Embeddings-Based Clustering for Target Specific Stances: The Case of a Polarized Turkey

Ammar Rashed , Mucahid Kutlu, Kareem Darwish, Tamer Elsayed,

Introduction

- On June 24, 2018, Turkey underwent early elections for the presidency and the parliament.
- We collected +108M stance-laden tweets and manually annotated ~6k users with party-level stances.
- Our assumption is that users with similar stances tweet in a semantically similar manner.
- Our novel unsupervised stance detection approach outperforms previous SOTA in both accuracy and speed.
- We provide a comprehensive framework to analyze fine-grained polarization between groups over various topics.

Data

- **Election Data (ED):** tweets including keywords pertaining to Turkish elections, between April 29, 2018 and June 23, 2018.
  - **Manual Labeling:** users explicitly specifing their party affiliation in their twitter handles or screen names.
  - **Label Propagation**: automatically labels users based on the tweets that they retweet.
- **Timeline Data (TD):**
  - On Dec. 28, 2018, we collected ~98.7M tweets from the timelines of ~86K pro, and ~115M by ~82K anti users.

Supporters	Users	Tweets
pro-Erdoğan	1,772	561,510
anti-Erdoğan w/o party affiliation	2,115	516,166
pro-CHP	29	171,201
pro-IYI	890	168,442
pro-HDP	354	61,274
Total	5,960	1,478,593

Manually labeled users

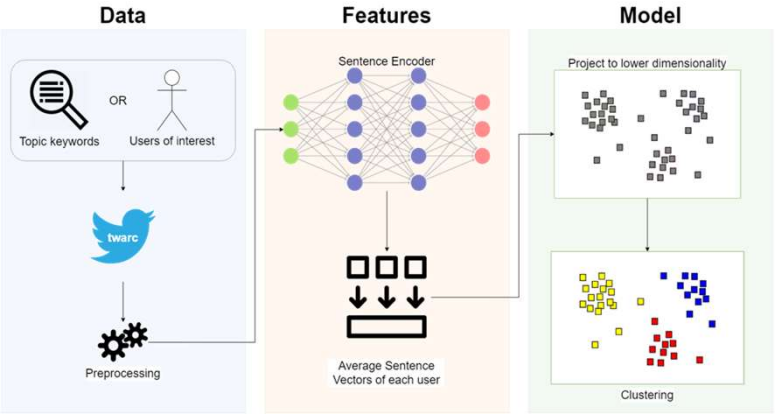
- **Trump Dataset:**
  - Oct. 25-27, 2018, tweets of 7,421 pro- and 6,310 anti-Trump users. (Darwish et al., 2019)

Conclusion

- We provide a novel fine-grained unsupervised stance detection approach.
- We provide a large Turkish stance detection dataset.
- We quantitatively and qualitatively analyze polarization between Twitter users in the 2018 Turkish elections.

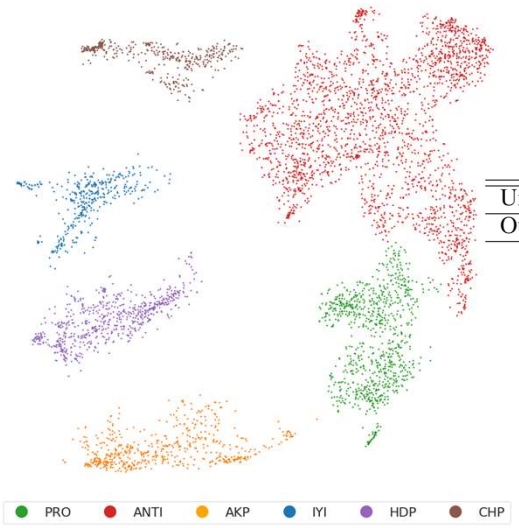
Approach

- User tweets are encoded with Multilingual Universal Sentence Encoder then averaged per user.
- User embeddings are projected into a 2D space using UMAP and then clustered with HDBSCAN.



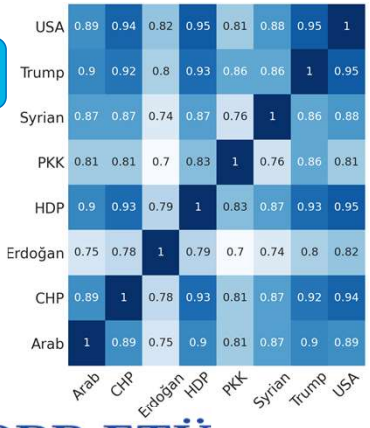
Results

- Detected stances were aligned with manually labeled party affiliations, without inputting # of clusters.



	Trump	ED
UnSup	0.83	**0.86**
Ours	**0.86**	0.84

Macro F1 with previous SOTA



Can we infer the stance on a topic from a different topic?

Adjusted mutual information between the clusters of different topics

QCRI معهد قطر لبحوث الحوسبة Qatar Computing Research Institute
HAMAD BIN KHALIFA UNIVERSITY جامعة حمد بن خليفة
TOBB ETÜ University of Economics & Technology
ÖZYEĞİN UNIVERSITY