Audio signal analysis of cow vocalizations in Python's Librosa with gradient boosting algorithms to identify abnormal sound signatures in lame cows

Richa Pathak and Jan Shearer
Iowa State University
Ames, Iowa, USA
pandey@iastate.edu

Objectives
Develop methods to record high-quality audio data from individual cows, investigate platforms to preprocess the audio signals, and build machine learning tools to identify abnormal vocalization.

Materials and methods
Raw audio data was collected from individual cows in dairy farms using the Zoom H4n Pro Handy Recorder. This series of Zoom recorders has sophisticated capabilities for field recording, such as XY microphones, preamplification modules, and enhanced noise cancellation. Each audio recording was captured for around 30 seconds and saved in a .wav file format within a local server. All the saved data were initially screened to discard the audio recordings of poor signal quality. Thereafter, data labeling was performed in the presence of an expert to label the audio recordings into different categories. The next stage was data preprocessing which included the steps of label encoding, feature scaling and database split (into Test, Train and Validation Sets). All models were built using python3. For our model development, we used Librosa which is a powerful Python package for music and audio signal analysis.

Results
The preprocessed audio data was fed into Librosa to extract several features, such as Short Time Fourier Transform (STFT), Inverse Short Time Fourier Transform (ISTFT), Instantaneous Frequency Spectrogram (IFGRAM), Mel Spectogram, Chromagram, and Tonnetz representation. We investigated in-built tools available for music and audio signal processing, such as ways to estimate the tempo and positions of beat events, along with parameters to optimize the sampling rate, frame rate and hop length. Power spectrum analysis was performed on the labelled audio recordings to estimate the spread of signal power across the frequency band. We applied a non-parametric approach based on both time-domain analysis and frequency-domain analysis. The Mel spectrograms were visualized. Subsequently, machine learning models were developed with the extracted features. The training data set was fed to machine learning models for appropriate classification of the audio signals, such as normal vocalization or pain-related vocalization. We applied Xtreme Gradient Boosting (XGB) and Light GBM (Gradient Boosting Method) for the sound classification. A web interface was used to visualize the audio recordings in time and frequency domains and display the prediction outcomes from the models. We achieved reasonably high accuracy with our fine-tuned XGB model for sound classification. The Light GBM achieved better accuracy with less execution time. As such, our initial results indicate that the Light GBM is a better classification model for processing large audio datasets and for real-time, on-farm edge computing applications.

Conclusion

Built on the Python's Librosa software package, we demonstrated an audio signal processing framework that is applicable to the dairy industry, particularly to identify abnormal cow vocalizations related to pain, distress, or lameness. We described methods for audio recording, labelling, preprocessing, feature set extraction, visualization, sound classification, and model development. Our results indicate that the available classification tools currently being used in other machine learning applications, such as the Light GBM, are capable of predicting abnormal cow vocalizations with reasonable accuracy.