

Team TextMiningMasters

Members: ndeena2, ajaw2, pwasal3, jiahuah2

Project: Reproducing a Paper, Mining Casual Topics in Text Data

1) **Progress made thus far:**

We have gathered and prepared the necessary data in order to perform one of the experiments stated in the paper. More specifically, we have parsed the NYT corpus and time series data from Iowa Electronic Markets 2000 Presidential Winner-Takes-All market using Python and stored the data into a data structure.

2) **Remaining tasks:**

Now that we've gathered and prepared the data the next step is to implement the iterative topic modeling with time series feedback algorithm stated in the paper, and feed in the data that we have prepared into that algorithm. Then the next step is to perform some data visualization to understand our results.

3) **Any challenges/issues being faced:**

Some parts of the algorithm are unclear to us. More specifically, steps 2 and 3. We are unsure on how to compute $\text{sig}(C, X, T)$ in step 2, and how to apply granger tests to each of the candidate topics for step 3.