Georgia Institute of Technology

INTA6450 Spring 2021

Enron Course Project

Part 1, Proposal

Wenyan Du

Group: 2 Username: wdu63

February 15, 2021

1 Wrongdoing – Abuse of Mark to Market Accounting

1.1 Mark to Market (MTM) Accounting

Mark to market is a method in accounting to estimate the value of assets and liabilities. It could be an accurate method as it is evaluating the company's current financial situation based on current market conditions. However, during unprecedented times, it may not precisely reflect the market. [1]

1.2 Enron's abuse of MTM

There were many wrongdoings in accounting that contributed to Enron's fall. For example, as mentioned in Dr. Borowitz's lecture about Enron, the company divested large amount of its activities to special purpose entities(SPEs). These SPEs activities were off the balance sheet and used to support Enron's income numbers[1, 2]. Compared to that, abuse MTM is the most astonishing wrongdoing I found from the book: "The smartest guys in the room: The amazing rise and scandalous fall of Enron"[3]. Below I listed a few cases according to Gwilliam (2008) [1]:

First, Enron marked the full market value of assets even when the projects were incomplete. Thus, that marked value was actually exaggeration of the true value. One example is when Enron recorded MTM revenue for Cuiaba power station in Brazil. It was recorded as $34 million and $31 million in the third and fourth quarters of 1999. However, the power station and the pipeline were never finished.

Second, Enron tended to mark the market value when the value would increase and less likely to record when the value went down. This intentional selectivity would definitely introduce bias in the finance report. For example, Enron recorded the increased stock held by JEDI in 2000, but didn't mark decreases in stock value in 2001.

Third, Enron incorrectly recognized income and cash flow via marking to market. En-

[1] https://www.investopedia.com/terms/m/marktomarket.asp

ron signed long term contracts (15 to 20 years) with external companies. The long term interests were evaluated and wrote into current market value. Then, it was recognized as earnings and cash flow. However, since Enron pre-assessed market values that would spread over the next 15 to 20 years, the actual market value for the current year could be less than that. Therefore, their so claimed valuation of earnings and cash flow didn't reflect the real numbers. This was both happened to Blockbuster and Eli Lilly. Here is a quote from the paper: "Thus, within the space of about one year, this investment which resulted in Enron reporting $111 million of gain and $115 million of funds flow from operations in the fourth quarter of 2000 and the first quarter of 2001, proved to be worthless." [1]

To conclude, even though MTM could be a rather accurate accounting method, Enron did find ways to abuse MTM, to artificially make their financial report looking nice. The key characteristics of this wrongdoing is that the marked value was improperly exaggerated and did not match with the real numbers. This is one of the key signs to look for in the next session of the proposal.

2 Signs of wrongdoing in emails

2.1 People involved

Emails sent or received by people related to or involved with MTM should get warning signs. These people are:

1) Enron's finance and/or accounting team

2) Enron's senior management team or whoever outside the finance/accounting that can also decide the MTM valuation

3) External companies that were influenced by MTM

2.2 Keywords

Keywords such as "mark to market", "MTM", "M2M", "market value", "revenue", "value of assets" could be on a general watch list. Within emails that having above words, we can further look for words like "record", "evaluate", "estimate", "set". Then, we can go through each of the MTM abuse case to identify words specifically related to that case. For example, "Cuiaba power station" and "incomplete" or "never finished" could be the keywords for the first MTM abuse case described in part 1.2 of this proposal.

2.3 Monetary numbers

As stated in the conclusion in part 1.2 of this proposal, the most direct evidence of MTM abuse is the mismatch between marked value and the reality. Therefore, as a special key word, monetary numbers in these emails from/to above list of people with above keywords are very important to identify this wrongdoing. For example, $111 million of gain and $115 million of funds flow could be signs of MTM abuse from the third case described in part 1.2 of this proposal. Here we should not only include the number, but also the dollar sign and word "million" as part of the sign.

3 Strategies to find emails with wrongdoing

Below is the general strategy:

First, since Enron started to use MTM from 1992 [2], I would select all emails dated starting 1992. Then, filter emails to only include the ones sent/received by people related to MTM.

Second, use the pre-identified keywords (including the monetary numbers) to select and label a small set of emails as "with wrongdoing" or "without wrongdoing". This is the training dataset.

Third, utilize natural language processing tools to pre-process all the emails. (details described in method section)

Fourth, apply supervised learning method to the training dataset to train the model. (details described in method section)

Fifth, apply trained model on the rest of the email.

Sixth, randomly select a few emails labelled by the model as "with wrongdoing" to manually check the model accuracy.

4 External information

Based on above sections, I list external information in order to find abuse of MTM:

1) List of personnel involved with MTM (both internal and external). This will help me to filter through sender and/or receiver of the emails.

2) Detailed cases of Enron MTM abuse. With this information, I can find emails specific to these cases. If I have information such as the year and month the case happened or specific people that were involved, I can find these related emails easier from the large dataset.

5 Methods

5.1 Data pre-processing

In Trivedi (2019) [4] paper, they used Enron dataset (ver. 5 & 6). An email in this dataset has a header and a body. In the header section, there are date, time, email sender and receiver, subject and content type. In the body section, there is the detailed content of that email.

Data pre-processing apply mostly to subject in header and the body section. We can directly use date, time, sender and receiver information in the dataset.

Referenced from Trivedi (2019) [4], I propose steps for data pre-processing:

1) Tokenization: Same as what we did for the python exercise, the first step is to break strings into words.

2) Stop-word removal: Prepositions, conjunctions, articles can be removed.

3) Lemmatization: Transfer words into basis forms. This includes removing the present particle or past tense and convert the capital letter into lowercase (This is practiced in the python exercise).

4) Representation: Generate a vector for each email based on a dictionary. As the word vector is usually very sparse, dimensional reduction methods will be used. This is one set of the features prepared to train the model.

5) To get additional feature, apply sentimental analysis to the emails and keep the output keywords for each email as another feature. Depends on which tool is selected, data pre-processing may or may been needed.There are plenty of free and paid software/tools for sentimental analysis. [2][3][4]

6) Concatenate the features generated from 4) and 5). Now the dataset is ready for data processing.

5.2 Data processing

Apply supervised learning algorithms to pre-processed training dataset (manually labelled emails). Algorithms suitable to this dataset could be SVM, random forest, boosting or regression.

---

[2] https://opennlp.apache.org/
[3] https://monkeylearn.com/
[4] https://www.meaningcloud.com/

6 Why my methods may not work & what to remedy

6.1 Why my methods may not work

There are multiple reasons that my methods may not work:

1) For supervised learning projects, the label of each data point is the ground truth. The label in this case would be whether an email contains wrongdoing (labelled as "with wrongdoing") or not (labelled as "without wrongdoing"). If the number of labelled emails are not enough for the training dataset or if the label is not accurate, there is a chance that the model will not be well trained. Then, the model won't be able to properly identify emails with wrongdoings when apply to the rest of the dataset.

2) During data pro-processing, I proposed to use vector to store words in an email and then dimensional reduction methods will be utilized to simplify these features. This helps to save running time and resources. However, if too many features has been get rid of, the model may not be well trained.

3) I proposed to include key words generated from sentimental analysis as additional features. However, these key words may not related to wrongdoings.

6.2 What to remedy

1) I will need to increase the number of emails labelled and enlarge the population for the training data set. At the same time, make sure the manual labelling is as accurate as it could be.

2) I can try to reduce different numbers of features. Hope to find the balance point where it won't influence the running and also there are enough features to build a reliable model.

3) Since sentimental analysis is generating very few additional features, not contributing to the majority of the dataset. If these features won't help with refining the model, I would discontinue using sentimental analysis.

References

[1] David Gwilliam and Richard HG Jackson. “Fair value in financial reporting: Problems and pitfalls in practice: A case study analysis of the use of fair valuation at Enron”. In: *Accounting Forum*. Vol. 32. 3. Elsevier. 2008, pp. 240–259.

[2] A Rashad Abdel-Khalik. “How enron used accounting for prepaid commodity swaps to delay bankruptcy for one decade: The shadowy relationships with big banks”. In: *Journal of Accounting, Auditing & Finance* 34.2 (2019), pp. 309–328.

[3] Bethany McLean and Peter Elkind. *The smartest guys in the room: The amazing rise and scandalous fall of Enron*. Penguin, 2013.

[4] Shrawan Kumar Trivedi and Shubhamoy Dey. “A modified content-based evolutionary approach to identify unsolicited emails”. In: *Knowledge and Information Systems* 60.3 (2019), pp. 1427–1451.