Detecting anxiety from short clips of free-form speech

Aditya Mantri, Gaurav Bansal, Beri Prashant
MNNIT

Contents

- Motivation
- Problem Statement
- Data
- Methods
- Experiments and Analysis
- Upcoming work

Motivation

- Mental wellbeing is quintessential for overall health but many social issues like stigma and monetary barriers exist.

- Mental assessments are intermittent and may be limited partly due to the episodic nature of psychiatric symptoms.

- The gravity of the problem combined with the massive scope of improving upon current methods motivates us to study the problem of detecting anxiety disorders through speech.

Problem Statement

- Given the raw audio recording of the user, predict if the user has anxiety disorder or not as measured by the GAD-7 [4] questionnaire.

- Score of 5 or greater on GAD-7 implies some form of anxiety

GAD-7				
Over the last 2 weeks, how often have you been bothered by the following problems? (Use "✔" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

(*For office coding: Total Score T____ = ____ + ____ + ____)

Data

- Generalized Anxiety Disorder (GAD) dataset from Kintsugi Mindful Wellness, Inc.

- Self audio journal entry from participating users.

- Labels obtained by scores of GAD-7 questionnaire filled by the users shortly after the recorded journal.

- 2257 labelled examples with raw audio and test scores

Methods

- Audio Features (GeMAPS)
- Transcript based classifier
- Wav2Vec embeddings
- Multi-modal anxiety detection

Audio Features (GeMAPS)

- Standard acoustic parameter set for various areas of automatic voice analysis including clinical speech analysis [3]
  - Jitter/Shimmer
  - Loudness
  - Pitch (F0)

- Transcript based features
  - Emotion
  - Sentiment

- Logistic regression model with L1 regularization

Results

Model	Precision	Recall	F1	AUROC
Random	0.50	0.48	0.49	0.50
GeMAPS features	0.63	0.53	0.58	0.66

Audio Features (GeMAPS)
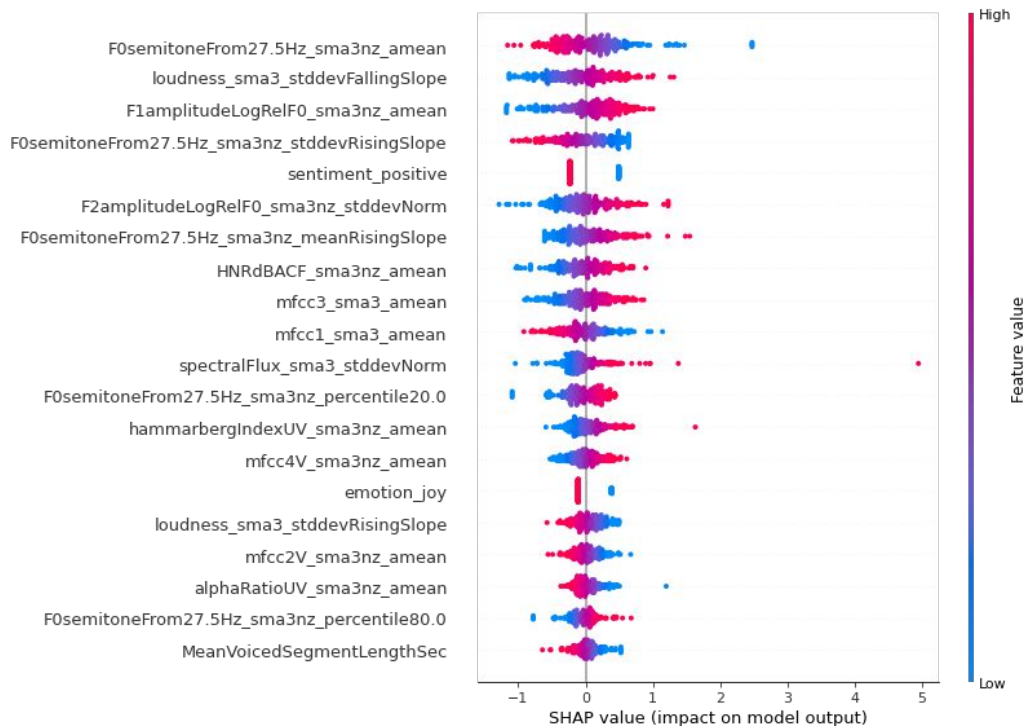
Transcript based classifier

- In this approach, we don't deal with audio data directly, but instead have a 2-step process:

- In the first step, we use Wav2Vec2-Base model to transcribe the given audio files. We then use pre-trained Roberta Large optimized for Semantic Textual Similarity (STS) to generate sentence-level embeddings for the transcripts.

- In the second step, the generated sentence embeddings and their corresponding Anxiety/No Anxiety labels are used to train a GBC model to serve as a binary classifier.

Results

Model	Precision	Recall	F1	AUROC
Random	0.50	0.48	0.49	0.50
GeMAPS features	0.63	0.53	0.58	0.66
Transcript classifier	0.64	0.57	0.60	0.68

Future work

- Use GAD-7 based sample weights to see if the model improves

- Also aim to qualitatively analyze and compare the performance of different models

References

1. Siriwardhana, Shamane, et al. "Jointly Fine-Tuning" BERT-like" Self Supervised Models to Improve Multimodal Speech Emotion Recognition." *arXiv preprint arXiv:2008.06682* (2020).
2. Baevski, Alexei, et al. "wav2vec 2.0: A framework for self-supervised learning of speech representations." *arXiv preprint arXiv:2006.11477* (2020).
3. Eyben, Florian, et al. "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing." IEEE transactions on affective computing 7.2 (2015): 190-202.
4. Spitzer, Robert L., et al. "A brief measure for assessing generalized anxiety disorder: the GAD-7." *Archives of internal medicine* 166.10 (2006): 1092-1097.
5. Schneider, Steffen, et al. "wav2vec: Unsupervised pre-training for speech recognition." *arXiv preprint arXiv:1904.05862* (2019).