Constructing credit auditing and control & management model with data mining technique

S.C. Chen *, M.Y. Huang

Department of Industrial Engineering & Management, National Chin-Yi, Institute of Technology, Taichung, Taiwan, ROC

ARTICLE INFO

ABSTRACT

The 2008 financial tsunami, hitting the globe across all types of industries, causing tides of bankruptcies and severe unemployment, had its epicenter at American subprime in the housing market. In fact, the US subprime storm was just a premonition, while the root cause of the financial tsunami lied in the oversupply of structured credit products. Credit card business, one of the structured credit products, which under an intensively competitive environment, have been released by many banks with high spread, high return, and easy-to-apply appeals to carter to consumers needs. In order to allure the customers, some banks even go to the extent as simplify the credit rating, which in turn has increased credit risk, causing high non-performing ratio, increased debt collection cost, and growing bad debt counts. Accordingly, credit risk auditing plays a vital role in the successful management of credit card business. In response to such needs, the present study aims to conduct analysis and investigation on the current status of the industry with CRISP-DM model. First, customers' demographic data and payment-related statistics were analyzed to identify feature variables, which were then sorted out as demographic data, debt data, payment rating etc. Next, by utilizing artificial neural network of data mining technique, the study tries to predict customer's regular pattern of consumption, payment and/or default and bad debt, and to develop a set of credit granting principle by employing the decision tree technique. Since data mining classification model has a greater power in discriminating credit card granting, it can thus be used to construct accurate credit variable rules and predictive model, to further improve credit checking effect and credit risk control. Using the credit auditing data of a certain bank as a case study, the study intends to verify that the model constructed by the researcher can effectively identify the potential key factors of its credit card granting rule, to minimize the cost loss of Model I and Model II credit business, and eventually enhance the stability and profitability of the bank's credit card business.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

To date, credit card has become one of the indispensible market phenomena of currency transaction system. This is particularly true of the world with a predominance of commercial interests, where either on the bank end or the consumers end, a finance management idea of "Enjoy first, pay later" is encouraged. Taking the department stores as an example, almost all of them offer interest-free installment at the anniversary day to attract the consumers to conduct advance consumption with their credit. The shadow of the duel-card (credit card and cash card) storm triggered 3 years ago, still lingers in the public mind. According to the latest statistics, until October 2008, total circulation of domestic credit cards amounts to 36.4 millions. And, this is only the new low since the 2005 credit card bad debt storm, because the number of newly issued cards were outnumbered by the cut and suspended

cards. Prior to the explosion of duel-card issue, domestic credit card market had experienced its peak period. 2005 set the historic record high of 45.49 million credit cards, which in 2006, dropped to a circulation of 38.32 million due to the impact of credit card bad debt. Since its official emergence in September 2005, total write-off of bad debt by various credit card issuing banks amounted to NT$13.4 billions. According to statistics released by the Financial Supervisory Commission, Executive Yuan in February 2006, sum total of the circulation interests of credit card and the granted loan balance of cash card had climbed to NT$76.49 billions, with 520 thousand overdue card holders, averaging default payment was NT$300 thousands per capita.

Oversupply of structured credit products and over expansion of credit have planted the root cause for the financial tsunami, created the most devastating global-scale financial crisis in nearly 50 years. Such incident makes it imperative that the banking industry should reexamine the way they judge and review the applicants' credits. Excessive issuance and overdue payments of credit cards have caused grave economic problems. Excessive use

---

* Corresponding author.
E-mail address: scchen@chinyi.ncut.edu.tw (S.C. Chen).

by card holders is not the sole cause for credit card problem. A more serious cause is the simplified credit rating among other reviewing processes that the bank used to win over customers in a competitive environment. This has created tens of billions of bad debt by the excessive consumption of insolvent card holders. Accordingly, financial banks, in conducting credit granting, should adopt a set of standards, stringent reviewing mechanism, and on the basis of revenues, try to make the right selection, to minimize the occurrence of bad debt, and to enhance the management performance of card issuance banks. A majority of previous literature focused on constructing credit card classification or behavior classification model with high accuracy, without taking into account the Model I and Model II errors resulted from misclassification. Here, Model I error refers to the misjudgment of applicants with good credit for high-risk group. Conversely, Model II error happens when applicants with bad credit are misclassified as low-risk group. As shown in the following Table 1:

In this study, the researcher intends to use the CRISP-DM 6-step cycle of improvement procedure to identify the influential factors causing default discrimination control in the reviewing process. Furthermore, by applying artificial neural network (ANN) and rule of decision tree, to cut down misjudged credit reviewing that cause bad debt resulted from credit expansion, and hopefully, to establish a set of relevant rules that can effectively eliminate those errors.

2. Definition of research model

To cope with the changing environment, many enterprises facing with the surging tide of IT development, expect to benefit from it by gaining some competitive edges. Nevertheless, upon introduction of IT system, they soon find themselves incapable of uncover the wealth of information stored in the huge databank. Thus data mining technique has become a scientific skill to excavate the knowledge and patterns concealed in the diversely complex mountain of data. In defining "data mining", Cabena (1998) explained that data mining is the process of effective accessing and extracting a large volume of information previous unknown, and provide the extracted information to his/her superiors for final decision-making. Berry et al. (1999) pointed out in their study that "data mining" is analyzing and finding meaningful relations or rules from a great amount of data in an automatic or semi-auto manner. Frawley, Piatetsky-Shapiro, and Matheus (1991) instead interpret "data mining" as the process of excavating from databank the non-apparent, implied, unprecedented, and yet may possibly be useful information. Grupe and Owrang (1995) regarded "data mining" as the act of dissecting facts and discover the new relations unknown to experts from the existing information. Hall, Mani, and Barr (1996), however, defined "data mining" as hunting/grabbing knowledge presented in a regular manner or other modes from the sea of data, by combining multiple techniques, such as data visualization, machine learning, statistics, and data warehousing.

In defining the research steps, the study tries to integrate the CRISP-DM 6-step cycle and the DMAIC 5-step data mining process, as shown in Fig. 1:
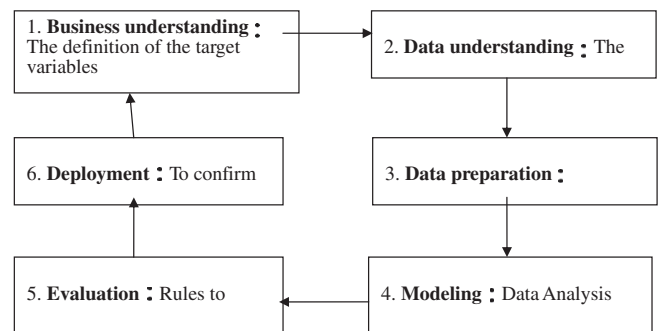
Table 1
Type I error and Type II error.

Actual class	Classified class	
	Good credit	Bad credit
Good credit	Accept	Type I error
Bad credit	Type II error	Reject

Fig. 1. CRISP-DM 6 steps in cycling and conform of DMAIC chart.

From Fig. 1, we find:

1. Commercial understanding and defining target variables: In this stage, we can have a clear understanding of the project's goal and need, to achieve the purpose and definition of data mining problems, and perfectly devise an appropriate plan and procedure;
2. Data understanding and defining conditional variables: In this stage, we must have initial collection and understanding of data, evaluate the quality of data, and propose hypotheses for the portions with possibly hidden problems. It is important that we understand the characteristics of data. For instance, in conducting correlation analysis, it is important to explore the positive, negative, or non-correlation between credit risk and consumer's occupation, position, seniority, and income.
3. Data preparation and measurement: In this stage, the raw data construct must be transformed into a final, useful data form, that is, convert them into an information format that can be constructed by instrumental software package. To achieve this purpose, data may be put through several conversions and compilations, depending on the construction need. The process includes table, record, selection of property and modeling, and changing and cleaning of the data fed into the tool kit.
4. Model designing and data analysis: In this stage, we have various models to choose and utilize from. Some of the models, however, may demand strict requirements for specific data forms, e.g., sequential, scattered data, data loss etc. Accordingly, we are often required to move back to the last stage to recheck and recompile the data.
5. Assessment and improvement of credit rules: In this stage, we built one or several models, and analyze those data. Before actual applying the model (s), an overall reviewing is needed to see whether there is any excluded important commercial considerations, in order to ensure that the model (s) meets the initial professional requirements. With this process, we can induce a set of practicable credit improvement rules.
6. Confirmation and Control: To confirm the application effect of the conclusion, to conduct continuous improvement based on the correctness/incorrectness of the responsive model, and to make mutual verification between the acquired knowledge and commercial understandings.

3. Defining target variables and input of variables

Data used in this study were collected from the consumers' basic credit checking data and post-checking consumption records of a certain bank. According to the credit card databank of the case bank, a total of 266,083 credit cards were then in circulation, among them, 123,592 were valid. Number of card issued in that month was 307, with an average monthly account of NT$ 2398,