Urban Sound Classification

Final Project Proposal

Reham Mohamed Aburas
Purdue University
raburas@purdue.edu

Habiba Farrukh
Purdue University
hfarrukh@purdue.edu

Pavani Gowtham
Purdue University
pguttula@purdue.edu

Aly Shehata
Purdue University
ashehat@purdue.edu

1. INTRODUCTION

In today's urban environments, we experience many different kinds of sounds that we, as humans, can instantly recognize and interpret. These urban sounds can include the sound of construction work, vehicle horns, street music, gun-shots, etc.

The ability to classify these sounds using Data Mining / Machine Learning algorithms opens the door for many future work and applications. For instance, self-driving vehicles can use urban sound classification to augment the vehicles understanding of its surroundings. If a vehicle can accurately classify a vehicle horn, for instance, it can then react to that sound if needed. Additionally, it can be used in a magnitude of other applications such as security, law-enforcement, and context-aware maintaining of noise-levels in crowded areas.

In this project, we propose to apply a set of different data mining algorithms to an existing urban sound dataset: https://www.kaggle.com/pavansanagapati/urban-sound-classification, and then perform a survey of the performance across the different algorithms. We plan to make methodical comparisons among our different algorithms in terms of their assumptions, parameter tuning procedures, and performance. Finally, we hope to gain enough insight on the data and the domain of urban sounds, while also attempting to apply our best performing algorithms to newly recorded, real-world examples.

2. PLAN OF ACTIVITIES

2.1 Literature Survey

Environmental sounds classification has gained momentum over the years. Early attempts have used signal processing and machine learning techniques as wavelets filterbank [6, 3], rule-based classifiers [10], SVMs and KNN [9], etc. These techniques use traditional audio features as cepstral coefficients, their derivatives, pitch, timbre, among others. Recently, deep convolutional neural networks (CNN) were applied on audio spectrograms [7, 5, 1, 4]. These approaches claim to capture new patterns to distinguish different types of sound, in which traditional audio features fail. Deep neural networks, however, are highly dependent on the size of the dataset. Therefore, data augmentation techniques were used in [7] a to overcome the data scarcity. Other approaches proposed using transfer learning methods [2] where they pretrain a model using large unlabeled sound and video dataset to learn the low level features and then use the pretrained model for the target task.

In this project, we plan to compare some of the different learning techniques for environmental sound classification and verify the assumption of CNNs superiority on traditional techniques. We also plan to test the effect of increasing the dataset size by applying data augmentation techniques on the learning task.

2.2 Data Collection and Exploration

UrbanSounds dataset was introduced with a sound taxonomy in [8]. The dataset contains about 18.5 hours of real-world manually annotated sounds across 10 classes. To construct this data, real recordings were collected from FreeSounds (an online sound repository containing over 160K user-uploaded recordings) that correspond to urban sound classes.

UrbanSouns8K is a subset of UrbanSounds dataset organized for classification tasks. This dataset is composed of short audio snippets with maximum duration of 4 seconds. The dataset assumes a single sound source, so each segment corresponds to a single class. To maintain uniform class distribution, the audio snippets for each class are limited to 1000 snippets. The total is 8732 labeled snippets (8.75 hours). The dataset is provided in folders divided into 10 folds.

To gain more insights on the data, we would extract different sound features' from the data and summarize these features statistics. We also plan to visualize these features among different classes, so that we can make predictions about the performance of different learning techniques.

2.3 Algorithm Design

Our task is to apply classification techniques on the UrbanSound dataset. For this task, first we will do preprocessing steps on the data for cleaning and extracting sound features. The dataset is composed of short snippets of raw

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

audio data. Cleaning filters such as low pass filters could be applied to remove noise. To highlight the difference in sound frequency distributions among the different classes, the data could be represented using spectrograms. Furthermore, sound features that are commonly used in literature could be extracted such as mel-frequency cepstral coefficient (mfcc), sound pitch, etc.

Next, we plan to evaluate and compare different classification techniques, including: SVM, Logistic Regression, Deep Learning, etc.

2.4 Implementation

For implementation, we plan to use available libraries in Python for sound processing and analysis, such as Librosa. We would also use some machine learning frameworks to implement our classifiers. We plan to survey about the different libraries available, such as Scikit-Learn, MLPython, Tensorflow, among others, and select the best libraries for the purpose of our task.

3. EVALUATION STRATEGY

The goal of our project will be to classify urban sounds within our dataset using various data mining algorithms. For our project, the evaluation will be mainly based on evaluating the performance of the different data mining algorithms (mentioned in the previous sections) on the given dataset. We will use the accuracies obtained on test set as a measure for the performance and compare the performance of the different algorithms. For this purpose, we will employ cross validation to determine the average accuracies for different classifiers. In addition to the basic implementation of the different data mining algorithms, we will also be looking into different optimization techniques like parameter tuning, increasing the size of the dataset etc. We will evaluate how these techniques affect the performance of the aforementioned classifiers.

In order to determine the best model suitable for our classification task, we will use hypothesis testing to compare different models obtained from the different algorithms. As an end goal, we want to record some real world urban sounds around us and evaluate the performance of the model, which performs best on the Kaggle dataset in the cross validation step, on this newly recorded data.

4. TIMELINE

Our team plans to have Bi-weekly meetings on every Tuesday and Friday throughout the project. The proposed timeline for the project is shown in Table 1.

Explore Datasets. Finalize the topic for project. Project proposal submission.	March 11th-17th 12 hours
Obtain dataset for the project. Slides for Project pitch.	March 18th-26th 8 hours
Concretize preprocessing steps. Complete required preprocessing.	March 27th-30th 12 hours
Implement 2 algorithms.	March 31st-April 7th 10 hours per Algorithm
Implement rest of the Algorithms. Evaluate the models.	April 8th -15th 10 hours for evaluation 10 hours per Algorithm
Evaluate Models on real world data. (if time permits) Project final report.	April 16th - 23rd 20 hours

Table 1: Proposed timeline for the project

5. REFERENCES

[1] D. M. Agrawal, H. B. Sailor, M. H. Soni, and H. A. Patil. Novel teo-based gammatone features for environmental sound classification. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1809–1813. IEEE, 2017.

[2] Y. Aytar, C. Vondrick, and A. Torralba. Soundnet: Learning sound representations from unlabeled video. In *Advances in neural information processing systems*, pages 892–900, 2016.

[3] J. T. Geiger and K. Helwani. Improving event detection for audio surveillance using gabor filterbank features. *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 714–718, 2015.

[4] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, et al. Cnn architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 131–135. IEEE, 2017.

[5] K. J. Piczak. Environmental sound classification with convolutional neural networks. *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2015.

[6] J. Salamon and J. P. Bello. Feature learning with deep scattering for urban sound analysis. *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 724–728, 2015.

[7] J. Salamon and J. P. Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24:279–283, 2017.

[8] J. Salamon, C. Jacoby, and J. P. Bello. A dataset and taxonomy for urban sound research. In *ACM Multimedia*, 2014.

[9] J.-C. Wang, J.-F. Wang, K. W. He, and C.-S. Hsu. Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 1731–1735. IEEE, 2006.

[10] H. Wu and J. M. Mendel. Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers. *IEEE Transactions on Fuzzy Systems*, 15:56–72, 2007.