



Corruption Prevention Network

CPN Special Edition News Update



Welcome to the Corruption Prevention Network Special Edition of News Update. This Special Edition takes you into the magical world of data mining and data mining services. CPN News Update provides a collection of current news, events and publications from various government and non-corporate organisations. If you would like to contribute to the Corruption Prevention Network News Update please email the Secretary at [cpn \[at\] corruptionprevention.net](mailto:cpn@corruptionprevention.net).

Special Data Mining Edition – What's it all about?

As always the CPN has arranged yet another way to give you the information you need to implement and enforce an effective Corruption Prevention Network in your organisation.

As such we are dedicating this latest edition to Data Mining. In organising this edition we sent out an email invitation to various organisations and people who work with data mining tools and asked that they provide an article based on experience and service.

In response we received four articles, including an introductory article on data mining and corruption prevention by Cicil Fonseka, Senior Systems Analyst for NSW Department of Health and former data mining lecturer at the University of

Index

Special Data Mining Edition – What's it all about?

Data Mining in Corruption Prevention
By Cicil Fonseka

An Introduction to Data Mining by Warwick Graco

KPMG – Fighting Fraud with K-Trace

PriceWaterhouseCoopers - Pro-active Fraud Detection

Disclaimer

Western Sydney and a introduction to data mining in general by Warwick Graco, Analytics Project, Change Program at the Australian Taxation Office.

The CPN continues to remain neutral on the subject as always, and as part of this neutral stance, we have arranged the articles in alphabetical order.

Enjoy!!!!!!!!!!!!!!

Index

Data Mining in Corruption Prevention

By Cicil Fonseka
Senior Systems Analyst
NSW Department of Health
(Previously Lecturer in Data Mining at the University of Western Sydney)

Data mining is identifying interesting patterns in large data sets. A simple example of data mining is that a retail bank finding out the customers who paid off a personal loan during the last 6 months and have a credit card account are more likely to get a home loan. The home loans promotions can target such customers and the data mining system used by the bank has discovered new information from the large customer and transactions data sets that was previously unknown.

Data mining has been defined as “the process of exploration and analysis by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules.” [1]

Currently available data mining software allows to analyse large volumes of raw data from business systems, applications, databases, web sites, and text based mediums.

Application areas: Data mining has been used in a wide variety of applications and industries including health care, fraud prevention, intrusion detection, character recognition and biometrics. Some examples are:

Telecommunications

Customer churn is a concern for telecommunications companies, ISPs and many other service providers. Data mining can identify reasons for churn and help reduce customer churn. Data mining has been used in lowering cost of customer acquisition and increasing total customer life-time

value.

Utilities

Utility companies provide critical goods and services as power and water to communities. Data mining has been effectively used to predict equipment failure quickly and more cost effectively, and then to perform preventive maintenance to minimize disruption in service.

Data Mining in Corruption Prevention

Corruption prevention is an area data mining technologies are not widely used. However the potential to use data mining in Corruption prevention is enormous.

When the corrupt behavior involves monetary transactions these are always recorded in systems within the organization and external systems. An analysis of the situation can define the non-corrupt behavior and data mining software can be modeled to look for or mine transactions different from the non-corrupt behavior.

For example, the abuse of corporate credit cards can be detected by data mining software by analyzing the transaction datasets for;

Type of merchant:

Merchant category codes such as cinemas, perfume shops, jewelry stores, pawn shops, and gambling places, merchant having the same name as the cardholder.

Nature of Transaction:

Personal use items luggage and accessories, sporting goods, sunglasses, holidays, gentleman clubs, cruise lines casinos etc.

Transaction amount:

High amount transactions, large number of transactions with one vendor, numerous transactions in round dollar amounts

Timing of transactions:

Holiday and weekend transactions, end of year transactions unusually high number of transactions made late at night, multiple transaction on the same day with one vendor.

As the example above indicates data mining can be used to detect corruption, prevent corruption and reduce loss generated by corruption.

An Introduction to Data Mining

Warwick Graco
Analytics Project
Change Program
ATO

Email: warwick.graco@ato.gov

Data mining is a branch of knowledge discovery and is concerned with finding relationships, patterns and trends in data – especially in large and complex datasets. What is specifically sought in data are ‘nuggets of gold’ or insights which raise profits, increase productivity, lower costs and achieve other positive outcomes. These nuggets are hidden in all large databases and the role of data mining is to find them.

Those doing mining activities use statistical and machine-learning techniques to find interesting patterns of scores and correlations in data and to make classifications and predictions. There are three types of machine learning. The 1st is *supervised or guided learning* where computers learn the relationships between inputs and an output such as the characteristics of horse races and the time horses take to run races. The model that is developed is applied to predict how fast horses will run in future races. The 2nd is *unsupervised or discovery learning* where the computer has nothing to guide it in its learning and instead has to find score configurations and correlations in data. An example is identifying the different combinations of items people purchase in supermarkets. The 3rd is *reinforcement learning* where desirable behaviour is rewarded and undesirable behaviour is punished. This type of learning is used to train robots. All three types of machine learning are used in data mining.

There are three types of temporal models developed by data miners. One is *retrospective* models that look back in time such as the purchases made using a credit card in the last month to assess if any were fraudulent. The 2nd is *inspective* models that classify the risk in a transaction such as an insurance claim. The 3rd is *prospective* models that predict what will happen in the future such as the number of suspect claims patients will make in the next 12 months. .

Data mining can be applied in a variety of fields such as industrial production, sports medicine, marketing and revenue collection. One prominent use of data mining is with fraud

detection. Two basic strategies used for this purpose including anomaly detection where behaviour that is out of character with a norm is identified and investigated. An example is a person who makes credit-card purchases that are not consistent with previous purchases. The other is signature approaches where like radar cases which have patterns indicative of fraud and abuse are identified. Data mining is also used with customer analytics for identifying different types of customers and their purchasing preferences.

Data mining is not a magic cure that will solve all organisational problems. It works best when there is a clearly defined problem that needs investigation and is based on careful analysis and review. Intelligence analysis should be applied to identify threats and opportunities while risk analysis is used to identify the risks associated with these issues. The intelligence process informs risk analysis. Those risks which cannot be mitigated are profiled to identify the modus operandi of each danger or vulnerability and their defining attributes such as the distinguishing characteristics of a fraudster and the frauds committed. This knowledge is used to explore data to find other cases and issues of interest and to develop classification and prediction models.

It is important to marry business knowledge with data-mining skills to ensure that those doing mining are guided in the work they do and that they receive feedback on the results they produce. The business experts act as navigators while the data miners are the drivers of the mining activities.

In conclusion, most large organisations have huge mountains of data most of which is unexplored. Data mining is a way of converting this data to knowledge thus enhances organisational performance.

[Index](#)



Fighting fraud with K-Trace

Incorporating advanced data mining and visualisation capabilities, K-Trace is a powerful new tool for controlling fraud and misconduct. Rod McKemmish and Matt Fehon tell us why.

New regulatory requirements and auditing standards are forcing organisations of all kinds to pay more attention to

fraud and other forms of misconduct. For example, the revised *Australian Auditing and Assurance Standards AUS 210, The Auditor's Responsibility to Consider Fraud in an Audit of Financial Statements*, expands the external auditor's responsibility for detecting fraud, and for assessing the effectiveness of an entity's fraud controls. Of course, as the standard notes, the prime responsibility for controlling fraud remains with an organisation's board and senior management.

Certain recent high-profile frauds have also sensitised capital markets to the issue. Because fraud always involves an element of concealment, its early detection can be difficult. However, fraud invariably raises warning signals for those who know *what* to look for, and *where* to find it. Forensic data analysis applies knowledge discovery techniques to the detection and prevention of fraudulent, irregular or unethical behaviours. It isolates the "red flags" that signal something is amiss.

In one recent example, an employee within an accounts payable department inadvertently discovered that they had write privileges to the master vendor information. Utilising this access, the employee proceeded to replace selected vendors bank account details with their own bank account number. Unfortunately, due to the lack of any suitable real time electronic auditing process these changes were not detected immediately. Fortunately however, the periodic use of Forensic Data Analysis by the organisation facilitated the identification of the newly established relationship between the employee and the vendor, resulting in a red flag being raised and the matter subsequently being investigated.

Unlike traditional detection and auditing techniques that rely on data samples, modern forensic analysis works with complete data sets. It reduces the risk that anomalous events will be overlooked or misinterpreted. It can be used on very large data sets, and on multiple databases. KPMG's approach to forensic data analysis combines data mining and data visualisation techniques to identify patterns and relationships that may indicate the existence of fraud or other irregular behaviour. It follows a six-step methodology.

KPMG Forensic Data Analysis Methodology KPMG has further developed this methodology into a forensic toolkit called *K-Trace*. It employs powerful data mining and data visualisation techniques. Using K-Trace, our forensic practitioners can search for fraud indicators, or red flags, and identify control weaknesses.

In comparison with older methodologies, K-Trace offers several important advantages.

- It features customised routines developed by experienced fraud investigators and forensic accountants.
- It contains specific routines for identifying irregularities within accounts payable, accounts receivable, payroll and expense reimbursement data sets.
- It allows the development of specific routines to meet individual client needs.
- It enables the review of existing fraud prevention and detection controls.
- It analyses 100 percent of data populations, including data from multiple sources.
- It offers a time and cost efficient approach.
- It builds on tested methodologies.

Why can't organisations adopt a DIY approach to forensic data analysis? Well, they can, but it's a bit like reinventing the wheel. Moreover, forensic data analysis is a highly specialised skill set based on an intimate knowledge of fraudulent behaviours, business processes and IT systems. Many organisations will find it more convenient and cost efficient to utilise the services of experienced professionals in the field.

For more information on K-Trace contact

KPMG Forensic
forensic@kpmg.com.au
1800 500 376

Rod McKemmish
Director,
KPMG Forensic
rmckemmish@kpmg.com.au
02 9335 7687

Matt Fehon
Senior Manager,
KPMG Forensic
mfehon@kpmg.com.au
08 9263 7539

Pro-active Fraud Detection

Most modern organisations store the majority of their financial information in vast electronic databases. Customer information, employee data, vendor data, journal entries and transactions themselves are stored within these databases. The complexity of modern day commerce, and the sheer volume of electronic information, provides the



opportunity for fraudsters to conceal their activities within the millions of items of valid data held within these databases.

Manual testing of data is rarely an effective or efficient solution, and hardly the job of time-pressed management or external auditors for those interested in identifying suspicious data. In addition, many organisations find it difficult to identify suspicious transactions as they do not possess the right tools or the relevant expertise to efficiently ‘connect the dots’ to identify them. However, automated fraud detection programs are available to assist an organisation to identify anomalous transactions or other data records that appear to be suspicious and therefore might be worth a ‘closer look’ or further investigation.

In order to identify suspicious transactions, organisations need to develop a methodology that enables them to:

- Match data ‘fields’ within their databases (‘data matching’)
- Compare data fields with external data sources or profiles
- Quickly drill down to the individual transaction level.

An appropriate methodology will allow for the easy identification of fraudulent or suspicious transactions that require further investigation. The process used will often involve the use of an expert and specific data mining software tools. These tools need to be adaptable and able to be used across many different databases; they also need to be able to handle large data volumes.

There are literally thousands of possible fraud detection tests that can be run in a typical organisation. The decision as to which tests to run depends on a number of factors, such as type of business, quality of data, number of employees, vendors and

customers, standards of internal control, past incidents and so on. However, these best tests are those that have been designed after 'reverse engineering' actual fraud incidents. Some of the more useful tests are as follows:

- Tests that match employee details (address, telephone, and particularly bank account) with vendor details
- ABN and TFN validity tests
- Duplicate transaction tests – particularly those matching invoice and payment details, rather than vendor details
- Tests that identify high proportions of sales credits per customer
- Payments to redundant vendors or employees
- Detailed matches against known fraud profiles (prisons, registered mail boxes, serviced offices).

The sheer volume of transactions entered into often means that without the utilisation of data mining software, suspicious transactions that require further investigation can remain undetected.

High risk areas in many organisations include purchasing, payment and expense records, although data mining testing can also yield results in sales, inventory, insurance claims, superannuation payments and entitlements and other areas of organisations where the potential for fraud can exist. The results of a testing program are a detailed list of questionable transactions, employees, suppliers and so on which need further, more detailed investigation.

One of the main criticisms around fraud detection testing is the volume of 'false-positive' results when large data sets are tested. This has been a problem for years as testing is all but useless if results are not followed up – most of the time, results will have a legitimate explanation – it is often a small proportion of results that prove to be illegitimate transactions. To overcome this problem, experts are now using more sophisticated data mining algorithms to conduct the testing.

These algorithms work by assigning different risk weightings to positive results for each test conducted. These risk weightings are then carried forward into each transaction. For example, a purchasing transaction in which:

- The vendor has a non-conforming ABN
- The vendor operates from a serviced office address
- The payment transaction is to a bank account shared by an employee
- The payment date is very close to the invoice date.

....would have a very high risk rating. In this way, it is theoretically possible to identify the single 'riskiest' transaction in a listing of many millions of transactions.

These refined results will complement internal and external auditors by providing a narrower and targeted listing of results to review. An automated fraud detection program run before an audit will also complement an organisation's existing schedule of audit visits, making the best use of valuable and often scarce resources. It is a tool which will quickly identify problem areas and can also be used where applicable, to audit the records of suppliers where a 'right to audit' exists. The process is simple, time-efficient and will not jeopardise the integrity of an organisations data or be disruptive to normal business operations.

For further information concerning the issues discussed in this article, please contact Kate Kuerschner on 61 7 3257 8885 or kate.kuerschner@au.pwc.com. Kate Kuerschner is an executive in the Investigations and Forensic services practice of PricewaterhouseCoopers, Brisbane specialising in the delivery of our fraud detection data mining service - suspicious transaction analysis, forensic accounting and fraud investigations. She is also a Chartered Accountant and Certified Internal Auditor. PricewaterhouseCoopers Investigations and Forensic Services Practice consists of approximately thirty staff across Australia, with backgrounds in law enforcement, civil investigation, computer crime and accounting.



Thank you to everyone who has contributed to this edition of E-News. We are hoping that the next Special Edition of E-News will focus on IT Investigations and the impact, if any, the new Workplace Surveillance Act has had. If you would like to contribute by writing an article or submitting information regarding a service you offer, please email [cpn \[at\] corruptionprevention.net](mailto:cpn[at]corruptionprevention.net).

Sarah Morrison

[Index](#)

Disclaimer:

Please note that information regarding the various publications and training have been taken directly from the various Internet sites cited. In some cases the information has been paraphrased. Advertising of Publications and Training is with permission of the provider. If any information is cited

incorrectly, or if people wish to no longer have their details cited in this publication or wish to add additional information, please email the CPN and we will amend accordingly.	
---	--
