

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

A framework for detecting financial statement fraud through multiple data sources

Darshan Dillon and Maja Hadzic

Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology, Perth, Australia
e-mail: (darshan.dillon, m.hadzic)@cbs.curtin.edu.au

Abstract— This project deals with how to detect fraud and non-compliance in financial statements in the present day in one of the biggest economies in the world, the U.S. Since it is mainly public companies that release detailed financial information, they are the focus. This project focuses on the top five market sectors where fraud is most common. It focuses on a variety of fraud types, but not on cases of deception that do not constitute fraud. A framework will be proposed which accounts for both structured data (the numbers in the balance sheet, income statement and cash flow statement) and unstructured data (the footnotes in these financial statements). It uses ontology-driven data mining techniques to do so.

Index Terms—financial statement fraud, ontology, data mining, public companies, revenue recognition

I. INTRODUCTION

The subject of financial compliance is as old as the finance and banking industry itself. A fair and orderly marketplace has always been a concern of ruling powers in every time in history. Recently, financial statement fraud has been highly topical. There have been a number of high-profile corporate collapses including Satyam, Enron and Bernard Lawrence "Bernie" Madoff, who lost \$ 50 billion of investors money operating a Ponzi scheme in the U.S. in 2009.

Given the prevalence of fraud and the significant effects on shareholders, employees and customers, there is a need for a systematic study. This paper seeks to propose such a study using innovative and recent techniques in the field of ICT such as multiagent systems, ontology's and data mining. In this project examination will be made of the U.S. market, many types of fraud (see Figure 2), the significance of fraud, market sectors where fraud is most common, conflicts of interest in the market that make fraud more likely, the type of information that is used to identify fraud and finally, useful strategies in identifying fraud. The project includes the design of an ontology and ontology-driven data mining techniques that can be used to identify financial statement fraud.

II. AREAS OF FINANCIAL STATEMENT FRAUD PREVIOUSLY STUDIED

A review of the literature in relation to fraud in financial statements reveals that publications exist in four basic areas.

- The direct and indirect costs of fraud. The direct costs

of fraud include the funds that are lost from company accounts the company as a result of the fraud. The indirect costs could include the reduced market capitalization of the company as a result of a reduced stock price. This area of the literature has extensive information, but most of it is anecdotal rather than rigorous research. This area is outside the scope of this project but may be of interest to other researchers.

- The mechanics of financial fraud in terms of a range of different fraud types. The literature in this area is fairly extensive and is to be a basis for which categories of financial statement fraud to examine. This research builds on that.

- The relationships and responsibilities of the different players in the market who have a stake in the issue of fraud. The area most commonly covered was the audit process. The conflicts of interest inherent in the company-auditor relationship were discussed and explored in the literature and changes to legislation in this context were considered. Less frequently discussed was the need for analysts and investors to have reliable information upon which to base their decisions. Finally, there was discussion of the role that regulators play in the examination of financial statements of public companies.

- The use of technology to detect fraud in financial statements. Papers were examined that covered data mining. No data mining algorithms were found that were ontology-driven for the purpose of detecting financial statement fraud. No methods were found which also considered unstructured information of the sort found in footnotes in financial statements. During this review no comprehensive examples of an ontology specifically focused on financial statement fraud was found. In many of these areas no existing research was found, or significant gaps in the research were found. This provides a clear motivation for this research.

A. Costs of committing fraud.

In terms of the costs of committing fraud a broad range of market participants suffer losses. Firstly, there are frequent references to the relative costs of a company of committing fraud [1]. The author notes the fact that a drop in the market capitalization of the company is generally about 500 to 1,000 times the value of the fraud. Albrecht notes a particular example in which a financial fraud of \$ 7 million led to a \$2 billion drop in market capitalization. This indicates that

people that choose to commit fraud to achieve marginal gains in the stock price can end up risking destruction of the entire company if they are caught. Albrecht continues on to do a role call of companies that have been caught in the net of regulators. It is clear from the companies listed, mainly in the U.S., that the non-financial costs are even higher. Once a scandal has been reported in the national press, the reputation of the company and its directors is usually destroyed.

The second group frequently affected are employees of the company [3]. In this example, the author notes that in the case of WorldCom, for example, there were 17,000 jobs lost. Albrecht [1] also indicates that the auditors have significant problems as a result of having responsibility for clients who commit fraud. The scandal of Enron and Arthur Anderson is a case in point. After having been indicted for obstruction of justice Arthur Anderson faced action in the courts and their loss of reputation was so serious that they went bankrupt.

The area which is not covered in great depth in the literature are the losses that shareholders bear as a result of relying on fraudulent financial statements except in so far as existing investments in the company will be heavily devalued due to a collapsing stock price.

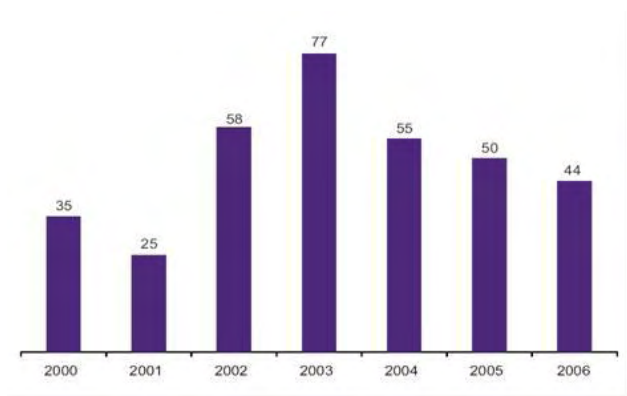


Fig.1 Total number of fraud cases prosecuted for 2000-2006 AAER S.E.C. releases from 344 AAERs in Deloitte report [5].

B. Mechanics of committing fraud.

In terms of the mechanics of committing fraud the literature included a Deloitte Consulting Company [5] report including a set of accounting specifics that commonly account for fraud. They included improper revenue recognition, improper capitalization of expenses, improper asset valuation and other items. This report is a very useful publication since it was based directly on Accounting and Auditing Enforcement Releases made by the S.E.C. between 2000 and 2006. This provides an excellent basis for analyzing existing fraud schemes and creating a method to detect them.

Powell [14] also gave some excellent broad categories of practices that permit a distinction between aggressive accounting practices and financial reporting fraud. This report talks about Managerial Intent and the fact that it can be analyzed in terms of GAAP Compliance, Materiality and

Measurement Subjectivity. The paper states about GAAP that it represents "Generally Accepted Accounting Principles which comprise the accounting rules, procedures, and practices for the recording of transactions." [14]. Notably though, GAAP does not cover every situation. Materiality is a discussion of whether an item should be disclosed in a financial statement. The basic principle relates to whether omission would effect the decision of the user of the financial statement. In terms of Quantitative Materiality "Kinney and Martin (1994) found the majority of proposed audit adjustments relate to detected misstatements that materially overstate earnings." [14]. In addition to this Qualitative Materiality covers "guidelines provide examples where it would not be appropriate to determine materiality based solely on the amount of an item." [14] The paper goes on to say that "auditors place more importance on the size, rather than the nature, of a financial manipulation." Powell & Albrecht [14, 1] give an excellent discussion of fraud related to manipulation of inventory/costs of goods sold.

Weaknesses identified in the current literature : four basic weaknesses were identified in the literature in this area: (1) There was scant literature that gave an exhaustive breakdown of the different fraud types. (2) Some of the literature gave a conceptual overview of the problem, but one that was not well suited to defining a usable solution to the problem. (3) There were infrequent use of detailed examples and (4) There was inadequate use of knowledge in the form of rules or ratios to expose fraud. Each one of these things is to be addressed in the research in an endeavor to fill the gaps.

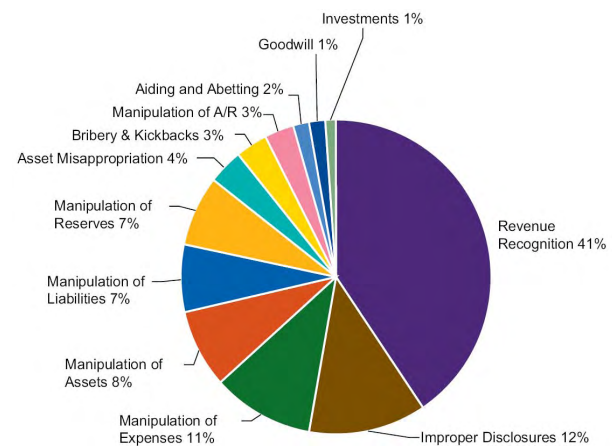


Fig.2 Breakdown of fraud schemes for 2000-2006 AAER S.E.C. releases by frequency found in Deloitte report [5].

C. Relationships and responsibilities of different players in the market.

Thirdly, the discussion of the relationships between the major players such as company directors, auditors and regulators is the most well-covered area in the literature even if papers are inconclusive about solutions. For example, Cooper [3] discusses the Sarbanes-Oxley Act 2002 and refers to some teething problems with it. It is the major piece of leg-

isolation in the U.S. that governs the responsibilities of auditors and companies engaged in the audit process. In particular, there is extensive discussion of the liability of auditors and the fact that they can be reasonably expected to do more when confronted with irregularities in the books of a company. This is an area that is to be modeled in the ontology and is to be factored into the data mining techniques used to identify fraud.

D. Use of technology.

Finally, a review was conducted of the technology to be used to build and formalize a framework. Papers were examined that featured a selection of methodologies commonly used to build ontology's. There are several methodologies that were identified from Corcho [4]. As this project proceeds a particular method for building the ontology is to be chosen. The literature for methods related to ontology evaluation (Hartmann [10]) was also reviewed. A particular method for ontology evaluation is to be chosen. The data mining conducted is to be ontology-driven. Chu [2] gave a summary of data mining methods for detecting financial statement fraud. No papers were found that focused on unstructured data in relation to financial statement fraud.

Spasic [16] discussed a variety of issues related to data mining of unstructured data. From the paper it was clear that such text mining often involves (any one or more of the following) information retrieval, information extraction and/or data mining itself. The paper presented a number of tools that are effective in one or more of these areas. Newer techniques were examined that made extensive use of a domain-centric ontology to assist with annotation and also to provide a template which is gradually populated to organize knowledge from a corpus. [17, 18, 9, 7, 15, 13, 19]

A review of ontology's and data mining algorithms that have been proposed and implemented in software tools is given in Kanellopoulos & Kingston [11, 12]. An ontology called ffPoirot was found online [6] that was indirectly related to this project. It dealt with illegal solicitation of financial instruments, fraudulent insurance claims, fraud by e-mail and illegal e-commerce activities specifically in relation to underpayment of VAT taxes. As such it was not focused on financial statement fraud as a whole but specifically on fraud related to VAT tax payment. No other ontology was found that focused on financial statement fraud. As mentioned earlier, data mining techniques were only found for detecting financial fraud in structured data. Even in this case, nothing was found implemented in a software tool.

Current weaknesses in the use of technology for fraud detection : (1) No systematic use of unstructured information was identified in the domain of financial statement fraud. (2) Research is immature in the use of data mining that is ontology-driven. Everyone is doing something different. (3) Very little actual implementation of ideas and algorithms in functional software. These three points are addressed in this research in order to fill the gaps.

III. SIGNIFICANCE

A. Social and Economic Significance

(1) Timely. There have been a number of spectacular corporate collapses that have occurred as a result of financial statement fraud in the past ten years. Fraud is on the increase and the costs are high in terms of shareholder losses and employee lay-offs. Perhaps most serious is the prospect of a loss of investor confidence in the market itself and therefore the possibility of capital flight. It is more imperative than ever that fraud is identified and prosecuted.

(2) Extensible. One goal of this research is to provide a framework within which existing fraud schemes can be conceptualized, detected and stored. It is an objective of the research to establish a repository for known fraud schemes and be able to add to it as additional fraud schemes are discovered.

(3) Relevant. Fraud schemes currently in use are to be identified. A flexible model is to be developed to define what is normal in financial statements for the top 5 market sectors where fraud is common. Financial statements are to be examined for many different size corporations. The research uses data from Accounting and Auditing Enforcement Releases made by the S.E.C. for the period from 2000 through to 2008. (see Figure 1) The data set used to test the software is to include much of this data.

B. Scientific Significance

(1) Taking account of unstructured as well as structured data. Since a financial statement may be 3 pages with 100 pages of footnotes, being able to examine both types of data and combine the results of both is essential to obtain a realistic idea of whether fraud is present in the financial statements being examined.

(2) An ontology dealing with financial statement fraud does not appear to exist. While there are plenty of ontology's in the finance & banking domain generally, there does not appear to be a single ontology that deals specifically and exhaustively with financial statement fraud in particular. To be able to classify the different types of financial statement fraud it is necessary to develop such an ontology. This is to be a major contribution and is therefore significant from a scientific point of view.

(3) Data mining which is ontology-driven to identify fraud. The data mining algorithms which are developed or identified are to be implemented in the software prototype that is built as a part of this project. Data mining will be performed in order to extract patterns from the structured data. These patterns will be used to modify the ontology. That is, ontology evolution will take place, not merely the population of ontology instances. The use of these ontol-

ogy-driven data mining techniques are a significant contribution and part of a very recent research effort in this area.

IV. METHODOLOGY

A large body of knowledge is available to be used in the study of financial statement fraud. In order to enable collective analysis of the available data, the data needs to be structured using a shareable form. Additionally, the form is used to give coherent structure to the data needs to be machine-readable. This will enable automatic programs to access, retrieve and analyze the data.

Ontology's have been proven to be a powerful tool to address abovementioned issues. They are highly expressive knowledge models and can be used to enable sharing, exchange and analyzing information from different resources by providing a consistent way of representing this information. An excellent definition of this is found in Gruber [8].

“An ontology is an explicit specification of a conceptualization” (and it is a shared conceptualization of a domain).

They capture domain concepts and relationships between those concepts, and can be used to bridge heterogeneities between different data resources in regard to their data format as well as term descriptions. Ontology's are machine-readable and computer programs such as data mining programs are able to process the ontology-structured information automatically.

A financial statement fraud ontology is to be designed to capture and represent knowledge related to financial statement fraud. This ontology is to enable consistent description of the data and support automatic data analysis such as data mining. This ontology is to assist in the interpretation of the data mining results, and will evolve as additional financial fraud schemes begin to operate in the market.

The proposed methodology consists of the following four steps :

Firstly, there is to be the specification and construction of a framework which involves data mining which is ontology-driven to examine financial statements in order to identify fraud. This will involve examining methods for building a domain-centric ontology and choosing one. Corcho [4] gave a review of methods for building ontology's. This framework will therefore include both a set of data mining techniques and an ontology. This should be able to account for structured and unstructured information.

Secondly, a software prototype is to be designed that takes financial statements as input and as output indicate whether a fraud scheme has been committed, whether the financial statements are abnormal or whether the company has been through an abnormal set of state transitions. This

design will specify the way that ontology-driven data mining techniques are to be implemented in order to detect financial statement fraud. The selection of tools to build the ontology and data mining algorithms is to be done in a flexible way.

Thirdly, evaluation of the ontology is to be performed using a well-established methodology. Hartmann [10] gave a review of methods for evaluating ontology's. It also involves randomly selecting publicly available S.E.C. data to see if the ontology is capable of capturing the full spectrum of fraud types. That is, verification of the ontology is to be done to see if it provides fairly complete conceptual coverage. Data mining rules are to be evaluated to see if they correctly identify fraud for the randomly selected data set. It is important to note that S.E.C. operatives are already using the very same publicly available data to identify financial statement fraud. While the framework may be an extension of the strategies and techniques used to identify fraud the data sources are the same as those being successfully used at the moment.

Finally, to evaluate the software produced an input set of data is to be chosen which includes cases where statements are regarded as fraudulent, and some which are not. The different outputs are to be sufficient to identify fraud for different cases in the test set. It will be particularly useful to see the correlation between cases of fraud already identified by the S.E.C. and those found by the prototype. It will also be useful to see if the tool identifies cases of fraud which were not found by the S.E.C. The results will be published in this project.

V. CONCLUSION

The basic focus of this project will be on financial statement fraud in public companies in the U.S. This project will make a major contribution in a number of ways. The research on financial statement fraud will combine an analysis of structured and unstructured data. It will develop a multi-pronged strategy to find different types of fraud and under different circumstances. A domain-centric ontology will be built that focuses on financial statement fraud in its many forms. The research will include the development of a framework of ontology-driven data mining algorithms to investigate financial statement fraud. It will develop a software product that proactively identifies financial statement fraud and implements many of the technologies and strategies above. This project aims to address every single one of these things to produce a theoretical body of research and a usable piece of software to advance the cause of identifying, understanding and prosecuting financial statement fraud.

VI. REFERENCES

- [1] Albrecht, W.S. "Business Fraud (the Enron Problem)." Brigham Young University 2005.
- [2] Chu, C-H. "A Review of Data Mining-Based Financial Fraud Research." In International Conference on Wireless Communications 2007, 5519-22: IEEE, 2007.
- [3] Cooper, J. "Financial Statement Fraud Corporate Crime of the 21st Century, ASIC ": ASIC, 2005.
- [4] Corcho, O. & Fernandez-Lopez, M & Gomez-Perez, A. "Methodologies, Tools and Languages for Building Ontology's. Where Is There Meeting Point ?" Data and Knowledge Engineering 46 (2003): 41-64.
- [5] Deloitte. "Ten Things About Financial Statement Fraud." Deloitte Forensic Centre, Deloitte Consulting, 2007.
- [6] ffPoirot. "Fraud-Based Ontology." <http://www.ffpoirot.org>.
- [7] Gaizauskas, R., Demetriou, G. & Artymiuk, P.J. & Willett, P. "Protein Structures and Information Extraction from Biological Texts: The Pasta System." Bioinformatics 19 no. 1 (2002): 1-9.
- [8] Gruber, T. "Toward principles for the design of ontology's used for knowledge sharing.", International Workshop on Formal Ontology. March, 1993.
- [9] Guarino, N. "Formal Ontology, Conceptual Analysis and Knowledge Representation", International Journal of Human-Computer Studies 43, no. 5-6 (1995).
- [10] Hartmann, J. "Methods for Ontology Evaluation ": Knowledge Web Consortium, 2005.
- [11] Kanellopoulos, D & Kotsiantis, S & Tampakas, V. "Towards an Ontology-Based System for Intelligent Prediction of Firms with Fraudulent Financial Statements." In Emerging Technologies and Factory Automation, 2007, 1300-07. , 2007.
- [12] Kingston, J. & Schafer, B & Vandenberghe, W. "Towards a Financial Fraud Ontology: A Legal Modeling Approach." Artificial Intelligence and Law 12, no. 4 (2004): 419-46.
- [13] Nédellec, C. & Nazarenko, A. "Ontology's and Information Extraction ": LIPN Internal Report, 2006.
- [14] Powell, L. "The Distinction between Aggressive Accounting and Financial Reporting Fraud : Perception of Auditors." In AFAANZ Conference 2005, 2005.
- [15] Smeaton, A.F. & van Rijsbergen, C.J. "Experiments on Incorporating Syntactic Processing of User Queries into a Document Retrieval Strategy." In International ACM SIGIR conference on Research and development in Information Retrieval, 31-51, 1988.
- [16] Spasic, I. "Text Mining and Ontology's in Biomedicine: Making Sense of Raw Text." Briefings in Bioinformatics 6, no. 3 (2005): 239-51.
- [17] Vallet, D & Fernández, M & Castells, P. "An Ontology-Based Information Retrieval Model." In The Semantic Web: Research and Applications: Springer Berlin / Heidelberg, 2005.
- [18] Yangarber, R & Grishman, R "Transforming Examples into Patterns for Information Extraction ", edited by Annual Meeting of the ACL : Workshop, 97-103, 1998
- [19] Zhang, W & Yoshida, T & Tang, X "Using Ontology to Improve Precision of Terminology Extraction from Documents." Expert Systems with Applications: An International Journal 36, no. 5 (2009): 9333-39.