ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Forecasting Business Failure in Highly Imbalanced Distribution based on Delay Line Reservoir

A. Rodan[1,2], P.A. Castillo[3], H. Faris[2], Ala' M. Al-Zoubi[2], A.M. Mora[4], H. Jawazneh[3] *

[1]Higher Colleges of Technology,
United Arab Emirates

[2]King Abdullah II School for Information Technology,
The University of Jordan, Jordan

[3]Department of Computer Architecture and Technology,
ETSIIT and CITIC, University of Granada, Spain

[4]Department of Computer Sciences and Technology, ESIT,
International University of La Rioja (UNIR), Spain

Abstract. Bankruptcy is a critical financial problem that affects a high number of companies around the world. Thus, in recent years an increasing number of researchers have tried to solve it by applying different machine-learning models as powerful tools for the different economical agents related to the company. In this work, we propose the use of a simple deterministic delay line reservoir (DLR) state space by combining it with three popular classification algorithms (J48, k-NN, and MLP) as an efficient and accurate solution to the bankruptcy prediction problem. The proposed approach is evaluated on a real world dataset collected from Spanish companies. Obtained results show that the proposed models have a higher predictive ability than traditional classification approaches (without DLR reservoir state), resulting in a suitable and efficient alternative approach to solve this complex problem.

1 Introduction

Predicting financial failure is a problem traditionally approached heuristically, which requires a wide knowledge about that company, and usually is carried out by means of accounting experts. Nowadays it has become a critical problem, approached in the recent years by many researchers and that, doubtlessly, it is a topic that also worries company shareholders. The interest in determining the financial status of a company is based on both the management, who can thus count on information to correct foreseeable financial distress, and the credit bureaus and other potential investors, who need this information to evaluate their investment [1].

Recently, several authors developed and applied different ANN models to financial and bankruptcy prediction, with a high degree of success, such as

*This work has been partially funded by projects EphemeCH TIN2014-56494-C4-3-P, DeepBio TIN2017-85727-C4-2-P and TEC2015-68752 (Spanish Ministry of Economy and Competitiveness and FEDER).

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Salchenberger et al. [2], who describes an ANN-based method that reaches results as good or even better than a logit model, observing at the same time less type I errors (i.e. false positives), but more type II errors (i.e. false negatives). In general, authors report better results when using ANN models, compared to classical statistical methods, as in [3, 4]. Another approach was followed by some researchers by proposing more complex models by combining different prediction methods in order to enhance the accuracy of the predictor. In this sense, Verikas et al. [5] proposed a hybrid method based on soft computing techniques for prediction, while Yeh et al. [6] created two-stages classifier by merging Rough set theory with SVM.

Reservoir computing (RC) [7] is a framework for computation using recurrent neural networks that solves the problems of gradient decent methods, where only the output (readout) layer is trained using any linear regression method. Echo state network (ESN) [7] is one of the simple and most effective forms of RC. In ESN all the input and hidden (reservoir) weights are fixed random drawn from a uniform distribution over a symmetric interval. Simple fixed deterministic versions of ESN was proposed by Rodan and Tino [8] that have a comparable results of the classical ESN on several benchmark datasets. One of theses simple deterministic models is the simple Markovian delay line reservoir (DLR) [8] that has been used in the literature for several tasks including Non-linear Communication Channel [7, 8] and Photonic hardware implementation of Artificial neural networks (ANNs) [9] with promising results.

In this paper we will use the reservoir state of DLR as an input for three popular classification algorithms (C4.5 decision tree algorithm (J48), k-nearest neighbors (k-NN), and multilayer perceptron neural network (MLP)). Moreover, we will also propose the use of an ensemble model based on the majority voting of the three classification models.

The remainder of this paper is structured as follows: Section 2 describes the problem and the dataset used in this paper. In section 3, the proposed methods are detailed. Section 4 describes the experiments and the obtained results, followed by a brief conclusion in Section 5.

2 Problem description

The sample used in this paper contains 470 non-financial firms taken from the Infotel database[1], including information about both successful and failed companies during six years sequentially (1998 to 2003). Thus, there are 2860 patterns, from which 62 correspond to financial failures or bankruptcy in those enterprises. The group of financial failures corresponds to those firms that had suspended payments or had declared legal bankruptcy.

In order to solve the bankruptcy problem, the dependent variable takes a value of 1 in the case of legal failure (F), and of 0 in the case of a healthy firm (H). The original dataset included 2860 instances, each one of them consists of

[1] http://www.infotel.es

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

39 independent variables of different types (categorical and numerical). After removing meaningless variables (such as internal codes), we adopted 33 variables, 27 of them are numeric and the remaining are categorical (some of those variables refer to financial information). The independent variables are quantitative, ratios taken from financial statements, along with qualitative information. In the dataset the proportion of bankrupt to healthy is very small so we face the problem of imbalanced dataset.

3 DLR reservoir state combined with classification models

For our proposed models, we use a DLR with fully connected input layer, where all input connections have the same absolute weight value $v>0$; the sign of each input weight is deterministically generated from aperiodic pattern (in our case π). In the reservoir layer, the units are organized in a line, where only elements on the lower subdiagonal of the reservoir matrix W have non-zero values $W_{i+1,i} = r$ for $i = 1...N-1$, where r is the weight of all the feedforward connections and N is the reservoir size. We drive the DLR with the input data and collect the reservoir states $\dot{X}$ using the following equation.

$$\dot{X} = \frac{1}{\alpha}(-ax + f(Vs + Wx + z)), \tag{1}$$

where f is the reservoir activation function ($tanh$ in this study), V is the input to reservoir weight matrix, s is the input data, z is zero-mean noise, $a \in [0,1]$ is the leaking rate parameter, and $\alpha > 0$ is the time constant. The reservoir state $\dot{X}$ of DLR will be used as an input to a classification model to produce a classification readout. This can be expressed as follows:

$$Y = Model(\dot{X}) \tag{2}$$

where $Model$ is a simple classification model. In this work four different models will be applied. The first three models are simple well-known classification models, i.e. C4.5 decision tree algorithm, k-NN and MLP. The fourth model is an ensemble majority classification model. Therefore, combining DLR with these classification models will produce four different hybrid models, which will be denoted as DLR-J48, DLR-k-NN, DLR-MLP, and DLR-Vote, respectively. The general idea of DLR reservoir state combined with classification models is depicted in Figure 1.

4 Experiments and results

The DLR part of the proposed framework is trained in a way that the optimal parameters including input weight value v, reservoir weight r, leaking rate a, zero-mean noise z, time constant α, and reservoir size N to be chosen by minimising the mean square error (MSE) values using linear regression based on the training dataset. For k-NN, the number of neighbours is set to 1. For MLP, the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

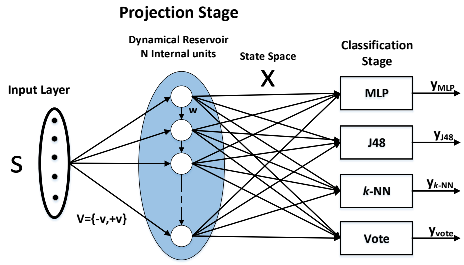

Fig. 1: An illustration of the proposed DLR reservoir state combined with classification models.

number of hidden nodes is set to half the number of input features, the learning rate is 0.3, the momentum is 0.2 and the number epochs is 500.

For training and testing, we split the dataset using stratified sampling, where half of the dataset is used for training and the rest is used for testing. Stratified sampling is important for imbalanced datasets to preserve the original ratio of the class labels. For evaluation, the following measures are calculated: Accuracy, Precision, Recall and the geometric mean of the recalls of both classes (G-mean).

In our experiments we follow two different scenarios:

Scenario I: We apply the DLR model in combination with a majority voting scheme of heterogeneous base classification algorithms (i.e k-NN, MLP, J48) independently. This approach will be compared to DLR in combination with each classifier independently, and it will be compared also with the performance of the base classifiers as well. All these classification methods will be trained on the training dataset without any oversampling.

Scenario II: Same classification methods will be applied this time on training datasets that are oversampled using Synthetic Minority Over-sampling Technique (SMOTE) at different ratios (i.e 100%, 200%, 300%, 400%, and 500%). The goal of this scenario is to study the performance of the proposed approach after handling the imbalanced class distribution.

Results without oversampling: The evaluation results of all classification algorithms based on the training dataset without oversampling are shown in Table 1. As it can be seen, all classification methods shows very competitive classification accuracy rate. However, since the dataset is highly imbalanced, these ratios do not give a realistic indication of the performance of the classifiers. For these reasons, the other measures are examined. Having a look at the Precision, Recall and G-mean ratios, it can be seen that all DLR based models noticeably outperform the basic classifiers. The best performing model is DLR-MLP with a precision of 48.8%, recall of 64.5%, and a G-mean of 79.7%.

Results with oversampling: In an attempt to improve the classification results obtained in scenario I, all classifiers are trained based on the training dataset oversampled at different ratios (i.e 100%, 200%, 300%, 400%, and 500%). The results of this experiment are shown in Table 2. In general, the results of all classification models have been significantly improved after applying the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Table 1: Evaluation results based on the original dataset (i.e without oversampling)

Classifier	Accuracy	Precision	Recall	G-mean
J48	0.973	0.333	0.226	0.473
MLP	0.974	0.350	0.226	0.473
k-NN	0.968	0.143	0.097	0.309
Vote	0.977	0.417	0.161	0.401
DLR-J48	0.970	0.300	0.290	0.535
DLR-MLP	0.978	0.488	0.645	0.797
DLR-k-NN	0.971	0.273	0.194	0.437
DLR-Vote	0.974	0.385	0.323	0.565

oversampling step. Moreover, all DLR based models still outperform the basic classifiers. Overall, the best classifier is DLR-J48 at oversampling ratio of 300%. It achieved a precision of 55.8%, recall of 93.5%, and G-mean of 95.9%.

Table 2: Evaluation results based on the dataset oversampled with different ratios.

Classifier	Accuracy	Precision	Recall	G-mean
SMOTE=100%				
J48	0.977	0.444	0.258	0.506
MLP	0.969	0.276	0.258	0.504
k-NN	0.966	0.200	0.194	0.436
Vote	0.977	0.450	0.290	0.537
DLR-J48	0.974	0.432	0.613	0.776
DLR-MLP	0.985	0.656	0.677	0.820
DLR-k-NN	0.971	0.321	0.290	0.535
DLR-Vote	0.977	0.474	0.581	0.757
SMOTE=200%				
J48	0.967	0.250	0.258	0.504
MLP	0.969	0.290	0.290	0.535
k-NN	0.967	0.214	0.194	0.436
Vote	0.973	0.316	0.194	0.438
DLR-J48	0.974	0.429	0.581	0.755
DLR-MLP	0.979	0.516	0.516	0.715
DLR-k-NN	0.971	0.310	0.290	0.535
DLR-Vote	0.978	0.500	0.484	0.692
SMOTE=300%				
J48	0.971	0.351	0.419	0.642
MLP	0.969	0.259	0.226	0.472
k-NN	0.965	0.194	0.194	0.436
Vote	0.972	0.304	0.226	0.472
DLR-J48	0.983	0.558	0.935	0.959
DLR-MLP	0.971	0.378	0.548	0.733
DLR-k-NN	0.972	0.333	0.290	0.535
DLR-Vote	0.980	0.543	0.613	0.778
SMOTE=400%				
J48	0.970	0.342	0.419	0.642
MLP	0.969	0.290	0.290	0.535
k-NN	0.964	0.161	0.161	0.398
Vote	0.978	0.474	0.290	0.537
DLR-J48	0.971	0.383	0.581	0.754
DLR-MLP	0.974	0.429	0.581	0.755
DLR-k-NN	0.972	0.364	0.387	0.617
DLR-Vote	0.979	0.514	0.581	0.757
SMOTE=500%				
J48	0.971	0.292	0.226	0.472
MLP	0.969	0.276	0.258	0.504
k-NN	0.962	0.171	0.194	0.435
Vote	0.974	0.350	0.226	0.473
DLR-J48	0.979	0.511	0.774	0.873
DLR-MLP	0.980	0.529	0.581	0.758
DLR-k-NN	0.971	0.353	0.387	0.617
DLR-Vote	0.980	0.531	0.548	0.737

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

5 Conclusions

In this work, the reservoir state of DLR has been used as an input for three popular classification algorithms, i.e. J48, k-NN, and MLP in order to efficiently solve the bankruptcy prediction problem. The proposed approach has been evaluated on a real world dataset taken from the Infotel database, including information about both successful and failed Spanish companies during six years sequentially. We have empirically demonstrated that the use of DLR reservoir state can lead to performance improvements in Forecasting Business Failure over the normal ensemble voting or other single classifier models including J48, k-NN, and MLP. For future work, we aim at experimenting and compare DLR to other networks like Echo state networks, Cycle Reservoirs with Regular Jumps and Extreme Learning Machines.

References

[1] B.S. Ahn, S.S. Cho, and C.Y. Kim. The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications, 18*, pages 65–74, 2000.

[2] Salchenberger L.M., E.M. Cinar, and N.A. Lash. Neural networks: A new tool for predicting thrift failures. *Decision Sciences, July/Agoust, 23(4): 899-916*, 1992.

[3] Clarence N. W. Tan. A study on using artificial neural networks to develop an early-warning predictor for credit union financial distresss with comparison to the probit model. *Neural Networks in Finance and Investing. R.R. Trippi and E. Turban Editors. ISBN:1-55738-919-5. pp. 329-365*, 2000.

[4] Tam K.Y. and M.Y. Kiang. Predicting bank failures: a neural network approach. *Neural Networks in Finance and Investing. R.R. Trippi and E. Turban Editors. ISBN:1-55738-919-5. pp. 267-301*, 2000.

[5] Antanas Verikas, Zivile Kalsyte, Marija Bacauskiene, and Adas Gelzinis. Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing*, 14(9):995–1010, 2010.

[6] Ching-Chiang Yeh, Der-Jang Chi, and Ming-Fu Hsu. A hybrid approach of dea, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, 37(2):1535–1541, 2010.

[7] Mantas Lukoševičius and Herbert Jaeger. Survey: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.*, 3(3):127–149, 2009.

[8] A. Rodan and P. Tino. Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, 22(1):131–144, 2011.

[9] G. Sande, D. Brunner, and M. Soriano. Advances in photonic reservoir computing. *Nanophotonics*, 6(3):561–576, 2017.