

Fraud detection model & application for credit card acquiring business based on data mining technology

Tiebin LIU^{1, a}, Shiping LIU^{2, b}

¹ China's Securities Regulatory Commission, Beijing, 100033, China

² Financial Research Center, Chinese Academy of Science University, Beijing, 100049, China

^aemail: liutb@csrc.gov.cn, ^bemail: shiping_liu@gbicc.net

Keywords: data mining; credit card acquiring; anti-fraud

Abstract. Large quantities of fake card transactions bring about huge risks to commercial banks' credit acquiring business, this paper uses data mining technology to build credit card acquiring fraud analysis model based on mass credit card transaction data and merchant materials, and also developed merchant fraud risk management system. The application of this system effectively reduces the frequency of fraudulent credit card transactions, and helps to minimize losses from merchant's fraud.

Introduction

With the substantial increase in bank acquiring transactions, credit card frauds become ever more rampant. Backwardness in merchant's risk management is becoming one of the biggest obstacles for the development and profit generation of credit card business. To reduce the losses brought by merchant risks, improve the automatic, scientific, pragmatic level in merchant risk management, it's urgently needed to introduce data mining technology to identify highly risky fraud card transactions in time, and carry out effective surveillance and management on merchant fraud risks.

The goal of credit card acquiring bank risk surveillance system is to carry out in-depth analysis on historical transaction information with advance data mining technology, find out the hidden knowledge and rules, develop a model for transaction fraud risk evaluation, identify fake and fraudulent credit cards, pinpoint transactions of fake and fraudulent credit cards to predict the probability of the credit card transaction being fraudulent, and provide scientific basis for intelligent anti-fraud strategy. For details, the system must be able to:

- Analyze massive credit card transaction data, establish fake card fraud identification model, reduce the time from occurrence of fraud to its identification & handling, and minimize losses caused by fake cards;
- Continuously track, study the new credit card fraud mode, implement rapid and flexible new anti-fraud strategy.
- Achieve management on merchant relationships, carry out risk management on merchant's basic situations by managing merchant material as well as recording merchant's tracking & feedback situation.
- Reflect statistics information of merchant's transaction and gain promptly with statistics reports, and reflect merchant's transactions intuitively with statistics reports and graphs.

This paper introduces the establishment process of credit card acquiring fraud analysis model, including business analysis, data research, data adjustment, modelling, model verifications etc., and finally gives advice on the model application.

Build fraud analysis model

Merchant fraud features analysis:

Main features of merchant fraud include:

1. Merchant size: The merchant size here is mainly reflected in two aspects: nature of the merchant company and operational floor area. Generally speaking, large merchants with

- standard management are comparatively safe, while the probabilities of fraud in smaller, newly established merchants are bigger;
2. Whether the merchant has intention of fraud: Generally speaking, small, private merchants who sell antiques, ginseng & medicine etc. have bigger intention of fraud;
 3. Nature of operations: The probabilities of fraud at luxury consumption sites are bigger;
 4. Transaction time: The probabilities of fraud in non-business hours are bigger;
 5. Transaction amount: The probabilities of fraud in large, integral transaction amount are bigger;
 6. Transaction goods: The probabilities of fraud in goods that are small in size, high in price, and can easily turned into cash are bigger;
 7. Transaction frequency: The probabilities of fraud in frequent transactions in short period of time with large dollar value are bigger;
 8. Successive transactions on cards from different countries: The probabilities of fraud in successive transactions on cards from different countries in short period of time with large dollar value are bigger;
 9. Has failed transaction record: The probabilities of fraud in transactions on the same card after a certain transaction has failed.

Decide data mining target

The “credit card fraud risk” in this project means the risk that the credit card used in overseas credit card transactions being a fake card within the bank acquiring merchant’s scope. And it’s defined: “ (1) Fake card risk specifically means overseas fake card risk; (2) All the bank’s acquiring merchant’s transactions on credit cards whose reason for transaction verification in the overseas credit card transaction verification record is “fake card transaction verification” are defined as “fake card transactions” .

The goal of data mining is to predict the probability of one credit card transaction being a fake transaction.

Data preparation

The modelling of data mining divides the data of overseas credit card transactions into 4 categories: part I is transaction data, part II is transaction data extracted from variables, part III is data of merchant’s material, part IV is target variable table. The transaction data mainly include: overseas credit card transaction data, indirectly linked transaction data, indirectly linked transaction data; Merchant’s material used in modelling include: special merchant table, overseas credit card merchant material table and EDC-VISA material, the contents include: terminal No., zip code, merchant type (MCC code) etc.; target variables include all card numbers of fake and fraudulent cards appeared in transaction verification table, all transactions on these cards in the transaction records are considered as target variables.

On the basis of technical data preparation, 18 initial variables are derived from the original data, including: card organization, transaction code, industry, region, number of successful card transactions on current day, number of successful card transactions in 3 days, number of failed card transactions on current day, card swiping time (in hours), card expiration date, transaction amount, total transaction amount on current day, average transaction amount on current day, maximum transaction amount on current day, minimum transaction amount on current day, total transaction amount in 3 days, average transaction amount in 3 days, maximum transaction amount in 3 days, minimum transaction amount in 3 days

The next step is to standardize the index itself, merchant, and MCC code industry for each continuous variable. Standardization means erasing influence of the dimensions of original variables through certain mathematical transformation. Here we erase influence of the dimensions by normal standardization (0,1) of the indexes.

Data level regression and IV conversion are done on discrete data. The advantage of IV (Information Value) conversion is that missing values or the value of 0 can be treated as a major category, without affecting other normal values. We can also figure out the correlation between this variable and target variable from the information value.

To have an objective assessment on the model, it's needed to divide the modelling data into training set and measuring set. We took transaction data in 2005 as training set and transaction data in the first half of 2006 as measuring set. The number of observations in training set is 793,593, in which 793,061 samples are good, 532 samples are bad. The number of observations in measuring set is 315,728, in which 315,502 samples are good, 226 samples are bad.

Since the bad sample concentration of target variables in the original samples is 0.07%, so oversampling technology is used to raise the bad sample concentration. Sampling can only be done in the training set, take all bad samples, and take certain percent of good samples. The oversampling ratio is 1: 200, there were 4,644 observations after oversampling, in which 4,112 are good, 532 samples are bad.

Build the model

Credit card fraud belongs to classification model, whose common algorithms include Logistics regression, neural network, decision tree. This project used three methods: Logistics regression, neural network, decision tree to build the model respectively, then choose a best model after comparing the modelling results. To assess the models objectively, we build the models with transaction data in previous year, and evaluated with transaction data in the second year.

To compare the advantage of these three methods, we build models with these three methods with identical data set, then compare the three methods at assessment nodes. Three classification models: decision tree, neural network and logistics regress are compared below, the comparison result is shown in the figure below:

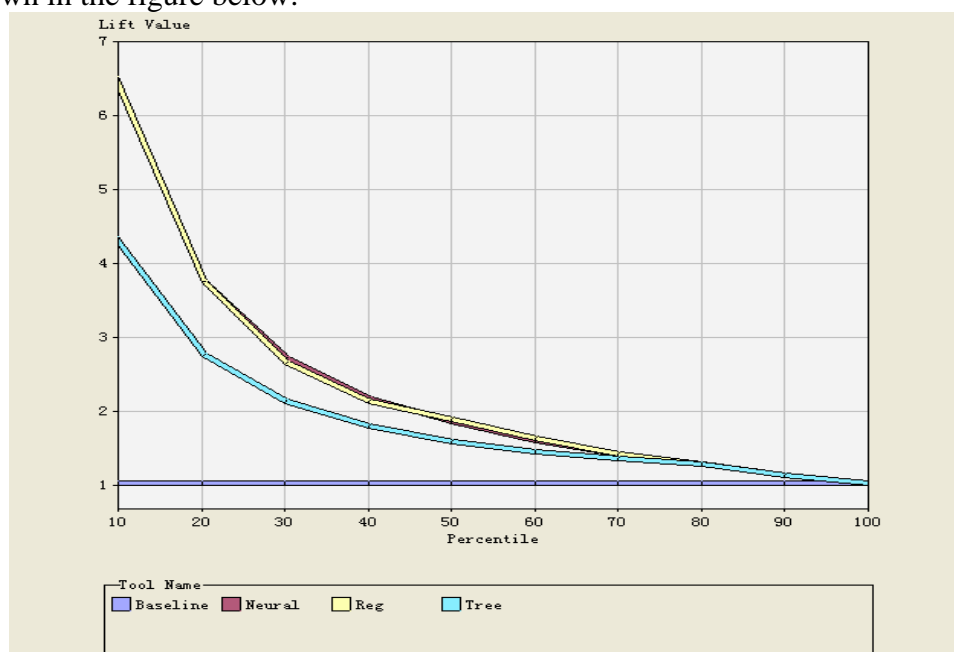


Figure 1- model lift chart

This figure is the model lifting comparison figure, representing the multiple of improved capture capability on each digit with and without modelling. The comparisons above show that the results of logistic regression and neural network are apparently better than decision tree, while the results of logistic regression and neural network are quite similar.

Modelling result

Previous tries show that the results of logistic regression model and neural network are apparently better than decision tree, while the results of logistic regression and neural network are quite similar. Because logistic regression can be better explained, so logistic regression is chosen as the final modelling method. By continuous fine tuning, the modelling result we finally have is shown in the table below:

$$\ln\left(\frac{P}{1-P}\right) = 1.88 + 4.76X_1 + 7.34X_2 + 3.74X_3 + 8.55X_4 + 7.44X_5 - 0.34X_6 + 0.11X_7 - 0.24X_8$$

Table 1- Meanings of modelling variants

No.	Index code	Index name
1	intercept	Intercept: target=0
X ₁	card_flag	Card type IV value
X ₂	industry	Industry IV value
X ₃	count3d	Number of transactions in three days IV value
X ₄	ft	Number of failed transactions on current day IV value
X ₅	exp_year2	Card expiration date
X ₆	sum3d_ms	Merchant's standardization of total transaction amount in 3 days
X ₇	max3d_ms	Merchant's standardization of maximum transaction amount in 3 days
X ₈	sum_mccs	MCC code standardization of merchant's transaction amount on current day

Advice on model application

The overseas credit card fraud model is established on the basis of sufficient understanding of credit card acquiring business, good predicting results are achieved when the model is verified with test data. Credit card risk control staff can take the modelling result as a reference for evaluating the risk degree of certain transaction. When evaluating the risk degree of certain transaction listed by the system according to rules of thumb, risk control staff can assess this transaction by taking the overseas credit card fraud model as a reference to make reasonable decisions. If a transaction complies with the preset rule of thumb, and is also listed in the transaction list provided by overseas credit card fraud model, then this the probability of this transaction being a fraud is very high, which needs special attention from risk control staff.

Data mining models do not remain unchanged after they are developed, with the passage of time, there may be changes in exterior environment, the technique, mode of fraud may also change, all these may reduce the applicability of the model, thus, continuous tracking is needed after the model is launched, and the model should be adjusted based on its feedback, this is a constant ongoing procedure.

Conclusion

Based on data mining technology, this paper achieved bank credit card acquiring risk surveillance system. The application of the system effectively improved bank acquiring organizations' capability on data development & use as well as fraud risk control, it can be widely used to help acquiring organizations to identify fraud risks, prevent fraudulent risks from merchants who lack credibility, and improved the risk management capability of financial institutions. The analysis system and analysis method brought forward in this paper are also generally applicable to insurance, securities, future, funds etc.

Reference:

- [1] Liu Shiping. Data mining technology and application [M]. Beijing: High Education Press, 2010.
- [102] Perols, Johan L., Detecting financial statement fraud: Three essays on fraud predictors,

multi-classifier combination and fraud detection using data mining[D]. University of South Florida ,2008.

[3] Wu X, Zhu X, Wu G Q, et al. Data mining with big data[J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(1):97-107.

[4] Yue D, Wu X, Wang Y, et al. A Review of Data Mining-based Financial Fraud Detection Research[C]// International Conference on Wireless Communications, NETWORKING and Mobile Computing. 2007:5519-5522.

[5] Malone J, McGarry K, Wermter S, et al. Data mining using rule extraction from Kohonen self-organising maps[J]. Neural Computing & Applications, 2006, 15(1):9-17.