Decision Support Systems 50 (2011) 595–601

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

A computational model for financial reporting fraud detection

Fletcher H. Glancy [a,*], Surya B. Yadav [b]

[a] Lindenwood University, 209 S. Kingshighway, Saint Charles, MO 63301, United States
[b] Dept. of Information Systems and Quantitative Sciences (ISQS), Rawls College of Business, Texas Tech University, Lubbock, TX 79409, United States

ARTICLE INFO

Available online 18 August 2010

Keywords:
Fraud
Financial reporting
Text mining
Quantitative model

ABSTRACT

A computational fraud detection model (CFDM) was proposed for detecting fraud in financial reporting. CFDM uses a quantitative approach on textual data. It incorporates techniques that use essentially all of information contained in the textual data for fraud detection. Extant work provides a foundation for detecting deception in high and low synchronicity computer-mediated communication (CMC). CFDM provides an analytical method that has the potential for automation. It was tested on the Management's Discussion and Analysis from 10-K filings and was able to distinguish fraudulent filings from non-fraudulent ones. CFDM can serve as a screening tool where deception is suspected.

1. Introduction

Corporate fraud has not been confined to the well-advertized cases of Enron, WorldCom, HealthSouth, etc. In all the AAERs we examined, the fraud detection was after years of abuse by senior management; and the Securities and Exchange Commission (SEC) did not detect it proactively. The continued pattern of fraud has shaken the confidence of the public in corporate America [28] both academics and auditing firms have been searching for ways to detect corporate fraud. While academic fraud research has examined many business areas [34], very little effort has been made to use quantitative approaches to examine textual data for automated financial reporting fraud detection. Phua et al. summarized the status of fraud research into four primary areas: internal, insurance, credit card, and telecommunications. In most of the internal fraud research, the object was to detect employee fraud or theft; financial reporting fraud involving senior management was not a major research focus. Most attempts to detect financial reporting fraud use financial ratios, applying various methodologies with varying results [6,28,30,31]. Phua et al. concluded that the use of unstructured data in fraud detection is essentially unexplored.

This paper proposes a quantitative model for detecting fraudulent financial reporting. The model detects the attempt to conceal information and/or present incorrect information in annual filings with the US Securities and Exchange Commission (SEC). The model uses essentially all of the information contained in a text document for fraud detection. A consistent and accurate screening tool would provide decision support for early detection of fraud; and hopefully, early detection will provide a deterrent to the commission of fraud.

In order to detect fraud, we must first define it. This would seem an easy task, but it is not always as straightforward as finding the dictionary definition [1]. We use the SEC's issuance of an Accounting and Auditing Enforcement Release (AAER) as a starting point for defining financial reporting fraud.

An AAER is an administrative proceeding or litigations release that entails an accounting or auditing related violation of the securities laws as enforced by the Securities and Exchange Commission (SEC). In the period from 2000 to 2008, the SEC issued 1700 AAERs. In the period from 2006 to 2008, they issued 555 AAERs. In this analysis, we use the term fraud when referring to an AAER to be a litigation release of an accounting or auditing violation where the SEC used the word fraud in describing the violation. We examined a sample of 74 AAERs from this period that charged companies with fraud. They showed that the average time between the identified initial fraud and the SEC filing charges was 7.26 years with a range of 3.6 to 11.6 years. The SEC charged that these companies committed fraud for an average of 4.1 years with a range of 1 to 12 years.

Churyk et al. [13] used qualitative content analysis of the required Management's Discussion and Analysis (MDA) part of the 10-K SEC filings to identify fraudulent filings. They were able to identify deceptive cues. SEC filings include several areas of text in addition to the MDA. All companies are required to explain anything in the operation that could have a significant impact on the future profitability of the company. Management's explanation is in the MDA; the accountant's explanation is in the notes to the consolidated financial statements. It has been proposed that if a company were to report information that would have potentially negative results on the company valuation, they would include it in the notes. The reason to include the information is that it is legally required, and failure to do so is a criminal act that could result in a long incarceration. Both fraud and failure to properly report are crimes; but reporting gives some protection to the accountants and auditors, which leaves only senior

* Corresponding author.
E-mail address: fglancy@lindenwood.edu (F.H. Glancy).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.08.010

management at risk. The idea that information can be concealed in either the MDA or in the notes, as it can be said "in plain sight", leads to the problem statement and subsequent research questions.

1.1. Problem statement

The status of fraud detection from the analysis of corporate financial statements can be described as follows:

1. Fraud detection is after the fact and officially recognized only after the SEC issues an AAER.
2. There are few proven quantitative methods to detect a potentially fraudulent filing [28]. We did not find any quantitative method that examined the text of the filing.
3. A high level of senior management compensation comes from stock options. The individual manager has high potential return from financial fraud and a low potential for detection. This leads to the conclusion that corporate fraud is likely to continue at the senior executive level [22].

There is little question that the incentive to commit fraud at the senior executive level exists in spite of the laws passed since the Enron failure [22]. From Agency Theory, the most common executive compensation methods promote fraud at the senior executive level. The options backdating scandals that came to light in 2006 re-enforces that the incentive for executive fraud continues to exist [35]. It has been proposed that the "New Economy" makes fraud a normal circumstance because of the requirement to predict and deliver continually increasing earnings [39]. With the apparent widespread financial reporting fraud, there is a need for improved detection.

There has been minimal use of data mining in the investigation of fraud in financial reporting [29]. Text mining is a subset of data mining that has been used infrequently with varying results [5,12]. The Churyk et al. [13] finding that a qualitative methodology can provide an indication of fraud begs the question of creating a quantitative analytical model for text that has the potential for automation. Text mining is an obvious possibility. "(Text mining) is useful because it provides an efficient, quantitative representation of each document." [2]. There is a need to put text mining into context to produce positive results [18]; this requires that supervised text mining is used for the initial model. This indicates that there is a potential for its use in fraud detection, but that new techniques will be required. We know that there are indications of fraud in the 10-K filings [37]. This is based on the requirement of all companies to explain anything in the operation that could have a significant impact on future profitability. This requirement creates a situation where, if the 10-K writers know of an instance or situation that can have an impact and do not report it, they knowingly deceive the reader [22]. This knowing deception has been shown to create internal conflict that manifests itself in several ways [23]. The 10-K writer has an internal conflict because the failure to report is a crime as is fraudulent reporting. However, reporting honestly can result in a dramatic decrease in stock price [39] and a personal loss of wealth as well as employment. Another possibility exists, to report the instance but to do so in a manner to reduce the apparent effect on the future business by use of affect modifiers and other linguistic methods. The concealing or covert revealing of information in financial reporting leads to our research issues.

1.2. Research issues

This article investigates the following issues:

1. Can the fraud be detected from the text of annual corporate SEC filings?
2. Can a quantitative and computational-based fraud detection model be developed that will provide a methodology for automating detection of potential fraud?

We propose an answer to these questions by creating a model for detecting financial reporting fraud in corporate annual SEC filings and testing it using additional SEC annual filings. The research approach used for this investigation was based on the Unified Research Methodology [4] and design science [25].

The paper is organized into the following six major sections. Section 2 reviews the financial reporting fraud literature, discusses the potential theoretical basis for fraud detection in text, and a tool used for pattern detection in text. In Section 3, we develop the computational model. In Section 4, we test the model. Section 5 presents the results of the testing and analyzes the results. In Section 6, we present the contributions of this work, and Section 7 presents our conclusions and potential future work.

2. Relevant literature

There are two separate research literature streams providing background and basis for this research; they are research into corporate financial reporting fraud and text-mining research. Both of these research areas are extensive, but there has been very little crossover research. In order to understand the basis for assuming it is possible to detect fraud from text, we examine the body of research into deception. First, we look at the theoretical basis for deception detection, fraud detection, and then at text mining.

2.1. Deception detection

McCornack [32] created the Information Manipulation Theory (IMT) using Grice's cooperative principle of communication and the maxims of expected quality, quantity, relevance, and manner (or clarity). The deception occurs when there is a covert violation of one of the maxims. IMT states that the four principles are independent, and violation of a single principle defines a deceptive communication. IMT was tested and the replication of McCornack's work showed that the maxims were not independent and that any deception would violate the quality maxim [26]. Intuitively, this is logical because violations of the quality maxim involve distortions or fabrications of information.

Burgoon and Buller [9] proposed the Interpersonal Deception Theory (IDT) as an explanation of how deceivers react when interacting with a person that is the target of deception. They define deception as "a deliberate act perpetrated by a sender to engender in a receiver beliefs contrary to what the sender believes is true to put the receiver at a disadvantage." They later refined deception as "a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver" [8]. They found that "deceivers were more uncertain, and vague, more non-immediate and reticent, showed more negative affect, displayed more arousal and non-composure, and generally made a poorer impression than truthtellers. Their behaviors also connoted greater formality and submissiveness." They found that deceptive interactions are dynamic; they change over time as the deceiver attempts to manage image and the victim's interpretation. The presence of deception and degree of suspicion affected the target individual's behavior and the degree they mirrored the deceiver's behavior. Empirical tests have confirmed the dynamics of interaction and the manifestations of deception. A main principle of IDT is that the face-to-face communication is interactive, and the deceiver alters the deception based on the feedback received from the receiver [10]. Burgoon et al. show empirically that interaction is a primary contributor to IDT; and without interaction between the deceiver and their subject, the deception is very difficult to detect. We find that other researchers use IDT as a theoretical basis for deception detection in other media.

George et al. [20] used IDT in combination with Media Richness Theory (MRT) as the theoretical basis for examining if alerting the receiver of the possibility of deception in CMC would increase the