WiMLDS

Women in Machine Learning & Data Science

Welcome to the 2nd Paris WiMLDS paper reading session!

agorize

We are happy to

Meet

You again!

The mission of WiMLDS is to **support and promote women and gender minorities** who are **practicing, studying,** or are **interested** in **machine learning** and **data science**

Let’s start!

@WiMLDS_Paris & #WiMLDSParis #WiMLDS
@Agorize

Code of Conduct

WiMLDS is dedicated to **providing a harassment-free experience for everyone**. We do not tolerate harassment of participants in any form. All communication should be appropriate for a professional audience including people of many different backgrounds. Sexual language and imagery is not appropriate.

Be kind to others. Do not insult or put down others. Behave professionally. Remember that harassment and sexist, racist, or exclusionary jokes are not appropriate.

Thank you for helping make this a welcoming, friendly community for all. https://github.com/WiMLDS/starter-kit/wiki/Code-of-conduct

Agorize

Agorize the leading open innovation challenge platform

WiFi Network :

_AgorizeGuest

Password: challengeACCEPTED!!!

agorize

Gender shades

Intersectional Accuracy Disparities in Commercial Gender Classification

Paper contributions

- New dataset composed of 1270 individuals, balanced
- First Intersectional demographic and phenotypic evaluation of face-based gender classification accuracy

Plan

1. Datasets
2. Classification
3. Applications

Datasets

Why is the dataset important

Example:

Accuracies of face recognition systems used by us law enforcement are systematically lower for people labeled female, black and 18-30

Existing Datasets

- IJB-A and Adience
- Disproportion of representation for gender and phenotypes
- Over representation of lighter males
- Under representation of darker individuals

IJB-A: most geographically diverse set of collected faces

PPB Dataset

Pilot Parliaments Benchmark

- Dataset balanced by gender and skin type
- From parliament pictures
- Countries with majority population at opposite ends of the skin type scale

Challenges

- Subjects’ phenotypic features can vary widely within a racial or ethnic category
- Racial and ethnic categories are not consistent across geographies
  - Racial and ethnic labels unstable => use skin type
- Fitzpatrick classification is skewed towards lighter skin
- Gender classifiers provided by companies : gender identity or biological sex?
  - PPB labeled as perceived as woman or man

Dataset Labeling

Skin type labels

- Labeled by the Fitzpatrick six point skin type scale
- Board-certified surgical dermatologist provided the definitive labels

Gender labels

- Based on name, gendered title, prefixes (Mr, Mrs...) and appearance on photo

Classification

Algorithms used

- 3 commercial gender classifiers: Microsoft, IBM, Face++
- Face recognition systems tend to perform better on their respective populations
- Microsoft : “advanced statistical algorithms”
- IBM and face++: “deep learning based algorithms”

Test methodology

Datasets are only used as testing benchmark.

Since proprietary algorithms, can't change training data.

Test evaluation

- Assess overall classification accuracy, male classification accuracy, female classification accuracy, etc
- Results detailed in more specific groups

Results

- Better performance on male faces than female faces
  8.1% − 20.6% difference in error rate
- Better performance on lighter faces than darker faces
  11.8% − 19.2% difference in error rate
- Worst performance on darker female faces
  20.8% − 34.7% error rate
- Best performance
  - Microsoft: lighter male faces (0.0% error rate)
  - IBM: lighter male faces (0.3% error rate)
  - Face++: darker male faces (0.7% error rate)
- The maximum difference in error rate between the best and worst classified groups is 34.4%

Applications

Examples of applications

- Helping determine who is hired, fired, granted a loan
- How long individual spends in prison
- Identify suspects
- Identify emotions from images of people's faces
- Understand and help people with autism
- Surveillance and crime prevention

Questions

Discussions