Environmental Audio Classification using Machine Learning

OVERVIEW

The goal was to analyze non-speech human sounds from the ESC-50 dataset, focusing on sound categories such as breathing, clapping, coughing, and more. The project explores the potential application in areas like baby monitoring and remote monitoring for individuals requiring constant care.

DATASET COLLECTION AND PREPROCESSING

The ESC-50 dataset was employed, and pandas and NumPy were used for data handling. Data was filtered based on specific sound categories and loaded using Pandas, then processed to extract audio features. Librosa facilitated feature extraction, where MFCCs, Spectral Centroid, Bandwidth, and Zero-Crossing Rate were computed for each audio sample.

DATA STRUCTURES AND PROCESSING

While no custom data structures were designed, inbuilt data structures such as NumPy arrays and Pandas DataFrames were pivotal. NumPy arrays stored features, enabling efficient manipulation and transformation. Pandas DataFrames facilitated data organization and filtering.

The utilization of these inbuilt structures provided a robust foundation for data representation, allowing seamless operations such as feature extraction, scaling, and train-test splitting.

ALGORITHMIC APPROACHES

Three distinct algorithms—K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machines (SVM)—were employed for classification tasks. Hyperparameter tuning was pivotal in enhancing model accuracy.

CODE WALKTHROUGH

The initial steps encompass data loading, filtering, and feature extraction from audio files using Librosa. This process involves extracting MFCCs, Spectral Centroid, Bandwidth, and Zero-Crossing Rate, which are concatenated into combined features. These features and labels are converted to NumPy arrays and scaled using Min-Max scaling. The dataset is then split into training and testing sets.

Subsequently, hyperparameter tuning was conducted to optimize the algorithms.

Following this, a common demonstration was implemented for all three algorithms. Functions for manual audio recording and feature extraction were defined utilizing the sounddevice library. The recorded sound was classified using respective algorithms - KNN, Random Forest, and SVM. The demonstration showcased how the algorithm predicts classes based on the manually recorded audio.

The project emphasized standardizing data, tuning hyperparameters, and applying machine learning algorithms for audio classification.

KNN AND RANDOM FOREST

For KNN and Random Forest, hyperparameters like the number of neighbors and estimators were optimized by exploring various values and selecting the best based on accuracy scores achieved via loops.

SVM WITH GRID SEARCH

SVM, being sensitive to multiple hyperparameters, underwent a Grid Search approach. Different combinations of hyperparameters, including C-values and kernels, were systematically tested, culminating in the identification of the best combination for improved accuracy. A maximum of 0.725 accuracy was achieved after the hyperparameter tuning.

CONCLUSION

The project emphasizes the significance of data processing in machine learning tasks. Although the core focus revolved around classification algorithms, the utilization of appropriate data structures and effective data processing methodologies was instrumental in achieving superior accuracy.

The code files provide insights into how these algorithms were implemented and how data was processed at various stages. The project's versatility across different algorithms reflects the importance of selecting appropriate models based on dataset characteristics and problem domains.