Ideology Detection for Twitter Users via Heterogeneous Types of Links

Yupeng Gu, Ting Chen, Yizhou Sun, Bingyu Wang
Northeastern University

Quiz! Can we predict whether the following people are conservative or liberal, simply based on who they follow/mention/retweet on Twitter?

- Alice follows 100 Democrats and 0 Republicans.
- Bob follows 100 Democrats and 90 Republicans.
- Carly *follows* 100 Democrats and *retweets* from Republicans for 100 times.
- Dave *follows* 32 Democrats, 18 Republicans; *mentions* Democrats 56 times, Republicans 79 times; *retweets* Democrats X times and Republicans Y times …

Problem Definition

Given a heterogeneous Twitter network with multiple types of links (*follow*/*mention*/*retweet*), our goal is to:

- Learn the political ideology for all users in the network. For example, can we predict whether a citizen is liberal/conservative/neutral? Where is a citizen's position in the political spectrum? Is Democrat U more liberal than another Democrat V?
- Determine the strength of each type of links in terms of ideology detection. For example, does *retweet* behavior reveal one's ideology more than *mention* behavior?

Model – Multi-Link Ideal Point Estimation Model (ML-IPM)

Our model is based on the following assumptions:

- Network homophily: people with similar ideology tend to have links in between.
- In directed networks, each user has two roles: link-sender and link-receiver, which correspond to two different features. When a user issues links to others (e.g. choose friends), her intrinsic feature (ideology) will be considered; when a user receives links, her impression feature will be considered.
- The weight of each link type should reflect its strength in explaining the corresponding network.

The probability of a link from *i* to *j* of link type *r*:

$$p\left(G_{ij}^{(r)}=1\right)=\sigma\left(\boldsymbol{p}_i^T\cdot\boldsymbol{q}_j^{(r)}+b_j^{(r)}\right):=\sigma_{ij}^{(r)}$$

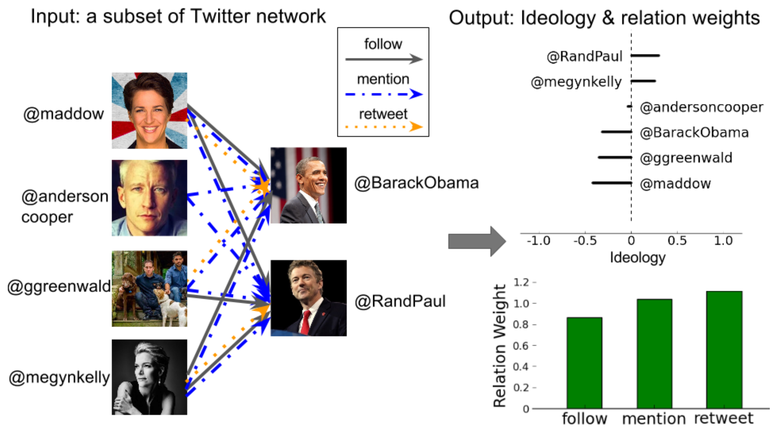
The likelihood of observing the network *r*:

$$l\left(G^{(r)}\right)=\sum_{i,j}I_{[G_{ij}^{(r)}=1]}\log\sigma_{ij}^{(r)}+I_{[G_{ij}^{(r)}=0]}\log(1-\sigma_{ij}^{(r)})$$

Maximize the following objective function w.r.t. model parameters $\{p_i\}_{i=1}^{N_1},\left\{q_j^{(r)},b_j^{(r)}\right\}_{r,j},\{w_r\}_{r=1}^{R}$

$$L=\sum_r w_r\cdot l(G^{(r)})$$

Overview



Experiments

Dataset: 46,477 politics-related Twitter users, including 487 congress people (labeled).

Method	Classification AUC
SL-IPM (follow)	0.980
SL-IPM (mention)	0.899
SL-IPM (retweet)	0.880
ML-IPM	**0.984**

Table 1: User Classification (on labeled politicians)

Method	Link Prediction AUC
AVER	0.606
B-IPM	0.770
SL-IPM	0.905
ML-IPM	**0.912**

Table 2: Link Prediction on Retweet Network

Link Type	w_r
follow	0.866
mention	1.035
retweet	1.117

Table 3: Relative Weights

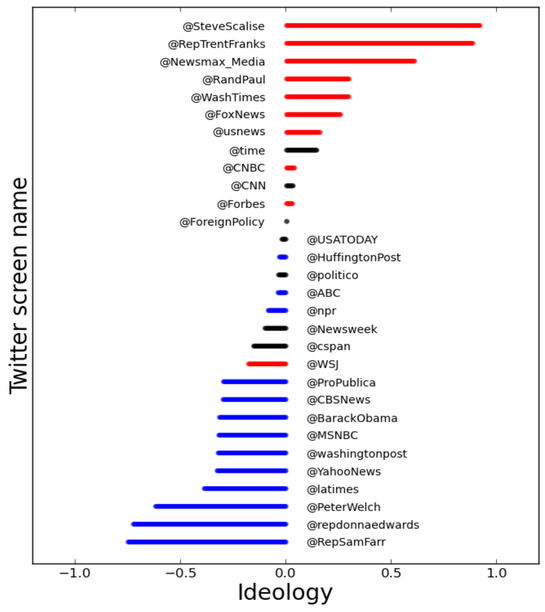


Figure 1: Detected Ideology for selected media accounts and some congressmen

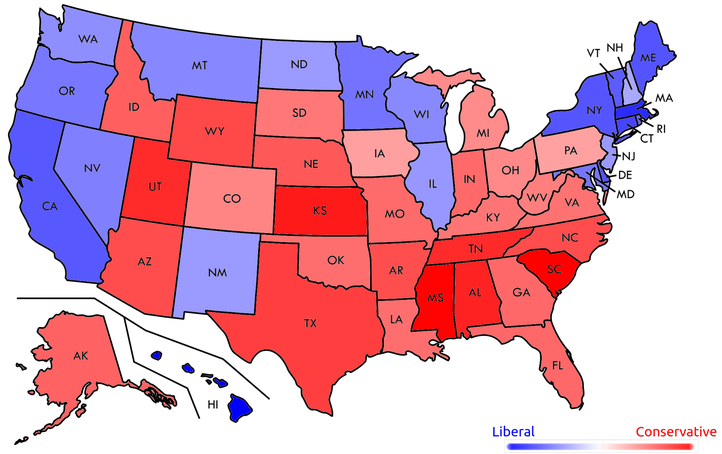


Figure 2: Average ideology score for users in different states