TFG informe inicial

Raquel

October 23, 2017

Contents

1 Data **1**
1.1 Initial Data 1
1.2 Generated Data 1

2 Análisis **3**
2.1 Intra sinset 3
2.2 Distribución total de las features 6
2.3 Distribución de las features por layer 7
2.4 Distribución de las features por synset 10
2.5 Features per image 12
2.6 Images per feature 14

1 Data

1.1 Initial Data

- Embedding matrix of size $(50000, 12416)$, con 62080000 features.
- **labels** Labels vector of size $50k$ which every label is in numeric format $(0, 999)$
- **synsets = synset0 synset1 synset2 ...** The set of synsets that we will analyze:

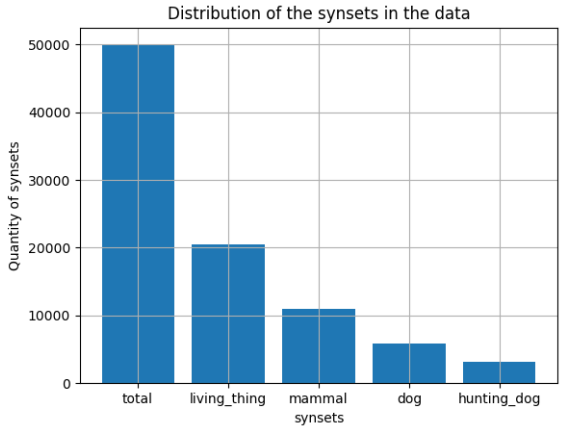
  $synsets = [living\_things, mammal, dog, hunting\_dogs]$
- **categories = {-1 0 1 }** The possible values of the features.

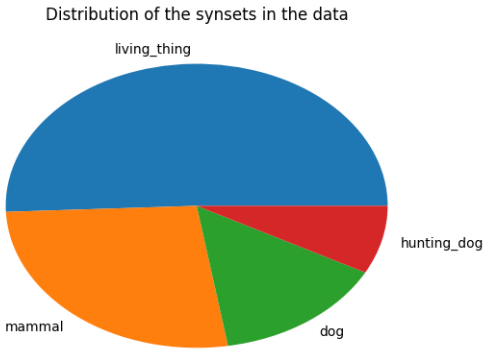
1.2 Generated Data

- **synset_index_hyponim.txt** A list with all the hyponims of every synset.
- **synset_index.txt**For each synset a list with the index of the elements of the hyponim list in the embedding.

- Un diccionario con la cantidad de imágenes que tiene cada feature para cada category.
- **features_per_image[synsets].pkl** dfd
- **features_per_layer[synsets].pkl**dsf
- **images_per_feature_per_synset[synsets].pkl**fsdfs
- **intra_synset[synsets].pkl**fdsfs

2 Análisis





2.1 Intra sinset

Dentro de living_things las imágenes se distribuyen como:

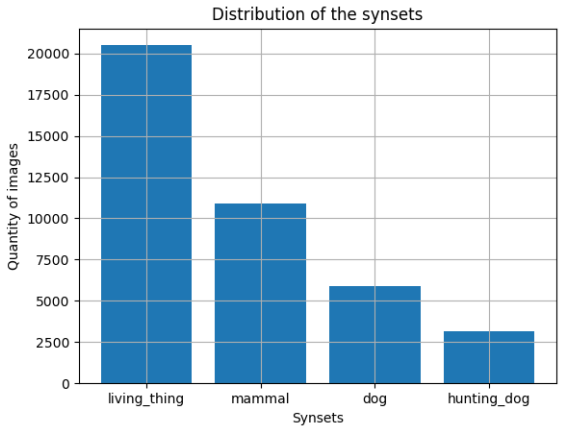
- el 53.1707317073% son mammal.
- el 28.7804878049% son dog.
- el 15.3658536585% son hunting_dog.

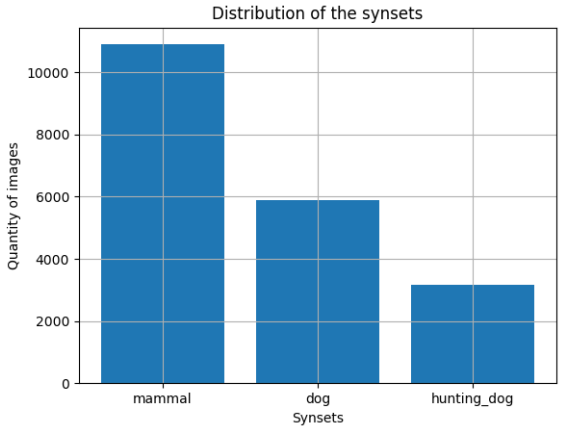
Dentro de mammal las imágenes se distribuyen como:

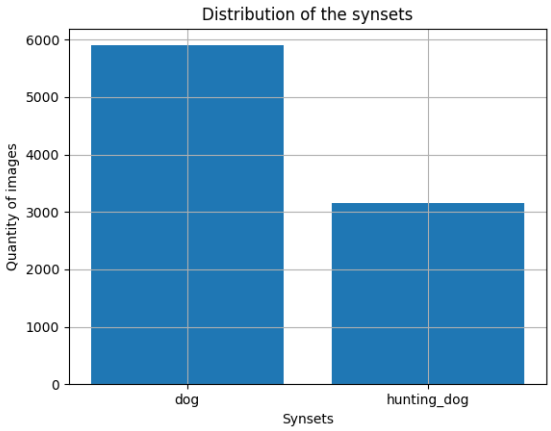
- el 54.128440367% son dog.
- el 28.8990825688% son hunting_dog.

Dentro de dog las imágenes se distribuyen como:

- el 53.3898305085% son hunting_dog.



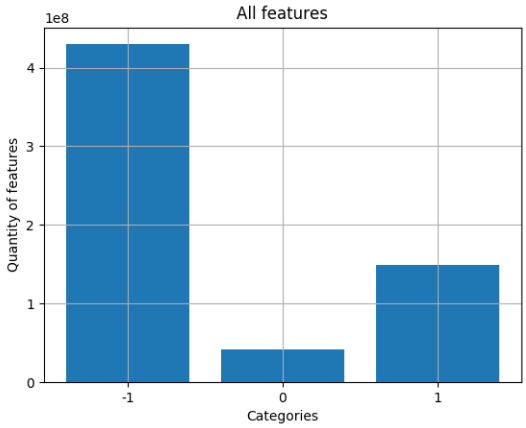



Distribution of the synsets
Quantity of images
6000
5000
4000
3000
2000
1000
0
dog
hunting_dog
Synsets

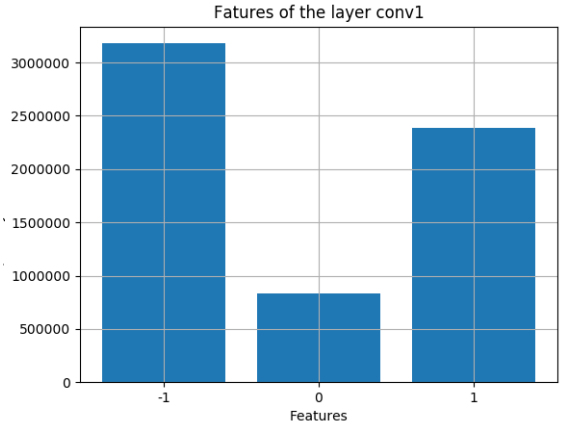
2.2 Distribución total de las features

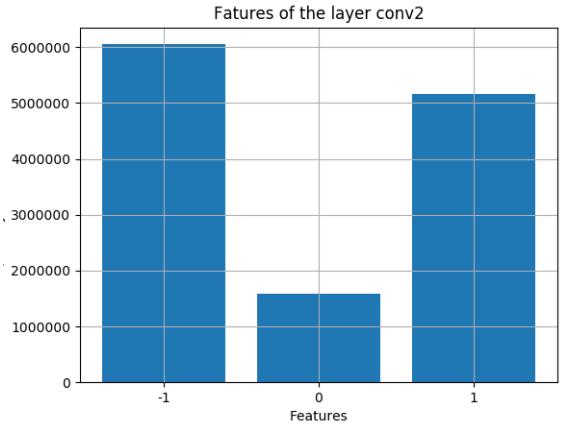
Del total de 620800000 features:

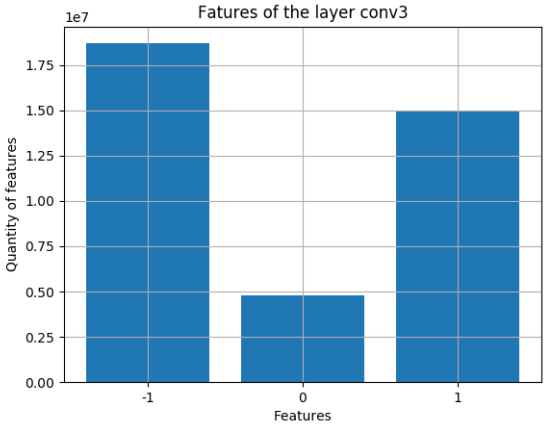
- Features de category -1: 419255437 el 67.534703125 %
- Features de category 0: 56916672 el 9.16827835052 %
- Features de category 1: 144627891 el 23.2970185245 %

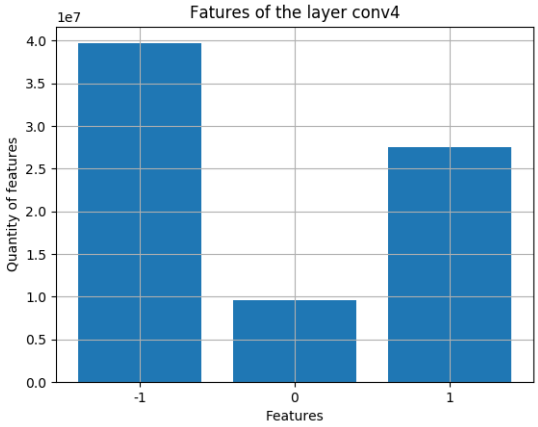


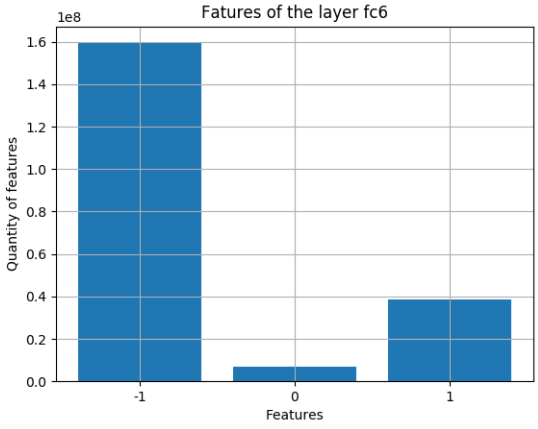
2.3 Distribución de las features por layer

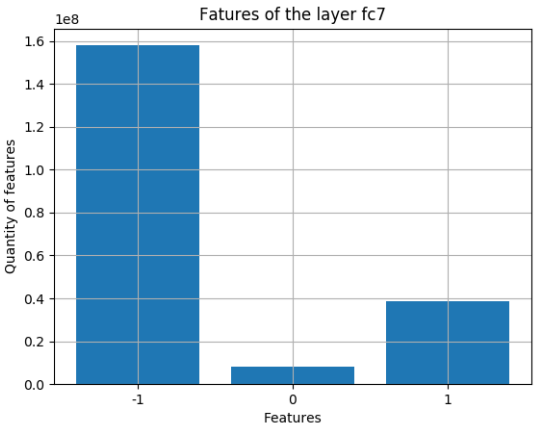




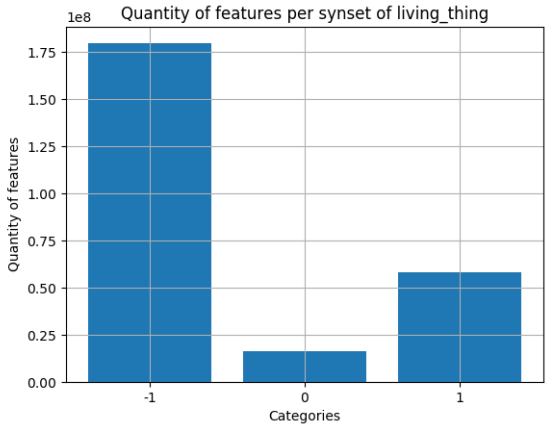

Fatures of the layer conv3
1e7
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Quantity of features
-1
0
1
Features

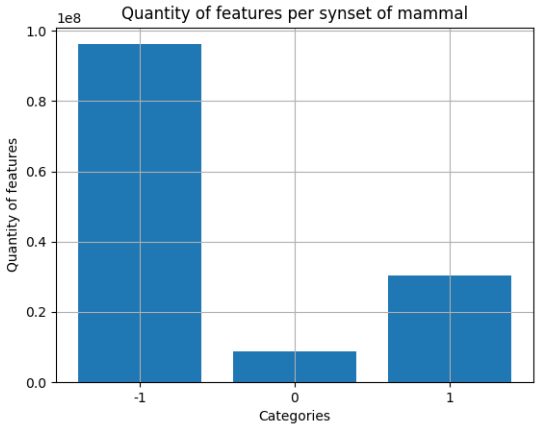

Fatures of the layer conv4
1e7
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Quantity of features
-1
0
1
Features

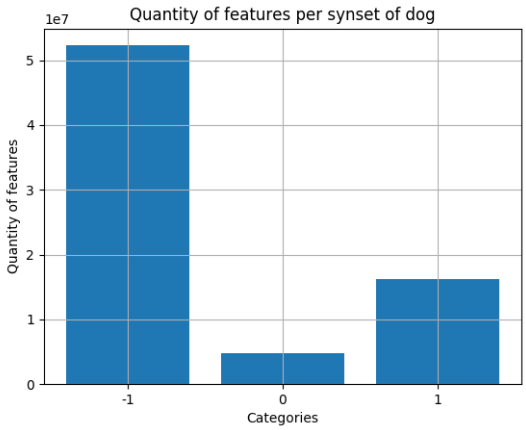

Fatures of the layer fc6
1e8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Quantity of features
-1
0
1
Features

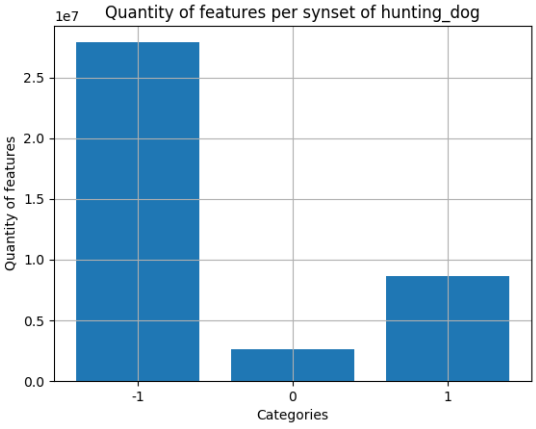

Fatures of the layer fc7
1e8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Quantity of features
-1
0
1
Features

2.4 Distribución de las features por synset



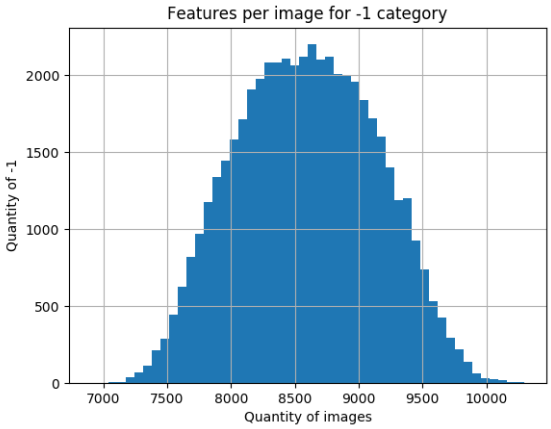


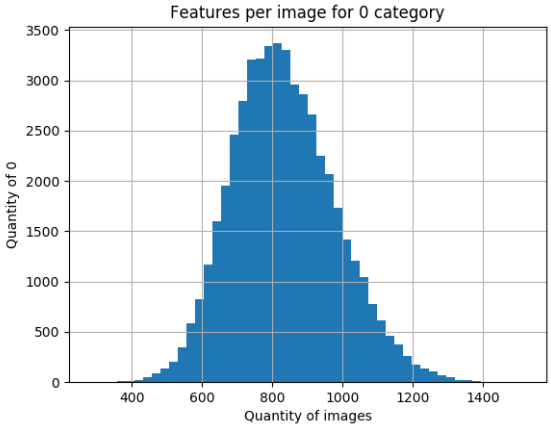

Quantity of features per synset of dog
1e7
5
4
3
2
1
0
Quantity of features
-1
0
1
Categories

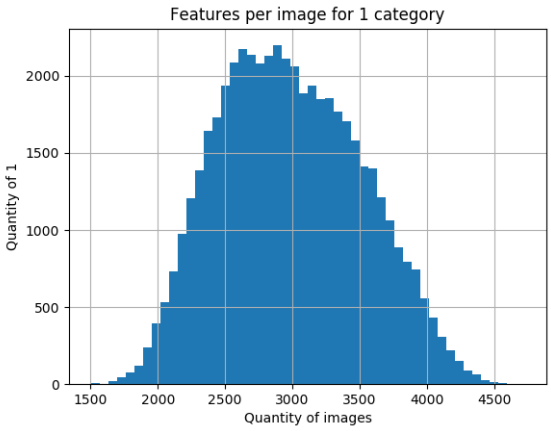

Quantity of features per synset of hunting_dog
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Quantity of features
-1
0
1
Categories

2.5 Features per image

There are for each image the total counting of features active, for the different possible categories $(-1, 0, 1)$:

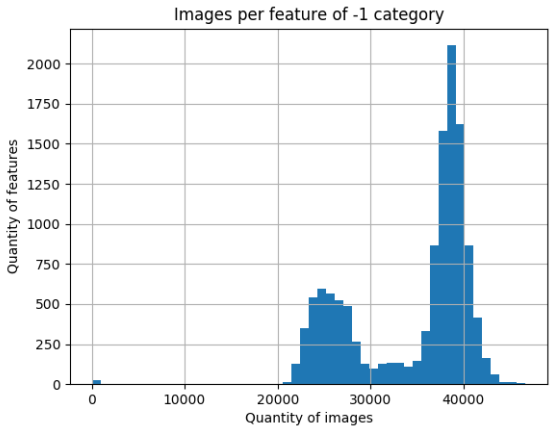


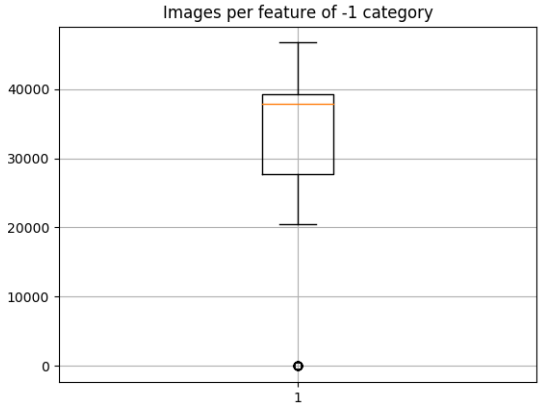


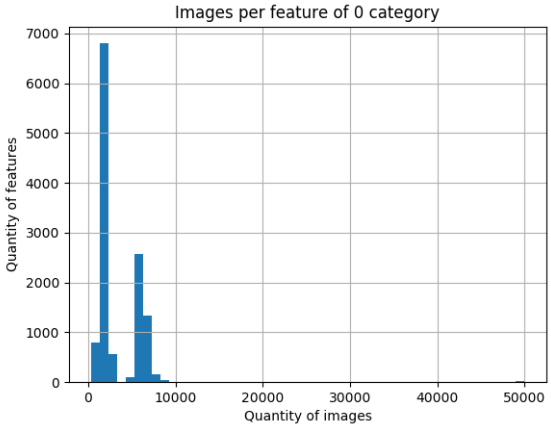

Features per image for 1 category
Quantity of 1
Quantity of images
0
500
1000
1500
2000
1500
2000
2500
3000
3500
4000
4500

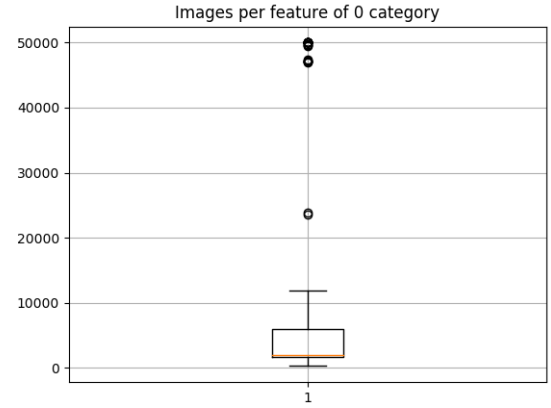
2.6 Images per feature

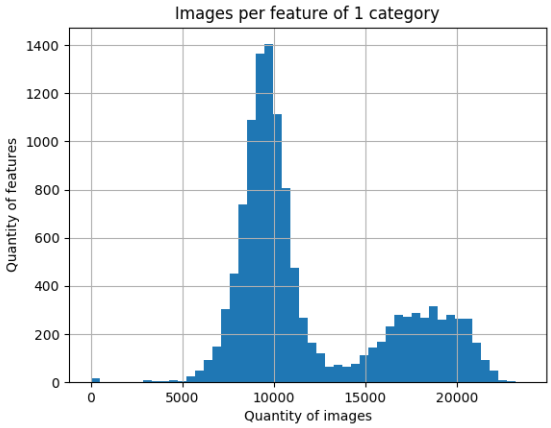
Now I calculate for each feature how many images activate in each specific category:

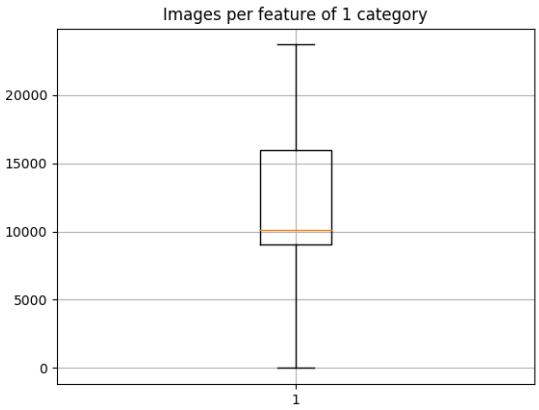





Images per feature of 0 category
Quantity of features
Quantity of images
0
1000
2000
3000
4000
5000
6000
7000
0
10000
20000
30000
40000
50000


Images per feature of 0 category
0
10000
20000
30000
40000
50000
1


Images per feature of 1 category
Quantity of features
Quantity of images
0
200
400
600
800
1000
1200
1400
0
5000
10000
15000
20000


Images per feature of 1 category
20000
15000
10000
5000
0
1