Exploring the filter bubble. The effect of Recommender Systems on Content Diversity

Recommender Systems

- A recommender system or a recommendation system is a subclass of information filtering system that seeks to predict the "rating" or "preference" that a user would give to an item.
- More data is reduced into smaller important data, which leads to important information
- Amazon once reported that 35% of its sales came from its recommendation systems.
- Netix in 2012 reported that 75% of what its users watched came from recommendations.

- Do recommender systems expose users to narrower content over time?
- How does the experience of users who take recommendations differ from that of users who do not regularly take recommendations?

Recommender Systems Cont..

- Isolate and measure the effect of accepting recommendations from recommender systems.
- Separate users into categories based on how often they actually consume recommended content.
- This separation lets us focus on users against a control group who use the same system but do not regularly follow recommendations.
- Introduce a method and metric for exploring changes in the diversity of consumed items over time.
- Provide quantitative evidence suggesting that users who take recommendations receive a more positive experience than users who do not.

Filter Bubble

- Eli Pariser termed it as Filter Bubble, describe the potential for online personalization to effectively isolate people from a diversity of viewpoints or content
- a self-reinforcing pattern of narrowing exposure that reduces user creativity, learning, and connection
- Filter bubble effect requires access to a longitudinal dataset and consumption of information items.
- Effect of filter bubble in recommender systems, reduces creativity and learning ability, and strengthens the belief of the user.
- Linden, one of the authors of Amazon's recommender system, suggested that narrowing user choices is not what personalization via recommender systems does.

The effect of Recommender Systems on Content Diversity

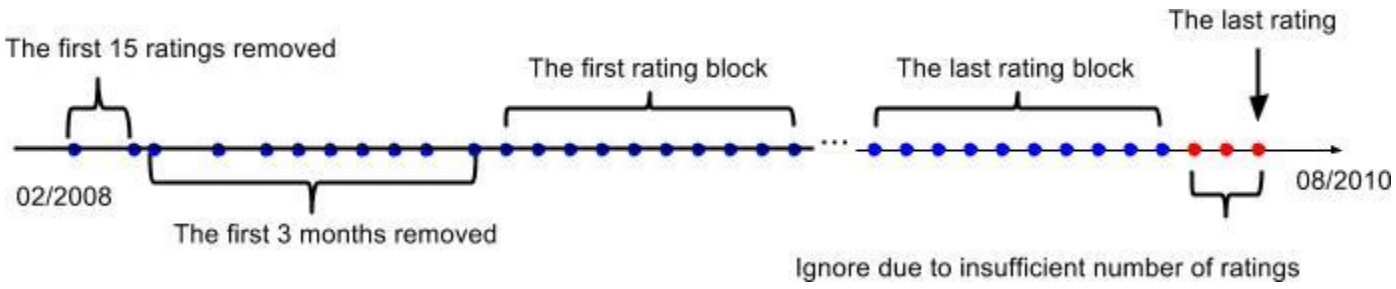
- datasets and discuss our methods for identifying recommendation takers and computing the content diversity of movies.
- Metric for exploring changes in the diversity of consumed items over time.
- we use data from MovieLens. MovieLens is a movie recommender system that has been in continuous use since 1997. As of September 2013.
- The Longitudinal dataset technique was being used in MovieLens

- MovieLens provides a feature called `Top Picks For You‘ that takes users to a page displaying movies the user has not seen, ordered from the highest predicted ratings to the lowest predicted ratings.

Content Diversity

- MovieLens started to log all user access to `Top Picks For You‘ pages and recommended movies with their respective positions in the recommendation lists at the time users accessing the page.
- MovieLens uses an item-item collaborative filtering (CF) algorithm
- An expressive way to compute content diversity.
- MovieLens has provided a feature that allows users to apply tags (words or short phrases) to movies.
- Tagging feature and the tags that MovieLens users have applied, built tag genome to help users navigate and choose movies where all dimensions, one, are the same as those of the compared movie.

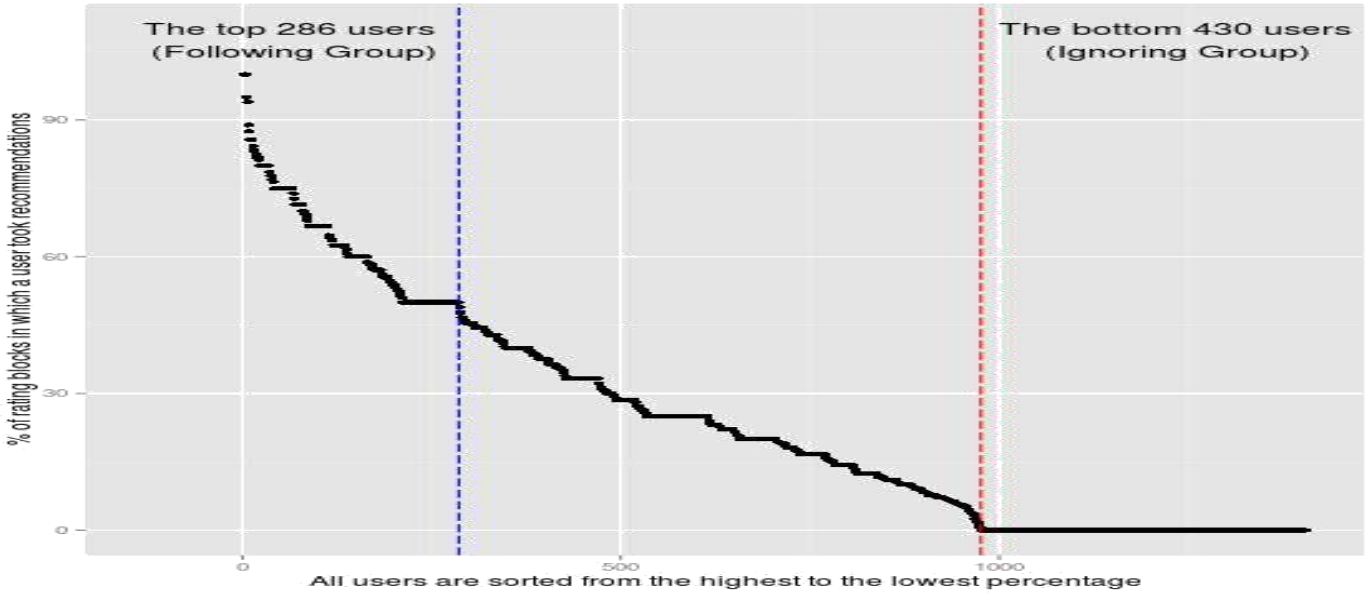
Identifying recommendation takers



- Analyse data in the period from February 2008 to August 2010 (21 months).
- Analyse this period because of missing log data from February 2007 to December 2007 and from October 2010 to May 2012
- To study the effect of `taking' recommendations, we need to classify the users in our dataset into those that do `take‘ recommendations and those that do not.
- To address both of the potential problems described above, we dene an interval as a block consisting of 10 consecutive ratings
- We only select users whose first ratings were in the analysed period (i.e. in the period of February 2008 – August 2010). We include only those users who have three or more ratings blocks in the analysed period.

Ignoring Group vs Following Group

- comparisons between two groups of users - one that consumes recommendations consistently over the time, and one that does not.
- The purpose of the study is to investigate the long term effect of using recommender systems on content diversity.



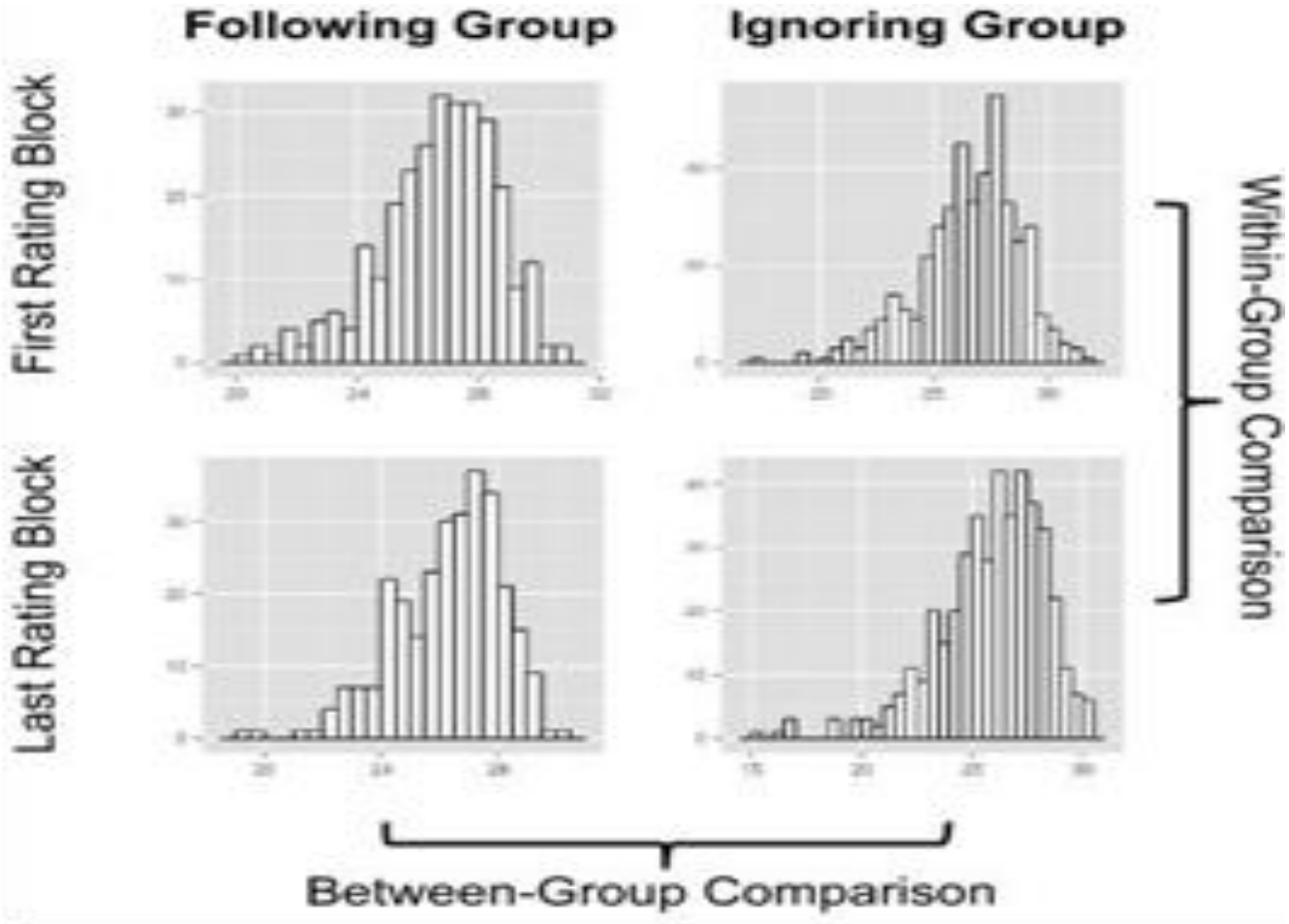
- With these per-user percentages computed, we rank our users from the highest percentage to the lowest percentage.
- the users who took recommendations in all of their rating blocks (i.e. percentage = 100%), are placed on top, those that did not take recommendations in any of their rating blocks (i.e. percentage = 0%) are placed bottom.
- Users who took recommendations in at least 50% of their rating blocks are classified as recommender takers and placed in the Following Group
- the Following Group consists of 286 users, and the Ignoring Group consists of 430 users.

Measuring Content Diversity

- We describe the tag genome data, our method to compute content diversity using tag genome and discuss why we use tag genome.
- $d(mi, mj) = \sqrt{\sum_{k=1}^{m} [rel(tk, mi) - rel(tk, mj)]^2}$
- The tag genome is an information space containing a set M of movies, and a set T of tags.
- Tag t describes movie m is expressed via the relevance score rel (t,m)
- Each movie mi is represented as a vector of size [T] where entry I, j is the relevance of tag j to movie I
- To measure the similarity of two movies, we compute the Euclidean distance between two movie vectors

Tags \ Movies	m_1	m_2	...	m_{n-1}	m_n
t_1	4.3	2.4	...	3.3	1.0
t_2	3.5	1.2	...	2.5	2.1
⋮	⋮	⋮	**rel(t, m)**	⋮	⋮
t_{n-1}	2.0	3.9	...	5.0	4.0
t_n	2.8	1.6	...	2.6	5.0

Group Comparison



- The narrowing effect actually was mitigated for users who appeared to follow the recommender. In other words, taking recommendations lessened the risk of a filter bubble.
- Recommendation-following users received more diverse top-n recommendation lists than non-following users.

Fairness Beyond Disparate Treatment & Disparate Impact
&
Learning Classification without Disparate Mistreatment

Fairness Beyond Disparate Treatment & Disparate Impact

- Anti-discrimination laws in various countries prohibit unfair treatment of individuals based on specific traits, also called sensitive attributes e.g., gender, race.
- These laws typically distinguish between two different notions of unfairness namely, disparate treatment and disparate impact.
- To account for and avoid such unfairness, introduce a notion of unfairness, disparate mistreatment, which is defined in terms of misclassification rates.
- When the ground truth for historical decisions is available, disproportionately beneficial outcomes for certain sensitive attribute value groups can be justified and explained by means of the ground truth.
- Notion of unfairness, disparate mistreatment, especially well-suited for scenarios where ground truth is available for historical decisions used during the training phase.

Fairness Beyond Disparate Treatment & Disparate Impact

	User Attributes		Ground Truth	Classifier's				Disp.	Disp.	Disp.
Sensitive	Non-sensitive		(Has Weapon)	Decision to Stop				Treat.	Imp.	Mist.
Gender	Clothing Bulge	Prox. Crime		C1	C2	C3				
Male 1	1	1	3	1	1	1				
							C_1	7	3	3
Male 2	1	0	3	1	1	0				
Male 3	0	1	7	1	0	1				
							C_2	3	7	3
Female 1	1	1	3	1	0	1				
Female 2	1	0	7	1	1	1				
							C_3	3	7	7
Female 3	0	0	3	0	1	0				

FORMALIZING NOTIONS OF FAIRNESS

		Predicted Label		
		y^ = 1	y^ = 1	
True Label	y = 1	True positive	False negative	P (y^ 6=yjy = 1) False Negative Rate
True	y = 1	False positive	True negative	P (y^ 6=yjy = 1) False Positive Rate
		P (y^ 6=yjy^ = 1) False Discovery Rate	P (y^ 6=yjy^ = 1) False Omission Rate	P (y^ 6=y) Overall Misclass. Rate

FORMALIZING NOTIONS OF FAIRNESS

- Avoiding disparate treatment
- $P(y \mid x, z) = P(y \mid x)$
- Avoiding disparate impact
- $P(\hat{y} = 1 \mid z = 0) = P(\hat{y} = 1 \mid z = 1)$
- Avoiding disparate mistreatment
- Overall misclassification rate (OMR)
- $P(\hat{y} \neq y \mid z = 0) = P(\hat{y} \neq y \mid z = 1)$
- False positive rate (FPR):
- $P(\hat{y} \neq y \mid z = 0, y = -1) = P(\hat{y} \neq y \mid z = 1, y = 1)$

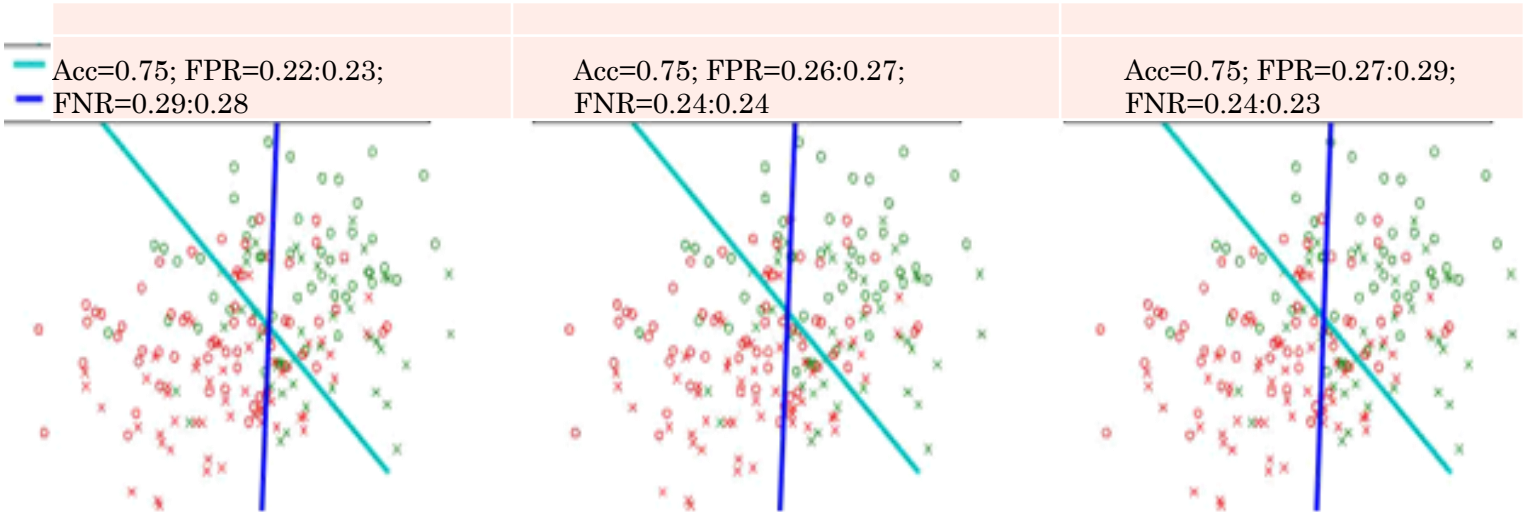
FORMALIZING NOTIONS OF FAIRNESS

- False negative rate (FNR)
- $P(\hat{y} \neq y \mid z = 0, y = 1) = P(\hat{y} \neq y \mid z = 1, y = 1)$
- False omission rate (FOR)
- $P(\hat{y} \neq y \mid z = 0; \hat{y} = -1) = P(\hat{y} \neq y \mid z = 1; \hat{y} = 1)$
- False discovery rates (FDR)
- $P(\hat{y} \neq y \mid z = 0, \hat{y} = 1) = P(\hat{y} \neq y \mid z = 1; \hat{y} = 1)$

CLASSIFIERS WITHOUT DISPARATE MISTREATMENT

- $L(\theta)\ P(y \neq y/z = 0) - P(y \neq y/z = 1) \leq £$

  $L(\theta) P(y \neq y/z = 0) - P(y \neq y/z = 1) \geq £$



(a) FPR constraints (b) FNR constraints (c) Both constraints

DISCUSSION AND FUTUREWORK

- Allows to avoid disparate mistreatment and disparate treatment simultaneously.
- Finally, we would like to point out that the current formulation of fairness constraints may suffer from the following limitations.
- The proposed formulation to train fair classifiers is not a convex program, but a disciplined convex concave program which can be efficiently solved using heuristic-based methods
- This approximation is expected to work well when a reasonable amount of training data is provided, it might be inaccurate for smaller datasets.