CS594 - Big Data Visualization & Analytics: Proposal

Roy G. Biv, Ed Grimley, *Member, IEEE*, and Martha Stewart

1 Introduction

Noise Pollution is a grievous problem which needs to be tackled, as it's a growing concern for many urban residents, many have reported that they suffered with behavioral and emotional consequences, such as difficulty in sleeping, relaxing and feeling annoyed, angry or upset [2] [1] [4].In order to mitigate this problem, there is a need to understand sound event detection. Sound event detection is defined as recognition of individual sound events in audio, e.g., "dog barking, engine exhaust noise" requiring estimation of onset and offset for distinct sound for sound event detection and identification of sound.

Applications for sound event detection can found in areas of Healthcare, security, audio and video-based indexing and retrieval. Sound class classification is usually approached as a supervised learning, with sound classes defined beforehand, we have taken a labelled Spatial and Temporal recording data which comprises of 3068 labelled 10 sec recordings from the Sounds of New York City (SONYC) acoustic network (An acoustic network is a method of positioning equipment using sound waves). Using this data, we plan to develop an application where the authorities or the user can narrow down the sounds generated at any particular location. We also aim to address the Mismatch of the testing data in this dataset.

This can be done by applying machine learning algorithm on any dataset and integrating it with sensors, data analytics for the development of machine learning systems for real world urban noise monitoring. The model build from the data could be used on neighborhoods to better understand noise in that location and help the authorities mitigate the issue. The many challenges for building the model would be like to separate the sound sources of interest, identifying the similar sounds compared to the other data in the dataset, identifying the main source for generating the sound among others.

2 Related work

Our application will be based on the research paper [3], which presents the process used to collect this data. SONYC has developed an acoustic sensor with high quality and low production cost to monitor the noise pollution levels across the city. The sensors follow DCASE (Detection and classification of acoustic scenes and events) to eliminate discrepancy. There are various other datasets like UrbanSound, UrbanSound8k that address this particular problem but have limited spacial and temporal data points. A VGGish model has been developed and trained using stochastic gradient descent to minimize cross-entropy loss. To eliminate over-fitting early stopping on validation set has been implemented. The overall AUPRC achieved by this model is 0.62 and 0.76 on different level classes, which performed poorly on music and non-machinery impact sounds.

According to [2] a supervised learning methodology is applied to real-life high quality recordings of 3-5 minutes with very little noise of 15 different acoustic scenes (lakeside beach, bus, cafe/restaurant, car, city center, forest path, grocery store, home, library, metro station, office, urban park, residential area, train, and tram) and two common enviornment areas (outdoor - residential areas and indoor - home). A mel frequency cepstral coefficient (MFCC) and Gaussian mixture model (GMM) was trained using expectation maximization algorithm. The overall accuracy achieved by the model is 72.5 % ranging from 13.9% for parks to 98.6% for office spaces.

The partitioning of the data was done based on the location of the original recordings. All segments obtained from the same original recording were included into a single subset - either development or evaluation. This is a very important detail that is sometimes neglected, and failing to recognize it results in overestimating the system performance, as the classification systems are capable of learning the location-specific acoustic conditions instead of the intended general audio scene properties,. The phenomenon is similar to the "album effect" encountered in music information retrieval, that has been noticed and is usually accounted for when setting up experiments. The cross-validation setup provided with the database consists of four folds distributing the 78 segments available in the development set based on location.

3 Data description

The dataset contains a training subset (13538 recordings from 35 sensors), and validation subset (4308 recordings from 9 sensors), and a test subset (669 recordings from 48 sensors). Each recording has been annotated using a set of 23 "tags" like "engine presence, machinery presence, non-machinery-impact presence, dog-barking-whining presence" [3].

The audio files used were recorded using the SONYC acoustic sensor network for monitoring urban noise pollution. In New York City, over 60 distinct sensors have been placed, accumulating the equivalent of more than 50 years of audio data, of which we present a small fraction. The data was sampled by picking the closest neighbors based on VGGish qualities of recordings with recognized classes of interest. All of the recordings are 10 seconds long and were made with the same microphones and gain settings.

The training, validation, and test subsets of the annotation data are contained in annotations.csv. Each row in the file represents one multi-label annotation of a recording—it might be a single citizen science volunteer's annotation, a single SONYC team member's annotation, or the SONYC team's agreed-upon ground truth (for more information, see the annotator id column description).

The sensors in the test set will not disjoint from the training and validation subsets, but the test recordings are displaced in time, occurring after any of the recordings in the training and validation subset. We plan to use the test data to find out the aggregate of mismatch by using a Multi Label classification Machine Learning Model.

4 Proposal

We propose to build an application which would aim at making the urban areas quitter by leveraging the spatial and temporal data captured by the sensors. We would use Multi Label Classification Machine Learning Algorithms on the given dataset to classify the audio files into various categorical sounds. We would also dwell more into understanding the mismatches which occur between the annotated and machine predicted data and also visually analyze the cause behind it. Overall, this tool would give authorities a better overview and insights of noises around the city.

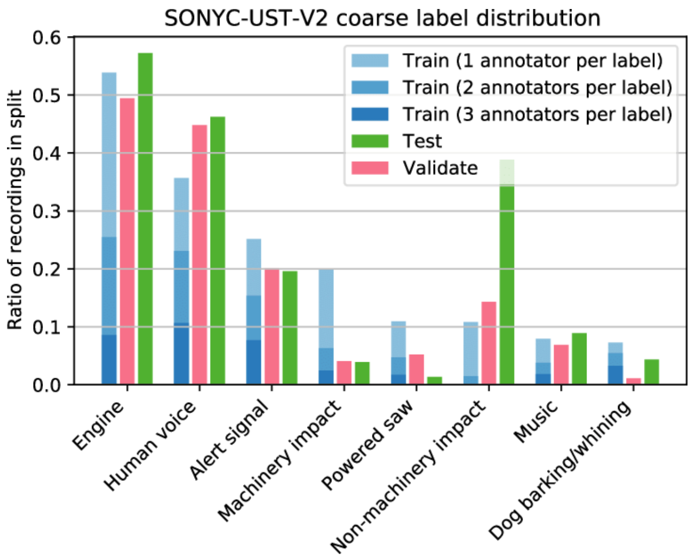

Fig. 1. Classes distibution in data

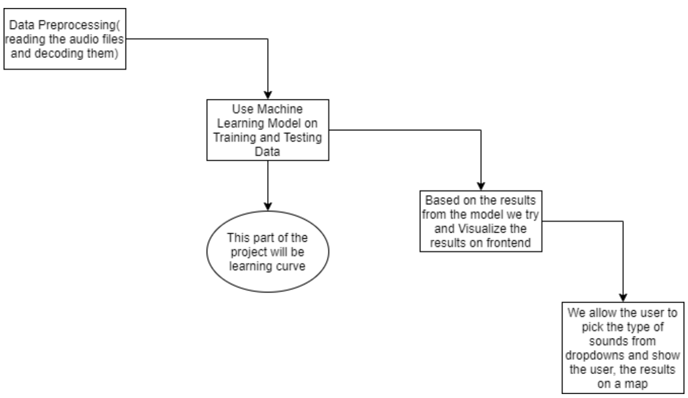

Fig. 2. Work flow diagram

5 Timeline

Week	Task
Week 1	First week is to Analyze and Preprocess the dataset
Week 2	Parsing the dataset and Extracting features from audio files
Week 3	Extracting features and Analyzing audio files
Week 4	Construct and apply the machine learning model on the given data
Week 5	Analyze and Infer results of the model
Week 6	Build a Backend system and develop API's for interaction with Frontend
Week 7	Develop API's in the Backend
Week 8	Develop Frontend component - SpatioTemporal analysis
Week 9	Develop Frontend component - Mistmatch analysis
Week 10	Develop Frontend component - Integrate User Interaction

References

[1] A. Bronzaft. Neighborhood noise and its consequences. Master's thesis, Survey Research Unit, School of Public Affairs, Baruch College, New York.

[2] T. K. S. M. S. Hammer and R. L. Neitzel. Environmental noise pollution in the united states: developing an effective public health response. Master's thesis, US, 2013.

[3] J. C. V. L. G. D. H.-H. W. J. S. O. N. Mark Cartwright, Ana Elisa Mendez Mendez1 and J. P. Bello. Sonyc urban sound tagging (sonyc-ust): A multilabel dataset from an urban acoustic sensor network. Detection and Classification of Acoustic Scenes and Events 201, October 2019.

[4] W. H. Organization. "burden of disease from environmental noise: Quantification of healthy life years lost in europe. Master's thesis, Europe, 2001.