

Project Progress Report

Team: Best Bots

S.No.	Task	Current Status	Remarks/Impediments
1	Access to DataSet by all members	Completed	
2	Understanding the dataset	Completed	How to deal with gaps in dataset. Raised @1273 on Piazza
3	Find parallel time series dataset	Completed	
4	Finding the libraries required for implementation	Completed	
5	Perform Preprocessing of data	90% Completed	
6	Perform LDA on news data set	10% Completed	Currently performing LDA on May data
7	Apply Granger Test to determine causality relationship.	75% completed	Matrix multiplication not providing expected results. Trying to do this using for loops
9	For each candidate topic apply Causality measure to find most significant causal words among top words in each Topic.	75% completed	Matrix multiplication not providing expected results. Trying to do this using for loops
10	Record the impact values of these significance words using Pearson correlations	Not Started	Currently being done in Granger test. Yet to implement Pearson coefficients. We may not need it.
11	Separate positive impact terms and negative impact terms	75% completed	Implementation is done. Testing in progress with actual data.
12	If orientation of words in prior step is very weak, ignore minor group	75% completed	Implementation is done. Testing in progress with actual data.
13	Assign prior probabilities proportions according to significance levels	75% completed	Implementation is done. Testing in progress.

14	Apply LDA to Documents using prior obtained	10% Completed	Research on parameters for gensim LDA.
15	Repeat until satisfying stopping criteria (e.g. reach topic quality at some point, no more significant topic change).	Completed	In our experiment iteration count is 5
16.	Make final project documentation report.	Not Started	
17.	Software usage tutorial presentation	Not Started	