

Get Recced! App Recommendations using Reviews and Ratings

Bhavesh Munot, Christopher Mureekan, Puneet Kohli, Sukhdeep Gill
CSCE 670 - Information Storage and Retrieval (Spring '18)

Motivation:

In today's world of ubiquitous digital content, the number of choices we have are overwhelming. In an effort to understand how recommender systems can make better predictions, we decided to explore the impact of review text on suggesting relevant items to users and helping them discover the content they are looking for. We hypothesized that using review text would produce better recommendations than traditional ratings-only based approaches to recommendation systems.

Dataset:

We used the "Apps for Android" 5-core Dataset which contains both product ratings and reviews. The dataset contains 752,937 entries with 87,271 users and 13,209 products (mobile apps). The data is from Amazon.com and has been collected by Julian McAuley, UCSD.

Methodology:

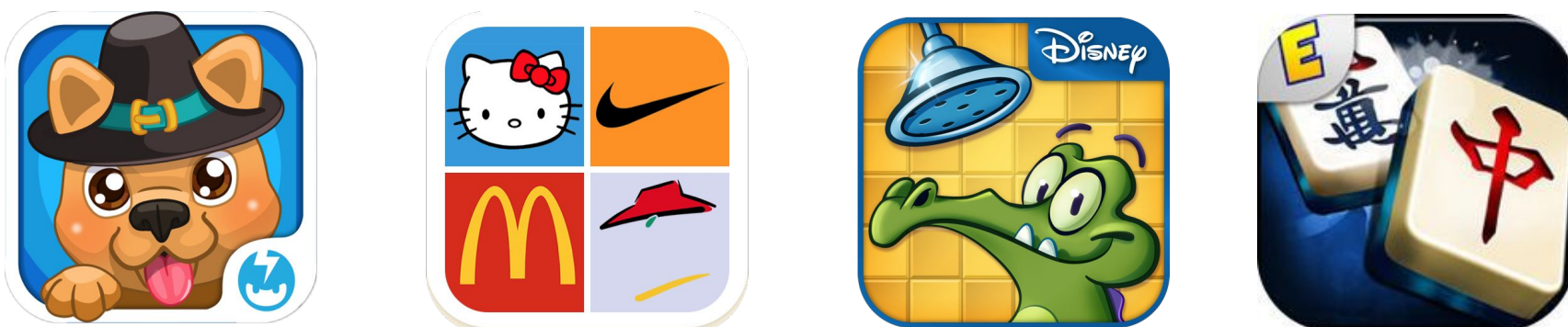
The **Hidden Factors and Hidden Topics** model (HFT) combines latent rating dimensions with latent review topics [1]. The hidden topics allow us to make predictions with high accuracy, even with a few reviews. We implemented the following models and compared their results to that of the HFT model on 2 datasets - The 5-core App dataset and a subset of it with 20k reviews:

- Global Mean Rating
- Mean Rating with Biases
- Collaborative Filtering (CF)
- Singular-Value Decomposition (SVD)
- Latent Dirichlet Allocation (LDA)

Example Recommendation:

For a user who had rated highly on mostly **children's games** as well as reward-based games such as the popular game show **"Wheel of Fortune"**, we were given the below suggestions:-

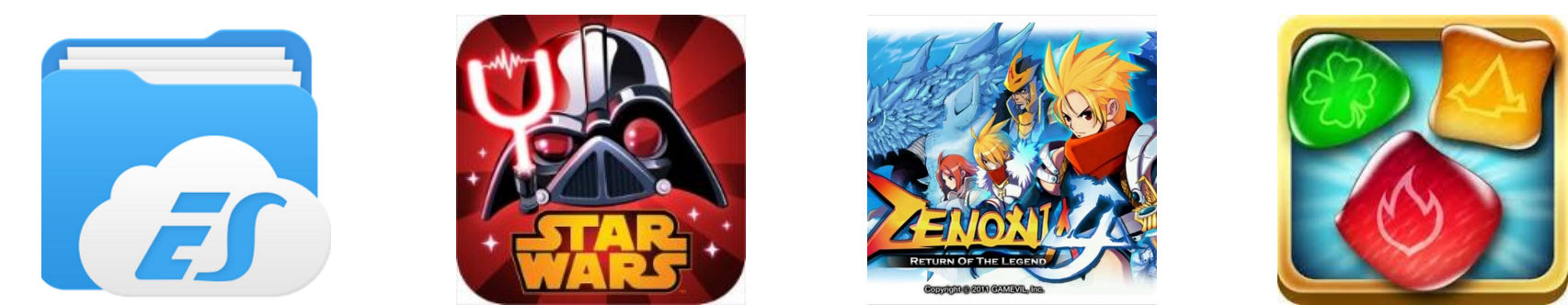
Global Mean Rating:



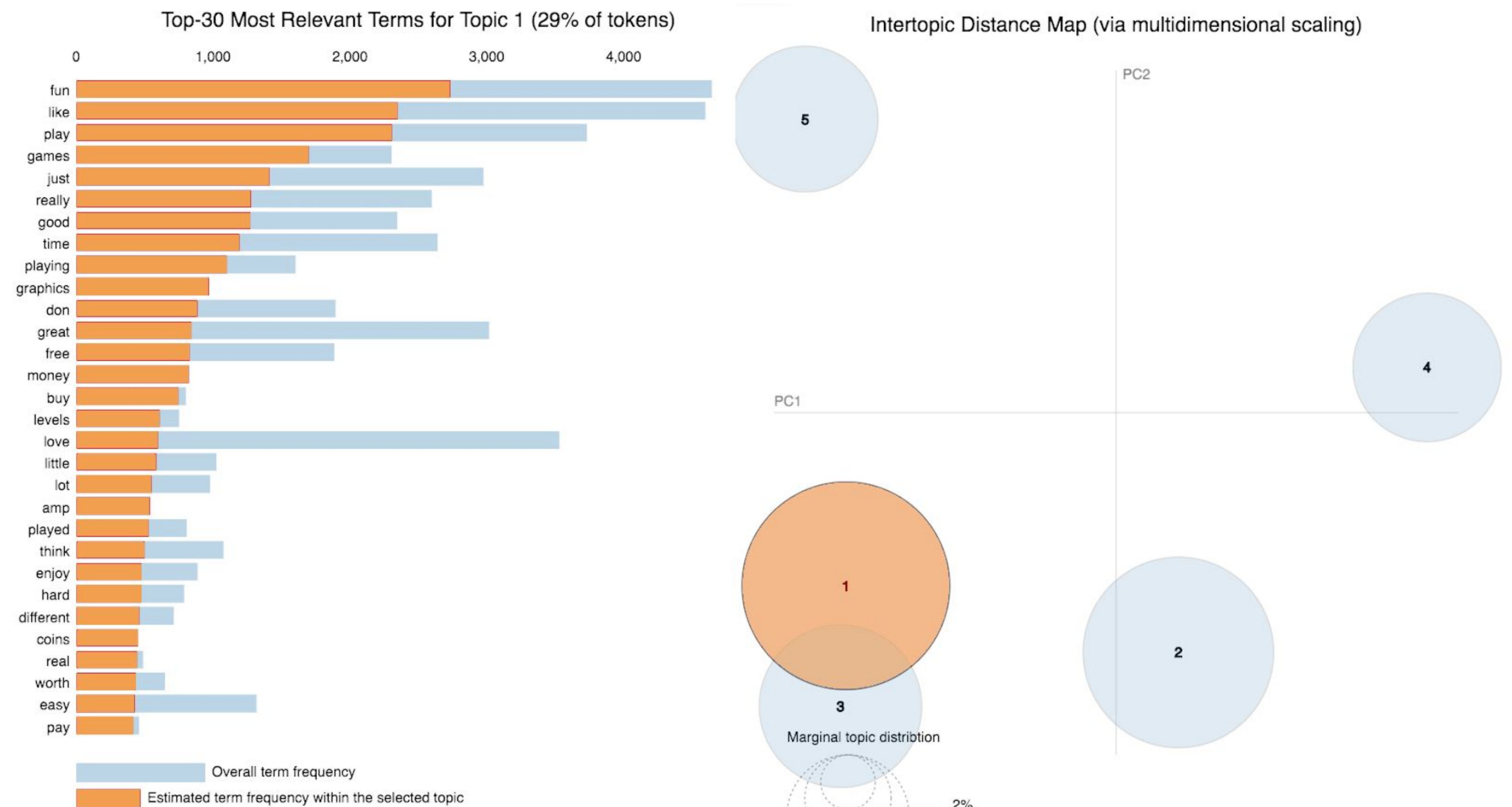
Collaborative Filtering:



Hidden Factors and Hidden Topics:



Topic modeling from review text



Results:

As visible from the table below, using the review text in combination with the numerical ratings hailed better results in terms of both RMSE and MAE. The HFT model performed the best with an RMSE of **1.1** on the whole 750K reviews, and an RMSE of **1.02** on 20K reviews. We did not run the Collaborative Filtering model or Latent Dirichlet Allocation model on the 750K dataset. Additionally, it is interesting to note that most of our baseline methods (using only ratings) performed either better or roughly the same on the larger dataset, whereas the HFT model performed slightly worse.

750K Reviews							
Metric	Mean	Biases	CF	SVD	LDA	HFT	% improve
RMSE	1.32	1.23	-	1.18	-	1.1	6.9%
MAE	1.06	0.93	-	0.91	-	0.85	6.5%
20K Reviews							
RMSE	1.26	1.37	1.33	1.18	1.4	1.02	13.5%
MAE	0.98	1.0	0.93	0.91	1.03	0.76	19.7%

Challenges:

- Creating a train-test split having each user and app from the test set present in the train set.
- LDA Topics are not clearly distinguishable for our dataset
- No ground truth for the predictions by using review texts only
- Creating an automated Amazon ASIN lookup tool which is usually a paid service

Takeaways:

- Using review topics in conjunction with ratings gives better recommendations than traditional methods
- It is difficult to make accurate predictions by relying on review text and biases alone (LDA)
- HFT is dataset dependent - some categories of items have reviews that better express the subjective tastes of the user/properties of the item
- Beyond a certain value, increasing the number of topics gives negligible improvements in RMSE
- App reviews are short and to-the-point, and mostly positive

Potential ethical issues:

- Recommending violent/inappropriate apps to children.
- Profile injection is a possible attack to push or nuke items.

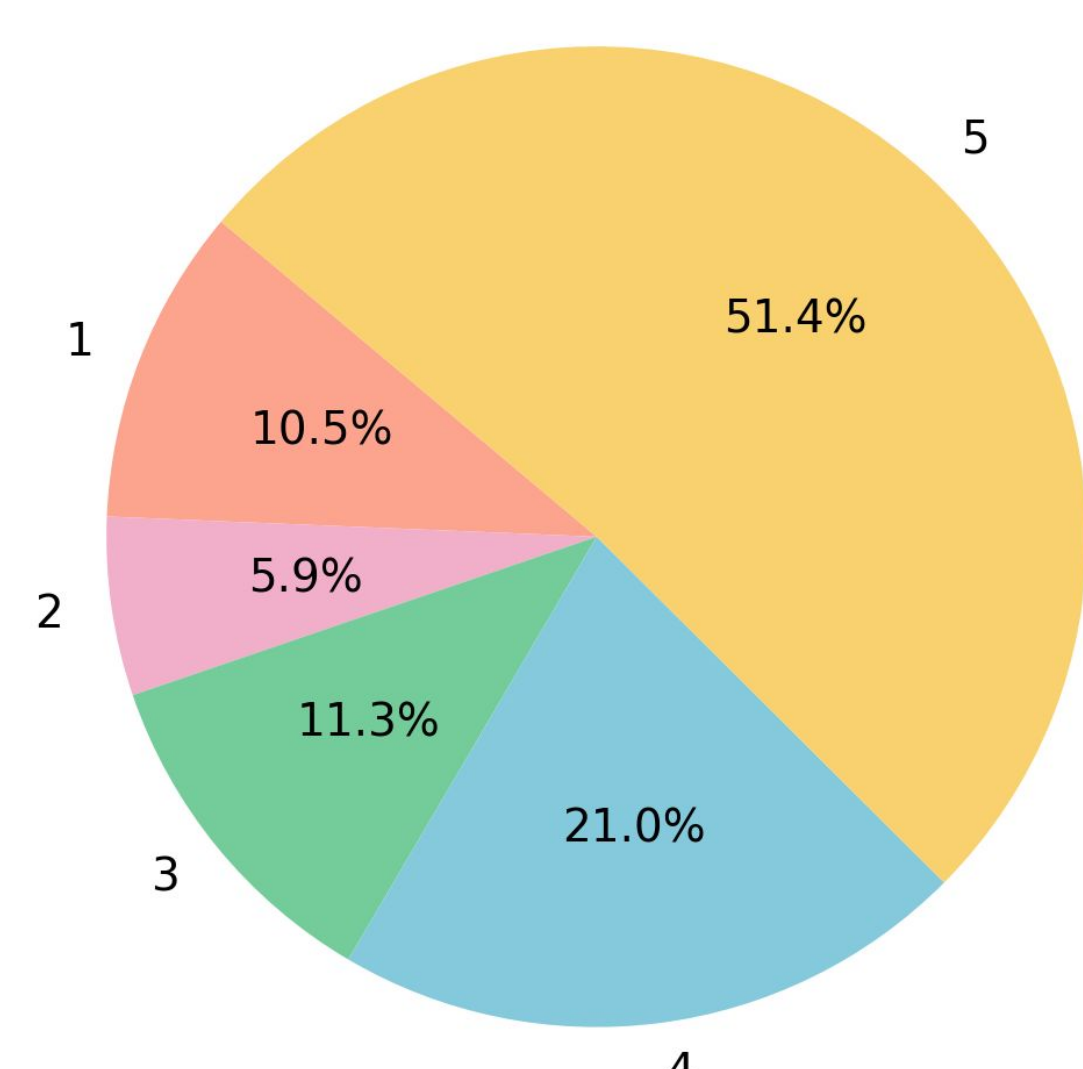
What's next?

- Word2Vec or GloVe can be used as part of hidden topic extraction from reviews
- The HFT model may be adapted to work with implicit data instead of ratings, along with reviews.

References:

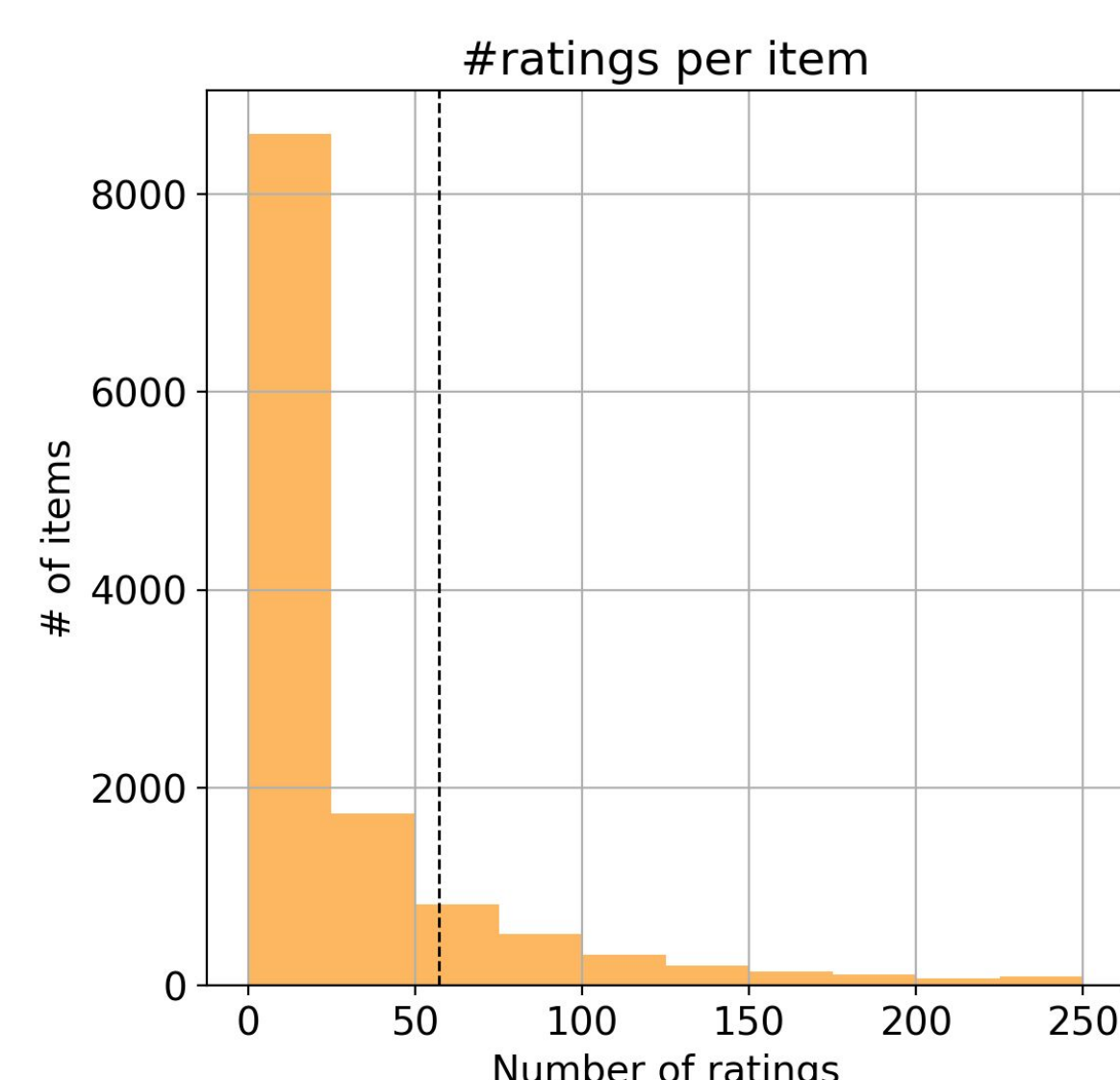
- [1] McAuley, Julian, and Jure Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text." Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013.
- [2] Yu, Kuifeng, et al. "Towards personalized context-aware recommendation by mining context logs through topic models." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2012.

Ratings Distribution



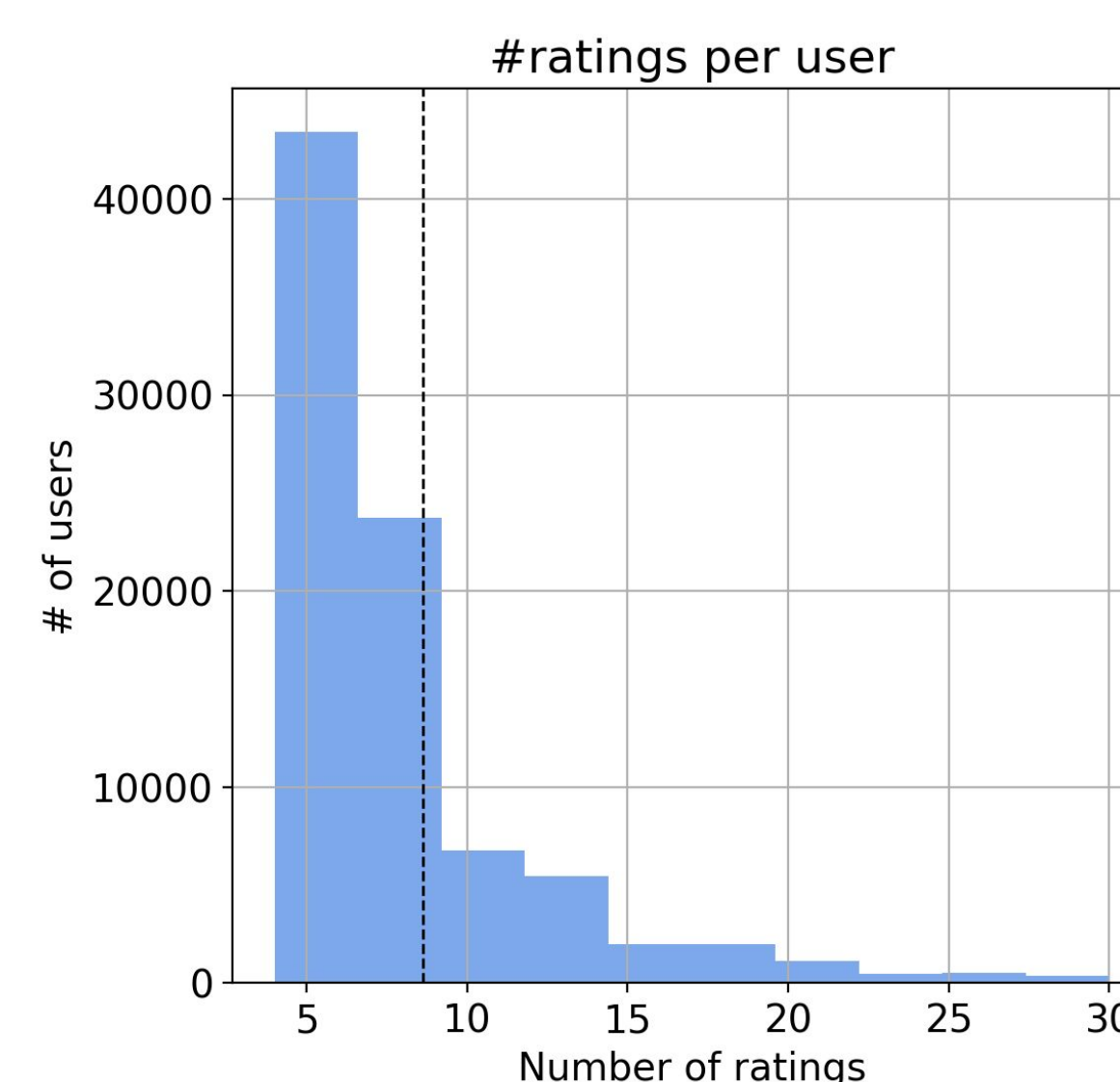
More than 50% of apps have a rating of 5

Ratings/Item



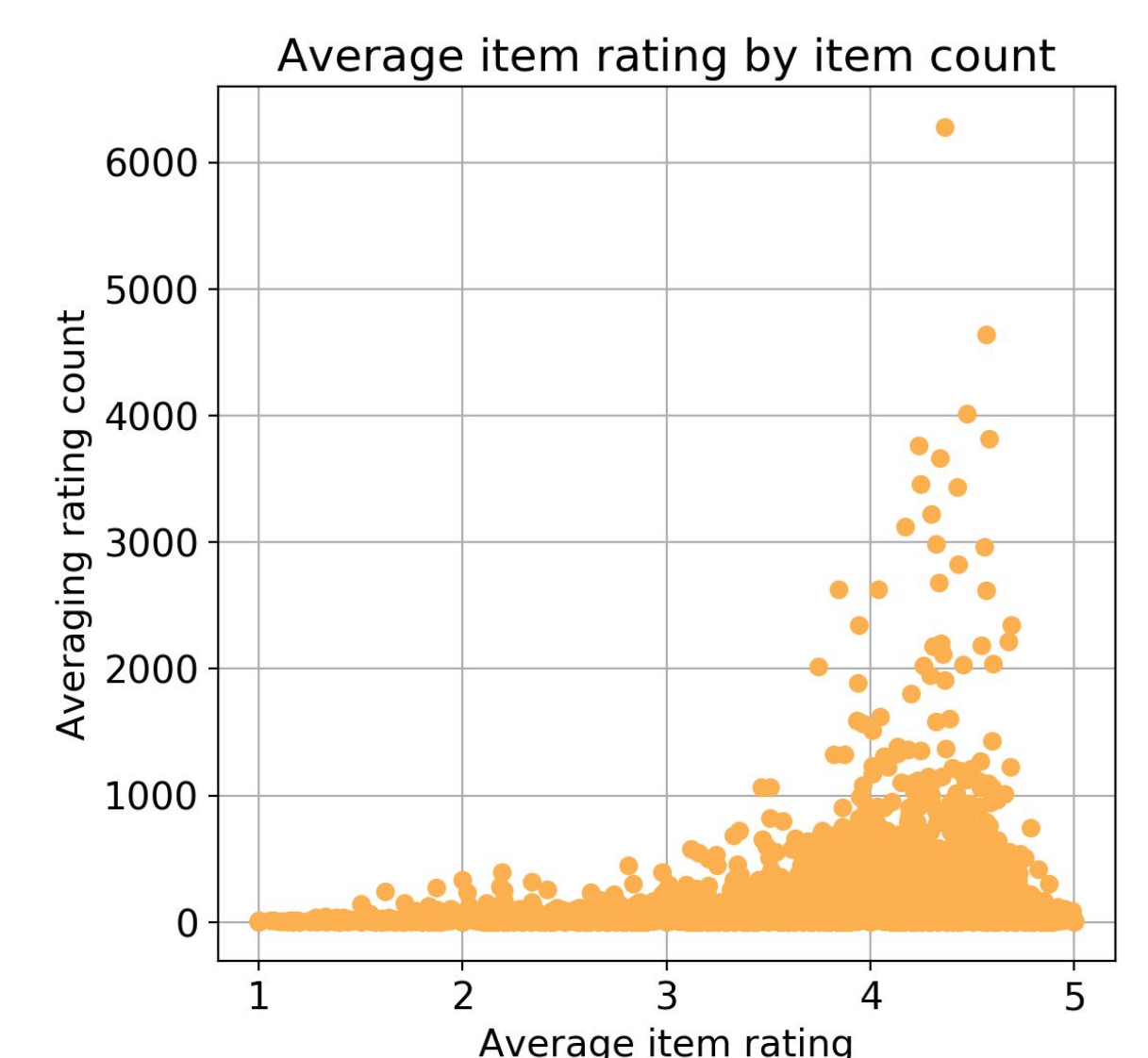
An item on average has 57 ratings/reviews

Ratings/User



A user on average has rated 9 items

Avg rating/count



The mean rating is 3.96 out of 5