Knowledge-Based Systems 36 (2012) 245–252

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring

Ling-Jing Kao [a], Chih-Chou Chiu [a,*], Fon-Yu Chiu [b]

[a] Department of Business Management, National Taipei University of Technology, Taiwan
[b] Institute of Commerce Automation and Management, National Taipei University of Technology, Taiwan

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form 5 July 2012
Accepted 10 July 2012
Available online 21 July 2012

Keywords:
Behavior scoring
Credit scoring
Bayesian
Latent variable model
Classification and regression tree

ABSTRACT

A Bayesian latent variable model with classification and regression tree approach is built to overcome three challenges encountered by a bank in credit-granting process. These three challenges include (1) the bank wants to predict the future performance of an applicant accurately; (2) given current information about cardholders' credit usage and repayment behavior, financial institutions would like to determine the optimal credit limit and APR for an applicant; and (3) the bank would like to improve its efficiency by automating the process of credit-granting decisions. Data from a leading bank in Taiwan is used to illustrate the combined approach. The data set consists of each credit card holder's credit usage and repayment data, demographic information, and credit report. Empirical study shows that the demographic variables used in most credit scoring models have little explanatory ability with regard to a cardholder's credit usage and repayment behavior. A cardholder's credit history provides the most important information in credit scoring. The continuous latent customer quality from the Bayesian latent variable model allows considerable latitude for producing finer rules for credit granting decisions. Compared to the performance of discriminant analysis, logistic regression, neural network, multivariate adaptive regression splines (MARS) and support vector machine (SVM), the proposed model has a 92.9% accuracy rate in predicting customer types, is less impacted by prior probabilities, and has a significantly low Type I errors in comparison with the other five approaches.

1. Introduction

Evaluating a customer's credit risk is crucial for financial institutions due to the high risks associated with inappropriate credit-granting decisions. It becomes an even more important task today after the recent financial crisis involving subprime loans. Three challenges are encountered by the bank while making credit-granting decisions. First, the bank would like to predict the future performance of an applicant accurately. Second, given information about current cardholders' credit usage and repayment behavior, financial institutions would like to determine the optimal level of product attributes, such as credit limit and annual percentage rate, for an applicant. Third, the bank would like to improve its efficiency by automating credit-granting decisions.

Financial institutions try to solve these three challenges by making heavy use of predictive scoring models. Depending on a financial institution's target activities and available customer information, the scoring models can be classified into credit scoring models or behavioral scoring models. A credit scoring model aims to make decisions about whether to grant new customers a particular financial product (e.g., credit), whereas a behavioral scoring model is designed to evaluate the credit of existing customers, i.e., how reliably customers keep up to date with payments.

Scoring models usually produce a binary credit rating which classifies customers into two groups (such as 'good/bad'; 1/0) according to classification rules and a set of explanatory variables. However, the choice of classification rules and explanatory variables is ad hoc and usually relies on managers' know-how and experience. Many statistical methods (e.g. logistic regression model; discriminant analysis) and artificial intelligence approaches (e.g. neural networks and rule-based approaches) have been used to construct scoring or fraud detection models [7,23,18,25,17,26]. These approaches have good credit or behavior scoring capabilities, but they are criticized for their black-box extraction of data feature vectors [19].

The objective of this paper is to provide a combined model that can overcome all three of these challenges involved in credit-granting decisions in the financial industry and that can overcome the drawbacks of traditional scoring approaches. It is called a "combined" model because the model we proposed is not only to

* Corresponding author.
E-mail addresses: lingjingkao@ntut.edu.tw (L.-J. Kao), chih3c@ntut.edu.tw (C.-C. Chiu), bdyucu@yahoo.com.tw (F.-Y. Chiu).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.07.004

apply methods in sequence but also to establish the linkage between the credit scoring model and the behavior scoring model. It allows us to study the decision of credit granting given both credit terms (APR and etc.) and credit usage behavior.

The idea of integrating multiple approaches is not new in literature. For example, Chen [4] integrated feature selection and CPDA-based rough set approach to classify Asian banks' credit rating. The combined approach proposed in this paper consists of two steps. First, given customer types (e.g., good or bad) determined by credit analysts in a bank, a hierarchical Bayes model is developed to estimate heterogeneous customer quality scores. A cardholder's quality score is a continuous latent variable which summarizes this cardholder's repayment decisions and credit usage behaviors. Then, given posterior estimates of quality value and customers' credit reports as well as demographic information, we use the Classification and Regression Tree (CART) algorithm to deduce decision rules that can be used to determine whether to grant an applicant credit, and to determine the optimal levels of product attributes, such as credit limits and annual percentage rate.

Data provided by a bank in Taiwan is used to illustrate the combined approach proposed in this paper. This data set consists of each credit card holder's credit usage and repayment data, demographic information, and credit report. The empirical study shows that demographic variables used in most credit scoring models have little explanatory ability with regard to a cardholder's credit usage and repayment behavior. A cardholder's credit history provides the most important information in credit scoring. We demonstrate that the proposed model can result in a 92.9% accurate prediction and the lowest Type I error.

The remainder of the paper is organized as follows: In Section 2, the proposed approach is developed by combining a Bayesian behavior scoring model and a CART-based credit scoring model. An empirical study of a credit card and its results are presented in Section 3. Conclusions are offered in Section 4.

2. A combined approach

2.1. Bayesian behavior scoring model

Let y_{ij} denote the credit rating of the ith cardholder assigned by credit analysts in a bank. y_{ij} equals one if the ith credit card holder was evaluated as creditworthy during jth time period. y_{ij} equals zero if the ith credit card holder was evaluated as not creditworthy during jth time period. Let z_{ij} be the quality score of the ith cardholder during the time period j. A cardholder's quality (z_{ij}) is a continuous latent variable that summarizes this cardholder's credit usage and repayment behavior and is assumed to associate with a stochastic component ε_i and the vector of independent variables X_{ij}. Let $X_{ij} = (x_{1ij}, x_{2ij}, \ldots, x_{kij})'$ be a vector of variables used to measure the "quality" of the ith cardholder during the jth time period. The behavior scoring model can be written as follows:

$$y_{ij} = \begin{cases} 1 & z_{ij} \geqslant 0 \\ 0 & otherwise \end{cases} \qquad (1)$$
$$z_{ij} = X_{ij}\beta_i + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma_i^2), \forall i = 1, 2, \ldots, H; j = 1, 2, \ldots, T$$

Eq. (1) is a latent variable model with binary responses. We use Eq. (1) to describe a cardholder's quality (z_{ij}) and its relationship to the corresponding discrete binary outcome variable (y_{ij}). Threshold value is set at zero and σ^2 is set to one to overcome the location and scale identification problems respectively. The latent variable model has been an important tool in the social sciences literature [1]. The binary observations are often assumed to be driven by unobservable behavior mechanisms that are associated with a set of explanatory variables. For example, Bartholomew [2] uses the latent variable to obtain insights into the structure of the responses in the Workplace Industrial Relations Survey. Hand and Crowder [10] proposed a latent variable model which measures the underlying quality of a customer from a retail bank's perspective. Economists also use latent variable models to study labor force participation, the choice of occupation, consumer choice among alternatives, etc. [9].

Both the literature and industry practice suggest that inter-payment time and repayment ability are good indicators to determine a customer's value [14,15]. Therefore, we assume that the vector X_{ij} in Eq. (1) consists of two independent variables (x_{1ij}, x_{2ij}), in which x_{1ij} represents repayment ability and x_{2ij} represents the inter-payment time of the ith customer at time period j. A customer has a higher chance of being in default if a long inter-payment time is observed or a low percentage of credit debts is paid. A customer is defined as being in default if the minimum payment is not met after passing the due date for a certain period of time (e.g., 120 days, depending on bank rules).

A customer's repayment ability can also be evaluated by computing the ratio between the payment and the entire outstanding balance. This ratio also represents the customer's liquidity. A consumer who can pay most of his balance is considered to have good liquidity. Depending on banks' particular rules, the payment made by a customer can be allocated into either principal or interest. Each bank also has its own method (i.e., average daily balance method) to compute the incurred interest.

Let w_{ij} be the ith customer's payments to the principal, while p_{it} is the ith customer's total principal, Q_{ij} is the ith customer's payments to the interest incurred from the loan, and s_{ij} is the ith customer's total interest incurred from the loan during the time period j. A metric for evaluating the ith customer's repayment ability is developed as follows:

$$x_{1ij} = \rho\left(\frac{w_{ij}}{p_{ij}}\right) + (1-\rho)\left(\frac{Q_{ij}}{s_{ij}}\right) \qquad (2)$$

In Eq. (2), the first term (w_{ij}/p_{ij}) represents a customer's principal repayment ability, and the second term (Q_{ij}/s_{ij}) is used to measure a customer's interest repayment ability. ρ represents the weight given to the importance of repayment ability. For example, if the ith customer is a convenience user who uses a credit card as a convenient payment device and pays off the outstanding balance every period, the transaction fees collected from merchants are the only contribution he brings to the bank. Then, by Eq. (2), his payment ability at this period will be 0.5. However, those customers who use credit and have incurred interest over time and repay most of their credit debts, can contribute the most to the bank and are considered to have good repayment ability. In our research, we assume that principal repayment ability and interest repayment ability are equally important.

The Gibbs sampler and data augmentation [8,11,20,22] were used to estimate the customer quality model. Let $Z_i = (z_{i1}, z_{i2}, \ldots, z_{iT})'$ and $X_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$. Given conjugate priors

$$\beta_i \sim N(\bar{\beta}, V_\beta),$$
$$\bar{\beta} \sim N(\mu_0, A_0),$$
$$V_\beta \sim IW(f_0, G_0),$$

posterior estimates of parameters can be obtained by generating draws from the following full conditional distributions.

$$[z_{ij}|\cdot] \propto \prod [y_{ij}|z_{ij}][z_{ij}|X_{ij}\beta_i] \sim Truncated\,Normal(X_{ij}\beta_i, 1)$$

$$[\beta_i|\cdot] \propto [Z_i|X_i\beta_i][\beta_i|\bar{\beta}, V_\beta]$$
$$\sim N((X_i'X_i + V_\beta^{-1})^{-1}(X_i'Z_i + V_\beta^{-1}\bar{\beta}), (X_i'X_i + V_\beta^{-1})^{-1})$$

$$[\bar{\beta}|\cdot] \propto \prod_{i=1}^{H}[\beta_i|\bar{\beta}, V_\beta][\bar{\beta}|\mu_0, A_0]$$
$$\sim N\left(\left(HV_\beta^{-1}+A_0^{-1}\right)^{-1}\left(V_\beta^{-1}\left(\sum_{i=1}^{H}\beta_i\right)+A_0^{-1}\mu_0\right), \left(HV_\beta^{-1}+A_0^{-1}\right)^{-1}\right)$$

$$[V_\beta|\cdot] \propto \prod_{i=1}^{H}[\beta_i|\bar{\beta}, V_\beta][V_\beta|f_0, G_0]$$
$$\sim IW\left(H+f_0, \left[\sum_{i=1}^{H}(\beta_i-\bar{\beta})(\beta_i-\bar{\beta})'+G_0\right]\right)$$

2.2. CART-based credit scoring model

Various types of prediction tools have already been successfully developed, including conventional statistical methods, non-parametric methods, and artificial intelligence methods. The CART algorithm developed by Breiman et al. [3] has been widely discussed and successfully applied and is the method we selected for use in this study. Basically, Classification and Regression Tree (CART), a non-parametric statistical procedure, is used primarily as a classification tool, where the objective is to classify an object into two or more populations. As the name suggests, CART is a single procedure that can be used to analyze either categorical or continuous data using the same technology.

The methodology outlined in Breiman et al. [3] can be summarized into three stages. The first stage involves growing the tree using a recursive partitioning technique to select variables and split points by means of a splitting criterion. Several criteria are available for determining the splits, including gini, twoing and ordered twoing. A more detailed description of these criteria can be found in Breiman et al. [3]. In addition to selecting the primary variables, we can also identify and select surrogate variables that are closely related to the original splits and that may be useful in classifying observations that have missing values for the primary variables.

After a large tree is identified, the second stage of the CART methodology uses a pruning procedure that incorporates a minimal cost complexity measure. The result of the pruning procedure is a nested subset of trees starting from the largest tree grown and continuing the process until only one node of the tree remains. Cross-validation or a testing sample will be used to provide estimates of future classification errors for each subtree. Cross-validation is used when only small numbers of data points are available for building the CART models.

The last stage of the methodology is to select the optimal tree, which corresponds to a tree yielding the lowest cross-validated or testing set error rate. Although the optimal tree has the smallest testing error, it could be identified as unstable because of its large size. To avoid this instability, trees with smaller sizes but comparable in accuracy (i.e. within one standard error) will be chosen as an alternative. This process is referred to as the one standard error rule and can be tuned to obtain trees of varying sizes and complexity. A measure of variable importance can be achieved by observing the drop in the error rate when another variable is used instead of the primary split. Basically, the more frequently a variable appears as a primary or surrogate split, the higher the importance score assigned. Please refer to Breiman et al. [3] and Steinberg and Colla [21] for more details regarding the model building process of CART.

Unlike the behavior scoring model in which an analyst is able to observe a cardholder's repayment and credit-usage behavior, a credit scoring model is built to predict a new applicant's behavior given all possible information a bank can collect from an applicant. Given that information, a bank looks for an efficient solution to predict the customer type and to determine credit limits and annual percentage rates.

Assume that a new customer i applies for a credit card. His demographic information (denoted by D_i) and credit history (denoted by CH_i) are collected to determine whether credit should be granted, what the credit limit (denoted by CL_i) should be, and how much APR (denoted by APR_i) should be assigned. To combine the behavior and credit scoring model, we assume that the estimated customer quality obtained from Eq. (1) is a function of applicant information and a bank's decision variables. That is, after obtaining the posterior draws of the latent customer quality ($\hat{z}_{ij}$) in Eq. (1), we can post-process these posterior draws by building up a CART-based credit scoring model as follows:

$$\bar{z}_i = \sum_{j=1}^{J_i}\hat{z}_{ij} = f(D_i, CH_i, CL_i, APR_i) \tag{3}$$

Eq. (3) allows us to derive decision rules that a bank can employ to automate this credit-granting process. That is, given an applicant's demographic information and credit history, we can grid search the optimal credit limit and APR for the maximum average customer quality value.

The Bayesian behavior scoring model and the CART-based credit scoring model are combined by the average latent customer quality. The average latent customer quality represents the expected long-term quality performance of a customer. Different from a binary credit rating, the continuous latent customer quality provides the magnitude of customer quality and allows considerable latitude for producing finer rules for credit granting decisions.

The other advantage of the CART-based credit scoring model is that the decision rules generated by the CART-based credit scoring model are transparent. Researchers can compare and study the behavior and portfolio of customers in each group. This advantage is not available if other credit scoring methods such as discriminant analysis, logistic regression, neural network, multivariate adaptive regression splines (MARS) or support vector machine (SVM) are applied. In those cases, threshold values of each decision rule will be very difficult to estimate by other structural parametric models.

3. Empirical application

3.1. Data

A dataset provided by a leading bank in Taipei, Taiwan, is used to illustrate the proposed combined model. This dataset contains the information of credit card holders whose applications were approved between 2008 and 2009 and have transaction records spanned over a complete year from January 1, 2010 to December 31, 2010. The transaction data contains account-level information, such as the transaction date, the amount of transaction, repayment date and the amount of repayment.

To ensure the consistency of data field and the requirement of HB model building, data preprocessing was needed. We first use a customer identification to match two raw datasets: one contains effective credit card account information of 25,328 credit card holders, and the other stores over 5 million individual transaction records for these accounts. To satisfy the requirement of HB model estimation, we only keep those credit card holders who have at least 8 repayment records within 1-year period. In this case, 8915 credit card holders were obtained. Then, we randomly select 2248 credit card holders, which is 25% of the valid sample, to build the behavioral scoring model. In this sample, the percentage of good customers is 26%, and the percentage of bad customers is 74%.

Table 1
Variable lists and summary statistics.

Gender 1 = female (1814 card holders) 2 = male (1737 card holders)	Annual Percentage Rate (APR) Mean = 17% Std deviation = 0.041
Marriage status 1 = married (1686 card holders) 2 = single (1865 card holders)	Credit limits Mean = 157651.09 Std deviation = 105609.89
Annual income Mean = 490108.42 Std deviation = 298625.38	
Education level 1. Ph.D. 2. Post-graduate 3. College 4. Vocational school 5. High school 6. Other	
Occupation code 101–106 White-collar 201–205 Professional 301–311 Military personnel, public school teacher, and government officers 401–409 Blue-collar 501–599 Service and others	
JCIC 1: the number of the short-term or mid-term non-secured loan 2: the longest period of a credit card ownership (month) 3: the number of banks that a customer ever use his cash advanced services in the past 12 months 4: the number of new query of credit reports in the past 1 month	

In this sample of 2248 credit card holders, who have at least eight repayment records within a 1-year period, each cardholder has his/her credit status (1 = good or 0 = bad credit) assigned by the bank analysts, six demographic variables, four credit variables from the Joint Credit Information Center (JCIC)[1] in Taiwan, and credit limits and APR granted by the bank. Among the credit card holders, 1248 data sets with respect to the ratio of good and bad credit were randomly selected as the training sample (estimating the parameters of the corresponding built scoring model), another 500 will be used to test the model (selecting the final scoring model), and the remaining 500 will be retained for validation (evaluating the classification capability of the built scoring model). Variables in this dataset are summarized in Table 1.

3.2. Estimation result of the Bayesian behavior scoring model

Customer creditworthiness is estimated by the Bayesian behavior scoring model illustrated in Section 2.1. One thousand Markov chain Monte Carlo (MCMC) iterations are required for model convergence. Draws from the last 500 iterations were used to evaluate the means and standard deviations of model parameters. Convergence was assessed by inspecting time series plots of model parameters and multiple initial values.

Fig. 1 displays the histogram of the average quality of each customer ($\bar{z}_i$). It shows that the proposed Bayesian model can classify customers into two customer groups. The group with $\bar{z}_i$ classified into the left first two bars are customers with bad credit evaluations (Group 1), and the remaining customers are classified into the group with good credit evaluations (Group 2).

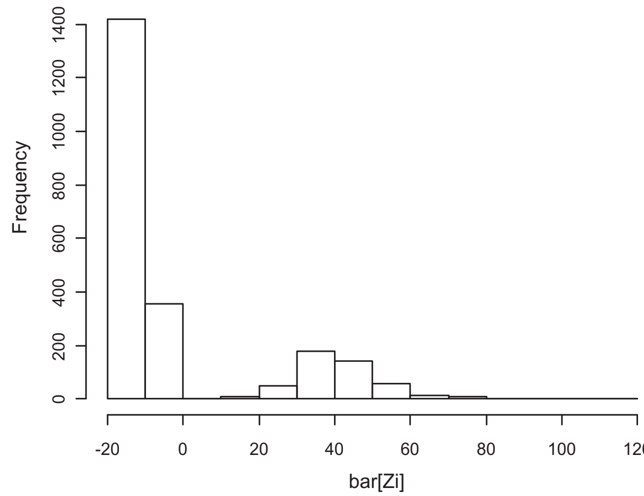

Fig. 1. Posterior mean of each cardholder's latent quality.

Fig. 2 compares the summary statistics of Group 1 and Group 2. By comparing the histograms of Group 1 and Group 2, we found that demographic variables except educational level showed almost no difference among the two groups. To do the comparison exactly, the Two-sample paired (Wilcoxon) Signed-Rank test for the six demographic variables in group 1 and group 2 is presented. The test results show that we cannot reject the null hypothesis (H_0: no difference in distributions) because the sums of signed ranks (W) for the five demographic variables (gender, age, marriage status, job position and annual income) all near zero and their *P* values are all greater than 0.15.

Both findings indicate that cardholders classified in Group1 and Group2 shared almost exactly the same characteristics in gender, age, marriage status, job position and annual income. This finding suggests that demographic variables used in most credit scoring models have little explanatory ability with respect to a cardholder's credit usage and repayment behavior.

Posterior estimates of consumer heterogeneity ($\bar{\beta}$ and V_β) are reported in Tables 2 and 3. Both of the covariates, repayment ability and inter-payment time, have posterior masses that are far from zero. This suggests that customer quality and repayment ability are positively associated, and that a short inter-payment time increases a customer's quality.

3.3. Variables' importance and deduction rules identified by CART

CART regression predictions are obtained by recursively splitting the sample and creating groups or clusters that are progressively more homogeneous than their ancestor nodes. The variables' importance and deduction rules (partial), generated using CART for the illustrated dataset, are summarized in Tables 4 and 5, respectively. The results in Table 4 suggest that the credit report from JCIC, followed by credit limits and APR, is the most critical in predicting customer quality. Demographic variables commonly applied in a conventional credit scoring model provide the least information for evaluating an applicant's future credit rating.

There are a total of 21 decision rules generated from the CART-based credit scoring model. Table 5 reports 4 decision rules among these 21. These decision rules cannot be obtained from other scoring models constructed by discriminant analysis, logistic regression, neural network, MARS and SVM. Threshold values of each decision variable will be very difficult to estimate using other

[1] JCIC is like a credit bureau in the United States.

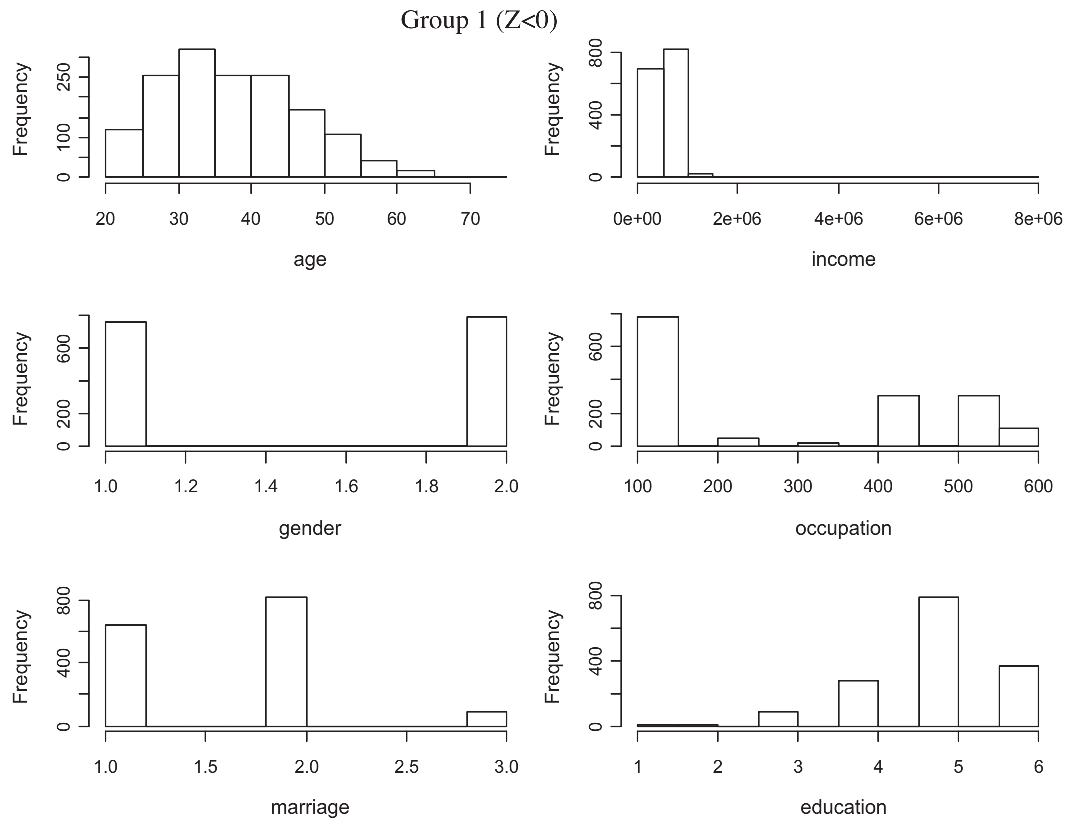

Fig. 2a. The summary statistics of Group 1.

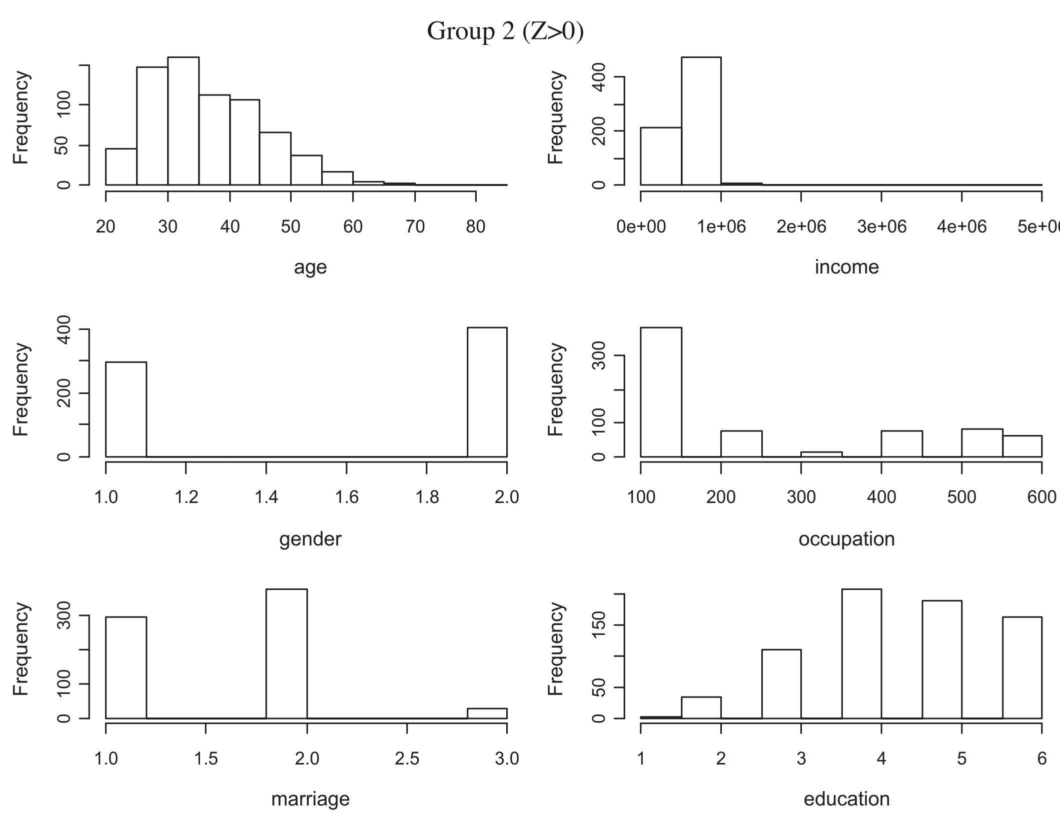

Fig. 2b. The summary statistics of Group 2.

Table 2
Posterior estimates of $\bar{\beta}$.

	Posterior mean	Posterior standard deviation
$\bar{\beta}_0$ (intercept)	0.35	0.028
$\bar{\beta}_1$ (repayment ability)	2.90	0.064
$\bar{\beta}_2$ (inter-payment time)	−11.61	0.200

Table 3
Posterior estimates of V_β.

	Intercept	Repayment ability	Inter-payment time
Intercept	1.007 (0.039)	0.077 (0.160)	−0.099 (0.144)
Repayment ability	–	2.496 (0.337)	0.182 (2.295)
Inter-payment time	–	–	2.523 (0.257)

Table 4
Importance of input variables in CART.

	JCIC1	JCIC2	JCIC3	APR	JCIC4	Credit limits	Educational level
Importance (%)	100.0	64.57	57.35	50.65	38.06	25.79	7.35

parametric approaches. The mean of the terminal node (3rd column in Table 5) represents the predicted value of customer quality for credit card holders in that terminal node [3].

Since the rules derived by CART are mutually exclusive, an analyst can apply the generated rules to the credit-granting decision. Given these decision rules, the optimal levels of APR and credit limit can be obtained by grid search. For example, if the range of values for the APR and credit limit parameters are 15–20% and 100,000–300,000, respectively, we can divide the range of values for each of the parameters into 10 discrete levels. That is (15.0%, 15.5%, ..., 20.0%) and (100,000, 120,000, ..., 300,000). Then, the grid search method can be simply employed to determine the "expected customer quality" at each point on the grid of parameter values with an applicant's JCIC information based on the generated rules. Consequently, the grid point that yields the largest value of customer quality can be identified and the optimal levels of credit limit and APR can be assigned.

3.4. Model prediction and model comparison

For comparison, the discriminant analysis, logistic regression, neural network, multivariate adaptive regression splines (MARSs) and support vector machine (SVM) approaches were applied to evaluate the identification capability of the proposed Bayesian–CART-based model. The prediction accuracy of each approach is evaluated by the binary classification that is commonly applied by banks. We let applicants with $Z \geqslant 0$ be customers with good credit, and applicants with $Z < 0$ be customers with bad credit. The binary classification is preceded as follows: First, the Bayesian behavior scoring model is used to estimate the average quality of each customer in the predictive sample, given his/her observed credit usage, repayment behavior, and credit rating assigned by the bank. The average quality obtained from a Bayesian behavior scoring model is considered the "true" expected customer quality. Second, assume that customers in the predictive sample are new applicants, and the bank has access only to their demographic information and credit history.

In applying neural network, we used back-propagation network (BPN) in building the classification model due to its flexibility and universal approximating capability [12,16]. It means that the BPN model has potential to capture the complex relationships between many factors that contribute to a certain credit status (1 = good or 0 = bad credit). As recommended by Cybenko [6], Hornik et al. [12], one-hidden-layer network is sufficient to model any complex system. Thus, in this study, the designed network model has only one hidden layer and one output node-the credit status of the customer. To determine the optimal number of hidden nodes, we follow the principle of Kolmogorov's theorem on multilayer neural networks [16], which suggests that, during the process of training a multilayer neural network, the number of hidden nodes should be twice the number of input space parameters. This, for this particular stage of evaluation (12 input nodes in the input layer), implies that the number of hidden units should be $2 \times 12 = 24$. This then leads to the construction of a {12-24-1} neural network model, which stands for 12, 24, and 1 neuron in the input layer, hidden layer, and output layer respectively. However, because Kolmogorov's Theorem is merely a recommendation for the reliably adequate training of a multi-layer neural network, it can be regarded as a good starting point.

Moving forward, because reducing the number of hidden nodes in a neural network architecture is considered an optimization procedure by some practitioners, a quick stint to reduce the number of hidden nodes is undertaken, and the result is summarized in Table 6. The first column of Table 6 contains information about the neural network architecture, the second column is the classification accuracy, and the third column is the number of training repetitions. A high accuracy implies that the model has good generalization ability while a small number of training repetitions implies that the model is unlikely to be subjected to too many rote learning – a phenomenon that could lead to poor generalization ability. Thus, the optimal network architecture could be identified by the balance between high accuracy and a small number of training repetitions. As shown in Table 6, {12-23-1} is the optimal network

Table 5
Deduction rule (partial) for the illustrated dataset.

Terminal node	Rule	Predicted customer quality mean
1	If (JCIC1* < 3 and JCIC2 $\leqslant$ 90 months and JCIC4 < 2)	54.95
2	If (JCIC1 > 3 and JCIC2 $\leqslant$ 72 months and JCIC3 > 2)	−0.21
3	If (JCIC1 > 2 and JCIC2 < 2 and 2 < JCIC3 < 4 and APR $\leqslant$ 0.135)	4.90
4	If (JCIC1 > 2 and 1 < JCIC2 < 3 and 2 < JCIC3 < 4 and 0.095 < APR $\leqslant$ 0.18875 and 144500 < CL $\leqslant$ 275,000)	43.24

*JCIC1: the number of the short-term or mid-term non-secured loan.
JCIC2: the longest period of a credit card ownership (month).
JCIC3: the number of banks that a customer ever uses his cash advanced services in the past 12 months.
JCIC4: the number of new query of credit reports in the past 1 month.

Table 6
Network hidden node reduction.

Network topology	Accuracy (%)	Repetitions
{12-24-1}	88.1	5
{12-23-1}	90.6	5
{12-22-1}	83.0	9
{12-21-1}	80.6	9
{12-20-1}	75.3	10

topology in this study. We did not provide the result for the number of hidden node beyond 20 because further reduction of hidden nodes number, for instance 15, either sharply increases the number of training repetitions in order to sustain high accuracy or sharply drops the accuracy when NN is trained at a small number of repetitions.

In building the SVM classification model, the radial basis function (RBF) defined in [5,24] was used herein since it is one of the most widely adopted kernel function. Because [5] pointed out that for multivariate d-dimensional problems, the RBF width parameter σ is set as $\sigma^d \sim (0.1, 0.5)$, where d is number of input variables, $\sigma = 0.8$ is used for all experiments in this study for simplifying the setting of parameter σ. In addition, the analytic parameter selection method proposed by Cherkassky and Ma [5] and the grid search proposed by Hsu et al. [13] are applied here for parameters setting of C and ε. We first use the analytic method to select a parameter set of C and ε. Then, the grid search uses the set as starting point for searching. The parameter set of C and ε which generate the minimum forecasting mean square error (MSE) is considered as the best parameter set.

The discriminant analysis and logistic regression credit scoring models was implemented using the popular SPSS software (SPSS, 1998). CART 4.0 (2001) and MARS 2.0 (2001) provided by Salford Systems are used, respectively, in building the CART and MARS credit scoring models. All the modeling tasks are implemented on an IBM PC with Intel Core Duo-T2400 1.83 GHz CPU processor. The detailed credit scoring results using the above-mentioned six modeling techniques were summarized in Table 7.

From Table 7, we can conclude that, in both the testing and validation sample, Bayesian–CART-based scoring model has better classification capability in terms of the average correct classification rate (92.9%). Consequently, based on the results from this data set, we can conclude that the credit scoring results of Bayesian–CART outperform the commonly utilized linear discriminant analysis, logistic regression, neural networks, MARS and SVM credit scoring models and hence provide efficient alternatives in conducting credit scoring tasks. Besides, as the classification accuracy of the validation samples is only slightly lower than those of the corresponding testing samples, it means that the built scoring models successfully unveil the hidden information buried in the data and hence can be successfully applied in the validation sample after the model building procedure.

The influence of prior probability of occurrence can also be observed in Table 7. Table 7 shows that, in the testing (the validation) sample, the prior probability of observing good customers is 26%(28%), and the prior probability of observing bad customers is 74%(72%). Both discriminant analysis and logistic regression are highly impacted by prior probability, and have a prediction accuracy of the case "bad customer" greater than 90% and a prediction accuracy of the case "good customer" less than 70%. By contrast, CART BPN, MARS and SVM were less affected by prior probability.

In order to demonstrate that the continuous latent customer quality ($\hat{z}_{ij}$) from the Bayesian behavior scoring model allows considerable latitude for producing finer rules for credit granting decisions, credit scoring tasks are performed on the bank dataset with cardholder's credit status (1 = good or 0 = bad credit) directly assigned by the bank analysts. The detailed scoring results were summarized in Table 8.

From Tables 7 and 8, it can be found that the identification results made by the Bayesian-based models are better than those generated by other non-Bayesian-based models. The comparison results showed that the combination of Bayesian and CART not only takes advantage of the superior capability of Bayesian latent variable model to estimate customer quality but also adds the benefit of developing decision trees that can provide critical insight into the cardholder's payment behavior and facilitate improved credit granting process.

Table 7
Credit scoring results of the six constructed models based on the Bayesian estimated average quality $\bar{z}_i$.

	Testing sample[a]		Validation sample[b]	
	{good–good} (%)	{bad–bad} (%)	{good–good} (%)	{bad–bad} (%)
Bayesian–discriminant	66.35	93.21	65.57	92.86
Bayesian–logistic	68.22	92.18	67.72	92.93
Bayesian–BNP	90.11	85.61	92.75	82.19
Bayesian–SVM	91.87	86.55	90.38	87.28
Bayesian–MARS	92.15	85.73	90.53	86.63
Bayesian–CART	96.12	89.10	95.69	90.11

[a] In testing sample, the numbers of good and bad customers are 131 and 369, respectively.
[b] In validation sample, the numbers of good and bad customers are 138 and 362, respectively.

Table 8
Credit scoring results of the six constructed models based on the dataset with credit status directly assigned by the bank analysts.

	Testing sample		Validation sample	
	{good–good} (%)	{bad–bad} (%)	{good–good} (%)	{bad–bad} (%)
Discriminant	67.64	83.49	69.23	84.69
Logistic	66.19	85.79	68.38	87.73
BNP	82.04	77.10	89.48	79.03
SVM	86.33	79.15	89.52	79.71
MARS	89.61	78.88	87.39	78.29
CART	91.23	78.34	90.28	80.38

3.5. Type I and Type II errors of the constructed models

In addition to comparing the accuracy of prediction rates and the influence of prior probabilities, we also use misclassification costs to evaluate the model's performance. It is apparent that the costs associated with Type I errors (a customer having good credit is misclassified as having bad credit) and Type II errors (a customer with bad credit is misclassified as having good credit) are significantly different. In general, the misclassification costs associated with Type II errors are much higher than those associated with Type I errors [25]. The difference can range from 5 to 1 up to 20 to 1. Therefore, Type II errors of the six models need to be compared in order to justify the overall credit scoring capability. Table 9 summarizes the Type I and Type II errors of the six built Bayesian-based models. According to the results from Table 9, CART has a competitive rate of Type II errors and a significantly low rate of Type I errors in comparison with the other five approaches.

Table 9
Type I and Type II errors.

	Testing sample		Validation sample	
	Type I error (%)	Type II error (%)	Type I error (%)	Type II error (%)
Bayesian–discriminant	33.65	6.79	34.43	7.14
Bayesian–logistic	31.78	7.82	32.28	7.07
Bayesian–BNP	9.89	14.39	7.25	17.81
Bayesian–SVM	8.13	13.45	9.62	12.72
Bayesian–MARS	7.85	14.27	9.47	13.37
Bayesian–CART	3.88	10.90	4.31	9.89

Hence, we conclude that Bayesian–CART not only has higher classification accuracy, but also lower Type I errors and hence can reduce the possible high risks associated with Type I errors.

4. Conclusions

Evaluating a customer's credit risk is crucial for a credit card issuing bank due to the high risks associated with inappropriate credit granting decisions. Three challenges are encountered by the bank in this decision-making process. First, the bank would like to predict the type of applicant accurately. Second, given current cardholders' credit usage and repayment behavior, financial institutions would like to determine the optimal credit limit and APR. Third, the bank would like to improve its efficiency by automating the process of credit-granting decision-making.

In this paper, a new approach is proposed to combine the Bayesian behavior scoring model and the CART-based credit scoring model to overcome these three challenges facing by credit card issuing banks. The Bayesian behavior scoring model and the CART-based credit scoring model are combined by the average latent customer quality. The average latent customer quality represents the expected quality performance of a customer. Different from the binary credit rating, the continuous latent customer quality score provides the magnitude of customer quality and allows considerable latitude for producing finer rules for credit-granting decisions.

The empirical study shows that the demographic variables used in most credit scoring models have little explanatory ability with respect to a cardholder's credit usage and repayment behavior. A cardholder's credit history provides the most important information in credit scoring. Compared to the performance of discriminant analysis, logistic regression, neural network, multivariate adaptive regression splines (MARS) and support vector machine (SVM), the proposed approach has a 92.9% accuracy rate in predicting customer types, is less impacted by prior probabilities, and has the lowest misclassification cost.

Acknowledgements

The research leading to these results has received funding from the National Science Council in Taiwan under grant agreement number NSC-99-2410-H-027-017-MY2.

References

[1] D.J. Bartholomew, M. Knott, Latent Variable Models and Factor Analysis, Arnold, London, 1999.
[2] D.J. Bartholomew, Scaling unobservable constructs in social science, Applied Statistics 47 (1998) 1–13.
[3] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth, Pacific Grove, CA, 1984.
[4] You-Shyang Chen, Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach, Knowledge-Based System 26 (2012) 259–270.
[5] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (1) (2004) 113–126.
[6] G. Cybenko, Approximation by superpositions of a sigmoid function, Mathematics of Control, Signals, and Systems 2 (1989) 303–314.
[7] V.S. Desai, J.N. Crook, G.A. Overstreet Jr., A comparison of neural networks and linear scoring models in the credit union environment, European Journal of Operational Research 95 (1996) 24–37.
[8] A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association 85 (1990) 398–409.
[9] William H. Greene, Econometric Aanalysis, Prentice Hall, Upper Saddle River, NJ, 2000.
[10] D.J. Hand, M.J. Crowder, Measuring customer quality in retail banking, Statistical Modelling 5 (2005) 145–158.
[11] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.
[12] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.
[13] C. Hsu, C. Chang, C. Lin, A practical Guide to Support Vector Classification, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2003.
[14] L.C. Jen, C.H. Chou, G. Allenby, A Bayesian approach to modeling purchase frequency, Marketing Letters 14 (1) (2003) 5–20.
[15] L.C. Jen, C.H. Chou, G. Allenby, The importance of modeling temporal dependence of timing and quantity in direct marketing, Journal of Marketing Research 46 (4) (2009).
[16] V. Kurkova, Kolmogorov's theorem and multilayer neural networks, Neural Networks 5 (3) (1992) 501–506.
[17] Tian-Shyug Lee, Chih-Chou Chiu, Yu-Chao Chou, Lu Chi-Jie, Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, Computational Statistics and Data Analysis 50 (2006) 1113–1130.
[18] P.F. Pai, M.F. Hsu, M.C. Wang, A support vector machine-based model for detecting top management fraud, Knowledge-Based Systems 24 (2011) 314–321.
[19] S. Piramuthu, Financial credit-risk evaluation with neural and neurofuzzy systems, European Journal of Operational Research 112 (1999) 310–321.
[20] A.F.M. Smith, G.O. Roberts, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods., Journal of the American Statistical Association Series B 55 (1993) 3–23.
[21] D. Steinburg, P. Colla, Classification and Regression Trees, Salford Systems, San Diego, CA, 1997.
[22] M.A. Tanner, W.H. Wong, The calculation of posterior distributions by data augmentation, Journal of the American Statistical Society 82 (1987) 528–540.
[23] L.C. Thomas, A survey of credit and behavioral scoring: forecasting financial risks of lending to customers, International Journal of Forecasting 16 (2000) 149–172.
[24] V.N. Vapnik, An overview of statistical learning theory, IEEE Transactions on Neural Networks 10 (5) (1999) 988–999.
[25] D. West, Neural network credit scoring models, Computer and Operation Research 27 (2000) 1131–1152.
[26] G. Wang, Jian. Mac, L. Huang, K. Xu, Two credit scoring models based on dual strategy ensemble trees, Knowledge-Based Systems 26 (2012) 61–68.