

Study doc: margin experiment

how does the margin affect the embedding space

Fabian Gröger
fabian.groeger@stud.hslu.ch

Wednesday 22nd April, 2020

Abstract

This experiment aims to show the affect of changing the margin α within a triplet loss on the DCASE dataset.

1 Introduction

A very important hyperparameter when training a triplet loss is the margin, denoted as α . The margin makes sure that the network is not allowed to output the trivial solution, where all the embeddings vectors are zero or contain the same values. Within the triplet loss function, it is used to put a limit on how far the network can push the negative sample away to improve the loss. Thus the distance of the negative sample has to be higher than the distance from the anchor to the positive sample plus the margin α . This experiment aims to show the importance of the margin as well as to find the optimal one for the DCASE dataset.

2 Hyperparameters

The hyperparameters used for this experiment are shown in table 1. The margin will be evaluated using a state of the art ResNet18 architecture on the DCASE dataset. The hyperparameters in section *Feature representation* as well as the sample rate are the default ones proposed by the organisers of the DCASE challenge within the baseline project. The margin

will be evaluated for six different values [0.3, 0.5, 0.7, 1, 2, 10].

Table 1: Hyperparameters used for the experiment

Hyperparameter	value
Dataset	DCASE
Model	ResNet18
Epochs	10
Batch size	256
Learning rate	1e-3
Margin	[0.3, 0.5, 0.7, 1, 2, 10]
Embedding dimension	512
Prefetch batches	2048
Random selection buffer	64
Shuffle dataset	True
<i>Multi threading</i>	
Number of generators	16
Number of parallel calls	16
<i>Audio sample</i>	
Sample rate	16000
Sample size	10
Sample tile size	2
Sample tile range	4
Convert to mono	True
<i>Feature representation</i>	
Feature extractor	LogMelExtractor
Frame length	480
Frame step	160
FFT size	1024
Number of Mel bins	128
Number of MFCC bins	13

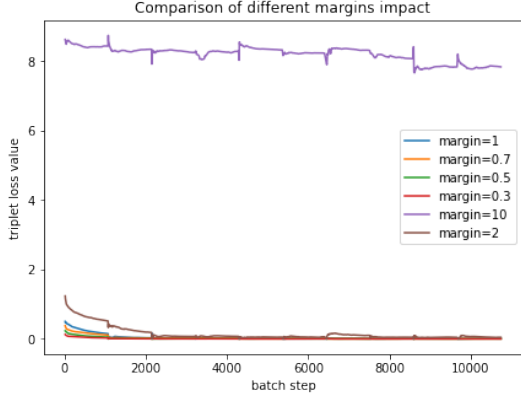


Figure 1: Plot of the triplet loss value of all margins

3 Results

Five models with the same hyperparameters, shown in table 1, were trained for ten epochs. The value of the triplet loss is the most important criteria for selecting the optimal margin since the margin has the most significant impact on the loss function. Figure 1 shows all the trained models in a single plot to visualise the impact on changing the margin. From this figure, the resulting embeddings improved from the trivial solution as the margin is increased. Thus as the margin is decreased the total number of triplets generated whose loss is higher than zero decreases, therefore, they do not contribute to the training of the model thus reducing the accuracy of the outputted embedding's.

There is a vast difference in the value of the loss values. The `margin=10` has by far the highest loss value of approximately 8, which is very intuitive since the distance to the opposite has to be higher than the distance to the neighbour plus the margin, which is in this case 8. This constraint is tough to satisfy, and therefore the loss is relatively high.

The smallest loss value is the one from the `margin=0.3`. The reason for that is the same as for the high margin, but vice-versa. The constraint it has to satisfy is that the distance to the opposite has to be higher than the distance to the neighbour plus the margin. This is very easily satisfied since it is a relatively

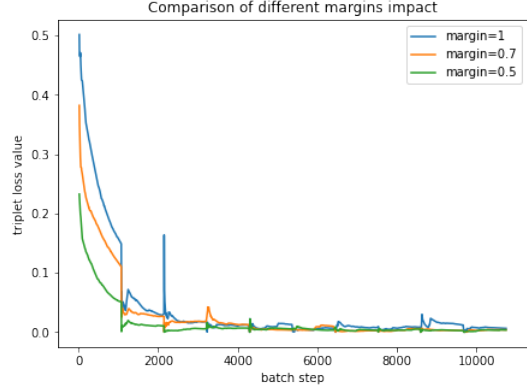


Figure 2: Plot of the triplet loss value of the margins=0.5, 0.7, 1.0

small distance.

The figure 2 shows the triplet loss of the margin values 0.5, 0.7, 1.0. It can be seen that the loss is fairly similar and approaches zero.

4 Conclusion

Since the highest and lowest margin can be omitted simply by examining the triplet loss plot because the lowest margin does not contribute to the training and the highest sets a constraint which can only be satisfied in a small number of cases, in the other cases it is a much harder decision since the loss value is fairly the same. However, when looking at the resulting embedding space, it can be seen that the `margin=1.0` distances between the centroids of each label are fairly equally distributed, which is the optimal outcome. Figure 3 shows the distance matrix of the `margin=1.0`. Whereas the other margins do not have such an equally distributed distance matrix, which indicates that one or more label can be distinguished better than others. However, the optimal solution should be that the distances from all clusters are fairly the same. Therefore the `margin=1.0` is chosen as the optimal parameter for the DCASE dataset and will be used in all of the further experiments as the standard hyperparameter.

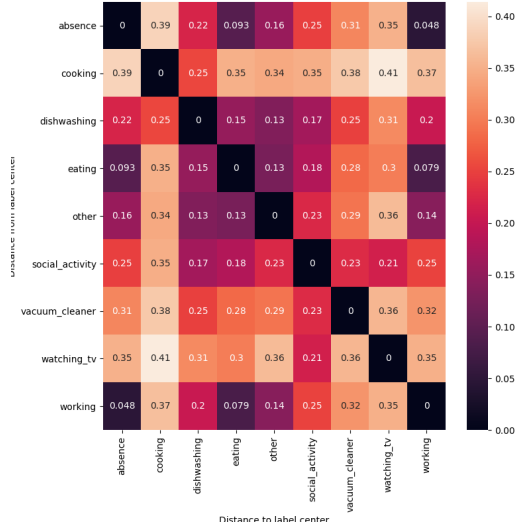


Figure 3: Distance matrix of the margin=1.0

5 Next steps

This experiment has to be also conducted for the music dataset since this parameter is heavily depended on the used dataset.