Chapter 16

Boosting Prediction Accuracy of Bad Payments in Financial Credit Applications

Russel Pears
Auckland University of Technology, New Zealand

Raymond Oetama
Auckland University of Technology, New Zealand

ABSTRACT

Credit scoring is a tool commonly employed by lenders in credit risk management. However credit scoring methods are prone to error. Failures from credit scoring result in granting loans to high risk customers, thus significantly increasing the incidence of overdue payments, or in the worst case, customers defaulting on the loan altogether. In this research the authors use a machine learning approach to improve the identification of such customers. However, identifying such bad customers is not a trivial task as they form the minority of customers and standard machine learning algorithms have difficulty in learning accurate models on such imbalanced datasets. They propose a novel approach based on a data segmentation strategy that progressively partitions the original data set into segments where bad customers form the majority. These segments, known as Majority Bad Payment Segments (MBPS) are then used to train machine learning classifiers such as Logistic Regression, C4.5, and Bayesian Network to identify high risk customers in advance. The authors compare their approach to the traditional approach of under sampling the majority class of good customers using a variety of metrics such as Hit Rate, Coverage and the Area under the Curve (AUC) metrics which have been designed to evaluate classification performance on imbalanced data sets. The results of our experimentation showed that the MBPS generally outperformed the under sampling method on all of these measures. Although MBPS has been used in this research in the context of a financial credit application, the technique is a generic one and can be used in any application domain that involves imbalanced data.

DOI: 10.4018/978-1-60566-754-6.ch016
INTRODUCTION

In today’s market place customers typically utilize credit to purchase a variety of consumer goods and automobiles. While credit terms and repayment periods vary the basic mechanism of evaluating the credit worthiness of customers follows a well-defined framework. Such a framework helps to assess the probability that the loan will be repaid in full by the customer at a future point in time.

Credit worthiness is usually assessed by five different categories of criteria, the first of which assesses customer characteristics. This is used to get a general idea of customer demographics. The second category is customer capacity to repay the loan. Customer capacities typically refer to the monthly surplus once all expenses have been met. The third category is collateral, which are valuable assets that can be pledged as security. The next category is customer capital, which includes individual investments, insurances, etc. The last category is condition, which covers other related situational facts such as market condition, social condition, etc.

Data on the above criteria is captured in an application form and may be assessed by a human credit analyst. However, due to rapid business expansion of credit products such as consumer credits, property mortgages, etc, the manual approval process tends to overwhelm credit analysts with too many credit applications (Abdou, Masry, & Pointon, 2007). Crook et al. (2006) shows that between 1970 and 2005, consumer credit outstanding balance in the US grew by 231% with a dramatic growth of 705% on property mortgages. As a result of the massive volumes involved, the manual credit analysis process is enhanced through the use of statistical methods (Servigny & Renault, 2004). A typical statistical approval method is credit scoring. Credit Scoring is defined as a set of tools that help to determine prospects for loan approval (Johnson, 2006).

After the credit application has been approved, lenders inform customers that their credit applications have been granted. This will generally lead to a customer signing a contract. On the contract, a payment schedule informs the customer of the amount and due date of payments on which the customer must repay the lender.

The majority of customers make their payments on schedule, but some customers do make late payments. Payments that are paid after the due date are called overdue payments. Collecting overdue payments may not be easy, depending on the willingness of customers to pay. If customers still want to pay their overdue payments, lenders may devise special schemes to facilitate loan repayment for such customers. In other cases, customers simply refuse to make their payments. As a result, such customers create collection problems. Overdue payments occur because credit scoring fails to filter all of the bad customers. We identify two related problems that will be addressed in this research. Firstly, the credit scoring process is imperfect causing overdue payments. Secondly, overdue payments directly give rise to payment collection problems.

The objective of this research is to provide solutions to both credit scoring and collection problems. The proposed solution is essentially a payment prediction of overdue payments at the next payment round in a bid to find potential overdue payments in advance. As a result, proactive action can be taken to pre-empt overdue payment. Payment prediction models built by classifiers show combinations of credit scoring parameters that characterize overdue payments. Such information can also be utilized to improve the current credit scoring method used.

We utilize classifiers based on machine learning approaches such as Logistic Regression, C4.5 and the Bayesian Network to classify customers into two categories, good customers who are predicted to make their payments on time and bad customers who either default on the loan or make late payments. The overdue payment prediction process is complicated by the fact that bad customers form a small but significant minority thus
Related Content

A Study of the Relationship between Freshman Composition and Student Performance in Intensive Writing Courses
Thomas K. Martin (2012). Teaching Cases Collection (pp. 375-396).
www.igi-global.com/chapter/study-relationship-between-freshman-composition/60862?camid=4v1a

Fuzzy Neural Network Models for Knowledge Discovery
www.igi-global.com/chapter/fuzzy-neural-network-models-knowledge/24214?camid=4v1a

Support Vector Machines for Business Applications
Brian C. Lovell and Christian J. Walder (2008). Mathematical Methods for Knowledge Discovery and Data Mining (pp. 82-100).
www.igi-global.com/chapter/support-vector-machines-business-applications/26134?camid=4v1a

Exploring Calendar-Based Pattern Mining in Data Streams
Rodrigo Salvador Monteiro, Geraldo Zimbã§, Holger Schwarz, Bernhard Mitschang and Jano Moreira de Souza (2010). Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications (pp. 342-360).
www.igi-global.com/chapter/exploring-calendar-based-pattern-mining/39600?camid=4v1a