Using Sentiment Induction to Understand Variation in Gendered Online Communities

Li Lucy & Julia Mendelsohn
Stanford University

Main Ideas

- Online communities could be better understood through vector-based representations that capture users, text, and affect.
- Sentiment can provide insight into variation that word-choice alone cannot. Members of online communities exploit linguistic resources to construct a wide array of gendered identities.

Background

- Communities: In sociolinguistics, communities of practice are characterized by their participants' shared beliefs and language styles (Eckert 2006). NLP research has also shown that online communities form collective linguistic norms (Danescu-Niculescu-Mizil et al. 2013).
- Gender: NLP often treats gender as a fixed biological variable, rather than a dynamic and social one (Butler, 1988; Nguyen et al., 2014; Herring and Paolillo, 2006).
- Sentiment Variation: Community-specific lexicon induction has revealed cross-community variation in small domains (Hamilton et al. 2016).

Data

- Explicitly gendered communities within top 400 subreddits
- Comments from May 2016 - April 2017
- These subreddits contain between 10^7 and 10^8 tokens.

Gendered Subreddits				
actuallesbians	*xxfitness*	*askwomen*	*femalefashionadvice*	*trollxchromosomes*
askgaybros	*mensrights*	*askmen*	*malefashionadvice*	

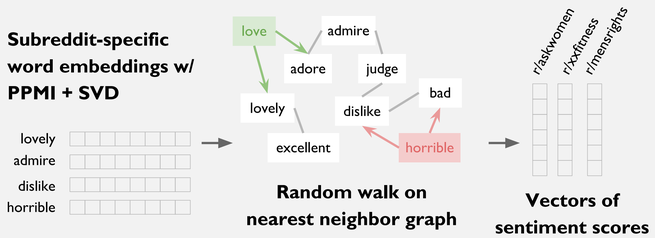
Approach

We represent subreddits in three different ways.

Text: Tf-idf on unigram counts + SVD → 100-dimensional vectors.

User: Tf-idf on commenter counts + SVD → 100-dimensional vectors.

Sentiment: SentProp with $\beta = 0.9$ (favors similar labels for neighboring words) and # of neighbors = 25 (Hamilton et al. 2016). Values have zero mean and unit variance, averaged over 50 bootstrap-sampled runs. Words without induced sentiment for particular subreddit are set to neutral.



Positive	love, loved, loves, awesome, nice, amazing, best, fantastic, correct, happy
Negative	hate, hated, hates, terrible, nasty, awful, worst, horrible, wrong, sad

Seed words for sentiment propagation (Hamilton et al. 2016)

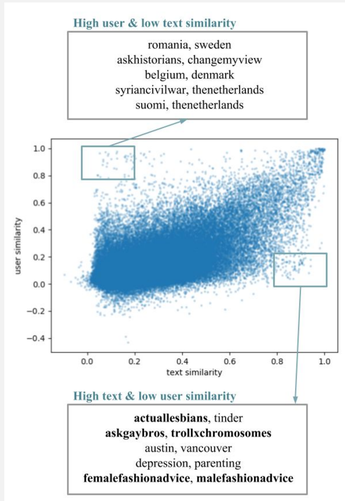
Text & Users

The geography of Reddit differs when defining subreddits through language or user membership.

User demographics are important motivators for community formation on Reddit.

Agglomerative clustering of text and user representations (k=20):

- Feminine subreddits all in same user cluster, with neighbors r/makeupaddiction and r/weddingplanning.
- Most gendered subreddits are in cluster containing personal topics, such as r/deadbedrooms and r/childfree.



Spearman corr: 0.549, $p < 0.0001$

Sentiment Similarities

highest pairs	cosine	lowest pairs	cosine
askmen, askwomen	0.6702	femalefashionadvice, mensrights	0.1802
askgaybros, askmen	0.6144	askwomen, malefashionadvice	0.1876
askwomen, trollxchromosomes	0.6003	malefashionadvice, trollxchromosomes	0.2162
actuallesbians, trollxchromosomes	0.5462	malefashionadvice, mensrights	0.2170
askgaybros, askwomen	0.5310	mensrights, xxfitness	0.2181

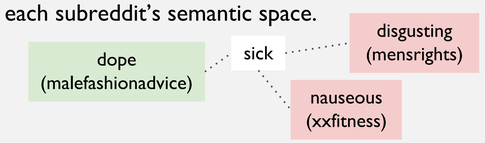
- r/askmen and r/askwomen have high sentiment similarity, low text similarity (0.287), near average user similarity (0.417).
- r/actuallesbians and r/trollxchromosomes have high text and user similarity (0.942 and 0.886), above average sentiment similarity
- Spearman corr. of text & sentiment similarity: 0.637 ($p < 0.0001$).

Community-Specific Sentiment Examples

word	subreddit	comments
loooove +	femalefashionadvice	expressive elongation is a female marker (Rao et al., 2010, Bamman et al., 2014)
men -	mensrights	users don't dislike men but instead focus on injustices towards men
thursdays +	femalefashionadvice	tradition for highlighting outfits for each day of the week; male community does not have this
trolls +	trollxchromosomes	re-appropriated a commonly negative term to refer to themselves
flu -	xxfitness	most negative words are physical ailments

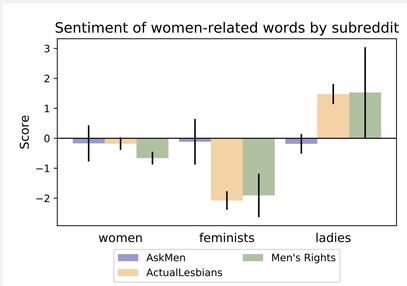
Variation Analysis

Some of the variation is due to polysemy: *sick*, with the 13th highest variance, encounters different neighbors in each subreddit's semantic space.

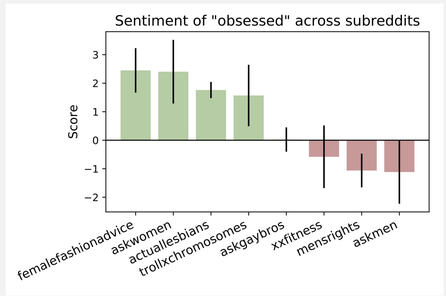


Though gender is often a helpful binary variable in NLP, its expression is not fixed across multiple contexts.

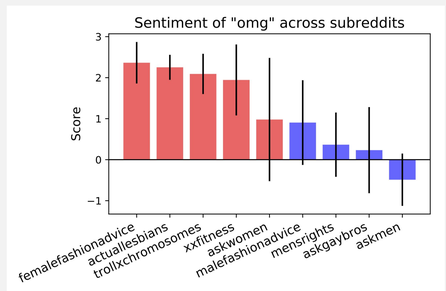
word	Volkova et al. (2013)	our work
overdressed	men (+), women (-)	malefashionadvice (-), femalefashionadvice (-)
weakness	men (-), women (+)	actuallesbians (+), xxfitness (-)



Comparing denotationally similar *women* and *ladies* and the related word *feminists* yields surprising results for r/actuallesbians and r/mensrights.



A word such as *obsessed* (5th highest variance) or *jealous* can be used to convey personal, positive passions or to make negative claims about others' mental states.



The word *omg* is commonly used by and associated with women (Bamman et al. 2014, Carpenter et al. 2017). Women also use it to convey highly positive affect.

Future Work

- Expanding to other communities: implicitly gendered subreddits (e.g. high user-overlap with gendered ones), comparing gendered subreddits (r/xxfitness) to non-gendered ones (r/fitness).
- Examining other semantic dimensions: arousal, emotions.

Conclusion

Online language does not vary according to a clearcut, binary perspective of gender. Sentiment can be a useful indicator of words' social meaning and community values, especially in the context of discussion content and user demographics.

Acknowledgements. We would like to thank Chris Potts, Will Hamilton, and Bill MacCartney.

References

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. Journal of Sociolinguistics, 18(2):135–160.
Judith Butler. 1988. Performative acts and gender constitution: An essay in phenomenology and feminist theory. Theatre journal, 40(4):519–531.
Jordan Carpenter et al. 2017. Real men don't say cute: using automatic language analysis to isolate inaccurate aspects of stereotypes. Social Psychological and Personality Science, 8(3):310–322.
Penelope Eckert. 2006. Communities of practice. Encyclopedia of language and linguistics, 2(2006):683–685.
William L Hamilton et al. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 595–605.
Susan C Herring and John C Paolillo. 2006. Gender and genre variation in weblogs. Journal of Sociolinguistics, 10(4):439–459.
Dong Nguyen et al.. 2014. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In Proceedings of COLING 2014, pages 1950–1961.
Delip Rao et al. 2010. Classifying latent user attributes in Twitter. In Proceedings of the 2nd international workshop on Search and mining usergenerated contents, pages 37–44. ACM.
Svitlana Volkova, Theresa Wilson, and David Yarowsky. 2013. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1815–1827.

Contact: lucy3@stanford.edu, jmendels@stanford.edu.