4th Annual Congress on MENTAL HEALTH

Cross-Corpus Multilingual Speech Emotion Recognition: Amharic vs. Other Languages

Retta Ephrem Afele
Northwest University, China

In a conventional speech emotion recognition (SER) task, a classifier for a given language is trained on a pre-existing dataset for that same language. However, where training data for a language do not exist, data from other languages can be used instead. We experiment with cross-lingual and multilingual SER, working with Amharic, English, German, and Urdu. For Amharic, we use our own publicly available Amharic Speech Emotion Dataset (ASED). For English, German and Urdu, we use the existing RAVDESS, EMO-DB, and URDU datasets. We followed previous research in mapping labels for all of the datasets to just two classes: positive and negative. Thus, we can compare performance on different languages directly and combine languages for training and testing. In Experiment 1, monolingual SER trials were carried out using three classifiers, AlexNet, VGGE (a proposed variant of VGG), and ResNet50. The results, averaged for the three models, were very similar for ASED and RAVDESS, suggesting that Amharic and English SER are equally difficult. Similarly, German SER is more difficult, and Urdu SER is easier. In Experiment 2, we trained on one language and tested on another, in both directions for each of the following pairs: Amharic↔German, Amharic↔English, and Amharic↔Urdu. The results with Amharic as the target suggested that using English or German as the source gives the best result. In Experiment 3, we trained on several non-Amharic languages and then tested on Amharic. The best accuracy obtained was several percentage points greater than the best accuracy in Experiment 2, suggesting that a better result can be obtained when using two or three non-Amharic languages for training than when using just one non-Amharic language. Overall, the results suggest that cross-lingual and multilingual training can be an effective strategy for training an SER classifier when resources for a language are scarce.

Biography

Dr. Ephrem Afele Retta is a distinguished Computer Science and Technology Ph.D. holder with over 8 years of industry expertise. Renowned for his prolific research, he has contributed significantly to diverse areas such as artificial intelligence, machine learning, deep learning, data science, and audio processing. Her impactful work is reflected in numerous publications in reputable journals and conferences. His dynamic career spans academia, research institutions, and industry, offering a comprehensive understanding of theoretical concepts and their practical applications. An active member of professional organizations, she is dedicated to advancing technology and fostering the next generation's growth through mentorship and knowledge-sharing initiatives. His visionary approach positions him as a leading figure in shaping the future of Computer Science and Technology.