




May 16th, 11:00 AM

Detecting Deception in Asynchronous Text

Fletcher Glancy

Oklahoma State University - Main Campus, fletcher.glancy_iii@okstate.edu

Follow this and additional works at: <https://commons.erau.edu/adfsl>

 Part of the [Business Law, Public Responsibility, and Ethics Commons](#), [Forensic Science and Technology Commons](#), [Information Security Commons](#), and the [Social Control, Law, Crime, and Deviance Commons](#)

Scholarly Commons Citation

Glancy, Fletcher, "Detecting Deception in Asynchronous Text" (2017). *Annual ADFSL Conference on Digital Forensics, Security and Law*. 2.

<https://commons.erau.edu/adfsl/2017/papers/2>

This Peer Reviewed Paper is brought to you for free and open access by the Conferences at Scholarly Commons. It has been accepted for inclusion in Annual ADFSL Conference on Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

EMBRY-RIDDLE
Aeronautical University™
SCHOLARLY COMMONS

(c)ADFSL



DETECTING DECEPTION IN ASYNCHRONOUS TEXT

Fletcher H. Glancy
Oklahoma State University
Spears College of Business
Stillwater, OK 74078-4011
fletcher.glancy_iii@okstate.edu

ABSTRACT

Glancy and Yadav (2010) developed a computational fraud detection model (CFDM) that successfully detected financial reporting fraud in the text of the management's discussion and analysis (MDA) portion of annual filings with the United States Securities and Exchange Commission (SEC). This work extends the use of the CFDM to additional genre, demonstrates the generalizability of the CFDM and the use of text mining for quantitatively detecting deception in asynchronous text. It also demonstrates that writers committing fraud use words differently from truth tellers.

Keywords: asynchronous communication deception, CFDM, deception, text mining, financial reporting fraud, identity theft, deception detection

1. INTRODUCTION

Deception is the intentional misleading of another such that they draw an inaccurate conclusion (Buller and Burgoon, 1996; Carlson, George, Burgoon, Adkins, and White, 2004; DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H., 2003). Fraud is criminal deception as defined by the laws of the country. Research has found deception and fraud to be important questions because the costs of deception and fraud are extensive. Investors lost billions of dollars when financial reporting fraud was uncovered at Enron, WorldCom, and Broadcom. Some of these investors were pension plans; the loss resulted in reduced pensions of the plan members. Many individuals lost their jobs, which cost taxpayers' unemployment compensation. While these three are widely known examples of financial reporting fraud, they are certainly not

the only ones. And yet, fraud is difficult to recognized. Often times, fraud will be committed for several years before it is discovered. Madov's Ponzi scheme lasted for more than 14 years before discovery (Markopolos, 2005).

A writer of deceptive text should know that the writing is deceptive. The problem has been how the reader can detect deceptive writing. In the past, many researchers attempted to find cues that indicate that writing is deceptive. DePaulo, et al. (2003) looked at 1338 different studies that attempted to find cues to that would confirm deception; 158 deception cues were identified (DePaulo et al., 2003). These cues were grouped into five different sets of predictions relating to liars and truth tellers. The predictive sets were: liars will be less forthcoming than truth tellers; liars will have more discrepancies than truth tellers; liars will be less positive and pleasant

than truth tellers; liars will be tenser than truth tellers; and liars' stories will have fewer common imperfections and less unusual content than truth tellers. The cues to these predictive sets were divided into the following groups: verbal, vocal, facial, active (body positioning), quantity, time, rate of speaking, pauses, types of words used, voice amplitude, nervousness, and other behavioral cues.

In the meta-analysis, twenty-five cues were statistically significant; and of the twenty-five, only thirteen were relevant to text (DePaulo et al., 2003). The cues relevant to text were shorter response length, providing fewer details, providing less sensory information, blocking access to information, increased response latency, less logical structure, greater internal discrepancy, fewer self-references, less immediacy (actives vs. passives and affirmatives vs. negatives), increased tentative constructions, less cooperative, more negative, and less contextual embedding. The cues only provide a potential method for detecting deception.

Glancy and Yadav (2010) developed a computational fraud detection model (CFDM) that successfully detected financial reporting fraud in the text of the management's discussion and analysis (MDA) portion of annual filings with the United States Securities and Exchange Commission (SEC). The method used text mining and clustering of the singular value decomposition (SVD) (Albright, 2004) of the document-term matrix. The CFDM was able to classify a MDA as fraudulent or non-fraudulent. The results were very highly significant, $p = 0.0059$. A sample of thirty MDAs—ten fraudulent and twenty non-fraudulent—were tested. Nine out of ten fraudulent MDAs tested as fraudulent. The company that was a false negative was retested using the MDA of the following year, and the retest was positive (i.e., fraudulent). Sixteen of twenty non-fraudulent MDAs tested as non-

fraudulent. The false positives may have been unidentified fraudulent reports; it is not possible to tell whether they are true false positives. The result was that the CFDM was very successful at classifying companies as fraudulent and non-fraudulent. The CFDM was quantitative, and it did not require human interpretation of the text.

We use the CFDM methodology (Glancy and Yadav, 2010) to answer the question: 'is the CFDM generalizable to other genres than the MDA?' We consider this question in the remainder of this article.

Section 2 discusses the CFDM in more detail and the methodology used to create the model. Section 3 applies the methodology to create models for detecting deception in three additional genres. Section 4 discusses conclusions, limitations, and future work.

2. COMPUTATIONAL FRAUD DETECTION MODEL (CFDM)

The CFDM (Glancy and Yadav, 2011) was created for detecting financial reporting fraud by text mining the management's discussion and analysis (MDA) of a company's annual report. The process used to create the CFDM is described below.

The first step was selecting fraudulent and non-fraudulent companies. The SEC's accounting and auditing enforcement releases (AAER) were reviewed. The SEC issues an AAER either during or at the conclusion of an investigation of accounting or auditing misconduct. The AAER states the reason for the SEC's charge of misconduct; these will normally require the company to restate their financial report. The companies that were selected as fraudulent were the ones the SEC accused of fraud, and they either admitted the fraud or were convicted of fraud in Federal Court. Most companies that committed

financial reporting fraud did so over a period of years. The annual report used for text mining was in the period the SEC identified in the AAER. If a company knows that they are under investigation by the SEC, they must include that knowledge in the annual report. The MDA selected was for the latest reporting period that did not mention an SEC investigation. This period was chosen because it was unlikely that the company knew at this point that the SEC was investigating them. The report that first mentioned the investigation was normally in the year prior to the SEC issuing the AAER, although in some cases the company knew of the SEC investigation for several years before the AAER was issued.

For each fraudulent company selected, a non-fraudulent company was selected. The first criterion for a non-fraudulent company was that they had not filed an amended financial report with the SEC in the 10 years preceding or proceeding the reporting year chosen for the fraudulent company. The second criterion was that they had the same SIC code (i.e., they were in the same industry). The third criterion was that they were approximately the same size. This criterion applied only if there was more than one company that met the first two criteria.

The second step was to prepare each MDA for text mining by eliminating all non-text items. These included tables of numbers, notes to tables, all headers and footers, punctuation, and proper nouns. The individual MDAs were saved as text files.

The third step was creating the document set and text mining. All documents, both fraudulent and non-fraudulent, were imported in SAS[®] Enterprise Miner[™] (EM) software. Common words also called stop words were eliminated (e.g., a, an, the, and, for). Punctuation was removed and all capitalization was converted to lower case. The

words were converted to terms by stemming. The process of stemming reduces the word to its base (e.g., absolutely becomes absolute, finalize and finally become final). After stemming, the words are referred to as ‘terms.’ Stemming as used in this text mining also identifies synonyms and separates words that have different meanings depending on the part of speech (e.g., a bank is a noun and synonymous with a financial institution, bank as a verb would refer to turning an airplane). Text-mining is performed by starting with the term document frequency matrix and then calculating the single value decomposition (SVD) (Albright, 2004) of the term-document matrix. The SVDs are clustered (de Ville, 2006; Gao and Zhang, 2005; Roiger and Geatz, 2003). Each SVD identifies a document. The clustering is bottom up (hierarchical) based on the distance between the SVDs. The identity of the documents remains after the SVD is created, but the individual terms in the document are not visible. However, the text-mining program can calculate the probability of a term being in a cluster and report the terms with the highest probability for a cluster. The explanation of the SVD is beyond this paper, but Albright (2004) provides a complete description of the SVD and its use in text-mining.

The fourth step is analyzing the results. The model had very good separation between the fraudulent MDAs and the non-fraudulent MDAs with 66 of 69 documents clustering correctly. The three that did not cluster correctly were false positives—companies meeting the non-fraudulent criteria but clustering with the fraudulent companies. The model is based on the SVDs of the documents used to create it. The model is used to test other MDAs to determine if they are fraudulent or not. When a fraudulent document is added to the original set of documents and the SVD process is repeated, it

should test as fraudulent by clustering with the fraudulent documents. A non-fraudulent document should test as not fraudulent and cluster with the non-fraudulent documents.

The model was used to classify an additional 31 MDAs, eleven fraudulent and twenty non-fraudulent. The companies were chosen using the same criteria that was used for the original fraudulent and non-fraudulent companies. Ten of the eleven fraudulent MDAs were correctly identified. Sixteen of the twenty non-fraudulent MDAs were correctly identified. The results were evaluated with the sign test (Conover, 1999), and the p -values for both tests were less than 0.01.

3. APPLICATION OF THE CFDM PROCESS

The CFDM was successful in detecting fraudulent MDAs. In order to demonstrate that the CFDM methodology is generalizable we investigate its use in an additional area in financial reporting and in fraudulent unsolicited email (spam). The CFDM process was given a different test in financial reporting fraud. The financial reporting fraud test was on the text of the notes to the financial statements of annual reports. The MDA is normally written by one of the chief executives of a company; we expect him to know if the report is fraudulent. Accountants write the notes to the financial statements; we do not know if the accountant who wrote the note knew that the report was fraudulent, but we would anticipate that a good accountant would at least suspect that the financial statements were incorrect. The similarities between the notes to the financial statements and spam are that both are asynchronous and the authors of notes to the financial statements in an annual report are essentially anonymous to the reader as are the authors of the spam.

We used spam e-mails because many unsolicited e-mails (spam) are fraudulent. The global cost of spam is estimated at \$130 billion annually (Jennings, 2009). The cost to US businesses is estimated at \$42 billion annually. We did two tests on spam. One spam test was on e-mail that attempted to steal identity, and the other test was on spam that attempted money theft. Both financial reporting fraud and fraudulent e-mails are asynchronous and relatively anonymous. In both domains, the reader can only make assumptions about the identity of the writer. Both are available electronically. Both are sent with the expectation of deceiving the reader. Both are often more than just deception; they are fraudulent.

An estimated 62 trillion spam are sent annually. Spam filters are biased toward legitimate e-mail and detect it at a rate of 99.99%. The bias toward detection of legitimate e-mail and preventing false positives or legitimate e-mail identified as spam (Yih, Goodman, and Hulten, 2006) causes the filters to pass on the recipient up to 5% of the spam. Of the estimated 62 trillion spam e-mails sent annually, spam filters will pass up to 3.1 trillion spam e-mails to the recipient's inbox.

In each of the three tests, the fraudulent documents were matched with legitimate documents from the same genre and in the same period. The total number of documents in each test varied from 69 to 110. The documents were analyzed using the CFDM methodology as described above (Glancy and Yadav, 2011). The sample size for each study was chosen for a statistical power over 90% with an effect size of 0.20 to 0.30 (Cohen, 1988). The CFDM was used to create the single value decomposition of the term-document matrix and to cluster the documents. The text mining clusters the document set into as many clusters as appropriate for the document set; the

maximum number of clusters for a document set is user specified and was set to forty for all studies.

3.1 Annual Financial Report Notes Test

We were able to use the same data set as Glancy & Yadav (2010). Instead of using the

MDA, we extracted the notes to the financial reports and followed the CFDM methodology. The results of the model were highly significant, p-value less than 0.001 with 3 false positives and three false negatives. The results are shown in Table 1.

Table 1.

Clustering results of the notes to annual financial reports test of the CFDM methodology.

Clustering Results	Correctly Classified	Incorrectly Classified
	63	6
Percentage correctly classified	91.3%	
False Positive	3	
False Negative	3	
p-value	2×10^{-13}	

The notes from 11 fraudulent annual reports and 20 non-fraudulent annual reports were tested using the notes model. Each set of notes was tested individually by adding them to the document set and then re-running the model. One of the 11 fraudulent documents was a false negative and three of the 20 non-fraudulent were false positives. The p-value was less than 0.001.

3.2 Spam Email Tests

The two tests on spam were on email that either tried to steal identity or money. We sent a request on Facebook for people to collect email that escaped their spam filter and made it into their inbox. The individual determined that the email was stealing identity or money and labeled it as such and forwarded it to the authors. Over 40 people responded and over 2000 money-theft spam and over 3000 identity-theft spam were collected. The two tests are discussed below.

3.2.1 Money Theft Spam Email Test

In order to compare money theft email to legitimate email that requests money, over 2,000 legitimate e-mails were collected from legitimate sources that asked the recipient for money (e.g., The Susan G. Komen Foundation, the March of Dimes, and the American Cancer Society). The spams and legitimate e-mail used in the study were randomly chosen from those collected. Originally the model was created with fifty-five spams and fifty-five legitimate e-mails, which were saved as plain text and put into a database. The results of the model were very highly significant, p-value less than 0.001, and are shown in Table 2. The number of emails was increased to 100 for each type of email (spam and legitimate) to increase the probability of an email being incorrectly identified and to increase the possible number of clusters. The model was replicated 20 times with the 100 emails randomly chosen from legitimate emails and 100 randomly chosen from money theft spam. The results in each test were very highly significant, p-value less than 0.001. In each trial, the number of

clusters was two, even though the allowable number of clusters was set to 40.

Table 2

Clustering results of the notes to money theft spam test of the CFDM methodology.

Clustering Results	Correctly Classified	Incorrectly Classified
	104	6
Percentage correctly classified	94.54%	
False Positive	2	
False Negative	4	
<i>p</i> -value	3.300×10^{-24}	

3.2.2 Identity Theft Spam Email Test

In order to compare the spam email to legitimate email, over 2,000 legitimate e-mails were collected from the sources that were spoofed by the spam, such as banks, investment companies, and credit card companies. The spams and legitimate e-mail used in the study were randomly chosen from those collected. The first model used fifty-five spams and fifty-five legitimate e-mails, which

were converted to plain text and put into a single database. The CFDM methodology was applied to the single database and the results were very highly significant, *p*-value less than 0.001, as shown Table 3. This test was also replicated 20 times with 100 randomly chosen spams and 100 randomly chosen legitimate email. The results were consistent with *p*-values for each model less than 0.001. Again, the allowable number of clusters was set to 40, but in every case the number of clusters was 2.

Table 3

Clustering results of the notes to identity theft spam test of the CFDM methodology.

Clustering Results	Correctly Classified	Incorrectly Classified
	107	3
Percentage correctly classified	97.27%	
False Positive	1	
False Negative	2	
<i>p</i> -value	3.325×10^{-28}	

3.3 Results of the CFDM Tests

The three tests of the CFDM process replicated the results of the original CFDM (Glancy and Yadav, 2011) in two domains and three genres. The CFDM created in each genre was successful at differentiating fraudulent text from legitimate or non-fraudulent text. The percentage of correctly identified documents

exceeded 91% in all tests. The average of the three tests was 94.37%. These results exceeded those of Goel, Gangolly, Faerman, and Uzuner (2010), using a support vector machine methodology; the average correctly identified in their three tests was 88.84%.

4. CONCLUSIONS, LIMITATIONS AND FUTURE WORK

All three tests on asynchronous text used the CFDM methodology to create models that were able to successfully distinguish legitimate asynchronous text from fraudulent text. In all cases the number of clusters was two, fraudulent and non-fraudulent. There were no additional clusters. This confirms the accuracy of the hierarchical clustering of the SVDs. This also confirms that the CFDM methodology is generalizable beyond financial reporting using the MDA. Repeating the CFDM methodology 20 times on money-theft spam and 20 times on identity-theft spam confirmed that the CFDM methodology is repeatable and has potential for developing a quantitative method of filtering email. This is especially significant when considering that the email used in the tests had already escaped detection by conventional email filters.

A potential limitation of this work is that only two genres were used for the testing, financial reporting and spam. Future work can include expanding the CFDM to additional asynchronous genre to further confirm the generalizability of the CFDM methodology. Another limitation is that although we have a quantitative method of detecting fraud in asynchronous text, we do not know why it works. The use of the SVD in text-mining does not allow inspection of the terms in each cluster and their frequency. It has been experimentally proven to be successful, but at this time we do not know what is in the text that allows fraudulent documents to cluster together when using the singular value decomposition of the reduced document text matrix. We can conclude that people committing fraud use words differently from those who are telling the truth. We can further suggest that the way those committing fraud

use words is subconscious; because if they were aware of the way they were using words, they would modify their writing to conceal the fraud. Further work will utilize the results from these tests to determine if a theoretical basis for explaining the success of CFDM can be developed.

REFERENCES

- Albright, R. (2004). Taming Text with the SVD. SAS Institute White Paper. Retrieved from <http://www.sas.com/apps/whitepapers/whitepaper.jsp?code=SDM5>
- Buller, D. B., and Burgoon, J. K. (1996). Interpersonal Detection Theory. *Communication Theory*, 6(3), 203-242.
- Carlson, J. R., George, J. F., Burgoon, J. K., Adkins, M., and White, C. H. (2004). Deception in Computer-Mediated Communications. *Group Decision and Negotiation*, 13(1), 5-28.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edition ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Conover, W. J. (1999). *Practical Nonparametric Statistics* (Third Edition ed.). New York: John Wiley & Sons, Inc.
- de Ville, B. (2006). *Decision Trees for Business Intelligence and Data Mining: using SAS Enterprise Miner* (1st ed.). Cary, NC: SAS Institute Inc.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129(1), 74-118.
- Gao, J., and Zhang, J. (2005). Clustering SVD strategies in latent semantic indexing. *Information Processing and Management*, 41, 1051-1063.
- Glancy, F. H., and Yadav, S. B. (2011). A Computational Model for Financial Reporting Fraud Detection Decision Support Systems, 50(3), 595-601.
- Goel, S., Gangolly, J., Faerman, S. R., and Uzuner, O. (2010). Can Linguistic Predictors Detect Fraudulent Financial Filings? *Journal of Emerging Technologies in Accounting*, 7, 25-46. doi: DOI: 10.2308/jeta.2010.7.1.25
- Jennings, R. (2009, February 10, 2010). Cost of Spam is Flattening - Our 2009 Predictions. Retrieved from <http://www.ferris.com/?p=322011>
- Markopolos, Harry (November 7, 2005). "The World's Largest Hedge Fund is a Fraud" (PDF). *Wall Street Journal Online*. Archived from the original (PDF) on August 18, 2009. Retrieved December 22, 2017.
- Roiger, R. J., and Geatz, M. W. (2003). *Data mining: a tutorial-based primer* (1 ed.). New York: Addison-Wesley.
- Yih, W., Goodman, J., and Hulten, G. (2006). Learning at Low False Positive Rates. Paper presented at the CEAS 2006 - Third Conference on Email and Anti-Spam, Mountain View, CA, USA.