Artif Intell Rev
DOI 10.1007/s10462-015-9434-x

Financial credit risk assessment: a recent review

Ning Chen[1,2] · Bernardete Ribeiro[3] · An Chen[4,5]

Abstract The assessment of financial credit risk is an important and challenging research topic in the area of accounting and finance. Numerous efforts have been devoted into this field since the first attempt last century. Today the study of financial credit risk assessment attracts increasing attentions in the face of one of the most severe financial crisis ever observed in the world. The accurate assessment of financial credit risk and prediction of business failure play an essential role both on economics and society. For this reason, more and more methods and algorithms were proposed in the past years. From this point, it is of crucial importance to review the nowadays methods applied to financial credit risk assessment. In this paper, we summarize the traditional statistical models and state-of-the-art intelligent methods for financial distress forecasting, with the emphasis on the most recent achievements as the promising trend in this area.

Keywords Financial credit risk assessment · Business failure · Ensemble computing · Cost-sensitive learning · Dimensionality reduction · Subspace learning

✉ Ning Chen
nchenyx@outlook.com

Bernardete Ribeiro
bribeiro@dei.uc.pt

An Chen
anchen@casipm.ac.cn

1 College of Computer Science and Technology (Software College), Henan Polytechnic University, 2001 Century Avenue, Jiaozuo 454003, Henan, People's Republic of China

2 GECAD, Instituto Superior de Engenharia do Porto, Instituto Politécnico do Porto, Rua Dr. Antonio Bernardino de Almeida, 431, 4200-072 Porto, Portugal

3 CISUC, Department of Informatics Engineering, University of Coimbra, Rua Silvio Lima-Polo II, 3030-790 Coimbra, Portugal

4 Safety and Emergency Management Research Center, Henan Polytechnic University, 2001 Century Avenue, Jiaozuo 454003, Henan, People's Republic of China

5 Institute of Policy and Management, Chinese Academy of Sciences, Beijing, People's Republic of China

Published online: 27 October 2015

1 Introduction

Financial crisis has become a severe problem to the world. In the past years, many small and medium enterprises (SMEs) already weakened by the collapse of growth, and announced zero profits or losses. An increasing number of large bankruptcies are systematically announced and the financial distress spread out of all type of firms across all industries. According to the Annual Survey of Insolvency and Constitutions Company Coface in Portugal, 80 % of SMEs that live off the domestic market suffered low profitability, difficulty of self-financing and high debt levels (Coface 2012). Bankruptcies had already risen by 33 % in 2011 which reached the historic levels, and 4519 companies were declared insolvent with a remarkable increase of 35.8 % compared to the previous year, while the start-ups raised only by 13.2 % at the same time. The great loss resulted from the bankruptcies led to considerable criticism on the functionality of financial institutions due to the inappropriate evaluation of credit risk. Most governments were forced to implement rescue plans for the banking systems with more effective credit risk assessment. It becomes a particularly challenging and important issue for banks and financial institutions to access the performance of customers. With this aim, financial credit risk assessment serves as the impetus to evaluate the credit admission or potential business failure of customers in order to make early actions prior to the actual financial crisis. The rest of this paper is organized as follows. Section 2 reviews the state-of-the-art methods widely applied to financial credit risk assessment. Section 3 focuses on the recently emerged multiple classier systems which have considerable potential for performance improvement of individual prediction models. Section 4 discusses cost-sensitive and imbalanced learning as one of the major topics in financial credit risk assessment, including the cost representation and cost-sensitive learning approaches. Section 5 emphasizes the importance of dimensionality reduction to financial databases and reports the advanced subspace learning methods in financial credit risk assessment. Section 6 illustrates the performance validation approaches for the sake of well comparison and deep interpretation of prediction results. The widely used performance metrics, statistical significance test, and visualization techniques are introduced. Lastly, the limitation of the present study of financial credit risk assessment are discussed as the future research directions in Sect. 7.

2 Financial credit risk assessment methods

The financial credit risk indicates the risk associated with financing, in other words, a borrower cannot pay the lenders, or goes into default. Accordingly, financial credit risk assessment intends to solve the problem stated as follows: given a number of companies labeled as bad/good credit or bankrupt/healthy, and a set of financial variables that describe the situation of a company over a given period, predict the probability that the company may belong to a high risk group or become bankrupt during the following years. The former problem is called credit rating or scoring, and the latter problem is called bankruptcy (failure) prediction or corporate financial distress forecast. Both of them are solved in a similar way as a binary classification task. In this review, the two categories of problems are collectively called financial credit risk assessment. The executive route and methodology of the financial credit risk assessment is generally described in Fig. 1. The data under investigation comes from the financial variables of companies. After preprocessing (filtering, missing value filling, normalization etc.), feature selection tools serve for dimensionality reduction through linear (e.g., PCA) or nonlinear (e.g., ISOMAP, LLE) transformation. Afterwards, statistical, intelligent, or ensemble learning models are applied to explore the patterns from the trans-

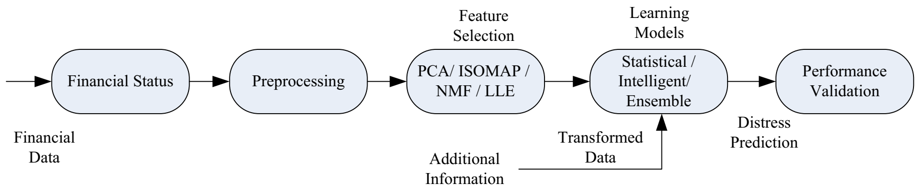

Fig. 1 Pipeline of financial credit risk assessment

formed data. The domain-specific information is integrated into the modeling algorithms to reinforce the exploration and understanding of the data. The information includes the cost preference of users, the class distribution of investigated data, the type of variables, the class label of observations, as well as some important privileged information. Finally, the prediction results are evaluated in terms of performance metrics for the sake of comparison among different classifiers.

The earliest research on financial credit risk assessment could be traced to FitzPatrick (1932) and the well-known Altman models (1968). To date a large variety of approaches have been employed to evaluate the creditworthiness of applicants using the traditional statistical methods or advanced machine learning methods. The overwhelming evidences found in numerous recent studies have shown that intelligent methods can markedly improve the accuracy of statistical methods without the reliance on restrictive assumptions. These techniques include artificial neural networks (ANNs), fuzzy set theory (FST), decision trees (DTs), case-based reasoning (CBR), support vector machines (SVMs), rough set theory (RST), genetic programming (GP), hybrid learning, and ensemble computing among others. On the other hand, a general conclusion is achieved that no method outperforms all others consistently across different data sets. Table 1 lists some recent reviews on financial crisis problem. The early review in the accounting and finance domain focused on statistics-based models (Dimitras et al. 1996; Hand and Henley 1997). Later machine learning techniques are studied extensively with more attentions. A complete review Bellovary et al. (2007) traced the historical summary of bankruptcy prediction studies and introduced the trends in this area. In another study, Ravi Kumar and Ravi (2007) reviewed the statistical and intelligent techniques of bankruptcy prediction in banks and firms. Lin et al. reviewed the development of state-of-the-art machine learning techniques for bankruptcy prediction and credit scoring from the year 1995 to 2010 (Lin et al. 2012). More recently, some researchers highlighted specific learning models as hot topic and promising trend in review articles. For example, Verikas et al. (2010) surveyed the hybrid and ensemble techniques able to improve the prediction accuracy of corporate bankruptcy, Wozniaka et al. (2014) discussed the approaches to construct multiple classifier systems, Jayanthi et al. (2011) reported the application of SVM and hybrid SVM for bankruptcy prediction, and Brabazon et al. (2012) introduced the application of neural computing to the widespread financial problems. Some studies are limited to the development of NNs techniques (Brabazon et al. 2012; Calderon and Cheh 2002; Vellido et al. 1999; Wong et al. 1997; Wong and Selvi 1998). With the ever-increasing magnitude of prediction models in the last years, there is a need to review the most recent research on financial credit risk assessment. Different from the previous reviews, we focus on the state-of-the-art approaches to the last three phases of credit risk assessment, introduce the most recent achievements in this area, and discuss some major topics including cost-sensitive learning, subspace learning and performance validation.

Table 1 Recent reviews on financial credit risk assessment

Year	Technique	Work
1996	Statistical methods	Dimitras et al. (1996)
1997	Statistical methods	Hand and Henley (1997)
1997	Neural networks	Wong et al. (1997)
1998	Neural networks	Wong and Selvi (1998)
1999	Neural networks	Vellido et al. (1999)
2000	Statistical and operational methods	Thomas (2000)
2002	Neural networks	Calderon and Cheh (2002)
2006	Statistical and machine learning methods	Balcaen and Ooghe (2006)
2007	Statistical and machine learning methods	Bellovary et al. (2007)
2007	Statistical and machine learning methods	Crook et al. (2007)
2007	Statistical and machine learning methods	Ravi Kumar and Ravi (2007)
2010	Hybrid and ensemble techniques	Verikas et al. (2010)
2011	Support vector machines	Jayanthi et al. (2011)
2012	Neural computing	Brabazon et al. (2012)
2012	Machine learning methods	Lin et al. (2012)
2013	Ensemble techniques	Wozniaka et al. (2014)
2014	Semi-parametric methods	Lam and Trinkle (2014)

2.1 Statistical models

The traditional statistical models comprise linear discriminant analysis (LDA), logistic regression (LR), multivariate discriminant analysis (MDA), quadratic discriminant analysis (QDA), factor analysis (FA), risk index models, and conditional probability model among others. The rational behind statistical models is to find an optimal linear combination of explanatory input variables able to model, analyze and predict corporate default risk. They tend to overlook the complex nature, boundaries and interrelationships of the financial variables, due to some strict assumptions such as linear separability, multivariate normality, independence of the predictive variables and pre-existing functional form. However, statistical models still belong to the most popular tools in some famous international rating agencies such as S&P, Moody's (MCO), and Fitch. In a recent study, Kouki and Elkhaldi (2011) showed the superiority of MDA and LR to neural network in medium horizon bankruptcy prediction of Tunisian firms. Kwak et al. (2012) used a multiple criteria linear programming (MCLP) to predict Korean bankruptcy after the 1997 financial crisis. They reported MCLP performed as well as MDA and LR, and comparably to DT and SVM in terms of overall prediction accuracy. A partial least square discriminant analysis (PLS-DA) model was applied to the prediction of USA banking crisis, resulting close performance to SVM (Serrano-Cinca and Gutierrez-Nieto 2013).

2.2 Semi-parametric methods

Parametric models specify the model structure a priori and define all parameters in finite-dimensional parameter spaces. A common statistical property of statistical methods is that they are fully parametric providing a clear interpretation of modeled process. On the other hand, non-parametric models determine the model structure from the data. Intelligent methods have a fully non-parametric specification of both the distributional form of variables and functional relations among them. Semi-parametric methods that represent a bridge between the statistical and intelligent models (Hwang et al. 2007, 2010; Cheng et al. 2010) can be considered as the middle way. Although clearly interpreting the modeled process, semi-parametric methods have more flexibility in model structure. Recently semi-parametric methods have become a new strand of bankruptcy prediction. Masten A. and Masten I. compared Logit parametric model, to the Klein and Spady (1993) semi-parametric model and to the non-parametric CART model for bankruptcy prediction combined with choice-based sampling. The results demonstrated semi-parametric method allowed for superior overall accuracy (Brezigar-Masten and Masten 2009). In Lam and Trinkle (2014) by combining the extant research streams of bankruptcy prediction modeling (e.g. either parametric or non-parametric) and the development of prediction intervals (semi-parametric) is possible to improve on the model accuracy. In that work, the Probit regression model forecasts the likelihood that a firm will go bankrupt and generates probability point estimates. The upper and lower bounds of the prediction intervals in consonance with the point estimates are used to improve the overall model accuracy. In a recent work Li et al. (2014) an approach combining a parametric binary logistic regression model (BLRM) and non-parametric models (e.g. SVM, DT) are put forward in a successful combination based on multiple discriminant analysis (MDA) to enhance overall performance on default prediction.

2.3 Artificial neural networks

Artificial neural networks indicate the family of neural networks of different structures, including multi-layer perceptron (MLP) or backpropagation neural network (BPNN), self-organizing map (SOM), learning vector quantization (LVQ), radial basis function network (RBFN), probabilistic neural network (PNN) among others. They have show inspiring performance in the application of bankruptcy prediction compared with other completive approaches. Their excellent capability of treating nonlinear data is beneficial to find the intrinsic patterns from complex financial data (Charalambous et al. 2000; Fu-yuan 2008). In principle, ANNs can approximate any function, but they have some weakness, such as lack explanation capability as a black-box algorithm, require consuming training time, may not provide optimal solutions, overfit to the training data. Huang et al. have demonstrated that MLP yielded excellent general performance on Taiwan and United States markets (Huang et al. 2004). Blanco et al. (2013) found that MLP outperformed the classic statistical models LDA, QDA, and LR in terms of misclassification costs for micro finance industry credit scoring. Likewise, MLP predicted the bankruptcy of Iranian companies with superior outcome to MDA (Rafiei et al. 2011). Characterized by simple topology and fast computation, RBFN received considerable well classification results in financial risk assessment (Chakraborty and Sharma 2007). In Chen et al. (2011) a stable credit rating model based on LVQ was applied to corporate failure prediction and credit risk analysis. SOM is a non-parametric neural network with the desirable combination of data abstraction and visualization. It was used to analyze and visualize the dynamic trajectory patterns of companies over several years defined on the 2D space through a two-step clustering process (Chen et al. 2013c). Korol (2013) compared the difference on bankruptcy risk forecasting between Latin America and Central Europe using DA, DT, and ANN models.

2.4 Support vector machines

Support vector machines transform the input vectors nonlinearly into a high-dimensional feature space through a kernel function so that the data can be separated by linear models. The principle of SVM is to search an optimal hyperplane able to classify the two classes and maximize the margin of the separation. Recently SMVs have gained wide popularity due to the good generalization performance on high-dimensional and a relatively small amount of data (Yang et al. 2011). Since the pioneer application directed to SVMs in bankruptcy prediction (Min and Lee 2005), there has been considerable interest in using SVM approaches to financial problems. SVMs are demonstrated particularly effective when the underlying data is typically nonlinear and non-stationary, therefore yield reasonably accurate models able to help on bank lending decisions and profitability (Jayanthi et al. 2011). Bae showed an SVM model with radial basis kernel function (RSVM) received outstanding performance for Korean manufacturing distress prediction (Bae 2012). By naturally separating the heterogeneous data into several structured groups with respect to the size and annual turnover of the firms, a corporate distress prediction model based on SVM+ achieved predominant predictability performance compared to the baseline SVM and multi-task learning with SVM (Ribeiro et al. 2012).

2.5 Decision trees

Decision trees are classic classification approach which infers a tree-shape decision structure compose of leaf nodes and decision nodes. The leaf nodes indicate the target attribute (class) of examples, and decision nodes specify the test on attributes with each branch corresponding to some specific values. DTs hold some advantages: easy interpretation of results, nonlinear estimation, non-parametric form, high accuracy, applicability to both continuous and categorical variables, and indication of important variables. There are different DTs algorithms including ID3, C45, CART, CHAID, MARS, ADTree among others. The well-known C4.5 and CART DTs have shown the applicability to bankruptcy prediction problems (Li et al. 2010). In another study, CART was used to select predictors, then a classic LR model was constructed based on the dummy variables in accordance with the tree nodes (Brezigar-Masten and Masten 2012). Delen et al. (2013) applied CHAID, C5.0, QUEST and CART to investigate the impact of financial ratios on firm performance. Out of the four DT algorithms, CHAID and C5.0 produced the best prediction accuracy.

2.6 Case-based reasoning

Case-based reasoning is able to to solve new problems and provide suggestions by recalling similar experiences. It is based on the K-nearest neighbors (KNN) principle that similar input should have the same output. A classic CBR system is composed of four steps: (1) extract KNN from the case base using a similarity kernel function (Retrieval); (2) make the prediction to the new sample according to the neighbors (Reuse); (3) validate the prediction of new sample (Revise); (4) add the new sample and its predicted class to the case base (Retain). CBR can be applied to small data set and allows the increment or deletion of samples. Although it was criticized in term of low accuracy compared with other methods such as NNs (Jo et al. 1997), the discriminant capability of CBR as a viable alternative method is affirmed for bankruptcy prediction. Park and Han (2002) improved a CBR model with feature weights derived from analytic hierarchy processing (AHP) for case indexing

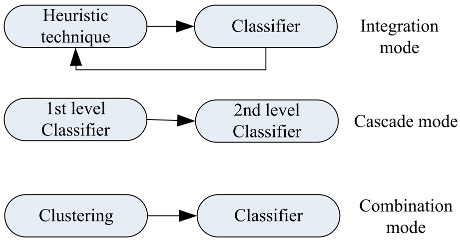

Fig. 2 Three modes to build hybrid systems

and retrieval, and Li et al. (2011) proposed a new CBR forecasting method based on the similarities to both positive and negative ideal cases.

2.7 Hybrid learning models

In a hybrid system, several heterogeneous techniques are exploited in the analysis but normally only one is employed for the final prediction. Figure 2 shows three approaches to build a hybrid system: integration mode, cascade mode, and clustering combination mode (Lin et al. 2012). In the first approach, the heuristic techniques are integrated with classification models for optimizing the prediction performance from several views: (1) tune the parameters of learning algorithm; (2) select the relevant features; (3) refine the samples for learning. Recently many researchers have shown that the performance of learning models can be significantly enhanced by the hybridization of heuristic techniques such as genetic algorithm (GA), annealing simulation (AS), particle swarm optimization (PSO), ant colony optimization (ACO). Min et al. used GA to optimize the input futures and parameters of an SVM model in order to further improve the prediction ability (Min et al. 2011). Likewise, Tsai et al. (2013) applied GA to perform both feature selection and instance selection in prior to an SVM and KNN classifier. They demonstrated that feature selection followed by instance selection in data preprocessing is helpful to the subsequent classification. Swarm intelligence represented by ant colony optimization and particle swarm optimization was employed to automatically select the optimal features used by the subsequent machine learning models (Lin et al. 2008; Marinakis et al. 2009. By integrated artificial bee colony (ABC) with SVM, the hybrid classification model induced an improvement on corporate credit rating of America (Chen et al. 2013a). The second approach to build a hybrid system is cascading different classifiers, where the output of the first-level classifier feeds to the second-level classifier as input. Ravisankar et al. used MLP, PNN, RST and GP to construct different two-phase hybrids, each taking one technique as feature selection method followed by another one as classifier. They reported the GP–GP performed superior to other hybrids in terms of accuracy, sensitivity, specificity and AUC (Ravisankar et al. 2010). Yeh el al. (Yeh et al. 2012) proposed a hybrid model, which combines random forests (RF) and RST to extract useful information, and applied KMV to evaluate the credit rating of corporations from the market-based information. Chuang (2013) developed several hybrids, namely RST-CBR (RST + CBR), RST-GRA-CBR (RST + Grey Relational Analysis + CBR), and CART-CBR (CART + CBR) and produced better prediction accuracy than using CBR alone. The results demonstrated the need to supplement CBR with other classification techniques. In the third approach, clustering is used as a preprocess step of classification to enhance the prediction accuracy. The non-representative samples can be eliminated by unsupervised clustering, then the clustering results are used as training data to identify the patterns by supervised classification (Jain et al. 1999). Peng et al. (2005) used CLUTO clustering to classify credit card dataset and two classification methods to improve the clustering results. Esfandiary et al. (2013) proposed a layered decision tree (LDT) approach which clusters the data and sorts them with respect to the importance, then divides the data into groups to construct one layer of DT. Ribeiro and Chen (2012a) proposed a two-step approach to predict the financial risk of French companies, first discovering the bicluster patterns upholding instances and features highly correlated, then exploiting a subspace learning model with constant regularization, followed by a SVM classification of the projected data. Recent work on clustering consensus Lourenco et al. (2015) that builds ensemble from a set of base clustering algorithms is raising interest able to discover new structures in data and hence improve the subsequent classification.

3 Multiple classifier systems

Multiple classifier system (MCS) is constructed by combining multiple classifiers, which are trained individually and aggregated in an ensemble manner for the final decision. The classifiers within the MCS are called base (component, or elementary) classifiers. The principle of a success MCS is to comprise appropriate elementary classifiers, which have high performance individually and low intercorrelation so as to ensure the effective fusion of the ensemble. MCS is regarded as a promising computing technique able to produce dominating prediction performance compared to individual models (Lorena et al. 2008; Tulyakov et al. 2008). A lot of prior studies claimed that MCS is actually beneficial to improve the standalone classifiers although it was not always reported superior to the single best classifier.

3.1 Construction of multiple classifier systems

So far a variety of strategies are commonly used to design a multiple classifier system. In general, these approaches can be divided into four categories shown in Fig. 3. From the viewpoint of the variance among elementary classifiers, ensemble members can be developed by: (1) varying the training data using resampling or replication techniques, for example Bagging and Boosting; (2) varying the feature sets of training data by means of feature selection methods, for example Random Subspace and Rotation Forest; (3) varying the parameters of a specific classifier, for example the kernel function of KNN classifier, the topology of a MLP neural network; (4) varying the type of classifier. These methods aim to build a set of base classifiers, which have high individual performance and diversity that leads to independent misclassifications. Recent efforts adopted several different strategies simultaneously to further increase the diversity of elementary classifiers. Wang and Ma (2012) combined Bagging and Random Subspace to form a hybrid ensemble SVM model. Sun and Li constructed a number of candidate classifiers trained by an SVM using different kernel functions (varying the parameters of a SVM classifier) and feature subsets (varying the features of training data), then selected the appropriate members with the consideration of both individual performance and diversity. The resulting SVM ensemble achieved superior performance to individual SVM (Sun and Li 2012). Erdal (2013) investigated three different ensemble approaches of DT, namely single ensemble, two-level ensemble of single strategy, two-level ensemble of different strategies, with the results showing that the two-level ensembles could noticeably advance the prediction accuracy of individual DT and single ensemble. From the perspective of fusion scheme, ensemble members can be combined in the manner of linear approach, nonlinear approach, statistical approach, or intelligent approach (Canuto et al. 2007). The linear combination calculates the sum and average of the output of base classifiers. The nonlinear combination aggregates the outcome of base classifiers through some rank-based methods such as simple

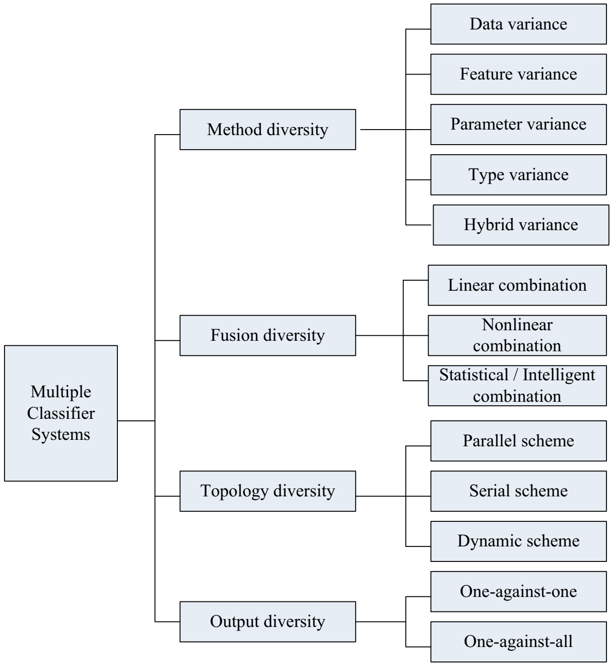

Fig. 3 Construction approaches of multiple classifier systems

majority voting and weighted voting. Additionally, the multiple classifiers can be combined using statistical methods (e.g., Bayesian combination) or intelligent methods (e.g., neural network, genetic algorithm). Zhang et al. (2013) built a finance early-warning model based on LR and SVM, whose outcome are combined by Dempster Shafer theory (DST), a general extension of Bayesian theory. They reported the method could well improve the prediction accuracy of single method for classification. From the viewpoint of topology, Finlay (2011) summarized the multiple classifier systems into three categories: static parallel systems, multi-stage systems, and dynamic classifier systems. Static systems, taking Bagging for instance, construct the component classifiers independently with the same input data, then combine them on the basis of individual classifier output in some manner. Multi-stage systems establish the component classifiers iteratively, with some parameters dependent on the previously generated classifiers. Boosting is a prominent member of this topology where individual classifiers are constructed in sequence. This serial topology is of benefit to enhance the capability of weak classifiers by focusing on the difficultly identified samples sequentially. Dynamic systems develop separate classifiers in different ranges of problem domain and choose the best one with respect to individual performance. From the viewpoint of output, multiple classifier systems can be built by the manipulation of individual classifier outputs. The decomposition is based on divide-and-conquer principle, which divides a complex problem into a set of sub-problems. By decomposing a multi-class classification problem into a set of binary classification problems, each member only classifiers some classes. The popular decomposition techniques include one-against-one and one-against-all. The binary classifiers are then combined through pairwise coupling, max–min rule or weighted voting (Wozniaka et al. 2014).

3.2 Diversity measures

Diversity among elementary classifiers, also called multicollinearity problem, is crucial to the success of an MCS. The individual classifiers produce different and independent misclassifications so that the combination hopefully achieves high accuracy. The diversity of an MCS can be measured by outcome diversity and structure diversity. Structure diversity evaluates how the classifiers are varied in the structure. Musehane et al. (2008) used Shannon–Wiener and Simpson to measure the diversity among multiple structurally different NNs. However, it is inapplicable to classifiers of different types. Outcome diversity evaluates how the classifiers are different in the outcome. The commonly used outcome diversity measures include pairwise measures and non-pairwise measures. Pairwise measures average the diversity between each classifier pair, including Q-statistic, correlation coefficient, disagreement, double fault. Alteratively, non-pairwise measures directly calculate the diversity of all elementary classifiers, including entropy measure, Kohavi–Wolpert variance, inter-rater agreement, difficulty index, generalized diversity, and coincident failure (Kuncheva and Whitaker 2003).

3.3 Multiple classifier systems in financial credit risk assessment

In the scope of financial credit risk assessment, MCS emerged in recent years to boost the performance of single prediction model. Many researchers have provided the supporting evidences that MCS methodology can produce better performance than individual credit risk assessment models (Deligianni and Kotsiantis 2012; Kuncheva 2004; Rokach 2010; Verikas et al. 2010; Zhou 2012). It is encouraging that the ensemble with some simple classifiers can produce prominent prediction results (Sun et al. 2011). Nanni and Lumini tested four ensemble strategies using four diversified classifiers on three financial data sets. They found random subspace produced the best performance in terms of AUC (Nanni and Lumini 2009). Li and Sun used CBR as the base classifier to construct an ensemble model based on diversified feature selection methods and weighted majority voting principle. The results confirmed the ensemble of multiple classifiers evidently improved the accuracy and stability of prediction (Li and Sun 2011b). Marqués et al. (2012) used resampling and feature selection techniques jointly to construct effective composite ensembles for credit scoring problems. Li and Sun generated diverse CBR predictors using random similarity kernel functions and produced significantly better performance than single CBR (Li and Sun 2013) on Chinese hotel business failure prediction. In Xie et al. (2013), the output of LR is introduced into SVM and MLP to achieve outstanding prediction for corporate distress in fashion and textile supply chains. GA was employed as a coverage optimization technique to solve the multicollinearity problem within an ensemble of multiple classifiers for bankruptcy prediction on Korea firms (Kim and Kang 2012). Some researchers implemented hybrid structural ensembles based on diversified types of learning algorithms. Table 2 lists some recent multiple classifier systems of hybrid structure. These studies indicated that a well-designed ensemble could inherit advantages and avoid disadvantages of the employed methods, and thus outperform the stand-alone classifiers. The advantage of hybrid structure within an ensemble is discussed in some studies. Canuto et al. investigated the relationship between the choice of ensemble members and accuracy through an extensive evaluation using different data sets and combination methods. They found that the hybrid ensemble structures outperformed the non-hybrid ensemble structures consistently in the accuracy (Canuto et al. 2007).

Table 2 Some hybrid structural multiple classifier systems

Work	Composite classifiers
Chen and Ribeiro (2013)	LR, SVM, C45, ADTree, KNN, MLP, Decision Table, RBF, Bayesian, Bayesian LR
Hung and Chen (2009)	DT, MLP, SVM
Ravi et al. (2008)	MLP, RBFN, PNN, SVM, CART, fuzzy rule-based classifier, PCA-MLP, PCARBF, PCA-PNN
Ravikumar and Ravi (2006)	ANFIS, SVM, Linear RBF, Semi-online RBF1 and Semi-online RBF2, Orthogonal RBF, MLP
Soltan and Mohammadi (2012)	MLP, RBFN, DT
Sun and Li (2008)	MDA, LR, NN, SVM, DT, CBR
Xie et al. (2013)	LR, SVM, MLP
Zhang et al. (2013)	LR, SVM

4 Cost-sensitive and imbalanced financial credit risk assessment

The real world financial failure prediction problem is typically characterized by cost-sensitivity and class-imbalance as a focus of recent study. Cost-sensitivity indicates that the misclassification costs are non-uniform, in other words, the cost due to classifying a bad credit (denoted as positive class) as a good one (denoted as negative class) is usually greater than that due to classifying a good credit as a bad one, so the performance of a classifier on the distressed class is consequently more important than the overall performance. Class-imbalance indicates that the class distribution of real world database is highly skewed, in other words, the bad credit samples (e.g., out-of-control companies) are overwhelmed by the good credit samples (e.g., well-managed companies) so that the classification tends to perform better on the majority class than on the minority class. However, the minority one is usually the class of primary interest in decision making, e.g., the distress in bankruptcy prediction. The two problems can be treated within a unified framework (Zhou and Liu 2006). On one hand, an applicable way to make an algorithm cost-sensitive is to intentionally imbalance the training data by means of resampling techniques. On the other hand, a solution to class imbalance problem is to train a cost-sensitive classifier with the misclassification cost of the minority class greater than that of the majority class. To manipulate the two problems simultaneously, a possible solution is to merge the class imbalance ratio with cost ratio through rescaling (Liu and Zhou 2006). However the equivalence of the two approaches has not yet been confirmed in terms of performance improvement. Experiments conducted in Chen et al. (2013b) using a real world French bankruptcy database found that both class imbalance ratio and cost ratio have prominent effect on the classification performance. A near-balanced training data set is favorable when a relatively uniform cost ratio is used, whereas a near-natural class distribution is favorable when a highly uneven cost ratio is used. In this review we introduce firstly the cost-sensitive learning approaches, and then the class imbalance classification methods for financial credit assessment applications.

4.1 Cost representation

With respect to the category of Turney (2000), there are different categories of misclassification cost: constant error cost of class, conditional error cost dependent on individual instance,

Table 3 A 2D misclassification cost matrix

Real class	Predicted class	
	Positive	Negative
Positive	0	C_p
Negative	C_n	0

Positive bad credit or bankrupt, *negative* good credit or healthy

time cost, feature cost and test cost. The present day research mostly concentrates on constant cost, which is simply represented and easily handled in practice. Constant cost indicates the instances of a class i have the same cost due to misclassifying as class j. Regarding the financial failure prediction which considers a binary classification, the misclassification cost can be represented in a 2D matrix in Table 3. The diagonal values of the matrix are 0, and the off-diagonal values are the cost of two kinds of errors where C_p denotes the misclassification cost associated with positive (bad credit or bankrupt) class, and C_n denotes the misclassification cost associated with negative (good credit or healthy) class. Normally, $C_p > C_n$ indicating the misclassification cost of a positive company is higher than that of a negative company.

4.2 Cost-sensitive learning approaches

By definition, cost-sensitive learning addresses the challenging problems in which there are asymmetric costs associated with different misclassification errors. Different strategies have been proposed to make a standard classification algorithm cost-sensitive, such as resampling, instance-weighting, instance-relabeling, post hoc threshold adjusting, and algorithm-level approach. Resampling is a traditionally approach to handle the asymmetric costs, either by over-sampling the costly class instances, by under-sampling the less costly class instances, or by systematically combining both over-sampling and under-sampling. Generally, resampling techniques can be categorized into random sampling and heuristic sampling. The former simply replicates or discards some instances at random, and the latter generates new instances or selectively removes some redundant instances based on the inherent characteristic of the data (Eitrich et al. 2007; Khalilia et al. 2011; Yin and Leong 2010). Although resampling is easily implemented, it distorts the distribution of the samples and might decrease the performance of classifiers. Instance weighting reweights the instances and then feeds them with weights to the employed classifiers. The weight of an instance is determined by the misclassification cost by assigning higher weight to the instance if the misclassification leads to more cost. The first attempt of this approach might be the greedy divide-conquer algorithm to induce a cost-sensitive tree (Breiman et al. 1984). Later the weighted C4.5 assigns the samples of different classes with different weights proportional to the corresponding costs (Ting 2002). Similarly, the cost matrix is incorporated into regularized least square (Vo and Won 2007). MetaCost provides a general solution to covert an arbitrary classifier into a cost-sensitive algorithm through relabeling the training samples with their estimated minimal-cost class (Domingos 1999). It is applicable to any number of classes and arbitrary cost matrix requiring no knowledge of the employed classifier. Post hoc threshold adjusting manipulates the output of classifiers learned in the standard way and makes prediction in a cost-sensitive manner. It selects a threshold, which moves the output towards the expensive class. Regarding DTs, the standard way is to label the leaf nodes with the common majority class. With the consideration of costs, the leaf nodes are alternatively assigned the class label which produces the minimal misclassification cost (Zadrozny and Elkan 2001). Threshold moving is applicable to classifiers of real-valued output. A bisection method is applied to neural network to find the optimal threshold which minimizes the total misclassification cost (Pendharkar (2008)). In addition to the general approaches applicable to most learning methods, the algorithm-level approaches attempt to make a specific classification method cost-sensitive by changing the learning methodology. Chen et al. (2009) studied a cost-sensitive LVQ for financial distress prediction by integrating the cost matrix into the learning of network. They further applied GA to optimize simultaneously the complexity and connected weights of LVQ network (Chen et al. 2011).

4.3 Classification of imbalanced financial data

The standard classification algorithms perform unsatisfactorily on imbalanced data due to the maximum-generality bias (Ting 1994), i.e., they tend to discover more general rules than the specialized rules that fit for the rare class. Classification of imbalanced data received extensive attention in data mining domain. A review Sun et al. (2009) presented a thorough introduction of reported research solutions to this problem addressing their advantages and constraints. Generally the reported approaches to classification of imbalanced data fall into two groups: data-level methods and algorithm-level methods. Data-level methods comprise a variety of sampling (to rebalance the class distribution) and boosting (to put more weights on the misclassified samples that usually belong to the rare class). The algorithm-level methods aim to strengthen the learning performance of existing algorithms with regards to the small class using some inductive bias, for example new pruning strategies of DTs (Zadrozny and Elkan 2001) and different penalty constants of SVM (Lin et al. 2002). Cost-sensitive learning is another widely used solution to imbalanced data classification that can be conducted in both data level and algorithm level. The cost-sensitive learning methods implemented by instance-weighting or instance-relabeling are considered as the data level solutions. An advanced solution is cost-sensitive boosting that introduces cost items into boosting ensemble, such as AdaC1, AdaC2, and AdaC3 (Sun et al. 2006, 2007). Generally the data-level approach is more preferable in practice because it is applicable to most existing classifiers with little change to the underlying learning strategy. However, it suffers from some difficulties, such as the setup of cost item when cost-sensitive learning is used. In some cases the cost item can be determined by the real values, such as the lost of a missing bankrupt company. Otherwise it can be set with respect to the relative importance of the costly class over the other. For example, the cost ratio (C_p/C_n) is set as 10 if the misclassification of a bankrupt company (positive class) is ten times important than that of a good one (negative class).

5 Dimensionality reduction and subspace learning

Dimensionality reduction is closely related to financial credit risk assessment in the sense that the financial databases usually comprise considerable variables ranging from one to 50s (Bellovary et al. 2007). For example, the widely used public databases Australia and German Frank and Asuncion (2010) contain 14 and 20 variables respectively. The commonly used corporate financial ratios comprise liquidity, solvability, productivity of labor and capital, margins, return on investment etc. The magnitude of variables leads to the dimensionality curse problem in data mining problems. Some recent empirical studies have explicitly highlighted the important role of dimensionality reduction with the conclusion that the selection of representative variables certainly increases the performance of prediction (Tsai 2009). Zhou et al. (2012) tested six feature ranking techniques with 21 quantitative models on two firm databases of USA and China. They found that although no feature ranking method can improve the performance of underlying classifiers consistently, the properly selected features could actually reduce the model complexity without the loss of performance. Subspace learning is essentially a dimensionality reduction approach to find a compact representation in a low dimensional data subspace of original high dimensional data. The subspace learning approaches include linear methods and nonlinear methods.

5.1 Linear methods

The linear subspace learning techniques include principal component analysis (PCA), singular value decomposition (SVD), and FA among others. They are effective when the data lies in or near in a linear subspace and produce improved discriminant results integrated with the classifiers (Pai et al. 2004). Chen and Vieira (2009) employed independent component analysis (ICA) to further improve the prediction accuracy of learning models. Li and Sun (2011a) employed PCA with several filer approaches for feature selection so that the subsequent statistical models can achieve reasonably well performance to predict the business failure.

5.2 Nonlinear methods

Linear methods assume the data lies in or near in a linear subspace, which however does not satisfy in financial data with complex nonlinear structures. Nonlinear subspace learning overcomes the limitations of linear methods, therefore decreases the generalization error of classification. Within the family of nonlinear methods, manifold learning shows the noteworthy applicability in the area of financial analysis. They are able to find the low-dimensional structures embedded in the high dimensional data space, so that subsequently the classification and prediction algorithms can be easily applied to the relatively low-dimensional spaces. The well-known manifold learning methods consist of isometric feature map (ISOMAP), locally linear embedding (LLE), Laplacian Eigenmaps (LE), and non-negative matrix factorization (NMF) amongst others. These techniques tend to cast the high-dimensional data into low-dimensional manifolds with few degrees of freedom and embedded intrinsic geometry by means of some nonlinear projections. Despite the application of nonlinear subspace learning in a wide range of areas, its development in financial and accounting problems has undergone minor advances. Only a few efforts have been devoted into this research topic. LLE and ISOMAP are demonstrated to produce superior business failure prediction accuracy to linear dimensionality reduction and feature selection methods such as PCA when integrated with SVM (Lin et al. 2011, 2013). Orsenigo and Vercellis (2013) employed PCA and double-bounded tree-connected ISOMAP as dimensionality reduction method, and found the nonlinear projection consistently provided more accurate predictions on six bank data sets. Ribeiro et al. (2008) demonstrated manifold learning is particularly suitable for bankruptcy prediction problem. They also incorporated the class information in an enhanced supervised ISOMAP algorithm (ES-ISOMAP) to uncover the embedded geometry structure of financial data (Ribeiro et al. 2008). More recently, they used graph regularized non-negative matrix factorization (GNMF) and spatially smooth subspace learning (SSSL) to extract the most discriminative features, subsequently construct a classification model for failure prediction (Ribeiro and Chen 2011). By combining biclustering with SSSL in a supervised learning manner, an SVM classification yielded further desirable results (Ribeiro and Chen 2012b). These empirical evidences attained support the fact that manifold learning is a valuable technique to enhance the prediction power of classification models.

Table 4 Contingency matrix of prediction results

Real class	Predicted class	
	Positive	Negative
Positive	tp	fn
Negative	fp	tn

Positive bad credit or bankrupt, *negative* good credit or healthy

6 Performance validation of credit risk assessment

6.1 In-sample and out-of-sample performance validation

In-sample and out-of-sample are two widely used approaches to evaluate the performance of prediction models. Most of the studies use in-sample performance validation, so the model is validated on the same sample on which the analysis is performed. There are also papers that perform out-of-sample validation, meaning that they test the model on a different sample rather than the original one that was used for estimating (Brezigar-Masten and Masten 2012). Cross-validation is a classic approach of out-of-sample validation. From the viewpoint of generalization error estimation, out-of-sample validation is generally considered less sensitive to outliers and more trustworthy than in-sample validation. In data mining tasks, in-sample is usually used for parameter optimization and model selection, while out-of-sample is used for performance evaluation (Hansen and Timmermann 2012). In bankruptcy prediction, out-of-sample is more relevant, since all the institutions that use bankruptcy prediction models at work actually face out-of-sample prediction. Li and Miu (2010) conducted an empirical study by terms of both in-sample and out-of-sample validation, showing the superior performance of a hybrid bankruptcy prediction model with dynamic loadings.

6.2 Performance metrics

Normally the result of credit risk assessment is a binary value: bad credit/bankrupt or good credit/healthy. Accordingly, the outcome of a classifier can be represented as a 2D contingency matrix shown in Table 4, where fn denotes the misclassification errors of bad credit/bankrupt companies as good credit/healthy (false negative error), fp denotes the misclassification errors of good credit/healthy companies as credit/bankrupt (false positive error), tp denotes the correct predictions of positive samples (true positive), and tn denotes the correction predictions of negative samples (true negative).

The performance of bankruptcy prediction models is usually evaluated in terms of some metrics. Table 5 lists some widely used metrics, such as accuracy (ACC), root mean squared error (RMSE), true positive rate (TPR), true negative rate (TNR), precision, F-score, area under ROC curve (AUC), lift, and SAR. In the case of asymmetric cost involved in the classification, expected misclassified cost (EMC), a cost-relevant measure, is favorable to incorporate the cost information through a tradeoff between two kinds of errors.

6.3 Comparison and interpretation of classification performance

In credit risk assessment, a less studied problem than developing powerful models is the performance validation of classifiers. According to the no free lunch theorem of machine learning, multiple classifiers are commonly generated to solve the same prediction problem. The comparison and interpretation of classification results becomes a key point.

Table 5 Performance metrics of prediction models

Metric	Definition	Description
Accuracy	$(tp+tn)/N$	Proportion of correct predictions out of the total samples
Root mean squared error (RMSE)	$\sqrt{\sum(P(x)-T(x))^2/N}$	How the predicted output (P(x)) deviates from the desired output (T(x))
True positive rate (TPR)	$\frac{tp}{tp+fn}$	Proportion of actual positives which are identified as positive
True negative rate (TNR)	$\frac{tn}{tn+fp}$	Proportion of actual negatives which are identified as negative
False positive rate (FPR)	$\frac{fp}{tn+fp}$	Proportion of actual negatives which are identified as positive
False negative rate (FNR)	$\frac{fn}{tp+fn}$	Proportion of actual positives which are identified as negative
Precision	$\frac{tp}{tp+fp}$	Fraction of true positives that are predicted as positive
Recall	$\frac{tp}{tp+fn}$	Fraction of true positives that are actually positive, i.e., TPR
Sensitivity	$\frac{tp}{tp+fn}$	Fraction of true positives that are actually positive
Specificity	$\frac{tn}{tn+fp}$	Fraction of true negatives that are actually negative
F-score	$2\frac{Precision*Recall}{Precision+Recall}$	Harmonic mean of Precision and Recall which reaches the best value at 1 and the worst score at 0
Area under ROC Curve (AUC)	–	Area under ROC (Receiver Operating Characteristic), a graphical plot of TPR versus FPR indicating if one classifier dominates another
Lift	–	Priority of a classifier predicting a positive sample over a random prediction
Calibration	–	How well a classifier is calibrated
Break-even point	–	Precision at a threshold where precision and recall are equal
SAR	$\frac{1}{3}(ACC+AUC+(1-RMSE))$	Combination of RMSE, AUC and Accuracy
Expected misclassified cost (EMC)	$C_p*\frac{fn}{N}+C_n*\frac{fp}{N}$	Tradeoff between FPR and FNR

N total number of samples, C_p misclassification cost of bankrupt company, C_n misclassification cost of healthy company

6.3.1 Comparison of classifiers on single metric

Statistical significance test is an important component of performance validation of multiple classifiers. It provides the evidence that one classifier is statistically better than another at a significance level. The commonly used significance test includes parametric test and non-parametric test. Parametric tests, such as the mostly used t test, determine if two sets of data are significantly different from each other in the mean. If the p value is small, then the null hypothesis is rejected, in other words, the two sets are significantly different. The parametric tests require the independency, normality, and homoscedasticity of investigated data, which are difficultly verified in machine learning studies (Garcia et al. 2010). Alternatively, non-parametric tests are particularly suitable for this task without these assumptions. Non-parametric statistical tests comprise a set of methods, such as Wilcoxon signed-rank test, Friedman test, Jonckheere–Terpstra test, and Kruskal–Wallis test. Wilcoxon signed-rank test calculates the absolute difference of each pair of classifiers, assigns the ranks according to the absolute differences, and then calculates the test statistic W as the sum of the signed ranks. The p value can be computed from W and N_r (number of pairs). If the p value is small, then the null hypothesis is rejected, indicating that there is significant difference between the two competing classifiers. Friedman test (Demsar 2006) is used for repeated measures analysis of variance by ranks to detect the differences among the classifiers. It defines the null hypothesis, as there is no significant difference among the performance of all classifiers. Let n be the number of problems, and k the number of classifiers. For each data set, the classifiers are ranked separately, where the best performance has rank 1, the second has rank 2, etc. If there are ties, the average rank is assigned to the tied values. We denote r_i^j as the rank of j-th classifier on the i-th problem. For each investigated classifier, the average rank is calculated as $R_j = \sum_i r_i^j / n$. The null hypothesis states that all classifiers behave similarly so that the average ranks should be equal. Therefore, Friedman statistic $\chi_F^2 = \frac{12n}{k(k+1)}[\sum_j R_j^2 - \frac{k(k+1)^2}{4}]$ is distributed according to χ_F^2 with $k-1$ degrees of freedom. If the resulting p value is less than a predefined level α (normally $\alpha = 5\,\%$), the null hypothesis is rejected. After the significant difference within the classifiers is verified, a post-hoc Nemenyi test is then undertaken for pairwise comparison. If their average rank between two classifiers differs at least the critical value $CD = q_\alpha\sqrt{\frac{k(k+1)}{6n}}$, the performance of the two competitors is significantly different at the α level.

6.3.2 Comparison of classifiers on multiple metrics

In most previous studies, the performance of classifiers is evaluated by single or a few metrics, such as the commonly used accuracy. However recent studies have shown the need to measure the quality of classification results from different perspectives. The difficulty of comparing and interpreting the results becomes particularly critical when facing with a multitude of empirical results of multiple performance metrics for various learning models upon different data sets. The study of performance metrics is still limited. Caruana and Niculescu-Mizil (2004) investigated the correlation among nine widely used performance metrics. According to their study, these metrics are divided into three groups: threshold metrics (accuracy and F-score), ordering metrics (AUC, precision, break-even point, and lift), and probability metrics (RMSE, cross entropy, and calibration). In another study, a component-wise aggregation method was designed to combine the results of different classifiers more precisely than the standard approaches such as simple average and win/loss/tie (Japkowicz et al. 2008). In order to reach the valid and useful conclusions, a proper way is probably to utilize multiple metrics and summarize the metrics through effective aggregation, clustering and visualization techniques. Multi-dimensional Scaling (MDS) is demonstrated particularly applicable to interpret the classification performance by projecting the high-dimensional metric data into a 2D space, easing the comparison among the studied classifiers from the relative positions (Alaiz-Rodriguez et al. 2008; García et al. 2012). In details, for each classifier a metric vector consists of the values of different data sets and metrics, then projected into a 2D space. Hence, each point indicates the performance of a classifier. Furthermore, an optimal vector is composed of the best performance for each dimension and also projected into the 2D space. In this way it becomes easy to compare the performance of different classifiers from the relative positions between the studied classifiers and the optimal one.

7 Conclusions and future remarks

In this review, we introduce the current of financial credit risk assessment research, discuss the traditional and advanced methods, and illustrate some major topics in the recent study. Despite the overwhelming research in this area, the present day study is limited in some aspects, which pose the open problems for future study. (1) A large number of nowadays research focus on the failure prediction problem (FPP), which is normally solved as a binary classification task taking the distribution of the outputs as bimodal instead of Gaussian. A well-structured credit-scoring scheme is more important in differentiating the degree of credit risk in the various credit exposures of a bank loan portfolio. This would allow a more accurate determination of the overall characteristics of loan portfolio, probability of default, value at risk and ultimately the adequacy of provisions for loan losses. (2) Most previous studies use corporate financial ratios for the construction of predictive models without the consideration of macroeconomic information. Due to the influence of macroeconomic conditions on the corporate performance, the models incorporating macroeconomic variables are expected to improve the prediction ability. (3) Although cost-sensitive learning has been studied using different strategies to reflect the cost preference in decision-making, most are focused on the simple class-dependent cost form and work well in binary classification problem. Efforts should be undertaken to adapt these algorithms to more complicated definitions of cost and multimodal classification problem. (4) Although performance validation starts to attract some attentions in credit risk assessment, it still needs further investigation to select appropriate metric measures and well interpret the large multitude of empirical results. In this sense, the mining from metric space becomes an urgent task with practical need for decision makers. (5) Most of the priori studies attempt to explore static models from historical financial statements. The bankruptcy trajectory receives little attention and only a few attempted to analyze the temporal sequence of financial statements. Trajectory data of financial statements makes possible to detect the time evolution of companies and recognize the trajectory patterns. Trajectory mining would be interesting as a future research direction in this area. (6) In most systems the financial rule extraction is completely algorithmic or automatic with little supervision and interaction of users. For the sake of acquiring useful and understandable knowledge, there exists a critical need to integrate the users into the black-box process through a visual and interactive framework.

Acknowledgments This work was partly supported by Natural Science Foundation of China (Contract No. 91024004).

References

Alaiz-Rodriguez R, Japkowicz N, Tischer P (2008) A visualization-based exploratory tool for classifier comparison with respect to multiple metrics and multiple domains. In: Proceedings of ECML PKDD, pp 660–665

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finance 23(4):589–609

Bae JK (2012) Predicting financial distress of the South Korean manufacturing industries. Expert Syst Appl 39(10):9159–9165

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

Bellovary J, Giacomino D, Akers M (2007) A review of bankruptcy prediction studies: 1930 to present. J Financ Educ 33:1–43

Blanco A, Pino-Mejias R, Lara J, Rayo S (2013) Credit scoring models for the microfinance industry using neural networks: evidence from Peru. Expert Syst Appl 40(1):356–364

Brabazon A, Dang J, Dempsey I, O'Neill M, Edelman D (2012) Natural computing in finance: a review. In: Rozenberg G, Back T, Kok J (eds) Handbook of natural computing. Springer, Berlin, pp 1707–1735

Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth, Belmont, CA

Brezigar-Masten A, Masten I (2009) Comparison of parametric, semi-parametric and non-parametric methods in bankruptcy prediction. IMAD Working Paper Series XVIII, vol 18

Brezigar-Masten A, Masten I (2012) CART-based selection of bankruptcy predictors for the logit model. Expert Syst Appl 39(11):10153–10159

Calderon TG, Cheh JJ (2002) A roadmap for future neural networks research in auditing and risk assessment. Int J Account Inf Syst 3(4):203–236

Canuto AM, Abreu MC, Oliveira LM Jr, Xavier JC, Santos AM (2007) Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles. Pattern Recognit Lett 28(4):472–486

Caruana R, Niculescu-Mizil A (2004) Data mining in metric space: an empirical analysis of suppervised learning performance criteria. In: Proceedings of the 10th international conference on knowledge discovery and data mining

Chakraborty S, Sharma SK (2007) Prediction of corporate financial health by artificial neural network. Int J Electron Finance 1(4):442–459

Charalambous C, Charitou A, Kaourou F (2000) Application of feature extractive algorithm to bankruptcy prediction. Int Jt Conf Neural Netw 5:303–308

Chen MY, Chen CC, Liu JY (2013) Credit rating analysis with support vector machines and artificial bee colony algorithm. In: Ali M, Bosse T, Hindriks K, Hoogendoorn M, Jonker CM, Treur J (eds) Recent trends in applied artificial intelligence, LNCS, vol 7906. Springer, Berlin, pp 528–534

Chen N, Chen A, Ribeiro B (2013) Influence of class distribution on cost-sensitive learning: a case study of french bankruptcy analysis. Int J Intell Data Anal 17(3):423–437

Chen N, Ribeiro B (2013) A consensus approach for combining multiple classifiers in cost-sensitive bankruptcy prediction. In: M.T. et al (ed.) 11th international conference on adaptive and natural computing algorithms (ICANNGA'13), LNCS, vol 7824. Springer, Berlin, pp 266–276

Chen N, Ribeiro B, Vieira A, Chen A (2013) Clustering and visualization of bankruptcy trajectory using self-organizing map. Expert Syst Appl 40(1):385–393

Chen N, Ribeiro B, Vieira A, Duarte J, Neves J (2011) A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. Expert Syst Appl 38(10):12939–12945

Chen N, Vieira A (2009) Bankruptcy prediction based on independent component analysis. In: 1st international conference on agents and artificial intelligence (ICAART09). pp 150–155

Chen N, Vieira A, Duarte J, Ribeiro B, Neves J (2009) Cost-sensitive learning vector quantization for financial distress prediction. In: Lecture notes in artificial intelligence (LNAI 5816). Springer, Berlin, pp 374–385

Chen N, Vieira A, Ribeiro B, Duarte J, Neves J (2011) A stable credit rating model based on learning vector quantization. Intell Data Anal 15(2):237–250

Cheng KF, Chu CK, Hwang R (2010) Predicting bankruptcy using the discrete-time semi-parametric hazard model. Quant Finance 10(9):1055–1066

Chuang CL (2013) Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. Inf Sci 236:174–185

Coface, for Safer Trade (2012) Risk assessment of Portugal. http://www.coface.com/Economic-Studies-and-Country-Risks/Portugal

Crook JN, Edelman DB, Thomas LC (2007) Recent developments in consumer credit risk assessment. Eur J Oper Res 183(3):1447–1465

Delen D, Kuzey C, Uyar A (2013) Measuring firm performance using financial ratios: a decision tree approach. Expert Syst Appl 40(10):3970–3983

Deligianni D, Kotsiantis S (2012) Forecasting corporate bankruptcy with an ensemble of classifiers. In: Maglogiannis I, Plagianakos V, Vlahavas I (eds) Artificial intelligence: theories and applications, LNCS, vol 7297. Springer, Berlin, pp 65–72

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

Dimitras A, Zanakis S, Zopounidis C (1996) A survey of business failures with an emphasis on prediction methods and industrial applications. Eur J Oper Res 90(3):487–513

Domingos P (1999) Metacost: a general method for making classifiers cost-sensitive. In: Proceedings of 5th ACM SIGKDD international conference on knowledge discovery and data mining. pp 155–164

Eitrich T, Kless A, Druska C, Meyer W, Grotendorst J (2007) Classification of highly unbalanced CYP450 data of drugs using cost sensitive machine learning techniques. J Chem Inf Model 47:92–103

Erdal HI (2013) Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction. Eng Appl Artif Intell 26(7):1689–1697

Esfandiary N, Azad I, Eftekhari Moghadam AM (2013) Ldt: layered decision tree based on data clustering. In: Proceedings of the 13th Iranian conference on fuzzy systems (IFSC). pp 1–4

Finlay S (2011) Multiple classifier architectures and their application to credit risk assessment. Eur J Oper Res 210(2):368–378

FitzPatrick PJ (1932) A comparison of the ratios of successful industrial enterprises with those of failed companies. J Account Res 10:598–605

Frank A, Asuncion A (2010) UCI machine learning repository. http://archive.ics.uci.edu/ml

Fu-yuan H (2008) A genetic fuzzy neural network for bankruptcy prediction in chinese corporations. In: International conference on risk management and engineering management (ICRMEM ’08). pp 542–546

Garcia S, Fernandez A, Luengo J, Herrera F (2010) Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Inf Sci 180(10):2044–2064

García V, Sánchez JS, Mollineda RA (2012) On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl Based Syst 25:13–21

Hand DJ, Henley WE (1997) Statistical classification methods in consumer credit scoring: a review. J R Stat Soc Ser A (Stat Soc) 160(3):523–541

Hansen PR, Timmermann A (2012) Choice of sample split in out-of-sample forecast evaluation. Economics Working Papers ECO2012/10

Huang Z, Chen H, Hsu CJ, Chen WH, Wu S (2004) Credit rating analysis with support vector machines and neural networks: a market comparative study. Decis Support Syst 37(4):543–558

Hung C, Chen JH (2009) A selective ensemble based on expected probabilities for bankruptcy prediction. Expert Syst Appl 36(3, Part 1):5297–5303

Hwang R, Cheng KF, Lee J (2007) A semi-parametric method for predicting bankruptcy. J Forecast 26:317–342

Hwang R, Ruey-Ching, Chung H, Chu C (2010) Predicting issuer credit ratings using a semi-parametric method. J Empir Finance 17(1):120–137

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv 31(3):264–323

Japkowicz N, Sanghi P, Tischer P (2008) A projection-based framework for classifier performance evaluation. In: Proceedings of European conference on machine learning and knowledge discovery in databases-part 1, vol 5211. LNCS Springer, Heidelberg, pp 548–563

Jayanthi J, Joseph KS, Vaishnavi J (2011) Bankruptcy prediction using SVM and hybrid SVM survey. Int J Comput Appl 33(7):39–45

Jo H, Han I, Lee H (1997) Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Expert Syst Appl 13(2):97–108

Khalilia M, Chakrabort S, Popescu M (2011) Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 11(17):51

Kim MJ, Kang DK (2012) Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. Expert Syst Appl 39(10):9308–9314

Klein RW, Spady RH (1993) An efficient semiparametric estimator for binary response models. Econometrica 61(2):387–421

Korol T (2013) Early warning models against bankruptcy risk for central european and latin american enterprises. Econ Model 31:22–30

Kouki M, Elkhaldi A (2011) Toward a predicting model of firm bankruptcy: evidence from the Tunisian context. Middle East Finance Econ 14:26–43

Kuncheva LI (2004) Combining pattern classifiers. Wiley, New York
Kuncheva LI, Whitaker CJ (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach Learn 51(2):181–207
Kwak W, Shi Y, Kou G (2012) Bankruptcy prediction for Korean firms after the 1997 financial crisis: using a multiple criteria linear programming data mining approach. Rev Quant Finance Account 38(4):441–453
Lam M, Trinkle BS (2014) Using prediction intervals to improve information quality of bankruptcy prediction models, chap. 8, pp 37–52
Li H, Adeli H, Sun J, Han JG (2011) Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction. Comput Oper Res 38(2):409–419
Li H, Sun J (2011) Empirical research of hybridizing principal component analysis with multivariate discriminant analysis and logistic regression for business failure prediction. Expert Syst Appl 38(5):6244–6253
Li H, Sun J (2011) Principal component case-based reasoning ensemble for business failure prediction. Inf Manage 48(6):220–227
Li H, Sun J (2013) Predicting business failure using an RSF-based case-based reasoning ensemble forecasting method. J Forecast 32(2):180–192
Li H, Sun J, Wu J (2010) Predicting business failure using classification and regression tree: an empirical comparison with popular classical statistical methods and top classification mining methods. Expert Syst Appl 37(8):5895–5904
Li J, Pan L, Chen M, Yang X (2014) Parametric and non-parametric combination model to enhance overall performance on default prediction. J Syst Sci Complex 27(5):950–969. doi:10.1007/s11424-014-3273-8
Li MYL, Miu P (2010) A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information. J Empir Finance 17(4):818–833
Lin F, Yeh C, Lee M (2013) A hybrid business failure prediction model using locally linear embedding and support vector machines. Rom J Econ Forecast 1:82–97
Lin F, Yeh CC, Lee MY (2011) The use of hybrid manifold learning and support vector machines in the prediction of business failure. Knowl Based Syst 24(1):95–101
Lin SW, Ying KC, Chen SC, Lee ZJ (2008) Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst Appl 35(4):1817–1824
Lin WY, Hu YH, Tsai CF (2012) Machine learning in financial crisis prediction: a survey. IEEE Trans Syst Man Cybern C Appl Rev 42(4):421–436
Lin Y, Lee Y, Wahba G (2002) Support vector machines for classification in nonstandard situations. Mach Learn 46:191–202
Liu XY, Zhou Z (2006) The influence of class imbalance on cost-sensitive learning: An empirical study. In: Proceedings of 6th IEEE international conference on data mining (ICDM06). pp 970–974
Lorena AC, Carvalho AC, Gama JM (2008) A review on the combination of binary classifiers in multiclass problems. Artif Intell Rev 30(1–4):19–37
Lourenco A, Bulo SR, Rebagliati N, Fred ALN, Figueiredo MAT, Pelillo M (2015) Probabilistic consensus clustering using evidence accumulation. Mach Learn 98(1–2):331–357
Marinakis Y, Marinaki M, Doumpos M, Zopounidis C (2009) Ant colony and particle swarm optimization for financial classification problems. Expert Syst Appl 36(7):10604–10611
Marqués A, García V, Sánchez J (2012) Two-level classifier ensembles for credit risk assessment. Expert Syst Appl 39(12):10916–10922
Min JH, Jeong C, Kim M (2011) Tuning the architecture of support vector machine: the case of bankruptcy prediction. Int J Manage Sci 17(1):1–116
Min JH, Lee YC (2005) Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Syst Appl 28(4):603–614
Musehane R, Netshiongolwe F, Nelwamondo FV, Masisi L, Marwala T (2008) Relationship between diversity and perfomance of multiple classifiers for decision support. Comput Res Repos. abs/0810.3
Nanni L, Lumini A (2009) An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert Syst Appl 36(2, Part 2):3028–3033
Orsenigo C, Vercellis C (2013) Linear versus nonlinear dimensionality reduction for banks credit rating prediction. Knowl Based Syst 47:14–22
Pai GR, Annapoorani R, Pai GV (2004) Performance analysis of a statistical and an evolutionary neural network based classifier for the prediction of industrial bankruptcy. In: IEEE conference on cybernetics and intelligent systems. pp 1033–1038
Park CS, Han I (2002) A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. Expert Syst Appl 23(3):255–264
Pendharkar P (2008) A threshold varying bisection method for cost sensitive learning in neural networks. Expert Syst Appl 34:1456–1464

Peng Y, Kou G, Shi Y, Chen Z (2005) Improving clustering analysis for credit card accounts classification. Lect Notes Comput Sci 3516:548–553
Rafiei FM, Manzari S, Bostanian S (2011) Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. Expert Syst Appl 38(8):10210–10217
Ravi V, Kurniawan H, Thai PNK, Kumar PR (2008) Soft computing system for bank performance prediction. Appl Soft Comput 8(1):305–315
Ravi Kumar P, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques: a review. Eur J Oper Res 180(1):1–28
Ravikumar P, Ravi V (2006) Bankruptcy prediction in banks by an ensemble classifier. In: IEEE international conference on industrial technology. pp 2032–2036
Ravisankar P, Ravi V, Bose I (2010) Failure prediction of dotcom companies using neural networkcgenetic programming hybrids. Inf Sci 180(8):1257–1267
Ribeiro B, Chen N (2011) Graph weighted subspace learning models in bankruptcy. In: Proceedings IEEE international joint conference on neural networks (IJCNN). pp 2055–2061
Ribeiro B, Chen N (2012a) Biclustering and subspace learning with regularization for financial risk analysis. In: Proceedings of international conference on neural information processing, part II, LNCS, vol 7664. pp 616–623
Ribeiro B, Chen N (2012b) Biclustering and subspace learning with regularization for financial risk analysis. In: T.H. et al. (ed.) Proceedings of the 19th international conference on neural information processing (ICONIP), part III, LNCS, vol 7665. Springer, Berlin, pp 228–235
Ribeiro B, Silva C, Chen N, Vieira A, Neves J (2012) Enhanced default disk models with SVM+. Expert Syst Appl 39:10140–10152
Ribeiro B, Vieira A, Duarte J, Silva C, Neves J, Liu Q, Sung A (2009) Learning manifolds for bankruptcy analysis. In: M. Köppen, et al. (eds.) International conference on neural information processing, vol 5506. LNCS, Springer, Berlin, pp 722–729
Ribeiro B, Vieira A, Neves JC (2008) Supervised Isomap with dissimilarity measures in embedding learning. LNCS 5197:389–396
Rokach L (2010) Pattern classification using ensemble methods. World Scientific Publishing, Singapore
Serrano-Cinca C, Gutierrez-Nieto B (2013) Partial least square discriminant analysis for bankruptcy prediction. Decis Support Syst 54(3):1245–1255
Soltan A, Mohammadi M (2012) A hybrid model using decision tree and neural network for credit scoring problem. Manage Sci Lett 2(5):1683–1688
Sun J, Jia M, Li H (2011) AdaBoost ensemble for financial distress prediction: an empirical comparison with data from Chinese listed companies. Expert Syst Appl 38(8):9305–9312
Sun J, Li H (2008) Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers. Expert Syst Appl 35(3):818–827
Sun J, Li H (2012) Financial distress prediction using support vector machines: ensemble versus individual. Appl Soft Comput 12(8):2254–2265
Sun Y, Kamel MS, Wang Y (2006) Boosting for learning multiple classes with imbalanced class distribution. In: Proceedings of the sixth IEEE international conference on data mining. pp 592–602
Sun Y, Kamela M, Wong A, Wang Y (2007) Cost-sensitive boosting for classification of imbalanced data. Pattern Recogn 40:3358–3378
Sun Y, Wong AC, Kamel MS (2009) Classification of imbalanced data: a review. Int J Pattern Recognit Artif Intell 23(4):687–719
Thomas LC (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. Int J Forecast 16(2):149–172
Ting K (2002) An instance-weighting method to induce costsensitive trees. IEEE Trans Knowl Data Eng 14(3):659–665
Ting KM (1994) The problem of small disjuncts: its remedy in decision trees. In: Proceedings of the tenth Canadian conference on artificial intelligence. pp 91–97
Tsai CF (2009) Feature selection in bankruptcy prediction. Knowl Based Syst 22(2):120–127
Tsai CF, Eberle W, Chu CY (2013) Genetic algorithms in feature and instance selection. Knowl Based Syst 39:240–247
Tulyakov S, Jaeger S, Govindaraju V, Doermann D (2008) Review of classifier combination methods. In: Marinai S, Fujisawa H (eds) Machine learning in document analysis and recognition, studies in computational intelligence, vol 90. Springer, Berlin, pp 361–386
Turney P (2000) Types of cost in inductive concept leaning. In: Workshop on cost-sensitive learning at 7th international conference on machine learning

Vellido A, Lisboa P, Vaughan J (1999) Neural networks in business: a survey of applications (1992–1998). Expert Syst Appl 17(1):51–70
Verikas A, Kalsyte Z, Bacauskiene M, Gelzinis A (2010) Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. Soft Comput 14(9):995–1010
Vo N, Won Y (2007) Classification of unbalanced medical data with weighted regularized least squares. In: Frontiers in the convergence of bioscience and information technologies. pp 347–352
Wang G, Ma J (2012) A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine. Expert Syst Appl 39(5):5325–5331
Wong BK, Bodnovich TA, Selvi Y (1997) Neural network applications in business: a review and analysis of the literature (1988–1995). Decis Support Syst 19(4):301–320
Wong BK, Selvi Y (1998) Neural network applications in finance: a review and analysis of literature (1990–1996). Inf Manage 34(3):129–139
Wozniaka M, Granb M, Corchado E (2014) A survey of multiple classifier systems as hybrid systems. Inf Fusion 16:3–17
Xie G, Zhao Y, Jiang M, Zhang N (2013) A novel ensemble learning approach for corporate financial distress forecasting in fashion and textiles supply chains. Math Probl Eng 23(2):388–400
Yang Z, You W, Ji G (2011) Using partial least squares and support vector machines for bankruptcy prediction. Expert Syst Appl 38(7):8336–8342
Yeh CC, Lin F, Hsu CY (2012) A hybrid KMV model, random forests and rough set theory approach for credit rating. Knowl Based Syst 33:166–172
Yin H, Leong T (2010) A model driven approach to imbalanced data sampling in medical decision making. Stud Health Technol Inform 160(Pt 2):856–860
Zadrozny B, Elkan C (2001) Learning and making decisions when costs and probabilities are both unknown. In: Proceedings of the seventh international conference on knowledge discovery and data mining. pp 204–213
Zhang L, Zhang L, Teng W, Chen Y (2013) Based on information fusion technique with data mining in the application of finance early-warning. Proc Comput Sci 17:695–703
Zhou L, Lai KK, Yen J (2012) Empirical models based on features ranking techniques for corporate financial distress prediction. Comput Math Appl 64(8):2484–2496
Zhou Z (2012) Ensemble methods: foundations and algorithms. CRC Press, Boca Racton
Zhou Z, Liu X (2006) Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Trans Knowl Data Eng 18(1):63–77