How Not to Write a Class Action "Reform" Bill

By John C. Coffee, Jr.

Understanding Runs in the Shadow Banking System

2 *By Kathryn Judge*

When Does Criminal Liabil Insider Trading Sense?

By John P. Anderso

THE CLS BLUE SKY BLOG
COLUMBIA LAW SCHOOL'S BLOG ON CORPORATIONS AND THE CAPITAL MARKETS

Detecting Risk Through Firms' Emails

By Sanjiv Das, Seoyoung Kim and Bhushan Kothari February 27, 2017

Comment

Recent advances in financial technology (FinTech) have dramatically transforme financial landscape with respect to the way we access, invest, and transfer financ In our recent article, we explore a promising avenue for the use of natural-langu processing in an effective yet non-invasive method by which to monitor the health and integrity of financial institutions and corporations in gen

Specifically, the continuous flow and abundance of corporate emails makes them a far more prevalent and timely indicator of escalating risk or than the numbers in quarterly financial statements. However, difficulties lie in systematically and efficiently drawing inferences from large amo unstructured text. On one hand, manual parsing ensures proper comprehension within the relevant context and field; however, this method is no intensive, but also raises substantial privacy concerns. Thus, a platform built to detect escalating risk or potential malfeasance, without actually individual employee emails, is highly valuable from a practical and ethical point of view. Furthermore, changes in the network of email interact also indicate sources ripe for further investigation, as fraudulent activity among corporate employees tends to occur in clusters.

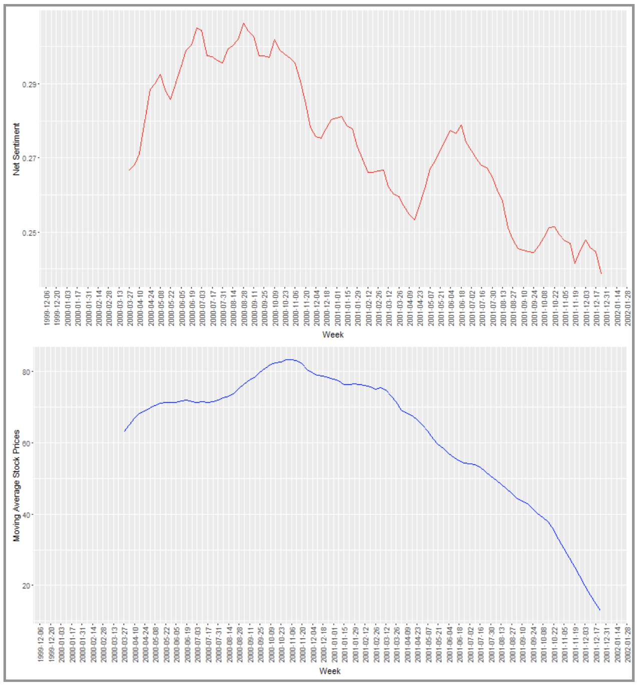
In an exploratory analysis, we develop an automated platform to parse a set of unstructured employee emails—specifically, Enron emails spann critical period from January 2000 through December 2001,[1] which is the two year period leading to Enron's demise. Based on the content ext these employee emails, we compute a net sentiment score over time using various context-dependent sentiment dictionaries for word classificat

In examining the net sentiment score alongside Enron's stock price over time, we observe that the net sentiment computed from email content g follows stock-price movements, with a marked decline in both measures toward the end of the Enron lifecycle (see **A1**). Interestingly, we also o the average email length also declines dramatically with stock prices over time (see **A2**), and in supplementary statistical analyses, we observe t length is a stronger predictor of subsequent price declines than the net sentiment conveyed by the message body itself. Perhaps as corporate risk malfeasance escalate, emails become increasingly shorter, as employees are less likely to include specific details in emails sent via the corporat

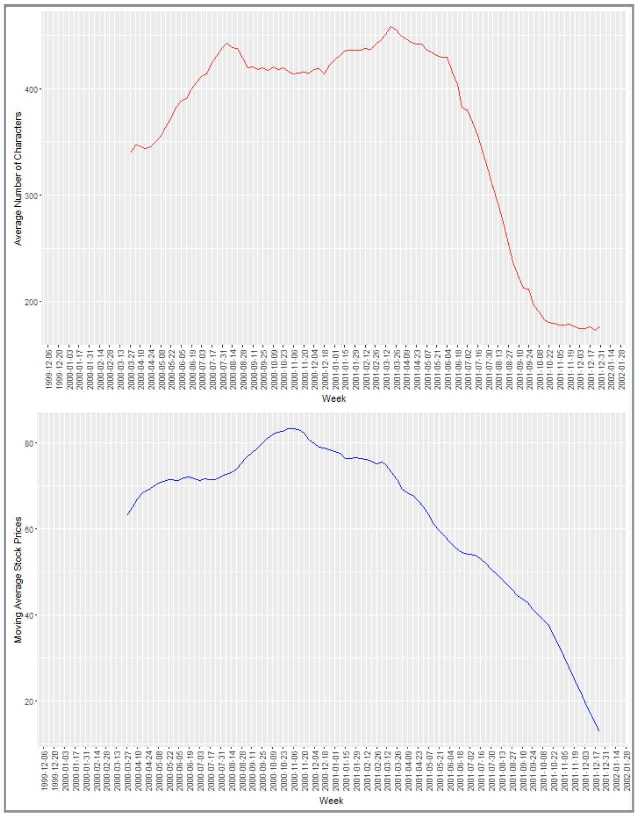
In future areas of exploration, we are interested in applying various models to uncover trending topics. For instance, generative statistical mode latent Dirichlet allocation (LDA), are predicated on the assumption that a given topic has varying probabilities of employing certain words. Thu techniques have the potential to expose important shifts in topics that may otherwise remain undetected, particularly in more novel, esoteric, or intentionally obscure subject matter, whereby specific phrases or vocabulary have not yet made their way into screening dictionaries.

We are also (and perhaps even more) interested in the non-verbal, interaction-based indicators of potential trouble, since email content may be e control or manipulate than the shifting patterns within the email network. For instance, changes in overall network connectedness or changes in employees are more pivotal players in the email network may be ripe sources for further exploration.

Overall, we suggest the promise of a regulatory technology (RegTech) approach by which to systematically parse email content and network str detect indicators of risk or malfeasance on a more timely and ongoing basis.



A1. Moving average stock prices and moving average net sentiment from email content over time (from Das, Kim, and Kothari (2017



A2. Moving average stock prices and moving average email length over time (from Das, Kim, and Kothari (2017))

ENDNOTE

[1] The Enron emails were first made publicly available by the Federal Energy Regulatory Commission (FERC) during its investigations into E practices, and were subsequently hosted on a dedicated site by Carnegie Mellon's computer science department. Over time, the emails have unc further cleaning as part of a redaction effort due to legal reasons as well as requests of affected employees. Thus, our study should be viewed as prescriptive manual, with implications for parsing, organizing, and analyzing unstructured email content, rather than as a positive study as to th workings of the Enron employee network.

This post comes to us from professors Sanjiv Das and Seoyoung Kim of Santa Clara University and Bhushan Kothari of Google. It is based on t article, "Zero-Revelation RegTech: Detecting Risk through Linguistic Analysis of Corporate Emails and News," available here.

Leave a Reply

Your email address will not be published. Requ are marked *

Comment

Name *

Email *

Post Comment

Columbia Law School Home | About | Contact | Subscribe or Manage Your Subscription

Powered by WordF