Can Pre-trained Models in the Clinical Domain Detect Social and Behavioral Determinants of Health from Tweets?

Minhwa Lee, BA[1], Zonghai Yao, MSc[1], Zhangqi Duan, BS[1], Avijit Mitra, MSc[1], Hong Yu, PhD[1,2,3,4]

[1]College of Information and Computer Science, University of Massachusetts Amherst, Amherst, MA, USA; [2] Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, USA; [3] Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA; [4] Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, USA

Introduction

Social and behavioral determinants of health (SBDH)[1,2] are the conditions of the environments in which people are born, live, work, and age, as well as individual-level behavioral determinants of health (e.g., smoking, substance use and alcohol consumption) and mental health status (e.g., depression and loneliness). Also, community-level factors of SBDH such as neighborhood environment have been shown to play an important role in individuals' health outcomes such as diabetes[3].

In this work, we investigated whether models trained on clinical corpora can generalize well on tweets for reliable SBDH prediction. Then, we examined whether there is association between the models' performance and the two following features: (1) the number of grammatical errors on tweets and (2) the neighborhood environment that the tweets are created within.

Methods

Our dataset contains the 100,068 tweets that have been created on November 2nd, 2021, written in English, created within the United States, tagged with geolocations, and not re-tweeted. Also, we obtained the list of SBDH keywords collected and curated by domain experts after literature reviews. In our dataset, 5,865 tweets contained at least one SBDH-related keyword. Also, Twitter users sometimes express their experience about health outcomes with figurative languages such as sarcasm and metaphor. Thus, we used the previous study's analysis[4] to search for health-mentioned tweets manually, which led to obtaining 1,485 tweets.

Based on the fact that tweets are not regulated by word choices and grammatical rules of standardized language in many cases, we considered correcting grammatical errors and misspellings in our dataset via a language-check API[5]. Thus, we ran on those grammatically corrected tweets in both keyword-mentioned and figurative-word tweets the following two models: (1) a pre-trained BioClinicalBERT[6] fine-tuned on the MIMIC-SBDH[1] dataset ('Model 1') and (2) a pre-trained RoBERTa[7] fine-tuned on veterans' 4,646 electronic health records ('Model 2') that we developed for the SBDH detection task.

Finally, we used Area Deprivation Index[8] (ADI) metrics to comprehensively judge the levels of disadvantage in the Census block groups that tweets in the dataset were created in. Higher ADI indicates a more disadvantaged neighborhood environment. In addition, we leveraged the US Census's geocoder[9] API to extract Census block-group information of each tweet's geolocation (e.g., 12-digit Federal Information Processing System (FIPS) geographic unit identifiers). Then, we matched each tweet with its corresponding block-group information and ADI.

For performance evaluation, the predictions from the two models were annotated by 10 people who read and acknowledged necessary concepts of SBDH before the annotation process. For detailed analysis, we manually selected 21 cases from the keyword-tagged tweets and 5 from the tweets with figurative words.

Results

The two clinical-domain models did not perform well in identifying multiple SBDH from the keyword-tagged subset of tweets. Among the 21 keyword tweets, the annotators disagreed with 84-92% of the prediction results from Model 1 and 38 -53% of the results from Model 2. Also, they disagreed with 80-100% and 20-40% of the results that Model 1 and 2 predicted on the 5 figurative tweets, respectively. This result confirmed that pre-trained clinical

models are not well-suited for SBDH detection from tweets. We believe that domain differentiating factors such as shortness, creative usage of words including metaphors, figurative expressions and hashtags, make it difficult to portray the user's health-related situations. Besides, it is almost impossible to detect SBDH from just one tweet without any previous history or better context. This also poses a serious challenge to human evaluation, even for the medical professionals.

Discussion

(1) Association between models' performance and grammar errors All four subsets with different levels of grammar corrections had an almost even number of tweets of which the annotators disagreed with Model 1's interpretations. However, from the total of 4 tweets of which annotators disagreed with the results from Model 2, all of them were identified to make at least four grammatical errors. Thus, making grammatical errors or misspellings could result in degrading Model 2' prediction capability.

(2) Association between models' performance and neighborhood environment Among the 19 tweets annotaters disagreed with the Model 1's results for, only 11 tweets were created in the block groups with valid ADI percentile, and eight of them were located in most disadvantaged neighborhoods (= ADI percentile > 70). The annotators also disagreed on the results of the three tweets from Model 2, which were created in mid-level neighborhoods of ADI between 30th and 70th percentile. We may confirm an association between the community-level environment and the clinical-domain models' performance on detecting SBDH from the corresponding locations' tweets.

(3) Validity of tweets as a reasonable health-related resource The public characteristics of Twitter could raise our doubts on whether people mention everything relevant to their life experiences, including socially unacceptable events. For example, the tweets mentioned with drug/alcohol/tobacco-related words were not primarily indicating that the users consumed them in reality: rather, those tweets were closer to slang or common language.

Our work called attention to developing pre-trained clinical-domain models that can also perform well on social media texts for SBDH detection. As a future work, we intend to improve these two models for social media text. We believe that in-depth analyses of tweets will be needed to develop the model that precisely understands the characteristics of users, such as the timeline and history of the tweets, temporal and spatial distribution of the tweets, and the community-level health outcome analysis. The continuing work will provide a robust pipeline for improving the current system of public health surveillance through social media texts.

References

1. Ahsan H, Ohnuki E, Mitra A, Yu H. MIMIC-SBDH: a dataset for social and behavioral determinants of health. Proceedings of Machine Learning Research. 2021;149:391–413.
2. Mitra A, Ahsan H, Li W, Liu W, Kerns RD, Tsai J, et al. Risk factors associated with nonfatal opioid overdose leading to intensive care unit admission: A cross-sectional study. JMIR Med Inform [Internet]. 2021 [cited 2022 Mar 9]; 9(11):e32851. Available from: https://medinform.jmir.org/2021/11/e32851
3. Walker RJ, Smalls BL, Campbell JA, Strom Williams JL, Egede LE. Impact of social determinants of health on outcomes for type 2 diabetes: a systematic review. Endocrine [Internet]. 2014;47(1):29–48. Available from: http://dx.doi.org/10.1007/s12020-014-0195-0.
4. Yadav S, Chauhan J, Sain JP, Thirunarayan K, Sheth A, Schumm J. Identifying depressive symptoms from tweets: Figurative language enabled multitask learning framework. arXiv [csCL] [Internet]. 2020 [cited 2022 Mar 8]; Available from: http://arxiv.org/abs/2011.06149
5. Language-check [Internet]. PyPI. [cited 2022 Mar 9]. Available from: https://pypi.org/project/language-check/
6. Alsentzer E, Murphy JR, Boag W, Weng W-H, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings [Internet]. arXiv [cs.CL]. 2019. Available from: http://arxiv.org/abs/1904.03323
7. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. arXiv [csCL] [Internet]. 2019 [cited 2022 Mar 9]; Available from: http://arxiv.org/abs/1907.11692
8. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the neighborhood atlas. N Engl J Med [Internet]. 2018;378(26):2456–8. Available from: http://dx.doi.org/10.1056/NEJMp1802313
9. Welcome to Geocoder [Internet]. Census.gov. [cited 2022 Mar 9]. Available from: https://geocoding.geo.census.gov/geocoder/