Algorithmic Biases in Facial Recognition Systems

Problem statement

Many facial recognition software today report much higher error rates detecting African female faces than Caucasian male faces. This happened because African faces are not well-represented in the training data set. When the subgroups inside the train set and the test set have different distributions, the disparity between the majority subset and the minority subset could lead to higher error rates for the latter.

This leads us to the question of algorithm fairness. What makes an algorithm fair? In statistics, a hypothesis achieves statistical parity if it treats the general population statistically similarly to the protected class. An algorithm is fair if the discrepancies among k different sub-populations are close to zero.

Project goal:
The goal of this project is to minimize classification accuracy discrepancies among different races when detecting gender using current facial recognition technologies. You will be working on constructing different train and test sets by sub-setting the original data set into different proportion of races. You will experiment with different techniques to work with imbalanced train and test sets and find the best approach that can minimize the differences in classification accuracy among different subgroups.

Techniques you could be working with include re-sampling, ensemble methods, boosting, or data augmentation techniques in neural networks. Things you could do:

- Preliminary analysis: Exploratory Data Analysis, Principle Component Analysis, Clustering, Support Vector Machines, etc.
- Testing different distributions when creating train/test data sets: For example, you could create a train set that is composed of mainly Caucasian people and use it to detect gender on a test set that has 30% Caucasian, 30% Asian and 40% African people. Or you could create a train set that has balanced racial profiles and use it to predict on a test set that contains mostly Asian people. Compare the differences in classification accuracy for these different models.
- Transfer learning: You will be applying transfer learning techniques on a pre-trained Convolutional neural network. (For example, pre-trained model from vggface2) You will remove the last fully-connected layer then treat the rest of CNN as a fixed feature extractor for the new data set. You will also experiment with fine-tuning techniques on the CNN with the provided data sets.

Data Recources

1. **VGGFace2**
   You will be working with a large-scale face dataset from University of Oxford visual geometry group (VGGFace2). The dataset contains 3.31 million images of 9131 subjects (identities), with an average of 362.6 images for each subject. Images are downloaded from Google Image Search and have large variations in pose, age, illumination, ethnicity and profession (e.g. actors, athletes, politicians). The whole dataset is split to a training set (including 8631 identities) and a test set (including 500 identities). The identities in the training set are disjoint with the ones in benchmark datasets IJB-A and IJB-B.
   http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/

2. **Faces in the wild**
   Faces in the Wild, consists of 30,281 faces collected from News Photographs. Included in the file faceData.tar.gz are a matlab file, FacesInTheWild.mat, and the face images stored by year/month/day/imgname.ppm. FacesInTheWild.mat contains two variables metaData (metaDatai gives the file name of face i and it’s label id), and lexicon (lexiconi gives the actual name of label i).
   http://tamaraberg.com/faceDataset/index.html/

3. **color FERET Database**
   The DOD Counterdrug Technology Program sponsored the Facial Recognition Technology (FERET) program and development of the FERET database. The National Institute of Standards and Technology (NIST) is serving as Technical Agent for distribution of the FERET database. The goal of the FERET program is to develop new techniques, technology, and algorithms for the automatic recognition of human faces. As part of the FERET program, a database of facial imagery was collected between December 1993 and August 1996. The database is used to develop, test, and evaluate face recognition algorithms..
   https://www.nist.gov/itl/iad/image-group/color-feret-database/

High-level project goals

1. Experiment with different techniques working with imbalanced train / test data sets.

2. Build CNNs for gender detection using different train / test sets. Compare the classification accuracy among sub-groups.

3. Describe the strengths and limitations of each CNN. Discuss how racial bias could be minimized using different techniques to detect gender.

References

1. VGGFace2: A dataset for recognising faces across pose and age. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman To appear in FG 2018.

2. Gil Levi and Tal Hassner, Age and Gender Classification using Convolutional Neural Networks, IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, June 2015

3. Jeremy Kun, ’Https://Jeremykun.Com/2015/10/19/One-Definition-of-Algorithmic-Fairness-Statistical-Parity/.’

4. Tamara L. Berg, Alexander C. Berg, Jaety Edwards, David A. Forsyth Neural Information Processing Systems (NIPS), 2004