Classifying Political Affiliation of Tweeter Users using Sentiment Analysis and Topic Models

- Mitali Nandargikar
- Saurabh Katkar
- Akshaya Hebbar

Problem Statement

- Ebola virus a cause of concern escalating to the level of panic among some people in United States.

- Use of sentiment analysis and topic modeling to analyze the level of concern by examining the tweets, and lexicon analysis and perspective bias to determine the political affiliation of the user.

Hypothesis

- A user's views and concerns about the Ebola virus might be influenced by
  - the news
  - the political, media personalities
  - close friends

- The said user's tweet will subsequently reflect the level of concern that he/she associates with the virus.

Overview of Method

- **Sentiment Analysis**
  - Usage of sentiment analysis along with topic models in understanding both text and links in Twitter.

- **Lexicon Analysis and Perspective Bias**
  - Users with similar political views may likely share a similar dialect and will mostly use the same choice words. Lexicon analysis filtered with perspective bias will help define affiliations

Overview of Method

- **Developing a Relational Topic Model**
  - For determining connections that exist between the tweeter user and other user characterized by the popularity and similarity between neighbors.

- **Classification**
  - **Generative Learning** (Latent Dirichlet Allocation (LDA)) – examining topic models
  - **Discriminative Learning** (Logistic Regression) – examining derived network structure.

    --- Based on different parameters, determine which is the better one.

Data

- Using REST API, the twitter data would be collected and filtered with the terms related to the project.

- **Attributes:**
  - Scores based on: Lexicon analysis, perspective bias, dialect used, probability of the user using certain catch words etc.

Related Work

- **Modeling Microblogs using Topic Models – Kriti Puniyani**
  - Usage of topic models to know whether two users tweet about similar topics is more useful in predicting links between them than standard network analysis metrics that ignore text

- **An analysis of perspectives in interactive settings - Dong Nguyen, Elijah Mayfield, Carolyn P. Rosé**
  - Analyzes a contributor's perspective bias through their lexical choice in political forum debates.

Tentative Timeline

- October 20: Project Proposal
- October 20 - 30: Data Collection
- November 21 - 6: Data Organization
- November 7 - 13: Implementation of Functions
- November 14 - 21: Debug, analysis of Final output
- November 22 - 28: Final Conclusion and review
- December 3: Project Presentation