CAPSTONE PROPOSAL

Urban Sound

Sai Likhith Kanuparthi

24th April 2020

TABLE OF CONTENTS

DOMAIN BACKGROUND

Environmental noise or noise pollution, which is defined as an unwanted or harmful outdoor sound created by human activity [1], is a growing problem in urban area. This noise pollution can affect the quality of life and health. Recent studies [2] have shown that exposure to noise pollution may increase the health risk. Therefore, decreasing the noise in the human environment can contribute to increase both the quality of life and health.

Environmental noise monitoring systems continuously measure the sound levels to quantify the noise level. However, it can be of interest to identify the type of noise source. By combining the noise level and the type of noise in real time we can describe the acoustic environment in a more complete way. Based on the outcome, actions can be taken to reduce the noise levels in urban areas. However, finding the type of noise can be a challenging task there an audio recording contains a mix of different noise sources. Here machine learning can be a helpful tool.

A way to build a real time monitoring system is the use of a low cost, small size, low power, wireless embedded device. Today, there is some attention to perform machine learning direct on these devices, so called machine learning on the edge, where we deploy a pre-trained neural network close to the sensor. However, because a limited memory footprint and compute resources the deployment of a neural network on an embedded device can be a challenge. Because these limitations trade-off needs to be made between memory footprint, compute power and the accuracy of the neural network.

PROBLEM STATEMENT

The design of a real time smart embedded monitoring system is, given the limit timeframe of this capstone project, out of the scope. The main objective of this capstone project is to take a first (small) step in the design of real time smart embedded system for environmental sound classification (ESC). In this capstone project we focus on the feature engineering step and the neural network design.

DATASET

An overview of suitable datasets for ESC can be found in [3-4]. For the project there are 3 datasets of interest: The Sounds of New York City (Sonyc), more specific the dataset used in the DCASE Challenge 2019 – Task 5 Urban Sound Tagging [5], ESC-50 [6] and the UrbandSound8k dataset [7-8]. For this project the UrbandSound8k is selected.

The UrbanSound8k is a dataset that contains 8732 labelled sound excerpts of urban sounds, in WAV format. These excerpts are less or equal to 4s. There are 10 classes defined: air conditioner, car horn, children plying, dog bark, drilling, engine idling, gun shot, jackhammer, siren and street music. The sampling rate, bit depth, and number of channels can vary from file to file. The files are pre-sorted into ten folds for cross validation and saved in folders named fold1 to fold10. Together with the dataset meta data is provided, as summarized in table 1.

Slice_file_name	The name of the audio file. The name takes the following format: [fsID][classID]-[occurrenceID]-[sliceID].wav, where: [fsID] = the Freesound ID of the recording from which this excerpt (slice) is taken [classID] = a numeric identifier of the sound class (see description of classID below for further details) [occurrenceID] = a numeric identifier to distinguish different occurrences of the sound within the original recording [sliceID] = a numeric identifier to distinguish different slices taken from the same occurrence
fsID	The Freesound ID of the recording from which this excerpt (slice) is taken
start	The start time of the slice in the original Freesound recording
end	The end time of slice in the original Freesound recording

salience	A (subjective) salience rating of the sound. 1 = foreground, 2 = background.
Fold	The fold number (1-10) to which this file has been allocated.
classID	A numeric identifier of the sound class: 0 = air_conditioner 1 = car_horn 2 = children_playing 3 = dog_bark 4 = drilling 5 = engine_idling 6 = gun_shot 7 = jackhammer 8 = siren 9 = street_music
class	The class name: air_conditioner, car_horn, children_playing, dog_bark, drilling, engine_idling, gun_shot, jackhammer, siren, street_music.

Table 1. Dataset meta data [8]

SOLUTION STATEMENT

A neural network will be built in the deep learning framework Keras [10] using the TensorFlow backend. The input features for training and testing are the Mel-Frequency Cepstral Coefficients (MFCC).

BENCHMARK MODEL

Salamon et [9] compares 5 different algorithms: decision tree (J48), k-NN (k = 5), random forest (500 trees), Support Vector Machine (SVM) (radial basis function kernel), and a baseline majority vote classier (ZeroR). The top performer is the SVM with a 67% classification accuracy for a 4 sec audio slice duration. The SVM will be used as benchmark model.

EVALUATION METRICS

To compare the result with [9] the evaluation metrics will be the average accuracy across all 10 folds.

PROJECT DESIGN

DATA ANALYSE (DA)

To get some insides in the data DA will be done on the 10 folds dataset.

FEATURE EXTRACTION

The feature extraction is performed using the PyTorch torchaudio [11] library or the libROSA [12] python package. The input features for training and testing are the MFCC.

CONVOLUTION NEURAL NETWORK (CNN) DESIGN

A CNN will be built using the deep learning framework Keras using the TensorFlow backend. Different architectures are explored and hyperparameter tuning will be performed.

CONCLUSION

Final conclusion will be drawn.

OPTIONAL

The X-Cube-AI [13] tool from STMicroelectronics will be used to get an idea about the performance of the neural network in case of deployment on an embedded device. The X-CubeAI package offers a way to validate the neural network both on a desktop PC and on an STM32 device, and a way to measure the performance on an STM32 device. The candidate target device is the B-L475E-IOT01A STM32L4 Discovery kit IoT node [14].

REFERENCES

[1] Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise, Jun. 2002.

[2] L. Poon, The Sound of Heavy Traffic Might Take a Toll on Mental Health, CityLab, [Online]. Available: https://www.citylab.com/equity/2015/11/city-noise-mental-health-traffic-study/417276/

[3] Datasets Environmental sounds, [Online]. Available: http://www.cs.tut.fi/~heittolt/datasets

[4] Detection and Classification of Acoustic Scenes and Events, [Online]. Available: http://dcase.community

[5] DCASE2019 – Task 5: Urban Sound Tagging, [Online]. Available: http://dcase.community/challenge2019/task-urban-sound-tagging

[6] ECS-50: Dataset for Environmental Sound Classification. [Online]. Available: https://github.com/karoldvl/ESC-50

[7] Urban sound, [Online]. Available: https://urbansounddataset.weebly.com

[8] Urbansound8k dataset, [Online]. Available: https://urbansounddataset.weebly.com/urbansound8k.html

[9] J. Salamon, C. Jacoby and J. P. Bello, A dataset and Taxonomy for Urban Sound Research, 22nd ACM International Conference on Multimedia, Orlando USA, Nov. 2014, [Online]. Available: http://www.justinsalamon.com/uploads/4/3/9/4/4394963/salamon_urbansound_acmmm14.pdf

[10] Keras. [Online]. Available: https://keras.io

[11] PyTorch. [Online]. Available: https://pytorch.org

[12] libROSA. [Online]. Available: https://librosa.github.io/librosa/

[13] "X-cube-ai: AI expansion pack for stm32cubemx," 2019. [Online]. Available: https:

//www.st.com/en/embedded-software/x-cube-ai.html.

[14] B-L475E-IOT01A, Discovery kit for IoT node, [Online]. Available:

https://www.st.com/en/evaluation-tools/b-l475e-iot01a.html?ecmp=tt9144_gl_link_dec2018