discriminant classifier, support vector machine and logistic regression), which are designed on the dissimilarity space instead of the classical feature space. Experimental results indicate that the prediction methods implemented with the dissimilarity representation perform considerably better than the same techniques when

Vicente García
División Multidisciplinaria en Ciudad Universitaria
Universidad Autónoma de Ciudad Juárez
32310 Ciudad Juárez, Chihuahua (Mexico)
E-mail: vicente.jimenez@uacj.mx

Ana I. Marqués
Dep. Business Administration and Marketing
Universitat Jaume I
12071 Castelló de la Plana (Spain)
E-mail: imarques@uji.es

J. Salvador Sánchez
Institute of New Imaging Technologies
Dep. Computer Languages and Systems
Universitat Jaume I
12071 Castelló de la Plana (Spain)
E-mail: sanchez@uji.es

Humberto J. Ochoa-Domínguez
Dept. Electrical and Computer Engineering
Universidad Autónoma de Ciudad Juárez
32310 Ciudad Juárez, Chihuahua (Mexico)
E-mail: hochoa@uacj.mx

applied onto the feature space, in terms of overall accuracy, true-positive rate and true-negative rate.

1 Introduction

In brief, bankruptcy refers to financial failure of a corporate or an individual, which not only leads to significant costs to shareholders and creditors but also may result in a considerable macroeconomic impact (Altman, 1993; Zopounidis and Dimitras, 1998). In order to avoid the financial losses associated with the failure, financial analysts have long seen the need for the early discovery of bankruptcy. This is the main reason why bankruptcy prediction is deemed as a subject of key relevance for financial institutions. As a consequence, improving the performance of existing techniques and building highly effective models have attracted the attention of many researchers and practitioners (Aziz and Dar, 2006).

A vast amount of techniques have been developed to help decision-makers and analysts in predicting financial failure. The most traditional approaches have been based on statistical and operational research methods (Balcaen and Ooghe, 2006), such as factor analysis (West, 1985), linear and multivariate discriminant analysis (Altman et al, 1977; Karels and Prakash, 1987), logit analysis (Ohlson, 1980; Jones and Hensher, 2004; Tseng and Lin, 2005), probit analysis (Zmijewski, 1984), linear and quadratic programming (Kwak et al, 2012), and data envelopment analysis (Cielen et al, 2004; Premachandra et al, 2009).

After the Basel II recommendations issued by the Basel Committee on Banking Supervision in 2004, financial institutions realized the need of using more complex systems based upon computational intelligence techniques. Unlike the statistical models, these methods do not assume any specific prior knowledge, but automatically extract information from past observations. Kumar and Ravi (2007) reported a comprehensive review of statistical and computational intelligence methods in the context of bankruptcy prediction. Among some other techniques, support vector machines (Shin et al, 2005; Min and Lee, 2005; Erdogan, 2013), genetic and evolutionary algorithms (Lensberg et al, 2006; Acosta-González and Fernández-Rodríguez, 2014), artificial neural networks (Wilson and Sharda, 1994; Sun and Shenoy, 2007; Cleofas-Sánchez et al, 2016; Zhao et al, 2016), rough sets (Slowinski and Zopounidis, 1995; Mckee, 2000), and hybrid and classifier ensembles (Verikas et al, 2010; Fedorova et al, 2013; Abellán and Mantas, 2014; Tsai, 2014) have received much attention. Many works have empirically compared and contrasted these soft computing methods (Alfaro et al, 2008; Chen, 2012; Olson et al, 2012; Erdal and Ekinci, 2013; Tsai et al, 2014).

All these statistical and computational intelligence techniques applied in the field of bankruptcy prediction are based on the assumption that samples

are represented by a set of features (explanatory variables), which defines a feature space. These features usually correspond to financial ratios and/or macroeconomics indicators, either represented as continuous variables or discretized in a straightforward manner as qualitative information. However, in a few cases the samples are described by means of qualitative variables whose values are gathered from expert judgments (Kim and Han, 2003).

Apart from the feature space, there exist other approaches to pattern representation that could also be exploited for very distinct financial applications. One is the dissimilarity representation, in which samples to be classified/predicted are derived from pairwise dissimilarities (distances from other samples in the data set) (Pękalska and Duin, 2002). The justification for constructing classifiers in a dissimilarity space is that a dissimilarity measure should be small for similar samples and large for distinct samples, thus allowing for efficient and more reliable discrimination of classes. Another important characteristic is that the dimensions of a dissimilarity space symbolize homogeneous types of information and therefore, all dimensions can be considered as equally relevant. On the other hand, for a complex problem, a simple linear prediction model in a dissimilarity space could separate the classes more easily than the same classifier in a feature space (Pękalska et al, 2002).

Taking into account the practical advantages of the dissimilarity representation over the classical feature-based one (Pelillo, 2013), this paper faces the problem of corporate bankruptcy prediction in a way different from that traditionally followed by the methods reported in the literature. As far as we know, the dissimilarity-based paradigm, which has shown to be truly effective on various real-life problems, has not been applied in the financial scenario. Accordingly, the present paper analyzes the performance of four standard linear classifiers built on the dissimilarity space for the discovery of corporate financial failure using a data set whose explanatory variables are qualitative, and compares them with their feature-based counterparts. The reasons for focusing this study on linear models are three-fold (Yuan et al, 2012): (i) they are good handling sparse data; (ii) they are easy to describe mathematically, computational simple and easy to interpret; and (iii) when applied to dissimilarity data, they often lead to very good performance (Pękalska et al, 2002).

The remaining of the paper is organized as follows. Fundamental concepts related to the dissimilarity representation are summarized in Section 2. The prediction methodology proposed in this paper is described in Section 3. Next, Section 4 introduces the bankruptcy database and describes the experimental set-up. Results are presented and discussed in Section 5. Finally, a number of concluding remarks and possible directions for future research are outlined in Section 6.

2 Dissimilarity Space

From a practical viewpoint, the bankruptcy prediction problem can be defined as a binary classification problem where a new input sample has to be catego-

rized into one of the predefined classes based on a number of observed variables or features related to that sample. Formally, it can be described as follows: Given a set of past observations $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where each example x_i is characterized by a vector of m features, $[x_{i1}, x_{i2}, \ldots x_{im}]$, and y_i denotes the class (bankrupt/non-bankrupt), then the bankruptcy prediction problem consists of constructing a model δ to predict the value y for a new input sample $\mathbf{x}$, that is, $\delta(\mathbf{x}) = y$.

Traditional prediction models rely on the description of examples through a set of explanatory variables. A reliable alternative to the feature (variable) space is the dissimilarity space proposed by Pękalska and Duin (2002), in which the dimensions are defined by vectors measuring pairwise dissimilarities between examples and individual prototypes from a given representation set $R = \{p_1, \ldots, p_r\}$, where r is its cardinality. This set can be chosen as the complete training set T, a set of generated prototypes, a subset of T that covers all classes, or even an arbitrary set of labeled or unlabeled samples (Pekalska et al, 2006). Although the representation set can be selected either in a systematic or in a random way, it has been shown that both strategies produce similar classification results (Duin et al, 1999).

Given a dissimilarity measure $d(\cdot, \cdot)$, which is required to fulfill the positivity $(d(x_i, x_j) > 0$ if x_i is distinct from $x_j)$ and the reflectivity $(d(x_i, x_i) = 0)$ conditions but it might be non-metric, a dissimilarity representation is defined as a data-dependent mapping function $D(\cdot, R)$ from T to the dissimilarity space. This means that every example $x_i \in T$ can directly be represented by an r-dimensional vector in the dissimilarity space, $D(x_i, R) = [d(x_i, p_1), \ldots, d(x_i, p_r)]$, that is, each dimension corresponds to a dissimilarity to a prototype from R. Therefore, dissimilarities between all examples in T to R are represented by a matrix $D(T, R)$ of size $n \times r$, which corresponds to the dissimilarity representation we want to learn from (Pękalska and Duin, 2005).

$$D(T, R) = \begin{bmatrix} d(x_1, p_1) & d(x_1, p_2) & \cdots & d(x_1, p_r) \\ d(x_2, p_1) & d(x_2, p_2) & \cdots & d(x_2, p_r) \\ \vdots & \vdots & \ddots & \vdots \\ d(x_n, p_1) & d(x_n, p_2) & \cdots & d(x_n, p_r) \end{bmatrix}$$

In general, a drawback related to the use of features is that completely different examples may have the same feature representation, which results in class overlap (examples that belong to different classes are represented by the same feature vectors). In the dissimilarity space, however, only identical examples (with the same class label) have a zero-distance, which means that there does not exist class overlapping. On the other hand, the dissimilarity-based classifiers may be robust against variations in scale (Duin and Pękalska, 2012). Note that in principle, any standard classifier can be built on the dissimilarity space in the same way as on the feature space.

3 Methodology

This section provides a general overview of the complete methodology for constructing the model and classifying new corporate samples. Figure 1 shows a flowchart of the learning and prediction processes for both a classical feature-based representation (black lines) and a dissimilarity-based representation (red lines).
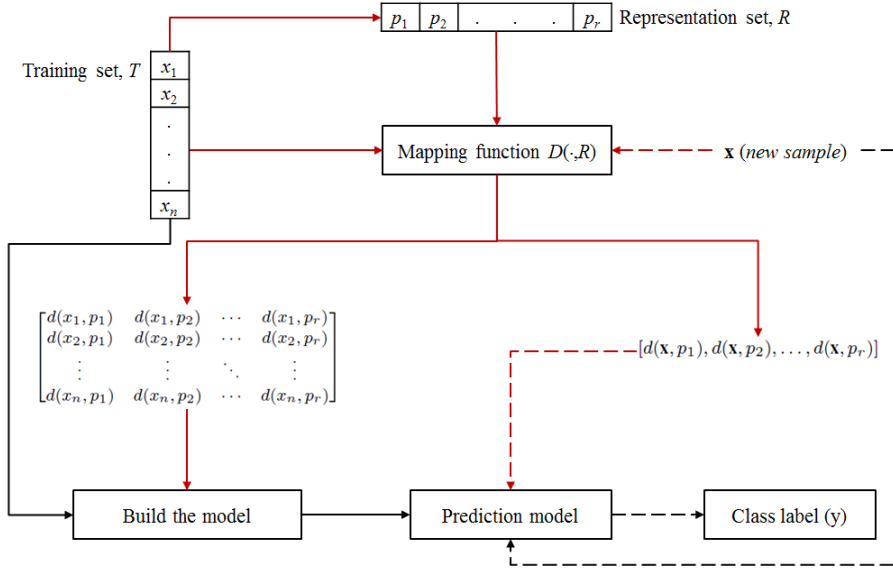

Fig. 1 Flowchart of the proposed methodology

Using a feature-based representation, the learning stage (continuous lines) simply consists of building the classifier with the training set T. In the case of a dissimilarity-based representation, the first step of learning consists of choosing a representation set R, whose prototypes will be used to measure the pairwise dissimilarities to the training examples in T. Next, the training set T is mapped into a dissimilarity space, which will be finally used to build the classifier.

In the testing stage (dashed lines), when a new sample $\mathbf{x}$ has to be classified, it is mapped into the dissimilarity space by calculating the dissimilarity between $\mathbf{x}$ and all prototypes in the representation set R, which results in a one-dimensional matrix (vector) $D(\mathbf{x}, R) = [d(\mathbf{x}, p_1), \ldots, d(\mathbf{x}, p_r)]$. This dissimilarity vector $D(\mathbf{x}, R)$ is passed through the prediction model for assigning a class label y to the new sample $\mathbf{x}$.

4 Database and Experimental Protocol

The database used in the present experiments was taken from the UCI Machine Learning Database Repository (Lichman, 2013). This is a subset of samples collected during the period 2001–2002 from one of the largest commercial banks in Korea (Kim and Han, 2003). It consists of 250 instances, with about 43% of them labeled as bankrupt. Each sample is represented by explanatory variables that correspond to levels (negative, average, and positive) of six qualitative risk factors (see Table 1) evaluated by loan officers. These risk factors are the ones established and used by the bank in order to estimate the default risk of manufacturing and service companies. Since all these variables were categorical, they were first converted into numeric values (negative = 1, average = 2, and positive = 3) as reported in the paper by Kim and Han (2003), and then these were normalized in the range $[0, 1]$.

Table 1 Description of the explanatory variables

Risk factor (variable)	Meaning
Industry risk (IR)	Stability and growth of the company, degree of competition within the company, and overall conditions of the company
Management risk (MR)	Efficiency and stability of management and organization structure
Financial flexibility (FF)	Ability of the company for financing from direct and indirect financial market and other sources
Credibility (CR)	Reputation of the company associated with credit history, reliability of information provided by the company, and the relationship with financial institutions
Competitiveness (CO)	Degree of competitive advantage determined by market position and the capacity of core technology
Operating risk (OP)	Volatility and stability of procurement, efficiency of production, and stability of sales

Even though the key question of this paper is not to select the most relevant explanatory variables, two feature ranking methods were applied to evaluate the usefulness of each variable: the ReliefF algorithm and the Pearson's correlation-based approach. The former evaluates the worth of a variable by repeatedly sampling an instance and considering the value of the given variable for the nearest instance of the same and different classes, whereas the latter evaluates the worth of a variable by measuring the correlation between it and the class. Results in Table 2 indicate that competitiveness (CO) is the most meaningful variable and the industry risk (IR) corresponds to the least relevant feature in terms of both ranking scores.

Bearing in mind that the purpose of this study is to compare both feature representations in the field of bankruptcy prediction, not to select the most meaningful variables, the experiments focused on four linear classifiers: the Fisher's linear discriminant (FLD), the linear discriminant classifier (LDC), a support vector machine (SVM) with a linear kernel and the soft-margin

Table 2 Relevance of the explanatory variables

	Ranking
ReliefF	CO (0.438) - FF (0.293) - CR (0.247) - OP (0.046) - MR (0.020) - IR (0.007)
Pearson	CO (0.204) - OP (0.159) - CR (0.152) - MR (0.101) - FF (0.048) - IR (0.016)

constant $C = 1.0$, and the logistic regression (logit) model (this is considered a classical econometric method that can be viewed as a reference approach for various financial applications). The performance of these techniques was explored both on the feature space (FS) and the dissimilarity space (DS). For the latter case, we chose the representation set R to be equal to a percentage of examples from the training set T, varying from 1% to 50% with a step size of 1. Here two variants were used: (i) the representation set was randomly drawn by picking examples from T without taking care of their class label (R-DS), and (ii) the representation set was created by randomly selecting the same proportion of examples from each class (RC-DS).

The common method to evaluate the performance of bankruptcy prediction systems when databases are small or medium sized corresponds to K-fold cross-validation because it appears to be a better estimator than other strategies, such as bootstrap with a high computational cost or re-substitution with a biased behavior (García et al, 2015). Here a stratified 5-fold cross-validation was applied: the data set was randomly divided into five stratified blocks of equal size; for each fold, four blocks were pooled as the training set, and the remaining part was used as an independent test set. Thus, the learning procedure was run a total of five times on different training sets and the results from predicting the class of the test samples were averaged across the five trials. Note that stratification allows to preserve the class proportions of the whole data set into each one of the blocks, thus reducing the prior probability of data set shift and the variance in the estimation process (Santafe et al, 2015).

In most financial applications, it is important to assess not only the overall accuracy of the model, but also the true-positive and true-negative hits because the misclassification costs are usually asymmetric (the cost of predicting a bankrupt sample as non-bankrupt is generally much higher than the opposite situation) (Caouette et al, 2008). The true-positive rate (or sensitivity) is the proportion of positive samples that are correctly predicted, whereas the true-negative rate (or specificity) is the proportion of negative cases that are correctly predicted. Note that we have considered that the bankrupt examples shape the positive class and the non-bankrupt ones form the negative class.

5 Results and Discussion

Figures 2– 3 display the accuracy, the true-positive rate (TPr) and the true-negative rate (TNr) averaged across the five runs. For each prediction model, we have plotted the results for the feature space and also the results of the two

variants for the dissimilarity space when varying the percentage of examples from T that have been chosen to generate the representation set R. Note that the line parallel to X-axis corresponds to the case of the feature space, which indicates that the results do not depend on the size of R because they were achieved by learning directly from the training set T. These plots show that the models built with any of both approaches to the dissimilarity representation perform much better than the respective feature-based classifiers.
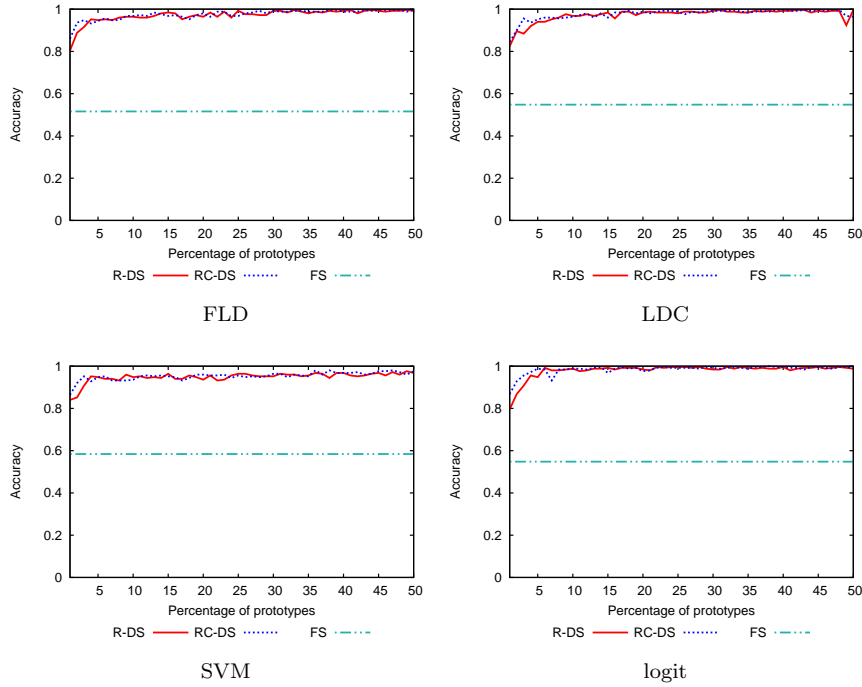

Fig. 2 Accuracy rates when varying the size of the representation set

If the focus is on the plots of Figure 3, it is remarkable and important to notice that differences between the dissimilarity space and the feature space are especially significant in the case of the true-positive rate, which refers to the number of hits on the most critical class because of the high cost of failing in the prediction of bankrupt samples.

When comparing R-DS and RC-DS, the plots in Figures 2– 3 indicate that in general, there do not exist differences in prediction performance, independently of the classifier used. However, when the percentage of prototypes is less than 5%, the option of generating the set R with the same proportion of examples from each class (RC-DS) performs slightly better than the R-DS variant.
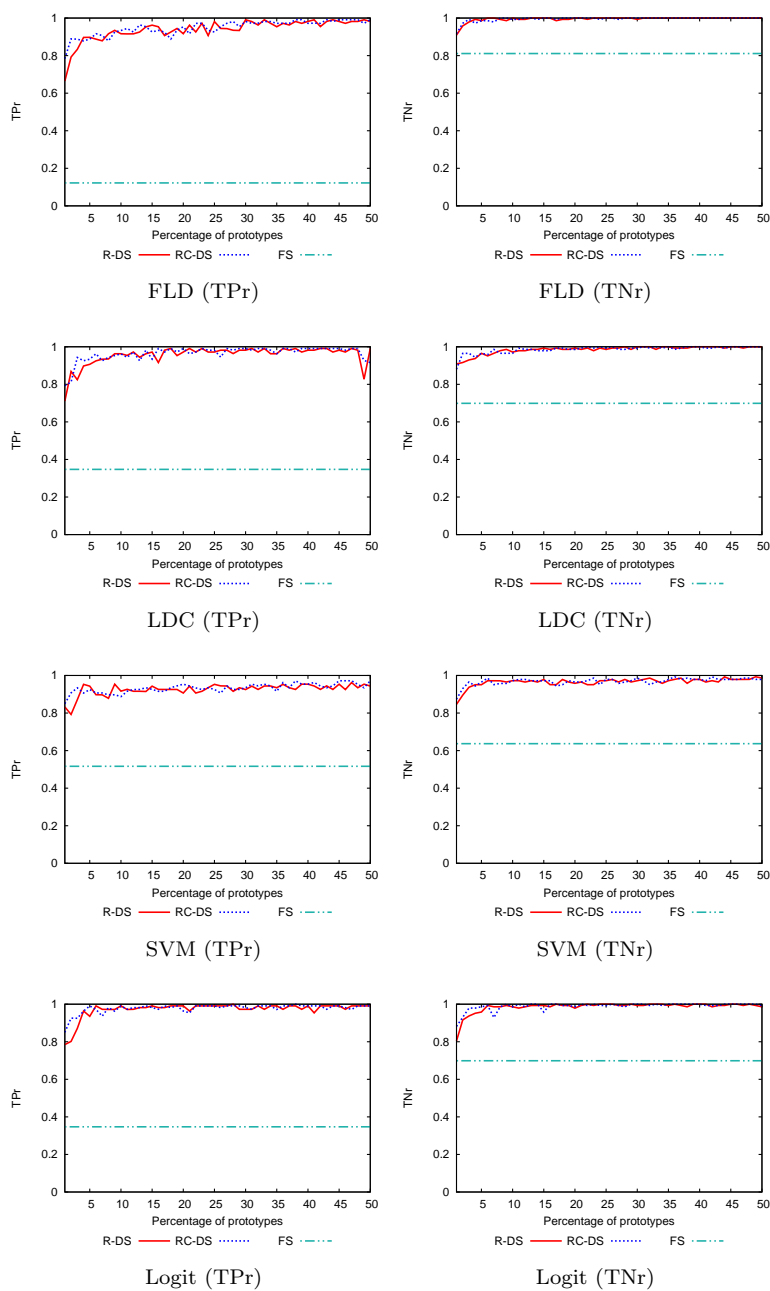
Fig. 3 True-positive and true-negative rates when varying the size of the representation set

Tables 3 and 4 report a summary of the experimental results for 10%, 20%, 30%, 40% and 50% of prototypes used to built the representation set.

Table 3 Summary of accuracy rates

	%	FLD	LDC	Logit	SVM
FS	–	0.5160	0.5480	0.5840	0.5480
R-DS	10	0.9640	0.9680	0.9480	0.9880
	20	0.9640	0.9840	0.9360	0.9840
	30	0.9920	0.9920	0.9520	0.9840
	40	0.9920	0.9920	0.9680	0.9960
	50	0.9920	0.9960	0.9680	0.9880
RC-DS	10	0.9680	0.9640	0.9360	0.9880
	20	0.9800	0.9880	0.9600	0.9760
	30	0.9920	0.9920	0.9640	0.9920
	40	0.9880	0.9960	0.9680	0.9960
	50	0.9960	0.9640	0.9760	0.9920

As can be observed, using a dissimilarity space instead of a feature space consistently produces considerable gains in terms of accuracy, true-positive rate and true-negative rate. In the case of accuracy, whilst the performance of the prediction models on the feature space is about 51%–58%, that on the dissimilarity space is about 96%–99%. Differences are even more significant when the performance is assessed by means of the true-positive rate, especially with the Fisher's linear discriminant model. On the other hand, various configurations of the dissimilarity representation yield 100% of true-negative rate. These results support the claim that the linear models generally lead to very high performance when they are built on the dissimilarity space.

Table 4 Summary of true-positive and true-negative rates

		TPr				TNr			
	%	FLD	LDC	Logit	SVM	FLD	LDC	Logit	SVM
FS	–	0.1221	0.3472	0.5165	0.3472	0.8111	0.6988	0.6369	0.6988
R-DS	10	0.9165	0.9065	0.9165	0.9909	1.0000	0.9719	0.9724	0.9862
	20	0.9165	0.9719	0.9065	0.9909	1.0000	0.9931	0.9581	0.9788
	30	0.9909	0.9818	0.9251	0.9727	0.9931	1.0000	0.9722	0.9931
	40	0.9818	0.9818	0.9537	0.9909	1.0000	1.0000	0.9791	1.0000
	50	0.9818	0.9909	0.9437	0.9909	1.0000	1.0000	0.9862	0.9860
RC-DS	10	0.9338	0.9623	0.8874	0.9909	0.9929	0.9653	0.9722	0.9860
	20	0.9537	0.9909	0.9537	0.9636	1.0000	0.9860	0.9650	0.9862
	30	0.9818	0.9909	0.9346	0.9818	1.0000	0.9929	0.9860	1.0000
	40	0.9719	0.9909	0.9537	0.9909	1.0000	1.0000	0.9793	1.0000
	50	0.9909	0.9156	0.9719	0.9909	1.0000	1.0000	0.9791	0.9931

To gain some insight into these results, we have projected the data onto a two-dimensional subspace through PCA. Figure 4 shows the scatter plots of the original feature space and the two variants of the dissimilarity space (for

the percentages of prototypes reported in Tables 3 and 4). In addition, as the size of the original training set is 250×6 (250 examples and 6 explanatory variables), we have also included the scatter plots of the dissimilarity representations obtained by random selection of six examples, which results in a matrix $D(T, R)$ of size 250×6. By this, one can compare the class distribution on both spaces under identical conditions (sizes).

As can be seen in Figure 4, the overlap between bankrupt and non-bankrupt examples is very high in the feature space, whereas both dissimilarity-based variants give rise to good separability between classes, irrespective of the size of the representation set R. The lack of separability between classes in the feature space may result in many false-positives or false-negatives, which helps to explain the low performance of the prediction models when they were applied on this space.
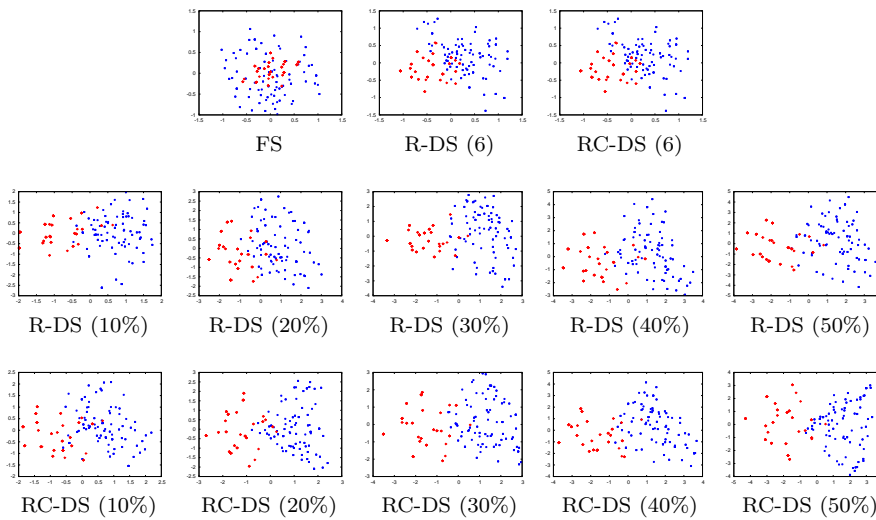

Fig. 4 Distribution of the bankrupt (red diamond) and non-bankrupt (blue star) classes in a two-dimensional space

6 Conclusions and Future Work

In the present study, we have explored the feasibility of applying the dissimilarity representation to effectively discriminate between bankrupt and non-bankrupt companies. To this end, four well-known linear prediction techniques (FLD, LDC, SVM and logit) have been implemented both on the feature space and the dissimilarity space and tested over a database generated by a commercial bank in Korea.

The experimental results have demonstrated that all the linear models here analyzed for bankruptcy prediction perform clearly better on the dissimilarity space than on the feature space in terms of accuracy, true-positive rate and true-negative rate. Projection of data onto a two-dimensional subspace has shown that the dissimilarity representation provides significantly higher separability between classes than the original feature representation, which allows to understand why the dissimilarity-based prediction models outperform their feature-based counterparts.

In the future, it would be of interest to perform further simulation studies that compare linear and non-linear prediction models on both the dissimilarity and the feature spaces. Other research directions might include the application of the methodology described in this paper to analyze the effects of class imbalance and data set shift on the dissimilarity-based models for bankruptcy prediction or even for other economic and financial problems. A final avenue for further research is to study the applicability of the dissimilarity representation to select the most relevant explanatory variables. This is a non-trivial problem that may require a significant effort, but deserves to be taken into account.

References

Abellán J, Mantas CJ (2014) Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. Expert Systems with Applications 41(8):3825–3830

Acosta-González E, Fernández-Rodríguez F (2014) Forecasting financial failure of firms via genetic algorithms. Computational Economics 43(2):133–157

Alfaro E, García N, Gámez M, Elizondo D (2008) Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. Decision Support Systems 45(1):110–122

Altman EI (1993) Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy. Wiley, New York

Altman EI, Haldeman RG, Narayanan P (1977) ZETA analysis: A new model to identify bankruptcy risk of corporations. Journal of Banking & Finance 1(1):29–54

Aziz MA, Dar HA (2006) Predicting corporate bankruptcy: where we stand? Corporate Governance 6(1):18–33

Balcaen S, Ooghe H (2006) 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. The British Accounting Review 38(1):63–93

Caouette JB, Altman EI, Narayanan P, Nimmo R (2008) Managing Credit Risk: The Great Challenge for Global Financial Markets. Wiley, Hoboken, NJ

Chen MY (2012) Comparing traditional statistics, decision tree classification and support vector machine techniques for financial bankruptcy prediction. Intelligent Automation & Soft Computing 18(1):65–73

Cielen A, Peeters L, Vanhoof K (2004) Bankruptcy prediction using a data envelopment analysis. European Journal of Operational Research 154(2):526–532

Cleofas-Sánchez L, García V, Marqués AI, Sánchez JS (2016) Financial distress prediction using the hybrid associative memory with translation. Applied Soft Computing 44:144–152

Duin RPW, Pękalska E (2012) The dissimilarity space: Bridging structural and statistical pattern recognition. Pattern Recognition Letters 33(7):826–832

Duin RPW, Pękalska E, de Ridder D (1999) Relational discriminant analysis. Pattern Recognition Letters 20(11–13):1175–1181

Erdal HI, Ekinci A (2013) A comparison of various artificial intelligence methods in the prediction of bank failures. Computational Economics 42(2):199–215

Erdogan BE (2013) Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. Journal of Statistical Computation and Simulation 83(8):1543–1555

Fedorova E, Gilenko E, Dovzhenko S (2013) Bankruptcy prediction for Russian companies: Application of combined classifiers. Expert Systems with Applications 40(18):7285–7293

García V, Marqués AI, Sánchez JS (2015) An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. Journal of Intelligent Information Systems 44(1):159–189

Jones S, Hensher DA (2004) Predicting firm financial distress: A mixed logit model. The Accounting Review 79(4):1011–1038

Karels GV, Prakash AJ (1987) Multivariate normality and forecasting of business bankruptcy. Journal of Business Finance & Accounting 14(4):573–593

Kim MJ, Han I (2003) The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. Expert Systems with Applications 25(4):637–646

Kumar PR, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operational Research 180(1):1–28

Kwak W, Shi Y, Kou G (2012) Bankruptcy prediction for Korean firms after the 1997 financial crisis: using a multiple criteria linear programming data mining approach. Review of Quantitative Finance and Accounting 38(4):441–453

Lensberg T, Eilifsen A, McKee TE (2006) Bankruptcy theory development and classification via genetic programming. European Journal of Operational Research 169(2):677–697

Lichman M (2013) UCI Machine Learning Repository. `http://archive.ics.uci.edu/ml/datasets/qualitative_bankruptcy`, Irvine, CA, University of California, School of Information and Computer Sciences

Mckee TE (2000) Developing a bankruptcy prediction model via rough sets theory. Intelligent Systems in Accounting, Finance & Management 9(3):159–173

Min JH, Lee YC (2005) Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications 28(4):603–614

Ohlson JA (1980) Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18(1):109–131

Olson DL, Delen D, Meng Y (2012) Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems 52(2):464–473

Pȩkalska E, Duin RPW (2002) Dissimilarity representations allow for building good classifiers. Pattern Recognition Letters 23(8):943–956

Pȩkalska E, Duin RPW (2005) The Dissimilarity Representation for Pattern Recognition: Foundations and Applications. World Scientific Publishing, River Edge, NJ

Pȩkalska E, Paclik P, Duin RPW (2002) A generalized kernel approach to dissimilarity-based classification. Journal of Machine Learning Research 2:175–211

Pekalska E, Duin RPW, Paclík P (2006) Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39(2):189–208

Pelillo M (2013) Similarity–Based Pattern Analysis and Recognition. Springer–Verlag, London, UK

Premachandra IM, Bhabra GS, Sueyoshi T (2009) DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. European Journal of Operational Research 193(2):412 – 424

Santafe G, Inza I, Lozano JA (2015) Dealing with the evaluation of supervised classification algorithms. Artificial Intelligence Review pp 1–42

Shin KS, Lee TS, jung Kim H (2005) An application of support vector machines in bankruptcy prediction model. Expert Systems with Application 28(1):127–135

Slowinski R, Zopounidis C (1995) Application of the rough set approach to evaluation of bankruptcy risk. Intelligent Systems in Accounting, Finance and Management 4(1):27–41

Sun L, Shenoy PP (2007) Using Bayesian networks for bankruptcy prediction: Some methodological issues. European Journal of Operational Research 180(2):738–753

Tsai CF (2014) Combining cluster analysis with classifier ensembles to predict financial distress. Information Fusion 16:46–58

Tsai CF, Hsu YF, Yen DC (2014) A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing 24:977–984

Tseng FM, Lin L (2005) A quadratic interval logit model for forecasting bankruptcy. Omega 33(1):85–91

Verikas A, Kalsyte Z, Bacauskiene M, Gelzinis A (2010) Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. Soft Computing 14(9):995–1010

West RC (1985) A factor-analytic approach to bank condition. Journal of Banking & Finance 9(2):253–266

Wilson RL, Sharda R (1994) Bankruptcy prediction using neural networks. Decision Support Systems 11(5):545–557

Yuan ZX, Ho CH, Lin CJ (2012) Recent advances of large-scale linear classification. Proceedings of the IEEE 100(9):2584–2603

Zhao D, Huang C, Wei Y, Yu F, Wang M, Chen H (2016) An effective computational model for bankruptcy prediction using kernel extreme learning machine approach. Computational Economics pp 1–17

Zmijewski ME (1984) Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research 22:59–82

Zopounidis C, Dimitras AI (1998) Multicriteria Decision Aid Methods for the Prediction of Business Failure. Kluwer Academic Publishers, Dordrecht, The Netherlands