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

by

Mitali Nandargikar

Saurabh Katkar

Akshaya Hebbar

Problem Statement:

The advent of Ebola virus in Africa and the news of a few cases being reported in the US has brought a cause of concern escalating to the level of panic among some people. Using sentiment analysis and topic modeling, we analyze the sentiment depicting the level of concern by examining the tweets of such users and coupled with lexicon analysis and perspective bias we 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 user is subjected to, the political, media personalities and close friends and their say on the issue. 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. Twitter hashtags would play an important role to capture meaningful associations that exist between topics.

- ## Lexicon Analysis and Perspective Bias

  Each political ideology has certain words that define the ideology. These words are called 'lexicons'. Also the twitter users will often show 'perspective bias' thus making sure they present a strong point on their side. This lexicon coupled with perspective bias will help measure the political orientation of text.
  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.

- ## 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.
  Eg: Person follows Ann Coulter, Sarah Palin, George Bush, Fox News…. Then right wing.
  Eg: Person follows Jon Stewart, Bill Maher, Hillary Clinton, MSNBC…… Then left wing.

- ## 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.