A DATA MINING APPROACH FOR CONSTRUCTING A REASSIGN CREDIT SCORING MODEL

Rong-Ho Lin , Department of Industrial Engineering and Management, National Taipei Univ. of Technology, No. 1, Sec., 3, Chung-Hsiao E. Rd., Taipei, Taiwan, ROC, 106, rhlin@ntut.edu.tw*

Chun-Ling Chuang, Department of Information Management, Kai-Nan University, No. 1, Kainan Road, Luzhu, Taoyuan, Taiwan, ROC 338, clchuang@mail.knu.edu.tw

Szu-Teng Huang, Department of Business Management, National Taipei Univ. of Technology, No. 1, Sec., 3, Chung-Hsiao E. Rd., Taipei, Taiwan, ROC, 106, ahbow911@yahoo.com.tw

ABSTRACT

The objective of this paper is to propose a two staged reassigning credit scoring model (RCSM) for solving the classification problem. First, the classification stage of RCSM is to classify applicants with good or bad credits. Second, the reassign stage of RCSM is built to reduce the Type I error by reassigning the rejected good credit applicants to the conditional accepted class. The result indicates that the RCSM model not only provides more accurate credit scoring than that of LDA, LR, CART, SVM and ANNs, but also contributes to increase business revenue by decreasing the Type I error.

INTRODUCTION & LITERATURE REVIEW

Data mining uses a broad family of computational methods that include statistical analysis, decision trees, neural networks, rule induction and refinement, and graphic visualization. All these data mining methods aim at solving the classification problems that play an important role in business decision making. Various credit scoring models have been developed by banks and researchers to solve the classification problems, such as linear discriminant analysis (LDA), logistic regression (LR), artificial neural networks (ANN), multivariate adaptive regression spline (MARS), classification and regression tree (CART), and case based reasoning (CBR).

First developed in 1936 as the earliest method to construct credit scoring model, LDA has often been criticized for its assumption of linear relationship between input and output variables which seldom holds and its sensitivity to deviations from the multivariate normality assumption [7]. In addition to LDA, LR is another common alternative to conduct credit scoring tasks [3]. Basically, the LR model is a technique in predicting dichotomous outcomes and does not require the multivariate normality assumption. However, both LDA and LR are based on the assumption that the relationships between variables are linear, making them less accurate in credit scoring. Free from this problematic assumption, ANN is usually adopted as a new alternative and proved to be more accurate than LDA and LR. However, ANN is also being criticized for its long training process in obtaining the optimal network and greater difficulty in identifying the relative importance of potential input variables and interpreting the results [6]. In addition to the above-mentioned methods, MARS can both work by itself as a classification technique and serve as an effective supporting tool for neural networks as the advantages of MARS can compensate the shortcomings of neural networks [2]. Except for being used in classification models, CBR is also being extensively applied in credit scoring models. Therefore, Many studies have been performed to increase the classification accuracy of various methods [1] [4] [5].

RESEARCH METHOD

The purpose of this study is to present a reassigning credit scoring model (RCSM) working through the two phases of classification and reassignment to promote the accuracy of classification. In the first phase, credit scoring stage, MARS is used to obtain key input variables of the ANN model to reduce the number of input nodes, simplify the network structure, and shorten the model building time. The ANN model is then used to classify credit card applicants into good and bad credit groups. In the second phase, reevaluating and reassigning stage, the rejected applicants are re-evaluated by the reassigning credit scoring model, using CBR to compare similarities between rejected credit card applicants based on the good and the bad CBR databases. If the value of SG (similarity retrieved from the good applicants database) is higher than that of SB (similarity retrieved from the bad applicants database), a rejected applicant is reassigned to conditional approval to reduce Type I error and increase the banking portfolio; otherwise, the applicant will be reassigned back to rejected group.

EMPIRICAL STUDY

The German dataset, consists of loans granted to a total of 1,000 credit card applicants with 20 independent variables, is adopted herein to evaluate the predictive accuracy of our reassigning credit scoring model and the capability of the CBR-based reassign classification.. Among them, 600 (420 good and 180 bad) applicants are randomly selected as the training sample, another 200 applicants (140 good and 60 bad) are used to test the model, and the remaining 200 applicants (140 good and 60 bad) are retained for validation.

A. Results of RCSM credit scoring

From the 20 original independent variables, MARS obtains 8 significant ones as the input nodes of the proposed model. The training of the network is also implemented with various learning rates and training lengths ranging from 1,000 to 10,000 iterations until the network converges. The network weights are also reset for each combination of the network parameters such as learning rates (from 0.01 to 0.4) and momentum (from 0.8 to 0.99). Several options of the ANN architectures for the testing data are evaluated, in which 8-16-1 is found to obtain better results, and the learning rate, momentum and number of training epochs are set to 0.1, 0.9 and 3000, respectively. The credit scoring results of the validation sample are summarized in Table 1.

B. Results of RCSM reassignment

For decreasing the potential errors of credit scoring and increasing the business revenue of the credit card issuer, a more productive measure may be adopted to reinforce or enhance the credit scoring model in a manner that gives applicants rejected after initial screening an opportunity of being re-evaluated and granted conditional acceptance via CBR-based method if his or her SG value is higher than the SB value. According to the results of credit scoring as shown in Table 1, two groups of applicants emerge to call for reevaluation and reassignment. Group 1 incorporates the 9 applicants approved in the original dataset but rejected after RCSM credit scoring while Group 2 covers the 34 applicants rejected both in the original dataset and after RCSM credit scoring. One applicant in Group 2 and eight applicants in Group 2 can be granted conditional approval after CBR-based reassignment; as a result, with a total of 9 applicants reassigned to conditional acceptance, the overall approval rate rises from 82.5% (165/200) to 86% (172/200) shown in Table 2.

Table 1: Results of RCSM credit scoring

Original Status	Status after RCSM Credit Scoring	
	Bad credit	Good credit
Good credit	9 (6.4%)	131 (93.6%)
Bad credit	34 (56.7%)	26 (43.3%)

Overall % Correct: 82.5%

Table 2: Results after RCSM Reassignment

Original Status	Post-Reevaluation Status	
	Bad credit	Good credit
Good credit	1 (0.71%)	139 (99.29%)
Bad credit	33 (55%)	27 (45%)

Overall % Correct: 86%

CONCLUSION

In order to evaluate the classification capabilities of the credit scoring models constructed by most four frequently used methods, the credit scoring results of the validation samples are summarized in Table 3. The results show that the proposed RCSM not only outperforms the commonly utilized LDA, LR, ANN and CART models but also provides a more efficient alternative in conducting credit scoring tasks. Therefore, this study compares five frequently used credit scoring approaches and demonstrates the advantages of applying MARS, ANNs and CBR to credit analysis. The study further proposes a reassigning credit scoring model (RCSM) to classify credit card applicants with greater efficiency and accuracy. Results of empirical studies indicate that our proposed approach is capable of not only reducing but virtually eliminating Type I error. Moreover, reassignment of originally rejected applicants to conditional approval helps both to prop up the approval rate of credit card application and to safeguard against occurrence of bad bet, enabling credit card issuing institutions to reduce credit risks and increase business revenue.

Table 3: Credit scoring results of the constructed models

Methods	Classified class		Type I error %	Type II error %	Accuracy rate%
	Bad-Bad	Good-Good			
LDA	45	107	24	25	76
LR	29	124	11	52	76.5
CART	44	111	21	27	77.5
ANN	33	126	10	45	79.5
RCSM	33	139	0.71	45	86

REFERENCE

[1] Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. Credit rating analysis with support vector machine and neural networks: A market comparative study. *Decision Support Systems*, 2004, 37, 543–558.

[2] Lee, T. S. & Chen I. F. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 2005, 28, 743–752.

[3] Laitinen, E. K. Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International Review of Financial Analysis*, 1999, 8, 97-121

[4] Li, Y. Predicting materials properties and behavior using classification and regression trees. *Materials Science and Engineering*, 2006, 433, 261-268.

[5] Ong, C. S., Huang, J. J. & Tzeng, G. H. Building credit scoring models using genetic programming. *Expert Systems with Applications*, 2005, 29, 41–47.

[6] Piramuthu, S. Feature construction for reduction of tabular knowledge-based systems, *Information Sciences*, 2004, 168, 1-4, 201-215.

[7] West, D. Neural Network Credit Scoring Models. *Computer & Operations Research*, 2000, 27, 1131–1152.