Machine Learning Engineer Nanodegree

Capstone Proposal

Filipe Reis
April 16th, 2017

Proposal

Domain Background

Humans have the great capability of using their senses to interact with the world. One specific interaction is through sounds, as one is capable of easily determining a situation by just listening to the ambient sounds, in other words, we perform well the task of classifying sounds. As machines become increasingly more present and important on our lives, the need of interaction with the ambient grows more relevant, making Sound classifiers important. Traditionally, sound classifiers have been widely specially for automated call center and applications alike but there is still great potential to apply these machines to many other fields, such as hearing aid systems Also, current sound classifiers are heavily based on classical preprocessing of data [1] and rudimentary machine learning methods, which is good because made even hardware-level implementations possible but has a performance tradeoff. Also, there has been incipient application of newer technologies such data intensive machine learning and deep learning but some recent work such as using deep learning to improve hearing aid [2] and deep learning techniques for large audio datasets [3].

As mentioned earlier, there is enormous potential to apply sound classifiers to hearing aid systems as its users may need more than just amplification of sounds, as our brains usually performs some filtering to what is heard and this filtering may also be compromised, generating hearing selectivity problems. This problem motivates me as I have people in my family with selective hearing problems and even I have a mild level of hearing loss on my right ear, which makes me extra excited with the potential that this application has.

Problem Statement

Sound classification is a task which traditionally relies heavily on dimensionality reduction strategies as its length is quickly converted into an

enormous number of samples due to the high sample rates necessary for acquisition without loss and aliasing. The traditional workflow for this classification is to calculate the Mel-frequency cepstral coefficients, which are coefficients obtained after transformations over the frequency response (resultant of a Fourier transform) and which approximate the human auditory systems' response [4] and use these coefficients as features on a machine learning classifier to generate labels corresponding to the sounds. Sound classification imposes a quantifiable, measurable and replicable problem as it takes a standard audio file as input as returns a label to it. As the training dataset is labeled and contains 10 different classes, if either a new or an existing sound of the dataset is presented, the system will return a label, which is the corresponding sound class of it and this label can be compared to the actual sound contained on the file. As for a typical classification problem, one can measure the accuracy of the results by comparing the predicted and original labels of the test files and even further tests could be carried out by introducing new samples from other sources.

Datasets and Inputs

The dataset considered for this project is the "Urban Sounds 8K", provided by NYU. This dataset was created from public samples available at freesound.org and is composed by 8732 wav (around 6.6GB) files divided into 10 unbalanced classes. This dataset excel others, such as AudioSet [4], by providing raw samples, which allow full experimentation on the data preprocessing stage.

As the data was collected throughout different contributions, it is not entirely uniform, varying specially regarding duration and sample rate, which impose a special challenge as the first implies on number of features not constant and the latter may change important characteristics of the model as the time interval between each sample gets different. One way of dealing with the different signal duration would be to select only the samples which have the most recurrent length, while other, and the chosen for this project, is to standardize the length during the preprocessing of the data.

The data is divided in 10 folds but will be reorganized into one group and then reordered into three sets, validation (70% of the total), train (20% of the total) and test (10%). Additionally, to provide better handling of data and allow for faster reprocessing, the three sets will be saved as pickle files as the time to load all sound files to memory is much higher than of loading three pickle blocks.

Solution Statement

The solution for the sound classification problem is a machine learning model trained using the UrbanSounds8K dataset which can be used to label new audio samples based on a python script, which will perform both the

preprocessing of data and the classification using the trained model saved on a pickle file. The specific machine learning model to be used will be chosen between Support Vector Machines, Decision Trees and Multi-Layer Perceptron, according to the best performance score obtained. The preprocessing of data will be performed using the python speech features [6] library, while the classification models will be performed using scikit-learn [7] library.

Benchmark Model

The creators of the dataset provide results for their own classifier trained on the data [5], which indicates a great comparison point for this project. The reference presents best results using Support Vector Machines using Radial Basis Function Kernel obtaining global accuracy 70%.

Evaluation Metrics

To evaluate the performance of both the benchmark and the solution model, the accuracy metric is ideal as its results are already presented on [5], making it ideal to compare to the calculation applied on the trained model.

Project Design

The project starts with adjusting the data distribution, by changing it from folds to three sets, training, test and validation and creating a pickle file for each subset, to reduce the time required to load the data for the rest of the process.

As audio data has high dimensionality, preprocessing the data is almost mandatory. For this process, calculation of both Mel-frequency cepstral coefficients, Principal Component Analysis are considered, as the first is a common strategy for audio preprocessing as it uses frequency-based analysis to generate coefficients which can simulate human's perception to audio, while the latter is a classical mathematical method for determining features with high variance, allowing fewer features to have the same significance on the result as the complete set.

To measure the improvement generated by each method, a standard SVM will be used to train the processed data, evaluate its performance gain over an unprocessed model and determine which feature selection method will be used.

Afterwards, different supervised learning methods are tested with default library parameters to allow for direct comparison and definition of a candidate for best model. The considered algorithms are Support Vector Machine Classifier, Classification Trees and multi-layer perceptron and were chosen due to their great generalization capacity.

After the best performer is defined, its parameters will be extensively varied aiming to obtain the best model, which will be saved on a pickle file with this optimal model.

References

[1] J. P. Bello, “Sound Classification: EL9173 Selected Topics in Signal Processing: Audio Content Analysis , NYU Poly,” [Online]. Available: http://www.nyu.edu/classes/bello/ACA_files/8-classification.pdf. [Acesso em 19 03 2017].

[2] D. Wang, “Deep Learning Reinvents the Hearing Aid,” IEEE Spectrum, 06 12 2016. [Online]. Available: http://spectrum.ieee.org/consumer-electronics/audiovideo/deep-learning-reinvents-the-hearing-aid. [Acesso em 19 03 2017].

[3] S. C. D. P. W. E. J. F. G. A. J. C. M. M. P. D. P. R. A. S. B. S. M. S. R. W. K. W. Shawn Hershey, “CNN Architectures for Large-Scale Audio Classification,” *International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE (2017),* 2017.

[4] Google, “AudioSet,” Google, 2017. [Online]. Available: https://research.google.com/audioset/download.html. [Acesso em 19 03 2017].

[5] C. J. Justin Salamon, “A Dataset and Taxonomy for Urban Sound Research,” [Online]. Available: https://serv.cusp.nyu.edu/projects/urbansounddataset/salamon_urbansound_acmmm14.pdf.