Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey

Antanas Verikas · Zivile Kalsyte · Marija Bacauskiene · Adas Gelzinis

Abstract This paper presents a comprehensive review of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction. A variety of soft computing techniques are being applied to bankruptcy prediction. Our focus is on techniques, namely how different techniques are combined, but not on obtained results. Almost all authors demonstrate that the technique they propose outperforms some other methods chosen for the comparison. However, due to different data sets used by different authors and bearing in mind the fact that confidence intervals for the prediction accuracies are seldom provided, fair comparison of results obtained by different authors is hardly possible. Simulations covering a large variety of techniques and data sets are needed for a fair comparison. We call a technique hybrid if several soft computing approaches are applied in the analysis and only one predictor is used to make the final prediction. In contrast, outputs of several predictors are combined, to obtain an ensemble-based prediction.

A. Verikas (✉)
Intelligent Systems Laboratory, Halmstad University,
Box 823, 301 18 Halmstad, Sweden
e-mail: antanas.verikas@hh.se

A. Verikas · Z. Kalsyte · M. Bacauskiene · A. Gelzinis
Department of Electrical and Control Instrumentation,
Kaunas University of Technology, Studentu 50, 51368 Kaunas,
Lithuania
e-mail: kalsyte@yahoo.com

M. Bacauskiene
e-mail: marija.bacauskiene@ktu.lt

A. Gelzinis
e-mail: adas.gelzinis@ktu.lt

Keywords Bankruptcy prediction · Ensemble · Committee · SVM · Neural network · Fuzzy sets · Decision trees · Case-based reasoning · Genetic algorithms · Rough sets · Hybrid techniques

1 Introduction

There is a large number of application areas of soft computing techniques in finances. Portfolio management, credit scoring, bankruptcy prediction, prediction of currency exchange rate, decision support systems for stock trading, and currency crises prediction are example application areas. Mochon et al. (2008) discuss a rationale of using soft computing techniques in finance and present a short introduction into several application areas.

A large body of soft computing applications in finances concern bankruptcy prediction and a large variety of soft computing techniques have been applied to bankruptcy prediction. Multilayer perceptron (MLP), radial basis function (RBF) networks, self-organizing maps (SOM), learning vector quantization (LVQ), support vector machines (SVM), relevance vector machines (RVM) (Ribeiro et al. 2006), probabilistic neural networks (PNN), decision trees (DT), Bayesian networks (BN), fuzzy decision trees (FDT), case-based reasoning (CBR), fuzzy logic (FL), rough sets (RS), genetic algorithms (GA), hybrid systems, and ensembles of predictors comprise a list of the most popular techniques applied.

We make a distinction between a hybrid system and an ensemble of predictors. We say a system is hybrid if several soft computing approaches are exploited for data analysis, but only one single predictor is applied to make a final decision. To obtain a final decision in an ensemble, outputs of several predictors are aggregated in one way or

another. Supervised learning is used to train a predictor (to estimate parameters of a predictor). It is worth mentioning, however, that in some cases there is no clear distinction between hybrid and ensemble-based systems. Suppose that we create a bankruptcy prediction system by combining a logistic regression (LR) model and an MLP. Let us assume that the LR output is used as an additional input to the MLP and the final prediction is made by the MLP. We call such a system hybrid. We can also create a system by training both LR and MLP first and then combining them, via weighted averaging for example. We call such a system ensemble-based. Tough, according to the definition given above, distinction between these two systems is not very evident, the distinction can be easily made for the vast majority of the reviewed papers.

There is a huge number of examples demonstrating that hybrid and ensemble-based systems, when properly designed, outperform a one predictor based system designed for solving a classification task. Therefore, our focus is on such type of techniques. We do not present any description of widely used techniques. However, a short description of some not so widely known aspects is given for the article to be self-explaining.

2 Previous reviews on soft computing techniques in finance

Reviews of past literature concerning soft computing techniques in business, financial engineering, and specifically bankruptcy prediction are available.

Artificial neural networks is one of the most popular soft computing tools used in financial engineering. A rather comprehensive review of past literature on neural network applications in business can be found in Wong et al. (1997, 2000) and Vellido et al. (1999). A review by Wong et al. (1997) covers journal articles published during 1988–1995. The following application areas were distinguished: accounting/auditing, finance, human resources, information systems, marketing/distribution, production/operations, and others. The area of finance is represented by 54 articles, several of them in the field of bankruptcy prediction. The authors emphasize that neural networks are often integrated with expert systems. Wong et al. (2000) review 302 journal articles published during 1994–1998. A significant decrease of publications in 1998 was observed when compared to three previous years. The articles are grouped into the same application areas as in Wong et al. (1997). There are 67 articles in the finance area covering more than 50 topics. The authors foresee that production/operations and finance will remain the most common research areas, concerning neural network applications in business, in the future. A survey by Vellido et al. (1999) covers the period 1992–

1998. The main areas covered by the survey are: accounting/auditing, finance, management, marketing, production, and others. The area of finance is mainly represented by bankruptcy prediction and credit evaluation. An MLP is the most frequently used network in all the areas. The authors emphasize that only a few studies concern integration of several models for predicting bankruptcy. Integration of neural networks within more general systems, like decision support systems or expert systems, is mentioned. It is emphasized that the disparity of sample sizes is very big in different studies, there are studies carried out with as few as 36 cases.

Zhang and Zhou (2004) discuss the main, financial applications specific, data mining issues and compare several data mining techniques from the financial applications prospective. The authors group existing applications of data mining in finance into the following six categories: prediction of stock market, portfolio management, bankruptcy prediction, foreign exchange market, fraud detection, and others. Five data mining techniques, namely, neural networks, genetic algorithms, statistical inference, rule induction, and data visualization are discussed. The study demonstrates that each technique is used in all the six categories of applications. Choice of data mining methods and suitable values of parameters governing the behaviour of the methods, scalability and performance, unbalanced frequencies of financial data, text mining, mobile finance, integration of multiple data mining techniques, heterogeneous and distributed data sources are identified as challenges and emerging trends for future research.

Refenes et al. (1997) present a review and guidelines for using neural networks in financial engineering. The paper describes a set of typical applications in financial engineering as well as a number of alternative ways to select features. Issues of dealing with non-stationary data, handling leverages in data sets, testing for misspecified models are also discussed in the paper.

Zhang et al. (1999) reviewed neural network applications in bankruptcy prediction. The authors point out that there are empirical studies showing that the performance of neural networks is not always superior to conventional statistical techniques. Herewith, the authors stress that in most studies, commercial neural network tools are used without clear understanding of the sensitivity of solutions to initial conditions. By applying a k-fold cross-validation and using a sample of 220 firms, the authors studied the robustness of neural networks in predicting bankruptcy in terms of sampling variability. The significantly better performance of neural networks than LR models was reported. Atiya (2001) also reviewed the applications of neural networks to predict bankruptcy. The author thoroughly discusses the financial ratios used by Altman (1968)

and stresses that these ratios are widely used as input features even for neural networks and other non-linear models. It is emphasized that though a prediction of a binary bankruptcy event is very useful, an estimate of the bankruptcy probability is very desirable. One more important issue, according to Atiya, is to consider macroeconomic indicators as input features to the neural network.

Though not related directly to financial applications, two useful reviews, regarding the use of neural networks for solving various prediction and classification problems, can be found in Zhang (2007) and Zhang et al. (1998). In a recent paper, Zhang (2007) discusses the most common pitfalls in using neural networks and suggests guidelines for practitioners. The non-linear non-parametric nature of neural networks and the lack of a uniform standard for designing neural network models are identified as two major factors contributing to pitfalls in neural network applications. The most common pitfalls occur in model building, model selection and comparison, due to overfitting and underfitting, small sample size, due to treating neural networks as totally unexplainable "black boxes". A comprehensive review on forecasting with neural networks can be found in Zhang et al. (1998). The authors focus on common modeling issues such as neural network architecture, training algorithm, data, performance measures.

A review of past works on the use of knowledge-based decision support systems (KBDSS) in financial management can be found in Zopounidis et al. (1997). A KBDSS is obtained by combining a decision support system (DSS) and an expert system (ES). The implementation of DSS and ES in different fields of financial engineering, such as financial planning, portfolio management, accounting, financial analysis, assessment of bankruptcy risk, is discussed first and limitations of these two approaches are identified. Then, the authors describe several examples of KBDSSs proposed for: stock portfolio selection and management, lending analysis, analysis of credit granting problems, and financial analysis. The authors argue that KBDSSs improve the decision-making process qualitatively by facilitating the understanding of the operation and the results of the system, ensuring the objectiveness and the completeness of the results, achieving the proper structuring of the decision analysis.

Rada (2008) has recently reviewed papers related to applications of expert systems and evolutionary computing in finance published in the *"Expert Systems with Applications"* journal. The review has shown that in the early 1990s authors were more apt to use expert systems tools, while in the mid-2000s evolutionary computation tools prevail. Regarding the financial application area, unexpectedly, in both periods financial accounting was more common than investing in stocks. The integration of the earlier knowledge-based techniques with the more recent

developments in evolutionary computing is foreseen as a promising research direction.

A chapter, written by Chalup and Mitschele (2008), of a handbook on information technology in finance presents a brief overview of kernel methods in finance. Dimensionality reduction, introduction to classification and regression, selection of kernel parameters, and survey of applications in finance are the issues considered in the chapter. Concerning dimensionality reduction, PCA, multidimensional scaling (MDS), kernel PCA, and Isomap are briefly described. The surveyed applications of kernel methods in finance are categorized into credit risk management and market risk management. The authors emphasize the potential of non-linear dimensionality reduction techniques in the analysis of financial data.

The list of business failure-related literature presented in Dimitras et al. (1996) contains 158 journal articles published in the period 1932–1994. The review, however, is limited to 47 articles presenting models and related to industrial and retail applications. The articles are classified according to industrial sector, financial ratios, and models or methods applied. The methods applied are categorized into eight groups: discriminant analysis, linear probability model, probit analysis, logit analysis, recursive partitioning algorithm, survival analysis, univariate analysis, and expert systems. There are 79 financial ratios identified and grouped into three categories: (1) profitability ratios, (2) managerial performance ratios, and (3) solvency ratios. The authors make a conclusion that the discriminant analysis is the most frequently used method and the most important financial ratios belong to the solvency category. A trend on using non-financial and qualitative variables, in addition to financial ratios, is also mentioned.

Dimitras et al. (1999) discussed the merits of rough sets and proposed an approach to bankruptcy prediction based on rough sets. The technique provides a set of decision rules used to discriminate between healthy and failing companies. The authors argue that the decision rules take into account the preferences of the decision maker and the technique discovers a relevant subset of features (financial characteristics) revealing all important relationships between "the image of a firm and its risk of failure". The rough sets-based approach outperformed the classical discriminant analysis and the logit analysis. The authors argue that transparency of decisions expressed in the form of decision rules and the possibility of using both quantitative and qualitative features make the rough sets approach superior over other existing methods.

As it has already been mentioned, we do not discuss stand-alone soft computing techniques, in this paper. A recent comprehensive review of intelligent and some statistical techniques applied to bankruptcy prediction can be found in Kumar and Ravi (2007). The intelligent

techniques are categorized into the following groups: fuzzy set theory, neural networks, support vector machines, decision trees, rough sets, case-based reasoning, data envelopment analysis, and hybrid. The general observation is that a majority of papers use many financial ratios as input features and only a few of the reviewed papers use the Altman's features. One more observation is that in majority of the studies, MLP outperformed other techniques, while SVM outperformed other techniques and MLP too. The sensitivity of the rough sets-based techniques to changes in data was pointed out. In general, ensembles outperformed individual models and a trend is towards using hybrid intelligent systems. Tough in a separate section of that paper, the authors discuss hybrid techniques, however, only 14 papers proposing such techniques were covered in the review. Moreover, much work has been done in this area since 2005.

3 Data preprocessing

Apart from data normalization, feature extraction, feature selection, and clustering are the main data preprocessing issues considered in the literature related to bankruptcy prediction.

A large number of features can be usually measured in various applications. Not all of the features, however, are equally important for a specific task. Some of the features may be redundant or even irrelevant. Usually better performance may be achieved by discarding such features (Fukunaga 1972). Moreover, as the number of features used grows, the number of training samples required grows exponentially (Duda et al. 2001). Therefore, in many practical applications we need to reduce the dimensionality of the data.

3.1 Feature extraction

Feature extraction aims at finding a mapping that reduces the dimensionality of the data being classified. The mapping found projects the N-dimensional data onto the M-dimensional space, where $M < N$. Mapping techniques can be categorized as being linear or non-linear. There are many methods of both types. Principal component analysis (PCA) (Bishop 2006), linear discriminant analysis (LDA) (Fukunaga 1972), classical MDS (Borg and Groenen 1997), and non-negative matrix factorization (NMF) (Lee and Seung 1999) are prominent linear techniques of feature extraction. These techniques attempt to reduce the dimensionality of the data by creating new features that are linear combinations of the original ones.

While PCA and LDA still remains the most popular linear dimensionality reduction techniques applied to

bankruptcy prediction data (Shin and Kilic 2006; Ravi et al. 2008; Ravi and Pramodh 2008), NMF is used in analysis of financial data with increasing frequency. Unlike PCA, NMF learns parts-based data representations. This occurs due to the non-negativity constrains allowing only additive, but not subtractive, combinations of the original data. Drakakis et al. (2008) have recently applied NMF to the problem of revealing underlying trends in the Dow Jones stock market data. The study demonstrated the ability of the method to cluster stocks in performance-based clusters. Szupiluk et al. (2007) applied NMF to integrate information from several models predicting the customer's behaviour.

Kernel principal component analysis (Shawe-Taylor and Cristianini 2004), Isomap (Tenenbaum et al. 2000), data-driven high-dimensional scaling (Lespinats et al. 2007), Sammon mapping (Sammon 1969), generative topographic mapping (Bishop et al. 1998), self-organizing maps (Kohonen 1990), curvilinear component analysis (CCA) (Demartines and Herault 1997; Lee et al. 2004), stochastic neighbor embedding (Hinton and Roweis 2003), locally linear embedding (Roweis and Saul 2000), kernel discriminant analysis (Shawe-Taylor and Cristianini 2004), and "autoencoder" (Hinton and Salakhutdinov 2006; Cottrell 2006) are prominent non-linear mapping techniques. Apart from SOM and kernel PCA, Isomap is also used in the analysis of bankruptcy data. Isomap builds on the classical MDS but seeks to preserve the so-called geodesic distances, instead of Euclidean distances preserved by the classical MDS. Ribeiro et al. (2008) have recently proposed using the supervised Isomap to distinguish between the distressed and healthy companies. Despite much fewer dimensions used by the Isomap, the achieved classification accuracy was comparable with the accuracy obtained from SVM and RVM. Lawrence has recently proposed a very promising, Gaussian process-based, non-linear mapping technique called Gaussian process latent variable models (GP-LVM) (Lawrence 2004, 2005). An extension of GP-LVM for classification was also developed recently (Urtasun and Darrell 2007). Like SOM and CCA, GP-LVM can be trained to exhibit the property of local distance preservation when mapping high-dimensional data onto a low-dimensional space (Lawrence and Quinonero-Candela 2006). Local data ordering in a low-dimensional space is a very useful property for exploring high-dimensional data.

3.2 Feature selection

Feature selection is a special case of feature extraction. Employing feature extraction all N measurements are used for obtaining the M-dimensional data. Therefore, all N features need to be obtained. Feature selection, in contrast,

enables us to discard $(N - M)$ irrelevant features. Hence, by collecting only relevant features, the cost of future data collecting may by reduced. Feature selection in general is a difficult problem. In a general case, only an exhaustive search can guarantee an optimal solution. The branch and bound algorithm (Narendra and Fukunaga 1977) can also guarantee an optimal solution, if the monotonicity constraint imposed on a criterion function used to assess the quality of a feature subset is fulfilled. A large variety of feature selection techniques that result in a suboptimal feature subset have been proposed (Kudo and Sklansky 2000; Verikas and Bacauskiene 2002). Genetic algorithms (Abdelwahed and Amir 2005; Ignizio and Soltys 1996; Wallrafen et al. 1996; Min et al. 2006; Ahn et al. 2006; Yeung et al. 2007) and rough sets (Zhou and Tian 2007; Ahn et al. 2000; McKee and Lensberg 2002) are the two most popular approaches to feature selection in hybrid and ensemble-based techniques for bankruptcy prediction. Classification accuracy is the most often used criterion to assess the quality of a subset of features in the selection process. However, criteria not related directly to the classification accuracy, like mutual information (Chan et al. 2006), are also used to assess the quality of a feature subset.

3.3 Clustering

Yao (2007) aiming to increase the bankruptcy prediction accuracy and to facilitate the SVM design, preprocesses data by Fuzzy C-Means (FCM) clustering and principal component analysis (PCA). A cascade FCM-PCA-SVM is trained and used to predict financial crises in Chinese companies. Ravi and Pramodh (2008) suggested using the so-called principal component neural network (PCNN). The network resembles the radial basis function network, with the difference that PCA is used instead of clustering in the first layer designed in an unsupervised way and the sigmoidal activation functions are used in the output nodes, instead of linear. The network is trained by stochastic optimization.

4 Hybrid techniques

4.1 Genetic algorithms in hybrid techniques

In bankruptcy prediction, GA are usually used to select a subset of input features, to find appropriate hyper-parameter values of a predictor (for example, the kernel width and the regularization constant in the case of SVM), or to determine predictor parameters (MLP weights, for example). In some applications, selection of both hyper-parameters and a subset of input features is integrated into one learning process.

Pendharkar and Rodger (2004) as well as Sai et al. (2007) used GA to train an MLP and then tested the neural network on bankruptcy prediction data. Abdelwahed and Amir (2005) developed a two stage technique for designing a bankruptcy prediction tool based on GA and an MLP. In the first stage, GA is used to select a subset of input features. Then, in the second stage, GA is applied to optimize the topology of the network. The final tuning of network weights is done by the gradient decent. Ignizio and Soltys (1996), and Wallrafen et al. (1996) combined MLP design, training and feature selection into one learning process based on genetic search.

Min et al. (2006) as well as Ahn et al. (2006) used GA to design an SVM-based technique for bankruptcy prediction. The selection of both SVM hyper-parameters and input features is integrated into one learning process based on genetic search. Chen and Hsiao (2008) as well as Wu et al. (2007) used GA to find SVM hyper-parameters. Van Gestel (2006), in contrast, find hyper-parameters for the least squares support vector machine (LS-SVM) by applying the Bayesian evidence framework (MacKay 1992; Gestel et al. 2002). Comparison of the efficiency of the GA- and the Bayesian evidence framework-based approaches to determination of the SVM hyper-parameters would be interesting.

Quintana et al. (2008) applied evolutionary programming to evolve the so-called evolutionary nearest neighbour classifier for bankruptcy prediction. The relevant number of the nearest neighbours to be used is determined through evolutionary programming. When testing on one data set, the authors have found that the classifier was more accurate than SVM or MLP.

Tsakonas et al. (2006) used GA to evolve a bankruptcy prediction system based on the so-called neural logic networks. An elementary neural logic network consists of a set of input nodes and an output node. Elementary networks can be combined to form larger networks. A three-valued logic is used. An output value [an ordered pair (x, y)] for a node of the neural logic network is given by:

$$(x, y) = \begin{cases} (1, 0) & \text{if } \sum_{j=1}^{N} w_j x_j - \sum_{j=1}^{N} v_j y_j \geq 1 \\ (0, 1) & \text{if } \sum_{j=1}^{N} w_j x_j - \sum_{j=1}^{N} v_j y_j \leq -1 \\ (0, 0) & \text{otherwise} \end{cases} \quad (1)$$

where N is the number of inputs and (w_j, v_j) is an ordered pair of weights. Both topology and parameters are determined by genetic search.

An interesting GA-based hybrid technique has recently been proposed by Hu (2008), and Hu and Tseng (2007). An MLP is the classifier used to predict bankruptcy. Nodes of an usual MLP aggregate input signals via a weighted sum. Nodes of the MLP suggested by Hu aggregate information

via the discrete Choquet integral. A non-additive fuzzy measure is used in the Choquet integral, instead of sum.

If we assume that Z is a non-empty finite set and g is a fuzzy measure on Z, the *discrete Choquet integral* of a function $h : Z \to \mathbb{R}^+$ with respect to g is defined as

$$C_g(h(z_1), \ldots, h(z_L)) = \sum_{i=1}^{L} [h(z_i) - h(z_{i-1})]g(A_i) \qquad (2)$$

where indices i have been permuted so that $0 \le h(z_1) \le \cdots \le h(z_L) \le 1, A_i = \{z_i, \ldots, z_L\}; h(z_0) = 0$, and L is the number of elements in the set Z (Grabisch 1996).

A set function $g : 2^Z \to [0, 1]$ is a *fuzzy measure* if

1. $g(\emptyset) = 0; g(Z) = 1$,
2. if $A, B \subset 2^Z$ and $A \subset B$ then $g(A) \le g(B)$,
3. if $A_n \subset 2^Z$ for $1 \le n < \infty$ and $\{A_n\}$ is monotonic in the sense of inclusion, then $\lim_{n \to \infty} g(A_n) = g (\lim_{n \to \infty} A_n)$.

In general, the ordinary fuzzy measure of a union of two disjoint subsets cannot be directly computed from the ordinary fuzzy measures of the subsets. Sugeno (1977) introduced the so-called λ-fuzzy measure, which allows such computation. Hu (2008) uses the λ-fuzzy measure and applies GA to train an MLP. A considerable improvement in bankruptcy prediction accuracy was obtained if compared to the accuracy obtained from an ordinary MLP.

4.2 Rough sets in hybrid techniques

In hybrid bankruptcy prediction techniques, rough sets are usually used to select input features. Zhou and Tian (2007) suggest combining the theory of rough sets and SVM. The SVM applied uses the wavelet kernel function. Therefore, the authors call the classifier the wavelet SVM. The Mexican hat wavelet is used to construct the SVM kernel. Rough sets are used to select input features. Cheng et al. (2007) have demonstrated that the bankruptcy prediction accuracy of the rough sets-based tool can be increased substantially by including a non-financial variable, auditor switching in this case, into the modeling process.

Aiming to increase bankruptcy prediction accuracy, Ahn et al. (2000) combined an MLP and the rough sets theory based technique. The rough sets theory based analysis is used for both feature selection and generation of rules. McKee and Lensberg developed a hybrid technique for bankruptcy prediction by combining the rough sets theory based model and genetic programming (McKee and Lensberg 2002). The rough sets theory is used to select the input features, while genetic programming evolve the model in the form of non-linear real-valued algebraic expressions of the features selected by the rough sets technique.

Bian and Mazlack (2003) proposed combining the fuzzy k-nearest neighbour algorithm (Keller et al. 1985) and the rough sets theory, to improve the accuracy of bankruptcy prediction. The authors demonstrated the increased prediction accuracy if compared to either the crisp or fuzzy nearest neighbour approach.

4.3 Hybrid systems of increased transparency

In general, an SVM (Vapnik 1998) or RVM (Tipping 2001) can provide near optimal performance. However, classifiers based on these techniques are not transparent enough and are often considered as "black boxes". Transparency is a very important issue sometimes. Aiming to increase the transparency, some researchers design fuzzy set theory-based techniques or incorporate SOM for data exploration and visualization purposes.

4.3.1 Fuzzy set theory-based techniques

Lu et al. (2006) aiming to obtain a transparent explanatory system for bankruptcy prediction, adopt the rule-based approach. Rules can be generated directly by a GA. However, to facilitate the designing process, the authors extract rules from a trained neural network. To obtain simple but substantial statements in classification rules, neural network weight pruning is carried out first. Then, the GA is applied, to obtain ultimate classification rules. Kumar and Ravi (2006) have also proposed a fuzzy rule-based bankruptcy prediction technique. The task of classifier design is formulated as a multi objective combinatorial optimization problem aiming to maximize the classification accuracy and to minimize the number of rules. The so-called modified threshold accepting technique (Ravi et al. 2001) is adopted to solve the optimization problem. In Jeng et al. (1997), bankruptcy predictions are obtained from a fuzzy decision tree, designed by combining the fuzzy set theory and decision tree construction based on inductive learning.

Neuro-fuzzy is a popular approach in various control and classification applications. By combining the fuzzy sets theory and the MLP, Gorzalczany and Piasta designed a neuro-fuzzy classifier for bankruptcy prediction (Gorzalczany and Piasta 1999). The fuzzy sets-based input module allows inputting both purely numerical data as well as qualitative, linguistic data that may be used to characterize the decision-making process. The authors demonstrated superiority of the neuro-fuzzy classifier over the rough sets-based technique, C4.5 decision tree, and the rule induction system CN2 (Clark and Niblett 1989). Lee et al. (2006) studied the efficiency of several training techniques applied to the POPFNN-CRI(S) fuzzy-neural network (Ang et al. 2003), which was then used to predict bankruptcy. As it is often the case in neuro-fuzzy approaches,

the network consists of five layers: input, antecedent, rule-base, consequence, and output.

Tung et al. (2004), aiming to predict bankruptcy and to identify the characteristics of financial distress, proposed the so-called Generic Self-organizing Fuzzy Neural Network (GenSoFNN). As many other fuzzy-neural systems, the proposed network also consists of five layers: input (fuzzifier) layer, antecedent matching layer, rule-based layer, consequent derivation layer, and output (defuzzification) layer. Parameters of the network are learned through the gradient decent. The base of IF-THEN rules designed during training provides insight into the contribution of the selected features (financial covariates) to the bankruptcy. Thus, it is possible to analyze reasons behind the bankruptcy and identify the symptoms of financial distress. Nonetheless, the slightly lower prediction accuracy obtained from the GenSoFNN if compared to MLP, the authors advocate using the GenSoFNN network due its transparency.

4.3.2 SOM in hybrid systems

Aiming to get a deeper insight into results obtained from a prediction tool, Serrano-Cinca (1996) created a SOM using the financial data (financial ratios) and superimposed the prediction results obtained from an MLP on the SOM. The obtained map, served as a convenient tool for visual inspection of the analysis results. Huysmans et al. (2006) have also combined MLP and SOM, aiming to exploit good data exploration properties of SOM. MLP is trained first using financial input data. The input data used to train SOM consist, however, of the financial input data augmented with the output of the MLP. When training SOM, the weighted Euclidean distance, given by Eq. 3, is used instead of the Euclidean one

$$||\mathbf{x} - \mathbf{m}||_2 = \sum_{j=1}^{N} w_j(x_j - m_j)^2 \qquad (3)$$

where N is the number of variables and w_j stands for the jth variable weight. A higher weight is assigned to the MLP output.

4.4 Combining traditional and soft computing techniques

Markham and Ragsdale design a hybrid system by augmenting the set of neural network input features with additional Mahalanobis distance measures (Markham and Ragsdale 1995). The authors demonstrate an improvement in the prediction accuracy if compared to a common neural network case. Piramuthu et al. (1998) apply constructive operators (multiplication and division, for example) to

original features and construct new features. A subset of the original and new features are then selected and used to train an MLP. Experimental tests performed using bankruptcy data demonstrated that the constructed features help improving the classification accuracy of the MLP. Lee et al. (1996) selected input features based on the multivariate discriminant analysis or ID3 tree and then used in a feedforward neural network to predict bankruptcy. A similar approach was also taken by Lee et al. (2002). The authors used the LDA for feature selection and also to generate an additional input to the MLP. The LDA output served as the additional input. Back et al. also experimented with various feature selection techniques followed by prediction based on the LDA, LR, or MLP (Back et al. 1996). To select features either LDA, LR, or genetic search was applied. According to the tests, MLP trained using features selected by the genetic search was the best approach.

Tseng and Lin (2005) have suggested combining LR and fuzzy regression called quadratic interval regression. The combined model called quadratic interval logit is characterized by a fuzzy parameter. The task of finding the fuzzy regression parameters is formulated as a linear programming problem. Case-based reasoning and information retrieval techniques were combined in the bankruptcy support system developed by Elhadi (2000).

5 Ensembles

Numerous previous works on prediction ensembles have shown that an efficient ensemble should consist of predictors that are not only very accurate, but also diverse in the sense that the predictor errors occur in different regions of the input space. Krogh and Vedelsby (1995) have shown that

$$E = \overline{E} - \overline{A} \qquad (4)$$

where E is the committee generalization error, $\overline{E}$ is the weighted average of the generalization errors of the committee networks, and $\overline{A}$ is the committee ambiguity.

Diversity of ensemble members can be achieved in the expense of ensemble accuracy. Thus, tradeoff between the accuracy and the diversity is desired (Kuncheva and Whitaker 2003). Achieving the tradeoff is a rather difficult task. For example, one always attempts to avoid over-fitting when designing a single predictor. However, Sollich and Krogh have shown that some over-fitting can be useful when designing an ensemble (Sollich and Krogh 1996). The authors found that in large ensembles of linear members one should use under-regularized members. An ensemble of such members benefits from "the variance-reducing effects

of ensemble learning" (Sollich and Krogh 1996). The authors expect the finding to carry over to an ensemble of non-linear members. To achieve the tradeoff when designing an ensemble of MLPs, negative correlation learning has been proposed (Liu and Yao 1999; Liu et al. 2000; Islam et al. 2003). The mean squared error function minimized during negative correlation learning is augmented with an additional term penalizing correlation between ensemble networks.

Splitting or splitting and weighting a data set by clustering (Verikas and Lipnickas 2002), bootstrapping (Breiman 1996), AdaBoosting (Freund and Schapire 1997), pasting votes (Breiman 1999), employing different subsets of features and different architectures are the most popular approaches used to achieve the diversity of ensemble members. A recent review on diversity creation techniques can be found in Brown et al. (2005). Since employing different subsets of features in different ensemble members affects both diversity of ensemble members and ensemble accuracy, integration of feature selection, selection of hyper-parameters, and training of ensemble members into one learning process is desired. An example of such approach to ensemble design can be found in Bacauskiene and Verikas (2004) and Bacauskiene et al. (2009).

The strategy used to aggregate predictors into an ensemble is one more issue greatly affecting the ensemble accuracy (Verikas and Lipnickas 2002; Kuncheva et al. 2001; Verikas et al. 1999; Liu 2005; Kuncheva 2002). Majority voting, averaging, and weighted averaging are the most popular aggregation techniques used in bankruptcy prediction. The rest of the survey is structured according to the aforementioned issues, most notably affecting the ensemble accuracy.

5.1 Creating diverse ensemble members

5.1.1 Using different feature subsets

Shin et al. (2006) promote the diversity of ensemble members using different techniques to select features for ensemble members. Two types of ensembles are investigated: a bagged ensemble consisting of 30 MLPs and a stacked one (Wolpert 1993) made of k-NN, C4.5 decision tree, and MLP. To promote diversity of ensemble members (RBF networks in this case), Chan et al. (2006) perform bagging and select features separately for each network trained on a separate bagged data set. Features being selected are those maximizing the mutual information between features and the class labels. When tested experimentally, ensembles built using averaging, weighted averaging, and majority voting provided approximately the same performance.

Yeung et al. (2007a, b) also design an ensemble of RBF networks to predict bankruptcy. Aiming to evolve diverse ensemble members (experts in different local regions of the input space), diversity is promoted during the GA-based feature selection process by including a diversity term in the fitness function. Features for all ensemble members are selected simultaneously by designing a chromosome of $L \times N$ genes, where L is the number of ensemble members and N is the dimensionality of the input space. The feature selection task is solved as the following optimization problem (Yeung et al. 2007):

$$\arg \max_{\{x_j\}_l \subseteq \{x_j\}, \forall l=1,\dots,L} \sum_{l=1}^{L} \psi_l \tag{5}$$

where $\{x_j\}$ and $\{x_j\}_l$ stand for the set of all features and the feature set used by the lth member, respectively, and ψ_l is the fitness function for the lth ensemble member. The fitness function is given by:

$$\psi_l = \frac{1}{R_{\text{SM}}^*(l)} + \lambda d(l) \tag{6}$$

where $R_{\text{SM}}^*(l)$ is the estimate of the local generalization error for the lth ensemble member, $d(l)$ stands for the diversity of the lth member, and λ is the regularization parameter. The diversity measure for the lth member is defined as:

$$d(l) = E_D\{f_l(\mathbf{x}) - E_D[f(\mathbf{x})]\}^2 \tag{7}$$

where $f(\mathbf{x})$ and $f_l(\mathbf{x})$ denote the ensemble output and the lth ensemble member output, respectively and E_D stands for expectation over the data set. The ensemble decision is obtained by aggregating the member decisions via the weighted sum rule.

5.1.2 Manipulating training data set

Alfaro et al. (2008) as well as Cortes et al. (2007) applied an ensemble of decision trees (Breiman et al. 1993) created using the AdaBoost algorithm (Freund and Schapire 1996; Freund and Schapire 1997). AdaBoost gradually increases the number of ensemble members. Training of subsequent members is gradually more and more focused on misclassified training data points. An output of an AdaBoost ensemble is given by a linear combination of outputs of single classifiers. When applying the AdaBoost ensemble of decision trees to the bankruptcy data, Alfaro et al. demonstrated a reduction of 30% of the test set error rate, when compared to the error rate obtained from a single MLP.

West et al. (2005) investigated the accuracy of ensembles made of 100 MLPs created using three different data manipulation strategies, namely, cross validation, bagging

(Breiman 1996), and AdaBoosting. When applied to bankruptcy data, no significant difference was found between the accuracy of the ensembles. However, tests performed by other authors using a large number of different data sets, have shown that ensembles created using the AdaBoost algorithm outperform the ones built using the other data sampling approaches (Bauer and Kohavi 1999). AdaBoost, however, is a rather complex algorithm. Breiman proposed a very simple algorithm, the so called Half & Half bagging technique (Breiman 1998). The Half & Half algorithm builds a committee incrementally. It uses random sampling to collect a new training data set that is half filled of data points correctly classified up to the present and half filled of misclassified data.

Yu et al. (2007) apply the bagging sampling technique to create different training data sets for training members of an SVM ensemble. A series of SVM with different hyperparameters are created using the training data sets and then aggregated into a committee by applying evolutionary programming. West and Dellana (2005) have also studied the influence of the diversity of members on the accuracy of a bagged ensemble.

Tsai and Wu (2008) obtained unexpected bankruptcy prediction results from an ensemble of MLPs diversified through training data set manipulation. Majority voting was the rule used to aggregate the ensemble members. On average, single classifiers showed a higher accuracy than the ensemble. This is probably due to the very small data sets used to train the ensemble members as well as due to the procedure applied to design the ensemble. Aiming to increase the prediction accuracy of an ensemble of MLPs, Shin and Kilic (2006) linearly transform the input features by applying the principal component analysis and use a fewer number of new features to train the networks. Horta et al. (2008) studied the problem of designing a classification ensemble for bankruptcy prediction in the context of class-imbalanced training data sets.

5.1.3 Using different architectures

Olmeda and Fernandez (1997), and Jo and Han (1996) were amongst the first to use an ensemble for bankruptcy prediction. In Olmeda and Fernandez (1997), an MLP, LDA, LR, Multivariate Adaptive Regression Splines (MARS), and C4.5 decision tree were combined into an ensemble. Two combination schemes were explored, voting and weighted sum. Genetic search was used to find the combination weights. Jo and Han (1996) and Jo et al. (1997) created an ensemble consisting of an MLP, LDA, and a case-based forecasting module. Weighted averaging was used to aggregate the members into an ensemble. The appropriate weight values were found experimentally by trial and error. In both works, improvement in prediction

accuracy was reported, when compared with the best single model. An MLP, LR, LDA, and C5.0 decision tree were combined into the weighted voting ensemble developed by Lin and McClean (2001). The weights were proportional to the prediction accuracy of the ensemble members estimated on the training data set. Only a slight improvement in the prediction accuracy was obtained from the ensemble if compared to the best single member, which was the decision tree in this application. Kim and Yoo (2006) used a linear combination of LR and MLP in their bankruptcy prediction application.

Hua et al. (2007) suggest combining SVM and LR. The SVM output range is divided into several intervals. If a decision made by the SVM is supported by LR with a large enough probability, the SVM decision is accepted. Otherwise, the decision may be modified depending on the interval the SVM output depends to.

Ravi et al. (2008) aggregated nine classifiers of different architecture, to build an ensemble for bankruptcy prediction. MLP, RBF, PNN, SVM, classification and regression trees (CART), fuzzy rule-based classifier, PCA-MLP, PCA-RBF, and PCA-PNN are the classifiers used to build the ensemble, where PCA means that data were preprocessed by PCA first. Majority voting and weighted averaging rules were used for the aggregation. Both ensembles outperformed the best single member, which was PCA-PNN. Aiming to create diverse ensemble members, Sun and Li (2008) have also used different architectures, namely LDA, LR, MLP, SVM, and CBR. The members were aggregated into an ensemble by the weighted majority voting rule.

5.2 Determining the number of ensemble members

Depending on the aggregation rule applied and the accuracy of ensemble members, the ensemble accuracy may greatly depend on the number of ensemble members. It was demonstrated that sequential forward selection of ensemble members may significantly improve the accuracy of averaging ensemble, when compared to the accuracy of ensemble obtained by averaging all the available ensemble members (Verikas et al. 2008). It was also demonstrated that the average ensemble accuracy may be increased substantially by designing data dependent ensembles, meaning that members included into such ensemble depend on the input data point being analyzed (Verikas et al. 2002; Santosa et al. 2008; Englund and Verikas 2005). Thus, dynamic selection of ensemble members is utilized. However, these issues have almost never been addressed in the bankruptcy prediction literature.

Ravikumar and Ravi (2006) experimented with ensembles created using a varying number of members. A set of seven classifiers was available: adaptive neuro fuzzy inference system (ANFIS) (Jang 1993), SVM, four types of RBF

networks, and MLP. The majority voting rule has been used to aggregate ensemble members. As expected, the optimal size and structure of the ensemble were data dependent.

5.3 Aggregating ensemble members

A variety of schemes have been proposed for combining multiple classifiers. The approaches used most often include the majority vote, averaging, weighted averaging, the Bayesian approach, the fuzzy integral, the Dempster-Shafer theory, the Borda count, aggregation through order statistics, probabilistic aggregation, the fuzzy templates, and stacked generalization (Kuncheva et al. 2001; Verikas et al. 1999; Liu 2005; Verikas and Lipnickas 2002; Kuncheva 2002; Wolpert 1993; Kittler et al. 1998; Xu et al. 1992). However, aggregation approaches used in bankruptcy prediction are most often limited to majority voting, averaging and weighted averaging.

Doumpos and Zopounidis (2007) applied the stacked generalization approach proposed by Wolpert (1993) to build an ensemble consisting of LDA, LR, PNN, SVM, the nearest neighbour classifier, the classification and regression trees (CART), and the quadratic discriminant analysis technique (QDA). The choice of the techniques is motivated by different learning capacity. Shin et al. (2006) used MLP as a meta-classifier to stack k-NN, C4.5, and MLP classifiers.

To aggregate MLPs into an ensemble, Shin and Lee assess the confidence α_i of the ith ensemble member in its prediction as Shin and Lee (2004)

$$\alpha_i = \max\{|0 - y_i|, |1 - y_i|\} \tag{8}$$

where y_i stands for the ith member output. In the case of conflicting predictions delivered by members of the ensemble, the ensemble output is given by the output of the member of the highest confidence.

6 Model assessment and selection

Usually, bankruptcy prediction is considered as a two-class (binary) classification problem. Assuming that the classes are labeled as negative and positive, and denoting the true and predicted class labels by $y = \pm 1$ and $\hat{y} = \pm 1$, respectively, a confusion matrix characterizing the performance of a classifier can be constructed as that shown in Table 1.

Table 1 A confusion matrix for a two-class classification problem

True\predicted	$\hat{y} = -1$	$\hat{y} = 1$
$y = -1$	TN	FP
$y = 1$	FN	TP

In Table 1, TN, FN, TP, and FP stand for true negatives, false negatives, true positives, and false positives, respectively. Several common metrics, characterizing the performance of a classifier, can be calculated from the confusion matrix: *sensitivity* (SE) (or true positive rate (TPR), also known as *recall*), *specificity* (SP) [or true negative rate (TNR)], false positive rate (FPR) (also known as 1-SP), and *accuracy* (AC) (Fawcett 2006; Waegeman et al. 2008):

$$SE = TPR = \frac{TP}{TP + FN} \tag{9}$$

$$SP = TNR = 1 - FPR = \frac{TN}{TN + FP} \tag{10}$$

$$AC = \frac{TP + TN}{N_- + N_+} \tag{11}$$

where N_- and N_+ stand for the number of data points in the negative and the positive class, respectively.

Accuracy (AC), FPR, and FNR (type-I error and type-II error) are the most widely used measures to asses the performance of bankruptcy prediction systems. To test the statistical significance of the difference obtained between two models, a p-value of the paired t-test applied to the cross-validation error rates (Doumpos and Zopounidis 2007; West et al. 2005; Tsai and Wu 2008) or McNemar's test (Ripley 1996; Gestel et al. 2006) is sometimes calculated.

Nowadays, a receiver operating characteristic (ROC) curve as well as area under the ROC curve (AUC) are increasingly used to characterize the performance of a binary classifier. A ROC curve is obtained by plotting the TPR versus the FPR. The curve depicts relative tradeoffs between benefits (TP) and costs (FP) (Fawcett 2006). However, this is not the case in the bankruptcy prediction techniques. ROC curves as well as AUC are used rather seldom in the analysis (Ribeiro et al. 2006; Gestel et al. 2006; Ravi and Pramodh 2008).

To compare AUC, Van Gestel et al. (2006) use the test of De Long et al. (1988) based on the theory of generalized U-statistics. Fawcett (2006) presents two algorithms for obtaining confidence intervals for ROC curves by averaging individual ROC curves created for a number of test data sets generated by cross-validation or the bootstrap technique (Efron and Tibshirani 1993, 1997). Yousef et al. (2005) suggest using the bootstrap-based estimator to estimate the AUC. The uncertainty of that estimate is also obtained from the same bootstrap samples.

The problem of selecting a model of appropriate complexity, the number of hidden nodes in an MLP for example, is often forgotten when developing soft computing techniques for bankruptcy prediction. Bootstrap sampling can be used to determine an appropriate model

complexity (Hastie et al. 2001; Verikas and Bacauskiene 2003; Kallel et al. 2002).

7 Discussion

A large variety of hybrid and ensemble-based soft computing techniques for bankruptcy prediction have been developed so far. Table 2 presents a selective survey of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction. The main model designing issues considered in different studies are provided in Table 2. The techniques developed are usually tested using one or very few data sets. Moreover, the disparity of sample sizes is very big in different studies and confidence intervals for the obtained prediction accuracies are seldom provided. Thus, fair comparison of results obtained in the different studies is hardly possible. Comparisons of various techniques on multiple data sets are required. Demsar suggests using the non-parametric *Wilcoxon signed-ranks test* to compare two classifiers and the *Friedman test* to compare several classifiers over multiple data sets (Demsar 2006).

Nonetheless the difficulty in comparing the reviewed techniques, one can make a general observation that ensembles, when properly designed, are more accurate than the other techniques. This is expected, since an ensemble integrates several predictors. In a successful ensemble design, a tradeoff between the ensemble accuracy and the diversity of ensemble members is achieved. Achieving the tradeoff is a rather difficult task and requires integration of feature selection, selection of hyper-parameters, and training of ensemble members into one learning process. GA suits well to accomplish such integration. Aiming to evolve diverse ensemble members, diversity can be promoted during the search process by including a diversity term in the GA fitness function. In bankruptcy prediction, GA are usually used to select a subset of input features, to find appropriate hyper-parameter values of a predictor, or to determine predictor parameters. Very few studies concern such integrated designing of ensembles. The ensemble accuracy may greatly depend on the number of ensemble members being aggregated and the aggregation rule applied. Data dependent dynamic selection of ensemble members is under-exploited issue in the bankruptcy prediction literature.

However, transparency of ensemble-based techniques is rather limited, when compared to RS or IF-THEN rules-based approaches. Transparency of decisions expressed in the form of decision rules, the possibility of using both quantitative and qualitative data that may be used to characterize the decision-making process are the advantages of RS and IF-THEN rules-based approaches. The base of rules designed during training provides insight into the

contribution of the selected features to the bankruptcy. Thus, it is possible to analyze reasons behind the bankruptcy and identify the main symptoms of financial distress. RS and IF-THEN rules-based techniques lend themselves well to creating KBDSSs. A KBDSS can facilitate the understanding of the operation and the results of the decision system, can help ensuring the objectiveness of the results, and structuring the decision analysis properly. Evolutionary computing based designing of KBDSSs can be a promising research direction.

Different studies indicate that the bankruptcy prediction accuracy can be increased substantially by including non-financial features into the modeling process and a trend is on using non-financial features, for example macroeconomic indicators and qualitative variables, in addition to financial ratios. A large number of features can be usually collected in various applications. Not all of the features, however, are equally important for a specific task. Some of the features may be redundant or even irrelevant. Therefore, in many applications we need to reduce the dimensionality of the data via feature selection or feature extraction. Genetic algorithms and RS are the two most popular approaches to feature selection in hybrid and ensemble-based techniques for bankruptcy prediction. For large feature sets, however, the GA-based feature selection can be very time consuming, especially if classification accuracy the estimation of which involves classifier training, is used to assess the saliency of a subset of features in the selection process. It is worth mentioning that classification accuracy is the most often used criterion to assess the quality of a subset of features. As to RS, the sensitivity of the approach to changes in data is an important issue.

Non-linear dimensionality reduction techniques offer a great potential for applications in the analysis of financial data. GP-LVM is a very promising non-linear mapping technique. An extension of GP-LVM for classification was also developed recently. GP-LVM can be trained to exhibit the property of local data ordering in a low-dimensional space when mapping high-dimensional data onto the low-dimensional space. Local data ordering is a very useful property, also characteristic to SOM and CCA, for exploring high-dimensional data. By providing ordered data maps, GP-LVM, SOM and CCA can facilitate the exploration and understanding of the results obtained from non-linear prediction techniques.

The non-linear nature of hybrid and ensemble-based models and the lack of a widely accepted procedures for designing such models are major factors contributing to pitfalls in applications of these technologies. Model building, model selection and comparison are the designing steps where the most common pitfalls occur due to small sample sizes, model over-fitting or under-fitting, the sensitivity of solutions to initial conditions. The problem of

Table 2 A survey of hybrid and ensemble-based soft computing techniques applied to predict bankruptcy

Techniques	Model designing issues
Hybrid techniques	
GA + MLP	GA used for: finding MLP weights (Pendharkar and Rodger 2004; Sai et al. 2007); selecting features (Back et al. 1996); selecting features and MLP topology (Ignizio and Soltys 1996; Wallrafen et al. 1996); selecting features then MLP topology (Abdelwahed and Amir 2005); finding weights of the fuzzy integral-based MLP (Hu 2008; Hu and Tseng 2007).
GA + SVM	GA used for: selecting SVM hyper-parameters (Chen and Hsiao 2008; Wu et al. 2007); selecting features and SVM hyper-parameters (Min et al.2006; Ahn et al.2006).
GA + k-NN	GA used for finding the appropriate k value (Quintana et al. 2008).
GA + NLN	GA used for finding the network topology and parameters (Tsakonas et al. 2006).
RS + SVM	RS used for selecting features (Zhou and Tian 2007).
RS + MLP	RS used for selecting features for MLP and generating decision rules (Ahn et al. 2000).
RS + GA	RS used to select features for algebraic expressions obtained by GA (McKee and Lensberg 2002).
RS + k-NN	RS used for estimating the class membership in a fuzzy k-NN classifier (Bian and Mazlack2003).
FS + MLP + GA	FS used to extract rules from MLP, which are then optimized by GA (Lu et al. 2006).
FS + MOCO	FS used to extract rules, which are then optimized by MOCO (Kumar and Ravi 2006).
FS + IL	FS and IL are combined to obtain a fuzzy decision tree (Jeng et al. 1997).
FS + MLP	FS and MLP are combined to obtain a neuro-fuzzy classifier (Gorzalczany and Piasta1999).
FS + BP	FS used to design a fuzzy-neural network trained by BP. The initial structure of the network is obtained by self-organization (Lee et al. 2006; Ang et al. 2003; Tung et al. 2004).
SOM + MLP	Prediction results obtained from MLP are superimposed on SOM visualizing the financial data (Serrano-Cinca 1996); SOM is trained using the financial data augmented with the prediction obtained from the MLP (Huysmans et al. 2006).
NT + MLP	Feature set for MLP is augmented with Mahalanobis distance measures (Markham and Ragsdale1995]; features are transformed by applying NT (Piramuthu et al. 1998).
LDA + MLP	Features are selected by LDA ands then used in MLP (Lee et al. 1996; Back et al. 1996); LDA is used to select features and also to generate an additional input to MLP (Lee et al. 2002).
LR + MLP	Features are selected by LR and then used in MLP (Back et al. 1996).
LR + FR	LR and FR are combined. Fuzzy regression parameters are found by solving a linear programming problem (Tseng and Lin 2005).
CBR + IR	CBR and IR techniques are combined (Elhadi 2000).
Ensemble techniques	
MLP + k-NN + C4.5	MLP, k-NN, and C4.5 are aggregated by stacking via a new MLP. To promote diversity, different techniques are used to select features for ensemble members (Shin et al. 2006).
MLP	Ensemble of 30 bagged MLPs (Shin et al. 2006).
MLP	Three different types of ensembles of 100 MLPs created using bagging, cross validation, or AdaBoosting (West et al. 2005).
MLP	To reduce dimensionality input features are transformed by applying PCA (Shin and Kilic 2006).
RBF	A bagged ensemble. Features are selected separately for each network trained on a separate bagged data set (Chan et al. 2006).
RBF	Diversity is promoted during the GA-based feature selection by including a diversity term in the fitness function. Features are selected simultaneously for all members (Yeung et al. 2007a, b).
DT	An ensemble of DT created using the AdaBoost algorithm (Alfaro et al. 2008; Cortes et al. 2007).
SVM	Bagged ensemble of SVMs, where different ensemble members use different hyper-parameters (Yu et al. 2007).
MLP+LDA+LR+ MARS+C4.5	Aggregation by voting and weighted sum are explored. GA is used to find the aggregation weights (Olmeda and Fernandez 1997).

Table 2 continued

Techniques	Model designing issues
MLP+LDA+CBR	Weighted averaging is used to aggregate the members into an ensemble (Jo and Han 1996; Jo et al. 1997).
MLP+LR+LDA+C5.0	The members are aggregated by weighted voting (Lin and McClean 2001).
MLP+LR	The members are aggregated via linear combination (Kim and Yoo 2006).
SVM+LR	If an SVM decision is not supported enough by LR, the decision may be modified (Hua et al. 2007).
MLP+RBF+PNN+SVM+ DT+FRB+PCA-MLP+ PCA-RBF+PCA-PNN	Nine different architectures are aggregated by the majority voting or weighted averaging rules. PCA means that data are preprocessed by PCA first (Ravi et al. 2008).
MLP + SVM + LDA + LR + CBR	The members are aggregated by the weighted majority voting rule (Sun and Li 2008). Diversity of the classifiers is studied.
ANFIS + SVM + MLP + four types of RBF	Experiments with a varying ensemble size. The optimal size and the structure were data dependent (Ravikumar and Ravi 2006).
LDA + LR + PNN + SVM + k-NN + QDA + CART	The members are aggregated by the stacked generalization approach (Doumpos and Zopounidis 2007). The choice of the techniques is motivated by different learning capacity.
MLP	The closeness of the MLP output value to either 0 or 1 is considered as confidence in the prediction (Shin and Lee 2004). The aggregation rule exploits the confidences.
MLP	MLPs diversified through training data set manipulation and aggregated by majority voting (Tsai and Wu 2008). On average, the best MLP was more accuracy than the ensemble, probably due to the very small training data sets.

NLN neural logic network, *MOCO* multi objective combinatorial optimization, *IL* inductive learning, *BP* backpropagation, *NT* non-linear transformation, *FR* fuzzy regression, *IR* information retrieval, *FRB* fuzzy rule-based classifier

selecting a model of appropriate complexity is often forgotten when developing soft computing techniques for bankruptcy prediction.

We hope that the comprehensive review of available techniques will help researchers to focus their attention on under-explored research fields. Large scale comparisons of various techniques, integration of multiple data mining methods and choice of suitable values of parameters governing the behaviour of the methods, scalability, feature selection for prediction ensembles, ensemble designing and adaptation in dynamic environments, integration of various ensemble designing steps into one learning process, unbalanced data sets, heterogeneous and distributed data sources, text mining, and estimation of the uncertainty of a binary bankruptcy prediction event are several important issues to consider.

References

Abdelwahed T, Amir EM (2005) New evolutionary bankruptcy forecasting model based on genetic algorithms and neural networks. In: Proceedings of the 17th IEEE international conference on tools with artificial intelligence (ICTAI05), IEEE Computer Society, pp 1–5

Ahn BS, Cho SS, Kim CY (2000) The integrated methodology of rough set theory and artificial neural network for business failure prediction. Exp Syst Appl 18:65–74

Ahn H, Lee K, Kim KJ (2006) Global optimization of support vector machines using genetic algorithms for bankruptcy prediction. In: Lecture notes in computer science, vol 4234. Springer, Heidelberg, pp 420–429

Alfaro E, Garcia N, Gamez M, Elizondo D (2008) Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks. Decis Support Syst 45:110–122

Altman E (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finance 13:589–609

Ang KK, Quek C, Pasquier M (2003) POPFNN-CRI(S): pseudo outer product based fuzzy neural network using the compositional rule of inference and singleton fuzzifier. IEEE Trans Syst Man Cybern B Cybern 33:838–849

Atiya AF (2001) Bankruptcy prediction for credit risk using neural networks: a survey and new results. IEEE Trans Neural Netw 12:929–935

Bacauskiene M, Verikas A (2004) Selecting salient features for classification based on neural network committees. Pattern Recogn Lett 25:1879–1891

Bacauskiene M, Verikas A, Gelzinis A, Valincius D (2009) A feature selection technique for generation of classification committees and its application to categorization of laryngeal images. Pattern Recognit 42:645–654

Back B, Laitinen T, Sere K (1996) Neural network and genetic algorithm for bankruptcy predictions. Exp Syst Appl 11:407–413

Bauer E, Kohavi R (1999) An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach Learn 36:105–142

Bian H, Mazlack L (2003) Fuzzy-rough nearest-neighbor classification approach. In: Fuzzy information processing society, 2003. NAFIPS 2003. 22nd international conference of the North American, pp 500–505

Bishop CM (2006) Pattern recognition and machine learning. Springer, Singapore

Bishop CM, Svensen M, Williams CKI (1998) GTM: the generative topographic mapping. Neural Comput 10:215–234

Borg I, Groenen PJF (1997) Modern multidimensional scaling: theory and applications. Springer, New York

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Breiman L (1998) Half & Half bagging and hard boundary points. Technical Report 534, Statistics Department, University of California, Berkeley

Breiman L (1999) Pasting small votes for classification in large databases and on-line. Mach Learn 36:85–103

Breiman L, Friedman JH, Olshen RA, Stone CJ (1993) Classification and regression trees. Chapman & Hall, London

Brown G, Wyatt J, Harris R, Yao X (2005) Diversity creation methods: a survey and categorisation. Inf Fusion 6:5–20

Chalup S, Mitschele A (2008) Kernel methods in finance. In: Seese D, Weinhardt C, Schlottmann F (eds) Handbook on information technology in finance, vol II. Springer, Berlin, pp 655–687

Chan APF, Ng WWY, Yeung DS, Tsang ECC, Firth M (2006) Bankruptcy prediction using multiple classifier system with mutual information feature grouping. In: 2006 IEEE international conference on systems, man, and cybernetics, Taipei, Taiwan, pp 845–850

Chen LH, Hsiao HD (2008) Feature selection to diagnose a business crisis by using a real GA-based support vector machine: an empirical study. Exp Syst Appl 35:1145–1155

Cheng JH, Yeh CH, Chiu YW (2007) Improving business failure predication using rough sets with non-financial variables. In: Lecture notes in computer science, vol 4431. Springer, Heidelberg, pp 614–621

Clark P, Niblett T (1989) The CN2 induction algorithm. Mach Learn 3:261–283

Cortes EA, Martynez MG, Rubio NG (2007) A boosting approach for corporate failure prediction. Appl Intell 27:29–37

Cottrell GW (2006) New life for neural networks. Science 313:454–455

DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845

Demartines P, Herault J (1997) Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. IEEE Trans Neural Netw 8:148–154

Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Dimitras AI, Zanakis SH, Zopounidis C (1996) A survey of business failures with an emphasis on prediction methods and industrial applications. Eur J Oper Res 90:487–513

Dimitras AI, Slowinski R, Susmaga R, Zopounidis C (1999) Business failure prediction using rough sets. Eur J Oper Res 114:263–280

Doumpos M, Zopounidis C (2007) Model combination for credit risk assessment: a stacked generalization approach. Annals Oper Res 151:289–306

Drakakis K, Rickard S, de Frein R, Cichocki A (2008) Analysis of financial data using non-negative matrix factorization. Int Math Forum 3:1853–1870

Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, New York

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman and Hall, London

Efron B, Tibshirani R (1997) Improvements on cross-validation: the .632+ bootstrap method. J Am Stat Assoc 92:548–560

Elhadi MT (2000) Bankruptcy support system: taking advantage of information retrieval and case-based reasoning. Exp Syst Appl 18:215–219

Englund C, Verikas A (2005) A SOM based model combination strategy. In: Wang J, Liao X, Yi Z (eds) Lecture notes in computer science, vol 3496. Springer, Berlin, pp 461–466

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874

Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: Proceedings of the 13th international conference on machine learning, pp 148–156

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139

Fukunaga K (1972) Introduction to statistical pattern recognition. Academic Press, New York

Gestel TV, Suykens JAK, Lanckriet G, Lambrechts A, Moor BD, Vandewalle J (2002) A bayesian framework for least squares support vector machine classifiers, Gaussian processes and kernel Fisher discriminant analysis. Neural Comput 14:1115–1147

Gestel TV, Baesens B, Suykens JAK, den Poel DV, Baestaens DE, Willekens M (2006) Bayesian kernel based classification for financial distress detection. Eur J Oper Res 172:979–1003

Gorzalczany MB, Piasta Z (1999) Neuro-fuzzy approach versus rough-set inspired methodology for intelligent decision support. Inf Sci 120:45–68

Grabisch M (1996) The representation of importance and interaction of features by fuzzy measures. Pattern Recognit Lett 17:567–575

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning: data mining, inference, and prediction (Springer series in statistics). Springer, New York

Hinton G, Roweis ST (2003) Stochastic neighbor embedding. In: Becker S, Thrun S, Obermayer K (eds) Advances in neural information processing systems, vol 15. MIT Press, Cambridge, pp 857–864

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507

Hu YC (2008) Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis. Knowle Based Syst 21:383–390

Hu YC, Tseng FM (2007) Functional-link net with fuzzy integral for bankruptcy prediction. Neurocomputing 70:2959–2968

Hua Z, Wang Y, Xu X, Zhang B, Liang L (2007) Predicting corporate financial distress based on integration of support vector machine and logistic regression. Exp Syst Appl 33:434–440

Horta RAM, de Lima BSLP, Borges CCH (2008) A semi-deterministic ensemble strategy for imbalanced datasets (SDEID) applied to bankruptcy prediction. In: Data mining IX: data minig, protection, detection and other security technologies. WIT transactions on information and communication technologies, vol 40, Spain, pp 205–213

Huysmans J, Baesens B, Vanthienen J, van Gestel T (2006) Failure prediction with self organizing maps. Exp Syst Appl 30:479–487

Ignizio JP, Soltys JR (1996) Simultaneous design and training of ontogenic neural network classifiers. Comput Oper Res 23:535–546

Islam MM, Yao X, Murase K (2003) A constructive algorithm for training cooperative neural network ensembles. IEEE Trans Neural Netw 14:820–834

Jang JR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665–685

Jeng B, Jeng YM, Liang TP (1997) FILM: a fuzzy inductive learning method for automated knowledge acquisition. Decis Support Syst 21:61–73

Jo H, Han I (1996) Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. Exp Syst Appl 11:415–422

Jo H, Han I, Lee H (1997) Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Exp Syst Appl 13:97–108

Kallel R, Cottrell M, Vigneron V (2002) Bootstrap for neural model selection. Neurocomputing 48:175–183

Keller JM, Gray MR, Givens JA (1985) A fuzzy k-nearest neighbor algorithm. IEEE Trans Syst Man Cybern 15:580–585

Kim MH, Yoo PD (2006) A semiparametric model approach to financial bankruptcy prediction. In: Engineering of intelligent systems, IEEE international conference on 2006, IEEE Press, pp 1–6

Kittler J, Hatef M, Duin RPW, Matas J (1998) On combining classifiers. IEEE Trans Pattern Anal Mach Intell 20:226–239

Kudo M, Sklansky J (2000) Comparison of algorithms that select features for pattern classifiers. Pattern Recognit 33:25–41

Kohonen T (1990) The self-organizing map. Proc IEEE 78:1461–1480

Krogh A, Vedelsby J (1995) Neural network ensembles, cross validation, and active learning. In: Tesauro G, Touretzky DS, Leen TK (eds) Advances in neural information processing systems, vol 7. MIT Press, London, pp 231–238

Kumar PR, Ravi V (2006) Bankruptcy prediction in banks by fuzzy rule based classifier. In: Proceedings of the 2006 first international conference on digital information management, pp 222–227

Kumar PR, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review. Eur J Oper Res 180:1–28

Kuncheva LI (2002) A theoretical study on six classifier fusion strategies. IEEE Trans Pattern Anal Mach Intell 24:281–286

Kuncheva LI, Whitaker CJ (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach Learn 51:181–207

Kuncheva LI, Bezdek JC, Duin RPW (2001) Decision templates for multiple classifier fusion. Pattern Recognit 34:299–314

Lawrence ND (2004) Gaussian process latent variable models for visualisation of high dimensional data. In: Thrun S, Saul LK, Schlkopf B (eds) Advances in neural information processing systems, vol 16. MIT Press, Cambridge, pp 329–336

Lawrence N (2005) Probabilistic non-linear principal component analysis with Gaussian process latent variable models. J Mach Learn Res 6:1783–1816

Lawrence ND, Quinonero-Candela J (2006) Local distance preservation in the GP-LVM through back constraints. In: Proceedings of the 23rd international conference on machine learning, Pittsburgh. ACM Press, New York, pp 513–520

Lee CH, Quek C, Maskell DL (2006) A brain inspired fuzzy neuro-predictor for bank failure analysis. In: 2006 IEEE congress on evolutionary computation, Vancouver, Canada, pp 2163–2170

Lee JA, Lendasse A, Verleysen M (2004) Nonlinear projection with curvilinear distances: isomap versus curvilinear distance analysis. Neurocomputing 57:49–76

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791

Lee KC, Han I, Kwon Y (1996) Hybrid neural network models for bankruptcy predictions. Decis Support Syst 18:63–72

Lee TS, Chiu CC, Lu CJ, Chen IF (2002) Credit scoring using the hybrid neural discriminant technique. Exp Syst Appl 23:245–254

Lespinats S, Verleysen M, Giron A, Fertil B (2007) DD-HDS: a method for visualization and exploration of high-dimensional data. IEEE Trans Neural Netw 18:1265–1279

Lin FY, McClean S (2001) A data mining approach to the prediction of corporate failure. Knowl Based Syst 14:189–195

Liu CL (2005) Classifier combination based on confidence transformation. Pattern Recognition 38:11–28

Liu Y, Yao X (1999) Ensemble learning via negative correlation. Neural Netw 12:1399–1404

Liu Y, Yao X, Higuchi T (2000) Evolutionary ensembles with negative correlation learning. IEEE Trans Evol Comput 4:380–387

Lu JJ, Tokinaga S, Ikeda Y (2006) Explanatory rule extraction based on the trained neural network and the genetic programming. J Oper Res Soc Jpn 49:66–82

MacKay DJ (1992) Bayesian interpolation. Neural Comput 4:415–447

Markham IS, Ragsdale CT (1995) Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis. Decis Sci 26:229–241

McKee TE, Lensberg T (2002) Genetic programming and rough sets: a hybrid approach to bankruptcy classification. Eur J Oper Res 138:436–451

Min SH, Lee J, Han I (2006) Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Exp Syst Appl 31:652–660

Mochon A, Quintana D, Saez Y, Isasi P (2008) Soft computing techniques applied to finance. Appl Intell 29:111–115

Narendra PM, Fukunaga K (1977) A branch and bound algorithm for feature selection. IEEE Trans Comput 26:917–922

Olmeda I, Fernandez E (1997) Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction. Comput Econ 10:317–335

Pendharkar PC, Rodger JA (2004) An empirical study of impact of crossover operators on the performance of non-binary genetic algorithm based neural approaches for classification. Comput Oper Res 31:481–498

Piramuthu S, Ragavan H, Shaw MJ (1998) Using feature construction to improve the performance of neural networks. Manage Sci 44:416–430

Quintana D, Saez Y, Mochon A, Isasi P (2008) Early bankruptcy prediction using ENPC. Appl Intell 29:157–161

Rada R (2008) Expert systems and evolutionary computing for financial investing: a review. Exp Syst Appl 34:2232–2240

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

Ravikumar P, Ravi V (2006) Bankruptcy prediction in banks by an ensemble classifier. In: Industrial technology, 2006. IEEE International Conference on ICIT 2006, pp 2032–2036

Ravi V, Reddy PJ, Zimmermann HJ (2001) Fuzzy rule base generation and its minimization via modified threshold accepting. Fuzzy Sets Syst 120:271–279

Ravi V, Kurniawan H, Thai PNK, Kumar PR (2008) Soft computing system for bank performance prediction. Appl Soft Comput 8:305–315

Refenes APN, Burgess AN, Bentz Y (1997) Neural networks in financial engineering: a study in methodology. IEEE Trans Neural Netw 8:1222–1267

Ribeiro B, Vieira A, das Neves JC (2006) Sparse Bayesian models: bankruptcy-predictors of choice? In: 2006 international joint conference on neural networks, Vancouver, pp 3377–3381

Ribeiro B, Vieira A, das Neves JC (2006) Supervised Isomap with dissimilarity measures in embedding learning. In: Ruiz-Shulcloper J, Kropatsch WG (eds) Lecture notes in computer science, vol 5197. Springer, Heidelberg, pp 389–396

Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge

Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326

Sai Y, Zhong CJ, Qu LH (2007) A hybrid GA-BP model for bankruptcy prediction. In: Proceedings of the international symposium on autonomous decentralized systems, Sedona, pp 473–477

Sammon JW (1969) A nonlinear mapping for data structure analysis. IEEE Tans Comput 18:401–409

Santosa EMD, Sabourina R, Maupin P (2008) A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. Pattern Recognition 41:2993–3009

Serrano-Cinca C (1996) Listed companies financial distress prediction based on weighted majority voting combination of multiple classifiers. Decis Support Syst 17:227–238

Shawe-Taylor J, Cristianini N (2004) Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge

Shin SW, Kilic SB (2006) Using PCA-based neural network committee model for early warning of bank failure. In: Lecture notes in computer science, vol 4221. Springer, Heidelberg, pp 289–292

Shin KS, Lee KJ (2004) Bankruptcy prediction modeling using multiple neural network models. In: Lecture notes in computer science, vol 3214. Springer, Heidelberg 668–674

Shin SW, Lee KC, Kilic SB (2006) Ensemble prediction of commercial bank failure through diversification of input features. In: Lecture notes in computer science, vol 4304. Springer, Heidelberg, pp 887–896

Sollich P, Krogh A (1996) Learning with ensembles: how over-fitting can be useful. In: Touretzky DS, Mozer MC, Hasselmo ME (eds) Advances in neural information processing systems, vol 8. MIT Press, Cambridge, pp 190–197

Sugeno M (1977) Fuzzy measures and fuzzy integrals: a survey. In: Gupta MM, Saridis GN, Gaines BR (eds) Fuzzy automata and decision process. North-Holland, Amsterdam, pp 89–102

Sun J, Li H (2008) Listed companies financial distress prediction based on weighted majority voting combination of multiple classifiers. Exp Syst Appl 35:818–827

Szupiluk R, Wojewnik P, Zabkowski T (2007) Ensemble methods with non-negative matrix factorization for non-payment prevention system. In: Proceedings of the 11th WSEAS International Conference on Systems, vol 2, systems theory and applications. Agios Nikolaos, Greece, pp 384–387

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323

Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244

Tsai CF, Wu JW (2008) Using neural network ensembles for bankruptcy prediction and credit scoring. Exp Syst Appl 34:2639–2649

Tsakonas A, Dounias G, Doumpos M, Zopounidis C (2006) Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming. Exp Syst Appl 30:449–461

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

Tung WL, Quek C, Cheng P (2004) GenSo: a novel neural-fuzzy based early warning system for predicting bank failures. Neural Netw 17:567–587

Urtasun R, Darrell T (2007) Discriminative Gaussian process latent variable model for classification. In: Proceedings of the 24th international conference on machine learning, Corvalis. ACM Press, New York, pp 927–934

Vapnik VN (1998) Statistical learning theory. Wiley, New York

Vellido A, Lisboa PJG, Vaughan J (1999) Neural networks in business: a survey of applications (1992–1998). Exp Syst Appl 17:51–70

Verikas A, Bacauskiene M (2002) Feature selection with neural networks. Pattern Recognit Lett 23:1323–1335

Verikas A, Bacauskiene M (2003) Using artificial neural networks for process and system modeling. Chemometrics Intell Lab Syst 67:187–191

Verikas A, Lipnickas A (2002) Fusing neural networks through space partitioning and fuzzy integration. Neural Process Lett 16:53–65

Verikas A, Lipnickas A, Malmqvist K, Bacauskiene M, Gelzinis A (1999) Soft combination of neural classifiers: a comparative study. Pattern Recognit Lett 20:429–444

Verikas A, Lipnickas A, Malmqvist K (2002) Selecting neural networks for a committee decision. Int J Neural Syst 12:351–361

Verikas A, Gelzinis A, Bacauskiene M, Hallander M, Uloza V, Kaseta M (2008) Combining image, voice, and the patient's questionnaire data to categorize laryngeal disorders. Artif Intell Med (submitted)

Waegeman W, Baets BD, Boullart L (2008) ROC analysis in ordinal regression learning. Pattern Recognit Lett 29:1–9

Wallrafen J, Protzel P, Popp H (1996) Genetically optimized neural network classifiers for bankruptcy prediction. In: Proceedings of the 29th annual Hawaii international conference on system sciences, Wailea, pp 419–426

West D, Dellana S (2005) Model selection strategies for ensemble solutions to bankruptcy detection. In: ICAI '05: Proceedings of the 2005 international conference on artificial intelligence, vol 1, Las Vegas, pp 46–52

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

Wolpert DH (1993) Stacked generalization. Neural Netw 5:241–259

Wong BK, Bodnovich TA, Selvi Y (1997) Neural network applications in business: a review and analysis of the literature (1988–95). Decis Support Syst 19:301–320

Wong BK, Lai VS, Lam J (2000) A bibliography of neural network business applications research: 1994–1998. Comput Oper Res 27:1045–1076

Wu CH, Tzeng GH, Goo YJ, Fang WC (2007) A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Exp Syst Appl 32:397–408

Xu L, Krzyzak A, Suen CY (1992) Methods for combining multiple classifiers and their applications to handwriting recognition. IEEE Trans Syst Man Cybern 22:418–435

Yao P (2007) Research of financial crisis prediction based on FCM-PCA-SVM model. In: ISCRAM CHINA 2007: Proceedings of the second international workshop on information systems for crisis response and management, Harbin, Peopel's Republic of China, pp 476–481

Yeung DS, Ng WWY, Chan APF, Chan PPK, Firth M, Tsang ECC (2007a) Bankruptcy prediction using multiple intelligent agent system via a localized generalization error approach. In: Service systems and service management. In: International conference on 2007, pp 1–6

Yeung DS, Ng WWY, Chan APF, Chan PPK, Firth M, Tsang ECC (2007b) A multiple intelligent agent system for credit risk prediction via an optimization of localized generalization error with diversity. J Syst Sci Syst Eng 16:166–180

Yousef WA, Wagner RF, Loew MH (2005) Estimating the uncertainty in the estimated mean area under the ROC curve of a classifier. Pattern Recognit Lett 26:2600–2610

Yu L, Lai KK, Wang SY (2007) An evolutionary programming based SVM ensemble model for corporate failure prediction. In: Lecture notes in computer science, vol 4432. Springer, Heidelberg, pp 262–270

Zhang GP (2007) Avoiding pitfalls in neural network research. IEEE Trans Syst Man Cybern C Appl Rev 37:3–16

Zhang D, Zhou L (2004) Discovering golden nuggets: data mining in financial application. IEEE Trans Syst Man Cybern C Appl Rev 34:513–522

Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14:35–62

Zhang G, Hu MY, Patuwo BE, Indro DC (1999) Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. Eur J Oper Res 116:16–32

Zhou JG, Tian JM (2007) Predicting corporate financial distress based on rough sets and wavelet support vector machine. In: 2007 international conference on wavelet analysis and pattern recognition, vol 1–4, Beijing, pp 602–607

Zopounidis C, Doumpos M, Matsatsinis NF (1997) On the use of knowledge-based decision support systems in financial management: a survey. Decis Support Syst 20:259–277