Urban Sound Classification
Final Capstone Project Proposal

Abstract

Sonic event classification is a field of growing research. Most of these researches focuses on music or speech recognition. There are scarce works on environment sounds with very few databases for labeled environment audio data.

The objectives of this project is to evaluate and train various machine learning models to classify the urban sounds into categories correctly.

Source

Audio data for this project is collected from UrbanSound8k, released by NYU CUSP. The data was sourced from field recordings uploaded to the Free Sound online archive.

The sources were selected from the Urban Sound Taxonomy based on the high frequency with which they appear in noise complaints as determined from the data provided by New York City's 311 service (over 370,000 complaints from 2010 to date). *Dataset Download Link*: Urban Sound 8K Audio Dataset

Data Set Information

The dataset is comprised of 8732 slices (audio excerpts) of up to 4 s in duration extracted from field recordings crawled from the Free Sound online archive. Each slice contains one of 10 possible sound sources: **air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, street music**. In addition to the sound excerpts, a CSV file containing metadata about each excerpt is also provided.

Problem Statement

Classify the audio files in urban setting and measure the performances of various models.

a. What feature extraction techniques should be used for optimal results?
b. How do the machine learning models compare against the neural network learning models?
c. Which model performed the best?

Challenges

The biggest challenge with this data set is feature extraction since we are dealing with the audio files in .WAV format. There are many files with less than 1 second duration, which calls for data cleanup methods.

Approach

a. First, perform exploratory data analysis on the audio files to quickly assess audio patterns.
b. Use feature extraction techniques for audio feature generation and embedding post processing.
c. Apply various machine learning based classification techniques to train the model to classify the audio file.
d. Evaluate and choose the best performer by measuring the effectiveness of different models.

Feature Extraction

Feature extraction is the most important part for designing a machine learning model. To extract the useful features from sound data, we used Librosa and VGG Audioset library. These two methods were compared in the previous clustering project "Unsupervised Learning".

Model Evaluation

We will evaluate 4 different models in this project – Support Vector Machine, Random Forest, Deep Neural Networks and Convolutional Neural Networks.

Steps to evaluate the model:

- Split the data set into 80% training data and 20% test data.
- With various hyperparameter optimization techniques, find the optimal parameters on the model's training data set.
- Train the model using the best parameters.
- Evaluate the model on the test data set
- View and analyze the performance metrics and the confusion matrix.
- Finally, compare performance of all the models and choose the best performing model based on accuracy rate, precision, recall rate etc.

Practical Uses

The automatic classification of audio events in an urban setting has a wide range of applications in areas such as,

- Audio Event Detection
- Music Retrieval
- Home security or Audio Surveillance
- Assisted living, elder or infant care

Project Report

View the complete Urban Sound Classification Project Report for more details.