Computer Science meeting Social Science

* Why?
  + President’s stance in military
  + Bill passing?
  + Ideology?
  + User’s stance
  + Changing of ideology -> prediction?
* Data-driven
  + GovTrack.us
  + Social Medias
    - Twitter
    - News (comments)
    - Online forums

Methodology

* Detection
  + Legislative Voting Network
    - * ­­­­Congress roll call data
        + Fed legislation -> the House & Senate -> Law
        + Data -> Network

Columns (Groups)

Voters, Bills, Words (content)

Arrows / connection

Voters -> Bills (yes, no)

Connection to words

Defining problem: algorithm

Input: legislative network

Output:

x(u): ideal points for politician u

a(d): ideal points for bill d

Coloring of data set -> political parties

Different lines (y axis) –> political issues

* + - Why?
      * Voters: different positions for diff topics (either left or right)
      * Trad matrix factorization -> no meaning for each dimension (line/topics)
      * Topics of bill -> biased – influencing politician’s voting
        + Voting behavior -> guiding the topics of bills (group)
        + Ex: health – subtopics (service and cost)

Often contradictory

* + - Topic-factorized IPM
      * Entities (columns): politicians, bills, terms
      * Links: vectors – connect each entity
        + (P, B) and (B, T)
        + Parameters to max: ideal points for pol & bills -> topic models - mapping
      * Voting (behavior) – prediction using intuitions
        + More similar of ideal points u and d -> higher prob of yea link -> higher weight
        + Alignment of ideologies of voter and bill

Predicting the actual distribution of the bills

1D: yes or no – simple -> predicting using prob.

* + - * Combination (linear combination of two avg log function over word links and voting links)
        + Optimization
        + Regulation -> fitting of data
    - Data Description
      * Roll call data of the legislative branch
      * Keep the version of the latest -> test the model with past data to see if it matches the newest data
        + Machine learning -> find the best fit to the result
      * Case studies
        + Different presidents’ stances of diff issues
        + Scatter graphs

Two different issues (overlap)

Questions: what’s the axis number?

Real number – sides of politics

* + Twitter
    - Input: nodes of people (subset of Twitter network)
      * Link: follow, mention, retweet
    - Output: ideology & relation weights
      * Follow: information, retweet: endorsement of ideologies
      * Weights of links -> scaling by importance to ideology
    - Challenges:
      * Model behaviors link to ideology?
      * Combining different types of behaviors?
    - Assumptions:
      * People following, mentioning, or retweet others with similar ideology score (biased)
      * True ideology score != public image ideology score
    - Example
      * Multiple types of link – vector
        + p(i): true ideology of user i
        + q(j,r): image ideology of user j for link type r
        + b(j,r): bias term of user j for link type r -> popularity of user j in link r
        + Function (linear combination)
      * Max the weighted sum of the likelihood function for each link
    - Data collection
      * 1. Seed users collection (manually collection from politicians)
      * 2. expand from seeds (followers and followees for each seed)
      * 3. filtering (set I – political related ~ 20+ followers/followees) vs set II – less political relate
    - Values: ranking and classification
      * Method, ranking accuracy, classification AUC (given a pair or link -> result if it exists ~ high accuracy percentage)
    - Case study: ideology map of U.S. – avg of each state
      * Geolocation info
    - Ideology of specific accounts (popular agency, politician)
    - Ideology of presidential candidates
* Joint Modeling of Content and Social Interaction
  + Analysis of text
    - News, online forums
  + Users input (key words) – linked between users -> output: user stances
    - Required contrasting topics -> visualization
  + Intuitions to the solution: different stances – different wording for same issue
  + Case study
    - CNN News Comments
* Co-evolution model of social network and opinion migration
  + Linking of people with similar ideology
  + Definition of similarity
    - Dot product similarity
    - Circular distance of ideologies
  + Gravity-based score functions
    - Opinion leader – more connection and less extreme ideology ->
      * Actors influenced by neighbors -> groupthink
  + Simulation
    - System level parameters
    - Small noise, fewer friends (communities -> opinion divergency
    - Big noise – random opinions – larger sparsity parameter
    - increase # of friends (diversity) + strong opinion leaders -> similar opinions that differ from others-> disagreement
    - Fit to congress co-data set
      * Use both to predict future results
* Why this presentation?
  + Address social interactions
    - Ubiquitous of online behaviors
    - Integration