

TCSS555 Project: User Profile in Social Media

Team 4: Iris Xia, Navyasree Petluri, Manish KC

This program built a system using python3 on Linux OS, for automatic recognition of the age, gender and personality trait scores of Facebook users given input data, including their status updates as text, profile pictures as image and pages liked as relation. Various of traditional machine learning and deep learning algorithms has been applied to predicted age, gender and personality trait scores from relation sources, age and gender from image sources and gender from text sources independently. Then the ensemble results came from the models with the best prediction performance from each source.

Getting started

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes.

Prerequisites

Operating system

```
Ubuntu Workstation 16.04 LTS
```

Python Version

```
Python3
```

Software and packages need to be installed

```
apt-get install python3
```

```
apt-get install python3-pandas
```

```
apt-get install python3-numpy
```

```
apt-get install python3-sklearn
```

```
apt-get install python3-keras
```

```
pip3 install tensorflow
```

Training and Testing Data Folder structure

```
/data/  
  training/  
    image/  
      train/  
        female/  
        male/  
        XX-24/  
        25-34/  
        35-49/  
        50-XX/  
      validation/  
        female/  
        male/  
        XX-24/  
        25-34/  
        35-49/  
        50-XX/  
    LIWC/  
    profile/  
    relation/  
    text/  
  public-test-data/  
    image/  
    LIWC/  
    profile/  
    relation/  
    text/
```

Program architecture

```
./  
tcss555.py  
tcss555_ensemble.py  
RelationPredictor_Final.py  
relation_Iris.py  
imageCNN1.py  
imageCNN2.py  
imageCNN3.py  
prediction_manish.py  
page_score.csv  
gender78.hdf5  
weights-improvement-gender-94-0.81.hdf5  
crop_model_best.hdf5
```

Installing

```
$ chmod +x tcss555
```

Running

```
$ tcss555 -i [input file path] - o [output file path]
```

eg. \$ tcss555 -i /data/public-test-data/ - o ~/output/

Result

xml files will be generated in the specified output file path with “.xml” extensions and user’s id value as the base file name.

e.g. Each file will be similar to the following outcome with newly predicted age_group, gender and personality traits for the id provided:

```
<user  
id="8157f43c71fbf53f4580fd3fc808bd29"  
age_group="xx-24"  
gender="female"  
extrovert="2.7"  
neurotic="4.55"  
agreeable="3"  
conscientious="1.9"  
open="2.1"  
>
```