Data search and discovery process for financial statement fraud

Mahmoud Lari Dashtbayaz ,
Assistant Professor in Accounting, Ferdowsi University of Mashhad, Mashhad, Iran

Shaban Mohammad,
Department of Accounting, Hakim Nezami University of Quchan , Quchan, Iran
E-mail: shasban1362@gmail.com

Abstract
Financial fraud in the form of significant public interest, media, investors, financial community and legislators to get oneself and for several large companies such as Enron fraud is known, Lucent and WorldCom have occurred in the past years. Fraudulent financial reporting in more than declarative through asset sales and profits and low self liabilities, costs and losses occur (Yue et al., 2007). Fraud, an important reason for the failure of many companies and particularly damaging to capital markets because investors, creditors and financial analysts in their decisions on the financial statements available to the public, they rely on and trust. Detect fraud in financial statements, it is difficult and complex issues discovered (Yue et al., 2007). The main objective of this paper is to provide an overview of data mining processes used to detect financial fraud, particularly fraud in the financial statements.
Keywords: Mining, fraud, financial statements, forecasting, regression, artificial neural networks

Introduction

Auditing standards, audit firms are required to manage the discovery of the fraud. This requirement, the need to explore effective fraud management increases. However, the discovery of the fraud management using normal audit procedures, is difficult. Finally, general managers, financial managers and accountants, auditors are deliberately trying to deceive (Fanning & Cogger, 1998). For those managers who recognize the limitations of an audit, audit procedures, standards may be inadequate. These restrictions require additional analytical procedures for the effective detection of fraudulent financial statements for all socioeconomic groups. Statistics and data mining techniques successfully to detect activities such as money laundering, credit card fraud e-commerce, telecommunications fraud, insurance fraud and computer influences have been used. Accepted definition of financial fraud, there is no universal (Ngai et al., 2010). Wang and colleagues (Wang et al. 2006) have defined fraud "Targeted interventions for illicit financial gain, contrary to the laws, rules, or politics." In recent years, various types of financial fraud, such as credit card fraud, corporate fraud and money laundering, has caused much concern and has attracted attention. Ngay and colleagues (Ngai et al., 2010) in the general classification, types of financial fraud has been divided into four categories, bank fraud, insurance fraud, securities and commodities fraud and other financial fraud. In practice, companies can distinguish two types of fraud. The first type is the misuse of assets, for example in the form of theft, embezzlement, falsification of expense accounts, personal use of company assets, and so on. And second, the fraudulent financial reporting. it has been reported this year. Fraudulent financial statements typically managed by or with the consent or knowledge to do so. Thus, manipulation of financial statements can be understood that the kind of fraud management and Willingham Elliott (Elliott & Willingham, 1980) is defined as "a purposeful fraud committed by management and the financial statements misleading material information to investors and creditors damages. " The auditor's responsibility is to determine to what extent financial reports in accordance with the accounting standards; Fraud risk assessment in financial reports and financial statements to identify fraud. On the other hand, the discovery Financial fraud is a difficult task that requires more effort than implementing standard audit procedures. Therefore, auditors for audit and help facilitate the detection of fraudulent financial statements of these tools and techniques New need. In this connection, a computer-based tool can be useful contextual help (Ata & Seyrek, 2009).

1. Data Mining and its duties.

With the advent of information technology to collect, store and process large amounts of data and facilitate possible, organizations are now using data from processes, customers and collect their environment, are in a better situation. But extracting meaning from the mass of data and its use is beneficial for organizational purposes, requires the use of advanced techniques, such as data mining is (Ata & Seyrek, 2009). Data mining is to discover previously unknown relationships that provide useful information (Hand et al. 2001). Bose and Mahapatra (Bose & Mahapatra, 2001) Data mining is the process of identifying interesting patterns in data bases that can be used in decision making, have been defined. Turban et al (Turban et al., 2007) have defined data mining process to extract and identify useful information and the subsequent acquisition of knowledge from

large data bases, statistics, mathematics, artificial intelligence and machine learning is used. Fraly and colleagues (Frawley et al., 1992) suggest that the purpose of data mining, information and useful Ghyrsryh of data stored in a large tank. Do data mining, business awareness of the statistical reliability of data and applications, but are not known in advance. This data should be accessible, relevant and sufficient. Also, data mining problem must be clearly defined (Phua et al., 2005). Mining in various industries such as healthcare, finance, retail, telecommunications, etc. To solve Msylhha and improve various aspects of the business are used. Fva and colleagues (Phua et al., 2005) note that a Jaaftadhtryn fraud detection applications in the mining industry and the government has become. In summary, data mining involves taking basic steps that should be taken carefully to the significant results obtained from analyzes conducted. The first step in data mining, clear definition of the problem. Data are collected and prepared for modeling. This step is very time consuming and basically involves filtering and manipulation of data. Then, a data mining model to examine the issue, it seems more appropriate than others, will be selected and used. At this stage, generally two sets of training and test data (approved) divided. Data extraction rules and relationships education model, used, while the data is verified (tests) are used to determine the rules created for a different set of data, the quality of the act. Based on the results of the implementation model, performance evaluation model. Consequently or modified or used to solve the problem (Ata & Seyrek, 2009). Application of data mining techniques to detect financial fraud Ngay and colleagues examined. They examined 49 published articles in journals and found that the investigators studied a series of six classes of functions / applications of data mining is used to detect financial fraud. These include classification, regression, clustering, prediction, detection of outliers and visualization. Each of the six floors, with a set of algorithmic approaches for data connections are supported by the following extract (Turban et al., 2007). This approach can solve the problems that are more complex. These classes can be broadly follows.

Classification:
The classification model is used to predict the class of objects known as the objects belonging to different classes are used to distinguish the cause. The pre-defined class titles, but distinct and not regular (Han & Kamber, 2006). Zhang and Zhou (Zhang & Zhou, 2004) argue that the process of classification and prediction models to identify a set of common features that characterize and differentiate the data classes or concepts. Classification methods include neural networks, Bayesian networks are a simple, decision trees, and support vector machines. Such classification tasks in credit card fraud detection, health insurance, car insurance and corporate fraud and other types of fraud over the Internet. Classification is one of the most popular models of learning in data mining applications to detect financial fraud (Ngai et al., 2010). The classification is a two-stage process. In the first step, using a training sample, a model is trained. This is an example of the number of rows (sub-group) and columns (attributes) are organized. One of the characters, the story, the characters, contain pre-defined categories, which indicates that each row belongs. This step is also known as supervised learning. In the second step, the model is trying things that do not belong to the training sample, classified as a test sample (confirmed) form (Kerkaus et al., 2007)

Clustering:
Of clustering to divide objects into groups / clusters that are significant in the sense used, so that a group of similar objects, and yet, very different from other objects in groups. Clustering, also known as segmentation and data scanning, and as a kind of unsupervised classification counts. According to Yvyh and colleagues (Yue et al., 2007), clustering analysis is concerned with the analysis of a dataset or fractionated (Usually multiple) into several groups so that points within the same group differ from each other as possible In the other group. Also, Zhang and Zhou argue that each cluster is a set of data objects Same cluster are similar to, but different from other objects in the cluster. The most common clustering methods are K nearest neighbor, Bayesian methods are simple and the layout of your map (Ngai et al., 2010).

Predictions:
 Forecast, future continuous numerical values based on the patterns of a given set of estimates. And Cambar Han (Han & Kamber, 2006) have stressed that the predictions, forecasts, it is a trait that should be continuous rather than qualitative (categorical). This attribute can be used to predict trait called. Neural networks and logistic model 12, the most common methods are used to predict (Ngai et al., 2010).

Discover Outliers:
 In order to measure the distance between the objects thrown from the discovery of a different way of discovering things that are inconsistent with the data set used. Data that appears to be different from the rest of the features are called Outliers. Outlier detection is / are contradictory, one of the most fundamental issues in data mining. A common method used in Outlier detection, self-learning algorithm is discounted (Yamanishi et al., 2004).

Regression:

Regression is a statistical method used to explore the relationship between one or more independent variables and the dependent variable (which is a constant value) is used (Han & Kamber, 2006). In many empirical studies, as a measure of logistic regression is used. Typically, the use of mathematical techniques such as regression, logistic regression and linear regression and use it to detect credit card fraud, insurance fraud and fraud by reporting companies (Ngai et al., 2010).

Illustration:

The researchers, from the ability to discover patterns in the human visual system uses a set of tools and applications that have the form of flexibility, data using the color, position, size and other visual characteristics, they encode. The use of imaging to detect complex patterns through the provision of clear data or functions (Ngai et al., 2010).

2. Data mining techniques to detect financial fraud

Average class / various data mining tasks mentioned above, many methods have been created from various fields such as artificial intelligence, pattern recognition, machine learning and statistics derived. In several studies, 26 mining methods have been used to detect financial fraud (Ngai et al., 2010). The most widely used methods to detect financial fraud include logistic regression models (most common), neural networks, Bayesian inference networks and decision trees that are important solutions to the problems inherent in the detection and classification of data, fraudulent offer (Ngai et al., 2010) . The 4-way, all of the "classification", which are discussed in more detailed form.

Regression:

The research literature data mining for fraud detection, the regression is the most commonly used method. Regression models Are used Logit (Logit), logistic - step by step, by using the multivariate generalized exponential and beta 2 (Yue et al. 2007) (EGB2). The logistic regression model, the most widely used model. Logistic model, a model Generalized linear regression double 25 for predicting variables can be used where little or Are qualitative. This model is essentially to solve the issues of car insurance fraud and corporate fraud (the Ngai et al., 2010). The idea behind this is that the regression using financial ratios, proportions to decide which model is obtained Associated with fraudulent financial statements. Fraudulent and non-fraudulent financial statements, including statements with data collection, we can Found that the form factors which significantly affect the companies and fraudulent financial statements to the The equation formulated. Model, based on the ratios of financial statements in the training phase as markers Been documented fraud, fraudulent and non-fraudulent companies to categorize groups will (Yue et al. 2007).

Artificial Neural Networks:

The neural network approach that uses a set of interconnected nodes, the function imitates the human brain. This The method is based on computer models of biological neurons. A multilayer neural network contains a large number of (Neurons) are linked together in a pattern of communication (Phua et al., 2005). This method greatly Classification, regression, clustering is used, and then, the most widely used data mining techniques to discover Financial fraud (Yue et al., 2007). First, the network using a set of test data to map Inputs and outputs are taught. The weights of the connections between neurons and networks established to determine New categories of data collection used (Phua et al., 2005). The benefits of this Are, first, that the method is adaptable. Second, this approach makes 26 models reliability and thirdly If you change the weight training, the classification process can also be modified. Neural networks for fraud Credit cards, auto insurance and corporate fraud used (Ngai et al., 2010). Chen and two (Chen & Du, 2009) using artificial neural networks, studied 68 companies in Taiwan Stock Exchange Respectively. With the use of financial and non-financial data, drafted a model of financial crisis. The results of their study shows Artificial neural networks are better than traditional statistical methods, to predict the financial crisis.

Bayesian inference network:

Bayesian inference network represents a set of random variables and conditional independence using a graph 27, in which the nodes represent variables guided 27 non-guided cycling accident and conditional independence between variables Determine (Kerkaus et al., 2007). Bayesian inference networks, often in credit card fraud detection, auto insurance, and Corporate fraud is used (Ngai et al., 2010).

Decision trees:

Decision trees, decision support tools are predicting a convincing picture of the consequences may make observations The (Han & Kamber, 2006). Decision trees, trees that are based on the values of adjectives topics

Classification. Leaves are expected symbol, each node in a decision tree represents an attribute of an object The classification of each branch represents the extent to which a node can actually provide and sharing features Show (Phua et al., 2005). Through a machine learning-based algorithms such as card 28 (CART), hl than 29 (ID3) algorithm and C 4/5 30 (C 4.5), the planting of trees. Decision trees are The typical credit card fraud, auto insurance and corporate fraud are (Ngai et al., 2010). Kerkaus and colleagues (Kerkaus et al., 2007), in their study of three methods that were utilized Neural networks, decision trees and Bayesian. They study the usefulness of these models to detect fraudulent financial statements Review and compare. 31 input vector, made up of financial ratios derived from the financial statements. These three Models were compared for their performance. Their sample was composed of 76 Greek manufacturing company 38 Cheaters and non-cheaters as 38 were classified as. Criteria for classification as dishonest, mainly Reports of the auditors and tax authorities. The auditors have expressed serious doubts about the accuracy of the accounts, According to the tax authorities for tax evasion by companies trying to do significant manipulation of financial statements, The companies listed on the Athens Exchange under suspension for reasons related to manipulate stock trading company Financial data and related court cases in fraudulent financial statements, are what Cheat codes are considered as one company. Krkas and colleagues (2007) reported that after Training models in the split sample test, the decision tree model with 96/2 percent accuracy, a neural network model with 100% Bayesian inference network model with 94/7 percent accuracy and precision of fraudulent companies have non-fraudulent companies Recognition Them.

3.a general framework for data mining algorithms
Although many data mining algorithms are used to detect fraud, but their use remains the traditional model of data mining feature selection, representation, collect and manage data, Pre, mining, processing and evaluating the performance follows. Yvyh et al (2007), features data mining techniques used for the specific purpose of financial fraud detection have been summarized in a general framework. Based on data analysis, financial fraud detection algorithms can be first divided into two main groups, cheaters and non-cheaters Data collected from companies, as well as the accounting data. In previous research, the focus is more on trying to uncover fraud in the collection of data was fraudulent and non-fraudulent (Yue et al., 2007).

4. Future Challenges
 Ngay et al (2010) suggest that one reason for the limited number of articles related to financial fraud detection (up to 49 years between 1997 and 2008), the research data is difficult to obtain. The problem is that before anything, and to start teaching model, to a series of fraudulent and non-fraudulent financial statements divided into two groups. Challenge identify fraudulent financial statements, financial fraud detection puts many obstacles in the way of research. Although data mining methods described above have generally been shown to be effective in detecting fraudulent financial statements, but their use for detecting fraud in financial statements, has many disadvantages and limitations of the application. Most existing data mining techniques to detect fraud in the financial statements, specific application domain specific constraints exist (Zhou & Kapoor, 2011). For example, although these methods are well developed for predictive modeling, but they have not developed to assess the effect well. Specifically, the test statistics are not yet available for some data mining techniques with which to assess the effects of independent variables on the dependent variables (Zhou & Kapoor, 2011). Therefore, analysis of Outliers to discover fraudulent patterns should be considered further. The lack of research on the use of financial fraud detection to detect Outliers Outliers may be due to the difficulty of discovery. In fact, the discovery of Outliers complex task that is not unlike taking a needle in a haystack. Unlike other methods of data mining, Outlier detection methods focus on finding common patterns associated with data objects, there are very few (Zhang & Zhou, 2004; Ngai et al., 2010). Also, the imaging methods can identify and provide considerable irregularity in the data. This can make it much easier identification and quantification of fraud schemes (Ngai et al., 2010). As Kapoor Zhou (Zhou & Kapoor, 2011) clearly caution, financial fraud detection using current detection methods, it is increasingly difficult. Zhou and Kapoor (2011), financial fraud detection based on data mining methods (such as regression, decision trees, neural networks, Bayesian networks) have been examined. In particular, their effectiveness and limitations of this method of data mining techniques in the creation of new fraudulent financial statements to conform with the discovery methods have ceased to criticize. The authors then propose a new method, an active exploration program that develops the potential fraudsters. Enabling a smart detection system for predicting, before any fraud unknown happen in the future, the power that creates new types of fraud financial statements are found to be effective. However, as the author himself admits ends, investigators discovered an active program in the future need to be designed so that it is effective and efficient.

Conclusion
In recent years, various types of financial fraud such as credit card fraud, corporate fraud and money laundering,

has caused much concern and has attracted a lot of attention towards. The area has witnessed significant Thvlhay financial fraud detection. Specifically, data mining has attracted wide attention to the increasing popularity achieved in the financial world. Reported successful applications of data mining and data mining research has shown that the effectiveness of the application and expanded. Accounting professional organizations as an important technology for data mining also acknowledged the new century (Zhou & Kapoor, 2011). The main methods used to detect financial fraud include logistic regression, neural networks, Bayesian inference networks and decision trees that are important solutions to the problems inherent in the detection and classification of data, fraudulent offer. One of the types of financial fraud, fraud management. Targeted fraud conducted by the management through the provision of financial statements misleading, harmful to investors and creditors. During the audit, the auditor should be possible to estimate the fraud management. Auditing profession, is faced with the challenge of managing a growing number of fraud cases. Advances in data mining methods for classification and prediction abilities are alleged, the auditors can ease the task of detecting fraud management (Kerkaus et al., 2007). . Detection of fraud indicators in the financial statements, significant effect on the determination of financial statement fraud. Despite the initial success of these systems and algorithms discovered by car pioneer in the discovery of fraudulent financial statements, non-continuous successful detection rate has dropped in recent years. Moreover, at a time when the financial statements fraud perpetrators, ways to circumvent the discovery programs have a car, there is an urgent need for a mechanism that is able to learn and use data mining techniques to facilitate public awareness of the industry. Also, the types of fraud and fraud patterns in different industries has changed over time. Fraud schemes have evolved to understand how important it is. Also, in order to anticipate changes and updates to keep the fraud by any means mechanical methods of fraud detection is important. In this regard, the study results may be important to develop a more robust business processes and mechanisms of adaptive management fraud detection / prevention / detection of fraud helpful.

References

1. Ata A., Ibrahim H. Seyrek, the Use of Data Mining Techniques in Detecting Fraudulent Financial Statements: An Application on Manufacturing Firms, Suleyman Demirel University, The Journal of Faculty of Economics and Administrative Sciences, Vol. 14(2), 2009, pp. 157-170

2. Bose I., R.K. Mahapatra, Business Data Mining — a Machine Learning Perspective ,Information Management, Vol. 39, 2001, pp. 211–225

3. Chen W.S. and Y.K. Du, Using Neural Networks and Data Mining Techniques for The Financial Distress Prediction Model, Expert Systems with Applications, Vol. 36, 2009, pp.4075–4086

4. Elliott R.K., and J.J. Willingham, Management Fraud: Detection and Deterrence, Petro celli Books, NewYork, 1980, p.4

5. Fanning K. and K. Cogger, Neural Network Detection of Management Fraud Using Published Financial Data, International Journal of Intelligent Systems in Accounting, Finance & Management, Vol. 7, 1998, pp. 21-24

6. Frawley W.J., G. Piatetsky-Shapiro, C.J. Matheus, Knowledge Discovery in Databases: An Overview, AI Magazine, Vol. 13, 1992, pp. 57–70

7. Hand D., H. Mannila, and P. Smyth, Principles of Data Mining, MIT Press, Cambridge, MA; 2001, p.1

8. Han J., M. Kamber, Data Mining: Concepts and Techniques (Second ed), Morgan Kaufmann Publishers, 2006, pp. 285–464

9. Kerkaus E., C. Spathis, Y. Manolopoulos, Data Mining Techniques for the Detection of Fraudulent Financial Statements, Expert Systems with Applications, Vol.32, 2007, pp. 995–1003

10. Ngai E.W.T., Yong Hu, Y.H. Wong, Yijun Chen, Xin Sun, The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of literature; Decision Support Systems, vol. 50(3), 2010, pp. 559-569

11. Phua C., V. Lee, K. Smith, R. Gayler, A Comprehensive Survey of Data Mining-based Fraud Detection research, Clayton School of Information Technology, Monash University, 2005

12. Turban E., J.E. Aronson, T.P. Liang, R. Sharda, Decision Support and Business Intelligence Systems, Eighth ed, Pearson Education, 2007

13. Wang J., Y. Liao, T. Tsai, G. Hung, Technology-based Financial Frauds in Taiwan: Issue and Approaches, IEEE Conference on: Systems, Man and Cyberspace, 2006, pp. 1120–1124

14. Yamanishi K., J. Takeuchi, G. Williams and P. Milne, On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms, Data Mining and Knowledge Discovery, Vol. 8, 2004, pp. 275–300

15. Yue D., X. Wu, Y. Wang, Y. Li and C. Chu, A Review of Data Mining-based Financial Fraud Detection Research, International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp. 5519–5522

16. Zhang D. and L. Zhou, Discovering Golden Nuggets: Data Mining in Financial Application, IEEE

Transactions on Systems, Man and Cybernetics, Vol. 34(4), 2004, pp.513-522

17. Zhou W., G. Kapoor, Detecting Evolutionary Financial Statement Fraud, Decision Support Systems, Vol. 50(3), 2011, pp. 570-576

18. Brown CE, Expert Systems in Public Accounting: Current Practice and Future Directions, Expert Systems with Applications, 3 (1), 1991, 3-18.

19. Connell NAD, Expert Systems in Accountancy: A Review of Some Recent Applications, Accounting and Business Research, 17 (67), 1987, 221-233.

20. Eining M.M. D.R. Jones, & J.K. Loebbecke, Reliance on Decision Aids: An Examination of Auditors' Assessment of Management Fraud, Auditing: A Journal of Practice and Theory, 16 (2), 1997, 1-18.

21. Gray G.L., T.E. McKee, & T.J. Mock, The Future Impact of Expert Systems and Decision Support Systems in Auditing, Advances in Accounting, 9, 1991, 249-273.

22. Green B.P., & J.H. Choi, Assessing the Risk of Management Fraud Through Neural Network Technology, Auditing: A Journal of Practice and Theory, 16 (1), 1997, 14-28.

23. Koh HC, Going Concern Prediction Using Data Mining Techniques, Managerial Auditing Journal, 19 (3), 2004, 462.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/   All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar