A Machine Learning Approach to Help Autistic Individuals Recognize Emotions in Vocal Conversation

Barrett, Arjun (School: The Harker School)
Lan, Alexander (School: The Harker School)

One in fifty-nine Americans suffers from autism spectrum disorder (ASD) and, as a result, has trouble with everyday social interaction. Without the ability to accurately read the emotions of their peers, they often appear abrasive, rude, or callous, so they struggle to build relationships, are excluded from social events, and face difficulties in the workplace. To help autistic individuals overcome these challenges, we developed a ConvLSTM neural network to recognize emotions in vocal conversation. We used the RAVDESS dataset, which contains audio recordings of voice actors emulating 8 core emotions (happiness, surprise, sadness, disgust, anger, fear, calm, and neutral) to train and test our models. We extracted mel-frequency cepstrum coefficients (MFCCs) from the audio signal to determine the salient features of the vocal tract characteristics in the recordings. The MFCCs were split into training and testing sets to more accurately represent a real-world scenario during evaluation. We generated 14 models with varying layer configurations and tweaked their hyperparameters for a total of 84 combinations using the Keras API for TensorFlow in Python 3.6. The best model configuration achieved 94% accuracy in classifying audio samples into one of four emotional categories. To collect real-world data, we created a web application for non-ASD individuals to record themselves emulating certain emotions to simulate situations ASD individuals face. Our model reached 93% accuracy on this real-world data relative to non-autistic people. In summary, our high-performance model has the potential to change the lives of millions of ASD individuals by determining emotion from speech nearly as accurately as normal people.

Awards Won: