

Using Artificial Intelligence to Develop a Lexicon-Based African American Tweet Detection Algorithm to Inform Culturally Sensitive Twitter-Based Social Support Interventions for African American Dementia Caregivers

Peter BROADWELL^a, Nicole DAVIS^b and Sunmoo YOON^{c,1}

^a *Center for Interdisciplinary Digital Research, Stanford University, USA*

^b *School of Nursing, Clemson University, USA*

^c *General Medicine, Department of Medicine, Columbia University, USA*

Abstract. We extracted 3,291,101 Tweets using hashtags associated with African American-related discourse (#BlackTwitter, #BlackLivesMatter, #StayWoke) and 1,382,441 Tweets from a control set (general or no hashtags) from September 1, 2019 to December 31, 2019 using the Twitter API. We also extracted a literary historical corpus of 14,692 poems and prose writings by African American authors and 66,083 items authored by others as a control, including poems, plays, short stories, novels and essays, using a cloud-based machine learning platform (Amazon SageMaker) via ProQuest TDM Studio. Lastly, we combined statistics from log likelihood and Fisher's exact tests as well as feature analysis of a batch-trained Naive Bayes classifier to select lexicons of terms most strongly associated with the target or control texts. The resulting Tweet-derived African American lexicon contains 1,734 unigrams, while the control contains 2,266 unigrams. This initial version of a lexicon-based African American Tweet detection algorithm developed using Tweet texts will be useful to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers.

Keywords. unigram, lexicon, social media, caregiver, disparity, dementia

1. Introduction

The prevalence of dementia is disproportionately higher among African Americans than Whites in the United States, and family caregivers of persons with dementia suffer significant physical and psychological symptoms due to caregiving burdens and stress. According to a systematic review, aspects of caregiving and as well as the caregiver's quality of life have been found to vary with the race and ethnicity of the family caregivers of persons with dementia; these differences manifest at the intrapersonal, interpersonal and environment levels. In particular, evidence from a large psychosocial intervention trial indicated that African American family caregivers received no measurable benefit from the intervention (null results), while Hispanic/Latino and white caregivers showed

¹ Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu

positive responses to the intervention [1]. Twitter has the potential to be a platform for providing culturally sensitive social support interventions among African American family caregivers of dementia; this was especially true during the Covid-19 pandemic, as the role of online technology and networking became critical to facilitate the exchange of knowledge and to alleviate loneliness among caregivers. Yet, racial and ethnic demographic information about Twitter users is largely unavailable. Although deep learning and other statistical techniques have been widely applied to infer demographic information on Twitter, those demographic detection algorithms tend to be unavailable to open science communities and/or require access to account details that would compromise individuals' privacy [2]. The purpose of this study is to develop a lexicon-based African American Tweet detection algorithm to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers.

2. Methods

In order to gain a broad perspective on African American-related discourse that would guide the development of a social media search lexicon, we collected texts from two data sources: historical literature authored by African American writers (e.g., poems, plays, novels) provided a source of "old" terms, while recent Tweets captured "new" terms (**Figure 1**). For the literature corpora, an American history librarian at Columbia University first identified databases and collections authored by African Americans; the titles of these collections included African-American Studies, Black Short Fiction and Folklore, Black Women Writers, Black Drama, Black Thought and Culture, The Black Panther Newspaper and Black Music. From those collections, we extracted 14,692 poems and prose writings by African American authors using a cloud-based machine learning platform (Amazon SageMaker) provided via ProQuest Text and Data Mining (TDM) Studio from March 15, 2019 to December 31, 2019 with the help of the project director at TDM Studio. We used this same platform to collect a larger "control" corpus of 66,083 writings by authors from other ethnicities, including poems, plays, short stories, novels and essays. Due to copyright regulations in the United States, most of the materials in both sets were written prior to the mid-20th century.

For the Tweet corpora, we first extracted 3,291,101 Tweets using top hashtags associated with African American-related discourse (#BlackTwitter, #BlackLivesMatter, #StayWoke). We also collected 1,382,441 Tweets as a control set (these were Tweets without the above hashtags; they could contain other hashtags or none at all) from September 1, 2019 to December 31, 2019, using the Twitter application programming interface (API). As a first step of corpus processing, we removed redundant messages (retweets) and applied a bot detection algorithm to exclude irrelevant Tweets which were automatically generated by non-human accounts. A total of 803,495 Tweets (24.41%) associated with African American-related discourse and 369,348 Tweets (26.71%) in the control group were identified as unique and non-bot generated Tweets [3]. Lastly, we combined statistics from log likelihood and Fisher's exact tests as well as feature analysis of a batch-trained Naive Bayes (NB) classifier to select lexicons of terms most strongly associated with the target or control Tweets. We used 10-fold cross validation to evaluate the consistency and accuracy of the NB classifiers. Resources including de-identified data, videos on how to extract data, and the corpus processing and analytical code are available at <https://osf.io/qruf3/>. The larger study was approved by the Institutional Review Board (IRB).

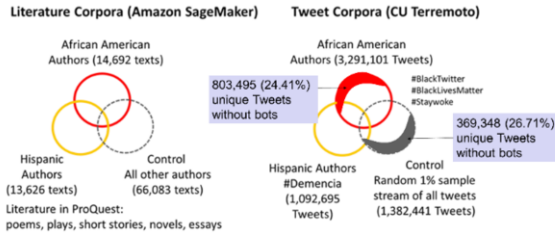


Figure 1. Quantities of texts used to create an African American lexicon and African American discourse-associated Tweet detection algorithm after applying a Twitter-bot detection process

3. Results

The “historical” African American literary-derived lexicon contains 1,643 unigrams (**Table 1**), while the Twitter-derived African American lexicon contains 1,734 unigrams and the control contains 2,266 unigrams. From 10-fold cross validation evaluation, the F-measure of the performance of the literary lexicon-based NB African American detection classifier was noted as 0.91 (area under the receiver operating characteristic curve: 0.97).

Table 1. Statistics of log likelihood, Fisher’s exact tests, and feature analysis of a batch-trained Naive Bayes classifier on sample terms in African American lexicon (*p-value =0.00)

term	aa freq	all freq	log ratio	fisher ratio	bayes nll	doc freq	log inv freq
black	28873	52819	1.85*	2.89*	-6.22	29880	1.71
white	22301	51368	1.47*	1.84*	-6.5	31968	1.64
old	20267	63743	1.08*	1.12*	-6.69	40211	1.41
woman	18606	40940	1.54*	2*	-6.59	24799	1.9
hair	7693	23581	1.11*	1.16*	-7.37	18011	2.22
hell	3786	10955	1.17*	1.26*	-7.88	8891	2.92
hurt	2756	7413	1.26*	1.42*	-8.11	6175	3.29
Jesus	1687	4298	1.33*	1.55*	-8.45	2973	4.02
Harlem	1586	1778	3.03*	19.77*	-8.33	1116	5
hoped	1140	2658	1.46*	1.8*	-8.8	2517	4.18
cops	796	1840	1.47*	1.82*	-9.04	1284	4.86
afro	656	757	2.94*	15.54*	-9.1	566	5.68

aa: African American; freq: frequency; nll: negative log likelihood; inv_freq: inverse document frequency

4. Discussion and Conclusiond

Although our lexicon-based classifiers for the historical literary corpus and the recent Tweet corpus were quite similar in their effectiveness at classifying held-out samples of pre-1950s literature and Tweets from 2019, respectively, each was less successful when applied to the other corpus, or when the training sources were combined to produce a unified classifier. This argues against the usefulness of historical texts for the classification of modern social media, at least with the methods employed in this study.

We also found that a lexicon composed of unigrams was more accurate in differentiating texts from held-out test samples of the two groups than lexicons composed of n-grams of various lengths. This result contrasts with recent studies reporting superior sentiment classification performance from n-gram based approaches [4]. Our finding is

likely due to the binary nature of the classification task and to the relatively terse nature of Tweets. In addition, although two domain experts in our team reviewed the final lexicon, further cleaning of an n-gram lexicon by multiple experts to retain culturally meaningful phrases (e.g., “family caregiver” as bi-gram) while removing irrelevant n-grams (e.g., “look at”), may improve its performance. Ongoing experiments involving classifiers using more sophisticated language models, specifically Bidirectional Encoder Representations from Transformers (BERT), have a strong likelihood of significantly altering the findings above.

This study developed a lexicon-based classifier to identify Tweets authored by African Americans to inform culturally sensitive Twitter-based intervention design for family caregivers of persons with dementia. As mentioned earlier, African Americans as a group experience disproportionately high rates of dementia. Relevant research literature urges the development of culturally sensitive educational interventions that carefully incorporate cultural traditions into family-oriented care strategies to deal with caregiving issues among African American caregivers, such as navigating not knowing the true wishes of a dementia patient. Furthermore, the Covid-19 pandemic has highlighted the role of social media in the propagation of health-related information. Nevertheless, the most effective content and framing strategies for culturally sensitive interventions using social media remain unclear. Identifying existing African American communities and discourse patterns on social media platforms like Twitter is the first basic step towards understanding a community and culture necessary to develop culturally sensitive interventions (e.g., terms, norms, culturally sensitive expressions, and topics). As a limitation, it is important to consider the ethical issues regarding the use of Twitter data in mental health surveillance [5]. Researchers must be aware of the unresolved distrust towards scientists and health professionals among African Americans in the U.S. due to historical factors. The Tuskegee experiment is perhaps the most infamous when evaluating the benefits and risks inherent to social media-based health research among African Americans. With these concerns in mind, we believe that our first version of a lexicon-based African American Tweet detection algorithm developed using literature and Tweet texts can be used both effectively and ethically to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers and future studies are needed to refine this algorithm.

Acknowledgments

U.S. federal grant TweetS2 R01AG060929 (P.I.: Yoon).

References

- [1] Alzheimer’s Association. Alzheimer’s disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. 2020.
- [2] Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PloS one*. 2015 Mar 2;10(3):e0115545.
- [3] Subrahmanian VS, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F. The DARPA Twitter bot challenge. *Computer*. 2016 Jun 13;49(6):38-46.
- [4] Dey A, Jenamani M, Thakkar JJ. Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*. 2018 Aug 1;103:92-105.
- [5] Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC medical ethics*. 2016 Dec;17(1):1-1.