An Investigation of Bias Through the Lens of Social Data

Presented by:
Kevin Cherry

What I hope to answer through this presentation…

1. What is fairness/bias look like in ML?
2. How does bias show up in data
3. What types of bias are there?
4. How does bias get into data through the ML pipeline?

We will be looking at all of these questions through the lens of Social Data

Why should you care?

If we ignore fairness, it can lead to...

- Limited access to resources for certain groups
- Misrepresentation of a group or individual
- Degraded user experience

What is Social Data?

Social Data is a general term for any digital traces left on the internet by users - a user’s internet “footprint”

- Social Media
- QA/Thread Platforms
- Collaborative Knowledge Platforms

How is Social Data Used?

There are 2 general use cases for the collection and analysis of Social Data…

1. Improving an app / user experience

2. Exploring social phenomena

What makes research valid?

Construct Validity - Do the measurements over our data actually measure what they think they do?

Internal Validity - Does our analysis of the measurements correctly lead to the conclusions?

External Validity - Can the findings be generalized to other situations?

Bias poses a threat to these validities

Bias in ML

When we talk about Bias in ML, we typically mean **Data Bias**

(Definition) Data Bias - A systematic distortion in sampled data that compromises its representativeness

Bias: Population Bias

Definition

Systematic distortions in demographics or other user characteristics between a population of users represented in a dataset or on a platform and some target population

Examples in Social Data

Different user demographics use different platforms

- Twitter is used most by men in urban areas
- Pinterest is used most by women

Different user demographics interface with the platform differently

- On Twitter, Germans are more likely to use hashtags
- Koreans are more likely to leave comments

Bias: Behaviour Bias

Definition

Systematic distortions in user behaviour across platforms or contexts, or across users represented in different datasets

Examples in Social Data

How users interact is platform dependent

- LinkedIn vs. Tinder

The type of content people like to interact with depends on the platform

Data on sites is created by user volition

- Misrepresenting the group that chooses <u>not</u> to interact

Quick Discussion

Moving away from social data…

Imagine you are trying to create a face detection system, and for your dataset you use only graduation photos.

Is there any bias would your data have?

Quick Discussion

Population Bias:

- Age, Race, Socioeconomic

Behavioural Bias:

- Smiling, Good lighting

Bias: Linking Bias

Definition

Behavioural bias that are expressed as differences in the attributes of a user's network

Ex.

Explicit Link - Users X and Y are "friends"

Other Links - User X commented on Y's post

Examples in Social Data

Behaviour based and connection based social links give very different results

- Users can follow a lot of people, but only interact with a few

Online links often depend on offline factors

- Users who are closer together geographically are more likely to interact with each other

Bias: Temporal Bias

Definition

Systematic distortions across user population or behaviour over time

Examples in Social Data

Populations and behaviours change

- Social media usage behaviour is much different than 5 years ago

Seasonal/periodic events

- Holidays and weather

Sudden unpredictable events

- Natural disasters

Failure to long term impact of an event

- A sports game vs. a natural disaster

Bias: Functional Bias

Definition

Biases that come from platform specific mechanisms (there are a limited number of actions a user can take)

Examples in Social Data

UI shapes user behaviour

- Hidden UI elements
- Youtube removing dislike count

Recommend algorithms shape user behaviour

- Users are more likely to interact with content presented to them

Bias can show up many ways during data collection

- **Who** you choose to collect data from
  (Population bias)
- **Where** you choose to collect data from
  (Behavioural bias, Functional bias)
- **What** type of data you collect
  (Linking bias)
- **When** your data is collected
  (Temporal bias)

Bias in the Data Pipeline: Data Collection

Acquisition - bias from the source

- APIs for social sites limit what data they give you
- Vague terminology in documentation “most relevant” to describe the returned data

Querying - bias in data requests

- APIs have limited query filters
- Query parameters can introduce bias
- Ex. Using the location of a post to determine where a user is from

Bias in the Data Pipeline: Data Processing

Cleaning - bias in data preparation

- Data representation

Ex. Removing images from data might leave posts with no text that can impact results

- Data normalization

Ex. Removing emojis can remove some emotional context from certain data (like sarcasm)

Bias in the Data Pipeline: Data Processing

Enrichment - bias in data labelling

- Manual labelling

Introduces subjectivity and noisy labels, as quality of the labelling varies

- Automatic labelling

Introduces systematic error, as NLP methods are not 100% accurate

Bias in the Data Pipeline: Data Processing

Aggregation - bias in collecting data into groups

- Can overemphasize or underemphasize certain users

Ex.By grouping topic interest by the number of posts, we overemphasise highly active users who make a lot of posts about that topic

- Results can be skewed by counting duplicate entries

Ex. Retweets vs. unique tweets

Bias in the Data Pipeline: Data Analysis (Model)

Inference and Prediction Models

Performance can depend on variation within dataset

- Preprocessing

Test and training set composition can introduce bias

- Sets consists of only males will be bias towards males

Class labels

Bias in the Data Pipeline: Evaluation

Evaluation Metrics

- Choice of metrics

By ignoring some metrics, we ignore certain failings of our model

- Metrics are aggregates

They are sensitive in the way the data is aggregated

- Metrics are not usually domain specific

Bias in the Data Pipeline: Remedies

Fairness metrics

- Group fairness, Predictive parity, etc.

Disaggregated Tests

- Testing for model performance on specific groups

Stakeholder conversations

- Personas, discussions, data analysis

Key Takeaways

- To achieve fairness we have to eliminate bias
- Bias can appear in many different forms in datasets (Population, Behavioural…)
- Bias should be considered in every stage of the data pipeline

Reference Paper

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019, July 11). Social Data: Biases, methodological pitfalls, and ethical boundaries. Frontiers. Retrieved November 25, 2021, from https://www.frontiersin.org/articles/10.3389/fdata.2019.00013/full#B21