Mitigating Confirmation Bias on Twitter by Recommending Opposing Views

Elisabeth Lex & Mario Wagner & Dominik Kowald
TU Graz & Know-Center
elisabeth.lex@tugraz.at & dkowald@know-center.at

ISDS Institute of Interactive Systems and Data Science

Know Center

SOCIAL COMPUTING

Abstract

People tend to perceive information so it confirms their existing beliefs, a phenomenon called **confirmation bias**. One countermeasure to confirmation bias is to give more prominence to opposing views. We propose a **content-based recommendation** approach to **increase exposure to opposing beliefs and opinions**. Our aim is to help provide users with **more diverse viewpoints on issues**, which are discussed in partisan groups from different perspectives.

Showcase & Goal

- Political debate on Twitter around the presidency of Donald Trump
- People discuss the same issues but from diverse perspectives
- We are interested in partisan groups: (i) pro-Trump stance and (ii) contra-Trump stance
- Our goal is to **make one group aware of the other's group perspectives on issues**
- We propose to use **content-based filtering (CBF)** to **recommend them tweets written by users from the other group** about the same issue

Evaluation Scenarios

1. **Standard**: recommends 10 best matching tweets regardless of user's stance
2. **Pro-Trump**: recommend only pro-Trump tweets
3. **Contra-Trump**: recommend only contra-Trump tweets
4. **Hybrid**: recommend half pro-Trump and half contra-Trump tweets

Recommender Approach

- Content-based filtering: Apache Solr[a]
- **15 most common trigrams** of user as proxy for personal preference
  - Example account "FxgFx"

Username	#Tweets	Stance	Trigrams (user preferences)
FxgFx	1199	contratrump-stance	'trump pay mexico', 'mexico would pay', 'pay wall meant', "trump's tax returns", 'showing true face', '#trumprussia #russiagate #resist' 'rep devin nunes', 'health care plan', 'house oversight committee', "can't wait til", 'get new orders', 'defund planned parenthood', 'make health insurance', 'bring candles back', 'gop members congress'

  - Recommend similar tweets

Trigram: "trump pay mexico" for user FxgFx

Nr	Username	Userstance	Text
1	ckenned	contratrump-stance	RT @SarahLerner: TRUMP: "Mexico will pay for the wall!" MEXICO: https://t.co/SbQPhhp6bA
2	CollinDemocrat	contratrump-stance	RT @chrislhayes: When he said Mexico was gonna pay for the wall, what he *meant* was homebound senior citizens are gonna pay for the wall.
3	bobedgar208	contratrump-stance	RT @JoyAnnReid: So "Mexico will pay for the wall" actually meant "we'll cancel healthcare for millions unless YOU pay for the wall."
4	concerned4theUS	Protrump-stance	Good! Now tell mexico to pay for the wall or else, f**k off! https://t.co/UIxFCWa0O8

[a] https://lucene.apache.org/solr/

Dataset Generation

- Set of **manually selected hashtags** to crawl initial dataset
  - **Pro-Trump**: #maga,#tcot, #americafirst, #trumptrain, #presidenttrump, #draintheswamp, #fakenews, #potus,#buildthewall, #presidentelect- trump.
  - **Contra-Trump**: #impeachtrump, #theresistance, #nobannowall, #resist, #trumprussia, #impeach45, #nottheenemy, #resistance, #notmypresident, #iamamuslimtoo, #nobannowallnoraids, #fakepresident, #dumptrump, #trumplies
- Create pro-Trump and contra-Trump issue stance vectors: concatenate all tweets of users from stance, normalize, tokenize and stem them & extract trigrams using TF-IDF
- Create user stances: concatenate user tweets & TF-IDF to extract user's trigrams
- Cosine similarity between issue stance vectors & user vectors to **verify if user actually belongs to issue stance**
- Removed accounts that used hashtags from both stances
- Filtered out managed accounts, as well as non-English accounts
- In total: **2,150 pro-Trump users with 2,615,140 pro-Trump tweets** and **3,522 contra-Trump users with 3,852,895 contra-Trump tweets**

Evaluation Measures

- **Beyond accuracy** metrics of recommender systems research:
  - **Diversity**: intra-list similarity. Sums all pairwise cosine similarities of the items in a given set and calculates the average of the sum. High if a set has many similar items, low otherwise
  - **Serendipity**: measures how surprising recommendations for a user are. Distance between recommended items and their expected content
- Plus: **average topic similarity** (i.e., pairwise cosine similarity between all users of an issue stance) to understand how diverse partisan groups are per se

Results

Issue stance	Recommendation variant	Serendipity	Diversity
Pro-Trump	Standard	.935	.560
	Pro-Trump	.943	.630
	Contra-Trump	**.951**	.695
	Hybrid	.946	**.728**
Contra-Trump	Standard	.924	.441
	Pro-Trump	**.957**	**.728**
	Contra-Trump	.925	.487
	Hybrid	.940	.701

- **Serendipity**: best results when recommending tweets from opposing view
- **Diversity**: in pro-Trump setting, hybrid wins
- However: in contra-trump setting, pro-Trump wins wrt diversity - why?
  - **Higher average topic similarity of contra-Trump users (44.6%)** versus 27.7% for pro-Trump users) in our dataset
  - Diversity lower if many tweets from a low diversity group mixed into the recommendations
  - Better diversity results: recommend fewer of the more similar contra-Trump tweets and more of the diverse pro-Trump tweets

References

E. Graells-Garrido, M. Lalmas, and D. Quercia. (2013). Data Portraits: Connecting People of Opposing Views. ArXiv e-prints (Nov. 2013). arXiv:cs.HC/1311.4658
Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M. (2017). Reducing Controversy by Connecting Opposing Views. In Proc. of WSDM '17.