Quantifying Depressed Social Media During COVID-19: Information Retrieval With ML & NLP

Brent D. Davis*[1], Dawn Estes McKnight[2], Rumi Chunara[3,4], Daniel J. Lizotte[1,5], and Alona Fyshe[2,6]

[1]Department of Computer Science, Western University, London, ON
[2]Department of Computing Science, University of Alberta, Edmonton, AB
[3]Department of Computer Science & Engineering, New York University, New York, NY
[4]Department of Biostatistics, New York University, New York, NY
[5]Department of Epidemiology and Biostatistics, Western University, London, ON
[6]Department of Psychology, University of Alberta, Edmonton, AB, Canada

1 Abstract

The ongoing pandemic continues to disrupt the normal functioning of society in numerous ways, and symptoms of depression are on the rise [Ettman et al., 2020]. In this work, we explored how analysis of social media can reveal changes in the number of authors presenting depressive symptoms on social media using Twitter and Reddit.

We first assessed the level of depressive symptoms expressed in a large set of tweets. While there are some efforts for identifying depressive symptoms in tweets, they are limited in scope and typically do not account for contemporary online discourse surrounding the experience of depression [Hasan et al., 2014, Pedersen, 2015]. To ensure that our assessment accounted for contemporary discourse, we extracted recent posts from `/r/Depression`, where symptoms and experience are a main topic of discussion. To further ensure that our assessment accounted for language that expresses depressive symptoms in a variety of contexts, rather than only when explicitly discussing the experience of depression, we also extracted all of the other Reddit posts of users who posted in `/r/Depression`. These user posts were extracted from all posts made by all authors in `/r/Depression` across all of Reddit for November and December 2019 (the most recent two months available in their entirety on Pushshift [Baumgartner et al., 2020]).

We then trained a GloVe word embedding on the posts made by users across Reddit who post in `/r/Depression` [Pennington et al., 2014]. Using the resulting word vectors, we then trained an author representation using the usr2vec [Amir and Wallace] method for both our `/r/Depression` authors and a sampled set of users to act as contrast against our archetypal example. This produces a high-dimensional representation of a user, based on a composite of the word representations we trained previously. Then, we used a linear kernel support vector machine (SVM) to find a separating hyperplane between these high dimensional representations of users who post in `/r/Depression` and the control set not active in `/r/Depression`. From here, we could use the SVM to directly classify unseen user representations; however, this is prone to bias, the classifications are challenging to explain, and training a representation for every new user is computationally expensive. We instead extracted vocabulary strongly associated with users who post in `/r/Depression` by taking the cosine of every word representation in the vocabulary of our word embedding with the decision direction the SVM produces. We took the most aligned words and used them to form a query for retrieving content written by depressed users. These words can be visualized and reviewed, mitigating bias and improving explainability. We call this method 'Archetype-based Information Retrieval' (AIR); our work is an example of using AIR to find depression-associated content, based on a similar approach for finding posts about substance abuse [Davis et al., 2019].

We created a query from the 200 most closely aligned-words and used BM25 to assign a score [Robertson and Zaragoza, 2009] to tweets from the Mega-COV [Abdul-Mageed et al., 2020] and official Twitter COVID-19 datasets. We took the top-scoring quartile of tweets from our search as being posts that indicate depressive symptoms. We sorted the tweets by the time they were posted, and looked for changes in the frequency of high-scoring matches to our query over time. We then ran topic models (Latent Dirichlet Allocation, Contextual) on tweets grouped by the month they were posted in and looked for consistencies and changes over time in the topics discovered by these automated approaches [Blei et al., 2003, Bianchi et al., 2020].

Future work will explore ties between social media metrics and traditional, offline metrics. We intend to group tweets by geotags and look for corresponding trends; it is an open question whether the local, municipal, provincial, federal or international situation regarding COVID-19 forms the primary stressors on individuals. This study lays the foundation for AIR as a tool for investigating COVID-19 impacts on mental health.

*Corresponding author: bdavis56@uwo.ca

References

M. Abdul-Mageed, A. Elmadany, D. Pabbi, K. Verma, and R. Lin. Mega-cov: A billion-scale dataset of 100+ languages for covid-19. *arXiv preprint arXiv:2005.06012*, 2020.

C. G. C. P. S. M. Amir, S. and B. Wallace. Quantifying mental health from social media with neural user embeddings. *Journal of Machine Learning Research, WC Track*, 68.

J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):830–839, 5 2020. URL https://www.aaai.org/ojs/index.php/ICWSM/article/view/7347.

F. Bianchi, S. Terragni, and D. Hovy. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv preprint arXiv:2004.03974*, 2020.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

B. D. Davis, K. Sedig, and D. J. Lizotte. Archetype-based modeling and search of social media. *Big Data and Cognitive Computing*, 3 (3):44, 2019.

C. K. Ettman, S. M. Abdalla, G. H. Cohen, L. Sampson, P. M. Vivier, and S. Galea. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA Network Open*, 3(9): e2019686–e2019686, 09 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.19686. URL https://doi.org/10.1001/jamanetworkopen.2020.19686.

M. Hasan, E. Agu, and E. Rundensteiner. Using hashtags as labels for supervised learning of emotions in twitter messages. In *ACM SIGKDD workshop on health informatics, New York, USA*, 2014.

T. Pedersen. Screening twitter users for depression and ptsd with lexical decision lists. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 46–53, 2015.

J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

S. Robertson and H. Zaragoza. *The probabilistic relevance framework: BM25 and beyond*. Now Publishers Inc, 2009.