College of Computing & Digital Media, DePaul University Chicago IL

Synthetic Attribute Data for Evaluating Consumer-side Fairness

Robin Burke
Jackson Kontny
Nasim Sonboli

Synthetic Attribute Data for Evaluating Consumer-side Fairness

College of Computing & Digital Media,

Department of Information Science, University of Colorado Boulder

Robin Burke
Jackson Kontny
Nasim Sonboli

Announcement

- Our CU Boulder research group

TRSL THAT RECOMMENDER SYSTEMS LAB

- that-recsys-lab.net
- We are interested in collaborations
  - especially in the area of fairness-aware and multistakeholder recommendation

Fairness-aware Recommendation

- Especially relative to users
- Requires demographic information
  - Which users are in the protected group?

The Problem

- The areas where fairness is important (employment, housing, etc.)
  - Are precisely those where user identity needs to be protected
  - Demographic attributes would be enable de-anonymization
- Public / standard recommendation datasets (with some exceptions) lack such features

One Solution

- Use data mining techniques to recover demographic attributes from the data
- But
  - That amounts to an attack against the anonymization of a particular data set
    - Probably a bad idea
  - Might violate terms of service (ex. XING challenge data)

Our solution

- Generate a synthetic attribute
    - Probabilistic labels for protected / unprotected group
    - Associated with some aspect of behavior
- Use as input to evaluate fairness-aware recommendation algorithms

FLAG algorithm

- Frequency-Linked Attribute Generation
- Assumption
    - Frequency of interaction is linked to protected / unprotected status
    - Support from studies in job seeking and other applications

Synthetic attribute (A/B)

- Group labels are drawn from a probability distribution
    - The membership probabilities are non-zero for both groups A and B
    - Supports non-deanonymization
- Feature should be correlated with user behavioral differences
    - In many datasets only behavior is known
- Data generator can be parameterized to account for
    - Different group sizes
    - Dissimilarity of groups in terms of behavior

XING dataset

- XING Challenge dataset
  - Career-oriented job networking site
  - Consisting of 10,000,000 interactions between users and job postings
  - Most attributes of users and jobs are anonymized
- Our sample
  - Region 7 only, Career Level 0
  - 3,000,000 interactions
  - 410k users with profile sizes between 1 and 30 interactions
  - Interactions follow a power-law distribution with an exponent of 1.45.

FLAG Algorithm

Basic idea

- Assume power law distribution of behavior
- Use a parameterized power law to assign B (protected group) labels with probability

$$f_B(i) = 1/i^{\alpha}$$

- Scale to achieve a given A/B expected propor

Expected value of sum of f_B

$$\mathrm{FLAG}_B(j) = \frac{\beta|U|}{j^{\alpha} \sum_{i=1}^{k} S(i)/i^{\alpha}}$$

Limitations

- Not every combination of α and β is feasible

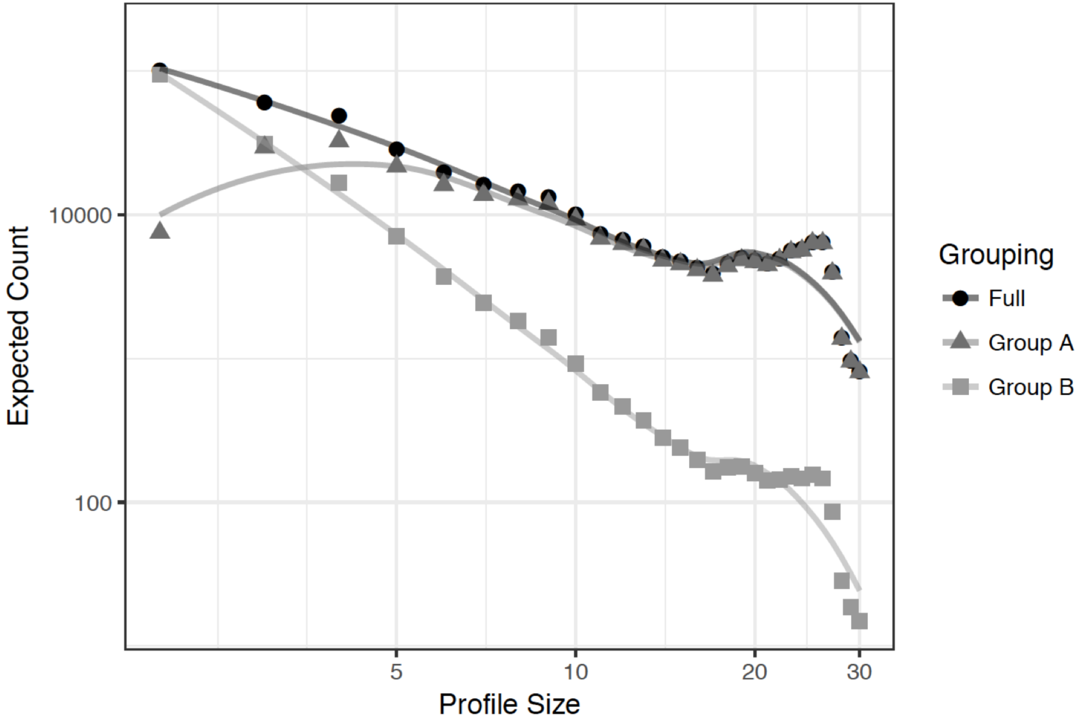
$$0 < \beta \leq \frac{E_f(|B|)}{|U| * f_B(1)} = \frac{E_f(|B|)}{|U|}$$

Results

XING dataset

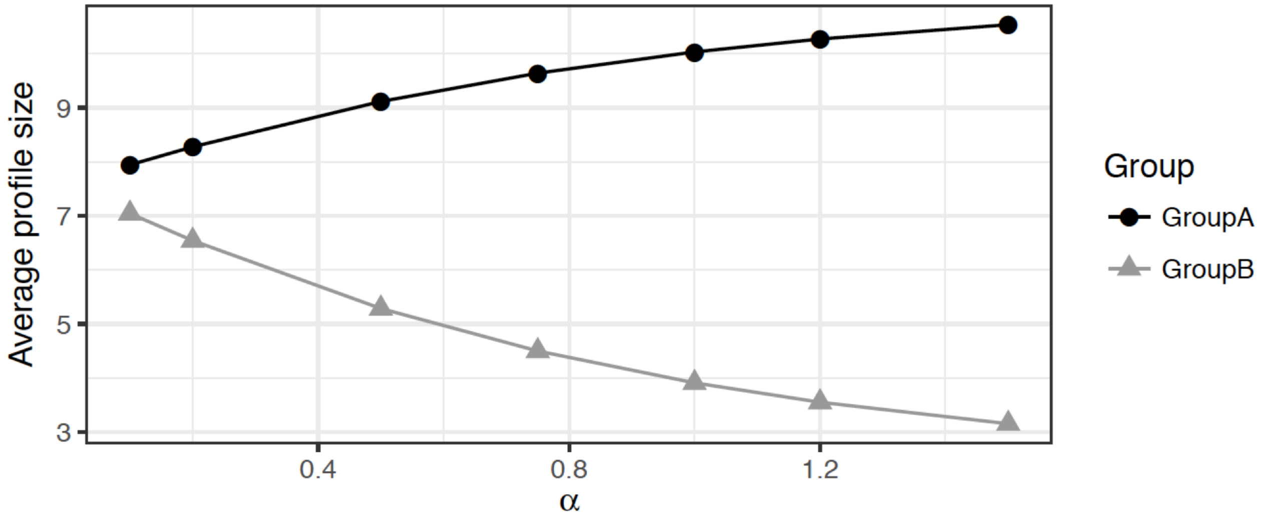
$\alpha = 1.45$

$\beta = 0.4$



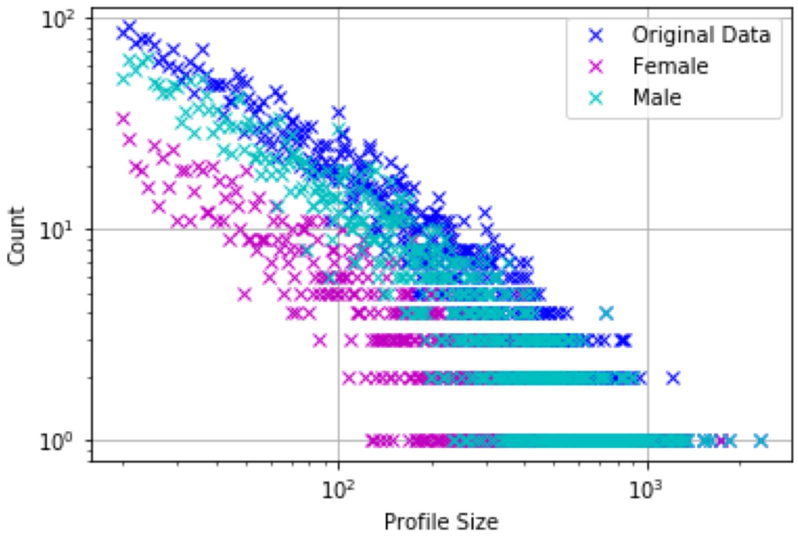
Legal values of α for β =0.4

As α increases, the behaviors of the two groups become increasingly different.



MovieLens 1M Dataset - User Attribute

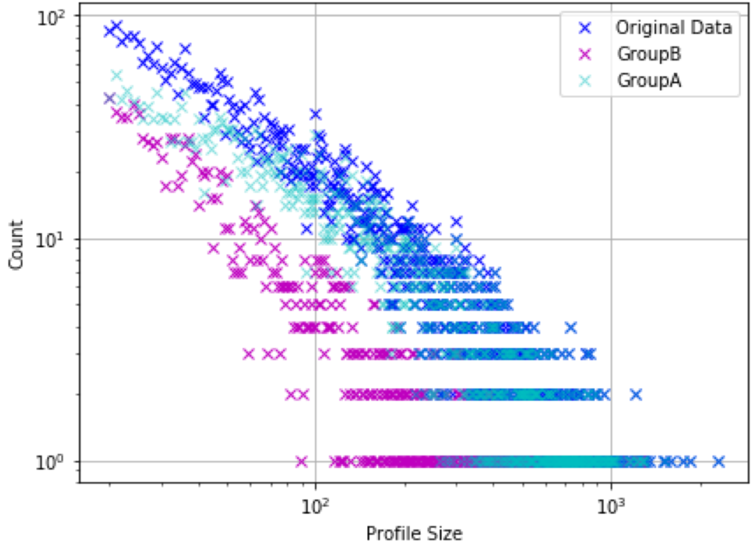
Sensitive attribute is **gender**



1709 females vs. 4331 males

Synthetic attribute

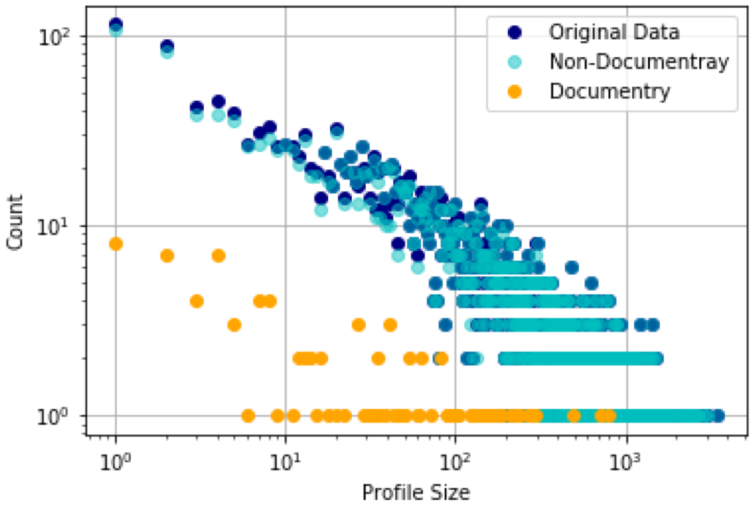
$\alpha = 0.23$, $\beta = 0.34$



1468 group B vs. 4592 in group A

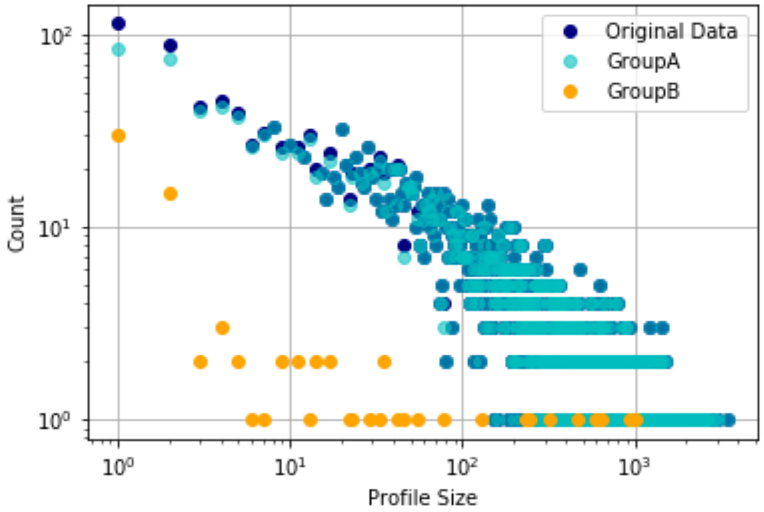
MovieLens 1M Dataset - Item Attribute

Sensitive attribute is **genre**



Synthetic attribute

α = 0.3, β = 0.1



110 documentary vs. 3706 non-documentary

Conclusion

Tradeoffs

- Benefits
  - No mapping to real demographics (just A/B labels)
  - Can adjust population characteristics to test the limits of fairness-aware algorithms
- Drawbacks
  - Correlations with other behavior traits not captured
    - By design!
- We believe this is a good compromise between enabling FATRec research and protected user anonymity

External validity

- Do demographic attributes follow the type of behavior distribution we assume?
    - Which attributes?
    - Which domains?
- Do results over FLAG-assigned attributes translate to real-world cases?
    - Real demographic attributes have correlations with other profile properties – ours may not

Fully synthetic data

- Instead of augmenting existing data
  - Compute new data set with characteristics similar to known data
  - Methodology used in social sciences
- Approaches
  - Borrow from context-aware recommendation: DataGenCARS (Rodríguez-Hernández, et al. 2017)
  - Bipartite ERGM (statnet)
  - Other ideas?

Results

Suggestions?

Data

Fairness

Ethical Issues

Limitations

Algorithm

Questions?