Towards Detecting the Gender-based Violence Victim Condition in Speech using Domain-Adversarial Training

Emma Reyner-Fuentes
ereynar@pa.uc3m.es
Department of Signal Theory and Communications, Universidad Carlos III de Madrid
Leganés, Madrid, Spain

Esther Rituerto-González
Esther.Rituertogonzalez@med.uni-muenchen.de
Department of Psychiatry and Psychotherapy, Section for Precision Psychiatry, Ludwig-Maximilian University & Max Planck Institute of Psychiatry
Munich, Bavaria, Germany

Carmen Peláez-Moreno
Department of Signal Theory and Communications, Universidad Carlos III de Madrid
Leganés, Spain

Department of Psychiatry and Psychotherapy, Section for Precision Psychiatry, Ludwig-Maximilian University & Max Planck Institute of Psychiatry
Munich, Bavaria, Germany

1 PROBLEM STATEMENT

Physical or sexual aggressions are considered manifestations of gender-based violence (GBV), a pervasive problem that affects mostly women and girls in our society. Different forms of GBV consistently lead to several negative effects in women’s behavior, as it has been recently explored [2], as well as to a range of mental illnesses globally, including anxiety, depression, suicide, Post-Traumatic Stress Disorder (PTSD), and substance abuse [4, 5, 8]. This pervasiveness of mental disorders is high among women who have suffered from some type of GBV [6, 9]. Their symptoms can be extremely distressing and can deteriorate a person’s ability to work, socialize, and carry out daily activities. Notably, PTSD is the most common sequela among GBV [1, 5, 12].

Speech technologies have proven to be extremely useful for biomedical applications, specifically for mental health assessments.

1. Both authors contributed equally to this research.


Due to being non-invasive, said technologies help reduce the influence of the diagnosis test in the patient’s perception and, therefore, allow to obtain a less biased response and better results. However, unlocking the full potential of speech technologies for biomedical applications is a great challenge, particularly when the speaker is anonymous. Personalization algorithms offer a compelling advantage across various speech-related tasks that depend on paralinguistics such as speech emotion recognition. However, relying on personal traits for medical applications may have confounding effects in the detection of mental states or health conditions. Moreover, the imperative to safeguard privacy, especially in sensitive domains like healthcare and medicine, cannot be overlooked.

In this paper, we present a groundbreaking study that delves into the speaker-agnostic detection of the gender-based violence victim condition (GBVVC) in speech data. Leveraging cutting-edge machine learning and domain-adversarial training strategies, our research unveils an innovative approach to enhance performance on the detection of this condition while not relying on speaker identity traits. We aim to tackle this crucial societal issue for the development of artificial intelligence-powered tools to assist and support mental health professionals.

2 METHODS

Drawing upon compelling evidence from prior research where we demonstrated that the GBVVC can be reliably detected from speech [10], we part from the hypothesis that the existing models fairly rely on speaker identity traits to predict such condition. Therefore, the work we hereby present aims to make the model disentangle the speaker identity-related information from that pertaining to the condition, causing the models to unlearn the speaker traits in the voice by means of adversarial training strategies in a minimax loss game.

For this purpose, we used a unique multimodal dataset, WEMAC+, comprising 39 gender based violence victims and 104 non-victims whose speech was recorded in Spanish after experiencing a variety of emotions in a virtual-reality setting. With said data, we assessed and compared different neural network model architectures and hyperparameters. The models are coded in Python with TensorFlow Keras.

KEYWORDS
Gender-Based Violence, Speaker-Agnostic, Speech Paralinguistics, Post-Traumatic Stress Disorder, Artificial Intelligence Therapy Assistants

ACM Reference Format:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

womENcourage 2024, June 26-28, 2024, Madrid, SPAIN

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX
3 RESULTS

We present the results of a 28.26% relative reduction in the model’s ability to identify the speaker (from 91% to 65%) which, although it does not erase the speaker’s information completely, means a severe attenuation of its influence. While achieving such attenuation, we were able to obtain a 6.39% relative improvement over the baseline GBVV detection accuracy compared to not using the domain-adversarial framework. This allows us to demonstrate that, in fact, neglecting the information about the speaker traits is beneficial for condition detection.

We also present one extra experiment intended to enlighten what was the model basing its predictions on. In our previous work [11] we saw some correlation between the accuracy of our model back then and the “Post-Traumatic Stress Disorder Revised Symptom Severity Scale”–or its acronym in Spanish, EGS-R [3]–score of the victims, being that the more pre-clinical symptoms of PTSD were experienced by a certain victim, more likely was the model to classify her as a victim. In this case, to check if the results were still following the hypothesis of being the pre-clinical symptoms of PTSD what allows the model to detect a GBVV, we checked the mean EGS-R score for those GBVV classified correctly and those who were not. We observed that the mean EGS-R score is higher for those victims who have been correctly classified (10.52 vs 6.9 for the domain-adversarial model/10.35 vs 8.07 for the baseline model), which is aligned with the aforementioned hypothesis. Plus, this gap enlarges when using the domain-adversarial model rather than the baseline one, meaning that removing the speaker information emphasizes the importance of the pre-clinical symptoms of PTSD for the model’s prediction.

4 SIGNIFICANCE

Our results prove that the more information the models manage to neglect, the better the GBVV condition prediction is, demonstrating that the models learn to focus in the more discriminative speech paralinguistic features for the detection of GBVV instead of subject-specific traits.

Additionally, we cross our model’s predictions with the victims’ psychological evaluation questionnaires and the results point towards a connection between the agnostic GBVV prediction and the pre-clinical Post-Traumatic Stress Disorder (PTSD) symptoms.

We encourage the research community into taking our research as a baseline to continue investigating on gender-based violence repercussions on the mental health of women.

ACKNOWLEDGMENTS

This work has been partially supported by the SAPIENTAE4Bindi Grant PDC2021-121071-100 funded by MCINAE10.13039/501100011033 and by the European Union “NextGenerationEU/PRTR”, PID2021-125780NB-I00 funded by AEI, the Spanish Ministry of Science, Innovation and Universities with the FPU grant FPU19/00448. The authors thank all the members of UC3M4Safety for their contribution and support, particularly to the UC3M4Safety Team Psychologist, Eva Rubio for her very insightful help on this paper.

REFERENCES


Received 14 February 2024; revised X XXXX 2024; accepted X XXXX 2024