

Discovering Fraud in Forensic Accounting Using Data Mining Techniques

Financial fraud has become a common phenomenon collapsing many companies, and weakening the economy. Fraud includes an array of irregularities and illegal acts characterised by **intentional** deception. Frauds related to receivables, vendor, payroll, and expense are some of common business frauds taking place in companies. In this context, forensic accounting plays an important role in detecting these frauds otherwise not found out in internal auditing by employing accounting, auditing, and investigative skills.

Forensic accounting is the action of identifying, recording, settling, extracting, sorting, reporting, and verifying past financial data or other accounting activities, for evidence to be suitable for the purpose of establishing accountability and/or validation of the activities of the firm. It differs from regular auditing. Auditing focuses on a sample of transactions, checks whether the information disclosed by the financial statements is supported by adequate material and makes qualification in audit report in case of any deviation, error, omissions, or exaggerated assertions etc. The basic objective of auditing is not to detect fraudulent transactions. Forensic accountants on the other hand searches for exceptions, oddities, irregularities, patterns, and suspicious transactions. There are no standard procedures, checklists and controls in forensic accounting to discover these frauds, as



- Dr. P.K. Panigrahi

The author is a senior faculty member in the Systems Department of Loyola Institute of Business Administration, Chennai. He can be reached at p_panigrahi@lycos.com

Forensic accounting is a growing field in the area of accounting which helps in detecting financial frauds. The tools and techniques employed in forensic accounting face a lot of challenges. This article discusses various issues related to current methodology and suggested a new technique called "Data mining" to help the forensic auditors in discovering and predicting frauds.

each type of fraud is unique to specific situation.

Forensic accounting uses some mathematical models such as Benford's Law, and Relative Size Factor (RSF) that are implemented in various computer-assisted auditing tools and audit command languages. Although this technique gives some relief to forensic accountants, the exponential growth of data,

complex fraudulent transactions, smartness of fraudsters and growth of technology pose tremendous challenges to the modeling techniques currently used in forensic accounting. In this situation data mining techniques help forensic accountants. Data mining is the process of discovering previously unknown and actionable trends, patterns and relationships in the data sets and also helps in modeling prediction. This article discusses various issues related to mathematical techniques currently employed in forensic accounting, and explains how data mining techniques can be applied in discovering as well as predicting suspicious transactions or fraud, with an illustration.

Techniques and Tools Used in Forensic Accounting

The popular mathematical techniques currently used in forensic accounting are Benford's law and theory of Relative Size Factor (RSF).

The objective of Benford's Law is to determine whether the field under study is free from any unintentional errors or frauds. *The basis of this law is that fabricated figures (an indicator of fraud) possess a different pattern from random (or valid) figures.* The steps for Benford's law are very simple. The left-most digit of the field is extracted, summarised for the entire population by classifying the first digit field and calculating its observed count percentage. Then the Benford's set is ap-

plied, followed by a Z-test. The significance of variance between the two population i.e. Benford's percentage numbers for first digit and observed percentage for first digit, for a particular level of confidence, is measured. If the data confirms to the percentages of Benford's Law, it is likely that there is no error or fraud. Benford set is available for all other digits, combination of digits, decimal as well as rounded numbers. The law is having few advantages such as scale invariance, non-requirement of supporting documents, but has many limitations. The law cannot be applied to categorical data, sample data and data with range limits. It can be applied to one variable at a time and the experiences gained, can neither be generalised nor can be used in other variables or other situations.

Relative Size Factor detects outliers or unusual data, which may be due to either simple errors or frauds. It is based on the basic concept that each field in any transaction has a normal range and any data falling outside the range is unusual or an outlier and need to be further investigated. The formula is very simple. It is measured as the ratio of the largest number to the second largest number of the given set. Although it is simple to calculate, it is not at all an effective and efficient technique. The outliers may be due to data entry errors and not frauds. On the other hand frauds are engineered in such a way that the data would be in the limits, which make impossible to be detected.

Auditors are currently using Computer Assisted Auditing Tools (CAATs) to deal with huge data sets, and to process complex transactions thereby saving time, and improving effectiveness. The tools help au-

ditor in implementing auditing procedures such as:

- a) testing details of transactions and balances,
- b) identifying inconsistencies or significant fluctuations,
- c) testing general as well as application control of computer systems,
- d) sampling programs to extract data for audit testing, and
- e) redoing calculations performed by the accounting systems.

The tools use various financial statements, analyse "ratios" and detect any unusual or anomalies. Audit Command Language is one of the popular software available currently. The software is an improved version of Microsoft Excel and having no limitations on the number of columns and rows.

Data Mining Techniques and Detection of Suspicious Cases

The techniques mentioned above confront several challenges. Each variable is analysed separately without recognising any relationship with other variables. Basically they detect "outliers" using some filtering conditions. On the other hand, all outliers are not frauds and a non-outlier may not constitute a genuine transaction. These techniques generate huge number of outliers so that it becomes impossible to investigate further. There is a need to narrow down the search space. Also wrong decisions such as considering a transaction as a fraud when it is actually not and approving a transaction as normal when in reality it is fraudulent are many. The crux of these limitations is that relationship between two or more variables or complete transactions but not any single variable or transaction, identifies a suspicious pattern. In

this situation, data mining has emerged as a strategic tool to identify different types of frauds and meet the above-mentioned challenges.

Data mining is a technique with an objective to mine large amount of data to discover previously unknown, action oriented, hidden trends, patterns and complex relationships. The technique studies past data, operates on all the variables and entire population, extracts variables of importance and uncovers patterns in the form of rules, and formulates model by using different techniques. The resulted rules or patterns help auditors to flag only those transactions most likely to be fraudulent.

Discovery, predictive modeling and deviation analysis are three major activities of data mining techniques. In "discovery", usual knowledge or patterns such as associations, trends and variations in data are discovered in the form of logic or rule without any pre-defined hypothesis about what the pattern may be. Some of the discovery techniques are association, sequencing, cluster analysis and Kohonen neural network. An association is a rule, which implies certain association relationships among a set of objects (such as "occur together" or "one implies the other") in a database. The technique creates rules that describe how often events have occurred together. The sequencing technique is the same as association, except that the time sequence of events is also considered. Cluster analysis, an unsupervised technique, segments a diverse group into mutually exclusive similar subgroups or clusters. The members of each group are as homogeneous as possible and different groups are as heterogeneous as possible. In predictive modeling, patterns discovered from the

data sets are used to predict the outcome, and to guess values for new data items. Deviation analysis applies the extracted patterns to find anomalies or unusual and specific cases.

Defining and detection of fraud is a very difficult task and any error in decision would create unhealthy situations. This article does not claim that data mining techniques are superior to human (i.e. forensic auditors) mental model (intuitions, judgments, foresight and experiences) in detecting frauds. On the other hand, the techniques help the auditors in automatically narrowing the search space, sharpening the decisions, reducing the errors and identifying complex frauds. Mental model and data mining models together improve the discovery and prediction of frauds.

Some of the general characteristics of data patterns of fraudulent transactions that need to be discovered, with the help of specific data mining tools from the accounting database, are discussed below.

- a) Transactions, where each variable contains data seemingly usual but the combination of variables look unusual, are likely to be fraudulent. The number of these types of transactions in general, is huge and hence difficult to detect rare fraudulent ones. We do not have past information or any reference to these types of transactions and hence one relies on domain knowledge. Certain combinations, which either should or should not occur, need to be focused. In this situation general-purpose association and sequencing techniques are applied on population transactions.
- b) There are some transactions,

where combination of entries is unusual with respect to a comparison or reference value(s), are also likely to be fraudulent. The reference value is determined by unsupervised cluster analysis. This is followed by an outlier analysis, to detect cases deviating from this reference value.

- c) Sometimes the value of a variable is unusually high or low. This may be due to data entry error or any other mistake and not necessarily a fraud. Descriptive statistics such as range, standard deviation and relative size factor etc. would be helpful in detecting this type of errors. Any data mining technique is not required to identify these irregularities.
- d) Accounting transactions are maintained in various files. In some occasions, one can find two or more seemingly unrelated records of two or more files having unexpectedly the same values for some or all of the fields of business relevance. Here there is a chance of fraud. This is one type of unexplained relationships and need unsupervised data mining technique, as there is no prior information about fraud. A cluster analysis on the variables in question is most appropriate data mining tool.
- e) In some situations, two or more unrelated records having unexpectedly the values showing a direct or indirect link between the records confirming that these records are related in some ways. This type of unexplained relationship is an indication of fraud. Unsupervised data mining technique and link analysis

are most appropriate to identify this type of fraud.

Once it is known what transactions are frauds and what are not using above discovery techniques, these outcome or historical data can be used further to build predictive model for detecting fraud. In predictive modeling, known outcome variable (fraud=Yes or No or degree of fraudulent characteristic=Threshold value) and a set of important variables that contributing the outcome, are modeled using a mathematical technique. Once the model is built, it is then used to predict a completely new transaction whether it is fraudulent or not and also identifies the variables of importance. This is called predictive data mining and it includes tasks such as classification, regression, time series analysis, and prediction. This is implemented by using regression, decision trees, or neural networks. A decision tree is a technique that generates a graphical representation of the model and accompanied by rules of the form "if condition then outcome".

Illustration: Fraud in Futures and Options Trading and Investment Company

In Future and Options (F & O) operations, a margin account is set up with a bank and a particular amount of fund is deposited in the bank account for trading. Trading is done through a group of qualified brokers on a daily basis. The banks are advised for providing appropriate margins to the brokers for the trades transacted. A fund manager manages the fund and scripts. Some of the important documents, database and registers maintained in the company are: a) detailed monthly bank

statement showing profit and loss and margin utilisation on a daily basis, the cumulative utilisation till date, and the margin account balance at the end of the day; b) comprehensive trade register includes information such as scripts traded, broker details, brokerage, margin adjustments, open positions for contracts not squared off, booked profits for closed contracts and notional profit for open contracts; c) broker statements with contract notes; and d) daily month-end balance and cumulative margin statement from banks. There is no concern for the company as the company earns a good profit supported by bank statement (daily month-end balance and cumulative margin) tallied with the fund position reflected in trade register. In a fraudulent environment the forensic auditor might observe the following: a) physical supporting documents might have been missed (e.g. contract notes for the trades in the trade register; detailed monthly bank statements b) Information might not be compiled in proper way; c) some contracts notes missing; and d) relevant computer data files deleted etc.

In this situation, some of the data mining techniques mentioned earlier can be applied to identify some red flags.

1) Discovery techniques can be applied to transactions with variables such as fund manager's profile, behaviour and performances, market positions, profit and loss in different types of scripts. Some information are already there in the accounting database while some are not available. A set of non-accounting information is also need to be collected. All these information should be collected in

different time frames. One might discover a) a profit in troubled period for some particular scripts, and b) a consistent profit figure in a speculative market etc, both unusual data, and are indication of fraud.

- 2) Various accounting transaction databases such as script database (script details), trading database (dates of transactions, time gaps, value of transactions, date of squaring) and broker database can be analysed to discover various relationships between records. With the help of domain expert, and unsupervised data mining techniques mentioned earlier, various complex relationship between records could be investigated. Some interesting patterns such a) large-value scripts traded and squared up on the same date through different brokers, b) certain trades squared up on the same day either for a small profit or a small loss, c) purchase and sale of script effected on the same date through different brokers at a very close interval of time and d) squaring up not done through the same broker etc., good indication of fraudulent transactions, can be discovered.
- 3) Links of purchase and sale of various scripts are analysed using link discovery techniques and outlier analysis is done to identify errors in the purchase and sales rates as well as quantities. Similarly errors in MIS reports can be identified. Some errors like "lesser losses or greater profits for some scripts" may be early indication of a major fraud in future.

Once a small data set is filtered, it can be manually analysed with experiences, intuitions, judgment and other forensic auditing tools and techniques. All the above techniques are explorative in nature which helps in prediction. The knowledge or pattern gained from the above analysis can then be easily applied to the larger data set as well as new transactions using predictive modeling. In this way, the forensic auditor can discover as well as predict some suspicious transactions.

Discovery techniques can be applied to transactions with variables such as fund manager's profile, behaviour and performances, market positions, profit and loss in different types of scripts.

Limitations

Data mining technique is not free from limitations. The technique requires huge amount of data and selection of appropriate variables. The input data has to be cleaned and transferred and hence limiting the use of online data mining. Fraudsters generally hide and destroy a lot of information required to detect frauds which limits use of some data mining tools. Also pattern derived from past may not continued to be the same in future. Certain types of frauds are difficult to be detected by data mining tools. Instead observation and deep insight of the forensic auditors into the problem help a lot. Therefore, data mining tools should be used only with the help of domain experts like forensic auditors and keeping in mind its limitations. □