

A MONEY LAUNDERING RISK EVALUATION METHOD BASED ON DECISION TREE

SU-NAN WANG^{1,2}, JIAN-GANG YANG¹

¹College of Computer Science and Engineering, Zhejiang University, Hangzhou 310027, China

²Shanghai Pudong Development Bank, Shanghai 200002, China

E-MAIL: wangsn@spdb.com.cn, yangjg@cs.zju.edu.cn

Abstract:

Money laundering (ML) involves moving illicit funds, which may be linked to drug trafficking or organized crime, through a series of transactions or accounts to disguise origin or ownership. China is facing severe challenge on money laundering with an estimated 200 billion RMB laundered annually. Decision tree method is used in this paper to create the determination rules of the money laundering risk by customer profiles of a commercial bank in China. A sample of twenty-eight customers with four attributes is used to induce and validate a decision tree method. The result indicates the effectiveness of decision tree in generating AML rules from companies' customer profiles. The anti-money laundering system in small and middle commercial bank in China is highly needed.

Key Words:

Anti-money laundering; Decision tree; Commercial bank

1. Introduction

Criminal activities, drug trafficking, smuggling, bribing and so on, can be highly profitable. Money generated by illegal activities must be made to look legitimate before it can be freely spent. Otherwise, it may be forfeited by the government. Money laundering is a process that takes illegally obtained or dirty money and puts it through a cycle of transactions or through various accounts in one bank or between banks. The cycling of the money makes the money appear to be from legitimate sources and the money cannot be traced back to its illegitimate source. Hiding legitimately acquired money to avoid taxation also qualifies as money laundering.

In 2005, China Anti-Money Laundering Monitoring & Analysis Center received 283,400 shares of the RMB suspicious transaction reports, and 1,988,900 shares of the foreign currency suspicious transaction reports related to 137.8 billions of RMB and more than one billion of US dollar in 4926 accounts [1]. The major part suspicious transactions came from the state-owned commercial banks

and the joint-stock commercial banks. Therefore commercial banks are facing severe challenge on money laundering in China today.

The developed countries already established some advanced online monitor systems for anti-money laundering (AML). For example, American Financial crime enforcement network Artificial Intelligence System (FAIS) [2] integrated intelligent human and software agents to identify potential money laundering on a very large data space. Artificial intelligence computer analysis system can greatly enhance the work efficiency and is an essential method for AML. However, the computer based AML technologies have not been used in Chinese commercial banks. An AML computer automatic monitor system is urgently needed. It is not appropriate that an AML artificial intelligence system from developed countries directly applied on Chinese immature financial market. Therefore, an artificial intelligence AML system must be established according to the characteristic of Chinese financial market. The researches of the computer technology for AML just started in China in recent years. To our best knowledge, we are the first one to apply artificial intelligence method into the AML domain in China.

Decision tree learning [3] is one of the most widely used methods for inductive inference since the 1960s. Since then, numerous researches have been conducted to improve the accuracy, performance and so on. ID3 [4] is considered as the milestone in decision trees. A decision tree can be viewed as a partitioning of the instance space. Each partition, called a leaf, represents a number of similar instances that belong to the same class. The split points of ID3 are chosen according to the most informative attributes of the data instances. Rules are thus can be extracted from the root to some leaf node. Each path from root to node provides one classification rules of the examples. In this paper, the risk rank is used in to determine the possibility that the customer launder money use the bank products and services.

2. Our approach of a decision tree method for the customer assessment

“Know you customer” program in banks asked by China Banking Regulatory Commission (CBRC) is trying to thwart the money launderer who launders dirty money using bank products and services. Unlike the most current researches which more focus on identifying the suspicious transaction directly, we divide the process into several steps. In this paper, we are more focus on the first step of AML which evaluates the potential money laundering risks of each customer. For instance, X bank is a midsize commercial bank in China with about 3 million individual and company accounts. The different categories of customers have different inherent possibility to carry on money laundering activities. For example, oversea payment of USD 10 millions regard as a normal fund transaction of a large-scale petroleum & chemical corporation to purchase crude oils, but it can be regard as an extremely suspicious fund transaction of a small import-export company. Therefore, different customer must adopt different AML policies.

When a customer opens an account, the bank usually suppose money laundering risk rank for the customer, according to the customer materials and the history experiences. The ranks usually divided into low, middle and high kinds. Presently, all Chinese middle-scale commercial banks depend on account manager to determine the risk rank of a new customer. No uniform standard and software can be used to complete the evaluation of the risk rank of customers. It is frequently happened that the customer was initialized a risk rank which was not consistent with the facts. For example, a customer was initialized as a low risk rank, but has discovered a lot of suspicious transactions of cash and internet services in later. Therefore, a data mining method for the assessment of the bank customer money laundering risks is urgently needed.

2.1. The sample data

160 thousands customers are currently recorded in the data warehouse, with each contains 52 attributes. For the demonstration purpose, 28 examples randomly selected from the data warehouse, and only 4 attributes which are considered more relate to AML are selected. In order to protect business secret, the full title of the company customers is deleted from primary datasheet, and the beginning numbers and the last numbers of the customer identification (ID) are hided and written in type of ‘X784X’ for an example, as showed in table 1.

Table 1. The customer data with AML composite risk

ID	Industry	Location	Busiz	Prods	Risk
X530X	IT	Beijing	Large	All	M
X624X	Domestic trade	Beijing	Small	All	M
X983X	Agriculture	Harbin	Middle	S/I	L
X634X	Domestic trade	Jiangsu	Middle	All	M
X382X	Manufacture	Fujian	Middle	D	M
X054X	Agriculture	Shandong	Small	S/I	L
X801X	Manufacture	Hunan	Large	All	L
X669X	Foreign trade	Guangdong	Small	ALL	H
X903X	Retail	Guangdong	Small	D	H
X957X	IT	Sichuan	Middle	D	L
X547X	Trade	Xi'an	Middle	All	M
X228X	Telecom Trade	Guangdong	Middle	D/S	H
X318X	Auto sales	Shandong	Middle	D/S	M
X196X	Manufacture	Xi'an	Listed	D/S	M
X252X	Manufacture	Hebei	Small	All	L
X227X	Medicine	Shanghai	Middle	S/I	M
X284X	Manufacture	Jiangsu	Listed	D	M
X625X	Foreign Trade	Guangdong	Large	All	H
X613X	Hi-Tech/Trade	Shanghai	Small	D/S	H
X898X	Foreign Trade	Shanghai	Small	OS	H
X135X	IT/Trade	Sichuan	Middle	D/S	H
X818X	Manufacture	Beijing	Middle	All	L
X283X	Foreign Trade	Yunnan	Middle	D	H
X982X	Domestic trade	Guangxi	Middle	All	H
X784X	Manufacture	Henan	Listed	All	L
X043X	Trade	Guangdong	Small	D/S	H
X881X	Manufacture	Hebei	Small	D/S	L
X376X	Media	Beijing	Small	C/I	H

Busiz-Business size, Prods-Products from Bank,

Risk-AML composite risk,

D-Deposit, D/S-Deposit/Settlement, All-Total product,

S/I-Settlement/Internet Bank, C/I-Cash/Internet bank,

OS-Oversea Settlement, H-high, M-Middle, L-Low.

2.2. The primary factors for money laundering risk of customers

Based on the supposition that laundering illegal money activities may be found in the bank accounts of company customer, bank manager always determines the money laundering risk rank with the integrated factors. Whether the customer is managing in a region, in where money laundering activities are frequently occur, the customer is processing large cashes in daily management, the customer is managing on international trades. Which kinds of the bank products and services dose the customer use? How

about the business size and manage capacity, as so on? Western commercial bank usually determines the risk of the customer from three factors, such as industry, service area and used products from bank. The listed company and large enterprise are quite reliable for money laundering, but some small business size companies usually are suspicious. Therefore, the business size is also included as an attribute for the determination of a Chinese company's money laundering risk. Four inherent anti-money laundering risks of the customers are taken into account as below:

- 1) Business and entity risk. What is the industry of the customer?
- 2) Location. Where the customer is located?
- 3) Business size of the customer.
- 4) Products and transaction risk. Which kind of products or services is being offered to the customer?

Consequently every customer has four attributes and an AML composite risk mark as showed in table 1.

2.3. Uniform ruled data

In order to succinctly demonstrate the decision tree learning, four attributes listed in the previous section are respectively marked as three ranks, which are 'low', 'middle' and 'high' as listed in tables 2-5. Thus, table 1 can be transformed into table 6, an extremely succinct table.

Table 2. Money Laundering risk for industry

Industry	AML risk
Manufacture	Low
Chemical	Low
Domestic Trade	Middle
Medicine	Middle
IT	Middle
Foreign Trade	High
Retail	High
Advertisement	High
Automobile Sales	High

Table 3. Money laundering risk for locations

Location	AML risk
Bohai Sea Rim	Low
Northwest China	Low
Northeast China	Middle
Yangtze River Delta	Middle
Central China	Middle
Pearl river Delta	High

Table 4. Money Laundering risk for business size

Business size	AML risk
Listed	Low
Large	Low
Middle	Middle
Small	High

Table 5. Money laundering risk for bank products

Bank's products	AML risk
Deposit	Low
Deposit and settlement	Middle
Internet bank services	High
Cash services	High
Loan services	High
Oversea settlement services	High
All services	High

Table 6. The ruled sample

ID	Industry	Location	Busiz	Prods	Risk
X530X	M	M	L	H	L
X624X	M	L	H	H	M
X983X	L	M	M	H	L
X634X	M	M	M	H	M
X382X	L	H	M	L	M
X054X	L	L	H	L	L
X801X	L	L	L	H	L
X669X	H	H	H	H	H
X903X	H	H	H	L	H
X957X	M	L	M	L	L
X547X	H	L	M	H	H
X228X	H	H	M	M	H
X318X	H	M	M	M	M
X196X	L	L	L	M	L
X252X	L	L	H	H	L
X227X	M	M	M	M	M
X284X	M	M	L	L	L
X625X	H	H	L	H	L
X613X	H	M	H	M	H
X898X	H	M	H	H	H
X135X	H	L	M	M	L
X818X	H	H	M	H	H
X283X	L	L	M	L	L
X982X	M	H	M	H	M
X784X	L	M	L	H	L
X043X	H	H	H	M	H
X881X	L	L	H	M	L
X376X	H	L	H	H	H

L-Low, M-Middle, H-Hig

3. Results and discussions

Decision tree learning is used to induce a knowledge tree which can help to determine company's money laundering risk. Quinlan's ID3 algorithm^[4] for constructing a smallest decision tree that best fits the sample is used in this paper. A C++ program developed by the authors is carried out to build the decision tree. In order to evaluate the performance of the decision tree, first twenty-one customers are used as a training sample for decision tree learning. Then the validity of the trained decision tree is tested on the holdout sample of seven customers. A decision tree for the decision of the customer's money laundering risk is induced as showed in figure 1 with fifteen determination rules. It can be easily find that the decision tree results of the holdout sample are consistent with the bank manager's mark except one customer's record.

A few interesting observations can be pointed out from the results. Industry is the most primary attribute in the four customer's attributes which affect the bank customer's money laundering risk. Location and business size of the customer and the products supplied from bank are other three sequential important attributes. The massive data exported from a commercial bank customer information system are preprocessed. Some information related with the money laundering activity is chosen out from the customer specifications and named as the attributes for decision tree learning. These attributes are categorized and graded according to the possibility of the customer who has one of these attributes will engage in the money laundering activities. So, a new ruled database can be established. Decision tree learning as a data mining method is used to find the judgment rules which will be stored in a knowledge database. The mined rules may be used in the bank real-time transaction system to help to judge an illegal transaction. Account opening system may automatically determine the money laundering risk rank of customers.

4. Conclusions and future works

Decision tree learning is a comparatively powerful method for inductive inference. An attempt is made in this paper to examine the validity of using decision tree learning method to find judgment rules for customer money laundering risk determination. The results indicated that money laundering risk of a bank customer can be determined through a method as described in this article. Out of 160 thousand current customer profiles, based on the rules generated in this paper, 12% customers are considered as high AML risk and needed to be further monitored in their future transactions. In future, the rules extracted from

this project will be added to the bank system which will be used to categorize AML risk of the new customer for this bank.

This article simply shows one kind of method using a data mining technology for customer money laundering risk judgment. A commercial bank has magnanimous customer information data, which increase and update and correct everyday. Therefore, the authors' future work will focus on the development of a highly effective real-time system using a data mining method or another artificial intelligence method, or simultaneously use these methods.

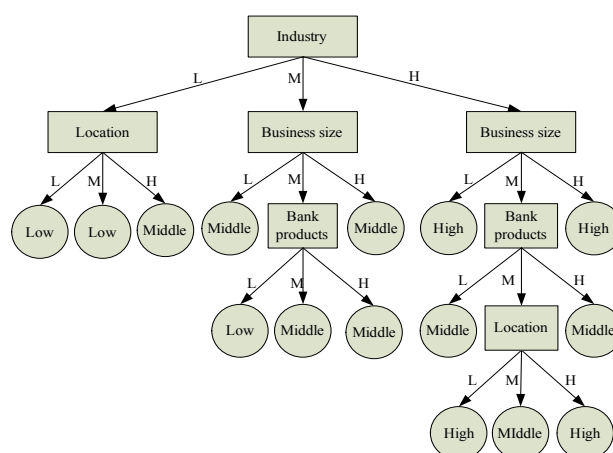


Figure 1. A decision tree for determination of customer's money laundering risk

Acknowledgements

The authors wish to acknowledge Ji Shen, a master graduate student in College of Computer Science and Engineering, Zhejiang University, for many useful discussions during the course of this study.

References

- [1]. The People's Bank of China, China anti-money laundering report 2005[R], China Financial Publishing House, Beijing, 2006.6. (In chinese)
- [2]. Ted E. Senator, Henry G. Goldberg, Jerry Wooton, etc., The financial crimes enforcement network AI system (FAIS) identifying potential money laundering from reports of large cash transactions[J], AI Magazine, Vol.16, No.4, pp. 21-39, Winter 1995.
- [3]. Safavin,S.R., Landgrebe,D. A survey of decision tree classifier methodology [J]. IEEE Transactions on Systems, Man and Cybernetics, Vol.21, No.3, pp.660-667, April 1991.
- [4]. Quinlan, J. R. Induction of decision trees, Machine Learning 1(1): 81-106, 1986.