Detection and Characterization of Stance on Social Media

Part2

Detection and Characterization of Stance on Social Media

Abeer Aldayel, School of Informatics, University of Edinburgh
Kareem Darwish, Qatar Computing Research Institute, HBKU
Walid Magdy, School of Informatics, University of Edinburgh

Stance Detection Modeling

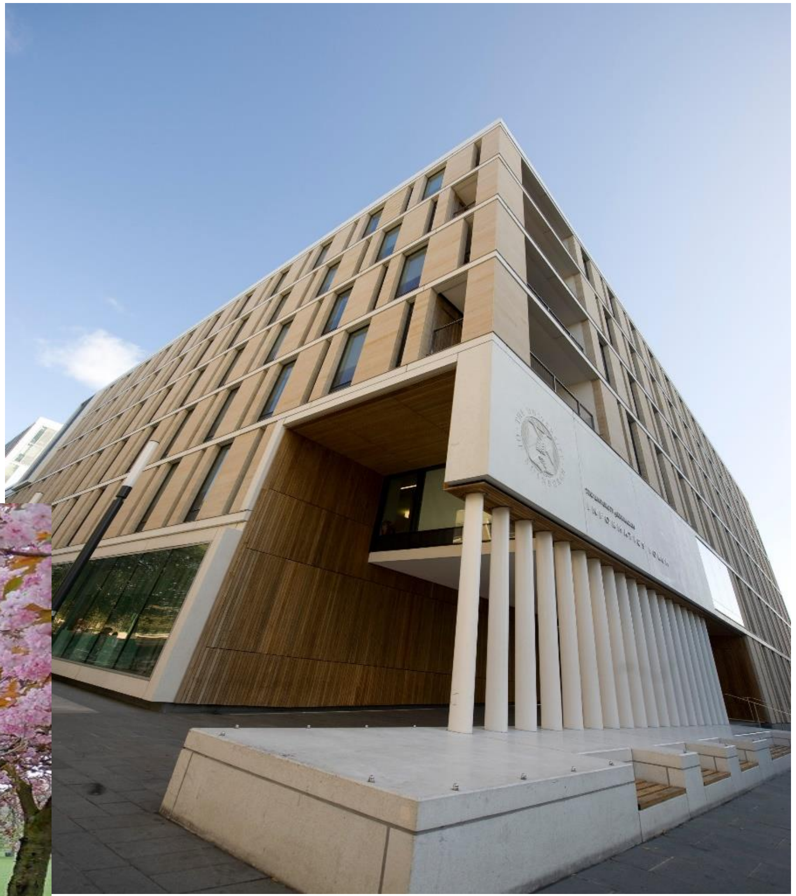
Abeer Aldayel
@AldayelAbeer

About me.

Abeer Aldayel https://abeeraldayel.github.io/ @AldayelAbeer

- Currently I am PhD candidate at University of Edinburgh, school of informatics.

- I am member of Social Media Analysis and Support for Hummanities (SMASH) research group.



School of Informatics, University of Edinburgh

My are of research:

- Computational Social Science
- Social computing
- Stance detection on Social Media

Meadows Park

Part2: Stance modeling in social media

2.1 Stance detection modeling

2.2 Most effective classification features

2.3 Stance detection methods
(supervised, semi-supervised, and unsupervised)

Part2.1

How to model the stance on Social Media?

How to model the stance in social media?

Stance on social media has been modeled using various online signals

1.What the users say [Content]

-Related to the topic

-Not-Related to the topic

2.Online behavior [Network Interactions]

-Users Similarity

-Users Heterogeneity

1. Textual content

- Textual entailment:

  -Text entails a stance to an entity (Text–>infavor/against an entity), if the stance to a target can be inferred from the given text.

  [Text ➔ favor/against an entity]

- Example:

  Target of analysis: legalization of abortion (LA)
  Text: The pregnant are more than walking incubators, and have rights!

  Entails

  A users is likely in-favor of Legalization of Abortion

How to model the stance in social media?

Stance on social media has been modeled using various online signals

1.What the users say
[Content]

Post level

-Related to the topic

-Not-Related to the topic

Features level

-Linguistics Features

-Vocabulary Choice

"No, hashtags are not a part of speech."

Textual content/ Post level

- Posts related to the topic
  - Using set of keywords related to the topic of analysis as the main input to stance detection model.
  - Example: topic: Brexit, keywords: Brexit, European Union, UK.

- Posts not related to the topic
  - Disentangling the topic from the viewpoint where the vocabulary is not only linked to the topic but to the individual attitude and characteristics.

    -Example: [Users Embeddings*]

    -Using users Text, compute a TF-IDF-weighted user-word matrix based on tweets.

*Benton, Adrian, and Mark Dredze. "Using author embeddings to improve tweet stance classification." *2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*.

Textual content/ Features

- ***Linguistics Features:***
  - Textual features→
    - n-gram modeling best score in Stance Semeval2016.
  - Sentiment polarity→
    - Negative/ Positive text polarity (sentiment analyzer)*.
  - Latent semantics→
    - Reduce the dimension of a given input such as mapping the sentences to predefined set of topics (topic modeling).

*Aldayel, Abeer, and Walid Magdy. "Assessing Sentiment of the Expressed Stance on Social Media." *International Conference on Social Informatics*. Springer, Cham, 2019.

Textual content/ Features

User's Vocabulary Choice:

- Hypothesis: individuals with same stance tend to use the same vocabulary choice to express their point of view.
- For instance, people who oppose abortion tend to use vocabulary such as pro-life to express their stance.

  - Example: [Klebanov et al. 2010] user stance detection model based on users vocabulary choice using a generative model and discriminative model using Naive Bayes and SVM.

2.Online behavior- Network Interactions

- Individuals views are related to their identity (Jaffe, 2009).

➔ Online interactions (Homophily)

How to model the stance in social media?

Stance on social media has been modeled using various online signals

1. What the users say [Content]

-Related to the topic

-Not-Related to the topic

2. Online behavior [Network Interactions]

-Users Similarity

- Users behavioral data
- Meta-data attributes

-Users Heterogeneity

- Rebuttal interactions

2. Online behavior- Users' similarity

The similarity between users considered a core property that helps in inferring stances.

- **Within the same platform:**
  - Common platform: Twitter
  - Common features: Retweeted accounts , used Hashtags.
- Case study: [*User similarity feature space*]*
  - Retweeted accounts; used hashtags; mentioned accounts ; shared URLs.
  - Graph reinforcement to calculate the similarity between the users

*Darwish, Kareem, Walid Magdy, and Tahar Zanouda. 2017. *Improved stance prediction in a user similarity feature space*. In Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, pp. 145-148. ACM, 2017.

2. Online behavior- Users' similarity

- **Multiple Platforms:**
- **Case study:** [Community Detection]*
  - Twitter + CreateDebate.com
  - Multi-layer graph model to represent profiles from different platforms and to extract communities.
  - Combining several types of online communities improves stance detection results.

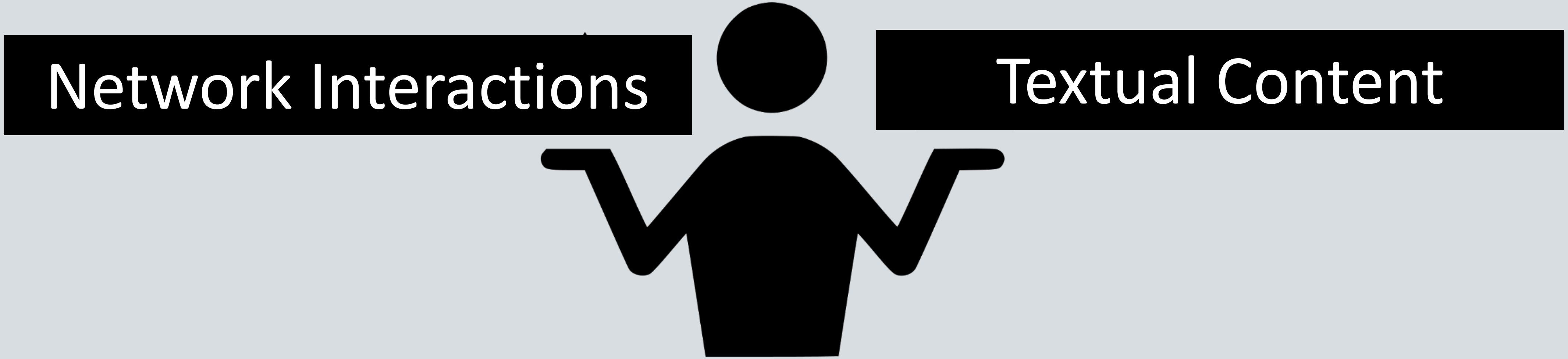
Ophélie Fraisier, Guillaume Cabanac, Yoann Pitarch, Romaric Besançon, and Mohand Boughanem. 2018. Stance Classification through Proximity-based Community Detection. In Proceedings of the 29th on Hypertext and Social Media (HT '18).

2. Online behavior- Users' dissimilarity

- Heterophily: tendency for individuals who differ from one another to make social connections.
- It is less common than homophily
- Case study: [Online debates]*
  - The tendency of a user to reply to the opposed viewpoint.
  - Rebuttal variable to model the reply attacks the previous author parent post.

*Amine Trabelsi and Osmar R Zaiane. 2018. Unsupervised model for topic viewpoint discovery in online debates leveraging author interactions. In *Twelfth International AAAI Conference on Web and Social Media*.

Part2.2 What is the most effective stance modeling?



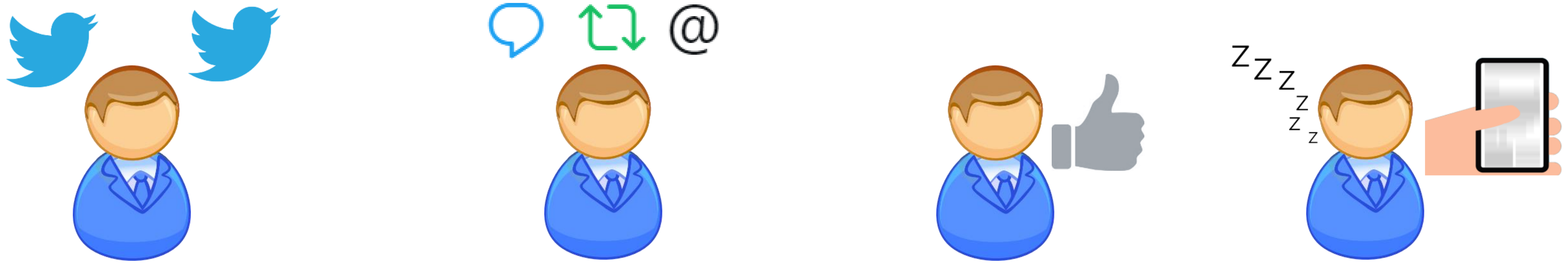
SemEval Stance Detection Task 2016

- **4K** tweets on **5 topics** labeled by stance {favor, against, none}
- Topics: Abortion, Atheism, Feminism, Clinton, & Climate Change
- **State-of-the-art**:
  - SVM + n-gram features → F-score **69%**
  - Other approaches: deep learning → F-score < 69%
  - Focus on content features only! → user discussed the topic!

- How detecting stance could be done if:
  - User never discussed the topic!
  - User never tweeted, but has some online activity!
  - User has no content and no activity!

*Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

Detecting Stance in Four Situations



On-topic content | General activity | Silent User | Passive User

TXT: tweet Text content

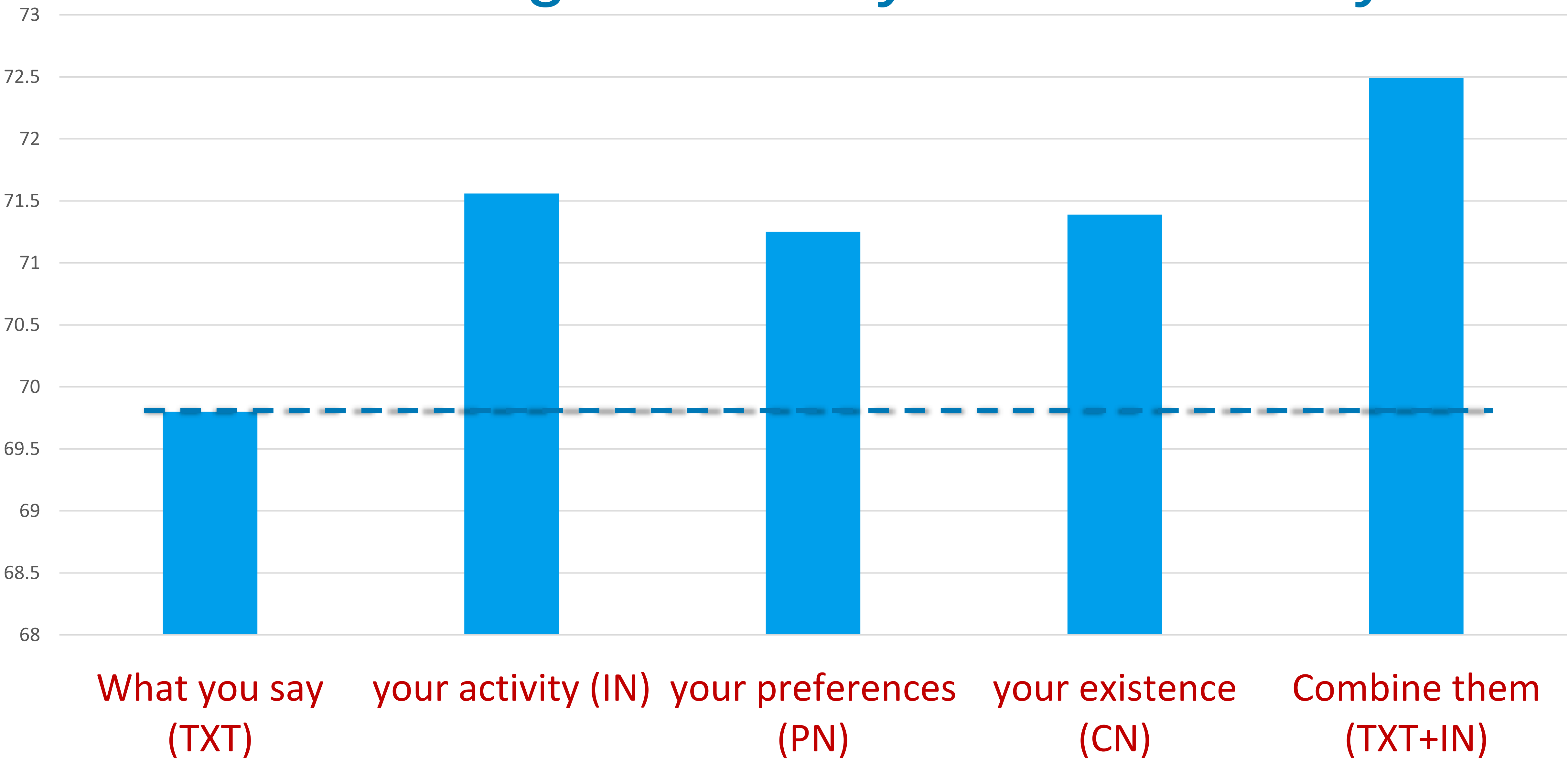
IN: Interaction Network → network user retweet, reply, mention*

PN: Preference Network → network in tweets user like

CN: Connection Network → network user follows

*Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

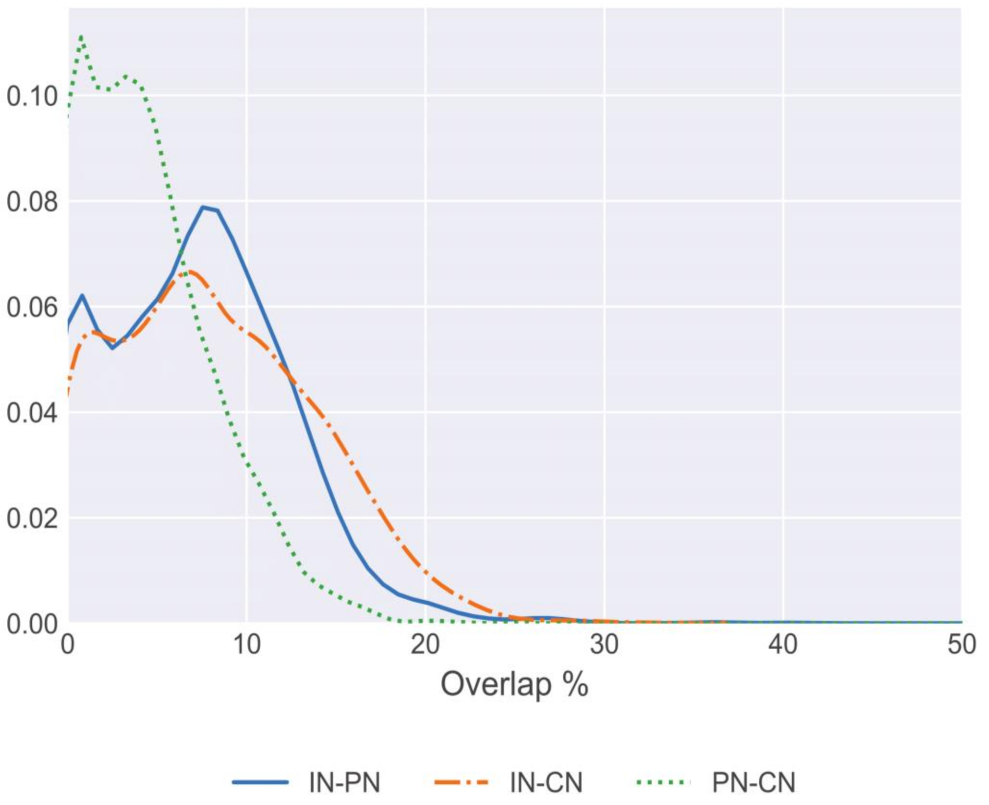
Detecting stance by online activity

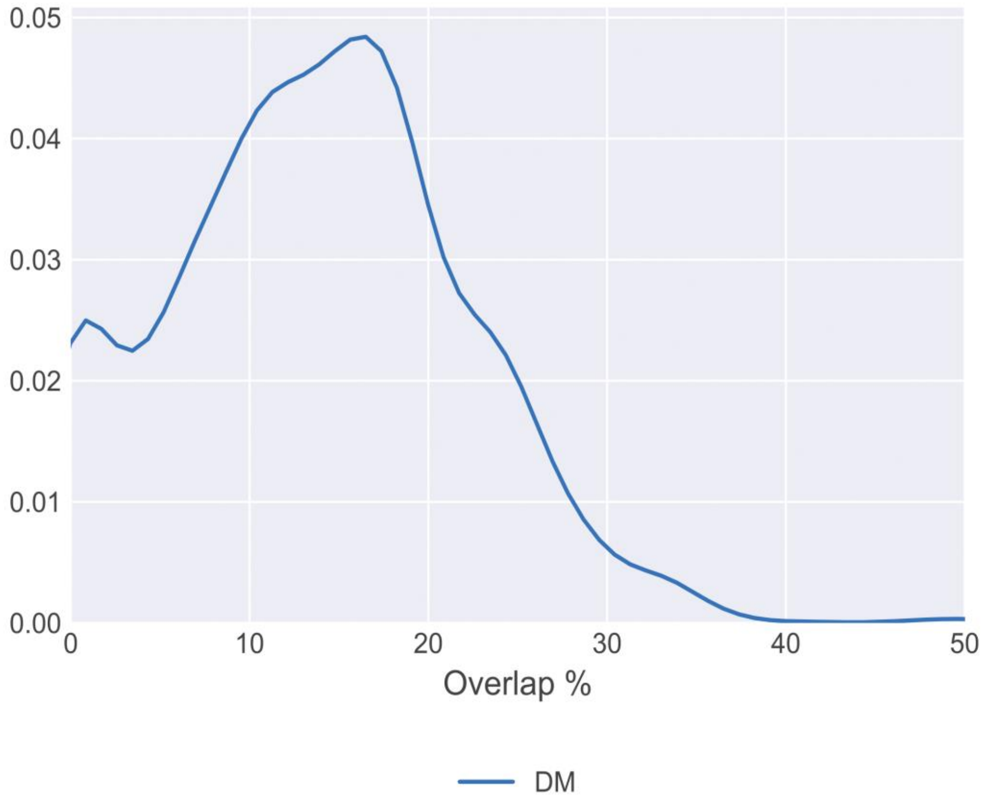


Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

Social dynamics of interactions

May be user interact with same account they network they like, which is the same network they follow!

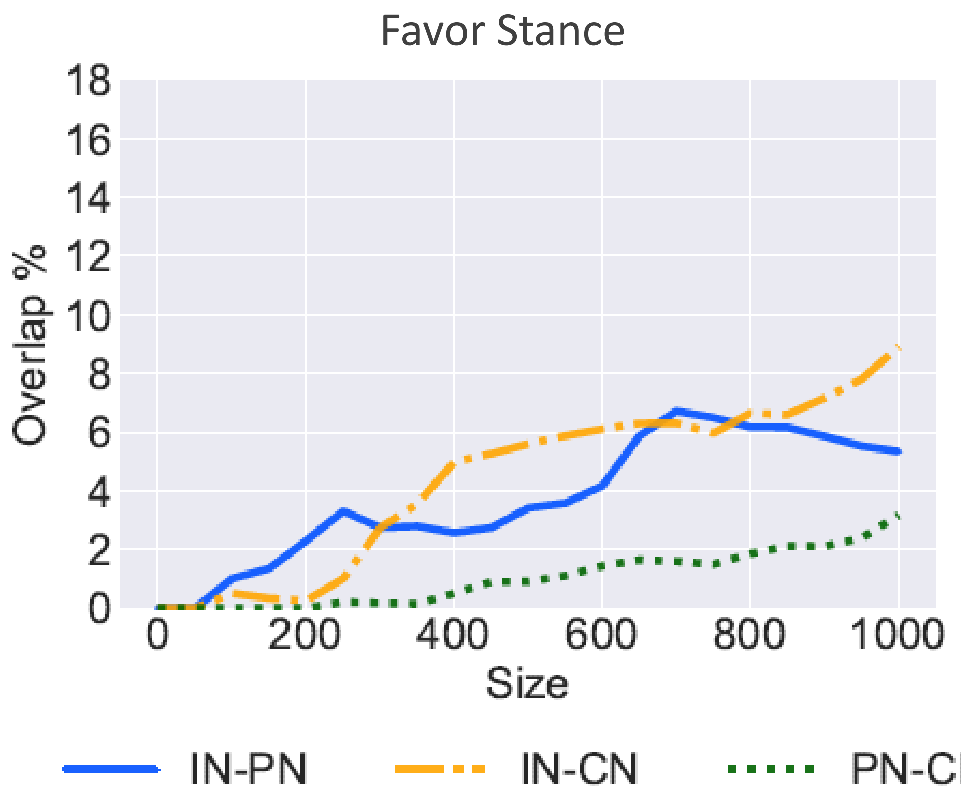


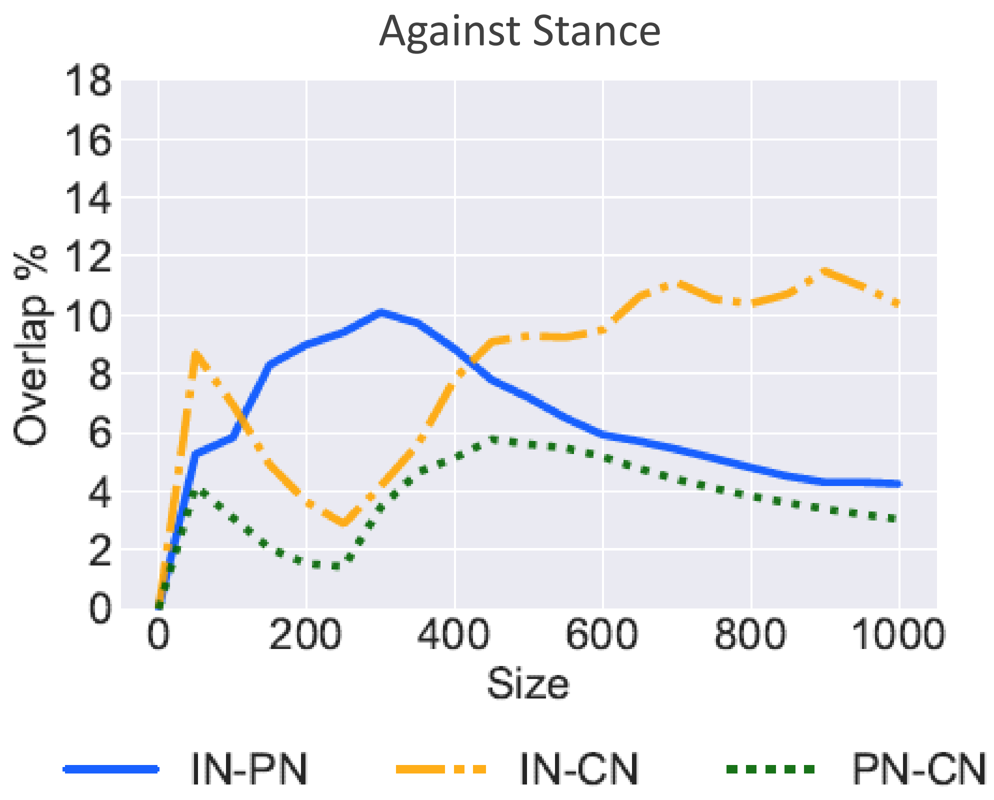


Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

Most Effective Features

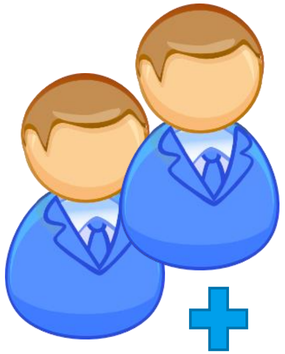
May be the small % of common network nodes are the most effective in detecting stance!





Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

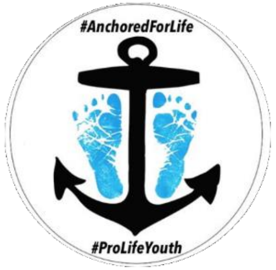
Most influential features



(CN)

Against

Legalization of abortion



@ProLifeYouth

@March_For_Life

Supporting

Hillary Clinton

@HillaryClinton

@ SheHasMyVote

Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc.* CSCW 2019

Most influential features [Supporting Hillary Clinton]

@

IN

PN

@RealDonalTrump

@WashTimes

@CBSNews

You are an idiot on so many levels, @realDonaldTrump

Context of the interaction ≠ Topic

RT@Telegraph: Prince Charles reveals his gardening inspiration: a hidden #BuckinghamPalace

RT @SkyNewsBreak: Former Labour Prime Minister Tony Blair has told Sky News Theresa May will win the General Election

Global scale analysis

Comparative analysis of the stance modeling

Stance Detection - Network | Content | Both

Dataset	NW	Content	Both	ML Algorithm
• Before Paris attacks	**85.0**	82.0	84.0	SVM [Magdy et al. 2016]
• Islam Dataset	84.0	76.0	**--**	SVM [Darwish et al. 2017a]
• Gun control, abortion and Obama care	81.9	60.6	**82.0**	Non-negative matrix factorization [Lahoti et al. 2018]
• SemEval stance dataset	71.6	69.8	**72.5**	SVM [Aldayel & Magdy 2019]
• SemEval stance dataset	61.8	62.8	**65.9**	RNN [Lynn et al. 2019]
• LIAC dataset	**67.8**	54.1	--	Generative Model [Rui et al. 2017]

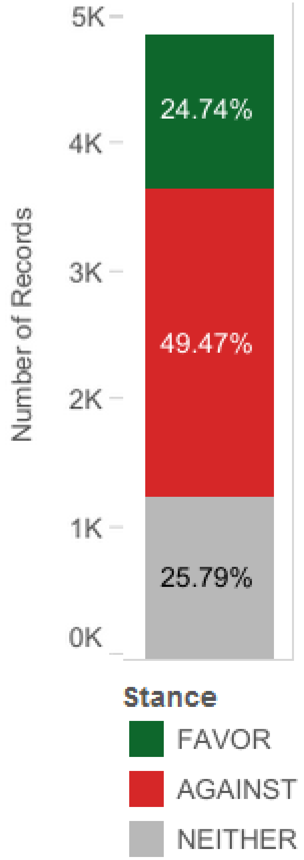
Key takeaways

- The accurate learning of the stances at the user-level representation(IN).
- Detecting stance for silent users is possible (PN and CN networks).
- Using Heterophily to model stance is less common than homophily.
- Please do not use sentiment analyzer to gauge online support!

Part2.3: Supervised/ Semi-supervised/ Unsupervised stance detection

Stance detection methods [Current state]

- Stance detection has been mostly approached using classification-based algorithm.
- Mostly by using supervised learning algorithms with huge dependency on human-annotated data.
- The widely used benchmark Stance Detection dataset :
  - SemEval stance 2016.
  - About 4000 tweet
  - 5 topics



SD- models on SemEval stance dataset

	Model	Features	F1	
Supervised learning	SVM	Content+NW	72.5	[Aldayel &Magdy 2019]
	RNN	Content+NW	65.9	[Lynn et al. 2019]
	BiGRU	Content	72.33	[Lahoti et al. 2018]
Transfer learning	BiGRU	**Noisy Labeling+ topic modeling**	60.80	[Rui et al. 2017]
	DAN	WordEmbeddings	35.2	[Ebner et al. 2019]

Key takeaways

- Using a topic based transfer learning achieved lower performance score in comparison with supervised learning techniques.

  ➔ further enhancement by adapting a non-topical aspects to enhance the current state of this methodology.

- Several studies show consistent improvement by using the network features instead of just using content of post only.

  ➔Crucial role of users online social attributes in the stance detection models.

Supervised Stance Detection

SD – Supervised Learning

- Train a text based classifier on Tweets:
  - Case study: Egyptian 2013 Coup (pro-coup/anti-coup)
    - Period of study: June 21 – Oct. 1, 2013
    - Tweet level classification
    - Features: word unigrams, word bigrams, hashtags
    - Labeled data: 1,000 tweets – pro/anti/neutral
    - Evaluation: 20-fold cross validation
    - Avg accuracy: 87%

- Borge-Holthoefer, Javier, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. *Content and network dynamics behind Egyptian political polarization on Twitter.* In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 700-711. ACM, 2015.

SD – Supervised Learning

- Train a text based classifier on Tweets:
  - Case study: Egyptian 2013 Coup (pro-coup/anti-coup)
    - Given that classification was done on tweets, we can observe changes in stance.
      - June-21: We will continue to revolt till we reach freedom. Gathering revolution from Alexandria to Cairo to oust Morsi, the sheep.
      - July-19: The Mohandseen march is closing the main streets till the police station #NoToMilitaryCoup
    - Percentage of change 2-3%

- Borge-Holthoefer, Javier, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. *Content and network dynamics behind Egyptian political polarization on Twitter.* In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 700-711. ACM, 2015.

SD – Supervised Learning

- Train a text based classifier on users:
  - Case study: Support for ISIS (ISIS vs. Islamic State)
    - User level classification
    - Features: word unigrams, hashtags, user mentions
    - Labeled data: > 14,000 users – pro/anti
    - Evaluation: 10-fold cross validation

	Precision	Recall	F1-measure
Pro-ISIS	89.6	83.7	86.6
Anti-ISIS	84.7	90.3	87.4
Average	87.2	87.0	87.0

    - We can predict who will end up supporting ISIS later with 87% accuracy

- Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. *#FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support.* First Monday 21.2 (2016).

SD – Supervised Learning

- Train a text based classifier on users:
  - Case study: Islamophobia in the US (pro/anti)
    - ISIS carries out terrorist attacks in Paris 11/2015
    - User level classification
    - Features: word unigrams/hashtags/user mentions/RT
    - Labeled data: 1,534 user tweets – pro/anti ➔ 44k users
      - I feel horrible that people who practice Islam have to apologize for the #ParisAttack - Muslim people aren't responsible; terrorists are.
      - Why are muslims even allowed out of their garbage countries? We need to take out the trash #KillAllMuslims #DeportAllMuslims #RemoveKebab

- Magdy, W, Kareem Darwish, Noura Abokhodair, Afshin Rahimi, Tim Baldwin. 2016. *ISISisNotIslam or DeportAllMuslims? Predicting Unspoken Views.* In Proceedings of the 8th ACM Conference on Web Science. 8th ACM Conference on Web Science.

SD – Supervised Learning

- Train a text based classifier on users:
  - Case study: Islamophobia in the US (pro/anti)
    - Can we predict who will have Islamophobic views?
    - Evaluation: 200 tweets before incident for training

	w/ prior views		w/o prior views	
	Hashtags	RT	Hashtags	RT
Positive Prec.	84	89	90	90
Negative Prec.	75	83	58	79

- Magdy, W, Kareem Darwish, Noura Abokhodair, Afshin Rahimi, Tim Baldwin. 2016. *ISISisNotIslam or DeportAllMuslims? Predicting Unspoken Views.* In Proceedings of the 8th ACM Conference on Web Science. 8th ACM Conference on Web Science.

SD – Supervised Learning

- Map users into latent space prior to classification:
  - Pick a set of “exemplar users”, and use similarity to them as features
    - Case study: Islamophobia dataset
      - Computed similarity based on: RT/Hashtags
      - Size of latent space: 100 users
      - Training set: 100 users, Test set: 2,607 users
      - Compared raw features vs. using the features to compute similarity

SVM Classifier	RT		HASH	
Measure	SIM	Raw	SIM	Raw
POS F1	***0.76***	0.69	0.72	0.63
NEG F1	***0.92***	0.90	0.90	0.88
Macro-F1	***0.84***	0.80	0.81	0. 75

- Darwish, Kareem, Walid Magdy, and Tahar Zanouda. 2017. *Improved stance prediction in a user similarity feature space*. In Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, pp. 145-148. ACM, 2017.

SD – Supervised Learning

- Advantages:
  - Simple

- Disadvantages:
  - Accuracy seems to be capped
  - Requires training data

- Observations:
  - **Works even with non-topical content**
  - **Network interactions (RT) better than actual content (hashtags)***
  - **People's views are durable**

Now over to… Kareem

...To be continued with part 2.3