Chapter 18
Fraud Detection and Corporate Filings

Sunita Goel
Siena College, USA

ABSTRACT

This chapter is focused on detection of fraud in organizations by using content-based analysis on the annual reports issued by firms. Unlike a variety of previous work on fraud detection that have used quantitative financial information, this research examines qualitative textual content in annual reports to decipher evidence of fraud embedded in these reports through careful examination of the tone, content, and emphasis across reports. The basic premise of this research is that organizations tend to camouflage negative findings to sound less damaging. The real intent of the writer is hidden in content but can be revealed through structured content analysis. Using a corpus of annual reports of companies where fraud has occurred and juxtaposed with companies where fraud has not been detected, this study systematically examines the differences in the use of language. The results of this study reveal that fraudulent annual reports exhibit themes of optimism, variety, complexity, activity, and passivity. On the other hand, non-fraudulent annual reports exhibit themes of certainty and realism.

INTRODUCTION

Detecting financial statement fraud is a difficult task as it occurs in less than 5 percent of all financial statements. However, the fact that financial statement fraud is relatively rare does not diminish the serious consequences that it can have on all stakeholders when such cases arise. Furthermore, the likelihood that fraudulent misstatements will be discovered on a timely basis is low as companies deliberately try to conceal them. Misstatements occur when financial statements are not presented in conformity with generally accepted accounting principles or when there is intentional falsification.
of financial statements and related disclosures to hide the effect of fraud. Unfortunately, there is no foolproof method of preventing financial statement fraud, but it is possible to discover evidence of fraud even when there is no prior suspicion. This is particularly important in light of the fact that when fraud is first suspected, it is likely to be more serious than it may initially appear. This research is focused on detection of fraud in organizations by using content-based analysis on the annual reports issued by firms. Previous research evidence suggests that textual content of corporate reports contains useful information that can be used to detect fraud (Cecchini et al., 2010; Goel et al., 2010; Humpherys et al., 2011; Goel and Gangolly, 2012). In particular, this research empirically examines whether themes associated with fraud can be detected from the text of corporate narratives using content analysis technique.

Content analysis is a well-known technique for making inferences by objectively and systematically identifying specified characteristics of a text (Holsti, 1969). In extant literature, it has been used for systematic and objective quantitative analysis of linguistic expression (written, spoken, or visual). It is typically used to investigate specified words and concepts within texts; it also allows for the systematic examination of large volumes of data with relative ease.

Content analysis methodology has also been used for efficient analysis of large textual datasets as it uses structured techniques to analyze data and make valid inferences from text (Weber, 1990; Neundorf, 2002; Krippendorff, 2004). Content analysis is not restricted to the manifest meaning but focuses on latent meaning as well (Holsti, 1969). In annual report narratives, explanations are often expressed within the framework of accounting language. The content analysis procedure involves interaction of two processes. The first involves the specification of content characteristics to be measured and the second process involves the application of rules for identifying and recording the characteristics when they occur in the data. These measurements serve as a reference for drawing inferences. Content analysis may involve both the structure and content of information. Structural (format) categories include length of the document, paragraph length, placement of information etc. Content related categories are often subjective—such as deception, tone, hostility, pleasure, anger, fear etc. An important decision in content analysis is the unit of analysis i.e. for narrative (verbal or written) it could be words, sentences, phrases, or paragraphs.

The rest of the chapter is structured as follows. The next section discusses the relevant literature highlighting key findings of earlier studies. The subsequent section presents an overview of research methodology, sample of qualitative data that is used and results of the experiments. The last section presents concluding observations indicating directions for future research.

BACKGROUND

The early studies on the use of qualitative textual content of corporate disclosures were hamstrung by the lack of large databases of financial statements, the lack of software for linguistic analysis, and lack of hardware powerful enough to support latest advances in multivariate statistical techniques in text and data mining. For the same reasons, there were virtually no uses of such qualitative analysis in the detection of fraud. We next present a brief account of the various studies that involves qualitative analysis of textual content of financial reports.
