Information Technology and Quantitative Management (ITQM2015)

Customers’ Risk Type Prediction Based on AnalogComplexing

Changzheng He, Bing Zhu, Mingzhu Zhang, Yuanyuan Zhuang*, Xiaoli He, Dongyue Du

Business School, Sichuan University, Chengdu 610065, P.R. China

Abstract

Credit card holder’s behaviour may change over time, which would lead to the change of risk type. This paper introduces a new method, AnalogComplexing (AC), to predict consumer’s risk type. Furthermore, the new method uses the observed historical process itself for prediction which does not need any information of input variables in unknown prediction period. The authors applied the new proposed method AC to a bank customer dataset from one city of Western China, and the empirical study shows that AC is significantly better than widely-used neural network in terms of prediction accuracy. The empirical results in indicates that AC is an effective method, and provide a new way to predict consumers’ risk type.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Consumer, Risk type, Prediction, AnalogComplexing

1. Introduction

One segment of credit cardholders, revolvers, has attracted many attentions [1-3] in the field of credit risk management. Though customers with a high risk of default are included in this segment, the most profits of credit card also come from them [4]. Revolvers contain a kind of customers who have delinquent records. Moreover, this kind of cardholders can be further roughly divided into two groups. For some users, the ones who do not meet minimum payment requirement for a long time (more than three successive months) are defined as high-risk group[4,5]; for the other users, they just delinquent occasionally because the factors like they forget to repay and will eventually pay off their debts, which can be defined as low-risk group[6,20].

In order to identify consumers’ risk type, several credit risk assessment models exist. Logistic regression and NN have been used extensively in marketing to solve credit risk problems [8-15]. Dreiseitl et al. [21] found logistic regression and neural network perform on about the same level more often than not, with the more flexible neural networks generally outperforming logistic regression in the other cases. To classify the
consumer’s risk type statically, these methods worked very well. What the results will be if take the dynamic of consumers’ risk type into consideration? Can we apply the methods to solve this problem? Unfortunately, when we do not know the value of inputs during the prediction period, the above methods need pre-estimate the value of input variables and the estimate values are used to predict the risk type. Through the two-stage prediction, more bias of the final results may generate.

In this article, we introduce the AnalogComplexing (AC for short) model which is a sequential pattern recognition method for predicting[16] into this area. This method is based on the assumption that typical situations of a time process will repeat in some form. That is, each actual period of state development of a given multidimensional time process may have one or more analogous periods in history. The main characteristic of AC is that forecasts are not calculated in the classical sense but selected from the table of observational data. That is, the observed process itself is used for forecasting.

Although AC has received increasing attention in various fields (e.g., for the meteorological forecasts, see [17]; for economy forecasts, see [18]), to the best of our knowledge, credit risk type literature does not contain any reference to such model. Therefore, we attempt to fill this gap by empirically investigating whether AC can challenge more traditional risk type forecasting models. In particular, we examine their performance in predicting credit card customers’ risk type for a bank.

To solve this problem, unlike logistic regression and neural network, AC does not need any information in prediction period when predicting the consumer credit risk type. Therefore, the bias generated by two-stage prediction can be eliminated. And unlike other literatures using the credit history records, personal and card information [6, 7] to establish the model, AC tries to only consider the information of consumer’s behaviour. Some variables of personal and card information, such as income and credit limit, may change over time but it is hard for us to obtain this change. These data limitation may omit bias in the modelling process. AC just uses the historical spending and repayment information which is recorded by the credit card issuer itself, so the limitation of the data can be eliminated.

This paper is structured as follows. First, the next section contains a description of the model. Then the three subsequent sections discuss and research design issues. In the ensuing section, results are presented and discussed. Last, we conclude the study.

2. Methodology of consumers’ risk type model

In this paper, we introduce AC model to predict the consumers’ risk type, and compare its prediction ability with artificial neural network (ANN). There are three benchmark models based on ANN. These benchmark models are (1) the one uses a customer’s historical information to train the model and actual value of input variables in forecasting period to predict the model (Benchmark Model 1); (2) a model uses a customer’s historical information to train the model and estimation value of input variables to predict the model (Benchmark Model 2); (3) a model brings other customer’s historical information into the model and uses estimation value of input variables to predict the model (Benchmark Model 3).

2.1. The AC model steps in predicting consumers’ risk type

One of the main assumptions of AC is that it is likely that any behaviour of process will repeat similarly over time. If we succeed in finding one or more sections of past behaviour (analogous pattern) which are analogous to the most recent behaviour trajectory of the process (reference pattern), the prediction can be achieved by combining the known continuations of the analogous patterns to a continuation of the reference pattern. There is a four-step-procedure for AC: (1) generation of alternate patterns; (2) transformation of analogues; (3) selection of similar patterns (pattern similarity measures and selection); (4) combining forecasts. More details about this method can be referred to [18].
2.2. Artificial Neural Network (ANN)

The Benchmark Model 1 uses the observations from June 2009 to November 2010 of a consumer \( c \) to train the model and the actual input variables’ value from December 2010 to May 2011 of him or her to predict the model. Sliding window is also used. Each six rows (months) is defined as a pattern. One pattern corresponds to one class label.

For the other two benchmark models, we assume that the inputs from December 2010 to May 2011 of all consumers are not known. Therefore, we estimate the value of 4 input variables of the six months (from December 2010 to May 2011) using Autoregressive Integrated Moving Average Model (ARMA) firstly. Then, we transform the 4 variables of the 6 months into 24 variables.

For the Benchmark Model 2, we just use the observations from June 2009 to November 2010 of consumer to train the model. The estimation 24 variables (from December 2010 to May 2011) are used to predict the risk type of the next six months.

The difference between the Benchmark Model 3 and the Benchmark Model 2 is that the third one brings other samples into the model, which is similar to AC. Thus we use the observations from June 2009 to November 2010 of the calibration samples and consumer \( c \) in the holdout validation set to generate all the patterns. All the patterns are used to train the ANN model. The estimation variables (from December 2010 to May 2011) are also used to predict the risk type of the next six months.

3. Empirical study and discussions

3.1. Empirical study setting

We used a data set that a bank in a city of western China provided. Our samples include information from credit applications and monthly statements of 280 special cardholders (who have delinquency records) from June 2009 to May 2011. The data set records cardholder delinquency, total balance, repayment histories, purchases and cash advances, and credit limits. In addition, the data set provides information on consumers’ demographics, such as income, educational level and residential status. We used 80% of the sample for the estimation and the remaining 20% as a holdout sample for validation purposes. That is, the calibration sample contains 224 consumers, and the validation sample includes 56 consumers.

The variable we attempt to predict is whether a credit card holder is high risk in the next six months after the sampling date. We coded the risk type as a dummy variable, where \( y = 0 \) if the consumers who are delinquent for three successive months in the six months and 1 if otherwise.

We use the ratio of the spending and repayment variables as the inputs of models in this paper (see Table 1). We believe that ratio of the spending and repayment variables can level out the differences among the consumers better. Moreover, these ratios capture a customer’s financial risk status to some degree [20].

Table 1. The input variables of AC and ANN

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Risk type</td>
</tr>
<tr>
<td>X1</td>
<td>Actual payment/total balance</td>
</tr>
<tr>
<td>X2</td>
<td>Actual payment/ debit note</td>
</tr>
<tr>
<td>X3</td>
<td>Expense/total balance</td>
</tr>
<tr>
<td>X4</td>
<td>Cash advance/expense</td>
</tr>
</tbody>
</table>
As we noted previously, consumer’s risk type may change over time. Among 280 consumers, 24 consumers’ risk type change from high risk to low risk while 14 consumers’ risk type change from low risk to high risk. We find that when consumer is high risk, his or her number of trade and transaction amount is lower than when he or she becomes low-risk one. Without doubt, the available amount may be lower when he or she is high-risk one. Therefore, we calculate the transaction rate (transaction rate = transaction amount / available amount) of the consumer. The results show that when consumers are high-risk ones, the mean number of trade is 2.1 per month and mean transaction rate is 16% per month, while when they change to low risk the mean number of trade is 4.1 and mean transaction rate is 22%. The consumers from low risk to high risk show the similar pattern.

3.1. Prediction and model comparison

In Table 2, we present the holdout hits from the proposed model and the benchmark models. AC achieves 52 hits (a correct prediction of a cardholder’ risk type is considered a “hit”) on the holdout validation sample while Benchmark Model 1 gives the result as well as AC. However, the actual value (from December 2010 to May 2011) Benchmark Model 1 used cannot be obtained when doing prediction in practice. How about using ANN to do really prediction? Benchmark Model 2 and 3 achieve 42 and 46 hits on the holdout validation sample respectively. It is obvious that AC is much better than Benchmark Model 2 and 3.

Table 2 Number of correct cardholder’ risk type in the holdout sample(N = 56)

<table>
<thead>
<tr>
<th></th>
<th>hits</th>
<th>hit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>52</td>
<td>92.85%</td>
</tr>
<tr>
<td>Benchmark Model 1</td>
<td>52</td>
<td>92.85%</td>
</tr>
<tr>
<td>Benchmark Model 2</td>
<td>42**</td>
<td>75%</td>
</tr>
<tr>
<td>Benchmark Model 3</td>
<td>46*</td>
<td>82.14%</td>
</tr>
</tbody>
</table>

Note, ** indicate significant at 1%; * indicate significant at 10%

To examine whether AC significantly outperforms the ANN, the McNemar’s test is used [19]. This test is a nonparametric test for two related samples. We can find that the null hypothesis (equality of forecasting accuracy) is rejected at the 1% and 10% significance levels respectively. Thus we can confirm that AC is significantly better than ANN in predicting consumer’s risk type.

Except the whole hits, there are two types of errors. In one case, researchers may misclassify a high-risk customer as a low-risk customer. In another one, researchers may misclassify a low-risk customer as a high-risk customer. Type I and type II hit rates (as shown in Table 3) are the percentage of correctly predicted low risk and high risk respectively. It is obvious that AC has the highest and most balance Type I and type II hit rate among the three models.

Table 3 Type I and type II hit rate of the models

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>.91</td>
<td>.94</td>
</tr>
<tr>
<td>Benchmark Model 2</td>
<td>.55</td>
<td>.88</td>
</tr>
<tr>
<td>Benchmark Model 3</td>
<td>.73</td>
<td>.88</td>
</tr>
</tbody>
</table>

Why Benchmark Model 2 and 3 perform worse than AC? For ANN, we use ARMA to estimate the inputs.
Although we use several formulas to forecast the variables and choosing the best ones as inputs, the two-stage prediction may result the more bias of ANN.

4. Conclusions

The delinquency rates of consumers in the credit card market are consistently higher than those in the other parts of the loan market [4]. It is important for a credit card company to identify consumer segments (high risk or low risk) to develop targeted marketing strategies. We propose an AC model to predict consumer’s risk type in the subsequent months in the credit card market. In our modelling framework, we just use the variable of the spending and repayment in each month to identify low-risk and high-risk consumers. We successfully apply our proposed model to a data set that includes the inception of a consumer’s credit card history of monthly spending and repayment.

In marketing practice, our method can benefit the credit card issuer by foreseeing consumers’ risk type in the subsequent months. Thus, the managers can control the credit risk better by effective credit supply policies. Also, our experiment shows how the credit card issuer can use our approach to predict the consumer’s risk type in the subsequent months. Further research could concentrate on the behaviour characters of different risk type cardholders to help the credit card company make better marketing strategy.

References