1_compare_features

February 7, 2020

1 Compare lists of feature pairs of the Balance Faces in the Wild (BFW) dataset.

Load table in `data/bfw-datatable.pkl` to extract all features and store in the datatable. Overwrites the table to `data/bfw-datatable.pkl`.

1.1 Add project code to PYTHONPATH, if not already there

Check that *path_package* is set to *code* directory on respective system

[3]:
```
%matplotlib inline
import numpy as np
import swifter
from sklearn.metrics.pairwise import cosine_similarity
# Load out custom tool for loading and processing the data
from facebias.iotools import load_bfw_datatable, save_bfw_datatable,␣
 ↪load_features_from_image_list

scorefun = np.dot # function to compare (or score) pairs of features with

dir_data = '../../data/bfw/'
dir_features = f'{dir_data}features/sphereface/'
f_datatable = f'{dir_data}meta/bfw-v0.1.5-datatable.pkl'
overwrite_pickle = False
```

1.2 Load the data

Read in the data as a pandas.DataFrame and show the first few rows.

[4]:
```
data = load_bfw_datatable(f_datatable)
data.head()
```

[4]:

	fold	p1	p2	\
0	1	asian_females/n000009/0010_01.jpg	asian_females/n000009/0043_01.jpg	
1	1	asian_females/n000009/0010_01.jpg	asian_females/n000009/0120_01.jpg	
2	1	asian_females/n000009/0010_01.jpg	asian_females/n000009/0122_02.jpg	
3	1	asian_females/n000009/0010_01.jpg	asian_females/n000009/0188_01.jpg	
4	1	asian_females/n000009/0010_01.jpg	asian_females/n000009/0205_01.jpg	

```
   label  id1  id2          att1          att2     vgg16  resnet50   senet50  \
0      1    0    0  asian_females  asian_females  0.820039  0.703258  0.679089
1      1    0    0  asian_females  asian_females  0.719199  0.523613  0.594268
2      1    0    0  asian_females  asian_females  0.732029  0.527567  0.643680
3      1    0    0  asian_females  asian_females  0.607093  0.348211  0.458883
4      1    0    0  asian_females  asian_females  0.629153  0.384273  0.494913

    a1  a2 g1 g2 e1 e2  sphereface
0   AF  AF  F  F  A  A    0.392526
1   AF  AF  F  F  A  A    0.354262
2   AF  AF  F  F  A  A    0.302028
3   AF  AF  F  F  A  A   -0.009217
4   AF  AF  F  F  A  A    0.132534
```

1.3 Load features and generate scores

First check if scores were calculated for each pairs; else, load and calculate

[5]:
```
# create ali_images list of all faces (i.e., unique set)
li_images = list(np.unique(data.p1.to_list() + data.p2.to_list()))

# read features as a dictionary, with keys set as the filepath of the image␣
 ↪with values set as the face encodings
features = load_features_from_image_list(li_images, dir_features,␣
 ↪ext_feat='npy')
```

[6]:
```
# score all feature pairs, because L2 norm applied on features dot is same as␣
 ↪cosine sim
data['sphereface'] = data.swifter.apply(lambda x: scorefun(features[x.p1],␣
 ↪features[x.p2].T), axis=1)
```

```
/Users/jrobby/miniconda3/envs/fairness/lib/python3.7/site-
packages/tqdm/std.py:658: FutureWarning: The Panel class is removed from pandas.
Accessing it from the top-level namespace will also be removed in the next
version
  from pandas import Panel
```

```
HBox(children=(FloatProgress(value=0.0, description='Pandas Apply', max=923898.0, style=Progres
```

[7]:
```
data.head()
```

[7]:
```
   fold                                 p1                                 p2  \
0     1  asian_females/n000009/0010_01.jpg  asian_females/n000009/0043_01.jpg
1     1  asian_females/n000009/0010_01.jpg  asian_females/n000009/0120_01.jpg
```

```
2    1  asian_females/n000009/0010_01.jpg  asian_females/n000009/0122_02.jpg
3    1  asian_females/n000009/0010_01.jpg  asian_females/n000009/0188_01.jpg
4    1  asian_females/n000009/0010_01.jpg  asian_females/n000009/0205_01.jpg

  label  id1  id2           att1           att2     vgg16  resnet50   senet50  \
0     1    0    0  asian_females  asian_females  0.820039  0.703258  0.679089
1     1    0    0  asian_females  asian_females  0.719199  0.523613  0.594268
2     1    0    0  asian_females  asian_females  0.732029  0.527567  0.643680
3     1    0    0  asian_females  asian_females  0.607093  0.348211  0.458883
4     1    0    0  asian_females  asian_females  0.629153  0.384273  0.494913

   a1  a2 g1 g2 e1 e2        sphereface
0  AF  AF  F  F  A  A    [[0.39252594]]
1  AF  AF  F  F  A  A    [[0.35426214]]
2  AF  AF  F  F  A  A    [[0.30202782]]
3  AF  AF  F  F  A  A  [[-0.009217382]]
4  AF  AF  F  F  A  A    [[0.13253382]]
```

[14]:
```
if not pathlib.Path(f_datatable) or overwrite_pickle:
    save_bfw_datatable(data, fpath=f_datatable)
else:
    print('Scores were in datatable. Will not overwrite by default')
```