CS 410 Progress Report

11/29/20

ludy2

Topic:

Reproducing a paper- *Mining Causal Topics in Text Data: Iterative Topic Modeling with Time Series Feedback*

Progress Made:

All progress thus far has been related to planning and researching. Collection of data from NY Times articles for evaluating the proposed algorithm has been completed. I have also done some research and thinking on how I will go about implementing the proposed algorithms (PLSA implemented based on the Lemur information retrieval toolkit) to run on the data.

Remaining Tasks:

Implement the proposed algorithms in python, run code on dataset with and without guidance from the time series, as well as with tweaked parameters (number of topics and strength of the prior), evaluate and analyze results and compare with results from the published paper. To do the data analysis, I will evaluate performance using Grainger tests. I will then be able to determine if I have successfully reproduced the results found in the paper, that "ITMTF finds topics that are both more pure and more highly correlated with the external time series than typical topic modeling, especially with a strong feedback loop."