Qualitative analysis

Interview with mister Emanuel Oehri

Fabian Gröger
fabian.groeger@stud.hslu.ch

Saturday 30th May, 2020

Abstract

To evaluate the performance of the embedding space trained on the music dataset, mister Emanuel Oehri will be interviewed. He is a DJ and further kindly provided the music dataset, and is therefore very familiar with the categories of the dataset and their corresponding genres. This document represents the interview guide of the interview from the 27.05.2020.

1 Project and status

The goal of the project is to create an embedding space for noise detection and music dataset by adapting triplet loss to an unsupervised setting. Therefore no label is used to train the model and will be only used to evaluate the model performance.

The embedding space should be able to represent meaning, and the distances of embedded samples should represent the similarity between them.

Both of the embedding space models use the same hyperparameters, except for the segment size, neighbouring range and sample rate. The difference in hyperparameters in mainly because the music dataset has the possibility of using larger segments and ranges since each song is available as full audio.

The experiments for the DCASE dataset (noise detection) are finished, and the embedding space was examined. To compare the performance with the results from the challenge, a simple logistic classifier was trained using the embedding space as an input. The best performing embedding architecture accomplished a macro-averaged F1 score of 62.9%, which is a 20% gap between the model of the thesis and the submitted results of the challenge. However, when examining the embedding space, a few interesting properties were found. Such as that the embedding space can:

- find misclassified sound segments in the dataset, by merely looking at the neighbourhood of each sample
- find microphone malfunctions in the dataset, by merely looking at the neighbourhood of each sample
- represent the underlying sound it represents and builds clusters, for example in the embedding space a "silence cluster" was observed

- generate audio files which "walk through the embedding space" from one label to the other

The experiments using the music dataset, provided by mister Emanuel Oehri, are now also finished and the embedding space was examined. While examining the embedding space, a few exciting properties were found and will be discussed in this interview.

2 Resources

Each sample (song) in the dataset is split into segments, to perform the triplet selection. All of the sound segments are 10s long, and the neighbouring range is 40s wide. This means that each song consists out of multiple 10s segments of a song in the dataset. As a neighbourhood, in the embedding space, most of the time, the closest three are considered. The embedding space consists of the entire music dataset, which means that it contains samples from both the training, evaluation and test set, which is completely normal when evaluating the performance of a model in an unsupervised setting.

- (R1) Playlist (Folder with samples), where each song was created by appending segments around the neighbourhood of the label centres (closest five)
  → given in the provided zip file, in the folder "DJ_centers"
- (R2) Playlist (Folder with samples), where each song was created by appending sound of segments together, where the neighbourhood is not consistent
  → given in the provided zip file, in the folder "DJ_neighbours"
- (R3) Playlist (Folder with samples), where each song was created by appending sound of segments together, to represent a walk through the embedding space from label to label
  → given in the provided zip file, in the folder "DJ_walk_through_space"
- (R4) Interactive visualisation of K-Means clustering in two- and three-dimensions:
  → https://chart-studio.plotly.com/dashboard/fabiangroger:10/view
- (R5) Interactive visualisation of three principal components:
  → https://chart-studio.plotly.com/dashboard/fabiangroger:10/view
- (R6) The combination of labels within each cluster of K-Means:

  1. 'MelodicHouseAndTechno': 28,
     'Techno_Raw_Deep_Hypnotic': 54,
     'Trance': 711

  2. 'DeepHouse': 183,
     'Electronica_Downtempo': 139,
     'IndieDance': 451,
     'MelodicHouseAndTechno': 362,
     'Trance': 31

  3. 'DeepHouse': 395,
     'Electronica_Downtempo': 112,
     'IndieDance': 99,
     'MelodicHouseAndTechno': 209,
     'Techno_PeakTime_Driving_Hard': 105,

'Techno_Raw_Deep_Hypnotic': 86,
'Trance': 115

4. 'DeepHouse': 93,
'Electronica_Downtempo': 164,
'IndieDance': 165,
'Techno_PeakTime_Driving_Hard': 49,
'Techno_Raw_Deep_Hypnotic': 521,
'Trance': 62

5. 'DeepHouse': 101,
'Electronica_Downtempo': 625,
'IndieDance': 43,
'MelodicHouseAndTechno': 91,
'Techno_Raw_Deep_Hypnotic': 53,

6. 'DeepHouse': 373,
'Electronica_Downtempo': 36,
'IndieDance': 300,
'MelodicHouseAndTechno': 644,

7. 'IndieDance': 34,
'MelodicHouseAndTechno': 55,
'Techno_PeakTime_Driving_Hard': 975,
'Techno_Raw_Deep_Hypnotic': 174,
'Trance': 91

3 Interview guide

1. Did you already have the chance to look at the resources provided before the interview?

   Emanuel Oehri: *Yes, I read this document and listened to some of the audio files to understand their structure.*

2. How similar would you rate the six different categories in the music dataset, from 1 to 10, where ten is very similar?

   Emanuel Oehri: *Overall I would say it's a 6 because they are all electronic and belong to dance music. If we would say it's a 1 compared to classic music or jazz for example. The both Techno categories might be a 9, whereas Techno to MelodicHouseAndTechno is a 7. The most special one is probably Electronica_Downtempo, followed by Trance and IndieDance.*

3. Would you consider the categories as genres or as subgenres of a genre? If subgenre, what would you consider to be the genre?

   Emanuel Oehri: *Sometimes people use the term "Techno" in a very broad sense. The overall genre of all these categories I'd call "Dance Music", whereas "Techno", "DeepHouse", "Electronica" and "Trance" are the main genres. MelodicHouseAndTechno and Techno_PeakTime_Driving_Hard and Techno_Raw_Deep_Hypnotic are subgenres of Techno. Downtempo is kind of a slower version of DeepHouse. IndieDance is rather special, since it can sound similar to MelodicHouseAndTechno, Deep House and maybe Electronica.*

4. The songs in the playlist R1 consists out of the neighbours of the cluster centre of each label.

   (a) Does every cluster represent a typical song within the category?

   Emanuel Oehri: *Yes, I think they represent the categories very well.*

   (b) Did you hear something, where a song stood out for a reason?

   Emanuel Oehri: *In DeepHouse there are segments included from 2 songs. At 30s starts the segment of the song that sounds to me very typical to DeepHouse - more than the other segments.*

5. The songs in the playlist R2 consists out of the neighbours where the neighbourhood label is not consistent.

   (a) When neglecting the label of each segment, do the segments of the created song sound similar?

Emanuel Oehri: *In most cases I would say they do.*

(b) Do you consider the similarity to be reasonable?

Emanuel Oehri: *Yes.*

(c) Was there a song which stood out, in a good or in a bad way?

Emanuel Oehri: *The first segment of* `DeepHouse_Sleeping Bag (Original Mix)_five_neighbours` *seems to be the end of a song. This is probably hard to classify but it sounds reasonable in DeepHouse. Segment at 20s in* `Electronica_Downtempo_Automatic Dub (Rex The Dog Remix)_five_neighbours` *would maybe fit better to Techno_PeakTime_Driving_Hard*

(d) Are there created songs which do not sound similar and therefore represent a failure in the embedding space? Why?

Emanuel Oehri: *In* `Techno_PeakTime_Driving_Hard_Anxiety (Original Mix)_five_neighbours` *segments starting at 20s sounds very much like DeepHouse.*

6. Categories "Electronica_Downtempo" and "MelodicHouseAndTechno" are not well separated and songs often have neighbours of both categories.

   (a) How similar would you rate the categories "Electronica_Downtempo" and "MelodicHouseAndTechno", from 1 to 10, where ten is very similar?

   Emanuel Oehri: *Rather not similar, so maybe a 4.*

   (b) Can you describe the similarity or dissimilarity?

   Emanuel Oehri: *Similarity: Both categories have a lot of melodic content. Dissimilarity: MelodicHouseAndTechno is faster, has more energy and most of the time big 4/4 bassdrums/kicks. Electronica_Downtempo is more laidback and groovy and may have a beat that is not standard 4/4 (as for most dance music songs)*

7. The category "IndieDance" and "DeepHouse" are not well separated and songs often have neighbours of both categories.

   (a) How similar would you rate the categories "IndieDance" and "DeepHouse", from 1 to 10, where ten is very similar?

   Emanuel Oehri: *Rather similar, so maybe a 7.*

(b) Can you describe the similarity or dissimilarity?

Emanuel Oehri: *DeepHouse is more groovy with sharp and clean sounds, whereas IndieDance is more noisy and its elements often sound similar to the low-quality sounds of the 80s. IndieDance is perhaps also a bit more melodic.*

8. The similarity between genres, by applying K-Means clustering to the embedding space with 7 clusters.

(a) Would you consider the genre combinations (R6) as reasonable?

Emanuel Oehri: *Yes.*

(b) Where do you think the embedding space still has some flaws?

Emanuel Oehri: *It works only if the segment has enough content in it for example during the intro/outro or at a break in the track. In Dance Music there is a very common structure of the base patterns. Usually there is the same 4/4 Kick, Hi-Hat und Snare Pattern. So it is hard to find similarities between such segments.*

9. Do these statements sound plausible? Rate from a scale from 1 to 10, where 10 is very plausible.

(a) 'MelodicHouseAndTechno' - 'DeepHouse' = 'IndieDance'?

Emanuel Oehri: *8.*

(b) 'DeepHouse' - 'Electronica_Downtempo' = 'MelodicHouseAndTechno'?

Emanuel Oehri: *2.*

(c) 'Techno_PeakTime_Driving_Hard' + 'Electronica_Downtempo' = 'DeepHouse'?

Emanuel Oehri: *8.*

(d) 'DeepHouse' - 'IndieDance' = 'Electronica_Downtempo'?

Emanuel Oehri: *4.*

(e) 'Techno_Raw_Deep_Hypnotic' - 'MelodicHouseAndTechno' = 'Techno_PeakTime_Driving_Hard'?

Emanuel Oehri: *3.*

(f) 'DeepHouse' - 'Techno_PeakTime_Driving_Hard' = 'MelodicHouseAndTechno'?

Emanuel Oehri: *2.*

(g) 'Electronica_Downtempo' + 'Techno_PeakTime_Driving_Hard' = 'DeepHouse'?

Emanuel Oehri: *7.*

(h) 'MelodicHouseAndTechno' - 'Techno_PeakTime_Driving_Hard' = 'IndieDance'?

Emanuel Oehri: *9.*

(i) 'IndieDance' - 'Techno_Raw_Deep_Hypnotic' = 'DeepHouse'?

Emanuel Oehri: *3.*

(j) 'DeepHouse' - 'Trance' = 'IndieDance'?

Emanuel Oehri: *2.*

(k) 'Electronica_Downtempo' - 'Trance' = 'DeepHouse'?

Emanuel Oehri: *3.*

(l) 'Techno_Raw_Deep_Hypnotic' - 'Trance' = 'MelodicHouseAndTechno'?

Emanuel Oehri: *3.*

10. The songs in the playlist R3 consist of segments to "walk" from one label to the other.

    (a) Did you find something which did not sound very good?

    Emanuel Oehri: *The step at 1m30s in `DeepHouse_to_Techno_PeakTime_Driving_Hard` is not very smooth. These segments should not be close to each other. Same as for at 1m30s in `Electronica_Downtempo_to_Trance`. This is probably because the genres are so different to each other. At 1m00s in `IndieDance_to_Techno_PeakTime_Driving_Hard` there's a 10s segment that doesn't fit.*

(b) Did you find something which sounded very good?

Emanuel Oehri: *Some of the transitions in the DJ walk through sounded very interesting (in a good way). Especially when the categories themselves were rather similar.*

11. Overall, was there something that stood out when listening to the songs?

(a) In a bad way?

Emanuel Oehri: *If the segments of an intro (or a break) of a song are being used there is not much content to classify it in some way, i.e. segments that for example contain only a beat or only atmospheric noises.*

(b) In a good way?

Emanuel Oehri: *I liked some transitions from one genre to another in the walk through. I could imagine playing these after another also in a DJ set.*

12. How would you describe the results of the embedding space?

Emanuel Oehri: *I think the results are pretty impressive. Overall the identified similarities make completely sense. When listening to the examples it was more like finding the flaws instead of finding the good things.*

13. Would you consider the experiment to be successful or not?

Emanuel Oehri: *The experiment is a success. k-Means shows 7 clusters for the 7 categories, each representing the most segments of another genre. Already this shows that it's working to some extent. However, similar songs don't need to be assigned to the same genre (I assume the music label can select the genre of a song they want to release). in R2 and R3 the examples show that segments not belonging to the same category are close to each other in the embedding space because they in fact sound similar in most of the cases.*

14. Would you pursue this idea further?

Emanuel Oehri: *Yes. I can think of some applications. For example I would love to create good, balanced, coherent and interesting playlists by just a single click.*

15. Would you like to see something else? For example, special combination or multiple combinations?

Emanuel Oehri: *Multiple combinations between categories that fit to each other such as DeepHouse, IndieDance, and MelodicHouseAndTechno. In a DJ set it is often nice to mix up similar genres but you rarely jump from Downtempo to Techno_PeakTime.*

16. Do you have any additional comments? Or things you want to mention?

    Emanuel Oehri: *Great work! I'd really love to use it somehow for song recommendations when preparing a DJ set / a playlist. It could find already forgotten songs in the library that fit well in the set. And it would save time to go through the library one by one because the classifier would do a pre-selection for you.*