Assessing Evaluation Bias amongst readers towards ardently favored/despised authors

Naman Sharma

Motivation

The key motivation behind this project is to provide some mathematical support to the arguments of a “cult-writer”, or “hated by the critics writer” etc.

Data Details:

Amazon Review Dataset (Julian McAuley et al) - Books
Total Size: 50Mn reviews, for 2.5 Mn books on Amazon up to 2018

Utilized for project: sample @ 100K reviews
Why? Only considered books with 50 reviews, done by reviewers with who’ve written 50 reviews, aka ‘50-core’

Task 1

Highlight Ardent Promoters and Detractors for each work (author-book)

Metric:

Reviewers that rated highly positive or very negative for more than 80% on a particular book.

How:

- K-Means clusters for items enough data available for 3 distinct clusters per book per author

- Measure the changes in ratings upon removal of the highlighted book-reviewer combo

Approach and Setup

Author	Book	User	Rating
XYZ	Abridged Bird Law	Deandra	1.5
XYZ	Abridged Bird Law	Frank	5
XYZ	Abridged Bird Law	Charlie	4.5
XYZ	Abridged Bird Law	Mac	1.5
XYZ	Denim Chicken and other recipes	Dennis	5
XYZ	Denim Chicken and other recipes	Mac	3
XYZ	Denim Chicken and other recipes	Charlie	2
XYZ	Denim Chicken and other recipes	Deandra	4.5
XYZ	Denim Chicken and other recipes	Cricket	3
XYZ	Denim Chicken and other recipes	Frank	5

Approach and Setup

Author	Book	User	Rating
XYZ	**Abridged Bird Law**	**Deandra**	**1.5**
XYZ	**Abridged Bird Law**	**Frank**	**5**
XYZ	Abridged Bird Law	Charlie	3.5
XYZ	Abridged Bird Law	Mac	4.5
XYZ	Denim Chicken and other dreams	Dennis	5
XYZ	Denim Chicken and other dreams	Mac	3.5
XYZ	Denim Chicken and other dreams	Charlie	4
XYZ	**Denim Chicken and other dreams**	**Deandra**	**2**
XYZ	Denim Chicken and other dreams	Cricket	5
XYZ	**Denim Chicken and other dreams**	**Frank**	**5**

Approach and Setup

Author	Book	User	Rating
~~XYZ~~	~~Abridged Bird Law~~	~~Deandra~~	~~1.5~~
~~XYZ~~	~~Abridged Bird Law~~	~~Frank~~	~~5~~
XYZ	Abridged Bird Law	Charlie	3.5
XYZ	Abridged Bird Law	Mac	4.5
XYZ	Denim Chicken and other dreams	Dennis	5
XYZ	Denim Chicken and other dreams	Mac	3.5
XYZ	Denim Chicken and other dreams	Charlie	4
~~XYZ~~	~~Denim Chicken and other dreams~~	~~Deandra~~	~~2~~
XYZ	Denim Chicken and other dreams	Cricket	5
~~XYZ~~	~~Denim Chicken and other dreams~~	~~Frank~~	~~5~~

Results 1

- The user clusters were bound by hard coded logic
- 99% of the reviews did NOT fall into the category. Reviewers fluctuated between 0.75-1.25 perc for all samples.
- The lack of significant results probably due to the high number of reviews

User Clusters:

	mean	std	count
Low	1.72	0.45	3.14%
Normal	3.79	0.41	36.66%
High	5.00	0.00	60.19%

Task 2

Evaluate distortions in product recommendations due to inflated/deflated ratings for particular authors and works

Metric:

Distortion =

1-(Recommendation Prior Removal AND Recommendation Post Removal) /Recommendation Prior Removal

How:

Fetch collaborative fill/item-based suggestions before and after dataset cleaning. Higher distortion means worse effects.

Results

Example:

{'0001844423', '0003302245', '0007284241', '0007350961', '0140431179', '0142406635', '045141411X', '0545123267', '1477809732', 'B00A9WRH3M'}

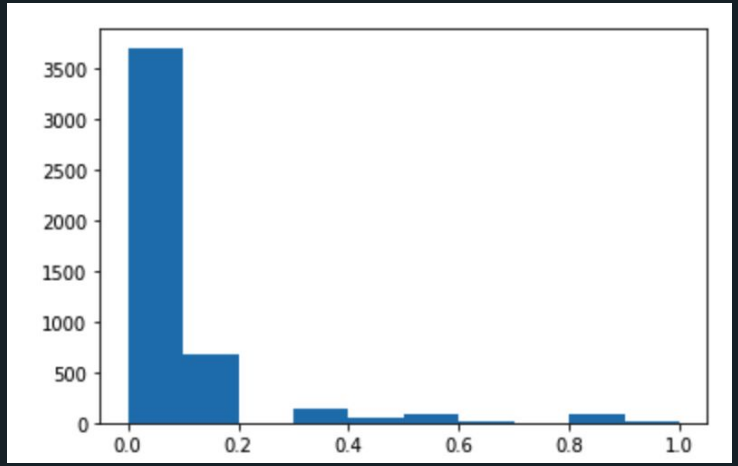
v/s

{'0001048767', '0003302245', '0007284241', '0007350961', '0140431179', '0142406635', '045141411X', '0545123267', '0545123267', 'B00A9WRH3M'}

Distortion = 0.8

Distortion over the dataset:11.3 [SD: 15.4]

Distortion Density:



Future Steps

- Results to be recalculated 5/10/20-core data, i.e., more data, along with categorical focus as well
- Robustness of recommendations should be tested with other learning/stationary models such as kNN, Matrix Factorization/SVD
- De-biasing is a fairly well explored area in the field of recommendations, with existing methods that provide some mathematical guarantees/outlines of effect
- Demographics affected, as well focus on authors that are most affected by the biases