Qualitative analysis

Interview with mister Emanuel Oehri

Fabian Gröger
fabian.groeger@stud.hslu.ch

Thursday 28th 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?

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

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

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?

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

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?

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

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

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

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?

   (b) Can you describe the similarity or dissimilarity?

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?

   (b) Can you describe the similarity or dissimilarity?

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?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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?

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

11. Overall, was there something that stood out when listening to the songs?
    (a) In a bad way?

    (b) In a good way?

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

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

14. Would you pursue this idea further?

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

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