ELSEVIER

Expert Systems with Applications 28 (2005) 743–752

Expert Systems with Applications

www.elsevier.com/locate/eswa

A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines

Tian-Shyug Lee[a,*], I-Fei Chen[b]

[a]*Graduate Institute of Management, Fu-Jen Catholic University, No. 510, Chung-Cheng Road, Hsinchuang, Hsin-Chuang, Taipei 24205, Taiwan, ROC*

[b]*Graduate Institute of Business Administration, Fu-Jen Catholic University, Hsin-Chuang, Taipei, Taiwan, ROC*

Abstract

The objective of the proposed study is to explore the performance of credit scoring using a two-stage hybrid modeling procedure with artificial neural networks and multivariate adaptive regression splines (MARS). The rationale under the analyses is firstly to use MARS in building the credit scoring model, the obtained significant variables are then served as the input nodes of the neural networks model. To demonstrate the effectiveness and feasibility of the proposed modeling procedure, credit scoring tasks are performed on one bank housing loan dataset using cross-validation approach. As the results reveal, the proposed hybrid approach outperforms the results using discriminant analysis, logistic regression, artificial neural networks and MARS and hence provides an alternative in handling credit scoring tasks.

Keywords: Credit scoring; Classification; Neural networks; Multivariate adaptive regression splines; Cross-validation

1. Introduction

Credit risk evaluation decisions are crucial for financial institutions due to high risks associated with inappropriate credit decisions that may result in huge amount of losses. It is an even more important task today as financial institutions have been experiencing serious challenges and competition during the past decade. When considering the case regarding the application for a large loan, such as a mortgage or a construction loan, the lender tends to use the direct and individual scrutiny by a loan officer or even a committee. However, if hundreds of thousands, even millions of credit card or consumer loan applications need to be evaluated, the financial institutions will usually adopt models to assign scores to applicants rather than examining each one in detail. Hence various credit scoring models need to be developed for the purpose of efficient credit approval decisions.

With the tremendous growth of the credit industry and the diversified loan portfolios nowadays, credit scoring has gained more and more attention as the credit industry can then benefit from on time decisions, reducing possible risks, improving cash flow, and insuring proper credit collections. Aiming to satisfy the above-mentioned needs, many different useful techniques, known as the credit scoring models, have been developed by financial institutions and researchers in order to solve the problems involved during the evaluation process. The objective of credit scoring models is to assign credit applicants to either a 'good credit' group that is likely to repay financial obligation or a 'bad credit' group, with high possibility of defaulting on the financial obligation, whose application should be denied. Therefore credit scoring lies in the domain of the more general and widely discussed classification problems (Anderson, 1984; Dillon & Goldstein, 1984; Johnson & Wichern, 2002). The classification problems where items/observations can be assigned to one of several known disjoint groups have long played important roles in business related decision making due to its wide applications in decision support, financial forecasting, fraud detection, marketing strategy, process control, and other related fields (Cabena, Hadjinaian, Stadler, Verhees, & Zanasi, 1997, Chen et al., 1996, Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Usually, credit scoring is applied to rank credit information based on the application form details and

* Corresponding author. Tel.: +886 2 2903 1111x2905; fax: +886 2 2901 8475.

E-mail address: badm1004@mails.fju.edu.tw (T.-S. Lee).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2004.12.031

other relevant information held by a credit reference agency. As the results, accounts with high possibility of default can be monitored and necessary actions can be taken in order to prevent the account from being default. In response, the statistical methods, non-parametric methods, and artificial intelligence approaches have been proposed to support the credit approval decision process (Desai, Crook, & Overstreet, 1996; Lee, Chiu, Lu, & Chen, 2002; Thomas, 2000; West, 2000).

After careful review of the crediting scoring literature, it can be concluded that linear discriminant analysis (LDA) and logistic regression were the two most commonly used statistical techniques in building credit scoring models. However, the utilization of linear discriminant analysis has often been criticized due to the assumptions of linear relationship between dependent and independent variables, which seldom holds, and the fact that it is sensitive to deviations from the multivariate normality assumption (Karels & Prakash, 1987; Reichert, Cho, & Wagner, 1983). Theoretically, quadratic discriminant analysis (QDA) should be adopted when the covariance matrices of the different populations are unequal. However, QDA seems to be more sensitive to the model assumptions than LDA and LDA has reported to be a more robust and precise method (Dillon & Goldstein, 1984; Sharma, 1996).[1] In addition to the LDA approach, logistic regression is another commonly utilized alternative to conduct credit scoring tasks. Basically, the logistic regression model was emerged as the technique in predicting dichotomous outcomes. Logistic regression does not require the multivariate normality assumption, however, the dependent variable exposed to a full linear relationship among independent variables in the exponent of the logistic function. Basically, both LDA and logistic regression are designed for the case when the underlying relationship between variables are linear and hence are reported to be lack of enough credit scoring accuracy (Thomas, 2000; West, 2000).

Artificial neural networks provide a new alternative to LDA and logistic regression in handling credit scoring tasks, particularly in situations where the dependent and independent variables exhibit complex non-linear relationships. Even though neural networks have reported to provide better credit scoring accuracy than those using LDA and logistic regression (Desai et al., 1996; Jensen, 1992; Lee et al., 2002; Piramuthu, 1999; West, 2000), it is, however, also being criticized for its long training process in obtaining the optimal network's topology, not easy to identify the relative importance of potential input variables, and certain interpretive difficulties and hence has limited its applicability in handling general classification and credit scoring problems (Craven & Shavlik, 1997; Lee et al., 2002; Piramuthu, 1999).

In addition to the above-mentioned techniques, multivariate adaptive regression splines (MARS) is another commonly discussed classification technique nowadays (Friedman, 1991). MARS is widely accepted by researchers and practitioners for the following reasons. Firstly, without the drawbacks of LDA and logistic regression, MARS is capable of modeling complex non-linear relationship among variables without strong model assumptions. On the other hand, unlike neural networks, MARS can capture the relative importance of independent variables to the dependent variable when many potential independent variables are considered. Thirdly, MARS does not need long training process and hence can save lots of model building time, especially when the dataset is huge. Finally, one strong advantage of MARS over other classification techniques is the resulting model can be easily interpreted. It not only points out which variables are important in classifying objects/observations, but also indicates a particular object/observation belongs to a specific class when the built rules are satisfied. The final fact has important managerial and interpretative implications and can help to make appropriate decisions.

Based on the above-mentioned modeling advantages of MARS, the authors believe that MARS should be a good supporting tool for neural networks as the technical merits of MARS are just the shortcomings of neural networks. Using MARS as a first-stage modeling tool with the obtained results being the inputs to neural networks should contribute to the success of the subsequent model building tasks. Focusing on improving the above-mentioned drawbacks of neural networks credit scoring models, the purpose of this study is to explore the performance of credit scoring with a two-stage hybrid modeling procedure using artificial neural networks and multivariate adaptive regression splines (MARS). The rationale underlying the analyses is firstly to use MARS in building the scoring model, the obtained significant variables are then used as the input variables of the designed neural networks model. Please note that, according to the knowledge of the authors, there still does not exist a theoretical method, which can optimally determine the appropriate input nodes of a neural networks model; MARS can be implemented as a generally accepted method for identifying important variables when many potential independent variables are considered. Finally, as the two-stage modeling procedure will use the obtained significant variables from MARS as input nodes, hence it can reduce the number of input nodes, simply the network structure, and shorten the model building time.

To demonstrate the feasibility and effectiveness of the proposed two-stage credit scoring procedure, credit scoring tasks are performed on one housing loan dataset. As cross-validation is the preferred procedure in testing the out-of-sample classification capability of the built classification model (Breiman, Friedman, Olshen, & Stone, 1984;

[1] Since LDA has reported to be a more robust method than QDA when the theoretical presumptions are violated, hence the LDA approach will be used in building the credit scoring model in this study.

Johnson & Wichern, 2002) when the dataset size is small, the five-fold cross-validation scheme will be conducted to test the scoring capability of the proposed model. Empirical results demonstrated that the proposed two-stage modeling procedure outperforms linear discrimintant analysis, logistic regression, multivariate adaptive regression splines and backpropagation neural networks. Besides, using MARS as a supporting tool for designing the network topology can help to identify important independent variables and contribute to the success of the subsequent policy design.

The rest of the paper is organized as follows. We will review the credit scoring literature in Section 2. Section 3 gives a brief outline of multivariate adaptive regression splines. The developments and the empirical results of credit scoring models using linear discriminant analysis, logistic regression, BPN neural networks, MARS and the two-stage hybrid method are presented in Section 4. To verify the effectiveness of the designed two-stage hybrid model, the comparison of the credit scoring results of the five built models in terms of the minimum expected misclassification cost criterion is summarized in Section 5. Finally, Section 6 addresses the conclusion and discusses possible future research areas.

2. Literature review

We will review the literature of credit scoring and the commonly used techniques in modeling credit scoring problems in this section.

2.1. Discriminant analysis

Fisher (1936) first proposed discriminant analysis as a classification technique. Up to date, it has been reported as the most commonly used technique in modeling classification and the credit scoring problems (Lee, Sung, & Chang, 1999; Thomas, 2000). As a matter of fact, discriminant analysis has been widely devoted to a considerably wide range of application areas, such as medicine, business, education, marketing research, finance, chemistry, biology, engineering and archaeology (Altman, 1968; Deakin, 1972; Kim, Kim, Kim, Ye, & Lee, 2000; Lee, Jo, & Han, 1997; Trevino & Daniels, 1995). In addition, Bardos (1998); Desai et al. (1996); Martell and Fitts (1981); Overstreet, Bradley, and Kemp (1992); Reichert et al. (1983) and Titterington (1992) also proposed using discriminant analysis in building credit scoring models.

Discriminant analysis is the first widely adopted statistical methodology in building credit scoring models. However, it has also been criticized for lack of classification precision due to the fact that it is primarily designed for capturing linear relationships among variables. Based on this drawback, researchers are forced to search for new alternatives. In summary, discriminant analysis provides the decision maker with a binary outcome of the credit applicant under study. The obtained result, though important, does not provide any estimate of the associated risk. Based of this idea, logic regression is proposed instead since it can estimate the associated probability of an applicant's credit status.

2.2. Logistic regression

Logistic regression is a widely used statistical modeling technique in which the probability of a dichotomous outcome is related to a set of potential independent variables (Cox & Snell, 1989; Hosmer & Lemeshow, 1989). The logistic regression model does not require the assumptions of discriminant analysis. However, Harrell and Lee (1985) found that logistic regression is as efficient and accurate as discriminant analysis. Besides, since logistic regression models can estimate the associated probability of an applicant's credit status and hence give a better understanding of the distribution of the financial risk than the discriminant analysis approach. The objective of a logistic regression model is to determine the conditional probability of a specific observation belonging to a class, given the values of the independent variables of that credit applicant.

Logistic regression models have been widely discussed in social research, medical research, design, control, bankruptcy prediction, market segmentation, and customer behaviors (Flagg, Giroux, & Wiggins, 1991; Laitinen & Laitinen, 2000; Lau, 1987; Suh, Noh, & Suh, 1999). Logistic regression has also been explored by Joanes (1993); Laitinen (1999); Westgaard and van der Wijst (2001); and Wiginton (1980) in building credit scoring models.

2.3. Artificial neural networks

A neural network is a system comprised of highly interconnected, interacting processing units that are based on neuro-biological models. Neural networks process information through the interactions of a large number of processing units (the neurons or nodes, we will use them interchangeably thereafter) and their connections to external inputs. A three-layer backpropagation neural networks (BPN) is shown in Fig. 1. BPN is a gradient steepest

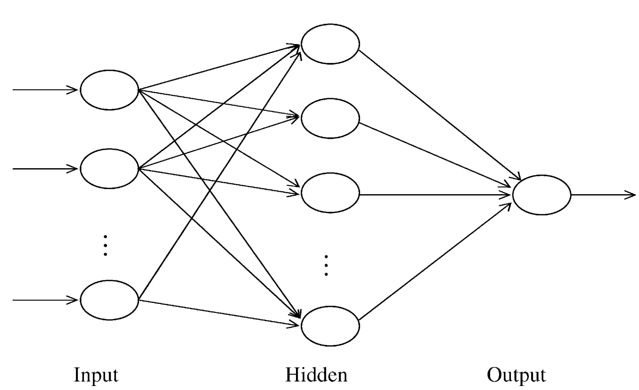

Fig. 1. A three-layer backpropagation neural networks.

descent training algorithm and has been the most often utilized paradigm to date in business applications (Vellido, Lisboa, & Vaughan, 1999). The network consists of a number of neurons connected by links. The nodes in the network can be classified as three different layers: the input layer, the output layer, and one or more hidden layers. The nodes in the input layer receive input signals from external sources and the nodes in the output layer provide the target output signals. For the gradient descent algorithm, the step size, called the learning rate, is crucial since smaller learning rates tend to slow down the learning process before convergence while larger ones may cause network oscillation and unable to converge.

Neural networks are increasingly found to be useful in modeling non-stationary processes due to its outstanding generalization capability (Anderson & Rosenfeld, 1988; Cheng & Titterington, 1994; Haykin, 1994; Repley, 1994; Stern, 1996). Based on these facts, neural networks have been widely used in engineering, science, education, social research, medical research, business, finance, forecasting and related fields (Chiu, Shao, Lee, & Lee, 2003; Lee & Chen, 2002; Lee & Chiu, 2002; Repley, 1994; Stern, 1996; Vellido et al., 1999; Zhang, Patuwo, & Hu, 1998). Neural networks have also been explored by Arminger, Enache, and Bonne (1997); Barney, Graves, and Johnson (1999); Deng (1993); Desai et al. (1996); Glorfeld (1996); Glorfeld and Hardgrave (1996); Jagielska and Jaworski (1996); Jensen (1992); Lee et al. (2002); Piramuthu (1999); Piramuthu, Shaw, and Gentry (1994); Richeson, Zimmermann, and Barnett (1994); Torsun (1996); and West (2000) in handling credit scoring problems. As neural networks are primarily designed to capture subtle functional relationship among variables, the majority of the above references have reported that the credit scoring accuracy of neural networks are better than those using discriminant analysis and logistic regression techniques.

3. Multivariate adaptive regression splines

Multivariate adaptive regression splines (MARS), a non-linear and non-parametric regression methodology, is first proposed by Friedman (1991) as a flexible procedure which models relationships that are nearly additive or involve interactions with fewer variables. The modeling procedure of MARS is basically inspired by the recursive partitioning technique governing classification and regression tree (CART, Breiman et al., 1984) and generalized additive modeling (Hastie & Tibshirani, 1990), resulting in a model that is continuous with continuous derivatives. MARS excels at finding optimal variable transformations and interactions, as well as the complex data structure that often hides in high-dimensional data. And hence can effectively unveil important data patterns and relationships that are difficult, if not impossible, for other methods to reveal.

MARS essentially builds flexible models by fitting piecewise linear regressions; that is, the non-linearity of a model is approximated through the use of separate linear regression slopes in distinct intervals of the independent variable space. Therefore the slope of the regression line is allowed to change from one interval to the other as the two 'knot' points are crossed. The variables to be used and the end points of the intervals for each variable are found through a fast but intensive search procedure. In addition to searching variables one by one, MARS also searches for interactions between variables, allowing any degree of interaction to be considered as long as the built model can better fit the data.

The general MARS function can be represented using the following equation (Friedman, 1991)

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ \quad (1)$$

where a_0 and a_m are parameters, M is the number of basis functions, K_m is the number of knots, s_{km} takes on values of either 1 or -1 and indicates the right/left sense of the associated step function, $v(k,m)$ is the label of the independent variable, and t_{km} indicates the knot location.

The optimal MARS model is selected in a two-stage process. Firstly, MARS constructs a very large number of basis functions to overfit the data initially, where variables are allowed to enter as continuous, categorical, or ordinal—the formal mechanism by which variable intervals are defined, and they can interact with each other or be restricted to enter in only as additive components. In the second stage, basis functions are deleted in the order of least contributions using the generalized cross-validation (GCV) criterion. A measure of variable importance can then be assessed by observing the decrease in the calculated GCV when a variable is removed from the model. This process will continue until the remaining basis functions all satisfying the pre-determined requirements. The GCV can be expressed as follows

$$\mathrm{LOF}(\hat{f}_M) = \mathrm{GCV}(M) = \frac{1}{N} \sum_{i=1}^{N} [y_i - \hat{f}_M(x_i)]^2 \Big/ \left[1 - \frac{C(M)}{N}\right]^2 \quad (2)$$

where there are N observations, and $C(M)$ is the cost-penalty measures of a model containing M basis functions (therefore the numerator measures the lack of fit on the M basis function model $f_M(x_i)$ and the denominator denotes the penalty for model complexity $C(M)$).

MARS can also handle the missing-value problems using dummy variable skills. By allowing any arbitrary shape for the function and interactions as well as using the above-mentioned two-stage model building procedure, MARS is capable of tracking very complex data structures that often hide in high-dimensional data. Please refer to Friedman

(1991) for more details regarding the model building process.

MARS has been widely used in modeling problems in the areas of forecasting and classification problems (De Gooijer, Ray, & Krager, 1998; Friedman & Roosen, 1995; Griffin, Fisher, Friedman, & Ryan, 1997; Kuhnert, Do, & McClure, 2000; Lewis & Stevens, 1991; Nguyen-Cong, Van, & Rode, 1996; Ohmann, Moustakis, Yang, & Lang, 1996). For a detailed list regarding the referred articles using MARS, the readers are recommended to login in to the website provided by Salford Systems http://www.salford-systems.com/MARSCITE.PDF for more details and descriptions.

4. Empirical Study

In order to verify the feasibility and effectiveness of the proposed two-stage credit scoring model using artificial neural networks and multivariate adaptive regression splines, one housing loan dataset provided by a local bank in Taipei, Taiwan is used in this study. There are totally 510 housing loan customers in the dataset with 459 good credit customers while the remaining 51 are bad credit customers. The 10% relative ratio of bad credit customers to total customers is very close to the national standard in Taiwan and hence should be a representative dataset in verifying the feasibility of the proposed scheme. Each bank customer in the dataset contains 18 independent variables which can be summarized in Table 1 and the dependent variable is the credit status of the customer-good or bad credit.

In order to minimize the possible bias associated with the random sampling of the training and testing samples, researchers tend to use *n*-fold cross-validation scheme in evaluating the classification capability of the built model. In *n*-fold cross-validation, the entire dataset is randomly split into *n* mutually exclusively subsets (also called folds) of approximately equal size with respect to the ratios of different populations. The classification model will then be trained and tested *n* times. Each time the model is built using $(n-1)$ folds as the training sample and the remaining single fold is retained for testing. The training sample is used to estimate the credit scoring model's parameters while the retained holdout sample is used to test the generalization capability of the built model. The overall classification accuracy of the built model is then just the simple average of the *n* individual accuracy measures. As cross-validation is the preferred procedure in testing the out-of-sample classification capability when the dataset size is small (Breiman et al., 1984; Johnson & Wichern, 2002) and the size of bad credit customers is only 51, the five-fold cross-validation will be adopted in this study. Therefore there are 102 customers in each fold of the dataset.

Table 1
List of independent variables in building the credit scoring model

Variables	Ratios and/or quantities
X_1	Gender
X_2	Age
X_3	Marital status
X_4	Educational level
X_5	Occupation
X_6	Years working at the current company
X_7	Monthly income
X_8	Monthly installment/monthly income
X_9	Loan type
X_{10}	Loan amount
X_{11}	Loan amount/house appraisal value
X_{12}	Special loan for government employees
X_{13}	Purpose of buying the property
X_{14}	Property type
X_{15}	Property age
X_{16}	Number of guarantors
X_{17}	Relationship between guarantor and guarantee
X_{18}	Credit status of guarantor

The neural network simulator Qnet 97 (1998), developed by Vesta Services, Inc., was utilized to develop the neural networks as well as the two-stage credit scoring models. The discriminant analysis credit scoring models will be implemented using the popular SPSS software (SPSS, 1998). And MARS 2.0 (2001) provided by Salford Systems are used in building the MARS credit scoring model. All the modeling tasks are implemented on an IBM PC with Intel Pentium III 800 MHz CPU processor. The detailed credit scoring results using the above-mentioned modeling techniques can be summarized as follows.

4.1. Neural networks model

Since Vellido et al. (1999) pointed out that close to 80% of business applications using neural networks will adopt the BPN training algorithm, this study will also use the popular BPN in building the credit scoring model. As recommended by Cybenko (1989); Hornik, Stinchcombe, and White (1989); and Zhang et al. (1998) that the single hidden layer network is sufficient to model any complex system, the designed network will have only one hidden layer. There are 18 input nodes in the input layer (please refer to Table 1 for more details) and only one output node, the credit status of the customer—good or bad credit. As the issue of determining the optimal number of hidden nodes is a crucial yet complicated one, the most commonly used way in determining the number of hidden nodes is via experiments or trial and error (Hecht-Nielsen, 1990; Lippmann, 1987; Tang & Fishwick, 1993; Wong, 1991). We, therefore, will also use the trial and error approach with the range from 15 to 50 neurons to determine the appropriate number of hidden nodes for the desired networks. The training of a network is implemented with various learning rates ranging from 0.0001 to 0.4 (almost all the network structure cannot converge with a learning rate greater than 0.4) and training lengths ranging from 10,000 to 300,000 iterations until the network converges. Network weights will be reset for each combination of the network parameters such as learning rates and momentum.

Table 2
Cross-validation results of the BPN credit scoring models

Fold number	Credit scoring results		
	{1–1}	{2–2}	Average correct classification rate
1	86.96% (80/92)	60.00% (6/10)	84.31% (86/102)
2	84.78% (78/92)	50.00% (5/10)	81.37% (83/102)
3	88.04% (81/92)	60.00% (6/10)	85.29% (87/102)
4	89.13% (82/92)	50.00% (5/10)	85.29% (87/102)
5	89.01% (81/91)	54.55% (6/11)	85.29% (87/102)
Mean	87.58% (402/459)	54.90% (28/51)	84.31% (430/510)

Here a class 1 customer is defined as a customer with good credit while a class 2 customer is the one with bad credit.

As the training of any neural network is itself a stochastic process, the reported neural network result is therefore the medium value (avoid possible extreme values due to better/poorly trained networks) of 20 repetitive trials. The network topology with the highest correct classification rate is considered as the optimal network topology.

Five neural networks credit scoring models were built and the classification results of the corresponding testing samples were summarized in Table 2. From the results in Table 2, we can observe that the average correct classification rates for the five folds are 84.31, 81.37, 85.29, 85.29, and 85.29%, respectively, with mean equals to 84.31%.

4.2. Multivariate adaptive regression splines model

In order to demonstrate the modeling results of MARS scoring models, the first built MARS model will be used as an illustrative example. The obtained basis functions and variable selection results of the illustrative example are summarized in Table 3. It is observed that monthly installment/monthly income, number of guarantors, loan type, loan amount/house appraisal value, and marital status do play important roles in deciding the MARS credit scoring models. Besides, according to the obtained basis functions and the MARS prediction function, it can be observed that the loan applicant with high monthly installment/monthly income, high loan amount/house appraisal value and the loan type 0 tends to become a bad credit customer while an applicant with more guarantors and marital status 0 likely to be a good credit customer. The above conclusions from the basis functions and MARS prediction function have important managerial implications since it can help managers/professionals design appropriate loan policies in acquiring the good credit customers. Besides, according to our knowledge, no other commonly used credit scoring modeling techniques possess this type of capability. The above-mentioned technical merits of MARS are one of the main concerns for the authors in designing this two-stage hybrid credit scoring model and these concerns are further verified in this illustrative example.

The testing results of the five built MARS scoring models can be summarized in Table 4. From the results in Table 4, we can conclude that the average correct classification rates for the five folds are 80.39, 78.43, 82.35, 81.37, and 82.35%, respectively, with the mean equals to 80.98%. Even though MARS exhibits the capability of identifying important independent variables, however, its classification capability is still not that good in comparison with BPN after comparing the results of Tables 2 and 4.

4.3. The two-stage hybrid credit scoring model

The single-layer BPN model will again be adopted in building the two-stage hybrid model. The input layer of the hybrid model contains the obtained significant independent variables of the MARS credit scoring model (please refer to Table 3 for more details) as the input nodes. The trial and error approach will again be used to determine the appropriate number of hidden nodes for the desired networks. The training of the network is also implemented with various learning rates and training lengths ranging from 10,000 to 300,000 iterations until the network convergences. The network weights are also reset for each combination of the network parameters such as learning rates and momentum. Again, in order to avoid possible extreme values due to better/poorly trained networks, the reported neural network result is the medium value of 20 repetitive trials. And the network topology with the highest

Table 3
Variable selection results and basis functions of MARS credit scoring model

Variable selection results		Basis function	
Variable name	Relative importance (%)	Equation name	Equation
Monthly installment/monthly income (X_8)	100.000	BF1	max(0, $X_8-5.000$)
Number of guarantors (X_{16})	89.740	BF2	max(0, X_{11}-88.000)
Loan type (X_9)	64.182	BF4	($X_{16}=1$ or $X_{16}=2$)
Loan amount/house appraisal value (X_{11})	62.260	BF6	($X_9=0$)
Marital status (X_3)	28.021	BF8	($X_{16}=0$)
		BF12	($X_3=0$)
MARS prediction function: $Y=0.807+0.005\times\mathrm{BF1}+0.068\times\mathrm{BF2}-0.918\times\mathrm{BF4}+0.119\times\mathrm{BF6}-0.681\times\mathrm{BF8}-0.089\times\mathrm{BF12}$			

In the MARS credit scoring model, $Y=0(1)$ is defined to be a good (bad) credit customer.

Table 4
Cross-validation results of MARS credit scoring models

Fold number	Credit scoring results		
	{1–1}	{2–2}	Average correct classification rate
1	82.61% (76/92)	60.00% (6/10)	80.39% (82/102)
2	81.52% (75/92)	50.00% (5/10)	78.43% (80/102)
3	85.87% (79/92)	50.00% (5/10)	82.35% (84/102)
4	84.78% (78/92)	50.00% (5/10)	81.37% (83/102)
5	85.71% (78/91)	54.55% (6/11)	82.35% (84/102)
Mean	84.10% (386/459)	52.94% (27/51)	80.98% (413/510)

Table 5
Cross-validation results of the proposed two-stage credit scoring models

Fold number	Credit scoring results		
	{1–1}	{2–2}	Average correct classification rate
1	86.96% (80/92)	60.00% (6/10)	84.31% (86/102)
2	84.78% (78/92)	60.00% (6/10)	82.35% (84/102)
3	88.04% (81/92)	60.00% (6/10)	85.29% (87/102)
4	88.04% (81/92)	60.00% (6/10)	85.29% (87/102)
5	89.01% (81/91)	63.64% (7/11)	86.27% (88/102)
Mean	87.36% (401/459)	60.78% (31/51)	84.71% (432/510)

correct classification rate is considered as the optimal network topology.

The prediction results of the built hybrid models are summarized in Table 5. From the results in Table 5, we can conclude that the average correct classification rates for the five folds are 84.31, 82.35, 85.29, 85.29, and 86.27%, respectively, with the mean equals to 84.71%. It can also be observed from Tables 2, 4 and 5 that the scoring accuracy of the two-stage model, even though is better than MARS, is only slightly higher than that of the model solely using BPN.

5. Comparison of results of different credit scoring models

In order to evaluate the effectiveness of the proposed two-stage hybrid credit scoring model, the classification results are also compared with those using linear discriminant analysis and logistic regression models. Table 6 summarizes the average classifications results of linear discriminant analysis, logistic regression, MARS, BPN, and the hybrid two-stage credit scoring models. It can be concluded, from Table 6, that the two-stage hybrid model has the best credit scoring capability in terms of the average correct classification rate.

Even though the average correct classification rate is an important criterion in evaluating the classification capability of a credit scoring model, it is quite well known that the prior probabilities (or simply priors) and the misclassification costs also have to be taken into account in order to

Table 6
Summarized credit scoring results of the five constructed models

Credit scoring model	Credit scoring results		
	{1–1}	{2–2}	Average correct classification rate
Discriminant analysis	77.34% (355/459)	58.82% (30/51)	75.49% (385/510)
Logistic regression	78.21% (359/459)	56.86% (29/51)	76.08% (388/510)
MARS	84.10% (386/459)	52.94% (27/51)	80.98% (413/510)
BPN	87.58% (402/459)	54.90% (28/51)	84.31% (430/510)
Two-stage hybrid model	87.36% (401/459)	60.78% (31/51)	84.71% (432/510)

obtain a scoring model with the minimum expected misclassification cost (Johnson & Wichern, 2002; West, 2000). Eq. (3) expresses the function in computing the expected misclassification cost when only two different populations are considered

$$\text{Cost} = C(2|1) \times P(2|1) \times \pi_1 + C(1|2) \times P(1|2) \times \pi_2 \quad (3)$$

where π_1 and π_2 are prior probabilities of good and bad credit populations, $P(2|1)$ and $P(1|2)$ measures the probability of making Type I errors (a customer with good credit is misclassified as a customer with bad credit) and Type II errors (a customer with bad credit is misclassified as a customer with good credit), and $C(2|1)$ as well as $C(1|2)$ are the corresponding misclassification costs of Type I and Type II errors.

In order to compute the expected misclassification costs of different scoring models, the estimates of misclassification probability and misclassification costs have to be done first. The most commonly adopted estimates for $P(2|1)$ and $P(1|2)$ are the fraction of good credit customers misclassified as bad credit customers and the fraction of bad credit customers misclassified as good credit customers. As to the estimates of misclassification costs, it is a challenging and complicated task as valid estimates may not be available. However, in credit scoring applications, it is generally believed that the costs associated with Type I error and Type II error are significantly different. In general, the misclassification costs associated with Type II errors are much higher than those associated with Type I errors. As recommended by Dr Hofmann who compiled the German credit data reported that the relative ratio of misclassification ratio associated with Type I and Type II errors is 1–5 (West, 2000), we will also use this relative cost ratio in computing the expected misclassification costs of the five built credit scoring models.

Table 7 summarizes the Type I and Type II errors of the five built models and the corresponding expected misclassification costs. From Table 7, we can conclude that the two-stage hybrid model has the best credit scoring capability in terms of the expected misclassification cost criterion in comparison with those of linear discriminant analysis, logistic regression, MARS and BPN neural

Table 7
Errors and the expected misclassification costs of the five built models

Model	Type I error (%)	Type II error (%)	Expected misclassification costs
Discriminant analysis	22.26	41.18	0.40624
Logistic regression	21.79	43.14	0.41181
MARS	15.90	47.06	0.37840
BPN	12.42	45.10	0.33728
Two-stage hybrid model	12.64	39.22	0.30986

The priors of good and bad credit populations are set as 0.9 and 0.1 using the ratio of good and bad credit customers in the empirical dataset.

network models (similar results should be obtained with larger Type II errors). Consequently, we can conclude that the credit scoring results of the proposed two-stage hybrid model outperforms the commonly utilized discriminant analysis, logistic regression, and neural networks models and hence provides efficient alternatives in conducting credit scoring tasks. Besides, even the hybrid model is only slightly better than the model solely using BPN, it should indeed be a better alternative as it can identify important independent variables which may provide valuable information for further managerial and related decision makings.

6. Conclusions and areas of future research

Credit scoring has become more and more important as the competition between financial institutions has come to a totally conflicting stage. More and more companies are seeking better strategies through the help of credit scoring models. And hence various modeling techniques have been developed in different credit evaluation processes for better credit approval schemes. Therefore credit scoring is one of the main application areas of classification problems that have drawn serious attention during the past decade. Many modeling alternatives, like traditional statistical methods, non-parametric methods and artificial intelligence techniques, have been developed in order to successfully handle the credit scoring tasks. Discriminant analysis and logistic regression are the most commonly utilized statistical credit scoring techniques, but often being criticized due to the fact of their strong model assumptions and poor credit scoring capabilities. On the other hand, the artificial neural networks is becoming a very popular alternative in handling credit scoring tasks due to its associated memory characteristic, generalization capability and outstanding credit scoring capability. Even with the above-mentioned advantages, it is also being criticized for its long training process in designing the optimal network's topology, hard to identify the relative importance of potential input variables and certain interpretive difficulties.

In order to improve the decision of the network structure and support the interpretive difficulties of the obtained results of neural networks credit scoring models, the purpose of this study is to explore the performance of credit scoring using a two-stage hybrid modeling procedure using artificial neural networks and multivariate adaptive regression splines. The rationale underlying the analyses is using MARS as a supporting tool to neural networks scoring model with a MARS credit scoring model first being built, then the obtained significant variables are used as the input nodes of the designed neural networks model. To demonstrate the feasibility and effectiveness of the proposed two-stage modeling procedure, credit scoring tasks are performed on one housing loan dataset from a local bank using the cross-validation approach. Analytic results demonstrate that the hybrid credit scoring model has the best credit scoring capability, in terms of the minimum expected misclassification cost criterion, in comparison with those of linear discriminant analysis, logistic regression, MARS, and BPN neural networks. Besides, even the hybrid model is only slightly better than the BPN model, it should indeed be a better alternative as it demonstrates the capability in identifying important independent variables and the corresponding prediction function, which is the major drawback of neural networks that limited its applicability in handling credit scoring tasks, may provide valuable information for further managerial and related decision makings on the top of its excellent credit scoring capability. The research findings provide efficient alternatives in conducting future credit scoring tasks.

Future studies should aim at collecting more important independent variables that may further increase the credit scoring accuracy. Using other classification techniques, like the classification and regression tree (CART), bagging and boosting, fuzzy discriminant analysis, and support vector machines (SVM), in evaluating their applicability to credit scoring models is also recommended. Integrating other artificial intelligence techniques, like genetic algorithms and/or grey theory, with neural networks in further refining the network structure and improving credit scoring accuracy are other possible directions for future studies.

Acknowledgements

This research was partially supported by the National Science Council of the Republic of China under Grant Number NSC 89-2213-E-030-018. The author is also grateful to those graduate students whose numerous efforts and time in collecting and preparing the dataset for this research. Special thanks also go to the anonymous bank in kindly providing the housing loan data that makes this research possible.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, *23*, 589–609.

Anderson, J. A., & Rosenfeld, E. (1988). *Neurocomputing: Foundations of research*. Cambridge, MA: MIT Press.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York, NY: Wiley.

Arminger, G., Enache, D., & Bonne, T. (1997). Analyzing credit risk data: A comparison of logistic discrimination classification tree analysis and feedforward networks. *Computational Statistics*, *12*, 293–310.

Bardos, M. (1998). Detecting the risk of company failure at the banque de France. *Journal of Banking and Finance*, *22*, 1405–1419.

Barney, D. K., Graves, O. F., & Johnson, J. D. (1999). The farmers home administration and farm debt failure prediction. *Journal of Accounting and Public Policy*, *18*, 99–139.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Pacific Grove, CA: Wadsworth.

Cabena, P., Hadjinaian, P. O., Stadler, R., Verhees, J., & Zanasi, A. (1997). *Discovering data mining from concept to implementation*. Upper Saddle River, NJ: Prentice-Hall.

Cheng, B., & Titterington, D. M. (1994). Neural network: A review from a statistical perspective (with discussion). *Statistical Science*, *9*, 2–54.

Chen, M.S., Han, J., and Yu, P.S. (1996). "Data Mining: An Overview from a Database Perspective," IEEE Transactions on Knowledge and Data Engineering, 8(6), 866–883.

Chiu, C. C., Shao, Y. J., Lee, T. S., & Lee, K. M. (2003). Identification of process disturbance using SPC/EPC and neural networks. *Journal of Intelligent Manufacturing*, *14*(3), 379–388.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data*. London: Chapman & Hall.

Craven, M. W., & Shavlik, J. W. (1997). Using neural networks for data mining. *Future Generation Computer Systems*, *13*, 221–229.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematical Control Signal Systems*, *2*, 303–314.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, *10*, 167–179.

De Gooijer, J. G., Ray, B. K., & Krager, H. (1998). Forecasting exchange rates using TSMARS. *Journal of International Money and Finance*, *17*(3), 513–534.

Deng, P.-S. (1993). Automatic knowledge acquisition and refinement for decision support: A connectionist inductive inference model. *Decision Sciences*, *24*(2), 371–393.

Desai, V. S., Crook, J. N., & Overstreet, G. A. Jr. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, *95*, 24–37.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis methods and applications*. New York, NY: Wiley.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*, 27–34.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179–188.

Flagg, J. C., Giroux, G. A., & Wiggins, C. E. (1991). Predicting corporate bankruptcy using failing firms. *Review of Financial Economics*, *1*, 67–78.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, *19*, 1–141.

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, *4*, 197–217.

Glorfeld, L. W. (1996). A methodology for simplification and interpretation of backpropagation-based neural network models. *Expert Systems with Applications*, *10*(1), 37–54.

Glorfeld, L. W., & Hardgrave, B. C. (1996). An improved method for developing neural networks—the case of evaluating commercial loan creditworthiness. *Computers and Operations Research*, *23*(10), 933–944.

Griffin, W. L., Fisher, N. I., Friedman, J. H., & Ryan, C. G. (1997). Statistical techniques for the classification of chromites in diamond exploration samples. *Journal of Geochemical Exploration*, *59*, 233–249.

Harrell, F. E., & Lee, K. L. (1985). A comparison of the discrimination of discriminant analysis and logistic regression. In P. K. Se (Ed.), *Biostatistics: Statistics in biomedical, public health, and environmental sciences*. Amsterdam: North-Holland.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman & Hall.

Haykin, S. S. (1994). *Neural networks: A comprehensive foundation*. New York, NY: Macmillan.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Menlo Park, CA: Addison-Wesley.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximations. *Neural Networks*, *2*, 336–359.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley.

Jagielska, I., & Jaworski, J. (1996). Neural network for predicting the performance of credit card accounts. *Computational Economics*, *9*(1), 77–82.

Jensen, H. L. (1992). Using neural networks for credit scoring. *Managerial Finance*, *18*, 15–26.

Joanes, D. N. (1993). Rejecting inference applied to logistic regression for credit scoring. *IMA Journal of Mathematics Applied in Business and Industry*, *5*, 35–43.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

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

Kim, J. C., Kim, D. H., Kim, J. J., Ye, J. S., & Lee, H. S. (2000). Segmenting the Korean housing market using multiple discriminant analysis. *Construction Management and Economics*, *18*, 45–54.

Kuhnert, P. M., Do, K.-A., & McClure, R. (2000). Combining non-parametric models with logistic regression: An application to motor vehicle injury data. *Computational Statistics and Data Analysis*, *34*, 371–386.

Laitinen, E. K. (1999). Predicting a corporate credit analyst's risk by logistic and linear models. *International Review of Financial Analysis*, *8*(2), 97–121.

Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction: Application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis*, *9*(4), 327–349.

Lau, A. H. L. (1987). A five-state financial distress prediction model. *Journal of Accounting Research*, *25*(1), 127–138.

Lee, G., Sung, T. K., & Chang, N. (1999). Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, *16*, 63–85.

Lee, H., Jo, H., & Han, I. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems with Applications*, *13*, 97–108.

Lee, T. S., & Chen, N. J. (2002). Investigating the information content of non-cash-trading index futures using neural networks. *Expert Systems with Applications*, *22*(3), 225–234.

Lee, T. S., & Chiu, C. C. (2002). Neural network forecasting of an opening cash price index. *International Journal of Systems Science*, *33*(3), 229–237.

Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, *23*(3), 245–254.

Lewis, P. A. W., & Stevens, J. G. (1991). Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *Journal of American Statistical Association*, *86*, 864–877.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, *April*, 4–22.

MARS 2.0—for windows 95/98/NT, Salford Systems, San Diego, CA, 2001.

Martell, T. F., & Fitts, R. L. (1981). A quadratic discriminant analysis of bank credit card user characteristics. *Journal of Economics and Business*, *33*, 153–159.

Nguyen-Cong, V., Van, D. G., & Rode, B. M. (1996). Using multivariate adaptive regression splines to QSAR studies of dihydroartemisinin derivatives. *European Journal of Medical Chemistry*, *31*, 797–803.

Ohmann, C., Moustakis, V., Yang, Q., & Lang, K. (1996). Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain. *Artificial Intelligence in Medicine*, *13*, 23–36.

Overstreet, G. A. Jr., Bradley, E. L. Jr., & Kemp, R. S. (1992). The flat-maximum effect and generic linear scoring model: A test. *IMA Journal of Mathematics Applied in Business and Industry*, *4*, 97–109.

Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research*, *112*, 310–321.

Piramuthu, S., Shaw, M. J., & Gentry, J. A. (1994). A classification approach using multi-layered neural networks. *Decision Support Systems*, *11*(5), 509–525.

Qnet 97—Neural network modeling for windows 98/NT, Vesta Services, Winnetka, IL, 1998.

Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business and Economic Statistics*, *1*, 101–114.

Repley, B. (1994). Neural networks and related methods for classification (with discussion). *Journal of the Royal Statistical Society, Series B*, *56*, 409–456.

Richeson, L., Zimmermann, R. A., & Barnett, K. G. (1994). Predicting consumer credit performance: Can neural networks outperform traditional statistical methods? *International Journal of Applied Expert Systems*, *2*(2), 116–130.

Sharma, S. (1996). *Applied multivariate techniques*. New York, NY: Wiley.

SPSS 1997—Statistic modeling for windows 95/98/NT, SPSS Inc., 1998.

Stern, H. S. (1996). Neural networks in applied statistics. *Technometrics*, *38*(3), 205–216.

Suh, E. H., Noh, K. C., & Suh, C. K. (1999). Customer list segmentation using the combined response model. *Expert Systems with Applications*, *17*(2), 89–97.

Tang, Z., & Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing*, *5*(4), 374–385.

Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risks of lending to customers. *International Journal of Forecasting*, *16*, 149–172.

Titterington, D. M. (1992). Discriminant analysis and related topics. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), *Credit scoring and credit control* (pp. 53–73). Oxford: Oxford University Press.

Torsun, I. S. (1996). A neural network for a loan application scoring system. *The New Review of Applied Expert Systems*, *2*, 47–62.

Trevino, L. J., & Daniels, J. D. (1995). FDI theory and foreign direct investment in the United States: A comparison of investors and non-investors. *International Business Review*, *4*, 177–194.

Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications*, *17*, 51–70.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research*, *27*, 1131–1152.

Westgaard, S., & van der Wijst, N. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research*, *135*(2), 338–349.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, *15*, 757–770.

Wong, F. S. (1991). Time series forecasting using backpropagation neural networks. *Neurocomputing*, *2*, 147–159.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35–62.