Pré-rapport:
Reconnaissance étages de batiment à partir des échantillons sonores

Massih-Reza Amini [a] and Junior N'nane [b]
[a] Department, University, City, Country
[b] Department, University, City, Country

Abstract

Keywords: Machine Learning; Sound; Classification format;

1 Introduction

Le but de ce projet est de pouvoir déterminer, à partir d'un enregistrement sonore, l'étage du batiment dans lequel l'enregistrement a été effectué.

Nous nous servirons de méthodes de machine Learninge et de traitement de signal.

2 Notions

Le son peut se définir comme étant une sensation auditive provoquée par des vibrations de l'air. VanDerveer proposa les critères suivants pour définir un son de l'environnement :

- un événement le produit ;
- il est le reflet d'un ou d'une série d'événements causaux ;
- son traitement est plus compliqué qu'un son pur généré en laboratoire ;
- il ne relève pas de la reconnaissance de parole (plus généralement de communication selon sa définition, la parole n'étant pas le seul type de son nous permettant de communiquer, par exemple les interjection ne font pas partie de la parole, mais sont pour autant un moyen de communication).

Les sons d'un environnement peuvent être catégorisés suivant plusieurs catégories: bruit, son naturel, son artificiel, parole, musique. Le son peut qualifé d'impulsionnel ou stationnaire, mais aussi périodique ou non-périodique.

L'enregistrement sonnore se présente comme un vecteur unidimensionnel possédant un grand nombre d'échantillons parseconde. Il est possible d'utiliser une fenêtre pour observer un signal sur une durée finie, on le multiplie par une fonction fenêtre d'observation. La fonction de Hammming permet d'améliorer les lobes secondaires.

Une fois le fénétrage effectué, il est possible d'extraire les paramètres acoustiques divisés en deux catégories: temporels et fréquentiels.

Les paramètres temporels peuvent s'obtenir avec des méthodes telles que : ZCR. Les paramètres fréquentiels eux s'obtiennent avec les méthodes telles que MFCC (Mel-Frequency Cepstral Coefficients) et SFRF (Sectral Rollof Point) et SC (Spectral Centroid).

La transformation de Fourier est une opération qui transforme une fonction intégrable sur en une autre fonction, décrivant le spectre fréquentiel de cette dernière. elle permet d'obtenir une représentation Temps-Fréquence d'un enregistrement sonore, appelée spectrogramme.

La classification d'évenements sonores se fait en général avec des méthodes de Machine Learning, on peut utiliser des méthodes supervisées et non-supervisées

3 Données

Les données mises à notre disposition sont de trois types:

- spectrogramme/sonogramme (image)
- fichier audio brut (.wav)
- enregistrement amplitude (image)

4 Methodes à explorer

Les methodes à explorer sont:

- entrainement de modèle à modalité unique

- early fusion (avec un modèle U-net pour apprendre une representation conjointe)
- late fusion (entrainer un modèle "on top" des modèles déjà existants

5 Outils/Ressources disponibles

Ici, nous allons lister les outils python disponibles pour traiter les enregistrements sonores notamment ceux intégrés dans la librairie pytorch.

MFCC [pytorch]

CNN with Pytorch using Mel features[pytorch]

How to Detect COVID-19 Cough From Mel Spectrogram Using CNN [keras]

Audio feature extraction [pytorch]

Environment Sound Event Classification With a Two-Stream Convolutional Neural Network

6 Travaux relatifs

Ici, nous presentons des travaux similaires (utilisant le spectrograme/sonograme et la representation MFCC pour de la classification).

Honk: A PyTorch Reimplementation of Convolutional Neural Networks for Keyword Spotting

Environmental sound classification using temporal-frequency attention based convolutional neural network

Rethinking CNN Models for Audio Classification

Fast environmental sound classification based on resource adaptive convolutional neural network

Environment Sound Event Classification With a Two-Stream Convolutional Neural Network

References