INTA 6450 spring 2021 Group 2 Course Project

Identify emails representing signs of wrongdoings from Enron email dataset

Hongru Liu, Min Zhang,
Wenyan Du, Zhuotong Liu

Apr 2021

1

Group Project Timeline

- March 8 – 14: Read through each others proposal
- March 14: First meeting, decided on timeline and plan
- March 15 – 21: To explore which method would generate the best results, implemented individual methods
- March 21: Second meeting, compared notes
- March 22 – 28: Further improvement on individual implementation
- March 28: Third meeting, compared individual results and decided on one topic to focus and combined everyone's method together.
- March 29 – Apr 4: Finished combining everyone's method and running for final results.
- Apr 4: Fourth meeting, confirmed final results and started on final report.
- Apr 5 – Apr 11: Finished writing and revising report.
- Apr 12– Apr 18: Finished constructing and revising PowerPoint.
- Apr 19 – deadline: Finished presentation video.

Definition of Wrongdoing: Abuse of Special-Purpose Entity (SPE)
What is SPE and what did Enron do with SPE?

The Special Purpose Entity is a fenced organization having limited predefined purposes and legal personality. [1] SPEs are typically utilized by companies to isolate the firm from financial risk.

- Enron used special-purpose entities (SPE) to hide the company's debts from investors and creditors.

- They used SPEs such as Chewco to transfer and hide their troubled assets.

- They used SPEs to "reward" employees.

1. Sainati, Tristano; Brookes, Naomi; Locatelli, Giorgio (2016-09-19). "Special Purpose Entities in Megaprojects: empty boxes or real companies? Literature Review". Project Management Journal. 48 (2): 55–73.

Analyze Enron's Email Corpus for the Abuse of SPE
Data cleanup

- ElasticSearch has certain limitations - it only returns up to 10,000 records each time.

- Python - After parsing those json files and conducting some data manipulation, we got a Pandas dataframe with 251,734 rows and 11 columns.

Analyze Enron's Email Corpus for the Abuse of SPE

Network Analysis

People of Interest (POI) recognition:

- Started with 4 individuals: Kennth Lay, Jeffrey Skilling, Andrew Fastow, and Ben Glisan.
- A network was modeled to show the relationship between people.
- 13 neighbors were found that have at least 5 emails with the 4 POI. (17 in total).
- We filtered the dataset and reduced the number of emails from 251,734 to 4,288.
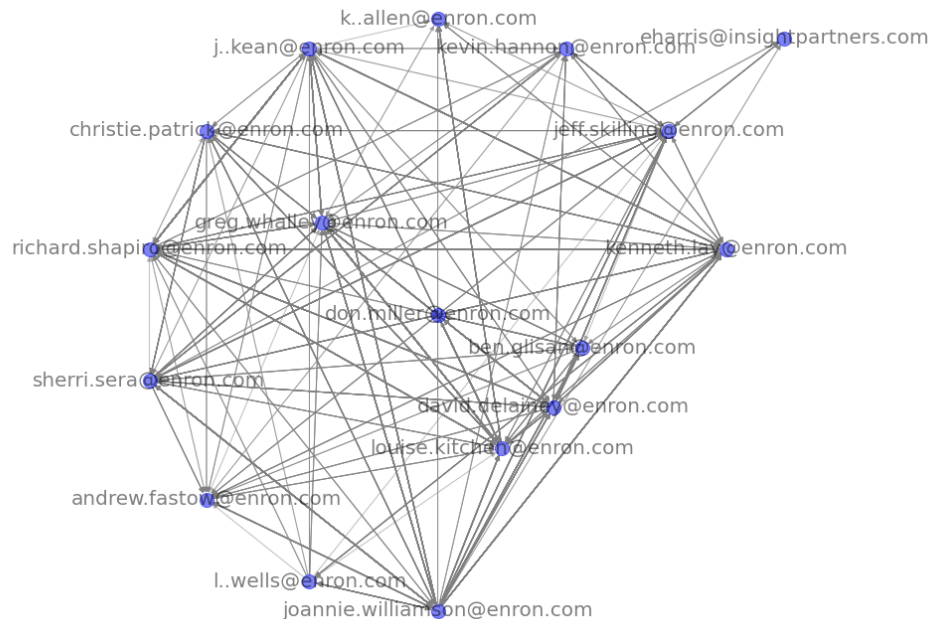
Analyze Enron's Email Corpus for the Abuse of SPE
Topic Modeling

- A topic modeling algorithm helped us to understand major topics of the emails and determine what emails may be relevant.

- Python NLTK library contains various utilities that allow us to effectively manipulate and analyze linguistic data. (tokenize, stem, lemmatize)

- We chose the number of topics to be 20 and build the Latent Dirichlet Allocation (LDA) model.

- The topic that has keywords such as "transaction", "trading", "legal" and "contract" was chosen.

Topic 20	"transaction" "enrononline" "trading" "credit" "system" "trade" "agreement" "product" "application" "data" "risk" "online" "counterparty" "commodity" "eol" "legal" "customer" "term" "contract" "question"

Analyze Enron's Email Corpus for the Abuse of SPE

Keywords Filtering

- Ben Glisan was identified
- After analyzing Ben's emails, we figured that "Raptor", which is associated with SPE, is the most critical word on this topic.
- Therefore, we applied "Raptor" as a filter word to our selected 4,288 emails from POI derived from step 2. It generated a list of 12 emails.

	from	to	cc	bcc	subject	text
2694	david.delainey@enron.com	wes.colwell@enron.com richard.causey@enron.com	ben.glisan@enron.com joseph.deffner@enron.com ...	ben.glisan@enron.com joseph.deffner@enron.com ...	Raptor et al	Rick, we have re-examined the portfolio as a r...
7188	ben.glisan@enron.com	richard.causey@enron.com rick.buy@enron.com ro...	david.gorte@enron.com george.mckean@enron.com ...	david.gorte@enron.com george.mckean@enron.com ...	RE: Raptor Debris	George & Gordon\n\nPlease work with RAC (Dave
26474	david.delainey@enron.com	mike.jakubik@enron.com			Re: Raptor	How do we immunize ourselves from the workouts...
38702	david.delainey@enron.com	richard.lydecker@enron.com	jesse.neyman@enron.com brandi.morris@enron.com	jesse.neyman@enron.com brandi.morris@enron.com	Brigham	Great job guys - keep it coming!\n\nRegards\n...
86200	david.delainey@enron.com	david.oxley@enron.com			Re: ENA Comp suggestions for Project Raptor	David , if this is going to cause you politica...
93255	david.delainey@enron.com	mike.jakubik@enron.com	raymond.bowen@enron.com	raymond.bowen@enron.com	Re: Raptor	That may work - I don't want to end up with a...
102229	david.delainey@enron.com	wes.colwell@enron.com joseph.deffner@enron.com			Catalytica	Guys, I suggest that we write it down in order...
119164	andrew.fastow@enron.com	john.lavorato@enron.com	louise.kitchen@enron.com	louise.kitchen@enron.com	RE:	John:\n\nThanks for the e-mail. I wasn't tryi...
126920	j..kean@enron.com	.palmer@enron.com			FW: Erroneous press	\n\n -----Original Message-----\nFrom: \tKoeni...
185833	ben.glisan@enron.com	greg.whalley@enron.com			Raptor Position Reports for 11/20/00	Greg,\n\nAttached are the daily position repor...
231041	david.delainey@enron.com	wes.colwell@enron.com	mark.frevert@enron.com john.lavorato@enron.com...	mark.frevert@enron.com john.lavorato@enron.com...	ROCE and Cost Iniatives	Wes, as per our discussion, I wanted to detail
243594	david.delainey@enron.com	joseph.deffner@enron.com david.gorte@enron.com...	andrew.fastow@enron.com rick.buy@enron.com lar...	andrew.fastow@enron.com rick.buy@enron.com lar...	AIG Fund	Guys, I would like to add the following commen...

Identified Emails

We selected 5 emails that show our target wrongdoing obviously. The index numbers of 5 emails are 38702, 86200, 93255, 102229, 185833.

- Email ID: 38702
  - Enron used raptor to hedge ENA and their specialists worked hard to balance severely distressed assets, which indicates that Enron exploited SPE in balancing their account.
  - SPE is not supposed to be used to balance mark-to-market losses.

- Email ID: 86200
  - David suggested that they pay compensation/rewards to VPs and directors assigned to the project using money in Raptor.
  - Using SPE, Enron was permitted to increase leverage without reporting debts on its balance sheet. However, it is improper to not follow regulations in using SPE.

Identified Emails

- Email ID: 93255
  - As Delainey wrote in this email "raptor may be a good accounting hedge", Enron employees exploited SPE as a hedge against their losses and SPE was abused to hedge their JEDI investment.

- Email ID: 102229
  - Here, Richard mentioned the value of Enron's Catalytica investment is affected by raptor capacity, not real earnings.
  - Enron's use of SPE violated Generally Accepted Accounting Principles (GAAP) SPE requirements.

- Email ID: 185833
  - Enron took the use of SPE to increase their credit capacity. This email indicates that the credit capacity of Raptor is $94 million.

Was your set of emails good?

- Yes! They all indicated our target wrongdoing.

Problems to Fix Next Time

- POI
  - Original dataset was shrunk 60 times based on POI filtering
  - We may loss some critical info or kept some useless info
  - Could consider change the initial POI or add a few more people in there

- Keyword selection
  - We were fortunate to found keyword "Raptor"
  - Filtering by "Raptor", we have 12 emails in total
  - We could also try to change the keyword or add more keyword if no obvious wrongdoing emails were found

Problems Unfixable

- Limitation of topic modeling
  - Topics were uncorrelated.
  - LDA does not model sentence structure.
  - No ground truth to validate the results.

Because of these limitations, we didn't get an expected topic group until topic 20.

Unless we implement supervised learning method together with topic modeling, the problem probably won't be fixed.

Summary

- We focused on abuse of Special-Purpose Entity (SPE) as the wrongdoing

- Our strategy includes 4 steps:

    Data preparation
    POI recognition – network analysis
    Keyword selection – topic modeling
    Fraud email identification – keywords filtering

- We were able to identify 5 emails for our target wrongdoing

- Change of initial POI and keyword selection could be the problems we fix next time

- We can't fix the limitations of topic modeling, but we can implement supervised learning to strengthen our method

Thanks for listening!

- Thanks for course instructor Jeff Borowitz, TAs and our peer students!

- All team members from group 2 wish you a happy summer!

- Stay safe and healthy!