Evaluating the Efficiency of Different Machine Learning Models in Locating the /s/ Phoneme

Finch, Ethan (School: Stillwater Area High School)

Natural Language Processing has led to the development of voice recognition technology, and taking speech analysis further can lead to the development of technology to aid in treatment of speech-language pathology. The purpose of this experiment is to train four different machine learning models and compare their performance in locating the /s/ phoneme in an audio file of a word by using the Area Under the Precision-Recall Curve (AUC-PR) as the scoring method. A total of 300 audio recordings, with 100 initial /s/ words, 100 medial /s/ words, and 100 final /s/ words were collected using Forvo. The 300 words were transformed into Mel spectrograms and split into 80% training, 15% validation, and 5% testing datasets (CloudFactory, n.d.). A Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) model, a Logistic Regressor, and a Gated Recurring Unit (GRU) were created and run through hyperparameter optimization. One model of each type was created, trained, and evaluated using the optimal hyperparameters. The hypothesis was the CNN would identify the location of the /s/ phoneme with the best AUC-PR score because of its ability to classify images. Each trained model was tested and compared to each other using the average AUC-PR of all words in the testing dataset. The LSTM had an average AUC-PR score of 0.695 on the test data, while the GRU, Logistic Regressor, and CNN all had an average AUC-PR score of 0.218, proving the hypothesis incorrect. The analysis and conclusion discuss considerations why the LSTM performed the best.

Awards Won: