What is this Sound?

Maynooth University - CS401 Project

Maynooth University
National University of Ireland Maynooth

Quentin Tardivon - 17182786

Niko Karuhla -

1. Project Goals

The goal of this project was to discover the usage of Tensorflow and to work with data in the Tfrecord format. We used for that the Audioset project, gathering 10 second audio clips extracted from Youtube with labels for each one. The final goal was to try to create a classifier able to tell apart acoustic events in audio files. Our plan was to train multiple binary classifiers for different classes of acoustic events and then combine the classifiers in order to have a multi-label classifier.

2. Realized Work

1. Extracting data

The first problem was to understand the format of the data and how to

work with it. The main advantage of tfrecord format is that you can save large amount of data with smaller disk usage. Indeed, the complete audioset balanced files is around 14mb against several gigabits for the audio features.

Google used for this type of file the protocol buffer which allow to organize the data for easier parsing and exchange. The main problem for this step was that the data format was not clearly described on the project page and we had to guess the data structure to parse the data. Indeed, if you do not have information about how the data was organized, it is impossible to retrieve your data from a tfrecord file.

2. Creating new data

In order to test the model, we wanted to create new data with the same data format as the Audioset project but with our own wav files. For that we use the vggish model available on github and tweak it in order to work with our wav files. We were able to successfully extract features from raw audio files, but we didn't use them for validation. We use the validation set from the Audioset project to try to evaluate our model.

3. Learning phase

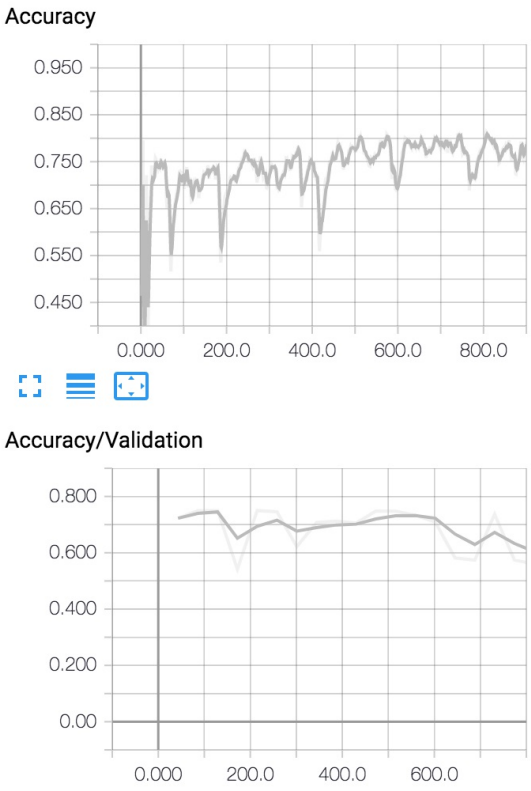
We encounter a number of problems during this phase linked to Tensorflow and its documentation. We had difficulties creating a working tensoflow program adapted to the shape of our data. We can see that the library is mainly oriented for data-scientist and not so much for software engineering student. When we succeeded to run a instance of tensorflow with sample

In the end, we used Tflearn, a high-level wrapper for Tensorflow which simplifies the creation of classifiers and computational graphs. For the training, we use a simple fully connected deep neural network for the first tests. We try different batch sizes and epochs with Adam optimizer. We used Tensorboard to vizualize learning data. We were able to observe

that the learning was non conclusive. With little training data our model overfit and was able to achieve an accuracy of 100%, but when we used many training examples our model learn to always output false.

A second run with different parameters show a level of validation increasing but the final test on the evaluation set keep its level around 46%. This result is not surprising with such a simplistic architecture and noisy data.

We can see on the following figure an improvement on our model with a validation set of 30% of data.



3. Possible Improvements

It would have been possible to work directly on the raw audio files after some minor treatment (mel spectrogram, FFT) to compare with the learning from features files. We also suspect that with only 25% of positive example and a low fidelity of our algorithm this cause the poor learning results. The idea was to equilibrate the dataset to a rough 50/50 repartition. It also may be possible to increase the number of

class to reequilibrate the learning.

4. Conclusion

With this project, we have been able to see different aspect of a machine learning problems: data acquisition and treatment, extracting data from binary formats, adapting the data to a model structure and trying to find good parameters to improve the learning. We conclude that we have seen too large for a first machine learning project and we spend more time trying to figure out problems linked to data than with the learning phase itself. However this project allowed us to see how Tensorflow works and how to read the errors messages of this library.