Bias in ML

Mijanur Palash

Bias in ML

- Bias in machine learning is an important challenge
- Machine learning algorithms can discriminate based on classes like race and gender

Researchers in MIT and Stanford[1] showed:

- Three commercially released facial-analysis programs from major technology companies demonstrate both skin-type and gender biases
  - Error rates in determining the gender of light-skinned men were never worse than 0.8 percent
  - For darker-skinned women, more than 20 percent in one case and more than 34 percent in the other two.

1. Joy Buolamwini and Timnit Gebru.Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency, pages 77–91. PMLR, 2018

Bias in ML

- A good model is dependent on a good dataset
- Without proper care a dataset can lack diversity
- Biased dataset will perform poorly with minority:
  - If most of the samples are white males, the model will fail for women and people of color

How Bias is Introduced

- Keyword searching in google is a popular method of collecting visual (image and video) data
- In our search with keyword “angry face”- 85% of the acceptable images appeared are male
- This pattern holds for other generic keywords like ”sad people”, ”happy human” etc.
- Therefore, a dataset prepared by collecting results from these types of keyword search results in bias
- Same applies to the volunteer choice for creating an acted dataset
  - Without careful selection of people from multiple genders and ethnic backgrounds, dataset bias can be easily incorporated into the model

Bias in Dataset

- One of the widely used facial emotion recognition dataset FER-2013 is suffering from keyword search bias

TABLE VII: Gender bias- number of images with male subjects per 100 images returned from gender neutral-keyword searches on Google and number of images with male subjects per 100 images on FER-2013 dataset.

Keyword	# Male in Google(%)	# Male in FER-2013
”Angry people”	84.7	70
”Fear face:	60.1	52
”Happy human face”	55.8	58
”Sad human face”	40.0	45

How to Reduce Dataset Bias:

- Better representation of minority groups by using specific keywords:
  - Using both “happy man face” and “happy woman face” instead of “happy face” keyword
- Choosing volunteers from diverse background
- To come up with new ML models which provide higher importance on less represented data samples