# Financial distress prediction using the hybrid associative memory with translation

L. Cleofas-Sánchez[a], V. García[b], A.I. Marqués[c], J.S. Sánchez[a,*]

[a]*Institute of New Imaging Technologies, Department of Computer Languages and Systems*
*Universitat Jaume I, 12071 Castelló de la Plana (Spain)*
[b]*División Multidisciplinaria de Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez*
*32310 Ciudad Juárez, Chihuahua (Mexico)*
[c]*Department of Business Administration and Marketing, Universitat Jaume I*
*12071 Castelló de la Plana (Spain)*

## Abstract

This paper presents an alternative technique for financial distress prediction systems. The method is based on a type of neural network, which is called hybrid associative memory with translation. While many different neural network architectures have successfully been used to predict credit risk and corporate failure, the power of associative memories for financial decision-making has not been explored in any depth as yet. The performance of the hybrid associative memory with translation is compared to four traditional neural networks, a support vector machine and a logistic regression model in terms of their prediction capabilities. The experimental results over nine real-life data sets show that the associative memory here proposed constitutes an appropriate solution for bankruptcy and credit risk prediction, performing significantly better than the rest of models under class imbalance and data overlapping conditions in terms of the true positive rate and the geometric mean of true positive and true negative rates.

*Keywords:* Associative memory, Neural network, Financial distress, Bankruptcy, Credit risk

## 1. Introduction

A large number of techniques have been developed to help decision-makers and analysts in predicting financial distress. Traditionally, decisions on credit risk of a corporate borrower were exclusively based upon subjective judgments made

---

*Corresponding author. Tel.: +34-964-728350.
*Email address:* sanchez@uji.es (J.S. Sánchez)

by human experts, using past experiences and some guiding principles [69]. However, two major problems with this approach are the difficulty to make consistent estimates and the fact that it tends to be reactive rather than predictive. The world financial crisis has led to increasing attention of banks and financial institutions on this question because of its significant impact on the decisions made [14], resulting in the development of numerous techniques to face the important challenge of credit risk and bankruptcy prediction from financial ratios using mathematical models. From the pioneer work by Altman [7], based on multivariate discriminant analysis, a variety of statistical and operations research methods have subsequently been applied to credit risk and bankruptcy prediction, including linear and logistic regression, multivariate adaptive regression splines, survival analysis, linear and quadratic programming, and multiple criteria programming. Most of these techniques typically rely on the assumptions of linear separability, multivariate normality and independence of the predictive variables, but they are very often violated in real-life problems [25, 34, 55].

Popular computational intelligence tools such as decision trees, neural networks, support vector machines, fuzzy systems, rough sets, artificial immune systems, and evolutionary algorithms are techniques that can deal with non-linearity. Besides, these methods are highly capable of extracting meaningful information from imprecise data and detecting trends that are too complex to be discovered by either humans or conventional systems. Despite various studies have concluded that no technique is clearly superior to other competing algorithms because it depends on the characteristics of the problem analyzed [13, 15, 16], different neural network architectures have shown good performance in comparison to other methods for a range of financial applications [10, 19, 48, 53, 78]. However, when the number of examples is relatively small, several works have demonstrated that the accuracy and generalization performance of a support vector machine (SVM) is usually better than that of statistical and other soft computing techniques [23, 24, 65, 67]. While typical neural networks used in this context are the multi-layer perceptron (MLP), the radial basis function (RBF) and the probabilistic or Bayesian network (BN), other neural models such as the associative memories have not been explored as yet.

The ability of human brain to make associations from partial information has historically attracted great interest among researchers, leading to a variety of theoretical neural networks that act as associative memories. An associative memory [39] is an early type of artificial neural network that relates an input vector $\mathbf{x}$ with an output vector $\mathbf{y}$. The functionality of associative memories is reached in two phases: learning and recall. The learning process consists of building a connection matrix $\mathbf{W}$ with a value for each association $(\mathbf{x}^k, \mathbf{y}^k)$. In the recall phase, an output vector $\mathbf{y}$, which corresponds to the most similar to the input vector $\mathbf{x}$, is

2

obtained from the associative memory. These models are powerful computational tools due to their conceptual and implementational simplicity, their strong mathematical foundation, and their capability of storing huge amounts of data that allow to properly recover the most similar patterns to an input vector with low computational efforts [77].

Representative examples of associative memories are lernmatrix [66], the linear associator [8, 38], the Moore-Penrose generalized inverse associative memory [40], the Hopfield network [28], the bidirectional associative memory [41], the fuzzy associative memory [42], the morphological associative memory [58], and the alpha-beta associative memory [2]. Some of these models have been used to solve very different problems. Sabourin and Mitiche [59] developed a Kohonen associative memory with selective multiresolution for OCR. A fuzzy associative memory was introduced to determine rock types from well-log signatures [17]. The bidirectional associative memory networks were used to find the relations between various cancers and elemental contents in serum samples with the aim of diagnosing cancer [81]. A hybrid classifier based on self-organizing maps and associative memories was designed for speaker recognition [31]. Zhang et al. [79] proposed a modular face recognition scheme by combining the wavelet subband representations and kernel associative memories. An associative memory based on the restricted Coulomb energy was also applied to human face recognition [49]. Namba and Zhang [50] devised an associative memory to recognize Braille images. A novel system for medical diagnosis based on associative memories was proposed by Aldape-Pérez et al. [5]. Itkar and Kulkarni [32] developed an efficient algorithm for mining frequent patterns using an auto-associative memory.

Apart from the associative memories just mentioned, Santiago-Montero [63] introduced the hybrid associative classifier and its extension, the hybrid associative classifier with translation (HACT). Both these associative memories are based on the learning phase of the linear associator and the recall phase of the Steinbuch's lernmatrix. This paper applies the HACT neural network to decision making problems for financial distress prediction and presents an empirical comparison with other popular prediction methods. To the best of our knowledge, this model has not been used for classification purposes, and even less in the context of finance and management. The aim of this paper therefore is four-fold:

1. To explore the capability of the HACT model in the prediction of bankruptcy and credit risk;
2. to analyze the behavior of this neural network under the presence of imbalance in class distribution, which constitutes a data complexity often neglected in financial applications;
3. to investigate how the class overlapping affects the performance of the asso-

ciative memory; and

4. to compare the performance of HACT with that of other prediction techniques.

From now on, the paper is organized as follows. Section 2 provides a review of works related to neural networks used for corporate bankruptcy and credit risk prediction. Section 3 introduces the fundamental concepts of the associative memories and describes the bases of the HACT model. The experimental set-up and databases are given in Section 4, while the results are discussed in Section 5. Finally, Section 6 presents the concluding remarks and outlines some directions for future research.

## 2. A review of neural networks applied to financial distress prediction

From the beginning of the 1990's, the development of artificial neural network technologies for bankruptcy and credit risk prediction problems has been the subject of considerable attention and research efforts. The first reference to using neural networks can be found in the paper by Odom and Sharda [51], showing that a three-layer feed-forward perceptron is more accurate and robust than multi-variate discriminant analysis. After this seminal work, many other studies have proposed the use of neural networks in credit scoring, bankruptcy or business failure prediction. For instance, Tam and Kiang [68] compared neural network models to linear discriminant analysis, logistic regression, nearest neighbors and decision tree for evaluating bank status. Salchenberger et al. [61] reported that the neural networks produced fewer or equal number of total errors, type-I errors and type-II errors compared to the logit model. Lacher et al. [43] investigated the use of the Cascade-Correlation neural network architecture and compared its performance with that of the multivariate discriminant analysis approach. Chang et al. [18] applied the theory and numerical algorithms of the BN to risk scoring and compared the results with traditional methods for computing scores and posterior predictions of performance variables.

Desai et al. [22] concluded that the MLP and the modular neural network can be especially useful to correctly predict the bad loans, but logistic regression models are comparable to the neural networks when the performance is measured by the percentage of good and bad loans correctly classified. West [74] analyzed the credit scoring accuracy of the MLP, the RBF network and several statistical techniques, suggesting that the MLP may not be the most accurate neural network model. An auto-associative memory trained with only data of non-bankrupt firms was developed by Baek and Cho [11]. Baesens et al. [12] used Markov Chain Monte Carlo search to learn unrestricted Bayesian network classifiers for credit scoring, which

4

gave a very good performance in terms of accuracy and area under the ROC curve. Also, Leong [46] showed that the BN performs well against logistic regression and MLP particularly with class imbalance, higher dimensions and a rejection sample, and it can be scaled efficiently when implemented onto a large data set.

The power of probabilistic and MLP neural networks was compared to that of discriminant analysis, probit analysis and logistic regression to evaluate credit risk in Egyptian banks [1]. Khashman [36] explored various back-propagation learning schemes to train three models (each with a different number of hidden neurons) of a three-layer supervised neural network. Angelini et al. [9] developed two neural network systems with a four-layer feed-forward topology, proving their applicability to credit risk prediction. An algorithm based on the threshold accepting meta-heuristic to train the principal component neural network architecture was investigated by Ravi and Pramodh [56], who inferred that their proposal outperformed other classifiers.

Lee and Chen [44] developed a credit scoring system using a hybrid modeling procedure with artificial neural networks whose input nodes were the variables obtained by multivariate adaptive regression splines. Hsieh [29] designed a credit scoring model that employed the SOM and K-means clustering algorithms to obtain the best inputs to a feed-forward MLP. Similarly, Lee et al. [45] explored the performance of credit scoring by integrating the linear discriminant analysis approach into a three-layer back-propagation neural network, revealing that the proposed hybrid approach converges much faster than the conventional neural network model and outperforms the discriminant analysis and logistic regression approaches. Cheng et al. [20] adopted an RBF to construct the financial prediction model and then carried out a logit analysis on the groups of similar firms present in the hidden layer of the network.

Chuang and Lin [21] proposed a reassigning credit scoring model involving two-stages: the classification stage builds a neural network-based credit scoring model, which classifies applicants with good or bad credits; then the reassign stage tries to reduce the type-I error by reassigning the rejected good credit applicants to the conditional accepted class by using a case-based reasoning classification technique. Khashei et al. [35] employed basic concepts of fuzzy logic and MLP neural networks to implement a hybrid binary credit risk prediction model, where fuzzy numbers were used so that the uncertainties and complexities in financial data sets can be better modeled. The emotional neural networks were successfully applied to credit scoring and evaluation [37], showing higher accuracy and lower computing time than the conventional neural models based on the back-propagation learning algorithm.

Another research direction refers to the application of ensembles of neural networks to credit scoring and bankruptcy prediction problems. For instance, West

et al. [75] investigated bagging and boosting with the MLP as the base classifier and found that the ensembles were superior to the single best model in most cases. Conversely, Tsai and Wu [72] concluded that the single best neural network is more suitable than multiple neural network classifiers. Yu et al. [78] employed bagging with three-layer back-propagation neural networks. Hung and Chen [30] proposed a selective ensemble of three classifiers (decision tree, back-propagation neural network and support vector machine) integrated with the concept of the expected probability, showing that it performs better than other stacking ensembles using the weighting or voting strategies. Tsai et al. [70] carried out an extensive comparison of ensembles using MLP, support vector machines and decision trees as the base classifiers for bagging and boosting, suggesting that boosting with decision trees perform significantly better than the other ensembles.

Apart from the many proposals for using different neural network architectures, we can find numerous comprehensive reviews and comparative studies related to credit risk and bankruptcy prediction where the neural network approaches have played an important role. Zhang et al. [80] reviewed the application of neural networks to predict corporate distress and studied their robustness in terms of sampling variability. Atiya [10] gave a thorough survey on the problem of bankruptcy prediction using neural networks and claimed the superiority of these models over other techniques. Verikas et al. [73] presented a comprehensive review of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction. One of the most complete benchmarking studies of classification algorithms in credit scoring is the one by Baesens et al. [13], which has further been extended with novel learning methods, performance measures and techniques to reliably compare different classifiers [47]. It is also interesting to highlight the paper by Brown and Mues [16], in which an exhaustive comparison of prediction techniques for imbalanced credit scoring data sets was presented.

## 3. Hybrid associative classifier with translation

In its most general form, an associative memory is a content-addressable neural network based on matrix algebra [39, 57] that maps input patterns (examples) to output patterns by using the $p$ different associated pattern pairs $(\mathbf{x}^k, \mathbf{y}^k)$ stored during the learning phase. The associative memory takes the form of a connection weight matrix $\mathbf{W} = [\mathbf{w_{i,j}}]_{\mathbf{m} \times \mathbf{n}}$ generated from a finite set of $p$ encoded associations, called fundamental set of associations, $\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \ldots, p\}$, where $\mathbf{x}^\mu \in \mathbb{R}^n$ are the fundamental input patterns of dimension $n$ and $\mathbf{y}^\mu \in \mathbb{R}^m$ are the the fundamental output patterns of dimension $m$. Then, $x_j^\mu$ and $y_i^\mu$ denote the $j$-th component of an input pattern $\mathbf{x}^\mu$ and the $i$-th component of an output pattern $\mathbf{y}^\mu$, respectively.

The associative memories can be of two types depending on the retrieved pattern: hetero-associative (e.g., lernmatrix and linear associator) and auto-associative (e.g., Hopfield network). A hetero-associative memory relates input patterns with output patterns of distinct nature and formats ($\mathbf{x}^\mu \neq \mathbf{y}^\mu$), while an auto-associative memory is a particular case where $\mathbf{x}^\mu = \mathbf{y}^\mu$ and $n = m$.

The HACT neural network is an associative memory that merges the learning (or encoding) phase of the linear associator with the recall (or decoding) phase of the Steinbuch's lernmatrix to exploit their strengths and improve the performance of the classifier. Basically the main advantages of HACT over its predecessors are: (i) the HACT model allows to operate with real-valued input patterns, while the lernmatrix only supports the binary values 0 and 1; and (ii) the input vectors are not required to be orthonormal, unlike those in the linear associator.

At this point it is worth stressing that the HACT approach cannot be deemed as a hybrid or combining prediction model in the sense of traditional hybridization. In general, hybrid learning methods are understood as systems that combine two or more different techniques in order to benefit from the synergistic effect between the individual components [71, 76]. For instance, a hybrid prediction model may consist of one unsupervised learner to pre-process the training data into homogeneous clusters and one supervised algorithm to build the classifier from the clustering result [29, 71], or it may use a feature selection strategy to choose the most relevant explanatory variables and then these are employed to design the predictor [44, 45, 56], or even it may be built from different cascading predictors in order to build an ensemble of classifiers [20, 21, 35]. However, as already described previously, the HACT model simply makes use of the fundamental ideas of two types of associative memories (one for the learning phase and the other for the recall phase), but there is not hybridization between them.

### 3.1. Learning phase

The learning phase of the HACT model, which is based on that of the linear associator, consists of constructing a matrix $\mathbf{W}$ such that when an input pattern $\mathbf{x}^\mu$ is presented, the stored pattern $\mathbf{y}^\mu$ associated with the input pattern is retrieved. This process comprises two basic steps:

1. For each association $(\mathbf{x}^\mu, \mathbf{y}^\mu)$ in the fundamental set, compute the outer product $\mathbf{y}^\mu (\mathbf{x}^\mu)^T$, where $(\mathbf{x}^\mu)^T$ is the transpose of the input vector $\mathbf{x}^\mu$.

2. Sum the $p$ outer products to obtain the matrix $\mathbf{W} = \alpha \sum_{\mu=1}^{p} \mathbf{y}^\mu (\mathbf{x}^\mu)^T$, where $\alpha$ is the normalizing constant (usually set to $1/p$). The $(i, j)$-th component of $\mathbf{W}$ is defined as $w_{i,j} = \sum_{\mu=1}^{p} y_i^\mu x_j^\mu$.

Unlike the standard hybrid associative memory, the HACT model incorporates an initial step in the learning phase, which consists of a translation of the coordinate axes to a new origin located at the centroid of the fundamental input patterns. The aim of shifting the fundamental set is the representation of the fundamental input patterns in a new $n$-dimensional space where patterns belonging to two different classes are located diametrically opposite to each other and the midpoint of the diameter is defined by the mean vector $\overline{x}$. This should allow for a better classification result because patterns of different classes will presumably be grouped quite far apart in different quadrants [63].

Let $A = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$ be a set of $n$-dimensional fundamental input patterns that belong to $m$ classes, and let $\hat{A} = \{\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \ldots, \hat{\mathbf{x}}^p\}$ be the corresponding set of fundamental input patterns that have been translated to the new origin $\overline{x}$. The implementation of the learning phase of HACT to construct the connection weight matrix $\mathbf{W}$ is described in Algorithm 1.

---

**Algorithm 1** HACT learning phase

---

1: $s \leftarrow 0$
2: **for all $\mathbf{x}^\mu \in A$ do**
3: $\quad s \leftarrow s + \mathbf{x}^\mu$
4: **end for**
5: $\overline{\mathbf{x}} \leftarrow s/p$ $\quad$ {$\overline{x}$ is the new origin of the coordinate axes}
6: **for all $\mathbf{x}^\mu \in A$ do**
7: $\quad \hat{\mathbf{x}}^\mu \leftarrow \mathbf{x}^\mu - \overline{\mathbf{x}}$ $\quad$ {Translate the input patterns to the new coordinate axes}
8: $\quad k \leftarrow 1$
$\quad\quad$ {Assign an output vector of size $m$ to patterns that belong to class $k$}
9: $\quad$ **while $k \leq m$ do**
10: $\quad\quad$ **if $class(\hat{\mathbf{x}}^\mu) = k$ then**
11: $\quad\quad\quad y_k^\mu = 1$
12: $\quad\quad$ **else**
13: $\quad\quad\quad y_k^\mu = 0$
14: $\quad\quad$ **end if**
15: $\quad$ **end while**
16: **end for**
17: $\mathbf{W} \leftarrow 0$
$\quad$ {Apply the learning phase of the linear associator}
18: **for all $(\hat{\mathbf{x}}^\mu, \mathbf{y}^\mu)$ do**
19: $\quad \mathbf{W} \leftarrow \mathbf{W} + (\mathbf{y}^\mu)(\hat{\mathbf{x}}^\mu)^T$
20: **end for**

---

### 3.2. Recall phase

Assume that the matrix $\mathbf{W}$ has been constructed by using Algorithm 1, then the classification of a new input pattern $\mathbf{x}$ will consist of two steps: (i) the application

of the same translation to $\mathbf{x}$ as that used in the learning phase (step 7) in order to obtain $\hat{\mathbf{x}}$, and (ii) the use of the recall phase of lernmatrix in order to assign $\mathbf{x}$ to a class.

The recall or decoding phase consists of determining the components of the vector $\mathbf{y}^\mu$ associated to a given input pattern $\mathbf{x}^\mu$. The $i$-th component $y_i^\mu$ of the class vector $\mathbf{y}^\mu$ is calculated by means of the following bipolar output function:

$$ y_i^\mu = \begin{cases} 1 & \text{if} \quad \sum_{j=1}^n w_{i,j}\hat{x}_j^\mu = \max_{h=1}^p \left[ \sum_{j=i}^n w_{h,j}\hat{x}_j^\mu \right] \\ 0 & \text{otherwise} \end{cases} \tag{1} $$

If a given input vector $\hat{\mathbf{x}}^\mu$ is assigned to class $k$, this expression leads to an $m$-dimensional output vector $\mathbf{y}^\mu$ with its $k$-th component equal to 1 ($y_k^\mu = 1$) and all the remaining components equal to zero ($y_j^\mu = 0$ for $j = 1, 2, \ldots, k-1, k+1, \ldots, m$).

It is worth pointing out that, despite an input pattern may contain errors and noise, the HACT model will still be able to retrieve the closest stored output pattern. Like most associative memories in general [52], the HACT network also shows a robust and fault tolerant behavior, which means that noise or errors will cause a certain decrease in generalization performance rather than a total degradation of the classifier effectiveness.

## 4. Experimental set-up

Nine data sets related to bankruptcy/creditworthiness have been employed in order to make a comprehensive comparison of the HACT model with four well-known neural networks (MLP, RBF, BN and the voted perceptron, VP), whose architectures and parameter settings are reported in Table 1. In addition, an SVM with a linear kernel (widely acknowledged as one of the best soft computing techniques) and the logit model (a classical econometric method) have also been included in this study. Note that, except for the HACT technique, the WEKA [27] and KEEL [3, 4] data mining and knowledge discovery suites have been used to conduct our experiments. Table 2 summarizes the main characteristics of the experimental databases, including the amount of explanatory variables with the number of categorical variables given in brackets, the percentage of default cases (a measure of imbalance in class distribution) and the Fisher's discriminant ratio (F1) whose value determines how well the two classes are separated from each other (the higher the value of F1, the easier the classification problem).

The Australian, German and Japanese data sets are from the UCI Machine Learning Database Repository (http://archive.ics.uci.edu/ml/). The Iranian data set comes from a modification to a corporate client database of a small

Table 1: Parameter values used in the prediction models

| | |
|---|---|
| MLP | Learning rate = 0.3; Momentum = 0.2; Training time = 500; Hidden layers = (variables + classes)/2 |
| RBF | Normalized Gaussian RBF; Center vectors of the functions determined using K-means clustering |
| BN | Initial count for estimating the conditional probability tables = 0.5; Naive Bayes network used as the initial structure; Bayesian Dirichlet score to evaluate the structure learned |
| VP | Linear kernel; Maximum number of alterations to the perceptron = 10000; Number of iterations = 1 |
| SVM | Linear kernel; Complexity = 1.0: Tolerance = 0.001; Epsilon = 1.0E-12 |
| Logit | Multinomial logistic regression; Quasi-Newton optimization method; Ridge value in the log-likelihood = 1.0E-8 |

private bank in Iran [60]. The Polish data set contains bankruptcy information of 120 companies recorded over a two-year period [54]. The SabiSPQ database [6] contains business information of 1180 companies whose accounts are placed on the Spanish Mercantile Registry. The Thomas data set [69] describes applicants for a credit product. The UCSD data set is a reduced version of a database used in the 2007 Data Mining Contest. The USA database [64] consists of the accounting statements from 8293 banks recorded by the Federal Deposit Insurance Corporation.

Table 2: Some characteristics of the data sets used in the experiments

| Data set | Variables | Default/non-default cases | % Default cases | F1 |
|---|---|---|---|---|
| Australian | 14(8) | 307/383 | 44.5 | 2.33 |
| German | 24(21) | 300/700 | 30.0 | 0.35 |
| Japanese | 15(9) | 296/357 | 45.3 | 2.54 |
| Iranian | 27(4) | 50/950 | 5.0 | 0.36 |
| Polish | 30(0) | 112/128 | 46.7 | 0.70 |
| SabiSPQ | 19(3) | 590/590 | 50.0 | 3.19 |
| Thomas | 12(4) | 323/902 | 26.4 | 0.18 |
| UCSD | 38(6) | 599/1836 | 24.6 | 0.42 |
| USA | 17(0) | 320/7973 | 3.9 | 0.94 |

The common way to assess the performance of financial distress prediction systems when databases are small or medium sized corresponds to $k$-fold cross-validation [26]. Accordingly, a 10-fold cross-validation has been adopted for the

experiments: each original data set has been randomly divided into ten stratified parts of equal (or approximately equal) size. For each fold, nine blocks have been pooled as the training data, and the remaining part has been employed as an independent test set. Stratification has been used to preserve the class proportions of the whole data set into each one of the subsets obtained by the sampling method, thus reducing the prior probability of data set shift and the variance in the estimation process [62]. The results from classifying the test samples using the training sets have been averaged across the ten runs and then evaluated for significant differences between models by means of statistical tests.

*4.1. Evaluation scores*

Standard performance evaluation scores in many financial applications as is the case of bankruptcy and credit risk prediction are usually calculated from a $2 \times 2$ confusion matrix (see Table 3), where each entry $(i, j)$ contains the number of correct/incorrect predictions.

Table 3: Confusion matrix for a binary classification problem

|  | Predicted as positive | Predicted as negative |
|---|---|---|
| Positive class | True Positive (TP) | False Negative (FN) |
| Negative class | False Positive (FP) | True Negative (TN) |

Most prediction systems often employ the accuracy as the criterion for performance evaluation. It represents the proportion of the correctly predicted cases (positive and negative), $Acc = (TP+TN)/(TP+FN+TN+FP)$, but it is strongly biased towards the majority class when data are skewed [33]. As bankruptcy and credit risk databases are commonly imbalanced, alternative measures have been used with the aim of obtaining a trade-off between the performance evaluation on both classes. A well-known example is the geometric mean of the true positive and true negative rates:

$$Gmean = \sqrt{TPR \cdot TNR} \qquad (2)$$

where $TPR = TP/(TP+FN)$ is the true positive rate (the percentage of defaulters that are correctly predicted) and $TNR = TN/(TN + FP)$ is the true negative rate (percentage of non-defaulters that are correctly classified).

Overall performance scores such as accuracy and geometric mean help decision-makers to compare firms or borrowers against each other, but neglecting the cost of different error types. Thus two of the most popular measures to predict corporate or credit default are the true positive and true negative rates, which concentrate only

on a part of the data and therefore, exhibit performance results on each class separately. This is especially important for this kind of financial applications because of the different misclassification costs associated to false positives and false negatives: the cost of predicting a defaulter as non-defaulter is generally much higher than the expected cost of false positives (non-defaulters classified as defaulters).

## 5. Experimental results and discussion

Table 4 reports the true positive rate averaged across the 10 runs for each database, the average values across all the databases and the Friedman's average rank for each neural network approach (the one with the lowest average rank has to be deemed as the best solution). The values for the best performing method in each database are underlined. Based on the Friedman's average ranks, the results reveal that the HACT model corresponds to the algorithm with the best performance, followed by MLP and the logit method. What is more interesting, however, is that the associative memory has been the best in 6 out of the 9 databases, demonstrating the benefits of applying this technique to the prediction of financial distress.

Table 4: True positive rate and Friedman's average rank

|            | HACT   | MLP    | BN     | RBF    | VP     | SVM    | Logit  |
|------------|--------|--------|--------|--------|--------|--------|--------|
| Australian | 92.50  | 80.78  | 81.09  | 72.31  | 89.60  | 92.53  | 86.97  |
| German     | 79.30  | 47.67  | 41.34  | 48.00  | 39.34  | 46.67  | 47.67  |
| Iranian    | 72.00  | 10.00  | 0.00   | 0.00   | 2.00   | 0.00   | 6.00   |
| Japanese   | 93.60  | 81.71  | 80.68  | 76.22  | 92.20  | 93.90  | 89.13  |
| Polish     | 80.45  | 73.39  | 75.05  | 68.85  | 66.89  | 75.23  | 72.42  |
| SabiSPQ    | 77.25  | 99.15  | 84.57  | 80.84  | 94.74  | 100    | 99.15  |
| Thomas     | 64.40  | 18.83  | 10.87  | 3.40   | 4.05   | 0.00   | 9.91   |
| UCSD       | 86.00  | 60.97  | 80.32  | 7.87   | 53.27  | 55.77  | 65.11  |
| USA        | 97.60  | 44.81  | 82.17  | 23.50  | 25.71  | 15.96  | 39.14  |
| Average    | 82.57  | 57.48  | 59.57  | 42.33  | 51.98  | 53.84  | 57.28  |
| Rank       | 1.8889 | 3.5556 | 4.2222 | 5.8889 | 4.8889 | 3.8889 | 3.6667 |

In order to make easier the comparison of the neural networks in terms of TPR, Figure 1 contains a spider plot showing the relationship between the true positive rate and the complexity of the problem: each radius represents a database and each star corresponds to a model. The databases have been placed in clockwise order from the most to the least complex problem based on the value of the percentage of default cases (a) and F1 (b). The performance of a model on a database is calculated as the distance from the center and therefore, a higher area corresponds to a better algorithm in terms of TPR. As can be seen, HACT has performed significantly

better than the other techniques especially with the most complex databases, while there have not been differences when applied to the "easiest" problems.



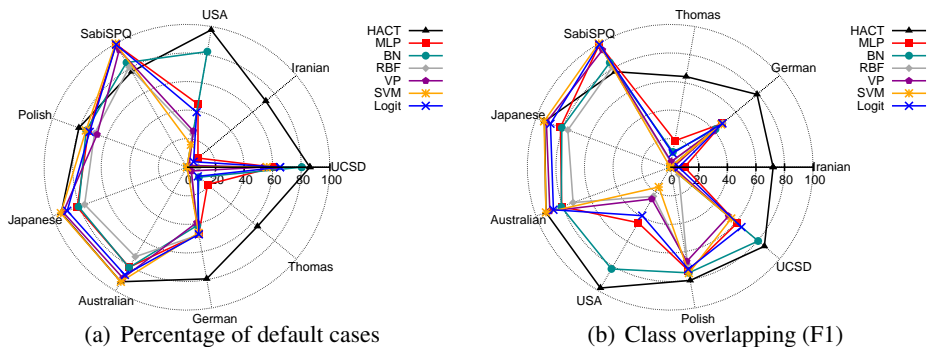(a) Percentage of default cases       (b) Class overlapping (F1)

Figure 1: Spider plot of true positive rates

Table 5 includes the true negative rates and the Friedman's average rank for each model. Again the best values in each database are underlined. It is interesting to compare these results with those in Table 4 with the purpose of better understanding the behavior of each method in function of class imbalance and overlapping. For instance, the only case where HACT has performed the worst in terms of TPR is the SabiSPQ database, which corresponds to a non-complex problem with perfectly balanced (50% of default cases) and well-separated classes (F1 = 3.19). Here logistic regression appears to be the best performing algorithm (TPR = 99.15, TNR = 99.49) since differences in true negative rates are not significant at all. Quite the contrary, when the classes are strongly imbalanced (Iranian, USA) and/or overlapped (German, Iranian, Thomas), the associative memory has made significantly better predictions on the default cases than any other method, as reflected by the high differences in the true positive rates. In these cases, the other techniques have in general achieved very high true negative rates (correct predictions on non-defaulters), but failed in the classification of the minority class. Three representative examples of this situation are the BN, RBF and SVM approaches applied to the Iranian database (5% of default samples and F1 = 0.36), which have reached a perfect classification of non-default cases (TNR $\simeq$ 100) while mispredicting the class of all default patterns (TPR = 0). Similar comments can be made with regard to the Thomas database (26.4% of defaulters and F1 = 0.18) where MLP, BN, RBF, VP, SVM and logit have achieved high true negative rates (94% – 100%) and very low true positive rates (0% – 19%), whereas the associative model has shown a significant trade-off between both performance rates.

With the aim of checking whether or not the TPR results are significantly dif-

13

Table 5: True negative rate and Friedman's average rank

| | HACT | MLP | BN | RBF | VP | SVM | Logit |
|---|---|---|---|---|---|---|---|
| Australian | 80.40 | 86.95 | 90.89 | 90.59 | 81.97 | 79.92 | 85.15 |
| German | 62.14 | 81.44 | 84.87 | 84.70 | 91.00 | 89.14 | 88.00 |
| Iranian | 56.40 | 98.50 | 99.47 | 100 | 99.89 | 100 | 99.16 |
| Japanese | 80.90 | 84.79 | 90.41 | 88.10 | 82.80 | 80.05 | 84.80 |
| Polish | 59.50 | 81.14 | 71.00 | 75.13 | 73.58 | 67.37 | 75.13 |
| SabiSPQ | 100 | 98.81 | 99.49 | 98.30 | 99.83 | 90.51 | 99.49 |
| Thomas | 47.00 | 94.02 | 97.47 | 97.47 | 97.69 | 100 | 97.12 |
| UCSD | 58.00 | 88.72 | 80.18 | 97.16 | 92.16 | 32.10 | 91.07 |
| USA | 53.70 | 99.08 | 93.19 | 99.26 | 99.68 | 99.91 | 99.41 |
| Average | 66.45 | 90.38 | 89.66 | 92.30 | 90.96 | 88.78 | 91.04 |
| Rank | 6.1111 | 4.5556 | 3.7778 | 3.0556 | 2.8889 | 3.9444 | 3.6667 |

ferent, the Iman-Davenport's statistic has been computed. This is distributed according to an $F$-distribution with $K - 1$ and $(K - 1)(N - 1)$ degrees of freedom, where $K$ denotes the number of models and $N$ is the total number of data sets. The $p$-value computed by $F(6, 48)$ was 0.002955595039, which is less than a significance level of $\alpha = 0.05$. Therefore, the null-hypothesis that all the prediction techniques perform equally well can be rejected. As the Iman-Davenport's statistic only allows to figure out differences among all methods, we have also carried on with the Holm's and Li's post hoc tests using the HACT model (the one with the lowest Friedman's rank) as the control algorithm.

Table 6: Post hoc comparisons over the results of Friedman's test for the true positive rates

| $i$ | Model | $z$ | $p$-value | Holm's | Li's |
|---|---|---|---|---|---|
| 6 | RBF | 3.927922 | 0.000086 | 0.008333 | 0.047279 |
| 5 | VP | 2.945942 | 0.00322 | 0.01 | 0.047279 |
| 4 | BN | 2.291288 | 0.021947 | 0.0125 | 0.047279 |
| 3 | SVM | 1.963961 | 0.049535 | 0.016667 | 0.047279 |
| 2 | Logit | 1.745743 | 0.080856 | 0.025 | 0.047279 |
| 1 | MLP | 1.636634 | 0.101707 | 0.05 | 0.05 |

Values in Table 6 show that the Holm's procedure rejects the null-hypothesis of equivalence for those methods that have an unadjusted $p$-value $\leq 0.0125$, proving that HACT has been significantly better than RBF and VP at a significance level of $\alpha = 0.05$. On the other hand, the Li's post hoc test rejects those hypotheses that have an unadjusted $p$-value $\leq 0.047279$, indicating that HACT has performed significantly better than RBF, VP and BN. Therefore, it is possible to conclude that

14

the associative memory is statistically equivalent to SVM, logit and MLP, while significantly better than the remainder of the neural networks in terms of the true positive rate.

For each database, Figure 2 displays all the prediction models in the space spanned by the true positive rate on the $x$-axis and the true negative rate on the $y$-axis. A method with perfect prediction will be located on the upper right corner (100% TPR, 100% TNR) of the plot. Therefore the closer the classifier is to the upper right corner, the higher the performance on both classes. However, in financial distress applications it is preferable not to miss a defaulter rather than a non-defaulter, which means that it is more important to maximize TPR (points close to the right side) than to maximize TNR (points close to the upper side).
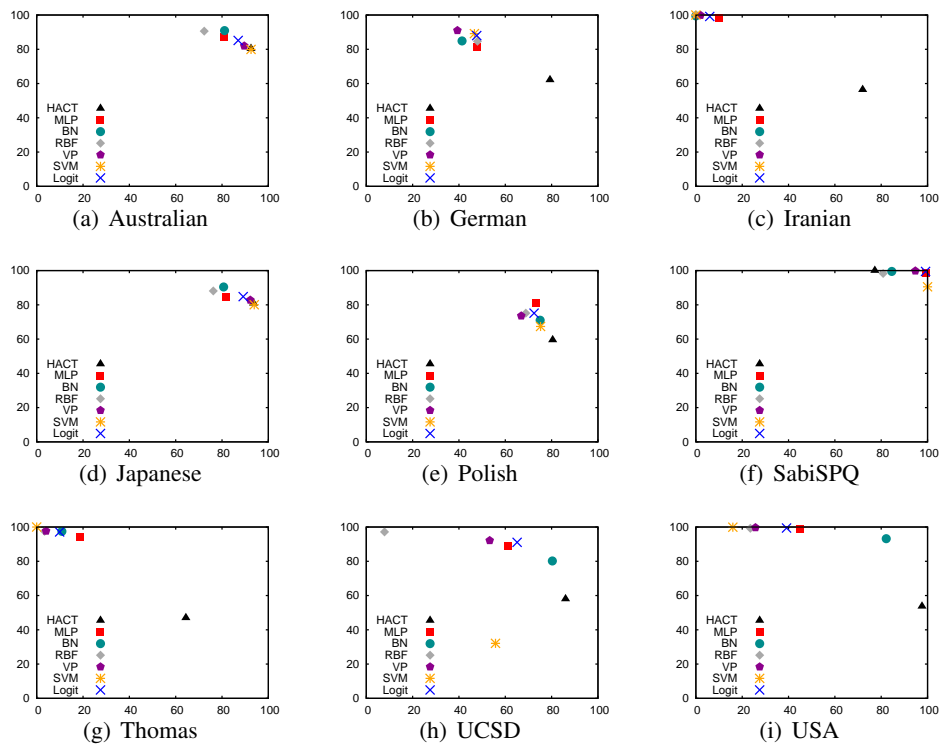


Figure 2: True positive rates versus true negative rates

In line with our previous findings, one can observe in Figure 2 that the associative memory mostly lies the closest to the right side of the chart and in general, it is not too far from the other models in terms of TNR. Very remarkable cases are the

15

Iranian and Thomas databases where all techniques except HACT are very close to the upper left corner of the plot, which reveals that they have misclassified almost all the default patterns. Note that the classes in these both databases are strongly imbalanced and highly overlapped. A similar behavior can be viewed on the German, UCSD and USA databases, which correspond to data with high/moderate imbalance and/or high overlapping between classes.

Table 7: Geometric mean and Friedman's average rank (the values for the best performing technique on each database are underlined)

|            | HACT   | MLP    | BN     | RBF    | VP     | SVM    | Logit  |
|------------|--------|--------|--------|--------|--------|--------|--------|
| Australian | 86.24  | 83.81  | 85.85  | 80.94  | 85.70  | 85.99  | 86.06  |
| German     | 70.20  | 62.31  | 59.23  | 63.76  | 59.83  | 64.50  | 64.77  |
| Iranian    | 63.72  | 31.38  | 0.00   | 0.00   | 14.13  | 0.00   | 24.39  |
| Japanese   | 87.02  | 83.24  | 85.41  | 81.94  | 87.37  | 86.70  | 86.94  |
| Polish     | 69.19  | 77.17  | 73.00  | 71.92  | 70.16  | 71.19  | 73.76  |
| SabiSPQ    | 87.89  | 98.98  | 91.73  | 89.14  | 97.25  | 95.14  | 99.32  |
| Thomas     | 55.02  | 42.08  | 32.55  | 18.20  | 19.89  | 0.00   | 31.02  |
| UCSD       | 70.63  | 73.55  | 80.25  | 27.65  | 70.07  | 71.67  | 77.00  |
| USA        | 72.40  | 66.63  | 87.51  | 48.30  | 50.62  | 39.93  | 62.89  |
| Average    | 73.59  | 68.79  | 66.17  | 53.54  | 61.67  | 57.24  | 67.29  |
| Rank       | 3.0000 | 3.3333 | 3.8889 | 5.8889 | 4.5556 | 4.7778 | 2.5556 |

The averaged geometric means of the performance on both classes in Table 7 show that the HACT method has achieved the highest balanced trade-off between the true positive and true negative rates. The associative memory has been the best in 4 out of the 9 databases, and it is still very close to the VP algorithm on the Japanese data. The fact that the probabilistic neural network, the RBF model and the SVM show a geometric mean of 0 on the very strongly skewed Iranian database is because these methods have obtained a TPR $= 0$ (i.e., they have predicted all cases as non-defaulters), as already discussed with the results in Tables 4 and 5.

The low Friedman's average rank of HACT in Table 7 corroborates that this neural model is also one of the best performing methods in terms of an overall score. In this case, however, logistic regression shows the lowest Friedman's average rank, while the MLP and BN models are very close to HACT. Finally, the RBF, SVM and VP approaches have been the techniques with the poorest performances, as already seen with the Friedman's ranks calculated on the true positive rate.

## 6. Conclusions and future work

From the first works in the beginning of the 1990's, the artificial neural networks emerged as an effective method for bankruptcy and credit risk prediction. They differ from classical financial prediction systems, such as the models based on statistical techniques, mainly in their black-box nature and in the assumption of a non-linear relation among variables. In this paper, the hybrid associative memory with translation has been explored and compared to other well-known neural models (MLP, RBF, BN and VP), a generally well-performing soft computing technique (SVM) and a common econometric model (logit).

The experimental results over nine real-life financial databases suggest that the associative memories can be an appropriate approach to prediction of financial distress, especially in the case of databases where the classes are strongly imbalanced and/or overlapped. The HACT neural network has obtained the highest true positive rates, which means that this model predicts the default cases better than the remainder of the methods here analyzed. When evaluated with the geometric mean of both rates, HACT has still shown to be an appropriate solution, revealing that the degradation of the true negative rate has not been abrupt.

There are several very interesting open research questions to work on. For instance, a more thorough and extensive analysis on the HACT model with a large pool of prediction techniques such as instance-based learners and decision trees deserves further consideration. Another avenue for future research is to study the performance of other well-known associative memories (e.g., the Hopfield network and the bidirectional associative memory) when applied to the prediction of financial distress. Other issues include to investigate the potential of ensembles of associative memories, both using these as the only base classifier and also together with other prediction models.

## References

[1] H. Abdou, J. Pointon, A. El-Masry, Neural nets versus conventional techniques in credit scoring in Egyptian banking, Expert Syst. Appl. 35 (2008) 1275–1292.

[2] M.E. Acevedo-Mosqueda, C. Yáñez-Márquez, I. López-Yáñez, Alpha–beta bidirectional associative memories: theory and applications, Neural Process. Lett. 26 (2007) 1–40.

[3] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework, J. Mult-Valued Log. S. 17 (2011) 255–287.

[4] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera, KEEL: A software tool to assess evolutionary algorithms to data mining problems, Soft Comput. 13 (2009) 307–318.

[5] M. Aldape-Pérez, C. Yáñez-Márquez, O. Camacho-Nieto, A.J. Argüelles-Cruz, An associative memory approach to medical decision support systems, Comput. Meth. Prog. Bio. 106 (2012) 287–307.

[6] E. Alfaro, N. García, M. Gámez, D. Elizondo, Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks, Decis. Support Syst. 45 (2008) 110–122.

[7] E. Altman, Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy, J. Financ. 23 (1968) 589–609.

[8] J.A. Anderson, A simple neural network generating an interactive memory, Math. Biosci. 14 (1972) 197–220.

[9] E. Angelini, G. di Tollo, A. Roli, A neural network approach for credit risk evaluation, Q. Rev. Econ. Financ. 48 (2008) 733–755.

[10] A.F. Atiya, Bankruptcy prediction for credit risk using neural networks: A survey and new results, IEEE T. Neural Networ. 12 (2001) 929–935.

[11] J. Baek, S. Cho, Bankruptcy prediction for credit risk using an auto-associative neural network in Korean firms, in: Proc. IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong, 2003, pp. 25–29.

[12] B. Baesens, M. Egmont-Petersen, R. Castelo, J. Vanthienen, Learning Bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search, in: Proc. 16th International Conference on Pattern Recognition, volume 3, Quebec City, Canada, 2002, pp. 49–52.

[13] B. Baesens, T.V. Gestel, S. Viaene, M. Stepanova, J. Suykens, J. Vanthienen, Benchmarking state-of-the-art classification algorithms for credit scoring, J. Oper. Res. Soc. 54 (2003) 627–635.

[14] T.R. Bielecki, M. Rutkowski, Credit Risk: Modeling, Valuation and Hedging, Springer, Berlin, Germany, 2004.

[15] P.L. Brockett, L.L. Golden, J. Jang, C.C. Yang, A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction, J. Risk Insur. 73 (2006) 397–419.

[16] I. Brown, C. Mues, An experimental comparison of classification algorithms for imbalanced credit scoring data sets, Expert Syst. Appl. 39 (2012) 3446–3453.

[17] H.C. Chang, H.C. Chen, J.H. Fang, Lithology determination from well logs with fuzzy associative memory neural network, IEEE T. Geosci. Remote 35 (1997) 773–780.

[18] K.C. Chang, R. Fung, A. Lucas, R. Oliver, N. Shikalo, Bayesian networks applied to credit scoring, IEEE T. Geosci. Remote 1 (2000) 1–18.

[19] W.S. Chen, Y.K. Du, Using neural networks and data mining techniques for the financial distress prediction model, Expert Syst. Appl. 36 (2009) 4075–4086.

[20] C.B. Cheng, C.L. Chen, C.J. Fu, Financial distress prediction by a radial basis function network with logit analysis learning, Comput. Math. Appl. 51 (2006) 579–588.

[21] C.L. Chuang, R.H. Lin, Constructing a reassigning credit scoring model, Expert Syst. Appl. 36 (2009) 1685–1694.

[22] V.S. Desai, J.N. Crook, G.A. Overstreet Jr., A comparison of neural networks and linear scoring models in the credit union environment, Eur. J. Oper. Res. 95 (1996) 24–37.

[23] Y. Ding, X. Song, Y. Zen, Forecasting financial condition of Chinese listed companies based on support vector machine, Expert Syst. Appl. 34 (2008) 3081–3089.

[24] B.E. Erdogan, Prediction of bankruptcy using support vector machines: an application to bank bankruptcy, J. Stat. Comput. Sim. 83 (2013) 1543–1555.

[25] M. Ezzamel, C. Mar-Molinero, A. Beech, On the distributional properties of financial ratios, J. Bus. Finan. Account. 14 (1987) 463–481.

[26] V. García, A.I. Marqués, J.S. Sánchez, An insight into the experimental design for credit risk and corporate bankruptcy prediction systems, J. Intell. Inf. Syst. 44 (2015) 159–189.

[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. Newsl. 11 (2009) 10–18.

[28] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, P. Natl. Acad. Sci. USA 79 (1982) 2554–2558.

[29] N.C. Hsieh, Hybrid mining approach in the design of credit scoring models, Expert Syst. Appl. 28 (2005) 655–665.

[30] C. Hung, J.H. Chen, A selective ensemble based on expected probabilities for bankruptcy prediction, Expert Syst. Appl. 36 (2009) 5297–5303.

[31] M. Inal, Y. Fatihoglu, Self organizing map and associative memory model hybrid classifier for speaker recognition, in: Proc. 6th Seminar on Neural Network Applications in Electrical Engineering, Belgrade, Yugoslavia, 2002, pp. 71–74.

[32] S. Itkar, U. Kulkarni, Efficient frequent pattern mining using auto-associative memory neural network, Brit. J. Appl. Sci. Technol. 4 (2014) 3160–3178.

[33] N. Japkowicz, Assessment metrics for imbalanced learning, in: Imbalanced Learning: Foundations, Algorithms, and Applications, John Wiley & Sons, Inc., Hoboken, NJ, 2013, pp. 187–210.

[34] G.V. Karels, A.J. Prakash, Multivariate normality and forecasting of business bankruptcy, J. Bus. Finan. Account. 14 (1987) 573–593.

[35] M. Khashei, M.T. Rezvan, A.Z. Hamadani, M. Bijari, A bi-level neural-based fuzzy classification approach for credit scoring problemss, Complexity 18 (2013) 46–57.

[36] A. Khashman, Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes, Expert Syst. Appl. 37 (2010) 6233–6239.

[37] A. Khashman, Credit risk evaluation using neural networks: Emotional versus conventional models, Appl. Soft Comput. 11 (2011) 5477–5484.

[38] T. Kohonen, Correlation matrix memories, IEEE T. Comput. C-21 (1972) 353–359.

[39] T. Kohonen, Associative Memory. A System-Theoretical Approach, Springer-Verlag, Berlin, Germany, 1977.

[40] T. Kohonen, M. Ruohonen, Representation of associated data by matrix operator, IEEE T. Comput. C-22 (1973) 701–702.

[41] B. Kosko, Adaptive bidirectional associative memories, Appl. Optics 26 (1987) 4947–4960.

[42] B. Kosko, Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice-Hall, Inc., Upper Saddle River, NJ, 1992.

[43] R.C. Lacher, P.K. Coats, S.C. Sharma, L.F. Fant, A neural network for classifying the financial health of a firm, Eur. J. Oper. Res. 85 (1995) 53–65.

[44] T.S. Lee, I.F. Chen, A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines, Expert Syst. Appl. 28 (2005) 743–752.

[45] T.S. Lee, C.C. Chiu, C.J. Lu, I.F. Chen, Credit scoring using the hybrid neural discriminant technique, Expert Syst. Appl. 23 (2002) 245–254.

[46] C. Leong, Credit risk scoring with Bayesian network models, Comput. Econ. (2015) 1–24.

[47] S. Lessmann, B. Baesens, H.V. Seow, L.C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, Eur. J. Oper. Res. 247 (2015) 124–136.

[48] T.H. Lin, A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models, Neurocomputing 72 (2009) 3507–3516.

[49] X. Mu, M. Artiklar, P. Watta, M. Hassoun, An RCE-based associative memory with application to human face recognition, Neural Process. Lett. 23 (2006) 257–271.

[50] M. Namba, Z. Zhang, Cellular neural network for associative memory and its application to Braille image recognition, in: Proc. International Joint Conference on Neural Networks, Vancouver, Canada, 2006, pp. 2409–2414.

[51] M. Odom, R. Sharda, A neural network model for bankruptcy prediction, in: Proc. IEEE International Joint Conference on Neural Networks, Washington, DC, 1990, pp. 163–168.

[52] G. Palm, On associative memory, Biol. Cybern. 36 (1980) 19–31.

[53] P.C. Pendharkar, A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem, Comput. Oper. Res. 32 (2005) 2561–2582.

[54] W. Pietruszkiewicz, Dynamical systems and nonlinear Kalman filtering applied in classification, in: Proc. 7th IEEE International Conference on Cybernetic Intelligent Systems, London, UK, 2008, pp. 263–268.

[55] V. Ravi, H. Kurniawan, P.N.K. Thai, P.R. Kumar, Soft computing system for bank performance prediction, Appl. Soft Comput. 8 (2008) 305–315.

[56] V. Ravi, C. Pramodh, Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks, Appl. Soft Comput. 8 (2008) 1539–1548.

[57] G.X. Ritter, P. Sussner, Associative memories based on lattice algebra, in: Proc. IEEE Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, volume 4, Orlando, FL, 1997, pp. 3570–3575.

[58] G.X. Ritter, P. Sussner, J.L. Diaz-de Leon, Morphological associative memories, IEEE T. Neural Networ. 9 (1998) 281–293.

[59] M. Sabourin, A. Mitiche, Modeling and classification of shape using a Kohonen associative memory with selective multiresolution, Neural Networks 6 (1993) 275–283.

[60] H. Sabzevari, M. Soleymani, E. Noorbakhsh, A comparison between statistical and data mining methods for credit scoring in case of limited available data, in: Proc. 3rd CRC Credit Scoring Conference, Edinburgh, UK, 2007.

[61] L.M. Salchenberger, E.M. Cinar, N.A. Lash, Neural networks: A new tool for predicting thrift failures, Decision Sci. 23 (1992) 899–916.

[62] G. Santafe, I. Inza, J.A. Lozano, Dealing with the evaluation of supervised classification algorithms, Artif. Intell. Rev. (2015) 1–42.

[63] R. Santiago-Montero, Clasificador Híbrido de Patrones basado en la Lern-matrix de Steinbuch y el Linear Associator de Anderson-Kohonen, Master's thesis, Centro de Investigación en Computación - Instituto Politécnico Nacional, 2003. In Spanish.

[64] C. Serrano-Cinca, B. Gutiérrez-Nieto, Partial least square discriminant analysis for bankruptcy prediction, Decis. Support Syst. 54 (2013) 1245–1255.

[65] K.S. Shin, T.S. Lee, H. jung Kim, An application of support vector machines in bankruptcy prediction model, Expert Syst. Appl. 28 (2005) 127–135.

[66] K. Steinbuch, Die lernmatrix, Kybernetik 1 (1961) 36–45.

[67] J. Sun, H. Li, Financial distress prediction using support vector machines: Ensemble vs. individual, Appl. Soft Comput. 12 (2012) 2254–2265.

[68] K.Y. Tam, M.Y. Kiang, Managerial applications of neural networks: The case of bank failure predictions, Manage. Sci. 38 (1992) 926–947.

[69] L.C. Thomas, D.B. Edelman, J.N. Crook, Credit Scoring and Its Applications, SIAM, Philadelphia, PA, 2002.

[70] C.F. Tsai, Y.F. Hsu, D.C. Yen, A comparative study of classifier ensembles for bankruptcy prediction, Appl. Soft Comput. 24 (2014) 977–984.

[71] C.F. Tsai, C. Hung, Modeling credit scoring using neural network ensembles, Kybernetes 43 (2014) 1114–1123.

[72] C.F. Tsai, J.W. Wu, Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Syst. Appl. 34 (2008) 2639–2649.

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

[74] D. West, Neural network credit scoring models, Comput. Oper. Res. 27 (2000) 1131–1152.

[75] D. West, S. Dellana, J. Qian, Neural network ensemble strategies for financial decision applications, Comput. Oper. Res. 32 (2005) 2543–2559.

[76] M. Woźniak, Hybrid Classifiers: Methods of Data, Knowledge, and Classifier Combination, Springer, Berlin, Germany, 2014.

[77] C. Yáñez-Márquez, Memorias Asociativas Basadas en Relaciones de Orden y Operaciones Binarias, Master's thesis, Centro de Investigación en Computación - Instituto Politécnico Nacional, 2002. In Spanish.

[78] L. Yu, S. Wang, K.K. Lai, Credit risk assessment with a multistage neural network ensemble learning approach, Expert Syst. Appl. 34 (2008) 1434–1444.

[79] B.L. Zhang, H. Zhang, S.S. Ge, Face recognition by applying wavelet sub-band representation and kernel associative memory, IEEE T. Neural Networ. 15 (2004) 166–177.

[80] G. Zhang, M.Y. Hu, B.E. Patuwo, D.C. Indro, Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, Eur. J. Oper. Res. 116 (1999) 16–32.

[81] Z. Zhang, H. Zhuo, S. Liu, P. de B. Harrington, Classification of cancer patients based on elemental contents of serums using bidirectional associative memory networks, Anal. Chim. Acta 436 (2001) 281–291.