A Two-Stage Cardholder Behavioural Scoring Model Using Artificial Neural Networks and Data Envelopment Analysis

I-Fei, Chen

Tamkang University, Department of Management Science and Decision Making, 151 Ying-chuan Road, Danshui Dist., New Taipei City 25137 Taiwan, enfa@mail.tku.edu.tw
doi:10.4156/ijact.vol3.issue2.11

Abstract

Since the databases that banks use for analysis of cardholders’ repayment behaviours are usually large and complicated, and the extant classification techniques hardly offer 100% correct classification accuracy so as to possibly incur a considerable loss associated with type II errors, the prediction of cardholders’ future payment behaviours has been still referred to as a difficult task in the credit industry.

This paper proposes a two-stage cardholder behavioural scoring model, with merits of artificial neural networks (ANNs) and data envelopment analysis (DEA), which not only enables banks to verify the ANNs predicted results of each cardholder’s future repayment behaviour as well as to identify creditworthy cardholders who is profitable with low risks, but also provides guidelines to improve contributions of each inefficient cardholder for card issuer profitability.

Keywords: Chi-square Automatic Interaction Detector (CHAID), Artificial Neural Networks (ANNs), Data Envelopment Analysis (DEA), Behavioural Scoring, Data Mining

1. Introduction

The profitability of credit cards has been shrinking over the past decade in that the outbreak of such events as non-performing loan storm, financial tsunami, regulations of reduction in the revolving interest rate etc. Furthermore, credit card market has appeared to be saturate so that the growth of card circulation is staying sluggish and the yields from interest or miscellaneous charge are inclined to diminish. To circumvent the increasingly fierce competition and time pressure, card issuers need to not only engage in long-term relationship management with their retained customers but also evade potential risks of default in advance.

To effectively predict the repayment behaviour and increase the contribution of cardholders, this paper aims to construct a cardholder repayment behavioural scoring model based on a two-stage method. In the first-stage, classification models are to be built using Chi-square automatic interaction detector (CHAID) and artificial neural networks (ANNs). In the second stage, data envelopment analysis (DEA) commonly used to evaluate management performance at corporate level is to be initatively applied in credit risk management to evaluate the efficiency of each individual cardholder and elicit some additional information that is generally hidden but valuable to credit card issuers. In other words, this proposed behavioural scoring model is expected to provide guidelines for identification of the correctly classified results (with low credit risk) and discernment of the efficient and inefficient customers. For the inefficient customers, it is also to prescribe the contribution improvement directions (with profitability exploitation) via DEA slack analysis. To examine the feasibility of the proposed model, a real-world dataset released by a Taiwan local bank is applied. In this empirical study, there are 700 cardholder accounts, each of which contains 34 variables including cardholder’s demographic profiles, purchase and transaction records, repayment history, etc. over six successive months as the observation period, one year before embarking on prediction, and every holder is to be predicted into three categories suggested by bank practitioners, namely good account who pay in full, revolver, and bad account (customers who default), using CHAID and ANNs classification models.

Based on 5-fold cross-validation, the average correct discrimination rate of each model on the 5 testing datasets is then estimated to find the better classification model and result. In the second stage, the better classification results derived in the previous stage are examined using DEA. Figure 1
A Two-Stage Cardholder Behavioural Scoring Model Using Artificial Neural Networks and Data Envelopment Analysis

I-Fei, Chen


illustrates the framework of the proposed two-stage behavioural scoring model. Through feedback, the relative efficiency, we attempt to not only ensure the discriminating power of the classification model but also identify the future repayment behaviour of each individual cardholder correctly predicted by the classifier for reduction in the possibility that the behavioural model may mislead banks to make their own customer policy due to misclassification. Thus, we estimate the relative efficiency of each group of cardholders classified by the ANNS model. Besides, DEA also provides suggestions on how to improve inefficient customers’ performance and contribution favorable to bank profitability.

Figure 1. Flowchart of steps in constructing a cardholder behaviour evaluation model
The remainder of this paper is organized as follows. Section 2 reviews literatures about behavioural scoring models; Session 3 gives a brief introduction to CHAID, ANNs and DEA; Session 4 presents a detailed discussion on the empirical results obtained from the proposed behavioural scoring model; Session 5 concludes the study with suggestions for further researches.

2. Literature Review

To avoid risks caused by incomplete credit investigation and costs incurred by debt collection operations, banks will conduct a comprehensive assessment of each customer’s application data (cross-sectional) as a measure of credit risk management. If an applicant’s risk exceeds the level they can tolerate, they should reject his/her application in the very first stage [1]. Generally speaking, this stage is relatively more emphasized by banks in the present. The importance of cardholder behaviour evaluation is considerable. If banks can accurately get hold of emergence of bad debts, they can save much cost of debt collection. For instance, banks can keep track of cardholders’ transaction records (longitudinal), including amount of repayment and spending history, and convert them into useful data. With these data, they can analyze cardholders’ purchase and transaction behaviours, apply proper credit controls, and detect abnormal use, fraud or loss of credit cards [2, 3, 4, 5]. If card issuers can predict non-performing credits in advance, they can effectively reduce various costs and enhance their competitiveness. [3] proposes a customer classification model based on credit and behaviour scoring. He conducts a cross-discipline integration of statistical theories and operations and takes into account customers’ economic conditions in the scoring to make up the technical insufficiency of conventional methods and process more complicated customer data. In the present, compared with credit scoring, behavioural scoring is less academically researched [6]. Hence, this paper attempts to apply credit scoring methods to behaviour scoring to construct a back-end credit review mechanism. This mechanism is expected to help financial institutions forecast customers’ payment behaviours and detect onset of non-performing credits.

Most previous researchers adopt the two-stage method for analysis of classification results. Compared with using one single classification instrument, they can obtain better solutions using the two-stage method [7, 8]. According to [9], the two-stage hybrids method demonstrates 80% classification accuracy and stronger prediction power than conventional single approaches. In addition, integration of conventional statistical analyses and ANNs is continuously applied and innovated in many research areas. For instance, considering inaccuracy of conventional approaches, such as discriminate analysis, logistic regression analysis, and ANNs, when applied to business failure prediction, [10] and [11] propose a DEA-based credit scoring model and use financial ratios to synthesize a firm’s overall performance into a single credibility score.

3. Research Method

The main idea of this paper is to propose a two-stage approach to construct a behavioural scoring model by integrating a classifier, ANNs or CHAID, for the classification sub-model and performance evaluation tool, DEA, for the relative efficiency sub-model as figure1 illustrates. From a general perspective, it manifests itself the significant merits of ensemble methods. This section is introducing the analysis techniques used in this paper.

3.1. Artificial neural networks (ANNs)

A neural network is a system comprised of highly inter-connected, 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. 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. Among ANNs, back-propagation network (ANNs) is most widely used. According to [12], 78% of business applications using neural networks use the ANNs training algorithms.

Neural networks are increasingly found to be useful in modeling non-stationary processes due to its outstanding generalisation capability and may possess superior prediction capability than other data mining techniques [13, 14]. Based on these facts, neural networks have been widely used in engineering, science, education, social research, medical research, business, finance, forecasting and related fields [8, 15]. Neural networks have also been explored in handling behavioural scoring problems [16, 17]. As neural networks are primarily designed to capture subtle functional relationship among variables, the majority of the above references have reported that the scoring accuracy of neural networks is better than those of other classification techniques.

3.2. Chi-square automatic iteration detector (CHAID)

Decision tree is a data classification technique, and CHAID is one of the decision tree algorithms. In this paper, CHAID is adopted for comparison with an ANNs-based classification model and to testify the applicability of this proposed two-stage behavioural scoring modeling procedure. CHAID uses a tree structure for classification and prediction. It is rule-based, easy to interpret, and can extract hidden or implicit information [18]. CHAID is an extension of automatic interaction detection (AID) proposed by [19]. AID is intended to mine the correlation between variables that affect responses and be focused on interactions between variables. In [20] integrates chi-square into AID to form CHAID and uses Bonferroni adjusted p-values as the basis for further splitting. In Chi-square test, homogeneous categories can be merged first and then split using a semi-hierarchical sequential search technique. When performing CHAID, responses to predictor variables are merged in pair and then split to ensure that the categories of responses to each predictor variable are minimal. The predictor variable that has the largest significance is first used to split the original sample into multiple subsets, which can be viewed as a population and further split according to the above process. This splitting process is recursively carried out until each subset of data consists entirely or dominantly of examples from one category. The difference between AID and CHAID in the splitting process lies in that AID can only split predictor variables into two categories, and CHAID is not subject to such limitation. Hence, CHAID has a higher data explaining power and is extensively applied in commercial statistical software.

As to predicting profitability from credit card delinquents, a recent study for example, [17] builds 26 ANNs models and 26 CHAID models. They estimate the utility value of each classification model and estimate the important variables that affect debt repayment of high-risk cardholders. Their empirical findings suggest that classifying customers into three groups is helpful for extraction of valuable information, and credit card companies can find potential customers by building this kind of models.

3.3. Data envelopment analysis (DEA)

DEA is an efficiency evaluation model developed by [21]. As a multi-objective decision-making tool, it analyses multiple input and output variables of DMUs (decision making unit) to figure out their relative efficiency. Before assessing each DMU’s efficiency, DEA does not presume the relationship between each input and output variable but compares the relative efficiency among DMUs to decide their efficiency value. Also, for inefficient DMUs, specific suggestions can be provided so that the composition of input and output items can be properly adjusted to achieve higher efficiency.

Since efficiency evaluation in DEA is based on the concept of Pareto optima, there may be more than one DMU judged as efficient. For instance, if there are 4 DMUs (DMUA, DMUB, DMUC and DMUD) whose input items are X1 and X2 while the output item is Y, then their relative positions can be illustrated as in Figure 2. In DEA, efficiency is computed on the basis of the envelope or efficient frontier, formed by all the observation values near the original point O, efficiency is obtained from the ratio between the distance from the original point to the relative point of the envelope and the distance from the original point to the observation point (optimal value=1). Therefore, it can be understood from
that the efficiency of DMUA is \( \frac{OA}{OA} \leq 1 \), efficiency of DMUB is \( \frac{OB}{OB} = 1 \), efficiency of DMUC is \( \frac{OC}{OC} \leq 1 \), and efficiency of DMUD is \( \frac{OD}{OD} = 1 \). In other words, DMUA and DMUC are inefficient DMUs while DMUB and DMUD are efficient DMUs. The slack variables of inefficient DMUs are not equal to zero, so the result of slack analysis can be adopted to improve the input or output items. Take DMUA for instance, the input value of resources \( X_1 \) and \( X_2 \) can be lowered to the level of DMUA’, so that DMUA can become an efficient DMU. The number of resources reduced is the value of slack variable in the input construct, suggesting aspects in which efficiency of DMUA can be improved in terms of resource input.

DEA has been widely applied to performance evaluation in many fields [22, 23, 24], especially in the financial industry [11]. Nevertheless, very few studies have been dedicated to the application of DEA in issues related to customer behavioural scoring models.

**Figure 2. DEA efficiency values**

**4. Empirical Study**

To examine the feasibility and effectiveness of the proposed two-stage behavioural scoring model using ANNs, CHAID and DEA, a credit card customer dataset released by a local bank in Taiwan is used in this study. There are totally 700 cardholders in the dataset with 500 good credit customers (class 0), 175 revolvers (class 1) and the remaining 25 are bad credit customers (class 3). The relative ratios of bad customers to total customer is 3.57% very close to the national standard in Taiwan and hence should be a representative dataset for testing the practicability of the proposed scheme. Each cardholder in the dataset contains 34 independent variables containing demographic characteristics, credit product types, repayment history, credit usage records and transaction data during the observation period, and the dependent variable is the credit status of performance period which is a year away from observation labeled as good, revolver, and bad.

In the first stage, we apply CHAID and ANNs to construct an optimal customer behavioural scoring model. The average correct discrimination rates of ANNs model and CHAID model are 83.83% and 82.92%, respectively. The difference is not large, but ANNs model is still slightly superior to CHAID model. We thus adopt Wilcoxon Rank Sum Test to ensure whether the difference between the two models reaches significance. Given a significance level at 0.05, the two-tailed test result indicates that the difference is not significant. In other words, although CHAID model is inferior to ANNs model in terms of overall correct discrimination rate, the important variables it derives for screening cardholders’ future payment behaviours are still powerful. Therefore, we will use these variables as the input and output variables in the second-stage DEA relative efficiency model.

The second-stage analysis with DEA was performed on Frontier Analyst 3 released by Banxia. In the previous stage, we used ANNs model to classify each subset of data based on five-fold cross validation. Each subset of test data was classified into three categories and every categorized group are used as decision making units (DMUs) in DEA. In this paper, 34 predictor variables were adopted.
Subject to the constraint of cost and time, we were unable to discuss and include all input and output factors. Besides, as indicated by Wilcoxon Rank Sum Test, the two classification models established in the first stage yielded no significantly different results. Therefore, we employed the important variables derived by CHAID model as the input and output variables in DEA.

To observe the managerial implications of DEA, we present the DEA efficiency analysis result of the three categories in one table (Table 1). The relative efficiency values of cardholders pre-classified into Category 0, Category 1, and Category 2 have been combined to facilitate examination. We combine them after estimating the relative efficiency of cardholders in each category, so this combination will not affect the overall efficiency value. As shown in the following Table, the initial sample consists of 700 cardholders. A total of 161 cardholders are efficient and pre-classified by ANNs model into Category 0 (115 persons), Category 1 (33 persons), and Category 2 (13 persons). Among these efficient cardholders, 112 persons have been correctly classified into \{0-0\}, 31 into \{1-1\}, and 13 into \{2-2\} by ANNs model. The overall correct discrimination rate is 96.89%. This explains that DEA can be used to increase ANNs classification capability, especially for bad accounts \{2-2\}, for which its correct discriminate rate reaches 100%. On average, ANNs discrimination results are verified 96.89% as correct classification when cardholders are efficient. In other words, an efficient customer was misclassified only as class 0 or class 1 by ANNs and all of efficient class 2 customers were not ever mistakenly categorized. Any of class 2 cardholder who is considered the most highly risky and will produce the most misclassification costs should be correctly predict his/her future repayment/credit status.

In a review of categories misclassified by the ANNs model from the standpoint of banks, we can find customers in categories \{0-1\} (3 persons) and \{1-0\} (2 persons) are profit-generating to banks but are misclassified. Through this analysis, banks can explore potential customers who can generate profit (revolving) or who have low risks (normal payment) and set up individualized card usage programs. For banks, developing long-term, stable, and positive customer relationships is also another way to increase their profits. Hence, through DEA efficiency value supporting, we can be more confident about ANNs results and confirm that efficient customers are creditworthy.

### Table 1. Integrated DEA efficiency values of the five subsets of data

<table>
<thead>
<tr>
<th>DEA relative efficiency analysis</th>
<th>Pre-classification of ANNs</th>
<th>Accuracy (%)</th>
<th>Overall classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good account Category 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revolver Category 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad account Category 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of cardholders in each category classified by ANNs</td>
<td>115</td>
<td>33</td>
<td>13</td>
</tr>
<tr>
<td>Actual category</td>
<td>Good account Category 0</td>
<td>112</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Revolver Category 1</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Bad account Category 2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to support for ANNs classification of efficient customers, DEA provides goal values for inefficient customers. From the standpoint of banks, this study uses output-oriented efficient analysis to help banks find directions for improvement in customer relationship management. By analyzing inputs of inefficient customers, banks can enhance the efficiency values of customers according to the goal values to increase their contributions. Take DMU No. 1327 as an example (see Table 2) whose efficiency is evaluated 25.36% classified by ANNs model into Category 0. This DMU
views DMU No. 214 in the same category as a benchmark. Through evaluation of this customer’s balance of spending and average total balance of the last 3 months, we can find there is considerable room for improving this cardholder in these two aspects. For the former, the customer should be provided more incentives to escalate to level 5 from level 1; the latter needs escalating from level 2 to level 9. Specifically, if the card issuer provides a customized card usage program for this customer to increase his/ her loyalty and frequency of using its credit card, this customer can reciprocally induce growth of his/ her spending as well as his efficiency and contribution to the card issuer.

Table 2. Goal values for DMU-1327 cardholder

<table>
<thead>
<tr>
<th>Efficiency 22.22%</th>
<th>Variable</th>
<th>Actual value</th>
<th>Goal value</th>
<th>Potential room for improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input variable</strong></td>
<td>Minimum repayment</td>
<td>2</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Output variable</strong></td>
<td>Balance of spending</td>
<td>1</td>
<td>5</td>
<td>400%</td>
</tr>
<tr>
<td></td>
<td>Average total balance of the last 3 months</td>
<td>2</td>
<td>9</td>
<td>350%</td>
</tr>
</tbody>
</table>

5. Conclusion

Relying on this two-stage behavioural scoring modeling scheme, DEA can not only verify classified results of ANNs model when DMUs are efficient, but also provide references for banks to make optimal decisions on credit extension. To cope with misclassification of ANNs model, DEA’s relative efficiency analysis can be applied to further differentiate customers in the same predicted category, and hence we can be more confident about ANNs results and confirm that efficient customers are creditworthy. As such, the loss incurred by type II errors can be reduced, and the probability that banks approve credits to the delinquent customers is decreased as well. DEA also can help decision makers get hold of additional information, such as which customer group is profit-generating or which customer group has low risks. For customers with an efficiency value under 100%, decision makers can use efficient customers as a benchmark to set up customized improvement programs, by which they can enhance each individual cardholder’s efficiency and contribution and the overall credit quality of the banks. Through integration of DEA, we provide banks a more substantive basis for improvement and managerial implications. Referring to suggestions for the future studies, inputs/ outputs of the second-stage DEA model should be the variables selected from the classifier (ie. ANNs), whose predicted results need to be verified, rather than those of the alternative classifiers (ie. CHAID) in the first-stage model so as to obtain better verifying capability.

6. References

A Two-Stage Cardholder Behavioural Scoring Model Using Artificial Neural Networks and Data Envelopment Analysis
I-Fei, Chen