Mining Causal Topics with Iterative Time Series Feedback DEMO

Faith Chung
CS 410
Fall 2020

Setup

Python Libraries used

```
import os
import sys
import xml.etree.ElementTree as ET
from gensim import corpora, models
from nltk.tokenize import RegexpTokenizer
from stop_words import get_stop_words
import numpy as np
from pandas import DataFrame
from statsmodels.tsa.stattools import grangercausalitytests
from collections import Counter
from scipy import stats
from datetime import datetime
import math
```

1. Install python modules

Most of the previously listed modules should be in stdlib. However, please install these modules if you have not already:

pip install gensim
pip install nltk
pip install stop_words
pip install numpy
pip install pandas
pip install statsmodels
pip install collections
pip install scipy

2. Clone git repo

In your terminal, execute:

git clone https://github.com/chungfaith1/CourseProject.git

3. “cd” into cloned repository

```
[Faiths-MacBook-Pro-49:CourseProject fchung1$ ls
 README.md              iem_data               log_12-13-2020.txt      nyt_data
 documentation          log.txt                main.py
[Faiths-MacBook-Pro-49:CourseProject fchung1$
```

4. Execute python script

In your terminal, execute:

python main.py

Note: This script will take 15-20 min to run

Demo

https://youtu.be/K2cQH4pPyBk

Results

1. log.txt

After execution is complete, you should see log.txt in the repository

```
[Faiths-MacBook-Pro-49:CourseProject fchung1$ ls
README.md               iem_data                log_12-13-2020.txt      nyt_data
documentation           log.txt                 main.py
[Faiths-MacBook-Pro-49:CourseProject fchung1$
```

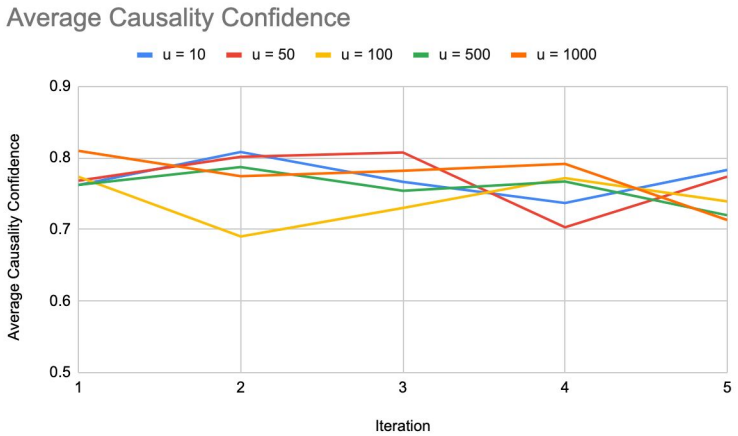
2. Log.txt (cont)

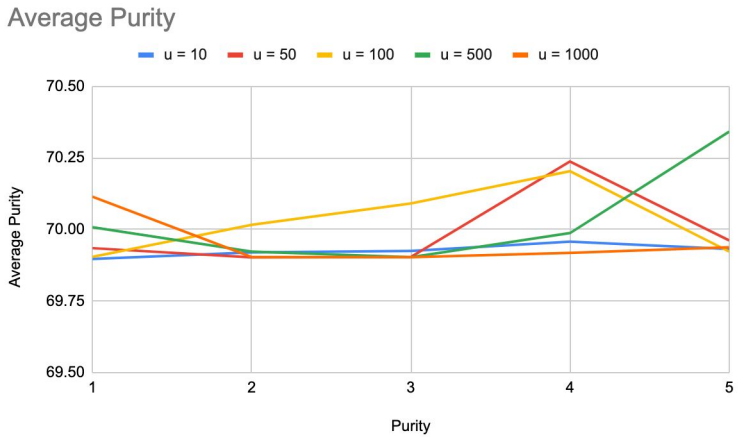
The log contains data for each iteration, which has a specific topic prior strength and number of topics.

Data include iteration #, avg causality confidence, avg purity, and top words from sig. yopics

```
iteration: 1
Average Causality Confidence: 0.8082756237415565
Average Purity: 69.91960032086311
Top words:
s, will, new, w, tax, gov, republican, lieberman, texas, one,
s, w, gov, republican, tax, today, one, new, t, republicans,
s, will, t, one, security, w, gov, new, social, government,
s, w, tax, gov, will, texas, one, republican, plan, state,
s, w, debate, gov, republican, one, voters, m, new, policy,
```

Avg. Causality Confidence and Purity - u





Avg. Causality Confidence and Purity - tn