A FOCUS ON DIFFERENT FRAUDS AND USING DATA MINING TO ENHANCE BUSINESS PROCESS IN BANKING SECTOR

B.MUNEENDRA NAYAK¹, NAVEEN KUMAR², R. MAHAMMAD SHAFI³
¹ RESEARCH SCHOLAR, DEPARTMENT OF CSE, UNIVERSITY OF ALLAHABAD, ALLAHABAD-211002.
² PROFESSOR, DEPARTMENT OF CSE, UNIVERSITY OF ALLAHABAD, ALLAHABAD-211002.
³ PROFESSOR & HEAD, DEPARTMENT OF MCA, SREE VIDYANIKETHAN ENGINEERING COLLEGE, TIRUPATI-517102.

ABSTRACT

Significant shifts in the business environment, economic volatility, changing customer and staff expectations, and the adoption of new technology make it increasingly challenging for banks to navigate technology strategy alternatives and prioritize technology investments. The banking industry around the world has undergone a tremendous change in the way business is conducted. Leading banks are using Data Mining (DM) tools for customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. This paper provides an overview of the concept of Data Mining and different frauds in Banking. Data might be one of the most valuable assets of any corporation, but only if it knows how to reveal valuable knowledge hidden in raw data. Data mining allows extracting diamonds of knowledge from the historical data, and predicting outcomes of future situations. It helps optimize business decisions, increase the value of each customer and communication, and improve customer satisfaction. Data mining is the process of extracting previously unknown information, typically in the form of patterns and associations, from large databases. Today, organizations are realizing the numerous advantages that come with data mining. It is a valuable tool, by identifying potentially useful information from the large amounts of data collected. An organization can gain a clear advantage over its competitors. The banking sector consists of public sector, private sector and foreign banks, apart from smaller regional and cooperative banks. In the market, various IT-based banking products, services and solutions are available. The most common of them are Phone Banking; ATM facility; Credit, Debit and Smart Cards; Internet Banking & Mobile Banking; SWIFT Network & INFINET Network; connectivity of bank branches to facilitate anywhere banking.

KEYWORDS: Fraud, Banking, Data Mining, Risk Management, Customer acquisition and management.

DATA MINING

Data Mining is the process of extracting knowledge hidden from large volumes of raw data. The knowledge must be new, not obvious, and one must be able to use it. Data mining has been defined as “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [1]. It is “the science of extracting useful information from large databases” [3]. Data mining is one of the tasks in the process of knowledge discovery from the database [5]. Fig. 1 shows the process of knowledge discovery. The steps involved in Knowledge discovery are [4,5]:

1. Data Selection: The data relevant to the analysis is decided and retrieved from the various data locations.
2. Data Preprocessing: In this stage the process of data cleaning and data integration is done.
3. Data Cleaning: It is also known as data cleansing; in this phase noise data and irrelevant data are removed from the collected data.
4. Data Integration: In this stage, multiple data sources, often heterogeneous, are combined in a common source.
5. Data Transformation: In this phase the selected data is transformed into forms appropriate for the mining procedure.
6. Data Mining: It is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used.
7. Interpretation and Evaluation: In this step, interesting patterns representing knowledge are identified based on given measures. The discovered knowledge is visually presented to the user. This essential step uses visualization techniques to help users understand.

INTERNAL FRAUD ANALYSIS - COMMONNESS OF CHAIN STRUCTURES

In collaboration with fraud experts, an example case study on real world data was defined and conducted. The results presented in this paper had to be made anonymous for confidentiality reasons. As the tabular view cannot be given, but is crucial for the assessment of structures, we try to partially compensate the information...
loss with more general and therefore non-critical explanations. In addition, it has to be stated that a detailed investigation includes further data sources and systems as for example the client contact history. The goal of this case study was to evaluate the following three aspects: Although transaction chains are known to play a major role in numerous analyzed fraud cases, the commonness of these structures in normal transaction behavior was widely unknown. By examining the frequency and context of transaction chains, their discriminative power in fraud detection may be estimated. How common are chains? Are there defined line-ups where chains emerge in normal, daily business? These and similar questions were considered in the case study.

Identification of a Reference Fraud Case
In particular, a reference case contained in the available data was proposed by fraud experts. The ChainFinder was configured without detailed knowledge of the case but common model knowledge. The ranking of the reference case in the ChainFinders suspiciousness scoring model (precision) and the portion of detected transactions being part of the case (recall) was evaluated.

Investigation Efficiency
This aspect evaluates the cooperation of the ChainFinder. The following questions were relevant: What is the average workload of investigating the produced alerts, in particular in comparison to existing fraud detection methods? May the focus on relational transaction structures instead of isolated transactions in combination with visualization improve investigation efficiency?

Data Basis and Configuration
The entire available dataset, representing a time window of two years, was scanned for chains of a dedicated transaction type, Manual Transaction Type (MTT). At this point, we will repeat the characteristics of MTT transactions relevant for internal fraud detection. Employees can trigger MTT transactions autonomously, that is, without customer order in written form. While being a valuable tool for un-bureaucratic and flexible customer service, MTT is prone to misuse. Above a defined threshold the additional control measures kick in. Therefore, transactions with an amount were excluded from this analysis. Below this threshold, fraud detection performs the task to impede misuse without decreasing the flexibility and value of the service. For this case study, the root set was defined by each customer featuring one or more outgoing MTT transactions. This definition is very general and was chosen with the intention to separate the evaluation of chain structures from other known model knowledge implemented in existing fraud filters. Furthermore, the scope of the search was limited to transactions between UCO Bank clients (no external transactions). However, chains leaving the bank (e.g. ending in a cash transaction or a MTT transaction to an external account) were retrieved using a second ChainFinder run where the initial structure based on internal chains suggested it. This approach appeared to be the best trade-off between performance and effectiveness for the rapid analysis of the entire data set given the limited resources for our study. A more focused search may incorporate external transactions from start. The most important configuration settings for the ChainFinder were defined as follows:

- Single chains of MTT transactions with an amount.
- Registration tracking enabled: single transactions forming a chain are registered by the same employee.
- Money remains at most seven days on intermediary accounts.
- Amount may differ within a chain up to 15%.

Figure-1: Real world data degree distribution between customers of UCO Bank
A simple scoring model based on the number of registered transaction chains was applied. A higher number of registered MTT transaction chains resulted in a higher interestingness score. For this case study, employees with a minimum of 5 registered chains were proposed for further analysis.

It has to be repeated that techniques involving other transaction types like document or signature forgery may also be used in internal fraud as good as in external fraud, possibly leading to chains and smurfing structures in a wide variety of transaction types. However, the inhibition threshold for these actions may be higher than for misusing MTT transactions, which are just one click away. The limitation to MTT transactions may therefore be reasonable for a starting point.

Commonness of Chain Structures
The overall results indicate the frequency of chain structures for MTT transactions: Figure-3 reveals that, before any application of exclusion criteria less than 2% of the MTT transactions in question are part of transaction chains. First insights motivated the definition of a number of exclusion criteria, which further reduced identified chains and led to a ratio of less than 1%. The average chain length is approximately 1.8.

This fact requires explanation. While the better part of the chains are of length 2, one transaction, in particular in high activity context, can repeatedly be accounted for multiple chains due to particle cloning. Figure-2 illustrates this effect.

Figure-2: A particle coming along transaction t will be cloned three times to travel along a, b& c. This results in 4 transactions that form 3 chains of length 2.

On the other hand, transaction chains up to length 6 were observed. The ratio of employees with at least 5 MTT-chains to all relevant employees which registered at least one MTT transaction is similar. From 10053 relevant employees, 100 were proposed for further analysis by the ChainFinder, leading to a ratio of 1%. Figure-4 shows the distribution of those employees according to the number of registered chains.

These results indicate that for the defined transactions of interest, transaction chains are rare. Of course, as long as it is unknown if any of the proposed employees is actually fraudulent, a statement on the discriminative power of chain structures in terms of fraud detection cannot be made. However, the low number of produced alerts makes investigation feasible and suggests potential under the observation that transaction chains are repeatedly present in analyzed fraud cases. Results were visualized and a selection was discussed with internal fraud experts. This led to the identification of previously unknown settings that produced a high number of chains but did not appear to emerge from fraudulent behavior.

Figure-3: Total number of relevant transactions and number of transactions in chains
Identification of a Reference Fraud Case

Table 1 shows the ten top scoring structures in this case study. Employee D turned out to be the perpetrator of the reference fraud case. The fact that the reference case is located in the “Top 5” of fraud scores on the overall data set may indicate a good precision. A discussion of this statement is given below. The evaluation also showed that the ChainFinder retrieved a substantially higher number of the transactions in the case in comparison with existing detection methods. With existing SQL-Filter, 4 transactions from this case triggered alerts. Manual analysis then showed that the whole fraud case actually consisted of more than 100 fraudulent transactions. In contrast, ChainFinder at first successfully identified approximately 60% of the fraudulent transactions and more than 90% after conducting the additional external chain search. It may be assumed, had the delinquent been only slightly more careful, the existing SQL-Filter wouldn’t have detected this quite extensive case but nevertheless it would have been found by the ChainFinder but this, of course, remains speculation.

It may be argued that the detection of a case basically consisting of chain structures is not a big accomplishment given the ChainFinder algorithm. We fully agree on this statement. However, as we decided on a pattern matching approach, our intention is not the identification of previously completely unknown structures, but to examine the potential of relational model knowledge in addition to the non-relational model knowledge in existing fraud detection monitors at UCO Bank. The fact that our algorithm was able to find the reference case is not surprising and goes without saying. The fact that it ranks among the Top 5 scoring structures even with a trivial scoring mechanism however shows the discriminative potential of this approach, which was previously unknown.

<table>
<thead>
<tr>
<th>Score (= number of chains)</th>
<th>Employee (anonymized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>241</td>
<td>A</td>
</tr>
<tr>
<td>54</td>
<td>B</td>
</tr>
<tr>
<td>40</td>
<td>C</td>
</tr>
<tr>
<td>38</td>
<td>D</td>
</tr>
<tr>
<td>30</td>
<td>E</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
</tr>
<tr>
<td>27</td>
<td>G</td>
</tr>
<tr>
<td>27</td>
<td>H</td>
</tr>
<tr>
<td>24</td>
<td>I</td>
</tr>
<tr>
<td>22</td>
<td>J</td>
</tr>
</tbody>
</table>

Table 1: The ten top scoring structures

DATA MINING IN THE BANKING INDUSTRY

The banking industry across the world has undergone tremendous changes in the way the business is conducted. With the recent implementation, greater acceptance and usage of “electronic” banking, the capturing of transactional data has become easier and, simultaneously, the volume of such data has grown considerably. It is beyond human capability to analyse this huge amount of raw data and to effectively transform the data into useful knowledge for the organisation. The enormous amount of data that banks have been collecting over the years can greatly influence the success of data mining efforts. By using data mining to analyse patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable. The banking industry is widely
recognizing the importance of the information it has about its customers. Undoubtedly, it has among the richest and largest pool of customer information, covering customer demographics, transactional data, credit cards usage pattern, and so on. As banking is in the service industry, the task of maintaining a strong and effective CRM is a critical issue. To do this, banks need to invest their resources to better understand their existing and prospective customers. By using suitable data mining tools, banks can subsequently offer ‘tailor-made’ products and services to those customers.

There are numerous areas in which data mining can be used in the banking industry, which include customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, cash management and forecasting operations, optimising stock portfolios, and ranking investments. In addition, banks may use data mining to identify their most profitable credit card customers or high-risk loan applicants. There is, therefore, a need to build an analytical capability to address the above-stated issues and data mining attempts to provide the answer.

Following are some examples of how the banking industry has been effectively utilizing data mining in these areas.

**Marketing:** One of the most widely used areas of data mining for the banking industry is marketing. The bank’s marketing department can use data mining to analyse customer databases and develop statistically sound profiles of individual customer preferences for products and services. By offering only those products and services that customers really want, banks can save substantial money on promotions and offerings that would otherwise be unprofitable. Bank marketers, therefore, need to focus on their customers by learning more about them. Bank of America, for instance, uses database marketing to improve customer service and increase profits. By consolidating five years of customer history records, the bank was able to market and sell targeted services to customers.

**Risk Management:** Data mining is widely used for risk management in the banking industry. Bank executives need to know whether the customers they are dealing with are reliable or not. Offering new customers credit cards, extending existing customers lines of credit, and approving loans can be risky decisions for banks if they do not know anything about their customers. Data mining, however, can be used to reduce the risk of banks that issue credit cards by determining those customers who are likely to default on their accounts. An example was reported in the press of a bank discovering that cardholders who withdrew money at casinos had higher rates of delinquency and bankruptcy. It is a common practice on the part of banks to analyse customers’ transaction behaviours in their deposit accounts to determine their probability of default in their loan accounts. Credit scoring, in fact, was one of the earliest financial risk management tools developed. Credit scoring can be valuable to lenders in the banking industry when making lending decisions. Lenders would not have expanded the number of loans they give out without having an accurate, objective, and controllable risk assessment tool. The examples of both a ‘good’ and ‘bad’ loan applicant’s histories can be used to develop a profile for a good and bad new loan applicant.

Data mining can also derive the credit behaviour of individual borrowers with instalment, mortgage and credit card loans, using characteristics such as credit history, length of employment and length of residency. A score is thus produced that allows a lender to evaluate the customer and decide whether the person is a good candidate for a loan, or if there is a high risk of default. Customers who have been with the bank for longer periods of time, remained in good standing, and have higher salaries/wages, are more likely to receive a loan than a new customer who has no history with the bank, or who earns low salaries/wages. By knowing what the chances of default are for a customer, the bank is in a better position to reduce the risks.

**Fraud Detection:** Another popular area where data mining can be used in the banking industry is in fraud detection. Being able to detect fraudulent actions is an increasing concern for many businesses; and with the help of data mining more fraudulent actions are being detected and reported. Two different approaches have been developed by financial institutions to detect fraud patterns. In the first approach, a bank taps the data warehouse of a third party (potentially containing transaction information from many companies) and uses data mining programs to identify fraud patterns. The bank can then cross-reference those patterns with its own database for signs of internal trouble. In the second approach, fraud pattern identification is based strictly on the bank’s own internal information. Most of the banks are using a ‘hybrid’ approach. One system that has been successful in detecting fraud is Falcon’s ‘fraud assessment system’. It is used by nine of the top ten credit card issuing banks, where it examines the transactions of 80 per cent of cards held in the US. Mellon Bank also uses data mining for fraud detection and is able to better protect itself and its customers’ funds from potential credit card fraud.

**Customer Acquisition and Retention:** Not only can data mining help the banking industry to gain new customers, it can also help retain existing customers. Customer acquisition and retention are very important concerns for any industry, especially the banking industry. Today, customers have so many opinions with regard to where they can choose to do their business. Executives in the banking industry, therefore, must be aware that if they are not giving each customer their full attention, the customer can simply find another bank that will.
Data mining can also help in targeting ‘new’ customers for products and services and in discovering a customer’s previous purchasing patterns so that the bank will be able to retain existing customers by offering incentives that are individually tailored to each customer’s needs. When Chase Manhattan Bank in New York began to lose customers to competitors, it began using data mining to analyse customer accounts and make changes in its account requirements, thereby allowing the bank to retain its profitable customers. Data mining is also being used by Fleet Bank, Boston, to identify the best candidates for mutual fund offerings. The bank mines customer demographics and account data along different product lines to determine which customers may be likely to invest in a mutual fund, and this information is used to target those customers. Bank of America’s West Coast customer service call centre has its representatives ready with customer profiles gathered from data mining to pitch new products and services that are the most relevant to each individual caller. Mortgage bankers are also concerned with retaining customers. The program uses leadingedge Internet technologies, predictive models, and customer-direct marketing to enable lenders to identify new customers and retain those that they already have.

SOFTWARE SUPPORT

Keeping in mind the usefulness and applicability of data mining techniques in various sectors, the software development companies have come up with various applications, which can automate the task of data mining. Some such software are:

STATISTICA Data Miner, A venture of StatSoft worldwide, is a revolutionary product in the data mining applications. It enables financial institutions to Detect patterns of fraud; Identify causes of risk; create sophisticated and automated models of risk; Segment and predict behavior of homogeneous groups of customers, Uncover hidden correlations between different indicators. 11Ants Analytics Ltd is a venture backed company located in Hamilton, New Zealand. 11Ants Analytics is committed to making advanced data mining accessible to non-technical users. They have built incredibly powerful data mining software which is deceptively simple to use. Data Mining with SAS® Enterprise Miner: SAS data mining software helps customers to: detect fraud; anticipate resource demands, increase acquisitions, curb customer attrition

CONCLUSION

Data mining is a tool used to extract important information from existing data and enable better decision-making throughout the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analysed and the information that is captured is used throughout the organisation to support decision-making. It is universally accepted that many industries (including banking, retail and telecom) are using data mining effectively. Undoubtedly, data mining has many uses in industries. Its practical applications in such areas as analysing medical outcomes, detecting credit card fraud, predicting customer purchase behaviour, predicting the personal interests of Web users, optimizing manufacturing processes etc. have been very successful. It has also led to a set of fascinating scientific questions about how computers might automatically learn from past experience. The retail industry is also realising that data mining could give them a competitive advantage. A majority of the banks in developing countries (particularly in the public sector) are not usually known to exploit their information “asset” for deriving business value through data mining and gain competitive advantage. But with progressive liberalisation of rules on entry for private and foreign multinational banks, under the GATS framework of WTO, competitive pressure on domestic banks is increasing. Thus, customer retention and acquisition will be an important determinant of the banks’ bottom lines. Those banks and retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision-making process will reap immense benefit and derive considerable competitive advantage to withstand competition in future.

REFERENCES

[5]. S.P. Deshpande, Dr. V.M. Thakare, ”Data Mining System And Applications: A Review”. 