Social and Language Influence in Wikipedia Articles for Deletion Debates*

Linda Wang
Cornell University

Abstract

Today, digital platforms are integral for group decision-making. This is especially evident on Wikipedia, where the curation of over six million articles[1] depends upon online discussion. In such a setting, structural (e.g votes) and linguistic (e.g. sentiment) features can be especially salient. Both can contribute to the likelihood of a discussant agreeing with the majority opinion; however, both can also be imperfect signals of a given discussion post's credibility. Thus, by precisely defining their relationship with individual users' behaviors, we may gain insights into herding-like behaviors and assist initiatives to improve user engagement and address knowledge gaps. This paper is a work-in-progress that attempts to achieve these objectives. Using a corpus composed of Wikipedia Articles for Deletion (AfD) debates, we first check that behavior consistent with herding is present in debates. We then construct a binary logistic model of user choice to assess the relative influence of structural versus linguistic features. Lastly, we propose two possible extensions: assessing how the influence of linguistic features changes in debates over articles about women, and conducting a lab experiment to check for causal relationships suggested by our previous analyses.

Keywords: group decision-making, voting, text mining, experiment, gender bias

Background

Several works have linked the outcome of AfD debates with vote sequences and voting language. For example, one study looked into the existence of herding effects and voter heterogeneity within debates. The authors found that an over- or under-expression of a particular vote type towards the start of a debate was associated with an over- or under-expression of that vote type, respectively, towards the end of a debate (Taraborelli and Ciampaglia, 2010). Another study encoded the rationales of each voting comment with a base BERT model, then trained a logistic regression classifier to predict a probability distribution over debate outcomes given a debate. This model indicated that early votes were highly predictive of debate outcomes, with the effect being especially evident for "keep" votes (Mayfield and Black, 2019). We hope to contribute to this literature by using a model to predict individual votes rather than debate outcomes; connecting structural and linguistic features of debates with gender gaps; and partially addressing the issues of selection into debates and endogeneity of user preferences.

Methods

Observational Study

Data

We utilize the Wikipedia Articles for Deletion Corpus included in the Cornell Conversational Analysis Toolkit (ConvoKit)[2]. This corpus is a ConvoKit-formatted version of the data released by (Mayfield and Black, 2019), a collection of 383,918 AfD debates that occurred between 1/1/2005 and 12/31/2018 on the English-language Wikipedia.

Behavior Check

Across debates, we look at the probability that the (k+1)th vote agrees with the majority of the preceding k votes, as a function of k. If herding-like behavior is present, we would expect to see the probability of agreeing with the majority increase with k.

Baseline Model

We construct a binary logistic model of user votes using subsets of features of the preceding votes. For simplicity, we focus on votes of "delete" and "keep", only. The structural features of interest are debate length (i.e. the total number of posts in a debate); presence of previous "delete" vote(s); presence of previous "keep" vote(s); and proportion of "delete" votes.

* I am grateful for the helpful comments from the WMF Research Team and many colleagues at Cornell. Thanks are due to Kiran Tomlinson for help in implementing the data analyses.

[1] https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

[2] https://convokit.cornell.edu/documentation/wiki-articles-for-deletion-corpus.html

The linguistic features of interest are post length; positive, neutral, and negative sentiment as scored with VADER (Hutto and Gilbert, 2014); presence of internal references (i.e. references to posts within a debate); presence of external references (i.e. presence of links outside the debate page); and use of Wikipedia AfD slang[3].

Extension: Gender

Prior work has indicated that over 25% of women biographies on the English-language Wikipedia are nominated for deletion each month, despite the fact that women biographies comprise only around 18% of total biographies (Tripodi, 2021). We would like to investigate whether such differences alter the relative influence of linguistic features in debates, as a natural extension of our current work. To do this, we will enhance our baseline model by adding gender features and re-running it on a subset of data consisting only of debates nominating biographies for deletion. To create the new features, we will use article titles and gender detection algorithms such as Wiki-Gendersort (Bérubé et al., 2020). However, a key feature for this kind of comparative analysis is article quality, which is unavailable in our data; thus, we will attempt to use another data source - Wikidata - as well as algorithmic assessments of article quality[4] to supplement the predictions of the enhanced model.

Extension: Lab Experiment

Given the difficulty in controlling for selection and user preferences in the data, we believe another possible extension would be a lab experiment. Broadly, each lab participant could see a series of AfD debates that have been manipulated to express at least one type of linguistic variation (e.g. higher frequency of external references). Another participant could see an alternative version of that series, manipulated to express the opposite linguistic variation(s) (e.g. higher frequency of internal references). External validity concerns may be ameliorated with operational definitions of "optimal" notability and article quality, which we are investigating at the time of this writing.

Preliminary Results

The preliminary results provide mixed evidence on the relationship between individual votes and structural and linguistic features. The behavior check indicates that a certain degree of "anti-herding" behavior may be present across AfD debates, especially in longer debates. Interestingly, however, our baseline model suggests that the effects of slang, sentiment, usage of links, and raw vote proportions on an individual vote differ depending upon the index of that vote in the debate.

[3] https://en.wikipedia.org/wiki/Wikipedia:Glossary

[4] https://meta.wikimedia.org/wiki/Research:Prioritization_of_Wikipedia_Articles/Language-Agnostic_Quality#V2

References

[Bérubé et al.2020] Nicolas Bérubé, Gita Ghiasi, Maxime Sainte-Marie, and Vincent Larivière. 2020. Wiki-Gendersort: Automatic gender detection using first names in Wikipedia. *SocArXiv*.

[Hutto and Gilbert2014] C. Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225.

[Mayfield and Black2019] Elijah Mayfield and Alan W. Black. 2019. Analyzing Wikipedia Deletion Debates with a Group Decision-Making Forecast Model. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), Nov.

[Taraborelli and Ciampaglia2010] Dario Taraborelli and Giovanni Luca Ciampaglia. 2010. Beyond notability. collective deliberation on content inclusion in wikipedia. In *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop*, pages 122–125. IEEE.

[Tripodi2021] Francesca Tripodi. 2021. Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 0(0).