



A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model

Bo-Wen Chi, Chiun-Chieh Hsu *

Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Keywords:

Dual scoring model
Mortgage behavioral scoring model
Credit bureau scoring model
Genetic algorithm
Logistic regression

ABSTRACT

Credit scoring model is an important tool for assessing risks in financial industry, consequently the majority of financial institutions actively develops credit scoring model on the credit approval assessment of new customers and the credit risk management of existing customers. Nonetheless, most past researches used the one-dimensional credit scoring model to measure customer risk. In this study, we select important variables by genetic algorithm (GA) to combine the bank's internal behavioral scoring model with the external credit bureau scoring model to construct the dual scoring model for credit risk management of mortgage accounts. It undergoes more accurate risk judgment and segmentation to further discover the parts which are required to be enhanced in management or control from mortgage portfolio. The results show that the predictive ability of the dual scoring model outperforms both one-dimensional behavioral scoring model and credit bureau scoring model. Moreover, this study proposes credit strategies such as on-lending retaining and collection actions for corresponding customers in order to contribute benefits to the practice of banking credit.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Credit scoring model is the most successful example of statistical model applied in financial institutions (Thomas, 2000). To achieve the purpose of risk control, application scoring models help banks to decide whether to grant credit to new applicants based on customers' characteristics such as age, education, marital status, and so on (Chen & Huang, 2003). In a similar vein, behavioral scoring models help banks to make use of analyzing existing customers' consumption behavior to evaluate future risk of delinquency and loss (Setiono, Thong, & Yap, 1998). Therefore, financial institutions try very hard to develop credit scoring models to separate customers' risk accurately, while concurrently effectively control risks to enhance the flexibility in funding use.

The causes of the US subprime mortgage crisis in 2007 were due to some reasons. First, US house prices began their steep decline after the peak in mid-2006, and the adjustable-rate mortgages began to reset at higher rates resulted in mortgage delinquencies soar. Second, financial institutions offered loans to higher-risk borrowers without appropriate review and documentation. The warning of subprime mortgage crisis made Taiwan's Financial Supervisory Commission to pay attention to see whether the banks' lending policy was too loose, and request banks to set up

a reasonable risk-based pricing policy. Therefore, it is essential to build up a mortgage scoring system to strengthen risk management mechanism.

The common approaches had been used to construct credit scoring models included linear discriminant analysis (e.g., Altman, Marco, & Varetto, 1994; Jo, Han, & Lee, 1997; Kim, Kim, Kim, Ye, & Lee, 2000), neural networks (e.g., Desai, Crook, & Overstreet, 1996; Mahlotra & Malhotra, 2003; Tsai & Wu, 2008; West, 2000; Zhang, Hu, Patuwo, & Indro, 1999), classification and regression tree (e.g., David, Edelman, & Gammerman, 1992; Feldman & Gross, 2005; Lee, Chiu, Chou, & Lu, 2006; Li, Ying, Tuo, Li, & Liu, 2004), and logistic regression (e.g., Dinh & Kleimeier, 2007; Laitinen & Laitinen, 2000; Steenackers & Goovaerts, 1989; Westgaard & Wijst, 2001). Altman (1968) applies multiple discriminant analysis to predict corporate bankruptcy. Altman's study has a sample size of 66, which are comprised of 33 manufacturing companies that have filed for bankruptcy from 1946 to 1965 and 33 non-bankrupt companies matched on a stratified random basis by both industry and asset size. Altman selects the five most significant variables from 22 financial ratio variables to construct a bankruptcy prediction model. These variables are working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total liabilities and sales to total assets. This model has an accuracy rate of 95% one year prior to bankruptcy, and an accuracy rate of 72% two years prior to bankruptcy. Atiya (2001) applies neural networks (NNs) to bankruptcy prediction for credit risk. He indicates that the use of

* Corresponding author. Address: No. 43, Sec. 4, Keelung Road, Taipei 106, Taiwan. Tel.: +886 2 27376766; fax: +886 2 27376777.

E-mail address: cchsu@cs.ntust.edu.tw (C.-C. Hsu).

equity-based indicators in addition to financial ratio indicators provides significant improvements in the prediction accuracy. Feldman and Gross (2005) apply classification and regression trees to analyze Israeli mortgage default data. They find that if the costs of accepting bad customers are more than rejecting good ones, borrowers' features are the powerful predictors of default rather than mortgage contract features. However, if the costs are equal, mortgage contract features are powerful predictors as well. The higher (lower) the ratio of the misclassification costs of bad risks compared to good risks, the lower (higher) are the resulting misclassification rates of bad risks, and the higher (lower) are the misclassification rates of the good risks. Steenackers and Goovaerts (1989) apply stepwise logistic regression model to find out the criteria for the selection of new credits. They use personal loans collected from a Belgian credit company during the period from 1984 to 1986. They list 19 characteristics of applicants and indicate the selecting criteria including age, with a telephone or not, duration of residence in current place, duration of employment period, standard of living area, occupation, civil servant or not, monthly income, ownership of the house, number of past loans, and length of the loan.

The above-mentioned methods use a one-dimensional approach to construct the credit scoring model. However, more recent studies have proposed to use a combined model or a combined score approach to enhance the predictability of the one-dimensional credit scoring model.

Altman et al. (1994) compare NNs with linear discriminant analysis (LDA) for Italian industrial firms using a large sample size. The results show that LDA outperforms NNs. However, they suggest integrating NNs and LDA due to the performance of NNs is subject to enhancement by integration with statistical approaches. Lee and Jung (2000) compare the forecasting ability of logistic regression (LR) with that of NNs to identify creditworthiness of urban and rural customers. The results show that LR has better prediction for urban customers and NNs have better prediction for rural customers. They also mention a combination of two models to enhance the forecasting accuracy. Koh, Tan, and Goh (2006) combine individual credit scoring models to build a combined model which performs better than the individual models. Nonetheless this study suggests that limitation is the possible difficulty in interpreting rules generated by the combined model. Furthermore, the combined model is unlikely to make sense because it is built on individual models that are likely to generate different rules.

Zhu, Beling, and Overstreet (2001) investigate the performance of a regression-based combination of application credit score and bureau credit score. The results show that the combined score outperforms each set of scores from its basis; however, this research also suggests that the dependence of the two individual scores should be taken into account.

Due to the difficulty in interpreting the rules generated by the combined model and the high collinearity between two consumer credit scores, we combine the bank's internal mortgage behavioral scoring model with the external credit bureau scoring model to construct the dual scoring model. In this way, we are able to make more accurate risk judgment and segmentation to further discover the parts which are required to be enhanced in management or control from mortgage portfolio. The results show that the predictive and distinguish ability of the dual scoring model outperforms both of the one-dimensional behavioral scoring model and credit bureau scoring model. Consequently, this study recommends applying the dual scoring model in bank's credit strategies.

The remainder of this paper proceeds as follows. Section 2 discusses research design and empirical methodology. Section 3 presents the sampling procedure, data, and empirical results. Sec-

tion 4 describes the application of bank's credit strategies based on the dual scoring model. A brief conclusion follows.

2. Analysis methodology

The literatures mostly emphasize on one-dimensional perspective to construct credit scoring model in order to measure customer risk. In this study, we use the bank's internal data to construct the mortgage behavioral scoring model. In addition, we utilize customers' credit bureau data from the Public Credit Registers (PCR) to construct the credit bureau scoring model, and combine the mortgage behavioral scoring model with the credit bureau scoring model to yield a two-dimensional dual scoring model. This dual scoring model is capable of more accurately segment customer risk and improve credit assessment. The development process of dual scoring model in this study is shown in Fig. 1:

Step 1: Data preprocessing.

In this step, we collect bank's internal data and credit bureau data from the PCR on existing mortgage customers, and engage in data cleaning, data integration, data transformation, data reduction, and generation of feature selection from candidate models for the above two types of data.

Step 2: Segmentation analysis.

In this step, we apply segmentation analysis to classify customers into homogeneous risk groups based on the customer characteristics, and thereby enhance the predictability of the model.

Step 3: One-dimensional credit scoring model building.

We build a one-dimensional credit scoring model respectively for segmented groups and calibrate several scoring models to derive a consistent score-to-odds relationship. The behavioral score and credit bureau score are divided into BHS1, BHS2, BHS3, BHS4, BHS5 and CBS1, CBS2, CBS3, CBS4, CBS5 risk ranks respectively in accordance with the degree of risk score.

Step 4: Dual scoring model building.

The five risk ranks of the mortgage behavioral scoring model and credit bureau scoring model are combined into a 5×5 risk matrix, and the Kolmogorov-Smirnov (K-S) statistic and Receiver Operating Characteristic (ROC) curve are used to measure the predictability of the credit scoring model.

Step 5: Credit strategy.

Based on the 5×5 risk matrix, we design corresponding credit strategies such as on-lending retaining and collection actions.

2.1. Data preprocessing

Data preprocessing is required to ensure data field consistency and determine the relative importance variables in credit scoring model building. In this study, we collect bank's internal data and credit bureau data from the PCR on existing mortgage customers. The bank's internal data include borrower characteristics, collateral characteristics and payment characteristics. The credit bureau data include mortgage customers' credit-related information of amounts owed, payment history, new credit, length of credit history, types of credit used, and account information. The data preprocessing will be proceeded as follows:

1. *Data cleaning*: Using the attribute mean to fill in the missing value, smooth out noisy data, identify or remove outliers, and resolve inconsistencies.
2. *Data integration*: Using a customer identifier to join different data sets.
3. *Data transformation*: Scaling attribute values to fall within a specified range (normalization), move up in the concept hierarchy on the numeric attributes (aggregation), move up in the

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات