

1. Discuss past and future government applications
2. Develop a more comprehensive related works section for Ch. 2
 - (a) Setup context more effectively
 - (b) More on general machine learning techniques
 - (c) Discuss analogous research

3. Investigate big spikes in the time graph

Spikes represent Gmail folders being ported to the Hume center using an old script, which was responsible for changing the timestamps. A more sophisticated system was developed for transferring emails in Summer 2014, and there are no more spikes after this point. The biggest spike (from Jan 2014) came from moving large folders from a single graduate student. Based on the folder names, these are unrelated to the Hume center and were sent before this person was a graduate student, therefore these emails will be removed from the dataset for the purposes of this analysis.

4. Calculate measure of non-symmetry of Hume adjacency matrix
 - (a) Briefly explain the case of people emailing themselves - is it from CC-ed emails?

5. Look into mutual information values of zero and explain

There were four significant digits available for each value, so it did not appear to be a rounding error. There seemed to be an arbitrary cutoff point around 0.22 bits. Lack of documentation makes it difficult to pinpoint the cause or correct for it, so I developed my own script to calculate the mutual information based on the description in Section 4.2. Table 4.1 has been updated to reflect the changes. Many of the features from the table were unchanged, but some are new.

6. Is there any source of data leakage? If not, explain disparity between classification and LOOCV results.
7. Discuss a naive guess/baseline for classification
8. Look into using a prior to represent the class distribution
9. Add more commentary on LOOCV results

- (a) What types of people are misclassified?

10. For hierarchy analysis:

- (a) Take direction into account for hierarchy connections
- (b) Add more edges

I changed this analysis to only consider sent emails. I added edges (increasing in transparency) to represent 2nd and 3rd strongest connections. These changes are shown in Figure 5.7. The more complex structure of the Hume center is now evident, but some of the divisions remain clear.

11. Add collective classification to future work